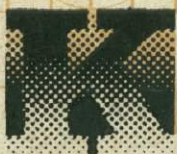
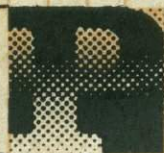


Е. К. ВАСИЛЬЕВ
М. С. НАХМАНСОН



качественный



Рентгенофазовый



Анализ



ИЗДАТЕЛЬСТВО «НАУКА»
СИБИРСКОЕ ОТДЕЛЕНИЕ

АКАДЕМИЯ НАУК СССР
СИБИРСКОЕ ОТДЕЛЕНИЕ
ИНСТИТУТ ЗЕМНОЙ КОРЫ

Е. К. ВАСИЛЬЕВ М. С. НАХМАНСОН

КАЧЕСТВЕННЫЙ РЕНТГЕНОФАЗОВЫЙ АНАЛИЗ

Ответственный редактор
д-р геол.-мин. наук *С. Б. Брандт*

4537



НОВОСИБИРСК
ИЗДАТЕЛЬСТВО «НАУКА»
СИБИРСКОЕ ОТДЕЛЕНИЕ
1986



Васильев Е. К., Нахмансон М. С. Качественный рентгенофазовый анализ.— Новосибирск: Наука, 1986.

Впервые обобщается обширный материал, полученный в результате изучения и анализа ряда систем автоматической идентификации, приведены материалы исследований авторов в области создания оригинальной системы автоматического рентгенофазового анализа, испытаний и тестирования системы, изучения качества информации, содержащейся в базе эталонных рентгенограмм. Рассматриваются математические основы качественного рентгенофазового анализа. Анализируются структура картотек эталонных рентгенограмм и их использование для решения задачи идентификации ручными методами.

Монография рассчитана на геологов, физиков, химиков, материаловедов и металлургов, а также студентов соответствующих специальностей.

Рецензенты *Г. Д. Феоктистов, А. А. Кашаев*

ПРЕДИСЛОВИЕ

Рентгенографический анализ поликристаллов является одним из основных методов исследования состава и структуры твердотельных соединений. В ряде случаев он дает уникальную информацию о фазовом составе и строении веществ, которая не может быть получена с помощью других аналитических методов. В последние полтора десятилетия значительно возросло практическое использование рентгенографии поликристаллов, которая постепенно становится методом производственного анализа руд, сплавов, синтетических материалов, полупроводников и изделий машиностроительной, металлургической, керамической, цементной и химической отраслей промышленности. Это сравнительно новое применение метода рентгенографического анализа, как и его традиционное использование в физике, геологии, неорганической химии и материаловедении, базируется на массовом внедрении в СССР и за рубежом автоматизированных дифрактометров, сопряженных с мини- и микроЭВМ.

Появление автоматизированных приборов, в которых имеется управляющая и обрабатывающая информация ЭВМ, предъявляет новые, более высокие требования к методикам исследования и интерпретации экспериментальных данных. Главное в них — четкая формализация методик и представление их в виде математических алгоритмов, которые могут быть реализованы на ЭВМ. Экспериментальные исследования в рентгенографии поликристаллов, как и во многих других аналитических методах, непосредственно не дают интересующей исследователя информации о составе, структуре и свойствах изучаемых объектов. Эта информация получается в результате обработки экспериментальных данных, причем в некоторых случаях, даже при условии применения ЭВМ, подобная обработка требует больших усилий и материальных затрат.

Особенность рентгенографического анализа заключается в многоцелевой направленности данного метода, позволяющего осуществлять такие виды исследований:

1. Определение параметров кристаллической решетки, которое сводится к индентификации рентгенограмм, вычислению и прецизионному уточнению параметров элементарной ячейки.

2. Определение фазового состава веществ или качественный рентгенофазовый анализ (РФА), осуществляемый путем идентификации рентгенограммы изучаемого соединения (смеси) по картотеке стандартных рентгенограмм.

3. Количественный рентгенофазовый анализ, включающий калибровку коэффициентов в уравнениях связи концентраций с измеренными интенсивностями аналитических линий и вычисление концентраций, составляющих исследуемый образец фаз.

4. Определение макроискажений (напряжений первого рода) по сдвигу рентгеновских линий.

5. Определение параметров тонкой структуры (субструктуры) — областей когерентного рассеяния и микронискажений (напряжений второго рода), которое производится путем анализа уширений профилей рентгеновских линий.

6. Текстуриный анализ — изучение анизотропии кристаллитов, возникающей в поликристаллических объектах (металлы, сплавы, горные породы) в процессе их деформации, термообработки, фазовых превращений и т. п., по характеру распределения интенсивности в дифракционных линиях рентгенограммы.

7. Полнопрофильный анализ, используемый для получения информации от атомной структуры соединений, т. е. для решения круга проблем, которые традиционно считались прерогативой рентгеноструктурного метода исследования монокристаллов.

Хотя решение перечисленных задач предъявляет некоторые специфические требования к этапу экспериментальных исследований, осуществляемых с помощью автоматизированного дифрактометра, в целом он сводится к получению профилей всех или некоторой части линий рентгенограммы путем сканирования по углу 2θ . Что касается методов интерпретации, то различия между ними очень значительны.

К настоящему времени в СССР и за рубежом опубликован ряд монографий, в которых рассматриваются все основные методы рентгенографического анализа поликристаллов, кроме автоматического качественного РФА и полнопрофильного анализа. Так, индцированию рентгенограмм посвящены подробные работы [44, 47], количественный РФА описан в [31], методы исследования субструктуры материалов рассмотрены в книгах [19, 80], текстуриный анализ изложен в монографии [5]. Необходимо также отметить пособия [63, 64], в которых излагаются практические методики индцирования, определения параметров элементарных ячеек, количественного и качественного РФА. Данная книга, посвященная автоматическому качественному РФА, продолжает описание основных методов рентгенографического анализа поликристаллов и послужит необходимым логическим дополнением к перечисленным работам.

Особенностью качественного РФА является необходимость обработки больших массивов справочной информации (картотек стандартных рентгенограмм), поэтому эффективность данного метода в научно-исследовательских и промышленных лабораториях определяется, прежде всего, применением ЭВМ. За рубежом первые системы автоматической идентификации появились в середине 60-х годов и к настоящему времени их насчитывается около двух десятков. В СССР, Чехословакии, Болгарии и ГДР аналогичные работы ведутся уже более 10 лет. Значительный опыт, накопленный при создании и эксплуатации систем автоматического качественного РФА, нуждается в теоретическом осмыслении, обобщении и оценке. Данную задачу, являющуюся основной, авторы попытались решить в этой книге.

Хотя первые системы автоматического качественного РФА появились в СССР и Чехословакии на 7—10 лет позже, чем в США, они не свелись к копированию таких известных программ идентификации, как программы Джонсона — Ванда, Фривела, Никольса, или к созданию каких-либо равноценных программ на несколько иной алгоритмической основе. Фактически в работах ученых СССР и Чехословакии разрабатывается общая математическая теория экспрессной и достоверной идентификации, из которой, как частные случаи, могут быть получены используемые сейчас алгоритмы автоматического качественного РФА. Основой данной теории является предложенная Фиалой в 1972 г. вторичная модель рентгендифракционного спектра. В рамках этой модели советскими специалистами в 1977—1983 гг. были разработаны такие важные методические элементы идентификации, как сжатие рентгендифракционных данных и представление их в двоичных шкалах, применение метода регуляризации и байесовского подхода на заключительном этапе анализа, понятие об идентифицирующей функции и способах ее исследования, структурирование картотеки стандартных рентгенограмм и оценка их качества, представление разнообразных критериев сравнения рентгенограмм в виде приближений универсальной меры близости векторов-спектров в векторном пространстве большой размерности, понятие о некорректности процедуры идентификации и т. д. Авторы надеются, что опубликование данной книги не только отметит приоритет советских и чехословацких ученых в ряде важных вопросов проблемы автоматической идентификации, но и привлечет внимание специалистов всех стран, работающих в этой области, к новым перспективным направлениям исследований.

Проблема идентификации исключительно широка и не сводится только к задаче автоматического качественного РФА. Эта проблема является одной из центральных в методах обработки на ЭВМ физико-химической информации. Задачи, аналогичные качественному РФА, возникают и успешно решаются в масс-спектрометрии, ультрафиолетовой, инфракрасной и Оже-спектроскопии, хроматографии и других областях анализа. В этой связи необходимо отметить работы [24, 75] по созданию уникальной системы идентификации молекулярных спектров, выполненные под руководством академика В. А. Коптюга, и публикации, посвященные анализу состава соединений с помощью масс-спектрометрических данных, ИК- и Оже-спектров [26, 27, 72]. По-видимому, на современном этапе развития рассматриваемая проблема уже требует консолидации усилий специалистов смежных областей для разработки наиболее оптимальных подходов к ее решению.

Важнейшая составная часть проблемы идентификации — формирование и хранение на машинных носителях картотеки стандартов (машинной базы стандартных данных), а также аттестация приводимых данных как общепринятых стандартов. Эти задачи, вообще говоря, выходят за рамки проблемы идентификации, поскольку база стандартных данных может использоваться не толь-

ко для качественного РФА. Прежде всего машинная база данных представляет собой хранилище важнейшей физико-химической информации, быстрейший доступ к которой возможен только с помощью ЭВМ. База данных является основой для выполнения широкого круга информационно-поисковых работ и реализации различных методик обработки экспериментальных данных. Если касаться только рентгенографии поликристаллов, то следует отметить, что представительная и достаточно полная база данных позволит решить задачу количественного РФА и в некоторых случаях значительно облегчит решение таких сложных задач, как индексирование рентгенограмм (с помощью подыскания аналогов в базе данных) и определение субструктуры (путем введения в базу набора эталонных линий, представляющих аппаратурную функцию прибора).

В свете изложенных выше соображений становится ясным, что внимание специалистов должно акцентироваться не только на проблеме идентификации, но и на исследовании базы данных с целью выяснения действующих в ней законов распределения информационных признаков, оценки качества данных, исключения недостоверной информации. В главах 5 и 6 приведены некоторые результаты оригинальных исследований, посвященных перечисленным вопросам.

Необходимо подчеркнуть, что идентификация является математически некорректной задачей, поскольку при ее решении исследователь вынужден оперировать физическими данными (спектрами), содержащими случайную ошибку измерения. Читатель должен помнить, что применение изложенных в книге методов ручного и автоматического РФА не всегда гарантирует решение задачи. Даже при использовании наиболее совершенных систем автоматической идентификации получение надежных результатов во многом зависит пока от искусства исследователя и наличия дополнительной информации о химическом составе, структуре и других свойствах изучаемых объектов.

Немаловажная задача книги — установление терминологии, достаточно понятной и удобной для всех специалистов, работающих в данной области. Естественная основа для подобной терминологии — математика; авторы, однако, в максимальной степени стремились не нарушать уже частично сложившегося смысла терминов, понятных физикам, химикам и материаловедам.

Глава 1, разделы 6.1 и 6.2 написаны авторами совместно. Главы 2 и 3 написаны Е. К. Васильевым, а 4, 5 и раздел 6.3 — М. С. Нахмансоном.

Авторы считают своим приятным долгом выразить благодарность ответственному редактору книги д-ру геол.-мин. наук С. В. Брандту, профессорам М. М. Уманскому и В. А. Франк-Каменецкому, а также Ю. А. Черному, В. К. Румянцевой, З. Ф. Ущаповской, Е. Д. Консону и Б. М. Щедрину за большую помощь при подготовке материалов этой книги и ценное обсуждение полученных результатов.

**ОСНОВЫ КАЧЕСТВЕННОГО
РЕНТГЕНОФАЗОВОГО АНАЛИЗА**

Информацию об элементном составе изучаемых объектов (горных пород, минералов, химических соединений, сплавов и т. д.) можно получить с помощью разнообразных аналитических методов. Однако лишь дифракционные методы (рентгеновский, нейтронографический или электронографический) обладают уникальной возможностью давать характеристику кристаллическим фазам. Наибольшее распространение на практике получил рентгеновский порошковый метод, который в основном используется в рентгенофазовом анализе. Вопросы, касающиеся рентгенофазового анализа как метода идентификации кристаллических фаз, рассматриваются отчасти в монографической литературе (различные руководства и учебные пособия) [1, 3, 21, 33, 38, 44, 49, 62—65, 77, 80, 84, 91, 114, 127], а также в статьях и обзорах [17, 76—78, 99, 104, 105, 169]. В этой главе излагаются основные принципы качественного рентгенофазового анализа, которые необходимы для правильного понимания и использования материала, относящегося к справочным стандартным рентгендифракционным данным, и рассматриваются способы организации справочных данных.

Авторы поставили себе задачу максимально подробного описания структуры и состава различных картотек и определителей, содержащих используемую для качественного рентгенофазового анализа информацию, современные требования, предъявляемые к ее содержанию, поскольку в отечественной литературе отсутствуют подобные описания и руководства.

**1.1. ПРИНЦИПЫ РЕНТГЕНОФАЗОВОГО АНАЛИЗА
И ИХ ФИЗИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ**

В основе рентгенофазового анализа лежат следующие принципы:

1) порошковая дифракционная картина является специфической характеристикой кристаллического вещества; 2) каждая кристаллическая фаза дает всегда одинаковый дифракционный спектр; 3) рентгендифракционный спектр от смеси индивидуаль-

ных фаз является суперпозицией их дифракционных спектров; 4) по дифракционному спектру смеси возможна количественная оценка соотношения кристаллических фаз, присутствующих в изучаемом образце.

В результате проведения исследований на дифрактометре может быть получена дифрактограмма (рентгендифракционный спектр) в виде заданной по точкам спектральной функции, выведенной на цифропечать, перфоленту, или в виде графического изображения этой функции на ленте самописца. Относительно этой информации далее будет употребляться термин «рентгендифракционный спектр»* вместо обычно используемых терминов «дифрактограмма», «порошковая рентгенограмма», «порошкограмма» и «дебаграмма». Использование термина «рентгендифракционный спектр» представляется более правильным, поскольку он отражает физическую специфику получения спектра и указывает, что дифрактограмма относится к информации спектрального типа. Тем самым подчеркивается информационная общность дифракционного спектра со спектральной информацией иной физической природы (оптическими, ИК-, УФ-, масс-спектрами, хроматограммами и т. д.), которая выражается в сходности процедур предварительной и окончательной обработки экспериментальных данных подобного типа. Фрагмент рентгендифракционного спектра представлен на рис. 1.1, а.

Мы будем также использовать в качестве единицы измерения значений межплоскостных расстояний и параметров элементарных ячеек внесистемную единицу — ангстрем ($1 \text{ \AA} = 0,1 \text{ нм}$).

Использование этой единицы измерения обусловлено тем, что 1) в ангстремах указаны величины дифракционных спектров справочных стандартов во всех базах данных; 2) в публикациях Международного союза кристаллографов такие данные приводятся в ангстремах; 3) в подавляющем большинстве справочников и аналогичных им изданий по физике, химии и кристаллографии величины атомного и субатомного порядка определены также в ангст-

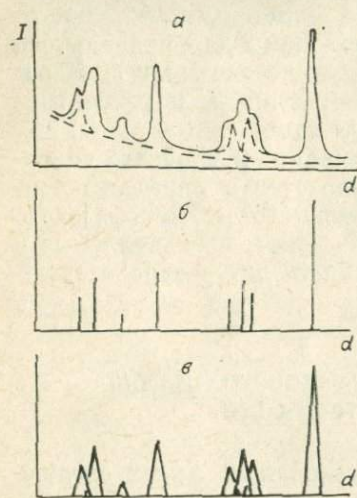


Рис. 1.1. Рентгендифракционный спектр. а — реальный спектр (штриховой линией показаны уровень фона и разделение наложенных линий на компоненты), б — штрих-диаграмма, в — модельное представление линий спектра в виде треугольных распределений.

* Использование термина «рентгендифракционный спектр» впервые было предложено М. М. Уманским.

ремах. Наконец, использование нанометров вместо ангстрем приводит к значительному увеличению объема печатной продукции.

Современные дифрактометры позволяют получить рентгенодифракционный спектр в диапазоне углов 2θ от 6—8 до 140—160° с шагом $\Delta(2\theta)$ 0,01—0,005°. Если считать, что половина диапазона получения спектра относится к областям фона, то информативные области спектра (области линий) будут содержать порядка $7,5 \cdot 10^3$ — $15 \cdot 10^3$ точек. Таким образом, для точного представления рентгенодифракционного спектра, содержащего информацию о форме рентгеновских линий, необходимо хранить массивы, содержащие тысячи чисел (значений интенсивности). Однако для решения конкретных аналитических задач, как правило, нет необходимости в хранении всего рентгенодифракционного спектра или его части в полном представлении, приведенном на рис. 1.1, а. Производится предварительная обработка спектра — отделение фона, определение положений пиков (по их максимумам или центрам тяжести), вычисление интегральных интенсивностей, результатом которой является сжатие спектральной информации примерно на два порядка. Такой сжатый рентгенодифракционный спектр представляется в виде совокупности пар значений $\{2\theta_i, I_i\}$ или $\{d_i, I_i\}$, где $2\theta_i$ или d_i определяет положение i -й линии в шкале углов 2θ или межплоскостных расстояний d , а I_i * есть интегральная интенсивность i -й линии. Графически сжатый спектр представляется в виде штрих-диаграммы (см. рис. 1.1, б).

Величины межплоскостных расстояний d определяются из уравнения Вульфа — Брэггов:

$$d = \lambda/2 \sin \theta, \quad (1.1)$$

где λ — длина волны используемого характеристического излучения, θ — угол скольжения.

Рентгенофазовый анализ по дифракционным спектрам от поликристаллических образцов представляет собой эмпирический метод, включающий количественное (!) сопоставление значений d изучаемого образца с соответствующими значениями d одного или более эмпирических справочных стандартов при качественном (!) сравнении интенсивностей линий в спектрах образца и стандартов. Из этого следует, что для проведения рентгенофазового анализа неизвестного вещества требуется набор надежных справочных стандартов кристаллических фаз, представленных совокупностью величин d и I .

Рассмотрим теперь термин «порошковые рентгенодифракционные данные» или «порошковые рентгенодифракционные данные эмпирического справочного стандарта». Под этим термином понимается совокупность следующих групп сведений об индивидуальной кристаллической фазе (химическом соединении, минерале, простом веществе и т. п.): 1) идентификатор кристаллической фа-

* Чаще всего I_i — значение пиковых интенсивностей.

зы (название, химическая формула, номер по какому-либо каталогу); 2) сведения о физико-химических свойствах фазы (внешний облик, цвет, оптические характеристики, твердость, плотность, и др.) и химическом составе; 3) сведения о структуре фазы (параметры элементарной ячейки, число формульных единиц, модель структуры); 4) сведения об условиях получения и способах обработки рентгendifракционного спектра; 5) рентгendifракционный спектр либо в сжатом виде (в форме штрих-диаграммы), либо в виде таблицы d , I и результаты его индцирования; 6) ссылки на использованную литературу.

Таким образом, порошковые рентгendifракционные данные объединяют рентгendifракционный спектр индивидуальной кристаллической фазы, результаты его обработки (индцирование, внесение поправок на систематические погрешности, сведения о структуре) и необходимую или полезную дополнительную информацию, полученную другими методами, подтверждающими ее индивидуальность, и повышающую надежность рентгendifракционного спектра.

Информация, которая должна быть включена в порошковые рентгendifракционные данные, как правило, неоднозначна. В ней обязательно должен присутствовать идентификатор индивидуальной кристаллической фазы и соответствующий ему рентгendifракционный спектр; что касается остальных сведений, то они могут полностью или частично отсутствовать. Объем информации определяет возможности использования порошковых рентгendifракционных данных. Естественно, чем более полная информация о кристаллической фазе содержится в порошковых рентгendifракционных данных, тем шире возможности их использования (в том числе и для решения информационно-справочных задач).

1.2. СПРАВОЧНЫЕ СТАНДАРТНЫЕ ДАННЫЕ И ПРЕДЪЯВЛЯЕМЫЕ К НИМ ТРЕБОВАНИЯ

Успешное выполнение рентгенофазового анализа, прежде всего, связано с наличием достаточного количества стандартных справочных данных, содержащих наборы d и I . Необходимость создания библиотек справочных данных для этой цели была понята сразу же после открытия порошкового метода, но только к 1938 г. были созданы первые такие библиотеки [5, 55, 125, 161], [16, 26—28], [4, 59, 105, 161] и разработаны рациональные способы проведения рентгенофазового анализа. К этому времени сложилось достаточно четкое представление о том, что основной рентгенометрической характеристикой индивидуальной кристаллической фазы являются межплоскостные расстояния и относительные интенсивности, образующие рентгendifракционный спектр. Выбор этих величин в качестве рентгенографических паспортных данных индивидуальной кристаллической фазы обус-

ловлен тем, что величины d не зависят от условий регистрации рентгendifракционного спектра, а значения интенсивностей I также мало зависимы от длины волны излучения, на котором снят рентгendifракционный спектр [76]. Исходя из этого, принято в качестве паспортной характеристики кристаллической фазы использовать ее рентгendifракционный спектр, т. е. совокупность величин $\{d_j, I_j\}$.

Естественно задать вопрос: каким требованиям должны отвечать эти данные? Наиболее важным требованием к стандартным данным является их высокое качество, т. е. достоверность. Критерий достоверности эмпирических справочных стандартных данных включает следующие требования [131].

1. Данные необходимо получить от известного вещества (кристаллической фазы), состав которого достаточно точно установлен; вещество является однофазным. Элементный состав определен химическим анализом, найдены оптические константы и другие физико-химические характеристики вещества.

2. Рентгendifракционный спектр должен обладать взаимной согласованностью, т. е. параметры элементарной ячейки, полученные из межплоскостных расстояний, обеспечивают согласие между измеренными и вычисленными значениями 2θ в пределах $0,1^\circ$, а линии большой и средней интенсивности разрешаются — расстояние между ними больше $0,2^\circ 2\theta$.

3. Рентгendifракционный спектр должен быть полным и содержать максимальное количество как сильных и средних, так и слабых линий в области углов $2\theta \leq 100^\circ$ для излучения $\text{CuK}\alpha$.

4. Интенсивности линий спектра следует оценить количественно.

5. При получении рентгendifракционного спектра должно быть уделено особое внимание исключению ошибок, вызванных предпочтительной ориентировкой и размером кристаллитов.

К сожалению, рентгеновские спектры веществ, содержащиеся в картотеках и определителях и публикуемые в оригинальных работах, далеко не всегда отвечают этим требованиям. Таким образом, использование порошковых рентгendifракционных данных в качестве справочных стандартов предполагает, что они подвергнуты апробации, которая гарантирует их достаточно высокое качество. Необходимо сразу же отметить, что апробация всех сведений, входящих в состав рассматриваемых данных, выходит за рамки рассматриваемой задачи. Так, существуют разночтения в наименованиях соединений (в частности, минералов), неоднозначности в написании сложных химических формул, неясности в составе некоторых соединений и т. д. Решение этих вопросов относится к компетенции соответствующих отраслей знания, например химии и минералогии. Таким образом, можно говорить об апробации и стандартизации рентгendifракционных спектров и той информации, которая получается из спектра путем его обработки.

В последние 4—5 лет в СССР и за рубежом интенсивно разрабатываются унифицированные методики проведения экспериментальных рентгендифрактометрических исследований и формулируются требования к условиям эксперимента, результаты которого могут быть признаны в качестве стандартных; более подробно этот вопрос будет рассмотрен в разделе 1.3 и главе 3. Следует, однако, отметить, что уже опубликовано несколько десятков тысяч спектров различных соединений, причем условия получения большинства из них далеки от эталонных в современном понимании. Разработка стандартных методик обработки рентгендифракционного спектра (прежде всего предварительной обработки, индцирования и определения параметров элементарной ячейки) только начинается. Таким образом, соответствующие результаты, приводимые в порошковых рентгендифракционных данных кристаллических фаз, используемых в настоящее время в качестве справочных стандартов, далеко не всегда отвечают предъявляемым к стандартам требованиям, тем более еще нет эталонной методики получения этих результатов.

Изложенные выше соображения доказывают, что качество порошковых рентгендифракционных данных пока в значительной степени зависит от квалификации экспериментаторов, получивших спектр, вычислителей, выполнивших расчеты по индцированию и определению параметров элементарной ячейки, и авторов картотек, выбравших из массы имеющейся информации данные того или иного исследователя. Итак, порошковые рентгендифракционные данные нельзя считать в полной мере стандартными; приложение к ним терминов «эталонный» или «стандартный» не выдерживает строгой критики и связано, прежде всего, с опубликованием этих данных в составе некоторого определителя или картотеки. В дальнейшем традиционный термин «стандартные порошковые рентгендифракционные данные» (или просто «стандартные данные») будет применяться с учетом сделанных выше оговорок. В литературе используются синонимы этого термина — «эталонные рентгенографические данные» и «порошковые дифракционные стандарты» (Powder Diffraction Standards).

Способы количественной оценки надежности рентгендифракционных спектров веществ и в особенности эмпирических справочных стандартов будут подробно изложены в одном из разделов главы 3.

1.3. ОПИСАНИЕ ФОРМАТОВ ДЛЯ ПРЕДСТАВЛЕНИЯ ПОРОШКОВЫХ РЕНТГЕНДИФРАКЦИОННЫХ ДАННЫХ СПРАВОЧНЫХ СТАНДАРТОВ

Традиционной формой публикации порошковых рентгендифракционных данных эмпирических справочных стандартов в определителях и картотеках является таблица, как правило, содержащая не только значения d , I и индексы Миллера для каждой линии спектра, но и некоторые дополнительные сведения,

d	$1a$	$1b$	$1c$	$1d$	$7a$			$8a$			$8c$
$1/I_1$	$2a$	$2b$	$2c$	$2d$	$7b$			$8b$			
Rad.	λ	Filter		Dia.	$d, \text{Å}$	$1/I_1$	hkl	$d, \text{Å}$	$1/I_1$	hkl	
Cut off		$1/I_1$									
Ref.					3						
Sys.		S.G.									
a_0	b_0	c_0	A	C							
α	β	γ	Z	D_x	4						
Ref.											
$E\alpha$	$n\omega/\beta$	$E\gamma$	Sign								
2V	D	mp	Color	5							
Ref.											
					6						
											9

Рис. 1.2. Формат карточки JCPDS.

образующие текстовую часть. Таблицы включают сведения, перечисленные в разделе 1.1. Таким образом, содержание таблицы справочного стандарта представляет пользователю определителями или картотеками дополнительную информацию, которая в известной мере повышает достоверность рентгенофазового анализа.

Схема описания справочного стандарта получила название формата. Формат обеспечивает унификацию данных, составляющих таблицу или карточку.

В мировой практике для рентгенофазового анализа широко используется картотека Объединенного комитета дифракционных стандартов (Joint Committee on Powder Diffraction Standards (JCPDS) — International Center for Diffraction Data) [141, 142]. Описание каждой индивидуальной кристаллической фазы в этой картотеке производится в стандартном формате, который дальше называется карточкой JCPDS. Совокупность таких карточек образует ежегодные выпуски (сет) картотеки. Формат карточки JCPDS оказался весьма рациональным, и форматы представления данных в других определителях и картотеках (в том числе и новый международный формат [156]) в целом похожи на него. Поэтому целесообразно вначале подробно рассмотреть структуру и содержание такой карточки.

Формат карточки JCPDS изображен на рис. 1.2. Карточка разделена на несколько полей, содержащих одну или несколько позиций для записи информации, относящейся к описываемой кристаллической фазе и образующей совокупность сведений, ко-

торые вместе с данными о рентгendifракционном спектре получили название «порошковые рентгendifракционные данные эмпирического справочного стандарта».

Перейдем к подробному анализу структуры формата карточки JCPDS.

Поле 1. В позициях $1a$, $1b$, $1c$ содержатся межплоскостные расстояния d , соответствующие трем сильнейшим линиям рентгendifракционного спектра, взятые в области углов $2\theta < 90^\circ$. В позиции $1d$ содержится максимальное значение d , надежное для данного соединения.

Поле 2. В позициях $2a$, $2b$, $2c$, $2d$ содержатся интенсивности линий, положения которых даны в позициях $1a$, $1b$, $1c$, $1d$. Интенсивности в картотеке JCPDS приводятся обычно в относительной стобалльной шкале, причем в сто баллов оценивается интенсивность наиболее сильной линии, $I_1 = 100$. Тогда интенсивности остальных линий описываются по формуле

$$I^{100} = \text{Бл. ц.} \left(\frac{I}{I_1} \cdot 100 \right). \quad (1.2)$$

Здесь Бл. ц. означает операцию округления до ближайшего целого, I , I_1 — исходные интенсивности линий спектра, I^{100} — интенсивность в стобалльной шкале.

Для того чтобы подчеркнуть относительный характер приводимых интенсивностей, в картотеке JCPDS принято обозначение интенсивностей I/I_1 . Поскольку ранее нами было принято для рентгendifракционного спектра обозначение $\{d_i, I_i\}$, необходимо подчеркнуть, что в данном контексте I_1 — не интенсивность первой по положению линии, а интенсивность самой сильной линии. В дальнейшем, чтобы избежать путаницы, примем следующие обозначения. Если линии спектра перечисляются в порядке убывания (или возрастания) d , то это обозначается как

$$\{d_i, I_i\} = d_1 I_1, d_2 I_2, \dots, d_n I_n. \quad (1.3)$$

Если линии спектра перечисляются в порядке убывания (возрастания) интенсивностей, то это обозначается как

$$\{I^{(j)}, d^{(j)}\} = I^{(1)} d^{(1)}, I^{(2)} d^{(2)}, \dots, I^{(n)} d^{(n)}. \quad (1.4)$$

В данном разделе мы сохраним за интенсивностью максимальной линии традиционное обозначение I_1 , поскольку оно приводится в карточке JCPDS; в дальнейшем же интенсивности первой (максимальной), второй, третьей и т. д. линий будут обозначаться $I^{(1)}$, $I^{(2)}$, \dots , $I^{(n)}$.

Поле 3 содержит информацию об условиях получения спектра и ссылку на источник, согласно которому в поле 9 приводится рентгendifракционный спектр. Информация в поле 3 сопровождается следующими обозначениями:

Rad. — источник рентгеновского излучения (материал зеркала анода);

- λ — длина волны рентгеновского излучения, Å;
 Filter — указание на материал фильтра;
 Dia. — диаметр рентгеновской камеры (мм), используемой для получения спектра фотометодом;
 Cut off — наибольшее значение d (Å), достижимое при получении спектра с используемой аппаратурой;
 I/I_1 — метод, используемый для получения или оценки величины интенсивностей;
 Ref. — ссылка.

Поле 4 содержит информацию о структуре соединения и ссылку на источник приведенных в поле данных. Информация в поле 4 сопровождается следующими обозначениями:

- Sys. — сингония;
 S. G. — пространственная группа;
 a_0, b_0, c_0 — линейные параметры элементарной ячейки, Å;
 A, C — осевые коэффициенты: $A = a_0/b_0, C = c_0/b_0$;
 α, β, γ — угловые параметры элементарной ячейки в градусах и десятичных долях градуса или в градусах и минутах;
 Z — число химических формульных единиц на единицу структуры (для элементов Z представляет число атомов на элементарную ячейку, для соединений — число формульных единиц на ячейку);
 D_x — рентгеновская плотность;
 Ref. — ссылка.

Поле 5 содержит ряд физических данных:

- $\epsilon_\alpha, n_{\omega\beta}, \epsilon_\gamma$ — показатели преломления;
 $\text{Sign}, 2V$ — знак оптической оси и угол между оптическими осями;
 D — измеренная плотность;
 mp — точка плавления по шкале Цельсия;
 Color — цвет образца;
 Ref. — ссылка на источник данных, приведенных в поле 5.

Информация в отдельных или во всех позициях может отсутствовать.

Поле 6 может содержать дополнительную информацию о результатах химического анализа, происхождении образца, температуре разложения соединения (DT), температуре возгонки (SP), фазового перехода (TP), значении температуры, при которой получен спектр, критерий качества и другие данные. Если в поле 6 приведена какая-либо информация, то обычно даются ссылки на ее источники.

Поле 7 содержит химическую формулу соединения в позиции 7a и его химическое название по номенклатуре ИЮПАК [35, 43, 58]. Эти данные вместе с номером карточки являются идентификатором соединения и крайне редко отсутствуют в карточке.

Тем не менее возможны ситуации, когда в поле 7 содержатся только химическая формула или только название соединения.

Поле 8 содержит структурную формулу соединения в позиции 8a и его общеупотребительное название в позиции 8b. Напомним, что для некоторых соединений существуют и широко используются общеупотребительные названия, не совпадающие с точными химическими наименованиями (например, каломель — хлорид ртути). Для минералов в этой позиции приводятся их названия. Для синтетических соединений, имеющих химический состав, подобный составу минерала, название минерала дается в круглых скобках.

В позиции 8c (правый верхний угол карточки) указывается признак достоверности приведенного в карточке рентгendifракционного спектра. Так, данные высокой достоверности обозначаются звездочкой (*), средней — символом *i* (эти данные характеризуются наличием индексов Миллера, интенсивности линий спектра оценены достаточно надежно), а низкой — кружком (o). Отсутствие одного из трех указанных выше признаков в позиции 8c означает, что данные по достоверности не аттестованы. Для рентгendifракционных спектров, полученных расчетным путем, в позиции 8c помещается символ C (calculation) в отличие от экспериментальных спектров. Такие карточки называются расчетными (C-карточки JCPDS).

Как подчеркивалось выше, наличие признака высокой достоверности не означает, что соответствующие данные прошли какую-либо стандартную аттестацию. Присвоение признаков достоверности осуществляется комитетом JCPDS, исходя из метода экспертных оценок. Таким образом, эта процедура содержит значительный элемент эмпирики. Как правило, признак высокой достоверности присваивается рентгendifракционным спектрам, полученным высококвалифицированными специалистами в известных лабораториях и на хорошем оборудовании. Следовательно, признак достоверности указывает, прежде всего, на степень доверия, которую комитет JCPDS питает к результатам, полученным конкретными исследователями. Подобное положение сохранится до тех пор, пока определение фазового состава на всех его этапах — пробоподготовка, съемка, интерпретация — будет включать элементы искусства. Опыт, интуиция и искусство исследователя всегда оказывают влияние на точность и достоверность получаемых результатов, поэтому принятая комитетом JCPDS система оценки данных, возможно, не так уже субъективна. Для оценки рентгendifракционных спектров соединения, включаемых в картотеку JCPDS, используются специальные пакеты программ (AIDS) [11]. Они позволяют производить проверку правильности сингонии, выявлять необычные соотношения и связи в кристаллической решетке, сравнивать экспериментальные и вычисленные значения *d*, уточнять индексы Миллера, вычислять критерий достоверности, приводить данные к стандартной установке.

Изложенные выше соображения следует воспринимать не как

критику используемой комитетом JCPDS методики оценки достоверности данных, а как предостережение читателю против слепого доверия к помеченным звездочкой спектрам.

Поле 9 содержит рентгенодифракционный спектр и результаты его индцирования (если индцирование удалось осуществить). Эта информация располагается в таблице из трех колонок под заголовками d , I/I_1 , hkl , причем линии перечисляются в порядке убывания величины d .

Значение интенсивности может сопровождаться дополнительными, отражающими специфические особенности линии, символами: b — широкая диффузная линия, d — двойная, n — отсутствующая в других источниках информации о данном соединении, nc — появление которой не объясняется предполагаемой элементарной ячейкой, ni — не индцирующаяся в элементарной ячейке, приведенной в карточке, β — с не определенной интенсивностью из-за присутствия или наложения β -линий, tr — очень слабая; + — кроме приведенных индексов возможны и другие, не поместившиеся в колонке, указание на множественность индексов hkl . Перед значением интенсивности слабых линий иногда ставится знак $<$; это означает, что наблюдаемая интенсивность данной линии меньше приводимого численного значения.

Индексы Миллера задаются в стандартной форме. Возможно указание нескольких наборов h , k , l к одной паре величин d , I .

Поле 10 выходит за формат карточки и содержит состоящий из двух частей номер карточки по картотеке JCPDS. Первая часть номера указывает на «сет» («набор» или «ящик»), вторая — на номер карточки в сете; сет комплектуется по годам. Номер карточки является основным идентификатором соединения в картотеке JCPDS и всегда присутствует в карточке.

В поле 9 стандартной карточки JCPDS может быть размещена информация о 40 линиях рентгенодифракционного спектра. Если число линий в спектре больше 40 (но не превосходит 80), то за основной карточкой следует карточка-продолжение аналогичного формата, содержащая букву A после номера в поле 10. В карточке-продолжении заполняются только поля 1, 2, 7, 8 и 9; содержащаяся в основной карточке на полях 3, 4, 5, 6 информация в карточке-продолжении не дублируется. В поле 9 карточки-продолжения содержатся линии спектра начиная с 41 и далее. Если рентгенодифракционный спектр содержит более 80 линий, то в конце таблицы, расположенной на поле 9 карточки-продолжения, дается примечание. В нем указывается, сколько еще линий (сверх числа, приведенного в двух карточках) содержится в спектре и до какого предельного значения d эти линии располагаются. Иногда оно имеет место на основной карточке, что обусловлено подобным примечанием в первоисточнике, на основании которого составлена карточка.

Данные, приводимые в ряде позиций карточки, могут помечаться знаком вопроса (?) (например, данные в позициях R_{ad} , S , G ., значения интенсивности и т. д.). Наличие символа « \circ » означает, что приводимые данные являются сомнительными.



Химическое (позиция 7b) и минералогическое названия (позиция 8b) приводятся в карточке JCPDS на английском языке, что создает для читателя дополнительные трудности при работе с картотекой и особенно при пользовании указателями названий соединений и минералов.

В завершение описания структуры порошковых рентгенодифракционных данных справочных стандартов, принятой в JCPDS, дадим в качестве примера вид основной карточки и карточки-продолжения для цейнерита (рис. 1.3). Хотя выше было дано подробное описание формата JCPDS, читателю необходимо учитывать, что в карточках, особенно опубликованных 25—35 лет назад (два первых выпуска вышли из печати в 1945 г.), могут встретиться некоторые отступления от стандартного формата. Так, в пределах полей 3, 4, 5 могут отсутствовать или быть перемещены некоторые позиции, а в полях 6 и 9 содержаться примечания самого разнообразного характера; обычно эти мелкие изменения и добавления не вызывают сложности у читателя. Формат для расчетных карточек JCPDS (помечены символом C — calculation) несколько отличается от стандартного. Для того чтобы лучше уяснить различие между стандартной и расчетной карточками JCPDS, рекомендуется сравнить их между собой и внимательно проанализировать все особенности C-карточки.

В связи с нарастанием в последние годы потока информации по индивидуальным кристаллическим фазам и появлением машинных методов рентгенофазового анализа возникла необходимость пересмотра требований к содержанию формата справочных рентгенодифракционных стандартов. Наиболее полно данный вопрос рассмотрен в работе, представленной Комиссией по кристаллографическим данным Международного союза по кристаллографии (МСК), в которой систематически изложены требования и рекомендации, предъявляемые к порошковым рентгенодифракционным эмпирическим справочным стандартам [126]. Дальнейшее уточнение и конкретизация таких рекомендаций сделаны в работах П. Бэйлиса [88] и Е. К. Васильева и др. [15]. Лишь в конце 70-х годов вопрос был разрешен путем создания форматов для представления порошковых данных, которые разработаны в СКТБ ИПМ АН УССР и Объединенном комитете порошковых дифракционных данных JCPDS [90]. Формат JCPDS рекомендован Комиссией по журналам МСК в качестве международного стандартного формата [156].

Основу обоих форматов составляет формат карточки JCPDS с учетом рекомендаций работ [88, 126]. Между форматами нет принципиальных различий. Имеющиеся различия носят композиционный характер и относятся в основном к текстовой части формата. Международный формат, приведенный в Приложении 1, имеет более четкую композицию. Вероятно, рациональность структуры формата также имеет определенное значение, способствуя более быстрому его распространению. Ниже приводится краткое сравнительное описание этих форматов и высказываются

17-150

d	10.3	3.59	5.20	10.3	$\text{Cu}(\text{UO}_2)_2(\text{AsO}_4)_2 \cdot 12\text{H}_2\text{O}$					
I/I_1	100	100	70	100	Copper Uranyl Arsenate Hydrate			Zeunerite		
Rad. Fe λ	Filter Mn		Dia. 57.3 mm		$d, \text{Å}$	I/I_1	hkl	$d, \text{Å}$	I/I_1	hkl
Cut off	I/I_1 Visual				10.3	100	002	2.08	60	0.0.10, 305
Ref. Walenta, Tscherm. min. pet. Mitt. <u>9</u> 111-74 1964					6.81	30	101	1.924	60	1.1.10
Sys. Tetragonal S.G. $P4/nno$ (126)					5.20	70	004	1.794	50	2.0.10, 325
					4.98	50	103	1.765	5	402
Ref. Ibid.					4.58	5	112	1.735	5	411, 0.0.12
a_0 7.18	b_0	c_0 20.79	A	C 2.896	3.59	100	105, 200	1.708	10	318
α	β	γ	Z 2	D_x 3.40	3.39	50	202	1.690	10	330, 413
Ref. Ibid.					3.16	10	211	1.661	5	
E α 1.586 $n_{D\beta}$ 1.615 E γ Sign					3.06	5	212	1.640	40	
					2.95	30	204	1.602	40	
2V D 3.39 mp Color Yellowish green					2.91	5	213	1.584	10	
Ref. Ibid.					2.86	10	116	1.562	20	
					2.74	30	107, 214	1.532	50	
Sample from Anton Mine, Black Forest, Germany					2.53	40	215, 220	1.427	30	
					2.46	30	222	1.378	40	
See following card					2.37	5	301, 216	1.362	30	
					2.32	20	118	1.355	30	
					2.26	40	303, 310	1.312	20b	
					2.21	40	109, 312	1.289	5	
					2.18	20	217	1.272	30b	

17-150a

d	10.3	3.59	5.20	10.3	<chem>Cu(UO2)2(AsO4)2·12H2O</chem>					
I/I_1	100	100	70	100	Copper Uranyl Arsenate Hydrate			Zeunerite		
Rad.	λ	Filter	Dia.		$d, \text{\AA}$	I/I_1	hkl	$d, \text{\AA}$	I/I_1	hkl
Cut off	I/I_1				1.245	40		0.998	30	
Ref.					1.221	5		0.995	40b	
					1.210	10		0.991	10	
					1.201	5				
					1.188	5				
Sys.		S.G.								
a_0	b_0	c_0	A	C	1.178	5				
α	β	γ	Z	D_x	1.165	40b				
Ref.					1.146	20				
					1.133	20				
					1.127	20				
$\epsilon\alpha$	$n\omega\beta$	$\epsilon\gamma$	Sign		1.117	5				
2V	D	mp	Color		1.092	50b				
Ref.					1.083	50b				
					1.059	40b				
					1.043	20b				
					1.036	40b				
					1.022	40bb				
					1.015	5				
					1.004	20				
					1.001	30				
See preceding card										

Рис. 1.3. Основная карточка и карточка-продолжение JCPDS для минерала цейнерита.

соображения по разработке нового формата, исключаящего их недостатки. В формате СКТБ ИМП АН УССР (Приложение 2) идентификатор соединения выделен в самостоятельный раздел, а в международном он входит в раздел «Характеристика образца». В последний включается химический анализ соединения и способ его получения, если это синтезированная фаза или местонахождение для минералов. Остальные физико-химические характеристики не детализированы и названы как «дополнительные характеристики образца». Раздел «Условия получения рентгенодифракционного спектра и способы его обработки» в международном формате содержит подробные характеристики прибора и условий регистрации спектра. В нем особое внимание уделено воспроизводимости и разрешению спектра. Формат СКТБ ИМП АН УССР содержит пункт «массовый коэффициент поглощения». Возможно, он представляет интерес для количественного рентгенофазового анализа, но из-за сильного влияния примесей на его численную величину и зависимости от длины волны едва ли он необходим. В число желательных данных международного формата включены пункты: порог обнаружения интенсивности, разрешающая способность по углу 2θ и воспроизводимость 2θ .

В разделе «Кристаллографические данные» (сведения о структуре соединения) оба формата практически содержат одинаковые пункты. К сожалению, в международном формате отсутствует важный пункт об изученности структуры соединения, который имеется в формате СКТБ ИМП АН УССР. В международном формате пункты этого раздела рассматриваются как желательные. Формат СКТБ ИМП АН УССР не содержит пункт «методика определения и уточнения параметров элементарной ячейки».

В сопоставляемых форматах имеются различия в содержании раздела «Таблица данных». Так, в международном формате обязательными являются значения $2\theta^0$ и I/I_0 и желательными $d_{изм}$, hkl , и $\Delta 2\theta = 2\theta_{изм} - 2\theta_{выч}$; в формате СКТБ ИМП АН УССР требуется приводить $d_{изм}$, $d_{выч}$, I/I_0 и hkl . Из сопоставления этих форматов вытекает, что они хорошо согласуются с ранее разработанным форматом карточки JCPDS. Сравнительный анализ форматов, рекомендаций к международному стандартному формату [90, 156] и обсуждение достоинств и недостатков показывают, что работу по совершенствованию формата следует продолжить.

В Приложении 3 приводится вариант формата, учитывающий результаты вышеприведенного анализа. Формат состоит из шести разделов (см. п. 1.1). В него входят все основные пункты международного формата с дополнениями, заимствованными из формата СКТБ ИМП АН УССР. Все пункты формата подразделяются на обязательные и желательные. Раздел «Таблица данных» включает колонки $2\theta_{изм}$, I/I_0 (обязательные), $d_{эксп}$, hkl , $\Delta 2\theta = 2\theta_{изм} - 2\theta_{выч}$ или $\Delta d = d_{эксп} - d_{выч}$ (желательные). Необходимость включения в таблицу величин $2\theta_{изм}$ связана с тем, что они получаются непосредственно из эксперимента и для них легче оценить инструментальные и физические погрешности, чем для $d_{эксп}$.

Действительно, величины $d_{\text{эксп}}$ вычисляются из уравнения Вульфа — Брэггов, а это приводит к дополнительной неточности, связанной с округлением. Эта неточность для больших значений 2θ может быть довольно значительной (до $0,5^\circ 2\theta$).

В последнее время наиболее общеупотребительным является формат JCPDS. Поэтому, отталкиваясь от данного формата, мы будем проводить далее обсуждение структуры порошковых данных, принятых в других определителях и в автоматизированных картотеках, а также рассматривать методики ручного и автоматизированного качественного рентгенофазового анализа. Однако перед тем как обратиться к этим вопросам, необходимо рассмотреть структуру и состав картотеки JCPDS и вспомогательных справочников (ключей, указателей или индексов), облегчающих работу с картотекой.

1.4. СТРУКТУРА И СОСТАВ КАРТОТЕКИ JCPDS И УКАЗАТЕЛЕЙ К НЕЙ

После выхода в свет статьи Ханавальта и др. [105] Американское общество по испытанию материалов (ASTM) предприняло работы, связанные с реализацией идей по рентгенофазовому анализу по порошковым рентгендифракционным спектрам, высказанным в этой публикации. В частности, это касалось проблемы идентификатора кристаллических фаз.

В 1941 г. при участии ASTM эти рентгендифракционные спектры были переизданы в формы картотеки, снабженной алфавитным указателем названий соединений и ключом (указатель Ханавальта), схема которого дана в [105]. Ключ содержал три самые интенсивные линии, химическую формулу соединения и номер карточки. В последующие годы картотека расширилась примерно до 2500 соединений. В рамках ASTM в 1943 и 1944 гг. она издана двумя выпусками. Для выполнения данной работы под эгидой ASTM был организован комитет, возглавляемый известным специалистом по рентгеноструктурному анализу проф. У. П. Деви. Позднее к работам по совершенствованию и накоплению порошковых рентгенографических данных присоединились научные и научно-общественные организации Канады, Великобритании и Франции. Комитет приобрел функции международной организации. С 1969 г. он стал полностью самостоятельным и картотека получила название «Порошковая дифракционная картотека Объединенного комитета порошковых дифракционных стандартов» (PDF JCPDS).

Одной из важнейших функций комитета JCPDS являются сбор и опубликование картотеки порошковых данных, выполняющих роль эмпирических справочных стандартов. Картотека состоит из двух независимых частей, включающих сведения о неорганических и органических соединениях. В дальнейшем, если это не

УС

Таблица 1.1

Динамика накопления информации в картотеке JCPDS (до 1969 г. картотека ASTM)

Номер сета	Год	Число соединений в сете	Номер сета	Год	Число соединений в сете
1	1943	1300	18	1968	1500
2	1944	1465	19	1969	1500
3	1950	1220	20	1970	1500
4	1952	900	21	1971	1500
5	1953	750	22	1972	1500
6	1955	725	23	1973	1500
7	1957	700	24	1974	1500
8	1958	825	25	1975	1500
9	1959	860	26	1976	1500
10	1960	850	27	1977	1500
11	1961	1000	28	1978	1500
12	1962	1000	29	1979	1500
13	1963	1000	30	1980	1500
14	1964	850	31	1981	1500
15	1965	1000	32	1982	1500
16	1966	1200	33	1983	1500
17	1967	1000	34	1984	1500

будет оговариваться особо, речь будет идти о картотеке для неорганических соединений.

Принцип формирования картотеки — хронологический. Все данные, поступающие в комитет JCPDS в течение года, публикуются в виде отдельного набора (сета) карточек. Картотека накапливается с 1941 г., а с 1957 г. (сет 7) начал ежегодный выпуск сетов, таким образом, в 1984 г. выпущен сет 34. Только сет 4—17 включали в себя меньше 1000 карточек, а с сета 18 каждый содержит 1500 карточек неорганических соединений. Они включают сведения для многих уникальных соединений. В табл. 1.1 приведено число карточек (соединения) в каждом сете по данным на 1984 г. Полное число содержащихся в картотеке соединений по данным на 1984 г. (сет 34) составляет примерно 40 000 карточек.

Необходимо заметить, что здесь и далее мы будем, как правило, отождествлять понятие «карточка» и «соединение» применительно к картотеке JCPDS. На самом деле, число реально существующих карточек всегда больше числа справочных стандартов или соединений, так как для некоторых соединений информация размещена на двух карточках — основной и карточке-продолжении, и для ряда соединений в картотеке присутствуют несколько основных карточек, содержащих данные, полученные либо разными авторами, либо теми же, но имеющие более подробное содержание на всех полях, кроме 1 и 8—10. В качестве примера второй ситуации укажем на карточки для низкотемпературного альбита 9—466 и 19—1184, рихтерита 10—456, 15—1068 и 20—484, апофиллита 19—82 и 19—944 и многие другие. Отсутствие в течение многих лет стандартных требований к представляемым в

картотеку данным приводило к включению в картотеку не всегда качественных данных. Комитет JCPDS стремится повысить качество данных, проводя работы в двух направлениях. Во-первых, в 1980 г. приняты новые требования к объему и качеству данных, представляемых в картотеку JCPDS; изложение этих требований проведено в разделе 4.3 и Приложении 1. Во-вторых, при комитете JCPDS функционирует редакторская группа, в задачу которой входит проверка новых данных, перепроверка устаревших, противоречивых и дублирующихся, а также оценка их качества по рассмотренной выше системе, используя, например, программу [111]. В этой связи начиная с 1950 г. комитетом JCPDS регулярно публикуются списки изъятых из картотеки данных (номера карточек).

Рассмотренные выше меры не всегда, однако, позволяют исключить дублирование данных в картотеке. Для некоторых синтетических и природных соединений (особенно минералов) рентгендифракционный спектр может отражать особенности, связанные с технологией синтеза соединения или со специфичной месторождений, в которых встречен данный минерал. В подобных случаях помещение в картотеку нескольких карточек для одного соединения может быть оправдано; эта информация представляет интерес, например, для минералогов и геологов.

Из изложенного выше ясно, что с физико-химической точки зрения картотека JCPDS является совокупностью бессистемно накапливаемых данных (карточек). Картотека не имеет четкой внутренней структуры, в которой отражались бы физические, химические и кристаллоструктурные свойства соединений. Как указывалось выше, информация в картотеке располагается в хронологическом порядке от сета к сету (это только тенденция, но не правило); таким образом, соединения (карточки) с последовательными номерами не имеют, как правило, ничего общего, если не считать, что они размещаются по алфавиту названий элементов в соединении.

Создание обоснованной внутренней структуры картотеки JCPDS является достаточно сложным и спорным вопросом. До сих пор не существует единой классификации неорганических химических соединений, подобной, например, классификации К. Линнея в биологии; отсутствует единая общепринятая классификация и более узкого круга объектов — минералов. До сих пор не ясно, можно ли создать такую классификацию минералов, которая была бы достаточно обоснованной с точки зрения всех изучающих минералы наук — кристаллографии, минералогии и геологии, структурной и неорганической химии, кристаллофизики и физики твердого тела. Следовательно, вопрос о создании структуры картотеки JCPDS и любой картотеки подобного типа выходит за рамки соответствующей специализации (в данном случае рентгеноструктурного анализа поликристаллов). Этот вопрос сводится в конечном счете к проблеме широчайшей общности, которая может быть решена на основе глубокого взаимодействия ряда отраслей

знания. Хотя примерно хронологический принцип, положенный в основу формирования рассматриваемой картотеки, является ее недостатком, вряд ли этот недостаток можно поставить в вину комитету JCPDS, поскольку явных, более всеобъемлющих и логически обоснованных принципов пока не существует.

Однако отсутствие обоснованной внутренней структуры картотеки JCPDS значительно усложняет ее использование. Оно сводится к поиску нужных соединений (карточек) при некоторых известных исходных параметрах. Проанализируем, какие виды поиска обычно осуществляются по картотеке JCPDS.

Первый тип — поиск соединения (карточки) по известному номеру. Он осуществляется путем выборки нужной карточки из соответствующего сета.

Второй тип — поиск по названию. Для его проведения необходимо иметь список «Название — номер», уже готовый для всей картотеки; естественно, что для облегчения поиска названия соединений должны перечисляться в алфавитном порядке.

Третий тип — поиск по химической формуле. Для него необходимо иметь список «Химическая формула — номер», в котором химические формулы соединений также упорядочены по алфавиту. Легко заметить, что составление списков для второго и третьего типов поиска означает фактически введение внутренней структуры картотеки в соответствии с алфавитом и названиями соединений или же в соответствии с алфавитом и химическими формулами соединений.

Под четвертым типом поиска можно обобщенно понимать поиск многопараметрический, т. е. по любым входным параметрам (кроме спектра) или любой их совокупности. Первые три типа являются частными случаями многопараметрического поиска.

Пятый тип — поиск по спектральным данным (по совокупности значений $\{d_i, I_i\}$), т. е. качественный анализ или идентификация неизвестного соединения по его спектру. Как и в предыдущем случае, он является многопараметрическим, поскольку при его проведении осуществляется проверка совпадения с данными справочных стандартов нескольких значений исходных параметров. Под исходными параметрами в этом случае понимаются величины $\{d_i, I_i\}$. Пятый тип поиска сложнее четвертого, поскольку в нем в качестве исходных данных и данных справочных стандартов выступают физические величины, которые не могут быть определены абсолютно точно (содержат ошибки измерения). В результате этот тип поиска допускает некоторую неопределенность: требуется не точное, а примерное совпадение исходных параметров и данных справочного стандарта (достаточно совпадения не всех параметров, а только наиболее важных — сильных линий).

Ускорения поиска во всех перечисленных выше случаях можно добиться путем выделения из основной картотеки отдельных ее частей, которые могут использоваться автономно в некоторой области знания. Входящие в такую отдельную выборку соединения

должны иметь нечто общее, а саму выборку необходимо четко определить. Примерами таких выборок могут служить минералы или совокупность данных для соединений некоторого элемента (алюмо- и оловосодержащие соединения, соединения железа, карбонаты и т. п.). Выборка может носить также сугубо специализированный характер, связанный с конкретным списком соединений, исследуемых на некотором предприятии.

Проведенный анализ существующих на практике типов поиска позволяет понять необходимость формирования в рамках картотеки ряда дополнительных указателей. Перед тем как перейти к описанию их структуры, рассмотрим перечень основных материалов, которые обычно объединяют под наименованием «Американская рентгенометрическая картотека», «Картотека JCPDS» или «Порошковый дифракционный файл» (Powder Diffraction File — PDF).

1. Ежегодные информационные сообщения (Featuring...) [142]. Они рекламного характера, содержат сведения о составе картотеки JCPDS, формах и способах ее поставки, цены на различные виды поставляемых информационных материалов. Наиболее подробная информация дается о последнем сете, накопленном в течение года, к которому относится выпуск данного информационного сообщения. Обычный объем его 30 с. Полное название этого документа есть в [142]. Следует еще раз подчеркнуть, что подобные информационные сообщения содержат сведения не только о последнем сете, но и обо всех материалах картотеки JCPDS.

2. Картотека JCPDS в традиционном печатном виде на карточках — Data Cards [140, 141].

3. Алфавитный указатель наименований соединений (указатель Дэви) (Alphabetical Index) [145] (см. рис. 1.4).

4. Числовой указатель Ханавальта [146] (см. рис. 2.1).

5. Числовой указатель Финка [147] (см. рис. 1.5).

Указатели Дэви, Ханавальта и Финка, как правило, представляют собой отдельные тома и образуют руководство для пользователя (Manual) картотекой JCPDS.

6. Выборка данных для минералов (в виде книг) [151].

7. Выборка данных для наиболее распространенных соединений [144].

8. Выборка данных для металлов и сплавов [150].

9. Картотека JCPDS на микрофишах — Microfiche System.

10. Картотека JCPDS — база данных на магнитной ленте Data Base Magnetic Tape; поставляется, как правило, вместе с программой идентификации Джонсона — Ванда и с соответствующим руководством для пользователя.

11. Картотека JCPDS (база данных) для автоматизированных порошковых дифрактометров.

В табл. 1.2, взятой из работы [130], приведены объемы полной картотеки JCPDS и отдельных выборок из нее для первых 28 сетов.

Алфавитный указатель наименований соединений состоит из трех частей. Часть первая содержит наименования неорганиче-

Число карточек в картотеке JCPDS и отдельных выборках из нее (сеты 1—28)

Картотека или выборка	Число карточек	Картотека или выборка	Число карточек
Полная картотека JCPDS	34 534	Минералы	3039
Неорганические соединения	24 971	Наиболее распространенные соединения	2378
Органические соединения	10 721	Порошковые дифракционные данные Национального бюро стандартов США	947
Металлы и сплавы	5 700		

ских соединений в соответствии с принятыми в химии правилами [42, 43, 48, 58], вторая — минералов, третья — списки исключенных из картотеки JCPDS карточек. Пользование указателем облегчается, если имеется справочник, позволяющий быстро устанавливать соответствие между английскими и русскими наименованиями соединений; для минералов в качестве такого справочника можно рекомендовать монографию [48] и дополнение к ней [42].

Алфавитный указатель является списком типа «Название — номер». С его помощью можно быстро разыскать номера карточек по названиям соответствующих соединений; задача поиска номеров карточек по химическим формулам соединений решается с помощью данного указателя только частично. Отдельного указателя «Химическая формула — номер» в справочном аппарате картотеки JCPDS нет, поэтому, если точное название соединения не известно, следует его установить, пользуясь химическими справочниками.

Фрагмент алфавитного указателя наименования соединений приведен на рис. 1.4. В каждой строчке указателя содержится следующая информация об одном соединении:

наименование соединения с указанием перед ним признака достоверности;

химическая формула соединения;

значение d для самой сильной линии, интенсивность которой помечается символом «х» (латинское 10);

					File No.	I/I_c
*	Barium Silicon Fluoride:	BaSiF ₆	3.59 _x	1.95 ₇	3.06 ₆	14—736 4.20
o	Barium Silicon Nitride:	BaSiN ₂	2.89 _x	3.01 ₈	2.79 ₈	27—41
*	Barium Silicon Sulfide:	(Ba ₂ SiS ₄)280	4.02 _x	3.01 ₉	3.39 ₇	22—92
c	Barium Silver:	(BaAg ₆)6H	2.44 _x	2.48 ₃	2.32 ₃	28—156
i	Barium Silver Fluoride:	BaAgF ₄	1.81 _x	3.42 ₉	2.08 ₉	26—181

Рис. 1.4. Фрагмент алфавитного указателя наименований неорганических соединений картотеки JCPDS.

	3.31—3.25	File No.	Fische No.
*	3.25 ₈ 3.23 _X 2.74 ₈ 2.08 ₃ 2.01 ₃ 2.00 ₃ 5.90 ₄ 4.57 ₄	20—864	1—425—B
o	3.31 ₄ 3.22 ₄ 3.13 ₆ 3.42 ₆ 0.00 ₁ 6.61 _X 4.19 ₆ 3.55 ₆	41—297	1—40—D
*	3.31 ₄ 3.22 _F 3.09 ₅ 3.02 _X 2.97 ₃ 2.90 ₃ 2.75 ₅ 3.34 ₄	9—346	1—32—F
c	3.31 ₄ 3.22 ₆ 3.09 ₅ 3.02 _X 2.97 ₄ 2.90 ₄ 2.75 ₅ 3.34 ₄	20—24	1—117—C
*	3.31 ₅ 3.22 ₅ 4.04 ₈ 7.76 _X 7.15 ₈ 6.53 ₈ 6.51 ₈ 3.57 ₅	49—1036	1—142—C

Рис. 1.5. Фрагмент указателя Финка карточки JCPDS.

значение d для второй сильной линии и ее интенсивность в десятибалльной шкале, которая помечается как нижний индекс к числу d ;

значение d для третьей по интенсивности линии и ее интенсивность аналогично второй линии;

номер соответствующей карточки JCPDS;

отношение I/I_c — корундовое число.

Численные указатели Ханавальта и Финка предназначены для проведения ручного качественного фазового анализа (пятый тип поиска). В разные годы детали построения этих указателей варьировались, их совершенствование продолжается по мере увеличения объема карточки JCPDS. Фрагменты этих указателей приведены на рис. 1.5 и 2.1.

Указатель Ханавальта построен следующим образом.

Рассматривается картотека, содержащая M карточек (соединений); величина M для картотеки JCPDS или специализированных представительных выборок из нее порядка тысяч или десятков тысяч. Из каждой карточки выбираются d и I для трех сильнейших по интенсивности линий, а также еще пять наиболее интенсивных линий, и каждой тройке линий приписывается номер карточки N_{JCPDS} :

$$d^{(1)}I^{(1)}N_{JCPDS}, d^{(2)}I^{(2)}N_{JCPDS}, d^{(3)}I^{(3)}N_{JCPDS}.$$

Таким способом формируется единый массив, содержащий $3M$ линий. Массив упорядочивается по убыванию d от 999,99 до 1 \AA . Затем вся область значений межплоскостных расстояний d разбивается на интервалы, причем в разные годы их число колебалось в широких пределах (46—в указателе [146]). Границы первого интервала 999,99 — 10 \AA , остальные 45 интервалов располагаются в области 9,99 — $1,00 \text{ \AA}$. Интервалы выбраны так, чтобы число линий (значений d), попадающих в каждый интервал, было примерно пос-

тоянным (в пределах 2000). Весь массив разбивается на группы, называемые группами Ханавальта. В главе 5 более подробно рассмотрена задача Ханавальта о построении интервалов в шкале d , содержащих равное число линий; здесь отметим только, что число линий, попадающих в разные группы Ханавальта, в указателе может колебаться.

В разбитом на 46 групп массиве каждая карточка встречается трижды, в соответствии с отобранными тремя сильнейшими линиями. К каждой из этих трех линий, находящихся в определенных группах и строках указателя, дописываются остальные две; порядок дописывания подчиняется закону циклической перестановки:

$$\begin{aligned} d^{(1)} d^{(2)} d^{(3)} \\ d^{(2)} d^{(3)} d^{(1)} \\ d^{(3)} d^{(1)} d^{(2)}. \end{aligned} \quad (1.5)$$

Таким путем образуется структура строки, состоящая из ряда позиций: в первой стоит одна из трех линий, определяющая попадание всей строки в данную группу Ханавальта, во второй размещается вторая по интенсивности линия, определяющая подгруппу, а третья служит ограничению вариантов при ручном поиске. Во второй и третьей позициях размещаются две остальные линии триады в порядке, заданном перестановочным законом; в четвертой находится номер карточки N_{JCPDS} . После этого строки в пределах каждой группы упорядочиваются по убыванию d линий, находящихся во второй позиции, образуя подгруппы; естественно, существовавшее ранее упорядочение по стоящему в первой позиции d разрушается. К 3 первым линиям, находящимся в 1, 2 и 3-й позициях и выделенным обычно жирным шрифтом, дописываются еще 5 сильнейших по интенсивности линий данного соединения и химическая формула; при этом N_{JCPDS} оказывается уже в десятой (предпоследней) позиции строки. Последнюю позицию в строке занимает корундовое число. Таким способом формируется окончательная структура указателя Ханавальта, несколько строк из которого представлены на рис. 2.1.

Информация о d и I дается в строчке указателя Ханавальта в таком же виде, как и в строчке алфавитного указателя: величины d приводятся в Å, а соответствующих I — в десятибалльной шкале в виде нижнего индекса к числу d ; символ «х» обозначает римскую 10.

Ежегодное увеличение объема картотеки JCPDS (до 40 000 карточек в 1984 г.) резко расширило указатель Ханавальта, что существенно усложнило процедуру ручного поиска при фазовом анализе и заметно ограничило его возможности. Поэтому начиная с 1982 г. записи вариантов (входов) представления дифракционного спектра каждого справочного стандарта в ключе подчиняются следующим правилам.

1. Все дифракционные спектры справочных стандартов картотеки JCPDS представлены (по меньшей мере один раз) парой величин (d_1, d_2), где d_1 — межплоскостное расстояние самой интенсивной линии и d_2 — межплоскостное расстояние следующей по интенсивности линии спектра.

2. Линии дифракционного спектра справочного стандарта повторяются в ключе дважды как (d_1, d_2) и (d_2, d_1), когда $I_2/I_1 > 0,75$, а $I_3/I_1 \leq 0,75$, т. е. $I_1 > I_2 > I_3$.

3. Линии дифракционного спектра справочного стандарта появляются в ключе три раза (d_1, d_2), (d_2, d_1) и (d_3, d_1), когда $I_3/I_1 > 0,75$, а $I_4/I_1 \leq 0,75$, т. е. $I_1 > I_3 > I_4$.

4. Линии дифракционного спектра справочного стандарта записываются в ключ четыре раза (d_1, d_2), (d_2, d_1), (d_3, d_1) и (d_4, d_1), если $I_4/I_1 > 0,75$.

В результате такого подхода среднее число записей на один спектр справочного стандарта составило примерно 1,7 вариантов.

Необходимо подчеркнуть, что входом в указатель Ханавальта являются две или три первые самые интенсивные линии, определяющие группу, подгруппы и ограничитель числа входов в подгруппе — третья линия (если она имеется); пять остальных линий и химическая формула соединения представляют дополнительную информацию, которая используется для более детального сравнения с изучаемым спектром и также ограничивает число карточек для дальнейшей работы по идентификации анализируемой пробы. Выходом из указателя Ханавальта служат номера карточек; по ним отбираются карточки для наиболее подробного сравнения с изучаемым спектром.

Указатель Финка внешне похож на указатель Ханавальта; при построении указателя Финка использован ряд правил, впервые предложенных Ханавальтом. Отличие между указателями заключается в том, что указатель Финка имеет четыре входа и информация для каждой карточки повторяется в нем четыре раза. Это обстоятельство делает указатель Финка более удобным при практической работе.

Перейдем теперь к рассмотрению специализированных выборок из картотеки JCPDS (см. табл. 1.2)*. Это картотеки «Минералы», «Металлы и сплавы», «Порошковые дифракционные данные Национального бюро стандартов» [143], опубликованные в форме книг, содержащие копии карточек и соответствующие указатели для ручного поиска. Книга «Наиболее распространенные соединения» содержит только указатели и является своеобразным «мини-ключом» для поиска этих фаз в картотеке JCPDS.

Картотека «Минералы» была опубликована в 1974 г. в виде двухтомника: собственно картотеки с алфавитным указателем и указателем Ханавальта для ручного поиска по карточкам первого

* Некоторое расхождение объемов специализированных выборок, приводимых ниже, с данными табл. 1.2 связано с тем, что Мак-Кэрти и Джонсон [130] приводят сведения для первых 28 сетов картотек.

тома. В этой картотеке было собрано около 1900 наименований минералов и их разновидностей (примерно 2500 карточек) из 23 выпусков картотеки JCPDS. В 1980 г. было издано в виде отдельной книги первое дополнение к подфайлу «Минералы», в которое вошли карточки данных о минералах из 24—29 выпусков картотеки JCPDS (около 850 наименований, более 1000 карточек). К нему прилагается том указателя Ханавальта на весь массив карточек минералов, вошедших в 29 выпусков картотеки JCPDS.

В 1978 г. был подготовлен и опубликован трехтомник выборки «Металлы и сплавы». Два тома содержат около 5700 карточек из 27 выпусков картотеки JCPDS, а третий включает алфавитный указатель и указатели Ханавальта и Финка для ручного поиска.

В 1977 г. опубликованы в отдельном двухтомнике порошковые дифракционные данные, полученные кристаллографической группой Национального бюро стандартов США*. Эти данные выбраны из 26 выпусков картотеки JCPDS. В первом томе содержатся 949 карточек для 938 неорганических и 11 органических веществ, дифракционные спектры которых получены в НБС по специально разработанным методикам, обеспечивающим их высокое качество. Ранее эти спектры публиковались в отдельных выпусках [157]. Второй том включает общий алфавитный указатель, алфавитные указатели минералов и органических веществ, а также указатели Ханавальта и Финка.

Выборка «Наиболее распространенные соединения», содержащая перечень примерно 2200 соединений, представляет собой книгу, в которой даны указатели Ханавальта и Финка для ручного поиска этих фаз и алфавитный указатель этих соединений. Она является вспомогательным указателем для ручного поиска при работе с полной картотекой JCPDS.

Картотека JCPDS также выпускается на микрофишах, для работы с которыми необходима специальная проекционная аппаратура.

Картотека JCPDS на магнитной ленте (база данных), естественно, не может быть непосредственно представлена в доступной человеческому зрению форме; для этого базу данных JCPDS надо распечатать на устройстве печати ЭВМ по специальной программе. Необходимо четко представлять несколько основных моментов, касающихся базы данных JCPDS: 1) она эквивалентна картотеке JCPDS в части числа соединений; все карточки, имеющиеся в картотеке, содержатся также в базе данных и 2) не эквивалентна картотеке JCPDS в части объема сведений об одном соединении. Элемент базы данных, относящийся к одному соединению (условная машинная карточка), содержит значительно меньше данных, чем традиционная карточка JCPDS (см. рис. 1.3). В машинной карточке содержатся: номер карточки N_{JCPDS} и соответствующего микрофиша, рентгенодифракционный спектр — таблицы значений $\{d_i, I_i\}$; химическая формула соединения; элементный состав (для

* Подробнее об этих данных сказано в разделе 1.5.

сплавов); химическое наименование; признак достоверности. Сведения о составе машинной карточки обычно приводятся в информационных сообщениях типа [142]. Таким образом, по своему составу элемент базы данных JCPDS стоит ближе не к карточке JCPDS, а к строке указателей Ханавальта и Финка, отличаясь от последних наличием полного спектра $\{d_i, I_i\}$, и названия соединения; 3) база данных JCPDS предназначена для проведения автоматического качественного анализа (пятый тип поиска) с помощью ЭВМ.

Из сказанного выше ясно, что в информационном смысле база данных JCPDS ни в коем случае не может заменить карточки JCPDS; для получения детальных сведений о структуре соединения, его свойствах, условиях получения спектра и авторах соответствующих работ необходимо обращаться к картотеке. База данных предназначена для эксплуатации на ЭВМ IBM-370 совместно с программой Джонсона — Ванда, которая будет рассмотрена в главе 4.

1.5. РЕНТГЕНДИФРАКЦИОННЫЕ СПРАВОЧНЫЕ СТАНДАРТЫ НАЦИОНАЛЬНОГО БЮРО СТАНДАРТОВ США [143, 157]

В Национальном бюро стандартов США (НБС) создана и эксплуатируется обширная картотека (свыше 60 000 соединений [133]), содержащая разнообразные характеристики химических соединений, в том числе органических минералов, металлов и сплавов. Одной из таких характеристик являются параметры элементарной ячейки и дифракционные спектры $\{d_i, I_i\}$. Эти данные для большинства содержащихся в картотеке веществ заимствованы из литературы.

В связи с различным качеством этих данных Кристаллографическая группа НБС, начиная с 1950 г., ежегодно публикует брошюры [157], содержащие дифракционные спектры кристаллических фаз, полученные экспериментально по специально разработанной для этой цели методике. Методика получения дифракционных спектров $\{d_i, I_i\}$ и описание текстовой части карточки приводятся во введении к каждой брошюре. Кроме того, в этом разделе есть сведения о стандартных веществах, используемых для определения значений 2 θ спектров, описываемых в их соединении. Эти стандартные вещества имеют параметры элементарной ячейки и значения 2 θ их дифракционных спектров, определенные с высокой точностью.

Спектры веществ $\{d_i, I_i\}$, публикуемые в брошюрах НБС, обладают достаточно высокой точностью и полнотой, так как они содержат линии как большой и средней, так и малой интенсивности. Кроме дифракционных спектров, полученных из эксперимента, в этих выпусках для веществ различных классов публикуются расчетные дифракционные спектры, получаемые из монокристалль-

ных структурных данных высокой точности. Количество карточек от выпуска к выпуску не постоянно, различно и число экспериментальных и расчетных спектров $\{d_i, I_i\}$.

Каждая карточка в этих выпусках имеет заголовок, состоящий из наименования химического соединения и его химической формулы. Для минералов и их синтетических аналогов в скобках приводятся их названия. Текстовая часть карточки почти не отличается от формата JCPDS.

Таблица данных имеет заголовок, в котором указаны вид излучения, его длина волны в Å, температура (°C), при которой получен дифракционный спектр, внутренний стандарт, использованный для корректировки значений 2θ , параметр его элементарной ячейки, а также четыре колонки: d — межплоскостные расстояния, Å; $I_{\text{отн}}$ — относительные интенсивности линий; hkl — индексы Миллера; 2θ — углы дифракции в градусах и его десятичных долях. При наложении линий спектра используются два способа указания на это; если налагается только две линии, то даются оба значения d и к ним добавляется символ M . При наложении большего числа линий значение d приводится один раз и вслед за индексами Миллера ставится знак $+$, интенсивность указывается только у первого значения d . Символ L (L — lesser) в колонке $I_{\text{отн}}$ указывает, что линия имеет интенсивность меньшую, чем единица. Количество линий спектра в таблице колеблется в широких пределах, достигая иногда 60—80.

Каждый выпуск данных эмпирических справочных стандартов НБС имеет несколько сводных указателей. Это — алфавитные указатели названий всех соединений, наименований органических веществ, минералов, а также формульный указатель органических веществ. Все они построены по единой схеме: наименование соединения, номер тома «Циркуляра 539» или выпуска «Монографии 25» (для нее к номеру выпуска приписывается индекс m) и страница. Формульный указатель органических веществ включает брутто-формулу, номер тома или выпуска и страницу, на которой находится карточка этого соединения. В выпуске карточки расположены по алфавиту наименования веществ.

Все данные справочных стандартов, подготовленные НБС, включаются в выпуски картотеки JCPDS и имеют высший признак достоверности — звездочку.

1.6. СТРУКТУРА И СОСТАВ ОТЕЧЕСТВЕННЫХ РЕНТГЕНОМЕТРИЧЕСКИХ ОПРЕДЕЛИТЕЛЕЙ

В СССР пионерами использования порошковых рентгенодифракционных спектров для идентификации кристаллических фаз были ленинградские кристаллографы из Федоровского института, которые разработали и издали первые рентгенометрические определители минералов [4, 50]. За период 1938—1984 гг. в СССР

издан ряд рентгенографических определителей, содержащих порошковые рентгендифракционные данные для различных классов минералов [14—16, 39, 46, 49, 51, 68, 69, 83]. Издание определителей минералов, по-видимому, связано с большим практическим значением задач по установлению фазового состава минерального сырья методом порошков. В отличие от карточек JCPDS отечественные рентгенометрические определители изданы в виде книг.

В первом отечественном определителе подробно рассмотрены требования к веществам, претендующим на роль эмпирических справочных стандартов, к структуре, содержанию и форме описания таблицы данных эмпирических справочных стандартов, включаемых в определитель. В соответствии с этими требованиями определитель делится на две части: в первой размещаются фактические материалы, т. е. таблицы дифракционных спектров кристаллических фаз, выполняющих роль справочных стандартов, во второй — справочно-поисковый аппарат.

Рассмотрим более подробно каждую часть определителя.

Фактический материал (таблицы данных), помещаемый в первой части определителей, черпается из двух источников: литературы и специально проводимых экспериментальных работ по получению дифракционных спектров эмпирических справочных стандартов в соответствии с требованиями работ [4, 49]. Совокупность сведений об отдельном минерале, приведенных в таблицах вышеперечисленных определителей, будем называть далее карточкой (по аналогии с принятой JCPDS терминологией).

Карточка в определителе В. И. Михеева и некоторых других авторов содержит следующую информацию: номер таблицы по определителю, русское и английское наименование минерала (соединения), химическую формулу соединения, местонахождение минерала, физико-химическую характеристику соединения — твердость, плотность (удельный вес), цвет, оптические данные и др., результаты химического анализа, кристаллохимическую формулу, некоторые кристаллографические данные — сингонию, пространственную группу, параметры элементарной ячейки, условия регистрации рентгендифракционного спектра, таблицы d , $I_{отн}$ и hkl , ссылки на литературу. Подробное описание содержания карточки дано во введении к определителю [49] и детализировано в [15]. Очевидно, что карточка в отечественных определителях в общем соответствует стандартной карточке JCPDS и включает, в принципе, все основные характеристики минерала и его дифракционный спектр.

Справочно-поисковый аппарат определителей состоит из ключа, списка литературы, алфавитного перечня минералов и оглавления. В каждой строчке ключа помещается 3—5 пар значений d и I наиболее интенсивных линий из дифракционного спектра каждой карточки. Значения d в строке указателя расположены в порядке убывания интенсивности; строки следуют в порядке убывания величины d . Три самые интенсивные линии в строке поочередно перемещаются на первое место по правилам циклической

Отечественные определители минералов

№ п/п	Класс соединений	Год издания	Кол-во карточек	Кол-во минералов	Ссылка
1	Общий определитель минералов				
	Том I	1957	905	876	[49]
	Том II	1965	332	300	[51]
2	Урановые и урансодержащие минералы	1960	123	105	[68]
3	Минералы, содержащие редкоземельные элементы	1961	138	85	[45]
4	Бораты	1969	232	110	[39]
5	Фосфаты	1974	249	225	[16]
6	Карбонаты	1980	151	116	[14]
7	Оловосодержащие минералы	1977	86	60	[13]
8	Урановые минералы	1981	234	144	[69]

перестановки и трижды появляются в соответствующем месте ключа в зависимости от величины d . Таким образом, ключ содержит $3M$ строчек, где M — число карточек в определителе. За набором пар d и I следуют название минерала и его химическая формула, последнюю позицию в строке занимает номер карточки в определителе.

Алфавитный перечень названий минералов содержит наименование минералов (иногда их химические формулы), три значения самых интенсивных линий спектра и соответствующие им номера карточек. Перечень и краткая характеристика отечественных определителей приведены в табл. 1.3.

Наиболее известным и еще используемым в практике рентгенофазового анализа минерального сырья является двухтомник «Рентгенометрический определитель минералов» В. И. Михеева [49, 51] (авторы второго тома В. И. Михеев и Э. П. Сальдау). В отличие от картотеки JCPDS в определителе карточки данных справочных стандартов минералов сгруппированы по классам, подклассам и группам. Например, класс «Сульфиды и сульфосоли», подкласс «Сульфиды», тип «Соединения A_2X ».

Ключ в отечественных определителях ближе всего соответствует указателю Ханавальта картотеки JCPDS, а алфавитный перечень минералов — указателю Дэви. В первых выпусках определителя минералов [49, 50] имелся также алфавитный перечень минералов на английском языке.

Совокупность определителей (см. табл. 1.3) является фактически отечественным аналогом выборки минералов JCPDS [51]. Некоторый недостаток отечественных определителей минералов — отсутствие в их карточках указания на номер эквивалентной карточки в картотеке JCPDS.

Совершенно иная структура представления информации, сравнительно с описанной выше, принята в определителе минералов [83]. Этот определитель, как и все определители, рассмотрен-

ные выше, не является чисто рентгенографическим, но содержит также физико-химическую информацию для приведенных в нем минералов. По структуре представления материала определитель [83] ближе к указателям Дэви, Ханавальта и Финка, чем к карточке JCPDS и перечисленным в табл. 1.3 определителям, отдельных карточек, относящихся к каждому минералу, определитель [83] не содержит. В нем дается только краткое описание каждого минерала в виде двух строк обширной таблицы, включающей минералы определенного класса. Определитель снабжен алфавитным перечнем минералов (на русском языке).

В пределах каждого класса минералы упорядочены по алфавиту в соответствии с их названиями. Для каждого минерала в двух строках таблицы дается следующая информация: название и химическая формула; сведения о структуре минерала (пространственная группа, сингония, параметры элементарной ячейки); оптические характеристики; плотность, твердость; значения d для нескольких главных линий рентгендифракционного спектра (обычно приводится шесть значений d). Совокупность перечисленных данных рассматривается автором [83] как диагностический спектр минерала, в котором рентгеноструктурная информация является важной, но не единственной составляющей.

Кроме того, в работах [22, 37, 46] опубликованы также таблицы дифракционных спектров, включающие по крайней мере три пронумерованных списка соединений, содержащих только химическую формулу и перечень $\{d_i, I_i\}$. Спектры в этих таблицах расположены по названиям элементов и комплексов (например, аммоний, NH_4) в русском алфавите. В первом списке [37] содержатся спектры 1299 соединений, во втором [46] — 1331 и в третьем [22] — 311 соединений. Основой таблиц [22, 37, 46] послужили дифракционные спектры соединений из работы [105], дополненные дифракционными спектрами из различных литературных источников, в том числе картотеки JCPDS. Эти списки спектров не заменяют рентгенографические определители и картотеку JCPDS, они носят сугубо учебно-иллюстративный характер.

Глава 2

МЕТОДЫ РУЧНОГО КАЧЕСТВЕННОГО РЕНТГЕНОФАЗОВОГО АНАЛИЗА

Необходимо подчеркнуть, что дифракционные спектры поликристаллических веществ — это экспериментально найденные величины и, значит, содержат в себе систематические и случайные погрешности. Поэтому экспериментатор должен отчетливо пред-

ставлять величины этих погрешностей, их закономерности и стремиться уменьшить их, с тем чтобы возможно надежнее и практически однозначно решать задачи качественного рентгенофазового анализа.

В этой главе будут рассмотрены основные методики ручного (рутинного) рентгенофазового анализа. Специфической особенностью этих методик являются опыт и интуиция аналитика-исследователя. На процедуру анализа также оказывает определенное влияние способ регистрации дифракционных спектров изучаемого образца и справочных стандартов.

2.1. ИНТЕРПРЕТАЦИЯ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Исторически первым способом регистрации дифракционных спектров от поликристаллических материалов был фотометод. Методики, основанные на использовании рентгенограмм, полученных фотометодом, применяются в лабораториях до сих пор. Как правило, фотоленки (дебаграммы) с дифракционными спектрами индивидуальных кристаллических фаз накапливаются в лабораториях, образуя с течением времени фильмотеки. Это фотоленки (дебаграммы), полученные от веществ (химических соединений, минералов, чистых металлов и сплавов и т. п.). Используя такую фильмотеку, исследователь имеет возможность проводить сравнение ранее полученных дифракционных картин с дифракционными картинками изучаемых поликристаллических веществ. Таким образом, по мере создания в лабораториях фильмотек возникли условия для получения ответа на вопрос: является ли изучаемый объект однофазным (мономинеральным) или представляет собой смесь нескольких фаз и каких именно?

Естественно, сравнение возможно при условии, что дифракционные картины получены на одном и том же излучении и в камерах одинакового диаметра. Фотообработка пленок должна производиться в идентичных условиях, чтобы ее усадка была примерно одинаковой, а образцы имели одинаковый размер (диаметр столбика).

Сняв рентгенограмму от анализируемого вещества в тех же условиях, что и рентгенограммы справочных стандартов, можно проводить их непосредственное сопоставление путем наложения пленок или построением интрих-диаграмм в удобном для работы масштабе.

В случае совпадения всех линий на рентгенограмме изучаемого вещества с линиями рентгенограммы одного из справочных стандартов изучаемому веществу можно присвоить наименование вещества справочного стандарта. Это наиболее простой вариант идентификации анализируемого вещества по рентгеновскому дифракционному спектру.

На практике исследователь чаще всего сталкивается с объектами, которые представляют смесь нескольких кристаллических фаз. В этих случаях приходится сравнивать рентгенограмму изучаемого вещества с рентгенограммами нескольких справочных стандартов. Возможны несколько вариантов решения этой задачи.

1. Непосредственное сравнение рентгенограммы изучаемого вещества с рентгенограммами справочных стандартов. После того как все линии на рентгенограмме анализируемого вещества окажутся сопоставленными с линиями рентгенограмм справочных стандартов, вопрос о фазовом составе изучаемого вещества станет решенным. Может оказаться, что не все линии на рентгенограмме изучаемого вещества будут идентифицированы с данными справочных стандартов. Речь идет, в первую очередь, о сильных и средних по интенсивности линиях на рентгенограмме. Основной причиной этого может быть то, что в фильмотеке недостаточно справочных стандартов и часть неидентифицированных линий принадлежит фазам, для которых в фильмотеке нет рентгенограмм.

2. Использование штрих-диаграмм анализируемого вещества и справочных стандартов, построенных в удобном для сопоставления масштабе, где по оси абсцисс отложены либо углы θ или 2θ , если рентгенограммы сняты на одинаковом излучении, либо величины d . В последнем случае не обязательно снимать рентгенограммы в одинаковых условиях, так как d не зависит от условий съемки. По оси ординат откладываются величины относительных интенсивностей линий. Удобно строить штрих-диаграммы на прозрачной бумаге, например на кальке, тогда существенно упрощается процедура идентификации.

Способ непосредственного сопоставления рентгенограмм или штрих-диаграмм изучаемого вещества и справочных стандартов весьма трудоемок и малопродуктивен. Составление фильмотеки справочных стандартов требует огромных затрат труда, а сама процедура сопоставления малопроизводительна и утомительна, так как она основана на переборе и сравнении множества рентгенограмм справочных стандартов. Эта методика на практике применяется тогда, когда у исследователей есть предположение о фазовом составе вещества или когда аналитик, имеющий большой практический опыт работы с фильмотекой, быстро находит идентичный изучаемому веществу справочный стандарт. Методика непосредственного сравнения рентгенограмм используется при идентификации изоструктурных соединений и изоморфных смесей.

2.2. МЕТОДИКИ РЕНТГЕНОФАЗОВОГО АНАЛИЗА, ИСПОЛЬЗУЮЩИЕ ВЕЛИЧИНЫ $\{d_i, I_i\}$

В работах А. К. Болдырева и В. И. Михеева [4, 49] для идентификации кристаллических веществ предлагается использовать дифракционные спектры вида $\{d_i, I_i\}$. Как уже указывалось, основное преимущество этого метода состоит в том, что

межплоскостные расстояния, рассчитанные из порошковых рентгенограмм, не зависят от основных условий эксперимента — длины волны излучения и диаметра камеры, в которой производится регистрация дифракционной картины. Благодаря этому стало возможным использовать в качестве справочных стандартов дифракционные спектры — $\{d_i, I_i\}$, получаемые в других лабораториях, на различном оборудовании и в разных условиях эксперимента.

Особенность этого метода состоит в двухэтапности процесса идентификации. На первом этапе по самым ярким линиям дифракционного спектра изучаемого вещества с помощью ключа отбираются вероятные справочные стандарты — находятся номера их карточек. На втором этапе сопоставляются полные спектры справочных стандартов с дифракционным спектром изучаемого вещества. На этой основе разработано несколько методик рентгенофазового анализа. Некоторые из них будут рассмотрены в последующих разделах этой главы.

2.2.1. Методика рентгенофазового анализа по В. Н. Михееву [49]

Эта методика идентификации анализируемого вещества по дифракционному спектру проводится в два этапа. На первом этапе с помощью ключа производится выбор справочных стандартов. Для этого из таблицы $\{d_i, I_i\}$ анализируемого вещества выбираются три самые интенсивные линии. Затем в ключе находится область, соответствующая межплоскостному расстоянию самой сильной линии с учетом возможной погрешности его определения, т. е. $d \pm \Delta d$. В этих пределах ведется поиск справочных стандартов по трем выделенным линиям спектра анализируемого вещества. Так как в ключе содержится пять значений d для каждого справочного стандарта, одновременно производится сопоставление значений d изучаемого вещества со значениями d справочных стандартов, и тем самым уже на первом этапе ограничивается число справочных стандартов, с которыми предстоит работать на втором этапе идентификации состава анализируемого вещества.

При правильно выбранных самых интенсивных линиях в дифракционном спектре изучаемого вещества может случиться, что в ключе, а значит и в картотеке отсутствует справочный стандарт, имеющий идентичный спектр, что дифракционный спектр анализируемого вещества и справочного стандарта идентичной фазы получен разными методами, например на дифрактометре и в Дебаявской камере. В этом случае самые сильные линии в спектрах будут различными. Для подтверждения данного тезиса в качестве примера ниже приводятся ключевые линии кварца, полученные в камере Дебая [49] и на дифрактометре [51]:

3,34(10), 1,813(9), 1,539(9), 1,372(8), 1,380(8) карточка 256 [49]
3,34(100), 4,26(35), 1,817(17), 1,541(15), 2,458(12) карточка 5-490 [51].

Этот факт подчеркивает необходимость учета условий получения дифракционных спектров, а именно способа их регистрации. В скобках указаны относительные интенсивности линий.

Когда изучаемое вещество является смесью нескольких кристаллических фаз, может оказаться, что среди выбранных линий в качестве ключевых будут взяты линии, принадлежащие разным фазам. Этот вариант ошибочного выбора ключевых линий исключается при сравнении выделенных линий из спектра анализируемого вещества с ключевыми линиями справочных стандартов, которых в ключе не три, а пять. Как правило, при сравнении с пятью ключевыми линиями довольно легко исключаются неправильно выбранные в триаде линии из дифракционного спектра анализируемого вещества. В результате у аналитика имеется возможность скомпоновать новую триаду для поиска следующей фазы. Как отмечено в [49], теоретически мыслимы и встречаются на практике такие случаи, когда в триаде сильных линий окажется линия, образовавшаяся в результате наложения двух компонентов средней интенсивности от разных фаз. В этом варианте по первым двум линиям невозможно отыскать в ключе вещество с такой сильной третьей. Наличие еще двух дополнительных имеющихся в ключе линий повышает надежность идентификации. Таким образом, когда имеется совпадение линий компонентов смеси, достоверность определения повышается, если метод предварительного отбора, подходящих справочных стандартов основан на большем числе (>3) ярких линий.

Первый этап идентификации может осуществляться также с помощью алфавитного указателя, когда имеются предварительные сведения о фазовом (минеральном) составе изучаемого вещества. Алфавитный указатель в работах [49, 145], кроме названий соединений, содержит три значения d самых интенсивных линий. Используя их, можно проконтролировать наличие предполагаемых фаз в изучаемом материале.

После того как на первом этапе по ключу или алфавитному указателю выбраны справочные стандарты, проводится второй этап идентификации: сравниваются полные дифракционные спектры справочных стандартов и изучаемого вещества. Для удобства проведения идентификации следует всегда составлять таблицу, в первых двух колонках которой записывается дифракционный спектр анализируемого вещества $\{d_i, I_i\}$, а в последующих парах колонок — данные дифракционных спектров справочных стандартов.

На втором этапе идентификации могут возникнуть следующие варианты: совпадают все линии дифракционных спектров анализируемого вещества и справочного стандарта и все линии дифракционных спектров изучаемого вещества с несколькими справочными стандартами; часть линий дифракционного спектра изучаемого вещества не совпадает ни с какими спектрами справочных стандартов.

Первый вариант решения самый простой и ответ на вопрос о составе изучаемого вещества оказывается практически однознач-

вым. Оно является однофазным и ему следует приписать название вещества справочного стандарта. Во втором варианте решения задачи приходим к выводу, что изучаемое вещество является смесью нескольких (двух или более) кристаллических фаз. Их названия определяются по названиям справочных стандартов, линии спектров которых совпали с линиями спектра и их интенсивностями анализируемого вещества. Третий вариант решения указывает на то, что в смеси присутствует фаза (или фазы), для которой в ключе нет необходимых данных. В этом случае следует попытаться выделить предполагаемую фазу (или фазы) и тщательно проанализировать ее свойства и рентгеновский спектр. Если будет доказана индивидуальность этого вещества, его дифракционный спектр становится паспортной характеристикой нового справочного стандарта.

Как пример использования методики качественного рентгенофазового анализа В. И. Михеева ниже дан анализ образца рудничной пыли. Известно, что горная порода, при бурении которой образовалась пыль, содержит полевой шпат, карбонаты и кварц. Из-за высокой дисперсности вещества пробы (10—0,1 мкм) не было возможности выполнить детальный микроскопический анализ. На рентгенограмме образца пыли, снятой на излучении CoK в камере 57,3 мм, было около 40 линий (табл. 2.1).

Знание примерного минерального состава горной породы позволило на первом этапе идентификации воспользоваться алфавитным указателем работы [49] для поиска идентификаторов (номеров карточек справочных стандартов кварца, карбоната и полевого шпата). Из этого указателя для кварца установлен номер карточки полного дифракционного спектра — 256. Затем сопоставили спектр справочного стандарта — кварца — с дифракционными данными образца, вписав его в таблицу. Из табл. 2.1 видно, что 18 сильных и средних по интенсивности линий дифракционного спектра кварца соответствуют линиям в спектре анализируемой пробы, что подтверждает его присутствие в пробе. После исключения этих линий из оставшейся части дифракционного спектра пробы были выбраны три наиболее интенсивные линии, а именно: 3,20(8), 3,52(4) и 4,10(5). По ключу найдено, что они соответствуют линиям дифракционного спектра альбита: 3,21(10), 4,11(6) и 3,52(7), карточка № 666. Сопоставление полного дифракционного спектра альбита с данными анализируемой пробы (см. табл. 2.1) показало, что 26 линий совпадают. Присутствие альбита в пробе вне сомнений.

Исходя из предварительных данных о составе породы, из которой взят образец, дальнейший поиск вели среди карбонатов по карточкам, приведенным в [49]. Для этого из оставшегося дифракционного спектра выделили линии 2,88(5), 4,10(5), 2,19(2) и др. Оказалось, что лучше всего из карбонатов согласуются с ними дифракционные спектры доломита (карточки № 425 и 425а).

Для оставшихся неидентифицированных линий 4,72(4), 3,52(4), 2,00(3), 1,564(4), 1,390(2), 1,226(5) по ключу было найдено, что

Результаты расшифровки дифракционного спектра рудничной пыли

Проба		Кварц		Доломит		Альбит		Клинохлор	
<i>d</i>	<i>I</i>	<i>d</i>	<i>I</i>	<i>d</i>	<i>I</i>	<i>d</i>	<i>I</i>	<i>d</i>	<i>I</i>
4,72	4	—	—	—	—	—	—	4,715	7
4,23	5	4,24	5	—	—	—	—	—	—
4,10	5	—	—	4,10	3	4,11	6	—	—
3,72	2	—	—	—	—	3,70	3	—	—
3,67	1	—	—	3,683	2	—	—	—	—
3,52	4	—	—	—	—	3,52	7	3,530	10
3,33	10	3,34	10	—	—	3,33	1	—	—
3,20	8	—	—	3,174	3	3,21	10	—	—
2,94	1	—	—	—	—	2,955	6	—	—
2,88	5	—	—	2,883	10	2,862	2	2,812	8
2,69	1	—	—	2,664	2	—	—	2,684	1
2,54	4	—	—	2,531	1	2,55	4	2,531	7
2,44	6	2,45	6	—	—	2,446	4	2,438	6
2,39	1	—	—	—	—	2,396	2	2,378	5
2,27	4	2,280	5	—	—	2,311	4	—	—
2,23	1	2,231	4	—	—	—	—	—	—
2,19	2	—	—	—	—	—	—	2,191	5
2,12	4	2,123	5	—	—	2,116	4	—	—
2,068	1	—	—	2,062	2	2,080	3	2,070	2
2,000	3	—	—	2,015	4	2,013	4	1,998	9
1,970	3	1,975	4	—	—	1,974	3	—	—
1,882	3	—	—	—	—	1,887	5	1,879	5
1,809	6	1,813	9	—	—	1,821	5	—	—
1,662	4	1,668	5	—	—	1,663	3	1,659	5
1,564	4	—	—	1,564	1	1,567	2	1,564	9
1,537	6	1,539	9	1,543	3	—	—	1,535	10
1,390	2	—	—	—	—	1,385	2	1,393	10
1,377	6	1,380	8	—	—	1,374	2	—	—
1,369	5	1,372	9	—	—	—	—	—	—
1,347	1	—	—	1,335	3	1,347	6	1,347	2
1,282	3	1,285	6	—	—	1,278	4	1,280	4
1,251	4	1,253	6	—	—	1,253	2	—	—
1,226	5	1,226	5	—	—	1,224	6	1,220	8
1,197	6	1,198	8	—	—	—	—	1,191	2
1,181	5	1,182	8	—	—	—	—	1,182	3
1,152	3	1,151	7	—	—	—	—	—	—
1,111	1	—	—	1,110	5	1,108	3	—	—

они ближе всего соответствуют линиям дифракционного спектра клинохлора (карточка № 859 — минерал из группы хлоритов). Сопоставление значений *d* и *I* клинохлора с дифракционным спектром изучаемой пробы позволило высказать суждение, что четвертым компонентом в этой пробе, вероятнее всего, является клинохлор. Но для однозначного решения лучше это подтвердить другими методами.

Итак (см. табл. 2.1), дифракционный спектр изученного вещества рудничной пыли вполне удовлетворительно согласуется со спектрами справочных стандартов кварца, доломита и альбита и в меньшей степени — клинохлора.

2.2.2. Методика рентгенофазового анализа с использованием указателя Ханавальта [146]

Идентификация изучаемого вещества по дифракционными спектрам при использовании этой методики также осуществляется в два этапа. Из таблицы $\{d_i, I_i\}$ выбирается пара самых интенсивных межплоскостных расстояний d . Поиск начинается с выбора «группы» в указателе Ханавальта, которая включает первое сильное значение d . Это d находится в первой колонке указателя, определяя тем самым «группу». Значение d второй сильной линии дифракционного спектра определяет «подгруппу». Оно отыскивается во второй колонке значений межплоскостных расстояний выбранной группы в указателе Ханавальта. Пары значений d и I позволяют сузить поиск по указателю. Для увеличения надежности идентификации в указателе приводится еще шесть линий каждого справочного стандарта, которые располагаются в строке по убыванию значения I . Наличие в указателе восьми значений d для каждого справочного стандарта сокращает их число в процедуре идентификации на первом ее этапе. При этом, если восемь значений d найденного справочного стандарта согласуются с линиями дифракционного спектра изучаемого вещества, идентификация является достаточно надежной.

Конечная ступень в этой процедуре состоит в сопоставлении полных дифракционных спектров анализируемого вещества и справочного стандарта из картотеки JCPDS, карточка которого находится по номеру, указанному в конце выбранной строки указателя Ханавальта.

Рассмотрим использование этой методики на примерах для одно- и многофазного материала.

Однофазное вещество

В табл. 2.2 приведены в упорядоченной форме значения d и I изучаемого вещества (две первые колонки). Выбираем пару сильнейших по интенсивности линий: 3,34 и 4,25. Принимая

Таблица 2.2

Результаты идентификации изучаемого вещества

Изучаемое вещество		Справочный стандарт		Изучаемое вещество		Справочный стандарт	
I	d	d	I	I	d	d	I
38	4,25	4,26	35	7	1,670	1,672	7
100	3,340	3,343	100	2	1,658	1,659	8
14	2,455	2,458	12	15	1,538	1,541	15
14	2,278	2,282	12	2	1,451	1,453	3
7	2,235	2,237	6	6	1,381	1,382	7
8	2,125	2,128	9	12	1,374	1,375	11
5	1,978	1,980	6	8	1,371	1,372	9
16	1,815	1,817	17	4	1,286	1,288	3

										File No.	I/I_c
i	3.33 _X	4.28 _X	3.24 _X	8.58 ₈	2.91 ₈	2.37 ₈	1.87 ₈	4.95 ₆	Cs _{0.35} V ₂ O ₈	30—385	
→ *	3.34 _X	4.27 ₄	3.19 ₂	2.70 ₂	7.28 ₂	4.91 ₂	1.82 ₂	3.13 ₁	CaAl ₂ Si ₂ O ₈ ·4H ₂ O	20—452	
*	3.30 _X	4.27 ₇	3.71 ₇	3.45 ₇	3.92 ₅	3.34 ₅	3.27 ₅	2.96 ₅	KFeSi ₃ O ₈	16—153	
	3.38 _X	4.26 _X	4.16 _X	1.68 _X	6.70 ₈	2.97 ₈	2.86 ₈	2.56 ₈	Cs ₂ VCl ₅ ·4H ₂ O	20—301	
→	3.34 _X	4.26 ₉	3.66 ₉	2.61 ₉	2.48 ₉	1.80 ₉	3.97 ₈	3.05 ₈	3PbCO ₃ ·2Pb(OH) ₂ ·2H ₂ O	9—356	
→ i	3.34 _X	4.26 ₈	2.13 ₈	7.40 ₆	2.57 ₆	2.03 ₆	3.49 ₄	2.24 ₄	Ca ₃ Mn(SO ₄) ₂ (OH) ₆ ·3H ₂ O	20—226	
→ i	3.34 _X	4.26 ₇	2.13 ₆	7.40 ₅	3.49 ₅	2.58 ₅	2.24 ₅	2.21 ₅	Ca ₃ Ge(SO ₄) ₂ (OH) ₄ ·4H ₂ O	19—225	
→ *	3.34 _X	4.26 ₄	1.82 ₂	1.54 ₂	2.46 ₁	2.28 ₁	1.38 ₁	2.13 ₁	α-SiO ₂	5—490	← 3.60
	3.30 _X	4.26 ₈	4.24 ₇	3.66 ₇	2.89 ₆	2.30 ₆	2.63 ₃	2.32 ₅	Sm ₂ O(CO ₃) ₂ ·H ₂ O	28—994	
i	3.31 ₉	4.25 _X	4.09 ₉	3.72 ₇	2.98 ₇	2.20 ₇	1.93 ₇	5.47 ₆	ClPtOF ₈	26—415	
i	3.30 ₉	4.25 _X	4.09 ₉	3.67 ₉	7.42 ₆	5.47 ₆	2.96 ₆	1.91	(ClF ₂ O)AsF ₆	26—1117	
	3.40 _X	4.24 ₆	3.00 ₄	5.72 ₃	2.65 ₃	2.85 ₂	2.37 ₂	0.00 ₁	KAuF ₄	27—395	
*	3.37 _X	4.24 ₃	2.32 ₂	2.12 ₂	2.45 ₁	1.84 ₁	1.83 ₁	1.54 ₁	GaPO ₄	8—497	
o	3.37 _X	4.23 _X	3.46 _X	3.00 _X	2.13 _X	6.00 ₈	2.69 ₆	2.45 ₆	CsSnFBrJ	30—360	
o	3.36 _X	4.23 ₅	1.64 ₅	2.72 ₄	2.44 ₃	2.22 ₃	1.93 ₃	3.14 ₂	Mo ₃ O ₈ ·xH ₂ O	21—574	
→ *	3.35 ₈	4.23 _X	4.45 ₇	1.80 ₇	2.29 ₆	3.07 ₅	1.67 ₂	1.59 ₂	Na ₂ SiF ₆	19—1193	
→	3.35 ₈	4.23 _X	3.12 _X	3.43 ₈	3.84 ₆	3.75 ₆	3.39 ₆	3.37 ₆	β-Cs ₂ Cu(SO ₄) ₂	28—288	
	3.41 ₈	4.21 _X	3.11 _X	2.75 ₈	7.58 ₇	5.60 ₇	4.46 ₇	3.31 ₇	Na ₂ SbF ₃ SO ₄	28—1038	

Рис. 2.1. Фрагмент указателя Ханавальта картотеки JCPDS. Стрелками отмечены вероятные стандарты.

во внимание экспериментальную погрешность определения меж-
плоскостных расстояний (например, $\Delta d = \pm 0,01 \text{ \AA}$), находим
группу, в которой производится поиск дифракционного спектра
справочного стандарта. Линия $3,34 (\pm 0,01) \text{ \AA}$ находится в группе
 $3,39 - 3,32$. Вторая сильнейшая линия $4,25 (\pm 0,02) \text{ \AA}$ определяет
подгруппу поиска в пределах выбранной группы по первой линии.

В указателе (рис. 2.1) находим семь вариантов, удовлетворяю-
щих выбранной паре линий. Испытывая каждый из них, выбираем
только один, в котором остальные шесть линий в ключе хорошо
согласуются с линиями дифракционного спектра изучаемого ве-
щества ($\alpha - \text{SiO}_2$, карточка 5-490). Для выполнения второго эта-
па идентификации извлекаем из картотеки эту карточку и сопо-
ставляем дифракционные спектры вещества и справочного стан-
дарта. Результаты сопоставления даны в табл. 2.2. На этом осно-
вании заключаем, что изучаемое вещество образовано одной фа-
зой — кварцем.

Многофазный образец

Для удобства работы и избежания пропусков и по-
вторений составляется таблица $\{d_i, I_i\}$, в первых двух колонках
которой помещается дифракционный спектр изучаемого вещества
(см. табл. 2.3) [146]. Оцениваются экспериментальные погреш-
ности значений d , с тем чтобы правильно определить группу в ука-
зателе Ханавальта. В рассматриваемом примере принята доста-
точной погрешность $\pm 0,01 \text{ \AA}$, хотя для $d \geq 4 \text{ \AA}$ и более вероятно
допустить $\Delta d = \pm 0,02 \text{ \AA}$.

Для поиска в указателе Ханавальта из дифракционного спек-
тра изучаемого образца (табл. 2.3) выбираем пару самых интенсив-
ных линий, а именно $3,41 (\pm 0,01)$ и $5,00 (\pm 0,02) \text{ \AA}$. По первой
линии находим в ключе группу $3,49 - 3,40 \text{ \AA}$, а по второй — под-
группу. В подгруппе $5,00 (\pm 0,02) \text{ \AA}$ имеется три удовлетвори-
тельных варианта:

3,41 _x	5,02 _x	6,63 _s	4,87 _s	2,71 _s	2,60 _s	5,91 _s	5,88 _s	Na ₃ AsO ₂ S ₂	12-1301
3,41 _x	5,00 ₇	4,01 ₇	3,23 ₇	6,00 ₅	3,06 ₅	3,72 ₃	2,50 ₃	Mg(H ₂ PO ₄) ₂	3-395
3,40 _x	5,01 _x	2,89 ₅	2,20 ₄	4,62 ₄	2,45 ₃	2,10 ₃	1,75 ₃	K ₂ PtF ₆	7-236

Однако в дифракционном спектре изучаемого вещества находятся
пары всем восьми линиям лишь для справочного стандарта K₂PtF₆
(карточка 7-236).

При сопоставлении его полного дифракционного спектра со
спектром изучаемого вещества найдено 11 парных линий (см.
табл. 2.3). Таким образом, после второго этапа идентификации
приходим к заключению, что в изучаемом веществе присутствует

Результаты идентификации анализируемого вещества

Исследуемое вещество		K_2PtF_6 , $N_{JCPDS} = 7-236$		$CdOHCl$, $N_{JCPDS} = 18-259$		$Ca_{12}Al_{14}O_{33}$, $N_{JCPDS} = 9-413$		$NaMnF_3$, $N_{JCPDS} = 18-1224$	
<i>I</i>	<i>d</i>	<i>I</i>	<i>d</i>	<i>I</i>	<i>d</i>	<i>I</i>	<i>d</i>	<i>I</i>	<i>d</i>
81	5,12			100	5,11				
97	5,00	100	5,01						
76	4,90					100	4,89		
34	4,61	40	4,62						
71	4,01							100	4,00
12	3,78					16	3,79		
100	3,409	100	3,40						
20	3,208					25	3,204		
17	3,180			20	3,18				
85	3,041			100	3,03				
36	3,007					50	3,000		
47	2,901	50	2,89						
22	2,889							30	2,88
75	2,839							100	2,83
15	2,787							20	2,78
17	2,710			20	2,70				
80	2,673					100	2,680		
23	2,561			10	2,56	18	2,556		
14	2,511	6	2,501						
63	2,449	30	2,45			50	2,450		
7	2,435							10	2,433
15	2,406							20	2,405
9	2,384							12	2,386
7	2,348					10	2,350		
68	2,326			80	2,33				
46	2,207	40	2,20						
32	2,186					40	2,190		
6	2,150							8	2,182
24	2,056	30	2,10						
7	2,050					10	2,054		
45	1,997							60	2,000
7	1,990			8	1,992				
24	1,936					30	1,940		
11	1,887	12	1,890						
6	1,846					8	1,850		
21	1,829			30	1,830				
18	1,809	8	1,807						
22	1,787							30	1,790
6	1,764					8	1,767	8	1,767
23	1,746	30	1,750						

фаза K_2PtF_6 . Из сравнения интенсивностей совпавших линий видно, что сильная линия 2,449 лишь наполовину объясняется интенсивностью линии 2,45(30) справочного стандарта. По всей вероятности, она получилась в результате наложения линий разных фаз с очень близкими межплоскостными расстояниями.

Из оставшихся неидентифицированных линий вещества (см. табл. 2.3, колонки 1 и 2) снова выделяем пару наиболее интенсив-

ных линий: $3,04(\pm 0,01)$ и $5,12(\pm 0,02)$ Å. Обращаемся к группе $3,04-3,00$ и подгруппе $5,12(\pm 0,02)$ в указателе Ханавальта. Находим только один вариант справочного стандарта:

$3,03_X$ $5,11_X$ $2,33_8$ $1,83_3$ $3,18_2$ $2,70_2$ $1,73_2$ $2,56_1$ CdОНCl 18-259

В дифракционном спектре изучаемого вещества находятся парные линии найденного справочного стандарта — CdОНCl. Сопоставление его полного дифракционного спектра с данными изучаемого вещества (см. табл. 2, 3, колонки и 5 и 6) убеждает нас, что это соединение является второй фазой изучаемого вещества.

После второго цикла идентификации в дифракционном спектре изучаемого вещества остался неидентифицированным ряд интенсивных линий. Выбираем из них пару наиболее ярких линий: $2,67(\pm 0,01)$ и $4,90(\pm 0,02)$ Å. По первой линии находим группу $2,69-2,65$, а по второй — десять подходящих вариантов, отвечающих этой паре линий.

$2,67_X$	$4,92_6$	$3,56_6$	$3,18_6$	$1,97_6$	$9,92_4$	$3,39_2$	$2,46_2$	$V_2WO_{7,5}$	24-1396
$2,69_X$	$4,91_9$	$3,22_7$	$0,77_8$	$3,02_6$	$1,95_6$	$2,20_5$	$1,67_5$	$Tb_3Al_2(AlO_4)_3$	17-735
$2,69_X$	$4,91_4$	$2,36_4$	$3,22_3$	$1,95_3$	$1,67_3$	$2,20_3$	$1,61_3$	$Dy_3Al_2(AlO_4)_3$	22-1093
$2,68_X$	$4,90_8$	$3,21_6$	$3,00_6$	$2,45_6$	$2,19_6$	$1,95_6$	$0,00_1$	$Al_5Y_3O_{12}$	30-51
$2,68_X$	$4,89_5$	$3,21_4$	$1,95_4$	$1,67_4$	$2,19_4$	$3,00_3$	$2,45_3$	$Al_{10}Ho_6O_{24}$	32-16
$2,68_8$	$4,89_X$	$2,99_4$	$2,44_4$	$2,19_3$	$3,20_2$	$1,66_2$	$1,60_2$	$11CaO \cdot 7Al_2O_3 \cdot CaF_2$	25-394
$2,68_X$	$4,89_X$	$2,45_5$	$3,00_5$	$2,19_4$	$1,95_3$	$1,66_3$	$1,60_3$	$Ca_{12}Al_{14}O_{33}$	9-413
$2,67_X$	$4,89_X$	$4,47_9$	$1,85_8$	$6,00_8$	$3,69_8$	$2,44_8$	$3,79_7$	$Na_3Co(NH_3)_2N_3$	11-54
$2,67_X$	$4,88_9$	$3,20_7$	$2,99_6$	$1,94_6$	$2,18_5$	$1,66_5$	$4,23_4$	$Tm_3Al_2(AlO_4)_3$	17-734

При сопоставлении остальных шести линий каждого из этих вариантов справочных стандартов сразу выпадают пять (строки 1—3, 8 и 9). Оставшиеся четыре варианта очень схожи (строки 4—7), и поэтому для них необходима проверка согласованности полных дифракционных спектров со спектром изучаемого вещества. Из сопоставления полных дифракционных спектров этих справочных стандартов (карточки 30-51, 32-16, 25-394 и 9-413) по причине наибольшей близости со спектром изучаемого вещества было отдано предпочтение справочному стандарту — $Ca_{12}Al_{14}O_{33}$ (карточка 9-413). Из табл. 2.3 видно, что более 12 линий его согласуются с данными спектра изучаемого вещества. Отметим, что линии $d = 2,561$ и $d = 2,449$ дифракционного спектра вещества не полностью согласуются по интенсивности с соответствующими линиями справочных стандартов CdОНCl и $Ca_{12}Al_{14}O_{33}$. Наложение их в спектре изучаемого вещества.

После трех циклов идентификации в дифракционном спектре изучаемого вещества остался еще ряд неидентифицированных линий, по которым был продолжен поиск. Среди них наиболее ин-

тенсивные линии $2,84(\pm 0,01)$ и $4,01(\pm 0,02)$ Å. Первая находится в группе $2,84-2,80$. Вторая линия, определяющая подгруппу поиска вариантов справочных стандартов, указала на четырнадцать вариантов, но только один из них по остальным шести линиям согласуется с данными спектра анализируемого вещества:

$2,83_x$ $4,00_x$ $2,00_6$ $1,61_4$ $2,88_3$ $1,79_3$ $1,65_3$ $2,78_2$ NaMnF_3 18-1224

После записи в табл. 2.3 линий дифракционного спектра из карточки этого справочного стандарта видно, что все линии изучаемого вещества удовлетворительно согласуются с линиями спектра данного справочного стандарта. На этом основании заключаем, что изучаемое вещество, представленное его дифракционным спектром, является смесью четырех фаз: K_2PtF_6 , CdOHCl , $\text{Ca}_{12}\text{Al}_{14}\text{O}_{33}$ и NaMnF_3 .

Таким образом, для выполнения процедуры идентификации потребовалось провести четыре цикла, проанализировав при этом 27 вариантов поиска по ключу Ханавальта. Рассмотренная методика ручного поиска была основана только на порошковых дифракционных данных, так как отсутствовала другая информация об изученном веществе.

2.2.3. Методика рентгенофазового анализа с применением указателя Финка [147]

Возникновение указателя Финка в картотеке JCPDS для идентификации веществ по их дифракционным спектрам обусловлено двумя причинами, ограничивающими использование рассмотренных выше методик: 1) сильное различие интенсивностей линий дифракционных спектров на рентгено- и электронограммах, вызываемое предпочтительной ориентировкой кристаллитов в препаратах для электронографии; 2) искажение интенсивности линий в дифракционных спектрах от многофазных веществ, связанное с наложением линий с близкими межплоскостными расстояниями от разных фаз.

В отличие от указателей М. И. Михеева [49] и Дж. Д. Ханавальта [146] у Финка [147] расположение ключевых линий дифракционных спектров справочных стандартов определяется только числовыми значениями и совершенно не зависит от их относительных интенсивностей.

Процедура пользования указателем относительно проста и достаточно прямолинейна. Она включает следующие операции. Составляется таблица с дифракционным спектром изучаемого вещества вида $\{d_i, I_i\}$. Из спектра выбирается пара значений межплоскостных расстояний высокой интенсивности d_s и d_{s-1} для использования их на первом этапе поиска. Для простоты в качестве d_s принимается наибольшее по величине значение d , имеющее большую интенсивность. По значению d_s в первой колонке указателя выбирают группу, учитывая при этом эксперименталь-

ные погрешности. Во второй колонке указателя по значению d_{s-1} определяется подгруппа и после этого сопоставляются межплоскостные расстояния остальных шести линий и интенсивностей справочного стандарта с линиями спектра изучаемого вещества. После достижения предварительной идентификации (и тем самым определения идентификатора — номера карточки в картотеке JCPDS) проводится сопоставление полного спектра справочного стандарта со спектром изучаемого вещества. Если на первом этапе не достигается предварительная идентификация, то: а) производится поиск подгруппы по другим значениям межплоскостных расстояний из дифракционного спектра изучаемого вещества, например, d_{s-2} , d_{s-3} и т. д., б) выбирается новое значение, и процедура поиска повторяется сначала. Если дифракционный спектр изучаемого вещества содержит немного линий, которые не сопоставляются с ключевыми линиями справочных стандартов, то следует при поиске использовать ранее проидентифицированные линии из спектра изучаемого вещества и повторять указанные выше операции до тех пор, пока не будет достигнуто соответствие всех линий спектров изучаемого вещества и справочных стандартов.

Из сказанного выше следует, что методика идентификации с помощью указателя Финка также может быть полициклической, как и методики В. И. Михеева и Дж. Д. Ханавальта. Рассмотрим применение этой методики на примере для идентификации многофазного образца. Дифракционный спектр изучаемого образца, подлежащего идентификации, представлен в табл. 2.4.

Выбираем из этого спектра пару линий: $d_s = 3,26 (\pm 0,01) \text{ \AA}$ и $d_{s-1} = 3,15 (\pm 0,01) \text{ \AA}$. По первой линии находим в указателе Финка группу 3,31—3,25, а по второй отыскиваем подгруппу. Однако при сопоставлении остальных значений d ключевых линий справочных стандартов из этой подгруппы с межплоскостными расстояниями изучаемого вещества получаем отрицательный результат. Из этого следует, что выбор пар линий сделан неправильно; эти линии принадлежат разным фазам.

Составляем новую пару $d_s = 3,26 (\pm 0,01) \text{ \AA}$ и $d_{s-1} = 2,82 (\pm 0,01) \text{ \AA}$ и повторяем поиск справочного стандарта в указателе. По первой линии находим группу 3,31—3,25, вторая дает нам подгруппу. Сравниваем остальные шесть ключевых линий справочных стандартов с линиями дифракционного спектра изучаемого вещества, результаты записываем в табл. 2.4 (колонка 3). Эти ключевые линии принадлежат NaCl (карточка $N_{\text{JCPDS}} 5-628$). Извлекаем данную карточку из картотеки JCPDS и сопоставляем полный спектр справочного стандарта со спектром изучаемого вещества. Из результатов этого сопоставления (см. табл. 2.4, колонки 4 и 5) следует, что первой фазой в анализируемом веществе является NaCl, так как все линии из карточки имеют пару в спектре изучаемого вещества.

Из оставшихся линий спектра вещества выбираем другую пару

Результаты идентификации образца

Изучаемое вещество		Указатель Финка		5-628	Указатель Финка		4-587
<i>I</i>	<i>d</i>	<i>d_I</i>	<i>d</i>	<i>I</i>	<i>d_I</i>	<i>d</i>	<i>I</i>
13	3,26	3,26 ₁	3,258	13			
100	3,15				3,15 _X	3,146	100
100	2,82	2,82 _X	2,821	100			
60	2,22		Нет пары		2,22 _o	2,224	59
55	1,99	1,99 _o	1,994	59			
23	1,82				1,82 ₂	1,816	23
2	1,70		1,701	5			
15	1,63	1,63 ₂	1,628	15			
8	1,57				1,57 ₁	1,573	8
20	1,41	1,41 ₁	1,410	6	1,41 ₂	1,407	20
1	1,29		1,294	1			
13	1,28				1,28 ₁	1,284	13
11	1,26	1,26 ₁	1,261	11			
7	1,15	1,15 ₁	1,1515	7			
2	1,11					1,1126	2
2	1,09		1,0855				
6	1,05				1,05 ₁	1,0490	6
2	0,997		0,9969	2			
2	0,995					0,9951	2
11	0,953		0,9533	1			
	0,949					0,9486	3
	0,940		0,9401	3			
	0,908					0,9083	1
	0,892	0,892 ₁	0,8917	4			
	0,873					0,8727	2
	0,860		0,8601	1			
	0,850		0,8503	3			
	0,841				0,84 ₁	0,8410	6
	0,829						
	0,814		0,8141	2			

линий с наибольшими значениями d : $d'_s = 3,15 (\pm 0,01) \text{ \AA}$ и $d'_{s-1} = 2,22 (\pm 0,01) \text{ \AA}$. Действуя уже известным способом, находим, что этим линиям соответствуют ключевые линии KCl — карточка JCPDS 4-587. Извлекаем ее из картотеки JCPDS и сопоставляем полный спектр найденного справочного стандарта с оставшимися неидентифицированными линиями спектра изучаемого вещества. Результаты сопоставления приведены в колонках 7 и 8 табл. 2.4. Итак, найдена вторая фаза изучаемого вещества.

Практически все линии дифракционного спектра изучаемого вещества проидентифицированы. Оно является смесью двух фаз: NaCl и KCl. Наличие Na, K и Cl подтверждается также качественным химическим анализом. Таким образом, имеется независимое подтверждение правильности рентгенофазового анализа этого вещества. Эти методики весьма рациональны и получили широкое признание у исследователей, занимающихся рентгенофазовым анализом.

ПОДГОТОВКА ДАННЫХ
ДЛЯ РЕНТГЕНОФАЗОВОГО АНАЛИЗА
(некоторые вопросы дифрактометрии)

Получение дифракционных спектров от поликристаллических веществ осуществляется с помощью соответствующей аппаратуры: рентгеновских камер Дебая — Шеррера, Гинье и др. при фоторегистрации и дифрактометров при регистрации с помощью детекторов излучения. Устройство камер и дифрактометров подробно рассмотрено в монографиях и учебных пособиях, перечисленных в начале главы 1, а также в технических описаниях, прилагаемых к установкам для рентгеноструктурного анализа. Новым этапом в развитии методов рентгенофазового анализа является использование дифрактометров, управляемых ЭВМ, которые обеспечивают сбор и обработку дифракционных спектров из изучаемых веществ на автоматизированном уровне. Благодаря дифрактометрам, управляемым ЭВМ, и методикам согласования профиля линий, в дифракционных спектрах создавалась благоприятная возможность повысить точность определения положения линий, вывести рутинный способ идентификации фаз на новый качественный уровень, повысить точность определения 2θ . Современные компьютерные методы интерпретации дифракционных спектров также направлены на преодоление трудностей, связанных с недостаточным разрешением рефлексов. Далее кратко рассматриваются вопросы подготовки и проведения эксперимента по регистрации рентгенограмм с целью уменьшения влияния различных факторов, действующих на точность и правильность определения положения и интенсивности линий дифракционных спектров изучаемых веществ и в особенности эмпирических справочных стандартов.

Качественный рентгенофазовый анализ расчлняется на следующие этапы: подготовка аппаратуры (ее юстировка и настройка), препарирование объектов исследования, оптимизация условий эксперимента, получение, расчет и обработка рентгенограмм. Юстировка и настройка дифрактометров, подготовка образцов к исследованию достаточно подробно описаны в литературе [20, 64, 84, 89, 127, 139, 152, 154, 159]. Ниже излагаются лишь некоторые вопросы дифрактометрии, необходимые для получения оптимальных условий эксперимента и правильной интерпретации дифракционных спектров, полученных при ручном рентгенофазовом анализе поликристаллических объектов.

3.1. ВЫБОР ОПТИМАЛЬНЫХ УСЛОВИЙ РЕГИСТРАЦИИ ДИФРАКЦИОННЫХ СПЕКТРОВ [29, 30, 32, 67, 71, 84, 127, 138, 159, 163—165, 168]

Качественный рентгенофазовый анализ базируется на точном определении положения и интенсивности линий дифракционных спектров. Особенности геометрии дифрактометров с фокусировкой по Брэггу — Брентано не обеспечивают одновременное выполнение этих требований. Приходится искать компромиссное решение, направленное на уменьшение действия противоречивых требований обеспечения высокой точности определения интенсивности и низкого порога обнаружения линий, с одной стороны, с прецизионностью и экспрессностью определения межплоскостных расстояний, с другой. Чтобы уменьшить статистические флуктуации счета при регистрации интенсивности дифракционных рефлексов, необходимо пользоваться широкими пучками, задавать большие постоянные времени и малые скорости сканирования. Для точного определения 2θ и достижения высокого разрешения линий при больших скоростях сканирования необходимы узкие пучки и малые постоянные времени.

Анализ проблемы учета влияния различных факторов на положение и интенсивность дифракционных линий с целью установления оптимальных условий регистрации дифракционной картины при рентгенофазовом анализе поликристаллических веществ показал, что наибольшее действие на смещение и искажение профиля линии оказывают горизонтальная и вертикальная расходимости первичного пучка рентгеновских лучей, проникновение излучения в глубь образца, конечное время зарядки и разрядки емкости интегрирующей цепочки интенсиметра, угловая скорость сканирования спектра. Из работ [13, 15, 97, 102—111] следует, что высоты ограничивающей и приемной щелей гониометра должны быть примерно равными ($H_f \approx H_{сч}$), а ширину приемной щели целесообразно выбирать примерно в два раза больше ширины проекции фокуса. Применительно к отечественным рентгеновским трубкам при угле выхода первичного пучка в 6° ширина приемной щели указана в табл. 3.1.

Из экспериментов следует, что увеличение ширины приемной щели вплоть до оптимального значения для каждого типа трубки не сопровождается заметным ростом интенсивности и не вызывает существенного искажения профиля линии.

Так как горизонтальная расходимость и длина освещенной части образца являются функцией угла дифракции, условия оптимальности интенсивности и формы профиля линии не одинаковы во всем интервале регистрации 2θ . В большинстве дифрактометров предусмотрено только ступенчатое изменение угла расходимости первичного пучка путем установки ограничивающих щелей разной ширины (мм): 0,25; 0,5; 1,0; 2,0; 4,0. Определение оптимальных условий регистрации дифракционной картины из-за быстрого изменения довольно сложно. Поэтому рекомендуется сту-

Характеристика рентгеновских трубок и оптимальная ширина приемной щели

Тип трубки	Мощность, кВт	Ширина проекции фокуса b_f , мм	Ширина приемной щели $b_{сч}$, мм	Фокус
БСВ-21	0,8	0,02	0,05	Острый
БСВ-22	1,2	0,04	0,10	
БСВ-23	1,5	0,10	0,25	Нормальный Широкий
БСВ-24	2,0	0,17	0,25	

печатое изменение ширины ограничивающей щели при постоянной ширине приемной щели в зависимости от области регистрации углов 2θ . Например, для трубок БСВ-23 и БСВ-24 с медным анодом при ширине приемной щели 0,25 мм в области 2θ от 10° до 20° следует использовать ограничивающую щель шириной 1 мм, для области $20-60^\circ$ $2\theta - 2$ и для области $60-100^\circ$ $2\theta - 4$ мм.

При изучении веществ, имеющих большие периоды идентичности, представляет интерес область углов $2\theta < 10^\circ$. Из-за ограниченной длины образца ($l_p \sim 25$ мм) едва ли можно достигнуть оптимальных условий регистрации без значительной потери светосилы прибора и, следовательно, существенного ограничения порога обнаружения фаз. Чтобы обеспечить хотя бы приблизительно оптимальные условия регистрации дифракционной картины в интервале $4-12^\circ$ 2θ , размер ограничивающей щели, определяющей l_p , выбирается из условия: проекция пучка не должна заметно превосходить размеры образца. Исходя из этого, рекомендуется использовать ограничивающую щель шириной 0,5 или даже 0,25 мм [67, 71].

Кроме рассмотренных выше инструментальных факторов, на оптимальные условия регистрации дифракционного спектра также влияют постоянная времени интенсиметра РС и скорость сканирования ω . При малых значениях этих величин резко возрастает время записи дифрактограммы и на ее качество заметное влияние оказывают флуктуации интенсивности, обусловленные статистической природой рентгеновского излучения. При больших значениях РС и ω происходят смещение профилей линий, их расширение и увеличение асимметричности. Для оценки оптимальной величины постоянной времени интенсиметра используется понятие времени смещения приемной щели на ее ширину W_t , которое прямо пропорционально ее ширине $\varphi_{сч}$ в градусах и обратно пропорционально скорости сканирования ω , т. е.

$$W_t = 60 \varphi_{сч} / \omega, \quad (3.1)$$

где $\varphi_{сч} = 57,3 \cdot b_{сч} / R$; $b_{сч}$ — ширина приемной щели, мм; R — радиус гониометра, мм. Для гониометра ГУР-8 ($R = 192$ мм) при

Оптимальные условия регистрации дифрактограмм от поликристаллических веществ для рентгенофазового анализа

Параметры прибора	ДРОН-3	ДРОН-1, ДРОН-2	УРС-50ИМ, ДРОН-0,5
Детектор	Сцинтилляционный или пропорциональный счетчик		
Рентгеновская трубка БСВ	23 и 24	22 и 23	6
Анод	Медь	Медь	Медь
Угол выхода первичного пучка, град	6	6	2—2,5
Напряжение, кВ	40	40	35
Ток, мА	30 40	25 30	12
Толщина фильтра, Ni, мкм		16—20	
Ограничивающая щель, мм			
ширина	2—4	2—4	2
высота	8—10	8—10	5—6
Щели Соулера, град	1,5	1,5	—
Приемная щель, мм			
ширина	0,25	0,1 0,25	0,25
высота	8—10	8—10	8
Скорость движения детектора, °/мин	2—4	2—4	1—0,5
Скорость протяжки ленты, мм/ч		600	
Диапазон интенсивметра, имп/мин	2000	2000	1000
Постоянная времени интенсивметра, с	2—5	2—4	4

$\omega = 2^\circ/\text{мин}$ и $b_{\text{сч}} = 0,1$ и $0,25$ мм соответственно $W_t = 0,9$ и $2,3$ с. Время смещения приемной щели на ее ширину служит косвенной характеристикой величины постоянной времени интенсивметра и указывает, что она должна быть малой. В работе [127] рекомендуется выбирать постоянную времени интенсивметра не более половины времени смещения приемной щели на ее ширину. В этом случае заряд и разряд емкости интегрирующей цепочки будут происходить достаточно быстро.

При постоянной интенсивности первичного пучка рентгеновских лучей и высоте ограничивающей и приемной щели их ширины определяют статистическую точность записи дифрагированной интенсивности на ленте самописца. Действительно, полное число квантов N , прошедших через приемную щель при сканировании малого углового интервала $\Delta 2\theta$, прямо пропорционально γ и $\varphi_{\text{сч}}$ и обратно пропорционально скорости сканирования ω , т. е.

$$N = k \frac{60 \cdot \varphi_{\text{сч}} \gamma}{\omega} \quad (3.2)$$

Относительное стандартное отклонение определяется величиной

$$\varepsilon_N = (N)^{-\frac{1}{2}} = \left(\frac{\omega}{k \cdot 60 \varphi_{\text{сч}} \gamma} \right)^{1/2} \quad (3.3)$$

пропорциональной скорости сканирования и обратно пропорциональной ширине ограничивающей и приемной щели. ε_N является мерой величины статистической ошибки при данных условиях ре-

гистрации дифракционного спектра. Таким образом, чтобы сохранить достаточным разрешение линий дифракционного спектра при высокой скорости сканирования, следует иметь ширину приемной щели $b_{сч} \approx 2b_f$ неизменной, но целесообразно увеличить ширину щели, ограничивающей первичный пучок.

Итак, в табл. 3.2 представлены рекомендации по оптимальным условиям регистрации дифрактограмм на отечественных дифрактометрах.

3.2. ОПРЕДЕЛЕНИЕ ПОЛОЖЕНИЯ ЛИНИЙ ДИФРАКЦИОННОГО СПЕКТРА

В рутинном рентгенофазовом анализе общепринятой является регистрация дифракционного спектра на ленту самописца. Положение линий на записи однозначно определяется с помощью некоторых реперов. В качестве реперов используются значения углов 2θ , соответствующие: 1) максимальной интенсивности профиля линии — $2\theta_{\text{макс}}$, 2) середине хорды, проведенной параллельно линии фона на некоторой высоте — $2\theta_{1/2}$ или 3) положению центра тяжести профиля линии — $\langle 2\theta \rangle$. На практике чаще всего используются два первых репера.

Нахождение центра тяжести или центроида $\langle 2\theta \rangle$ для локализации положения профиля линии связано с большими затратами труда и времени и пока не применяется в рутинном рентгенофазовом анализе. Несомненное достоинство центра тяжести линии как репера — определение 2θ с точностью $\pm 0,001^\circ$, так как погрешности поддаются математической обработке [164]. Положение вершины профиля линии $2\theta_{\text{макс}}$ определяется значительно быстрее, да и само это понятие более наглядно. Однако пока нет аналитического решения вопроса о погрешностях, вызывающих смещение вершины под действием различных факторов. Для результатов средней точности можно воспользоваться формулами поправок для центроидов [138, 163] или использовать семейства графиков эмпирических поправок, определенных для смещения пиков при различных условиях регистрации дифракционных спектров веществ с известными параметрами элементарной ячейки. В качестве реализации этой возможности Национальное бюро стандартов США предложило использовать ряд веществ, для которых определены с высокой точностью параметры элементарной ячейки: кремний [109, 112], синтетический фторфлогопит [110]. Для этой же цели можно рекомендовать кварц [34], некоторые силикаты с большим параметром c [71], стеарат свинца [118], α -стеариновую кислоту [23, 81] и др. Использование порошков этих веществ в качестве внутреннего стандарта позволяет учесть практически все погрешности и определить положение линий на дифрактограмме с погрешностью не более $\pm 0,01^\circ 2\theta$ в диапазоне углов 2θ от 3 до 140° .

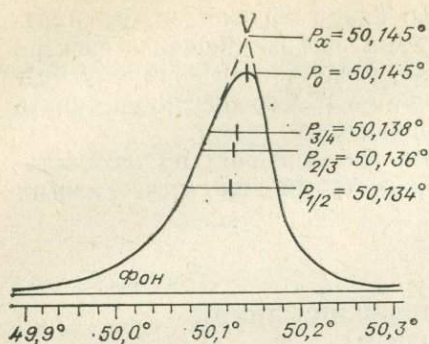


Рис. 3.1. Способы локализации положения профиля линий [147].

Вершину профиля линии $2\theta_{\text{макс}}$ на дифрактограмме можно определить: 1) по видимой ее точке P_0 или 2) по точке P_x пересечения двух касательных, проведенных к наиболее близким прямолинейным участкам профиля (рис. 3.1) [147]. Недостаток первого способа заключается в том, что погрешность вблизи этой точки статистически неопределенна, и случайные флуктуации интенсивности делают невозможной ее локализацию. В некоторой степени данная трудность преодолевается при использовании второго способа.

Использование репера $2\theta_{1/2}$ для локализации профиля линии на дифрактограмме обеспечивает большую точность и воспроизводимость результатов измерений. Хорда, обычно проводится на высоте $P_{1/2}$, $P_{2/3}$, $P_{3/4}$, $P_{4/5}$ от уровня фона. Лучше использовать два последних значения, особенно для линий неразрешенных $\alpha_1\alpha_2$ -дублетов.

Все эти способы локализации линий спектра приводят к несколько различающимся результатам, но величина разброса между ними не превышает $\Delta 2\theta \sim \pm 0,01^\circ$. Из сопоставления вариантов локализации положения профиля линий вытекает, что для измерения большего числа линий на дифрактограмме следует пользоваться «средней точкой хорды», проведенной на высоте не менее $P_{3/4}$. Однако ни один из рассмотренных способов не дает точного определения положения линии, свободного от субъективных качеств аналитика. Для исключения систематических погрешностей в определении 2θ профиля линий следует использовать съемку дифрактограмм с внутренним стандартом.

3.3. ОЦЕНКА КАЧЕСТВА ДИФРАКЦИОННЫХ СПЕКТРОВ ЭМПИРИЧЕСКИХ СПРАВОЧНЫХ СТАНДАРТОВ

В разделе 1.2 перечислены основные требования, которым должны отвечать дифракционные спектры справочных стандартов и отмечено, что содержащиеся в картотеках, определителях и оригинальных работах данные не в полной мере отвечают требованиям правильности, полноты и точности. С одной стороны, это связано с отсутствием количественных критериев, характеризующих эти понятия, а с другой — с отсутствием четких метрологических оценок, дающих объективную их характеристику.

Наиболее наглядным и надежным параметром достоверности дифракционного спектра является абсолютное среднее значение $|\Delta 2\theta| = \frac{1}{n_p} \sum |2\theta_{\text{изм}} - 2\theta_{\text{выч}}|$, для всех n_p разрешенных линий [93, 94], где $2\theta_{\text{изм}}$ и $2\theta_{\text{выч}}$ непосредственно получены из измерений и рассчитаны по параметрам элементарной ячейки углы дифракции. Причем в число n_p не включаются линии, образованные наложением более чем одного рефлекса, относительные интенсивности которых не установлены.

Другим параметром, характеризующим полноту дифракционного спектра эмпирического справочного стандарта, может служить отношение $n_{\text{набл}}/n_{\text{возм}}$, где $n_{\text{набл}}$ и $n_{\text{возм}}$ — соответственно число линий дифракционного спектра, полученного из эксперимента и рассчитанного по квадратичной форме, в соответствии с требованиями погасаний по пространственной группе. Близость этого отношения к единице является хорошей рекомендацией тому, что все возможные hkl и $2\theta_{\text{выч}}$ учтены. Если оно существенно больше единицы, нельзя с уверенностью утверждать, что дифракционный спектр принадлежит однофазному веществу. Однако в экспериментальном спектре линий всегда будет меньше, чем в расчетном, так как самые малоинтенсивные окажутся ниже порога обнаружения регистрирующей системы, либо часть слабых линий или линий, имеющих межплоскостные расстояния меньше некоторой величины (например, меньше $1,2 \text{ \AA}$), могут быть исключены из таблицы главным образом по формальным признакам, связанным с ограниченностью объема статьи.

Когда же не известны параметры элементарной ячейки, сингония и пространственная группа, определенные из монокристалльных исследований, невозможно оценить полноту дифракционного спектра вещества. Для множества кристаллических фаз отсутствуют пригодные для определения этих характеристик монокристаллы. Для преодоления этого препятствия Де-Вольф [166] предложил критерий надежности индцирования рентгendifракционных спектров от поликристаллических веществ в форме

$$M_{20} = \frac{Q_{20}}{2|\Delta Q|n_{20}}. \quad (3.4)$$

Здесь $Q_{20} = \frac{1}{d^2}$ — численное значение Q_{20} , зарегистрированного на рентгенограмме, $|\Delta Q| = \frac{1}{n_{20}} \sum |Q_{\text{набл}} - Q_{\text{выч}}|$ — абсолютное среднее наблюдаемых и вычисленных значений, Q_i и n_{20} — число теоретически возможных независимых линий вплоть до наблюдаемой двадцатой. Величина M_{20} будет большой, если $|\Delta Q|$ будет малым или когда n_{20} окажется близким к 20. Чем больше величина M_{20} , тем выше вероятность того, что интерпретация правильна. Предполагается, что величина критерия $M_{20} < 10$ указывает на сомнительный результат интерпретации рентген-

дифракционного спектра, а величина $M_{20} \geq 10$ будет гарантировать правильность индцирования спектра. Лестер и Липсон [128], анализируя надежность интерпретации рентгендифракционных спектров, указывают, что для повышения эффективности этого критерия надежности необходимо соблюдать следующие условия: 1) дифракционный спектр следует снимать с внутренним стандартом, для которого точно известны углы дифракции. Такая регистрация позволит исключить систематические погрешности, связанные с условиями эксперимента; 2) до начала индцирования спектра необходимо оценить точность определения Q_i ; 3) все экспериментальные величины Q_i должны согласовываться с их вычисленными значениями в пределах необходимой точности; 4) для определенных значений параметров элементарной ячейки и измеренной плотности число формульных единиц должно быть целым числом; 5) критерий M_{20} должен быть больше 10 или равным.

Несколько иной критерий надежности предложен Смитом и Снайдером [153], а именно

$$F_n = \frac{1}{|\Delta 2\theta|} \cdot \frac{n_{\text{набл}}}{n_{\text{возм}}} \quad (3.5)$$

где $n_{\text{набл}}$ и $n_{\text{возм}}$ — соответственно число зарегистрированных и возможных дифракционных линий вплоть до n -й и $|\Delta 2\theta| = \frac{1}{n} \sum |2\theta_{\text{изм}} - 2\theta_{\text{выч}}|$. Критерий F_n включает точность определения положения линий рентгенографического спектра (первый сомножитель) и полноту зарегистрированной дифракционной картины (второй сомножитель). Таким образом, чем выше точность или чем полнее зарегистрированный спектр, тем выше величина F_n . Единицей измерения F_n являются обратные градусы. Из этого следует, что если $F_n = 100$, то средний разброс $|\Delta 2\theta| \leq 0,01^\circ$, если $F_n = 50$, то $|\Delta 2\theta| \leq 0,02^\circ$.

Донней [93] считает, что более наглядными критериями достоверности дифракционных спектров являются не M_{20} или F_n , а набор величин, кстати входящих в F_n , а именно: $D_{n_p} = \frac{1}{n_p} \sum |2\theta_{\text{изм}} - 2\theta_{\text{выч}}|$, $n_{\text{возм}}/n_{\text{набл}}$ и n_p — число линий на рентгенограмме, не являющихся результатом наложения линий с близкими значениями 2θ . Причем D_{n_p} — основной критерий, позволяющий судить о качестве дифракционного спектра, причем разброс $\Delta 2\theta$ не должен превышать $3D_{n_p}$.

Эти критерии фактически не являются действительной мерой точности, так как значения $2\theta_i$, используемые для определения средних величин $\Delta 2\theta$, находятся из параметров элементарной ячейки, которые получают из уточненных по МНК экспериментально определенных углов. В действительности M_{20} , F_n и триада Донней — мера внутренней согласованности экспериментальных данных. Абсолютную точность в измерении углов 2θ можно

Оценка погрешности Δd для различных интервалов межплоскостных расстояний d и условий получения дифракционных спектров, Å

$d, \text{Å}$	Дифрактометр			Камера 57,3 мм
	НБС США	с внутренним стандартом	без внутреннего стандарта	
20—15	0,1—0,05	0,5—0,25	0,6—0,4	1,1—0,6
15—10	0,05—0,03	0,25—0,13	0,4—0,15	0,6—0,3
10—8	0,03—0,04	0,13—0,08	0,15—0,10	0,3—0,2
8—5	0,014—0,006	0,08—0,03	0,10—0,04	0,2—0,07
5—3	0,006—0,002	0,03—0,01	0,04—0,013	0,07—0,02
3—2	0,002—0,001	0,01—0,005	0,013—0,006	0,02—0,01
2—1,5	0,001—0,0005	0,005—0,0025	0,006—0,003	0,01—0,005
1,5—1,0	0,0005—0,0002	0,0025—0,001	0,003—0,001	0,005—0,002

достигнуть только при использовании внутреннего стандарта с точно известными параметрами элементарной ячейки.

Однако всегда следует помнить, что анализируемый дифракционный спектр плохого качества можно успешно сравнить с превосходящими стандартными справочными данными и, наоборот, всегда исследователь или аналитик будут испытывать неуверенность, если дифракционный спектр образца сравнивается со спектрами ненадежных справочных стандартов.

Фривел [101] для оценки качества дифракционных спектров предложил использовать признак достоверности — K . В основу его положено утверждение, что погрешность измерения углов θ зависит от характеристик используемых в эксперименте камер или дифрактометров и прежде всего их эффективного радиуса. Причем если $\Delta\theta = 0,0000$ рад, то $K = 0$. Отсюда, если $10^4 \cdot \Delta\theta = 1, 2, 4, 6, 8, 10, 12, 15, 20$ рад, соответственно $K = 1, 2, 3, 4, 5, 6, 7, 8, 9$. То есть чем меньше величина K , тем меньше $\Delta\theta$ и тем выше качество дифракционного спектра эмпирического справочного стандарта. Проведенный Фривелом анализ 26 сетов картотеки JCPDS показал, что данные Национального бюро стандартов США характеризуются $K = 2$ и $\Delta\theta \sim 0,01^\circ$, для дифрактограмм, снятых с внутренним стандартом, $K = 3-5$ и $\Delta\theta \sim 0,03-0,046^\circ$, для спектров из дифрактограмм, снятых без внутреннего стандарта, $K = 5-7$ и $\Delta\theta \sim 0,046-0,068^\circ$. Рентгенограммы, снятые в камере с $D = 57,3$ мм без внутреннего стандарта, характеризуются $K = 7-9$ и $\Delta\theta \sim 0,068-0,115^\circ$. На основе этих данных в табл. 3.3 приведены значения Δd для различных интервалов межплоскостных расстояний.

Сравнение этих данных с оценкой погрешности определения d , полученных из дифрактограмм [84], показывает, что оценки $\Delta\theta$ и соответствующих им Δd вполне пригодны для оценки погрешностей межплоскостных расстояний в ручном рентгенофазовом анализе.

Если определение положения линий 2 θ в дифракционном спектре поликристаллических веществ достаточно хорошо разработано

и имеются методики оценки их точности, то определение интенсивности линий является серьезной проблемой рентгенофазового анализа. По данным различных исследователей, величина погрешности в оценке интенсивности может достигать десятков процентов, существенно ограничивая надежность идентификации при использовании спектров $\{d_i, I_i\}$. Причем наибольший вклад в погрешность оценки интенсивности вносят предпочтительная ориентировка и вариации размера кристаллитов. Кроме того, заметную неопределенность в интенсивность вносит наложение линий, имеющих близкие значения 2θ .

Глава 4

СИСТЕМЫ АВТОМАТИЧЕСКОГО КАЧЕСТВЕННОГО РЕНТГЕНОФАЗОВОГО АНАЛИЗА

Приведенное в первых главах изложение методических основ ручного качественного РФА, описание используемых картотек справочных стандартов и аппаратуры позволяют перейти к подробному рассмотрению систем автоматического качественного рентгенофазового анализа. В настоящей главе дается обзор подобных систем, наиболее широко применяемых в международной и отечественной практике. Этот обзор послужит нам материалом для подробного анализа систем автоматического РФА, проведенного в главе 5, в ходе которого будут рассмотрены теоретические аспекты проблемы идентификации спектров.

4.1. ПРИМЕНЕНИЕ ЭВМ ДЛЯ РЕШЕНИЯ ЗАДАЧИ КАЧЕСТВЕННОГО РФА

Качественный РФА является, как подчеркивалось ранее, процедурой поиска и выбора (Search — match), цель которой — идентификация исследуемого спектра с суперпозицией стандартных спектров, разыскиваемых и выбираемых из картотеки справочных данных. Из-за большого объема картотеки процедура идентификации часто оказывается трудоемкой (даже при наличии облегчающих задачу поиска указателей типа Дэви, Ханавальта и Финка). Появление электронно-вычислительных машин позволило с середины 60-х годов перейти к интенсивной раз-

работке методов и программ автоматической идентификации рентгендифракционных спектров. При этом усилия исследователей были направлены на решение двух основных проблем:

1) представление в удобном машинном формате содержащихся в различных картотеках и определителях стандартных порошковых данных и их накопление в ЭВМ, т. е. формирование машинной базы данных;

2) логическая и математическая формализация методов качественного РФА и создание эффективных алгоритмов автоматической идентификации.

Большинство карточек JCPDS содержит 0,5—1 килобайт информации; в связи с этим формирование машинной базы данных объемом в несколько тысяч соединений требует оперирования с массивами данных (файлами) в несколько мегабайт. Для хранения таких массивов используются внешние запоминающие устройства ЭВМ — накопители на магнитных лентах и дисках (НМЛ, НМД). При проведении процесса поиска и сравнения информация считывается с магнитных накопителей в оперативную память ЭВМ определенными порциями (блоками). Блок может содержать одну карточку или несколько, так как разместить в оперативной памяти машины всю базу данных на несколько тысяч соединений невозможно даже для наиболее мощных современных универсальных ЭВМ. Таким образом, автоматическая идентификация с использованием достаточно крупной машинной картотеки сводится к просмотру базы данных по частям, т. е. связана с непрерывным и многократным чтением информации с НМД или НМЛ. Обмен информацией между магнитным накопителем и оперативной памятью ЭВМ — относительно длительный процесс, который определяет в конечном счете быстродействие программ автоматической идентификации.

Рассмотренное выше обстоятельство привело к значительному упрощению исходного формата JCPDS и других подобных форматов (см. гл. 1) при переносе информации из соответствующих картотек в машинные базы данных. Практически все системы автоматической идентификации используют базы данных специального формата и структуры, которые позволяют увеличить быстродействие процедуры поиска и выбора. Часто база данных состоит из нескольких файлов последовательно усложняющейся структуры, которые используются на начальном, промежуточных и окончательном уточняющем этапах идентификации. Необходимо отметить, что возникновение идеи о построении процедуры идентификации в виде многоэтапного процесса не связано с использованием ЭВМ; она заложена уже в методике ручного качественного анализа, когда предварительный отбор фаз осуществляется с помощью указателей Дэви, Ханавальта и Финка, а окончательный этап включает сопоставление исследуемого спектра с отобранными стандартами по картотеке JCPDS. Структура баз данных некоторых систем идентификации построена в соответствии с традиционной методикой ручного анализа: первый файл аналогичен указа-

телям Ханавальта и Финка и содержит 3—8 сильнейших линий стандарта, второй файл — картотеке JCPDS и содержит 20—40 линий стандарта и более.

Переход к упрощенному машинному формату осуществлялся в первых системах автоматической идентификации, как правило, путем исключения из карточки JCPDS всех данных, кроме ее номера, спектра (т. е. набора $\{d_i, I_i\}$) и иногда химической формулы или элементного состава. Опыт эксплуатации первых систем показал, что крайне желательно сохранить в машинной базе данных также информацию о названии соединения, признаке достоверности данных и некоторые другие сведения.

Как будет описано дальше, формат и файловая структура баз данных тесно связаны с алгоритмами процедуры поиска и выбора. Наиболее важными для исследователя моментами процесса автоматической идентификации является возможность быстрого просмотра базы данных и получения численной характеристики степени совпадения (меры сходства) стандарта с изучаемым спектром. Предварительные соображения относительно первого момента были уже изложены; что касается второго, то достоверность получаемой численной характеристики зависит от принятой при проведении идентификации модели спектра и сложности вычисляемого критерия соответствия спектров.

Остановимся теперь на самом понятии системы автоматического качественного РФА или автоматической идентификации. Это понятие отнюдь не исчерпывается только программой, осуществляющей качественный анализ на ЭВМ, наоборот, необходимо подчеркнуть прежде всего системный характер подобного рода разработки. Система автоматической идентификации в минимальном объеме обязательно включает следующие элементы: 1) машинную картотеку (базы данных), иногда состоящую из нескольких файлов; 2) программные средства, предназначенные для формирования, пополнения и корректировки базы данных (эти средства не всегда нужны и доступны пользователю, но они необходимы разработчику); 3) программу идентификации — основной рабочий инструмент пользователя; 4) совокупность тестов — контрольных задач, позволяющих проверить правильность функционирования программы идентификации; 5) методику проведения автоматической идентификации.

Если формат машинной базы данных приближается к рассмотренным в главе 1 полным форматам (например, к формату JCPDS), то на ее основе можно реализовать все пять типов поиска, описанных в разделе 1.4. В этом случае при разработке надлежащих программных средств система идентификации превращается в информационно-поисковую систему (ИПС), с помощью которой можно осуществлять не только качественный анализ, но и выборку интересующих данных по задаваемой совокупности параметров.

Первая система автоматической идентификации была разработана Фривелом в 1964—1968 гг. [100]. В дальнейшем Фривел совместно с Адамсом и Рахбергом значительно усовершенствовал

систему [102, 103] и рассмотрел вопросы, связанные с ее практическим применением [101]. Вскоре после первой публикации Фривела появилось сообщение о разработке Николюсом еще одной системы автоматической идентификации (1966 г.). Работа Николюса опубликована в труднодоступном советском исследователям издании [134], в связи с чем подробный анализ его системы нами не проводится. Наконец, в 1967—1978 гг. появились основополагающие работы Джонсона и Ванда [123, 124], в которых рассматривались важные методические вопросы автоматической идентификации

В последующих разделах системы идентификации описываются в следующем порядке: база данных, служебные программы (если они имеются), алгоритм идентификации, форма выдачи результатов анализа (выходные форматы). В оригинальных работах часто применяются разные обозначения для одних и тех же величин; кроме этого, иногда используются многосимвольные обозначения величин, взятые из текстов соответствующих программ. Такие многосимвольные обозначения (идентификаторы) используются в программировании, однако при математическом и логическом описании процедуры идентификации представляется более удобным и правильным пользоваться обычными математическими обозначениями. В дальнейшем будут применяться следующие основные обозначения: обратное межплоскостное расстояние — H ; окна совпадения линий в шкалах d или H — Δd или ΔH ; номер стандарта по картотеке JCPDS — N_{JCPDS} ; критерии соответствия спектров — K, F, R ; список вероятных фаз, определенных в результате проведения идентификации — S_B ; условная концентрация или максимальная концентрация по Дэви (Daví maximum concentration) — C или DMC; спектры маркируются «о» сверху для образца и индексом «s» сверху для стандарта. Принятые в оригинальных работах обозначения сохранены там, где это не нарушает связности и последовательности текста.

Отметим, что, несмотря на двадцатилетнее развитие систем автоматической идентификации, в литературе практически отсутствуют обзоры по рассматриваемой проблеме (кроме кратких обзоров [36, 78]). Материал, предлагаемый читателю в этой и последующих главах, является, таким образом, первым подробным сопоставительным анализом различных методов автоматической идентификации.

4.2. СИСТЕМА ZRD

Система ZRD была разработана Фривелом [100] для решения задач качественного РФА, возникающих в практической деятельности сотрудников Доу Кемикл Корпорейшн (Dow Chemical corporation). Особенность системы — обязательное наличие информации о химическом составе исследуемого вещества, поэтому, как указывает Фривел в [100], для использования его системы

необходимо вначале исследовать химический состав пробы методами оптической или рентгеновской флуоресцентной спектрометрии.

База данных. В 1964 г. картотека JCPDS включала около 8000 неорганических соединений. Из них Фривелом было отобрано 1359 стандартов, представляющих интерес для аналитиков Доу Кемикл Корпорейши; таким образом, Фривел впервые сформировал специализированную машинную картотеку целевого назначения.

База данных системы включает два файла — ZD и ZN. Файл ZD строится следующим образом. Из каждого стандартного спектра выбираются три значения $d^{s,(1)}$, $d^{s,(2)}$, $d^{s,(3)}$, соответствующие трем сильнейшим дифракционным линиям. Затем составляется список R_s всех химических элементов, входящих в данный стандарт (элементы обозначаются атомными номерами) и список $\{Z_s\}$ сложных полиатомных групп (радикалов, комплексов, лигандов и т. п.). Каждая полиатомная группа кодируется численным номером, принимающим значения от 1 до 5000. Таким образом, стандартная карточка ZD превращалась в файле в краткую запись следующего вида:

$$\{Z_s\} N_s \{R_s\} d^{s,(1)} d^{s,(2)} d^{s,(3)}. \quad (4.1)$$

Здесь целое число N_s является номером, устанавливающим соответствие стандарта в файле ZD с его номером по картотеке JCPDS.

Файл ZD состоит из ряда последовательных блоков, каждый из которых соответствует определенному элементу ($Z = 3$ — блок для соединений с литием; $Z = 4$ — блок для соединений с бериллием и т. д.). Каждый блок состоит из трех частей. Часть первая содержит все записи вида (4.1), упорядоченные по возрастанию $d^{s,(1)}$ — значений для сильнейшей линии; вторая — по возрастанию $d^{s,(2)}$, а третья — по возрастанию $d^{s,(3)}$. В работе [100] приводится конкретный вид одного из блоков файла ZD (для значения $Z = 34$, селен).

Для построения файла ZN из каждого стандартного спектра выбираются десять сильнейших линий (или все линии, если их меньше десяти) и формируется запись следующего вида:

$$N_{\text{JCPDS}} I^{s,(1)} d^{s,(1)} \dots, I^{s,(10)} d^{s,(10)}. \quad (4.2)$$

Как и первый файл, файл ZN состоит из последовательных блоков, соответствующих элементам от $N = 3$ (литий) до $N = 93$ (нептуний). Каждый блок содержит все записи вида (4.2), упорядоченные по возрастанию номеров стандартов N_{JCPDS} .

Рассмотренная структура базы данных системы Фривела довольно сложная. Цель, которая преследуется подобным построением базы данных, заключается в ускорении работы программы идентификации.

Алгоритм идентификации. Задаются спектр исследуемого образца $\{d_h^o, I_h^o\}$, список входящих в образец элементов $\{Z_o\}$, список полиатомных групп $\{R_o\}$ и некоторые параметры поиска. Иден-

тификация осуществляется в два этапа. Первый, предварительный, этап идентификации проводится с помощью файла ZD. Из списка $\{Z_o\}$ выбирается первый химический элемент и начинается поиск по соответствующему блоку файла ZD. Рассматривается первая часть блока и для двадцати сильнейших линий спектра образца определяются стандарты, первые линии которых удовлетворяют неравенству:*

$$d^o - \Delta d \leq d^{s(1)} \leq d^o + \Delta d. \quad (4.3)$$

Окно совпадения линий Δd полагается пропорциональным d^2 . Очевидно, что структура файла ZD позволяет определить нужный блок и отобрать удовлетворяющие неравенству (4.3) стандарты без просмотра всего файла.

Отобранные вероятные стандарты подвергаются затем трем проверкам — Z-, R- и D-тестам. Тест Z заключается в проверке совпадения элементов списка $\{Z_s\}$ с какими-либо элементами списка $\{Z_o\}$. Аналогично при проведении теста R определяется совпадение списка $\{R_s\}$ с некоторыми полиатомными группами, перечисленными в R_o . При проведении теста D с помощью неравенства вида (4.3) среди двадцати сильнейших линий спектра образца определяются линии, соответствующие $d^{s(2)}$ и $d^{s(3)}$ рассматриваемого вероятного стандарта.

Описанная выше процедура проводится для всех элементов из списка Z_o . В результате отбираются стандарты, удовлетворяющие следующим условиям: 1) список элементов стандарта является подмножеством списка элементов образца; 2) список полиатомных групп стандарта представляет собой подмножество соответствующего списка для образца; 3) три сильнейшие линии стандарта $d^{s(1)}$, $d^{s(2)}$, $d^{s(3)}$ наблюдаются среди двадцати сильнейших линий спектра образца.

На выходе первого этапа идентификации формируется список номеров вероятных фаз N_s . На втором, заключительном, этапе идентификации из списка Z_o выбирается первый элемент и следует обращение к соответствующему блоку файла ZN. Из содержащихся в этом блоке записей вида (4.2) рассматриваются только те, номера которых попадают в список вероятных фаз, определенных на первом этапе. Для этих стандартов проверяется соответствие линий (десять сильнейших) спектру образца в смысле выполнения условия вида (4.3). Затем по интенсивностям совпавших линий вычисляется условная концентрация:

$$C = \frac{\sum_{k=1}^{n_o} I_k^o}{\sum_{j=1}^{n_s} I_j^s}. \quad (4.4)$$

* В обозначениях Фривела окно совпадения линий образца и стандарта Δd выражается в виде суммы ошибок определения положения линии образца и соответствующей линии стандарта.

Здесь линии спектра образца нумеруются индексом k ; линии стандарта — индексом j ; знак штрих на суммах означает, что суммирование ведется только по совпадающим линиям спектров образца и стандарта.

Величина C используется для приведения спектров образца и стандарта в одну систему отсчета по интенсивности. На втором этапе Фривелом применяется разностный алгоритм идентификации (метод вычитания) — интенсивности линий стандарта, умноженные на C , вычитаются из спектра образца. Получающийся разностный спектр имеет вид: соответственно для совпавших линий и прочих

$$I_k^p = \begin{cases} \Delta I_k = |I_k^o - CI_j^s| \\ I_k^o \end{cases} \quad (4.5)$$

Таким образом, из исследуемого спектра исключается одна фаза, и для разностного спектра процедура поиска и сопоставления повторяется. Теоретически, если все последовательно исключаемые фазы определены правильно, разностный спектр после отбора всех истинных фаз окажется пренебрежимо малым. Для определения этого момента в системе Фривела задается пороговое значение интенсивности.

Предположим, что k' -я линия спектра образца является наложенной (суперпозицией линий различных составляющих фаз); тогда величина $\Delta I_{k'}$ может не оказаться малой. Для определения этой ситуации в системе задается еще один параметр; если $I_{k'}$ превосходит величину этого параметра, k' -я линия исследуемого спектра считается наложенной. Такие линии вносят ошибку в условную концентрацию, вычисленную по формуле (4.4). Чтобы избежать ошибки, величина C пересчитывается вновь, но вместо интенсивности наложенной линии $I_{k'}$ в верхнюю сумму в (4.4) подставляется интенсивность соответствующей линии стандарта, умноженная на условную концентрацию, — CI_j^s . Эта процедура повторяется несколько раз, пока в результате данного итерационного процесса не будет получено устойчивое значение C .

Выходные форматы. В результате работы программы идентификации на печать выводится таблица, в которой все пары значений d^o , I^o сопоставлены с соответствующими линиями первой, второй и т. д. найденных стандартных фаз. Интенсивности линий образца и стандартов даются в приведенной форме (в сопоставимой шкале); в последней графе таблицы печатаются величины отклонений по интенсивности для каждой линии спектра образца. Величины этих отклонений характеризуют качество идентификации. Пример выходной таблицы системы Фривела дан в [100].

Рассмотренная выше система была реализована на языке алгол-60 на вычислительной машине «Burrough-5000» с пользователем накопителя на магнитной ленте и получила наименование ZRD-1965. Время идентификации составляло от 2 до 5 мин. С учетом

того, что база данных системы включала только по 10 линий для 1359 стандартов, такое время является довольно большим.

В 1968 г. на ЭВМ «Burrough-5500» была разработана новая версия системы ZRD [102], база данных которой включала 752 стандарта, взятых из публикаций Национального бюро стандартов США (NBS). Для хранения и работы с базой данных использовался накопитель на магнитном диске.

В 1974 г. была создана значительно модернизированная версия системы, получившая наименование ZRD-1974 [103]. Эта система реализована на ЭВМ IBM-370-158 и отличается от более ранней версии следующими моментами.

1. База данных системы содержит информацию только для 298 наиболее часто встречающихся на практике стандартов. Список их был составлен на основании 30 000 определений фазового состава, осуществленных в Доу Кемикл Корпорейшн в течение 8 лет.

2. Для проведения предварительного этапа идентификации вместо файла ZD используется файл значительно более простой структуры. Он содержит записи вида (4.1) для 298 стандартных фаз, упорядоченные по возрастанию $d^{s(1)}$. Разбиения файла на блоки по химическим элементам не производится; таким образом, он состоит всего из 298 записей. В файл добавлена информация о химических формулах или наименованиях соединений.

3. При проведении идентификации учитывается вес совпадения линий образца и стандарта в зависимости от величины d . Для оценки весовой функции построена кривая интегральной плотности распределения сильнейших линий $d^{s(1)}$ для 298 стандартов.

Форма таблицы результатов идентификации системы ZRD-1974 такая же, как для ZRD-1965. Время поиска по 298 фазам составляет 8 с.

4.3. СИСТЕМА ДЖОНСОНА—ВАНДА

Система Джонсона — Ванда [123, 124] разработана под эгидой Комитета по химическому анализу методами порошковой дифрактометрии (Joint Committee on Chemical Analysis by Powder Diffraction Methods of the ASTM). Проблемы формирования базы данных и автоматизации процесса качественного анализа рассматривались Джонсоном и Вандом в более общей постановке, чем Фривелом. Необходимо подчеркнуть, что система Джонсона — Ванда с самого начала была ориентирована на применение в любых научных исследованиях, в ходе которых возникает необходимость в проведении качественного РФА. Авторы стремились стандартизовать процедуру автоматической идентификации, сделать ее экспрессной, достаточно достоверной и максимально независимой от влияния решаемых пользователем конкретных задач и специфики исследуемых соединений. Подобная постановка проблемы позволила Джонсону и Ванду разработать ряд важных методических приемов идентификации, которые используются и в настоящее время.

База данных. Независимость процедуры идентификации от специфики круга исследуемых соединений может быть достигнута формированием максимально полной базы данных. В связи с этим база данных системы Джонсона — Ванда включает рентгендифракционные спектры для всех соединений, входящих в картотеку JCPDS. К моменту опубликования первой работы [123] база данных содержала, в соответствии с картотекой JCPDS, 12 000 справочных стандартных спектров; в течение последующих лет она ежегодно пополнялась — в нее включались все новые спектры, получаемые комитетом JCPDS. В настоящее время объем базы данных около 40 000 соединений [142]. Формат ее в современном варианте был рассмотрен выше в разделе 1.4. Еще раз подчеркнем, что база данных системы Джонсона — Ванда (или, как она именовалась в разделе 1.4, машинная картотека JCPDS) по объему информации об одном соединении не эквивалентна полной картотеке JCPDS; она содержит только те сведения, которые необходимы для проведения качественного анализа.

Большой объем машинной картотеки JCPDS, используемой в системе Джонсона — Ванда, потребовал разработки экспрессных методов идентификации, которые определили специфику формата и структуры базы данных. Был предложен новый формат представления спектра в целочисленном виде. Положения линий задаются не в прямом d -пространстве, а в обратном пространстве H , где

$$H^s = \text{Бл. ц.} \left(\frac{1000}{d^s} \right). \quad (4.6)$$

Интенсивности переводятся из стобалльной шкалы в сглаживающую логарифмическую шкалу.

$$I^{s,lg} = \text{Бл. ц.} (5 \lg I^{s,100}). \quad (4.7)$$

Преимущество данных шкал заключается в том, что величины окон ΔH и ΔI^{lg} слабо зависят от диапазона изменения H и I^{lg} .

В первых версиях системы Джонсона — Ванда [123, 124] база данных включала два файла: инвертированный PS и прямой PSI. Использование инвертированного файла для ускорения процедуры поиска является одним из важных моментов рассматриваемой системы. Максимально возможному диапазону изменения d 0,6—200 Å соответствует диапазон изменения H от 1666 до 5. Файл PS строится путем последовательного перечисления всех целочисленных значений H , причем каждому значению H приписываются номера N_{JCPDS} всех стандартных фаз (карточек), в спектре которых встречается линия с таким положением. Инвертированный файл имеет следующую структуру:

$$\begin{aligned} H^s &= 5\{N_{JCPDS}\}_5, \\ H^s &= 6\{N_{JCPDS}\}_6, \\ H^s &= 7\{N_{JCPDS}\}_7 \end{aligned} \quad (4.8)$$

и так далее до последнего максимального значения $H = 1666$. Здесь $\{N_{JCPDS}\}_5$ означает список номеров карточек, содержащих спектры с линией $H = 5$; аналогичный смысл имеют обозначения $\{N_{JCPDS}\}_6$, $\{N_{JCPDS}\}_7$ и т. д.

Отметим, что идея организации базы данных в виде инвертированного файла уже была заложена в указателе Ханавальта. В принципе, подобные структуры достаточно часто используются при создании информационно-поисковых систем в различных отраслях знания.

Прямой файл PSI содержит совокупность машинных карточек, упорядоченных по возрастанию номеров JCPDS. Вид машинной карточки следующий:

$$N_{JCPDS} \{H_j^s, I_j^{s,lg}\}. \quad (4.9)$$

Список $\{H_j^s, I_j^{s,lg}\}$ может содержать до 98 линий спектра. Величины $H_j^s, I_j^{s,lg}$ находятся в одной ячейке машинной памяти в упакованном виде. Для файла PSI пересчет положений линий в шкалу H осуществляется по более точной, сравнительно с (4.6), формуле

$$H^s = \text{Бл. ц.} \left\{ \frac{10^4 000}{d^s} \right\}. \quad (4.10)$$

Служебные программы. Первая служебная программа обеспечивает ввод данных с перфокарт и построение прямого файла на магнитной ленте в виде совокупности блоков, каждый из которых содержит 20 карточек. После построения файла PSI на МЛ осуществляется проверка правильности записи карточек в порядке возрастания их номеров N_{JCPDS} .

Вторая программа использует файл PSI для формирования на отдельной МЛ промежуточного неупорядоченного инвертированного файла, включающего все пары $\{H^s, N_{JCPDS}\}$. При формировании этого файла можно указать граничное значение интенсивности $I_{гр}$ такое, что в промежуточный файл из файла PSI будут включены только линии с интенсивностью $I^{s,lg} \geq I_{гр}^{lg}$.

Третья служебная программа упорядочивает промежуточный файл по возрастанию H и формирует инвертированный файл PS на отдельной МЛ. Файл PS содержится на МЛ в форме совокупности блоков, каждый из которых содержит по четыре строки вида (4.8).

Алгоритм идентификации. Как и в системе ZRD, идентификация в системе Диксона — Ванда осуществляется в два этапа. В ЭВМ вводится спектр исследуемого образца и величина окна совпадения линий ΔH ; можно также ввести еще ряд параметров, определяющих условия поиска. Спектр образца преобразуется в формулу $\{H_h^o, I_h^{o,lg}\}$. Первый, предварительный, этап идентификации проводится с помощью инвертированного файла PS. Из последнего выбираются все строки вида (4.8), значения H^s

для которых совпадают с положениями линий спектра образца H^0 в пределах заданного окна:

$$H^0 - \Delta H \leq H^s \leq H^0 + \Delta H. \quad (4.11)$$

Величина окна полагается постоянной: $\Delta H = \text{const}$.

Поскольку записи (4.8) в файле PS упорядочены по возрастанию H^s , отбор удовлетворяющих условию (4.11) строк не требует последовательного просмотра всего инвертированного файла. Для заданного значения H^0 можно сразу определить совокупность представляющих интерес строк с номерами от $H^0 - \Delta H$ до $H^0 + \Delta H$. Таким образом, инвертированный файл обеспечивает экспрессное проведение предварительного этапа отбора вероятных фаз.

После отбора из файла PS всех строк, удовлетворяющих (4.11), по полученному массиву вычисляется частотный критерий соответствия K . Частотный критерий равен относительному числу линий стандарта, совпавших с линиями спектра образца в смысле выполнения (4.11):

$$K = \frac{n_{\text{совп}}}{n_s} \cdot 100\%. \quad (4.12)$$

Здесь $n_{\text{совп}}$ — число линий стандарта, совпавших со спектром образца; n_s , как и ранее, обозначает полное число линий стандарта.

На второй этап идентификации пропускаются стандарты, у которых $K \geq K_{\text{гр}}$ и среди совпавших линий присутствует не менее $l_{\text{гр}}$ сильнейших. Граничные значения $K_{\text{гр}}$ и $l_{\text{гр}}$ являются параметрами метода; они должны задаваться исследователем, либо, по умолчанию, принимаются равными некоторым определенным величинам. Так, величина $l_{\text{гр}}$ обычно равнялась двум или трем.

На втором, заключительном этапе идентификации из прямого файла PSI выбираются машинные карточки стандартов, прошедших первый этап. Для каждого вероятного стандарта рассчитывается условная концентрация (максимальная концентрация по Дэви) DMC, которая принимается в качестве критерия близости спектров. Величина DMC определяется как максимальное отношение интенсивности линии образца k интенсивности соответствующей линии стандарта:

$$\text{DMC} = \max' \left(\frac{I_k^{0,100}}{I_j^{s,100}} \right). \quad (4.13)$$

Знак штрих на символе \max' означает, как и ранее в формуле (4.4), что рассматриваются отношения только для совпадающих линий; индексы k и j нумеруют линии образца и стандарта по порядку их следования в спектрах.

Так как интенсивности в системе Джонсона — Вауда заданы в логарифмической шкале, то вместо (4.13) используется выражение

для логарифма DMC:

$$\lg \text{DMC} = \frac{I_k^{o, \lg} - I_j^{s, \lg}}{5} = \frac{\Delta I_{\text{совп}}^{\max}}{5}. \quad (4.14)$$

Здесь $\Delta I_{\text{совп}}^{\max}$ — максимальная разность логарифмических интенсивностей совпадающих линий. Полагается, что интенсивность линии образца всегда меньше или равна интенсивности соответствующей линии стандарта *, таким образом, $\Delta I_{\text{совп}}^{\max}$ — всегда величина отрицательная. Удобнее все же рассматривать положительную величину $\Delta I_{\text{совп}}^{\max} = \max' |I_k^{o, \lg} - I_j^{s, \lg}|$, тогда окончательный вид DMC будет следующим:

$$\text{DMC} = 10 \frac{\Delta I_{\text{совп}}^{\max}}{5} \cdot 100\%. \quad (4.15)$$

Поскольку логарифмические интенсивности в формуле (4.14) представляются числами 0, 1, 2, ..., 9, то разность $\Delta I_{\text{совп}}^{\max}$ также представляется числом из этого ряда. Таким образом, в системе Джонсона — Ванда возможно только десять значений логарифма DMC: 0, $-\frac{1}{5}$, $-\frac{2}{5}$, ..., $-\frac{9}{5}$ и, соответственно, десять дискретных значений DMC (%): 100, 63, 39, 25, 15, 10, 6, 3, 2, 1.

Выходные форматы. На печать выводятся следующие сообщения: 1) исходные параметры поиска; 2) рентгендифракционный спектр исследуемого образца; 3) результаты идентификации — число совпавших линий, величина DMC и соответствующий номер N_{JCPDS} стандарта.

Результаты идентификации выдаются не в форме таблицы, а полным текстовым сообщением [123], в котором многократно повторяются одинаковые фразы. Таблица соответствия по линиям спектров образца и вероятных стандартов отсутствует.

В работах [123, 124] описаны три первые версии системы Джонсона — Ванда. Первая версия была реализована на ЭВМ IBM-7074 на языке фортран-II; быстроедействие процедуры идентификации составляло 5 мин (для базы данных, включающей 12 000 стандартов). Вторая и третья версии реализованы на ЭВМ IBM-360/50, IBM-360/65, IBM-360/67 на языке фортран-IV, при этом время идентификации сократилось до 2 мин. При всех этих версиях использовались накопители на магнитных лентах.

Система Джонсона — Ванда была принята комитетом JCPDS в качестве базовой и в 70-х годах интенсивно эксплуатировалась десятками пользователей. В результате обобщения опыта использования системы были разработаны ее новые версии, число которых к 1977 г. достигло 18. Список всех версий с их краткой характеристикой дан Джонсоном в [122]. Почти все версии реализова-

* Это условие может нарушаться, если линия спектра образца наложенная, а также если образец монофазный и из-за ошибок в измерении интенсивности некоторые линии спектра образца сильнее соответствующих линий стандарта.

ны на ЭВМ IBM-360/67; версия 18 разработана на новой ЭВМ IBM-370/165 и может также эксплуатироваться на ЭВМ IBM-360/370, UNIVAC-1100, CDC 6000/7000. Последние версии системы работают с накопителем на магнитном диске.

Развитие системы Джонсона — Ванда в течение 10 лет привело, наконец, к созданию стабильного варианта (версии 18), для которого было разработано руководство пользователя [121]. Сравнительно с описанными выше первыми вариантами система претерпела значительные изменения. При этом основным усовершенствованием является введение новых критериев соответствия спектров — критерия близости стандартного и исследуемого спектров по положениям линий $F(d)$ (близость по d или по H); критерия близости спектров по интенсивностям $F(I)$ (близость по I); комбинированного критерия достоверности идентификации F (фактор надежности). Все эти критерии рассчитываются на втором, заключительном, этапе идентификации. Для сведения читателей, которые пожелают обратиться к оригинальным источникам, необходимо заметить, что в работах [122—124] формулы для критериев $F(d)$, $F(I)$ и F не приводятся. Их вид можно установить из работы Дисмора [92], посвященной практическому применению системы, и из обзора Фиалы [125], который излагает алгоритм идентификации Джонсона — Ванда, ссылаясь при этом на описание десятой версии системы [125].

Критерии $F(d)$ и $F(I)$ рассчитываются по формулам:

$$F(d) = 1 - \frac{\sum_{i=1}^{n_{\text{совп}}} |H_i^s - H_i^o|}{n_{\text{совп}} \cdot \Delta H}, \quad (4.16)$$

$$F(I) = 1 - \frac{\sum_{i=1}^{n_{\text{совп}}} |I_i^{s,lg} - I_i^{o,lg}|}{\sum_{i=1}^{n_{\text{совп}}} I_i^{s,lg}}. \quad (4.17)$$

Для простоты в формулах (4.16), (4.17) применена иная система нумерации линий, чем в формулах (4.4), (4.13); индекс i нумерует совпавшие линии спектров стандарта и образца в порядке их следования. Сравнение спектров осуществляется в пределах задания спектра исследуемого образца.

Фактор надежности вычисляется как произведение частотного критерия K , определяемого по (4.12), и критериев $F(d)$, $F(I)$:

$$F = K \cdot F(d) \cdot F(I). \quad (4.18)$$

Вероятный стандарт выводится на печать, если для него $F > F_{\text{гр}}$. Граничное значение $F_{\text{гр}}$ можно задавать в параметрах поиска; по умолчанию полагается, что $F_{\text{гр}} = 35\%$.

Кроме рассмотренной выше модификации системы, в ее базу данных были добавлены файлы, содержащие дополнительную

информацию, — наименования и химические формулы соединений, сведения о достоверности стандартов и т. д. Введение химической информации позволило, как и в системе ZRD, учитывать в процессе идентификации известный элементный состав образца. Однако, в отличие от системы ZRD, задание сведений о химическом составе в виде списков присутствующих и заведомо отсутствующих в образце элементов в системе Джонсона — Ванда не является обязательным.

Из полной базы данных были выделены подбазы, включающие сведения только о сплавах, минералах, органических, неорганических и наиболее широко распространенных соединениях (так называемый мини-файл). Введена возможность отбрасывания слабых линий стандартов. Для этого в исходных данных задается пороговая величина отсечки по интенсивности $I_{отс}$, после чего в процедуре идентификации учитываются только те линии стандартов, интенсивность которых $I^s \geq I_{отс}$. Значительно улучшен также выходной формат, которому придана форма таблицы со следующей информацией для каждой вероятной фазы:

$$N_{JCPDS} \quad \begin{array}{c} \text{Химическая} \\ \text{формула} \end{array} \quad n_{совп} \quad K \quad I_{отс} \quad DMC \quad F(d) \quad F(I) \quad F. \quad (4.19)$$

Наконец, в систему Джонсона — Ванда была введена методика вычитания спектров стандартных фаз (сходная с рассмотренной в разделе 4.2), хотя ранее Джонсон и Ванд критиковали в [123] применение данной методики в системах Фривела и Никольса.

Несомненно, система Джонсона — Ванда является одной из крупнейших систем, а также и наиболее распространенной среди пользователей разработкой в рассматриваемой области. Процесс ее совершенствования был длительным; он характеризовался разработкой многочисленных версий системы, созданием ряда методических элементов процедуры идентификации, отказом от некоторых из этих элементов и включением их в систему снова, на более высоком уровне понимания их значения. Таким образом, история совершенствования системы Джонсона — Ванда является прекрасной иллюстрацией сложности рассматриваемой проблемы автоматизации качественного рентгенофазового анализа.

4.4. СИСТЕМА PDSM CIS

В 1976 г. база данных JCPDS на магнитной ленте, рассмотренная в разделе 1.4, была включена в состав Химической информационной системы (CIS). Разработанные к этому времени системы идентификации Фривела, Никольса и Джонсона — Ванда были признаны неудобными и слишком дорогими в эксплуатации. Это обстоятельство послужило отправной точкой для разработки в рамках CIS новой системы идентификации [132].

Химическая информационная система (NTH — EPA Chemical Information System) была создана в начале 70-х годов Национальным институтом здоровья США (National Institute of Health NTH) при участии Агентства охраны окружающей среды (Environmental Protection Agency — EPA) и является одной из крупнейших систем США в области сбора, хранения и использования физико-химической информации. Состав, структуры и коммерческое применение системы на 1977 г. рассмотрены в работе [107]. Система разработана на базе мощной ЭВМ PDP-10.

Все программные средства CIS ориентированы на общение с пользователем в интерактивном режиме. Последний обеспечивает в процессе проведения поиска и анализа гибкую и устойчивую двустороннюю связь пользователя с ЭВМ. Благодаря этому исследователь имеет возможность давать системе подсказки, направлять процессы поиска и расчетные процедуры в желаемую сторону, активно вмешиваться в машинный анализ. Преимущества подобных человекомашинных систем отмечались неоднократно [2, 28, 52, 57, 66]; они позволяют, с одной стороны, эффективно использовать опыт и интуицию исследователя и, с другой, рационально задействовать скоростные возможности ЭВМ. В целом наблюдается значительный выигрыш во времени решения задачи. Особенно удобен интерактивный режим для решения некорректных задач, к которым относится и проблема идентификации. Конкретной формой выражения интерактивного режима является диалог между пользователем и ЭВМ, поэтому подобные системы часто называют диалоговыми.

Изложенные выше соображения поясняют необходимость разработки новой системы идентификации в рамках CIS. Поскольку CIS совершает сделки на продажу информационных услуг, система идентификации должна быть выгодной в коммерческом отношении. Система должна быть также интерактивной, поскольку это — один из основных принципов построения CIS, чрезвычайно удобный для пользователей.

Перейдем к рассмотрению новой системы идентификации CIS, получившей название PDSM (Powder Diffraction Search — Match System). Сообщение о ее разработке большой группой специалистов с участием Джонсона появилось в 1979 г. [132]. Однако эта работа весьма краткая и не содержит детального описания процесса поиска и сравнения. В частности, в ней не приводятся вид критериев соответствия спектров образца и стандарта, в связи с чем мы ограничимся только общим описанием системы.

База данных. В системе PDSM база данных включает поисковый файл и дополнительный, в котором содержатся названия и химические формулы стандартов. Поисковый файл строится из базы данных JCPDS специальной программой. При построении поискового файла в него отбираются только те линии стандартных спектров, которые наиболее значимы для процесса поиска. Критерием отбора значимых линий является не только их высокая ин-

тенсивность (как в указателях Ханавальта и Финка), но и вероятность распределения линий по d .

Поисковый файл состоит из двух частей. В первую часть для каждого стандарта включаются до 15 наиболее значимых линий спектра со значениями $d^s \geq 1,2 \text{ \AA}$. Если стандарт имеет линии с $d^s < 1,2 \text{ \AA}$, то во вторую часть поискового файла могут быть включены до 8 наиболее значимых линий из этой области; при этом в машинной карточке стандарта, содержащейся в первой части файла, делается соответствующая ссылка. Положения линий задаются в шкале H согласно выражению (4.6). Форма задания интенсивностей в [132] не описывается, но, по-видимому, она также аналогична принятой в системе Джонсона — Ванда.

Алгоритм идентификации. Идентификация осуществляется в два этапа. Исследуемый спектр $\{d_h^o, I_h^o\}$ может быть введен в ЭВМ вручную (с удаленного терминала — дисплейной станции), с перфолент, МЛ и непосредственно с прибора. В процессе ввода можно осуществлять коррекцию спектра, изменяя и исключая некоторые значения d^o, I^o . Задаются параметры поиска, в целом аналогичные используемым в последних версиях системы Джонсона — Ванда. Для ускорения процедуры поиска имеется возможность задания химической информации и указания специализированной подбазы (минералы, органические или неорганические соединения, сплавы, номер сета JCPDS). Поиск осуществляется по первой части поискового файла путем последовательного сравнения стандартов с исследуемым спектром. Обращение к записи во второй части файла производится только в том случае, когда стандарт и исследуемый спектр имеют линии в области $d < 1,2 \text{ \AA}$. Сравнение спектров по положениям линий осуществляется с постоянным окном ΔH с использованием условия вида (4.11).

Критерии отбора вероятных стандартов на предварительном и заключительном этапах идентификации в [132] не описаны. На втором этапе вычисляются основной критерий соответствия стандартного и исследуемого спектров (индекс подобия SI) и условная концентрация. При вычислении критерия SI производится приведение спектров стандарта и образца в сопоставимую шкалу по интенсивности. Можно осуществлять сравнение спектра образца со спектром отдельного стандарта; для этого в режиме диалога дается соответствующее указание и задается N_{JCPDS} интересующего стандарта. Алгоритм идентификации системы PDSM предполагается дополнить введением методики вычитания вероятного стандарта.

Выходные форматы. На печать выводится таблица результатов идентификации, каждая строка которой содержит следующую информацию для вероятного стандарта:

$$N_{\text{JCPDS}} \quad \text{SI} \quad C \quad n_s \quad n_{\text{совп}} \quad n_{\text{отс}} \quad \begin{array}{l} \text{Химическая} \\ \text{формула} \end{array} \quad \begin{array}{l} \text{Название} \\ . \end{array} \quad (4.20)$$

Здесь C — условная концентрация, n_s — число линий стандарта в интервале задания спектра образца, $n_{отс}$ — число линий стандарта, отсутствующих в спектре образца.

Совокупность критериев SI , C , n_s , $n_{совп}$, $n_{отс}$ характеризует степень соответствия стандарта исследуемому спектру. Стандарты упорядочиваются по убыванию степени соответствия и выводятся последовательными группами, содержащими по десять строк вида (4.20). Однако в [132] не указано, каким образом осуществляется упорядочение стандартов по совокупности нескольких критериев, а изучение приведенных в работе таблиц выдачи результатов показывает, что вероятные стандарты выводятся в достаточно произвольном порядке. Вопрос упорядочения вероятных стандартов по степени их соответствия спектру образца является исключительно важным, так как фактически это означает замену пороговых критериев Джонсона — Ванда (4.12), (4.18) (критериев типа да/нет, соответствует/не соответствует) на процедуру построения идентифицирующей функции. Этот подход был впервые теоретически проанализирован в работе [28]. Позднее многие исследователи пришли к необходимости построения в результате качественного РФА идентифицирующей функции, которая получила в англо-американской научной литературе название диаграммы достоверности (figure-of-merit или FOM). Последние публикации в области машинного качественного РФА, рассмотренные в разделах 4.5—4.8, показывают, что подход с построением FOM стал общепринятым. Таким образом, необходимо отметить как достоинство системы PDSM то обстоятельство, что данный подход, хотя и в неявной форме, был в ней заложен.

Окончательный вывод о составе образца делается исследователем на основании анализа таблицы результатов (4.20). Как отмечалось выше, в системе PDSM задание исходных данных и поиск ведутся в диалоговом режиме, что позволяет удобно провести несколько вариантов поиска при различных исходных значениях поисковых параметров. В качестве примера быстрого действия системы PDSM [132] приводится поиск по части подбазы органических соединений, включающих 777 стандартов, который длился 17 с. С учетом мощности ЭВМ PDP-10 такое время представляется довольно значительным.

4.5. СИСТЕМА ФИРМЫ ФИЛИПС

Обобщение опыта разработки и эксплуатации систем Фривела, Никольса и Джонсона — Ванда послужило основой для создания во второй половине 70-х годов ряда систем идентификации, как оригинальных, так и существенно использующих результаты работ Фривела и Джонсона — Ванда. Одна из таких крупных систем — PDSM CIS — описывалась в предыдущем разделе.

В рассматриваемый период проявилась такая общая тенденция развития систем идентификации, как ориентация их на мини-ЭВМ. Естественно, что необходимым условием возможности реализации на мини-ЭВМ системы автоматического качественного РФА является достаточно полная комплектность ЭВМ. В ее состав обязательно должны входить долговременные накопители на МЛ или МД, основной функцией которых является хранение базы данных. Постоянное возрастание мощности мини-ЭВМ, включение в их состав разнообразных терминалов позволяют, с одной стороны, увеличивать функциональные возможности и быстродействие систем идентификации и, с другой, развивать новые методики автоматического качественного РФА. Эти две тенденции легко проследить по проведенному в ряде последующих разделов описанию систем идентификации, базирующихся в основном на мини-ЭВМ. Данные системы разработаны в США, ФРГ, Нидерландах, и мы рассмотрим их примерно в хронологическом порядке. Необходимо также отметить, что в конце 70-х годов были созданы серийные системы идентификации крупнейшими приборостроительными фирмами Филипс, Сименс и ИВМ. Эти системы входят в состав больших программных пакетов, обеспечивающих управление приборами и обработку информации (поиск пиков, качественный и количественный РФА, определение субструктуры и т. д.) в масштабе реального времени. Все системы, разработанные приборостроительными фирмами, используют базу данных JCPDS для формирования собственных файлов специализированного формата.

В 1979—1982 гг. появились сообщения, описывающие две системы идентификации, разработанные фирмой Филипс (Philips, Netherlands — USA). Эти системы ориентированы на мини-ЭВМ и являются составными частями пакетов прикладных программ, обеспечивающих управление автоматическими дифрактометрами в процессе эксперимента, сбор и предварительную обработку данных (отделение фоновой составляющей, поиск пиков, разделение дублетов, вычисление интегральных интенсивностей), качественный и количественный РФА, анализ профилей линий с целью получения информации о параметрах субструктуры соединений. Первая система разработана в 1978 г. в Нидерландском отделении фирмы Филипс (Philips Application Laboratories for X-ray and Nuclear Analysis) и очень кратко описана в нескольких проспектах фирмы Филипс [86, 87, 167]. Система реализована на базе мини-ЭВМ фирмы Филипс P851M (объем памяти — 32 К, разрядность слова — 16 бит) и предназначена для использования в комплекте с дифрактометром APD-10. База данных (стандартный мини-файл) хранится на кассетном магнитофоне и может состоять из стандартов, взятых из картотеки JCPDS, других литературных источников и из результатов собственных измерений. Информация, хранящаяся для одного стандарта, включает N_{JCPDS} , название или химическую формулу соединения и до 50 пар значений $\{H^s, I^s\}$, где H^s определяется по (4.10). На одной кассете исследо-

ватель может сформировать необходимую ему в практической деятельности специализированную базу данных объемом до 2000 стандартов. Для проведения автоматической идентификации необходимо задать ряд параметров поиска, в том числе окно ΔH , граничное число совпадающих сильных линий $l_{гр}$, максимально допустимое число сильных линий стандарта, лежащих вне диапазона задания спектра образца, и т. д. При поиске последовательно просматривается файл стандартов и при обнаружении совпадения осуществляется вывод на печать номера N_{JCPDS} соответствующего стандарта, названия или химической формулы стандарта, значений критерия соответствия спектров и других показателей качества идентификации. Окончательный отбор фаз осуществляется исследователем. Скорость поиска невелика и составляет в среднем 1 мин на просмотр ста стандартов. Выражение для критерия соответствия системы идентификации APD-10 в [86, 87, 167] не приводится, но, по-видимому, это одна из модификаций широко распространенного критерия Джонсона — Ваида (4.18).

Вторая система идентификации фирмы Филип разработана в США Дженкинсом и др. [116, 117, 148, 149] (Philips Electronic Instruments Inc.) и предназначена для использования в комплекте с дифрактометром APD-3600 [116]. Система включает базу данных и две программы идентификации — SANDMAN (Search And Match/Nova) и CASSAM (Computer Aided Search And Match). Дифрактометр APD-3600 выпускается в трех основных вариантах, имеющих номера 00, 01 и 02. В варианте 00 используется мини-ЭВМ класса Nova с объемом оперативной памяти 32К слов (слово — 16 бит), двумя накопителями на гибких МД объемом по 0,3 М байта, графическим терминалом и другими устройствами ввода — вывода. Данный вариант используется в основном для количественного анализа, однако имеются ограниченные возможности проведения идентификации по базе данных небольшого объема. В варианте 01 в комплектность ЭВМ входят два НМД объемом по 5 М байт, а в варианте 02, кроме наличия двух НМД, увеличена оперативная память до 64К слов. Две последние модификации APD-3600 позволяют в полном объеме обрабатывать программы SANDMAN и CASSAM с проведением поиска по базе данных, включающей несколько десятков тысяч стандартов.

База данных системы APD-3600 строится по полной базе JCPDS; таким образом, база APD-3600 включала в 1978 г. 32 000 стандартов [116], в 1979 г. — 34 000 [117] и так далее. Для построения базы данных могут использоваться и другие источники, кроме JCPDS. Структура базы данных детально не описывается, однако изучение приведенных в работах [116, 117, 149] выходных таблиц и графиков показывает, что для каждого стандарта в базе данных содержится соответствующий N_{JCPDS} , химическая формула и сведения о значительном числе линий спектра — по крайней мере, для всех линий с интенсивностью $I^{s,100} > 10$.

Программа SANDMAN описана в [148, 149], причем детали поисковой процедуры не рассматриваются, но указывается, что в

данной программе реализован такой же алгоритм идентификации, как в системе PDSM CIS. SANDMAN является интерактивной программой, обеспечивающей автоматическую идентификацию спектра образца. Процесс общения исследователя с программой строится на базе набора команд, важнейшими из которых являются следующие:

ввода исследуемого спектра $\{d_h^o, I_h^o\}$ (вручную или из программы предварительной обработки); манипулирования с исходным спектром $\{d_h^o, I_h^o\}$ (вывод, визуализация, корректировка и т. д.); задания поисковых параметров; задания условий поиска по подбазам и химической информации (сплавы, неорганика, органика, минералы, неминералы, база данных NBS, специальная база данных пользователя, подбаза указанного номера и т. д.); SEARCH запуска полной автоматической процедуры поиска и сравнения (аналогичной реализованной в PDSM CIS); IDENTIFY, по которой выполняются те же действия, что и по команде SEARCH, плюс производится подбор такой совокупности фаз из списка вероятных стандартов, суперпозиция которых наилучшим образом соответствует исследуемому спектру; KNOWNs, которая используется для указания перед проведением процедуры поиска номеров известных фаз (например, номера N_{JCPDS} соединения, которое использовалось как внутренний стандарт при съемке спектра); TEST, позволяющая ввести в процедуру сопоставления стандарт указанного номера, если он не был выбран в результате поиска; MATCH, которая используется для указания номеров нескольких стандартов и построения таблицы их сопоставления по линиям со спектром образца.

Перечисленные команды достаточно полно представляют функциональные возможности программы SANDMAN. Особенностью программы является автоматический выбор окна соответствия линий, исходя из ошибок определения положений пиков, полученных в программе предварительной обработки. Выходная таблица результатов подобна формам (4.19), (4.20), но выводимые на печать стандарты строго упорядочены по убыванию основного критерия соответствия (вид которого, как и в работе [132], не приводится). Таким образом, одним из результатов работы программы SANDMAN является построение идентифицирующей функции. Программа затрачивает около 2 мин на анализ тысячи стандартов при идентификации четырехфазного образца.

Программа CASSAM, описанная в [117], позволяет сопоставить в интерактивном режиме исследуемый спектр со спектрами избранных исследователем стандартов. Для сопоставления используется высокоскоростной графический дисплей с большой степенью разрешения. Поскольку в дальнейшем этот способ идентификации получил довольно широкое распространение, приведем одну из возможных форм выдачи информации на экран дисплея (рис. 4.1).

Стандартные спектры выводятся на экран дисплея из той же базы данных, которую использует программа SANDMAN. Про-

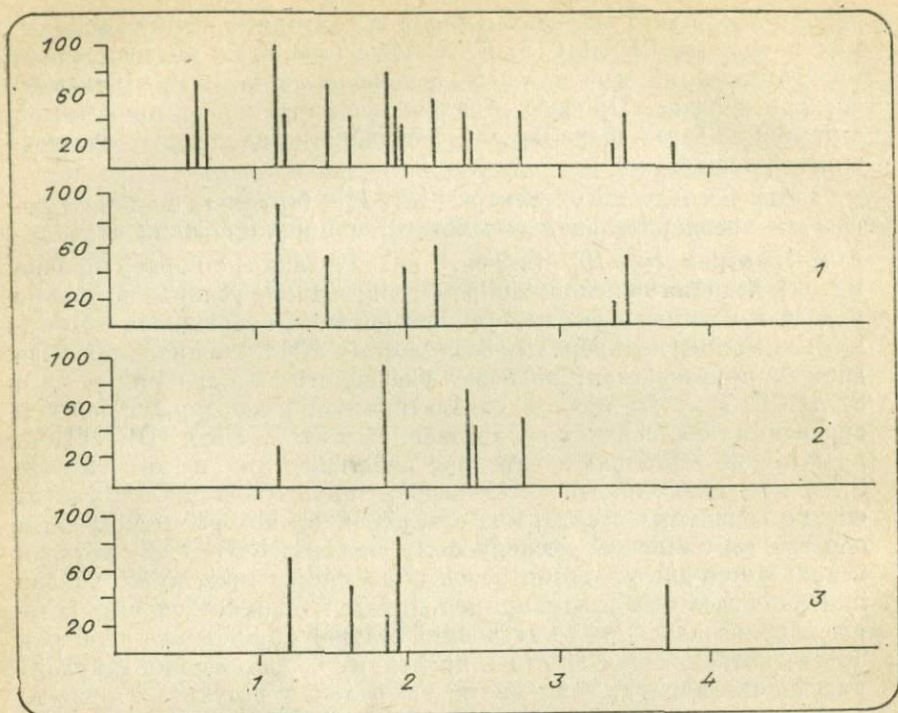


Рис. 4.1. Сопоставление на экране дисплея исследуемого спектра (вверху) со спектрами трех стандартов с заданными номерами картотеки JCPDS (1, 2, 3).

грамма CASSAM позволяет сопоставить исследуемый спектр с тремя стандартными спектрами одновременно (как это схематически изображено на рис. 4.1); в других подобных системах экран дисплея может делиться не на четыре, а на две части. В этом случае возможно только попарное сопоставление спектров образца и эталона.

CASSAM позволяет осуществлять графическое сопоставление спектров в двух формах. В первом случае исследуемый спектр переводится в вид штрих-диаграммы и сопоставляется со штрих-диаграммами стандартов; именно такая ситуация изображена на рис. 4.1. Во втором случае из штрих-диаграмм стандартов реконструируется некоторое приближение к реальным спектрам путем искусственного размытия их линий — например, с помощью распределения Гаусса. Тогда реальный спектр образца может быть сопоставлен с восстановленными спектрами стандартов.

Программа CASSAM позволяет также производить ряд манипуляций с сопоставляемыми спектрами. Исследуемый спектр может быть исправлен на фон, предусмотрено также разделение $K_{\alpha_1-\alpha_2}$ -

дублетов и исключение K_{α_2} -компонент. Возможно также масштабирование интенсивностей линий и варьирование их ширины. Реализована методика графического вычитания спектра стандарта из спектра образца. В этом случае в верхней части экрана дисплея размещается исследуемый спектр, под ним — спектр вычитаемого стандарта и ниже — результаты вычитания. Вычитание стандартов можно производить последовательно. Оно осуществляется в двух описанных выше формах — для штрих-диаграмм и для реальных спектров.

4.6. СИСТЕМА ФИРМЫ СИМЕНС IDENT

К 1980 г. специалистами фирмы Сименс (Siemens AG, ФРГ) был разработан обширный пакет программ, обеспечивающих обработку на мини-ЭВМ результатов рентгendifракционного исследования поликристаллов (Siemens DIFFRAC System). Пакет DIFFRAC позволял осуществить автоматический качественный РФА в двух вариантах. В первом поиск осуществлялся по полной базе данных JCPDS, содержащей к этому времени сведения для 35 000 соединений. Однако он требовал значительного машинного времени (от 30 мин до 1 ч) и сравнительно редко применялся на практике. Данные обстоятельства послужили отправной точкой для разработки в рамках DIFFRAC второго варианта программного обеспечения для качественного РФА — системы IDENT [120].

Система IDENT обеспечивает проведение интерактивной идентификации по специализированной базе данных, включающей спектры соединений, необходимых исследователю в практической деятельности. Система включает набор программ, обеспечивающих управление базой данных, получение на приборе и предварительную обработку (поиск пиков) рентгendifракционных спектров и решение задачи идентификации. Вместе с остальными программами пакета DIFFRAC система поставляется серийно Аналитическим рентгеновским отделом фирмы Сименс.

База данных. Детальная структура поискового файла [120] не описана. Отмечается, что для каждого стандарта в базе данных хранится следующая информация: номер стандарта, название соединения, список входящих в соединение элементов — до 10 (приводится вместо химической формулы соединения), положения трех сильнейших линий, корундовое число, спектр соединения. Спектр может содержать до 50 пар d^s, I^s , причем положения линий задаются в целом виде как $\frac{10\,000}{d^s}$ (см. формулу (4.10)), а интенсивности — в стандартной столбальной шкале. Рекомендуются авторами [120] размер специализированной базы данных — до 500 соединений.

Служебные программы обеспечивают управление базой данных. Программа FETSN позволяет переписать в специализированную

базу требуемые машинные карточки из полной базы JCPDS путем указания их номеров N_{JCPDS} . При переписи производится преобразование карточек JCPDS в принятый в системе IDENT формат. Программа CORRECT обеспечивает корректировку информации для любого стандарта, а также добавление в машинную карточку IDENT специфических данных, которых не имеется в исходных карточках JCPDS. Важнейшим из таких данных является корундовое число.

Программа ADD используется для расширения специализированной базы путем добавления в нее новых машинных карточек. Информация о добавляемых стандартах может быть взята из картотеки JCPDS, литературных источников, собственных измерений.

Наконец, программа LIST позволяет вывести базу данных на печать полностью или частично. Вывод можно осуществлять путем задания списка элементов или заданием нужных номеров стандартов. Фактически LIST является поисковой программой, обеспечивающей выборку нужных стандартов по их химическому составу или по номерам.

Алгоритм идентификации. Не останавливаясь на программах системы IDENT, обеспечивающих измерение и предварительную обработку спектров (программы MEASURE и PEAK SEARCH [120]), перейдем непосредственно к процедуре идентификации, реализуемой программой SEARCH + MATCH. Программа рассчитана на идентификацию не более пяти фаз за один проход. Идентификация осуществляется в два этапа. На первом этапе сравниваются спектры стандартов и образца по положениям трех сильнейших линий стандартов, по-видимому, в рамках общепринятого условия (4.11). Предварительный этап описан в [120] очень кратко, поэтому остается неясным условие пропуска стандартов на второй этап идентификации; вероятно, для этого необходимо достаточно хорошее совпадение двух или трех линий.

На втором этапе для каждого вероятного стандарта вычисляется комбинированный критерий соответствия следующего вида:

$$F = X \cdot \frac{C}{0,1 + F(I)} \cdot \frac{n_{\text{совп}}}{n_0} \cdot \frac{n_{\text{совп}} - n_{\text{несовп}}}{n_s} \quad (4.21)$$

Критерий F является произведением четырех составляющих, которые имеют следующий смысл. X обеспечивает возрастание критерия F для стандартов, химический состав которых соответствует составу образца. Величина X принимает три значения: $X = 1$, если химические данные не используются; $X = 2$, если известен один химический элемент и по нему имеется совпадение составов стандарта и образца; $X = 3$, если известны два и более элементов и по ним имеется совпадение.

Второй член содержит концентрацию C и критерий близости спектров по интенсивности $F(I)$. Величина C рассчитывается через корундовое число; при отсутствии в карточке корундового

числа вместо него берется единица. Критерий $F(I)$ равен усредненной по числу линий стандарта n_s разности логарифмических интенсивностей $I^{s,lg}$ и $I^{o,lg}$. В [120] $F(I)$ дается в форме

$$F(I) = \frac{1}{n_s} \sum_{i=1}^{n_s} \lg \frac{CI^{s,100}}{I^{o,100}}. \quad (4.22)$$

Здесь интенсивности линий стандарта и образца приведены в сопоставимую шкалу отсчета путем умножения $I^{s,100}$ на концентрацию C .

В случае идеального совпадения $F(I) = 0$. Чтобы не допустить во втором члене в (4.25) деления на нуль, в знаменатель дроби добавляется величина 0,1.

Третий член в (4.21) является частотным критерием типа (4.12), но $n_{совп}$ относится не к числу линий стандарта, а к полному числу линий спектра образца.

Четвертый член равен отношению разности совпавших и несовпавших линий стандарта к числу линий стандарта. Для того чтобы комбинированный критерий F был положительен, необходимо совпадение более 50% линий стандарта: тогда $(n_{совп} - n_{несовп}) > 0$.

Стандартная фаза считается вероятной, если значение F положительно, т. е. реально, превосходит заданное в программе малое число 0,01:

$$F > F_{гр} = 0,01. \quad (4.23)$$

Значение $F_{гр}$ можно изменить.

Возможно использование методики вычитания вероятной стандартной фазы с повторением процедуры идентификации для остатка спектра образца. Цикл вычитания проводится до четырех раз, так как за один проход программы можно определить не более пяти фаз.

Значения комбинированного критерия упорядочиваются по убыванию величины F ; таким способом строится идентифицирующая функция (FOM). Первые 25 значений идентифицирующей функции могут быть выведены и проанализированы исследователем. Более подробные сведения о выходных форматах в [120] отсутствуют.

Интерактивный режим системы IDENT предоставляет исследователю следующие возможности: 1) сравнение на экране дисплея исследуемого спектра и спектра указанного стандарта; 2) отбраковку неудовлетворяющих стандартов в процессе поиска; 3) указание на включение в процедуру сравнения интересующего стандарта; 4) согласование интенсивностей соответствующих линий образца и стандарта.

Система IDENT требует довольно значительных вычислительных мощностей. Система разработана в лаборатории исследования материалов фирмы Сименс на базе мини-ЭВМ PDP-11/34, рабо-

тающей в комплексе с автоматическим дифрактометром Siemens D 500. Мини-ЭВМ имеет оперативную память 64К слов, два НМЛ объемом по 5М байт, дисплей, устройство печати и цветной графопостроитель. Язык программирования — фортран; IDENT функционирует в рамках штатной операционной системы RT-11. Время анализа составляет 1—2 мин; по-видимому, это время следует отнести к идентификации по базе данных рекомендованного в [120] объема — до 500 стандартов.

4.7. СИСТЕМА ФИРМЫ ИВМ

Пакет прикладных программ фирмы ИВМ, предназначенный для управления прибором и решения задач предварительной и окончательной обработки рентгендифракционных спектров [113], был создан в основном к 1981 г. Частью этого пакета является система автоматической идентификации, разработанная в Научно-исследовательской лаборатории фирмы ИВМ (IBM Research Laboratory) Пэрришем и др. [108, 138]. Система реализована на универсальной ЭВМ IBM-370/168 и на мини-ЭВМ фирмы IBM Series/1 (интерактивный вариант системы). Состав мини-ЭВМ IBM Series/1 (необходимый для функционирования системы идентификации: объем оперативной памяти — не менее 96К байт; объем НМД — 13,9 М байт (используется для хранения программ и исходной базы данных JCPDS); графический дисплей, алфавитно-цифровой дисплей, устройство печати. Язык разработки программ — фортран.

База данных системы ИВМ строится из базы данных JCPDS; вероятно, обе базы эквивалентны в части объема информации для одного стандарта. Для минимизации времени поиска база данных системы ИВМ содержит четыре подбазы: неорганика (IN), органика и металлоорганика (OR), минералы (MI), металлы и сплавы (ME). Каждая подбаза содержит три секции: в первой собраны наиболее часто встречающиеся стандарты, во второй — все стандарты

Таблица 4.1

Распределение стандартов по секциям подбаз системы ИВМ

Подбаза	Секция			Полный объем
	1	2	3	
IN	2332	3870	19 193	25 395
OR	1346	878	8 547	10 771
MI	608	313	2 099	3 020
ME	363	471	4 815	5 649

высокой достоверности (помеченные звездочкой), не включенные в первую секцию, в третьей — все оставшиеся стандарты. В табл. 4.1 показано распределение стандартов по секциям четырех подбаз. Поскольку некоторые стандарты могут входить в несколько подбаз, то полный объем базы ИВМ составляет около 45 000 карточек, т. е. на 30% больше объема исходной базы JCPDS (34 000 стандартов).

Рассмотренное устройство базы данных системы IBM позволяет, в случае успешной идентификации, ограничивать процесс поиска только первой или первой и второй секциями. База данных может также включать специализированные подбазы пользователя, содержащие его собственные данные или данные из любых литературных источников.

Служебные программы. Программа SMUTIL 1 позволяет накопить на НМД архив спектров соединений неизвестного состава, получаемых непосредственно с дифрактометра. Архив может включать до 380 спектров в формате $\{d_k^o, I_k^o\}$. Любой из них может быть вызван из архива и введен в систему идентификации.

Программа SMUTIL 2 используется для создания в рамках базы данных специализированных подбаз пользователя.

Программа SMLOOK обеспечивает реконструкцию спектра стандарта указанного номера, т. е. переход от штрих-диаграммы $\{d_k^s, I_k^s\}$ к реальному виду спектра путем размытия линий. Такой реконструированный спектр стандарта может быть выведен на графический дисплей (полностью или по частям) и выдан на печать. Напомним, что подобная же операция реализуется в программе CASSAM системы фирмы Филипс.

Алгоритм идентификации. Спектр исследуемого образца может быть введен в ЭВМ тремя способами: непосредственно с дифрактометра (после соответствующей предварительной обработки), из архива (с НМД), вручную с алфавитно-цифрового дисплея. задается также ряд параметров поиска, в том числе указание на требуемую подбазу, значение окна совпадения линий, значение $l_{гр}$, число стандартов с наивысшими значениями идентифицирующей функции, которые необходимо вывести на печать (но не >20), химическая информация.

В процессе поиска последовательно просматривается указанная в параметрах поиска подбаза данных. Идентификация осуществляется в два этапа. Предварительный этап проходят только те стандарты, у которых не менее трех сильнейших линий совпадает со спектром образца, т. е. $l > l_{гр} = 3$. Величину $l_{гр}$ можно изменить. Рекомендуется задавать $l_{гр} < 3$ при работе со спектрами, полученными от текстурированных образцов. На предварительном этапе можно осуществлять селекцию по химическому составу. Для этого в параметрах поиска могут быть заданы два списка присутствующих и отсутствующих элементов (до восьми элементов в каждом списке). Отсутствующие могут быть заданы также в виде трех групп, включающих по два элемента каждая, которые указывают на начало и конец совокупности отсутствующих в образце элементов. На предварительном этапе идентификации исключаются стандарты, которые не содержат хотя бы одного элемента из списка присутствующих либо содержат один и более элементов из списка отсутствующих. Необходимо отметить, что система идентификации IBM ориентирована на обработку результатов, получаемых в процессе комплексного исследования образца на рентге-

новском флуоресцентном спектро- и дифрактометре, поэтому элементный состав образца, как правило, известен.

Для стандартов, прошедших фильтры предварительного этапа, на заключительном этапе рассчитывается значение идентифицирующей функции F (FOM). Формула для расчета критерия F в [108, 137] не приводится, однако можно попытаться хотя бы приблизительно восстановить ее по тексту статьи [128]. При вычислении F учитываются следующие факторы: доля совпавших линий в спектре образца $n_{\text{совп}}/n_o$; доля совпавших линий стандарта $n_{\text{совп}}/n_s$; критерий близости спектров по $d - F(d)$; критерий близости спектров по $I - F(I)$; отношение суммы интенсивностей совпавших линий стандарта к полной интенсивности линий стандарта

$$\sum_{i=1}^{n_{\text{совп}}} I_i^s / \sum_{j=1}^{n_s} I_j^s.$$

Таким образом, комплексный критерий соответствия спектров в системе IBM представляется, возможно, в форме

$$F = \frac{n_{\text{совп}}}{n_o} \cdot \frac{n_{\text{совп}}}{n_s} \cdot F(d) \cdot F(I) \left(\sum_{i=1}^{n_{\text{совп}}} I_i^s \left| \sum_{j=1}^{n_s} I_j^s \right. \right). \quad (4.24)$$

Как и во всех рассмотренных ранее методиках, критерий (4.24) рассчитывается для линий стандарта, попадающих в интервал задания (съемки) спектра образца. Вид критериев $F(d)$, $F(I)$, вероятно, подобен приведенным выше выражениям (4.16), (4.17), (4.21).

Выходные форматы и работа с ними. После заключительного этапа идентификации на печать выводится таблица, содержащая данные для каждого вероятного стандарта:

Номер по порядку	Химическая формула или название	Признак достоверности и специальные пометки спектра	N_{JCPDS}	F	$n_{\text{совп}}$	n_s .
------------------	---------------------------------	---	--------------------	-----	-------------------	---------

(4.25)

Строки выходной таблицы упорядочены по убыванию F .

Затем исследователем производится окончательные уточнения фазового состава образца в режиме диалога с системой с использованием графического дисплея. Фактически эта часть процедуры качественного РФА можно рассматривать как третий этап идентификации. Интерактивные возможности, предоставляемые системой IBM на этом этапе, в общих чертах совпадают с реализованной в программе CASSAM методикой.

Для характеристики быстродействия системы IBM в [108, 137] приводятся следующие примеры. Идентификация неорганической трехфазной смеси, спектр которой содержал 20 линий,

заняла 4 с машинного времени на ЭВМ IBM-370/168 по первой секции подбазы неорганических соединений IN (2332 стандарта). Идентификация органической трехфазной смеси (53 линии) на IBM-370/168 по первой секции подбазы OR (1346 стандартов) также длилась 4 с. Химическая информация в этих двух случаях не использовалась. Для мини-ЭВМ IBM Series/1 показатели времени значительно больше. Так, идентификация трехфазного неорганического образца (43 линии) по первой секции подбазы IN (2332 стандарта) продолжается 257 с без задания химической информации и 90 с при известном элементном составе образца.

4.8. СИСТЕМА СНАЙДЕРА

В 1980 г. на очередной Денверской конференции по применению рентгеновского анализа Снайдером была рассмотрена оригинальная система, разработанная в Институте керамики (New York State College of Ceramics Alfred Iniversity) [155]. Система ориентирована на широкий класс мини-ЭВМ в относительно небольшой комплектности; так, необходимый объем НМД составляет не более 5 М байт. Работа Снайдера [155] представляет большой интерес, поскольку в ней сделан ряд обобщений относительно реализации процедуры идентификации на мини-ЭВМ и, главное, рассмотрено большое число практических примеров.

База данных системы Снайдера построена в соответствии со следующими принципами.

1. Для хранения базы данных большого объема (в принципе — всей базы данных JCPDS) используется специальный двоичный формат для сжатого представления пар d^s , I^s .

2. Для ускорения процесса поиска в базе данных выделяются три подбазы, содержащие наиболее часто встречающиеся при анализе стандарты, часто встречающиеся стандарты и все стандарты JCPDS.

3. На предварительном этапе поиска используется инвертированный файл.

Значения d^s в диапазоне 0,488—1000 Å переводятся в шкалу H (в целые d -коды, согласно терминологии [155]) по формуле (4.6). Величины H^s могут принимать любое целое значение в диапазоне 1—2048, для кодировки которого достаточно 11 двоичных разрядов. Преобразование (4.6) огрубляет значения d^s , однако средняя ошибка этого огрубления в шкале углов 2θ составляет 0,025°, а максимальная не превосходит 0,05° (для излучения меди). В 16-разрядной ячейке мини-ЭВМ можно разместить значение H^s в 11-разрядном представлении; 5 оставшихся разрядов используются для размещения значения I^s , которое дается от 0 до 30 единиц. Кроме совокупности упакованных в машинное слово значений H^s , I^s , в машинной карточке базы данных содержатся также

химическая формула, признак стандарта (минерал, неорганическое соединение и т. д.) и номер N_{JCPDS} стандарта. Рассмотренный вариант упаковки позволяет сжать базу данных JCPDS (28 севт, 35000 стандартов) от 15 до 2,5 М байт.

База данных системы Снайдера состоит из трех подбаз. Первая подбаза MICRO включает 298 наиболее часто встречающихся стандартов, отобранных Фривелом (система ZRD-1974 [100]). Вторая подбаза MINI содержит около 2500 стандартов — выборку для наиболее распространенных соединений JCPDS. Третья подбаза MAXI включает 28 секций, соответствующих сетам JCPDS, и содержит взятые из полной базы данных JCPDS сведения о 35 000 стандартов. Каждая из трех подбаз состоит из двух файлов: указателя и основного. Файл-указатель является инвертированным файлом типа Ханавальта; в нем содержатся упорядоченные значения сильнейших линий стандартов $d^{s,(1)}$ и соответствующие им N_{JCPDS} . Основной файл является прямым и содержит машинные карточки в рассмотренном выше сжатом формате.

Алгоритм идентификации. Исследуемый спектр вводится в мини-ЭВМ вручную либо накапливается в результате работы сопряженного с ЭВМ автоматического дифрактометра и поступает в систему идентификации с выхода программы предварительной обработки. Имеется возможность введения параметров поиска — химической информации, признака соединения, граничного значения критерия соответствия $F_{\text{гр}}$.

Алгоритм идентификации в системе Снайдера исключительно сложный, для его пояснения дается рис. 4.2, отсутствующий в оригинальной работе [155]. Алгоритм является традиционно двухэтапным, но эти этапы могут многократно повторяться в зависимости от складывающейся при поиске ситуации. Первая попытка идентификации осуществляется при помощи малой подбазы MICRO. С НМД считывается инвертированный файл-указатель MICRO. Устанавливается $k = 1, 2, 3$, означающий, что будет проводиться поиск по трем первым сильнейшим линиям спектра образца $H^{o,(1)}, H^{o,(2)}, H^{o,(3)}$. Значение окна совпадения линий устанавливается в шкале 2θ равным $\Delta\theta = 0,1^\circ$ и пересчитывается в шкалу H . Для первой линии спектра образца $H^{o,(1)}$ с помощью неравенства типа (4.11) по файлу-указателю определяются блоки основного файла MICRO, содержащие такие стандарты, у которых первые линии $H^{s,(1)}$ совпадают с первой линией $H^{o,(1)}$.

На заключительном этапе идентификации с НМД считываются нужные блоки основного прямого файла MICRO и для содержащихся в них вероятных стандартов проверяется выполнение трех условий: 1) совпадение признаков стандарта и образца (если признаки указаны); 2) совпадение по химическому составу (если задана химическая информация); 3) совпадение по положениям для всех сильнейших линий стандарта с $I^{s,100} \geq 50$.

Эти три проверки сходны с Z , R и D — тестами системы Фривела. Для прошедших проверку стандартов вычисляются значения

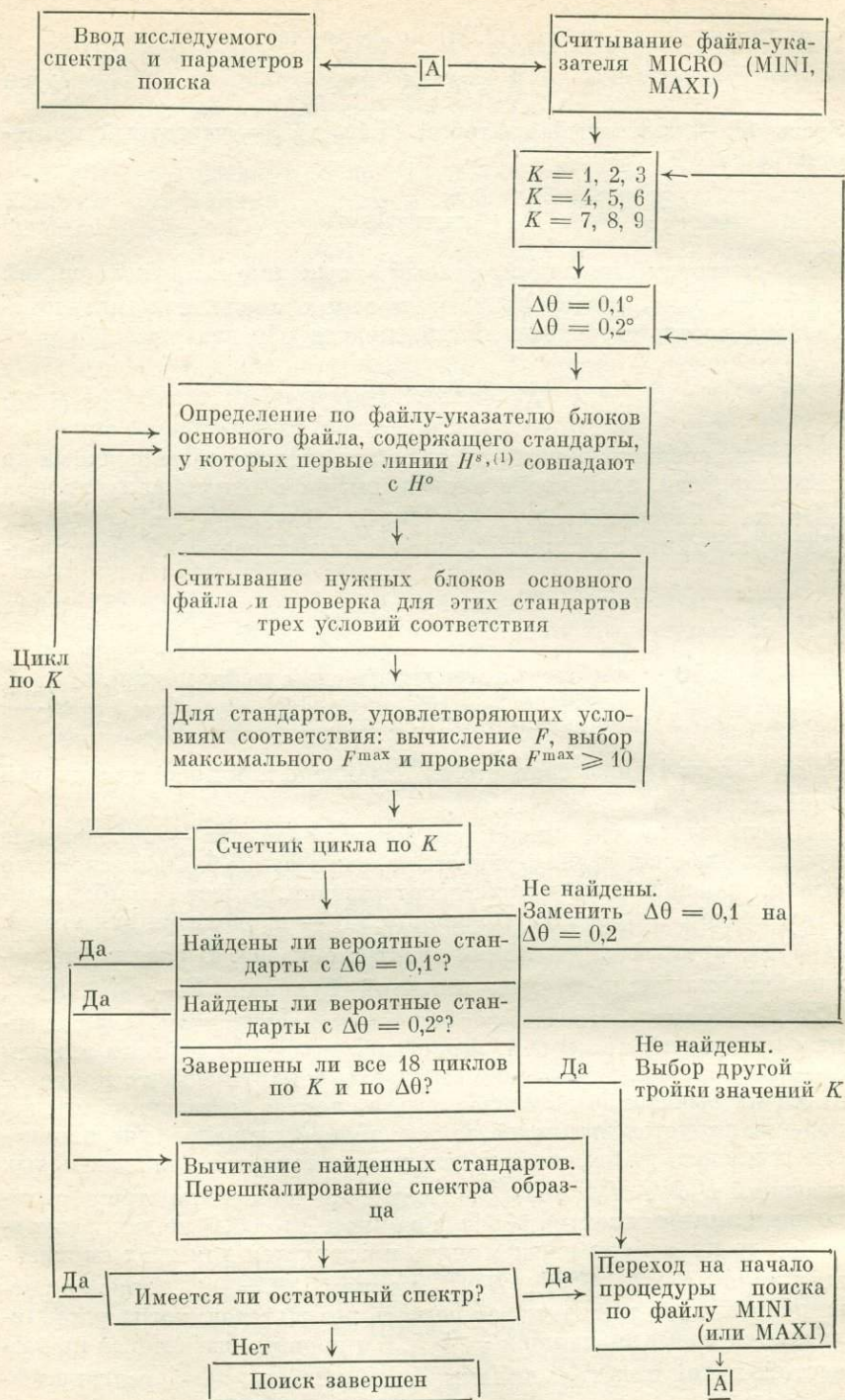


Рис. 4.2. Структура поиска в системе Снайдера.

функции идентификации (FOM) по формуле

$$F = K_1 \cdot K_2 \cdot K_3^2. \quad (4.26)$$

Здесь K_1 — частотный критерий (4.12), K_2 — частотный критерий вида

$$K_2 = \frac{n_{\text{совп}}}{n_0} \cdot 100\% \quad (4.27)$$

и K_3 — частотный критерий, равный отношению к n_s числа линий стандарта $n'_{\text{совп}}$, интенсивности которых совпали с интенсивностями соответствующих линий спектра образца

$$K_3 = \frac{n'_{\text{совп}}}{n_s} \cdot 100\%. \quad (4.28)$$

Линии спектров образца и стандарта считаются совпавшими по интенсивности, если после приведения их в одну шкалу отсчета разница интенсивностей $|I_i^o - I_i^s| \leq 5$. Величина $n'_{\text{совп}} \leq n_{\text{совп}}$. Подчеркнем, что критерий K_3 вводится в выражение (4.31) в квадрате.

Среди вычисленных значений F выбирается максимальное и для него проверяется условие:

$$\begin{aligned} F_1^{\text{max}} &\geq 10 \quad (\text{при отсутствии химической информации}), \\ F_1^{\text{max}} &\geq 7 \quad (\text{при задании химической информации}). \end{aligned} \quad (4.29)$$

Здесь индекс 1 указывает, что F_1^{max} найдено при поиске по первой линии спектра образца $H^{o,(1)}$.

После этого вся описанная выше процедура — определение нужных блоков по файлу-указателю, считывание блоков основного файла, проверка трех условий соответствия, расчет F , выбор максимального F и проверка условия (4.29) — проводится для второй $H^{o,(2)}$ и третьей $H^{o,(3)}$ линий спектра образца. В результате может быть получено до трех значений функции F — F_1^{max} , F_2^{max} , F_3^{max} , которые соответствуют вероятным стандартам, определенным при поиске по первой, второй и третьей сильнейшим линиям спектра образца. Если такие стандарты обнаружены, то производится их вычитание из исследуемого спектра, остаточный спектр перешкалируется на новое максимальное значение интенсивности и процедура поиска и сравнения повторяется для трех новых значений $H_{\text{нов}}^{o,(1)}$, $H_{\text{нов}}^{o,(2)}$, $H_{\text{нов}}^{o,(3)}$. Если же поиск по трем сильнейшим линиям спектра образца с окном $\Delta\theta = 0,1^\circ$ не дал результатов, то процедура поиска опять повторяется для трех сильнейших линий $H^{o,(1)}$, $H^{o,(2)}$, $H^{o,(3)}$ с окном $\Delta\theta = 0,2^\circ$. Поиск с расширением окна также может не дать результатов; тогда устанавливаются новые значения $k = 4, 5, 6$ и производится поиск по четвертой, пятой и шестой линиям $H^{o,(4)}$, $H^{o,(5)}$, $H^{o,(6)}$ исследуемо-

го спектра, сначала с окном $0,1^\circ$, затем — с $0,2^\circ$. Если успех не достигнут, то производится поиск с $k = 7, 8, 9$ по седьмой, восьмой и девятой линиям спектра образца.

Таким образом, рассмотренная процедура поиска и сравнения может осуществляться до 18 раз для линий от $H^{0,(1)}$ до $H^{0,(9)}$ с двумя значениями окна. Если поиск не дал результатов, то, очевидно, в малой подбазе MICRO отсутствуют составляющие образец фазы. Тогда процедура повторяется сначала для подбазы MINI, а затем — MAXI. Естественно, случаи, когда восемнадцатикратное сканирование подбазы MICRO не позволяет определить вероятный стандарт, сравнительно редки. Как правило, поиск по MICRO позволяет определить хотя бы одну, а, возможно, две — четыре фазы. После вычитания вероятных стандартов и исчерпывания возможностей поиска по подбазе MICRO можно произвести поиск по подбазе MINI, если имеется остаточный спектр. Поиск по MAXI также целесообразно делать только в том случае, если после сканирования MINI остаточный спектр образца не сведен практически к нулю. Иерархический поиск Снайдера позволяет, путем сканирования подбаз MICRO, MINI и MAXI, последовательно исключать из исследуемого спектра спектры составляющих фаз, причем исключение осуществляется в порядке понижения вероятности возможного присутствия фаз в образце.

Рассмотренная процедура иерархического поиска резко снижает время идентификации. Система Снайдера выполняет поиск по трем сильнейшим линиям исследуемого спектра в среднем за 5 с по файлу MICRO (~ 300 стандартов), за 27 с по файлу MINI (~ 2500 стандартов) и за 4,22 мин по файлу MAXI ($\sim 35\,000$ стандартов). Однако на практике поиск по трем сильнейшим линиям достаточно часто бывает несостоятельным; возможны ситуации, когда поиск даже по девяти сильнейшим линиям в подбазе MICRO не дает результатов [155]. Поэтому Снайдер, после обсуждения ряда контрольных примеров, дает более реалистические оценки среднего времени идентификации: около 4 мин при иерархическом поиске и около 25 мин при поиске только по полной подбазе MAXI.

В [155] не приводится вид выходного формата, но, по-видимому, в результаты идентификации для каждого вероятного стандарта включаются соответствующий номер N_{JCPDS} , химическая формула и показатели качества соответствия — значения F , условной концентрации, частотных критериев.

4.9. СИСТЕМА ФИАЛЫ

Важный этап развития автоматического качественного РФА связан с работами чехословацкого исследователя Фиала [95—98], который внес значительный вклад в формирование математической модели процедуры идентификации.

В отличие от рассмотренных выше систем качественного РФА, главное в работах Фиала заключается не в создании конкретной

программной системы или базы данных, а в попытке математического анализа проблемы идентификации и подборе подходящего математического аппарата, в рамках которого данную проблему удобно описывать. Таким аппаратом является векторная и матричная алгебра, которая широко применяется в рамках методов распознавания образов или решения задач идентификации в смежных областях — например, в масс- и ИК-спектроскопии [25]. Фиала внес данные представления в область идентификации рентгendifракционных спектров, положив тем самым начало новому, аналитическому подходу к проблеме. Итак, рассмотрим основные моменты подхода Фиала.

Рентгendifракционному спектру $\{d, I\}$ соответствует n -мерный вектор, который формируется следующим образом. Диапазон анализа D от 1 до 10 Å равномерно разбивается на промежутки по $\delta d = 0,01$ Å, длина которых в среднем соответствует точности определения d . Если в i -й промежуток попадает линия спектра, то значение i -й компоненты вектора полагается равным интенсивности линии, в противном случае — нулю. Таким образом, спектр представляется в виде 900-мерного вектора:

$$I(\tilde{d}) = (0, \dots, 0, I_1, 0, \dots, 0, I_2, 0, \dots, 0, I_{n_s}, 0, \dots, 0). \quad (4.30)$$

Здесь \tilde{d} обозначает аргумент d , принимающий только дискретные значения 1; 1,01; 1,02; и т. д.; n_s — число линий в спектре.

Вектор (4.30) соответствует штрих-диаграмме $\{d, I\}$ (см. рис. 1, б); можно, однако, заменить каждую линию штрих-диаграммы прямоугольным или треугольным распределением [97] (см. рис. 1, в), задаваемым в пределах полуширины или ширины линии по основанию. Такой реконструированный спектр также легко представить в виде вектора $I(\tilde{d})$, i -я компонента которого равна:

$$I_i = I(\tilde{d}_i) = \begin{cases} 0 & \text{вне области задания линии,} \\ f(\tilde{d}) & \text{— в области задания линии.} \end{cases} \quad (4.31)$$

Здесь $f(\tilde{d})$ является функцией, аппроксимирующей линию прямоугольным, треугольным или гауссовым распределением.

Совокупность векторов спектров рассматривается в n -мерном унитарном векторном пространстве R_n , которое в пределе, при $\delta d \rightarrow 0$, переходит в бесконечномерное гильбертово пространство. Тогда вектор, соответствующий спектру образца, можно представить как суперпозицию векторов эталонов:

$$I^o(\tilde{d}) = \sum_{j=1}^M C_j I_j^s(\tilde{d}). \quad (4.32)$$

Здесь j нумерует стандарты в базе данных, полное число которых равно M .

Считается, что образец может состоять только из фаз, содержащихся в базе данных. Коэффициенты C_j , которые соответствуют отсутствующим в образце фазам, равны нулю. Коэффициенты для входящих в образец фаз имеют положительные значения.

Выражение (4.32), записанное для каждого дискретного значения \tilde{d} , разворачивается в систему n линейных уравнений ($n = 900$) с M неизвестными C_1, C_2, \dots, C_N ($M \sim 3 \cdot 10^4 - 4 \cdot 10^4$). Искомые коэффициенты $\{C_j\}$ являются решением этой системы, но не единственным, так как число неизвестных больше числа уравнений — $M > n$. Можно увеличить точность задания векторов путем уменьшения δd до 0,001 или 0,0001 Å и перехода к размерностям 9 000—90 000, но это физически не обосновано, так как экспериментальная техника пока не позволяет получать значения d с точностью до 0,0001 Å. Однако если бы даже удалось согласовать значения n и M путем увеличения n или уменьшения M , то все равно осталась бы проблема решения системы уравнений (4.32) большой размерности — порядка сотен или тысяч. Реализовать такую процедуру на ЭВМ весьма трудно.

Для решения проблемы Фиала предлагает иной путь, связанный с радикальным уменьшением M в (4.32) до значений $M \ll n$. Для этого предлагается с помощью традиционных методов, например процедуры Джонсона—Ванда, отобрать список вероятных стандартов S_s , содержащий $N_s \sim 30-40$ фаз. Тогда (4.30) можно заменить системой уравнений, записанной только для вероятных стандартов:

$$I^o(\tilde{d}) = \sum_{j=1}^{N_s} C_j I_j^{s, \text{вер}}(\tilde{d}), \quad (4.33)$$

где $N_s \ll n$. Систему (4.33) можно решать с помощью метода наименьших квадратов. Таким образом, будет найдено наилучшее приближение спектра образца в смысле минимума нормы вектора, представляющего разность между левой и правой частями выражения (4.33).

Из изложенного выше ясно, что метод Фиала не позволяет работать с базой данных большого объема и минимизировать время поиска. Его можно рассматривать как дополнительный этап, завершающий процедуры идентификации, рассмотренные выше в разделах 4.2—4.8. Однако эти процедуры не гарантируют корректности списка вероятных фаз по крайней мере в двух ситуациях.

1. В список S_s может не войти истинная фаза, что приведет к искажению решения системы (4.33) в части остальных истинных фаз, попавших в список.

2. В список S_s , кроме истинной фазы, могут попасть стандарты со сходными спектрами, что приведет к плохой обусловленности решения системы (4.33).

Оба указанных фактора могут, естественно, действовать одновременно. Таким образом, использовать рассмотренную методи-

ку в качестве основы окончательного этапа идентификации следует с большой осторожностью.

Изложенные выше соображения показывают, что основная ценность работ Фиала заключается не столько в практическом применении соотношений (4.32) и (4.33), сколько в самом методическом подходе к проблеме идентификации в рамках векторной алгебры и теории линейных пространств. Дальнейшее развитие этих представлений, как будет показано ниже, не только позволяет разработать новые эффективные методики идентификации, но и дает исследователю мощное средство изучения структуры и статистических особенностей базы данных.

4.10. СИСТЕМА ФАЗАН

Вторая половина 70-х годов отмечена интенсивным развитием систем автоматической идентификации в СССР. Причины этого явления заключаются в массовом внедрении автоматических и полуавтоматических дифрактометрических систем типа ДРОН, АДП-1, обеспечивающих экспрессное получение спектральных данных для РФА, и значительном росте качественных и количественных показателей отечественных ЭВМ. В этот период был разработан ряд систем идентификации, наиболее крупные из которых, системы ФАЗАН и АРФА, подробно рассматриваются ниже.

Система ФАЗАН (фазовый анализ) разработана физическим факультетом и Научно-исследовательским вычислительным центром Московского госуниверситета и подробно описана в ряде публикаций [6—11]. Система ориентирована на универсальные ЭВМ БЭСМ-4, БЭСМ-6, ЕС.

База данных. При разработке системы ФАЗАН была создана сравнительно небольшая специализированная база данных, включающая сведения примерно о тысяче стандартов — металлических соединений, сплавов, используемых в машиностроительной промышленности. В дальнейшем, при разработке крупной базы данных БЭРД в рамках системы АРФА, появилась возможность использования данной информации совместно с программами идентификации ФАЗАН. Для этого была создана специальная программа, перекодирующая принятый в БЭРД формат машинной карточки в формат поискового файла ФАЗАН. В настоящее время пользователям системы ФАЗАН на ЕС ЭВМ практически полностью доступна вся информация из базы данных БЭРД (около 10 000 стандартов); структура БЭРД будет подробно рассмотрена в следующем разделе.

База данных системы ФАЗАН содержит два прямых файла специального формата: файл-указатель и файл-библиотеку. Положения линий задаются в шкале H согласно формуле (4.6). Величины интенсивностей в файле-указателе задаются в загрублен-

ной шкале P с четырьмя градациями.

I^{100}	100-71	70-41	40-21	20-0
P	4	2	1	0

Файл-указатель содержит совокупность машинных карточек, упорядоченных по возрастанию номеров N_{JCPDS} . Карточка содержит информацию о пяти сильнейших линиях стандарта и имеет вид

$$N_{JCPDS} \{H_j^s, P_j^s\}. \quad (4.34)$$

Величины интенсивностей в файле-библиотеке задаются в 15-балльной шкале:

$$I^{s,15} = \text{Бл. ц.} (0,15 \cdot I^{s,100}). \quad (4.35)$$

Вид машинной карточки файла-библиотеки:

$$N_{JCPDS}, \{\text{дополнительная информация}\}, \{H_j^s, I_j^{s,15}\}. \quad (4.36)$$

Вместо N_{JCPDS} в (4.34), (4.36) возможна ссылка на другой источник информации. Дополнительная информация может включать химическую формулу или список входящих в стандарт элементов, название, указание на используемое при получении спектра излучения, признак достоверности данных. Список может содержать до 60 линий спектра.

Служебные программы. В систему ФАЗАН входят служебные программы, обеспечивающие построения и редактирование файлов, формирование карточек указателя (4.34) и библиотеки (4.36), диагностику ошибок в стандартных данных, выявление одинаковых стандартов в библиотеке с целью их последующего исключения.

Алгоритм идентификации. В системе ФАЗАН идентификация осуществляется в три этапа. В ЭВМ вводятся с перфокарт исследуемый спектр, параметры поиска (ΔH , $I_{отс}$, возможно также задание химических ограничений) и управляющие коды, определяющие режим работы системы. На первом, предварительном, этапе идентификации с помощью файла-указателя производится анализ по пяти сильнейшим линиям стандартов. Карточки указателя последовательно сопоставляются с исследуемым спектром; совпадение линий определяется по неравенству (4.11). Вероятные стандарты должны удовлетворять следующим условиям: 1) число совпавших линий $n_{совп} = 3$, если имеются три линии стандарта с $P^s > 0$; 2) $n_{совп} = 2$, если имеются только две линии с $P^s > 0$; 3) $n_{совп} = 1$, если имеется только одна линия (первая) с $P^s > 0$ ($P_1^s = 4$); 4) в числе совпавших линий стандарта должна присутствовать хотя бы одна из двух первых линий.

Для удовлетворяющих данным условиям стандартов вычисля-

ется критерий соответствия

$$P = \frac{\sum_{i=1}^{n_{\text{совп}}} P_i^s w(H_i)}{\sum_{j=1}^5 P_j^s w(H_j)}. \quad (4.37)$$

Весовая функция $w(H)$ принимает значения, обратные функциям плотности распределения $N(H)$ в шкале H пяти сильнейших линий стандартов. Стандарт пропускается на второй этап идентификации, если

$$P \geq P_{\text{гр}}, \quad (4.38)$$

где граничное значение $P_{\text{гр}}$ задается исследователем.

На втором этапе идентификации осуществляется работа с файлом-библиотекой. Анализируются вероятные эталоны, прошедшие первый этап. Особенностью системы ФАЗАН является большое число разнообразных критериев соответствия спектров, рассчитываемых на втором этапе идентификации. К их числу относятся [8]:

- 1) частотный критерий K типа (4.12);
- 2) частотный взвешенный критерий $K_{\text{взв}}$, причем в качестве весов линий используются величины их интенсивностей;
- 3) коэффициент корреляции Q спектров образца и стандарта, вычисляемый по линиям спектра образца, свободным от наложений;
- 4) критерии перекрытия спектров образца и стандарта с аппроксимацией линий прямоугольными и треугольными распределениями R_1 и R_2 (см. рис. 1, в);
- 5) вероятностный критерий

$$W = \frac{1 - \prod_{i=1}^{n_{\text{совп}}} \left(1 - \frac{1}{2\Delta H \cdot N(H_i)}\right)}{1 - \prod_{j=1}^{n_s} \left(1 - \frac{1}{2\Delta H \cdot N(H_j)}\right)}. \quad (4.39)$$

Произведения функции плотности распределения линий $N(H)$ на величину $2\Delta H$ дают полное число линий стандартов библиотеки, попадающее в интервалы $2\Delta H$ с центрами в H_i, H_j ;

6) условная концентрация C , которая вычисляется, в отличие от (4.4), как среднее арифметическое отношений интенсивностей совпавших линий образца и стандарта:

$$C = \frac{1}{n_{\text{совп}}} \sum_{i=1}^{n_{\text{совп}}} I_i^o / I_i^s. \quad (4.40)$$

Знак штрих означает, что в (4.40) учитываются только свободные от наложений линии исследуемого спектра: таким образом, $n'_{\text{совп}} \leq n_{\text{совп}}$, где $n_{\text{совп}}$ — полное число совпавших линий стандарта. Линия образца считается наложенной, если $(I_i^o/I_i^s) > 1,3$. Кроме критерия C , который имеет смысл средней условной концентрации, вычисляется оценка C_{max} , равная максимальному отношению I_i^o/I_i^s .

Идентифицирующая функция F , в порядке убывания которой упорядочиваются вероятные эталоны, строится следующим образом:

$$F = QR_1W. \quad (4.41)$$

Вычисление указанных выше критериев производится по линиям стандарта с интенсивностями $I^s \geq I_{\text{отс}}$ и расположенных в пределах задания исследуемого спектра. На втором этапе можно использовать также для отбора химическую информацию. Стандарт считается прошедшим второй этап идентификации, если значения рассчитанных для него критериев соответствия не ниже некоторых граничных величин. Таким образом, на втором этапе формируется список S_e вероятных стандартов, ранжированных по F . Этот список, вместе с критериями соответствия, выдается на печать и служит исходной информацией для третьего этапа идентификации. Максимальная длина списка S_e — 50 фаз. Перед обращением к третьему этапу список может быть дополнительно откорректирован исследователем.

Третий этап идентификации позволяет уточнить фазовый состав путем решения системы уравнений вида (4.33) для N_e вероятных стандартов, перечисленных в списке S_e . Поскольку, как указывалось в разделе 4.9, система (4.33) может быть плохо обусловленной, для ее решения используется метод регуляризации Тихонова [73]. Стандарты, для которых значения C_j , найденные при решении (4.33), отрицательны или не превосходят некоторого малого граничного значения, отбрасываются, и система уравнений решается повторно по уточненному списку вероятных фаз. Третий этап идентификации не является обязательным, его можно не проводить, если полученные на втором этапе результаты удовлетворяют исследователя.

Выходные форматы и дополнительные возможности. На печать выводится следующая информация:

1) таблица результатов идентификации на втором этапе, содержащая для каждой вероятной фазы N_{JCPDS} дополнительную информацию (в соответствии с формой (4.43)), значения критериев соответствия;

2) таблица окончательных результатов идентификации (с учетом третьего этапа), в которой для каждой вероятной фазы приводится значение C_j и соответствующая среднеквадратичная погрешность;

3) таблица сопоставления исследуемого спектра со спектрами вероятных фаз (в шкале H , I).

В системе ФАЗАН имеется возможность вывода на алфавитно-цифровом печатающем устройстве сопоставленных штрих-диаграмм исходного спектра образца и реконструированного по результатам анализа. Подобная графическая иллюстрация проведенной идентификации оказывается полезной, поскольку в системе ФАЗАН, в силу ее ориентации на универсальные ЭВМ, нельзя реализовать рассмотренную выше интерактивную методику анализа спектров, сопоставляемых на экране дисплея.

4.11. СИСТЕМА АРФА

Система автоматизированного рентгенофазового анализа — АРФА была разработана в 1977—1982 гг. в рамках комплекса методико-математического обеспечения рентгенографического анализа поликристаллов (ММО РАП). Как и некоторые из рассмотренных выше систем, комплекс ММО РАП обеспечивает выполнение целой совокупности задач, связанных с исследованием порошков, начиная от управления дифрактометрами типа ДРОН, АДП-1 в режиме реального времени, предварительной обработки дифракционных спектров и заканчивая полной обработкой данных (индицирование, определение параметров кристаллической решетки, качественный и количественный фазовый анализ, определение субструктуры соединения методом гармонического анализа формы рентгеновской линии). Система АРФА может использоваться как в составе комплекса ММО РАП, так и автономно. Алгоритмы системы и ее применение для анализа ряда объектов подробно рассмотрены в работах [18, 40, 41, 53, 55, 57].

Система АРФА предназначена для эксплуатации на универсальных ЕС ЭВМ (в режиме пакетной обработки) и на мини-ЭВМ М-6000, СМ-1 (в интерактивном режиме).

Система АРФА включает базу эталонных рентгенографических данных БЭРД, подсистемы служебных программ, информационного поиска и идентификации. Последняя включает две программы идентификации — POISF и АКФА.

База данных. К 1984 г. БЭРД содержала сведения примерно для 10 000 наиболее широко распространенных соединений. Подбор соединений в БЭРД осуществляется методом экспертной оценки со стороны ряда научно-исследовательских организаций геологического, металлургического, материаловедческого и химического профилей. Из полного состава БЭРД можно выделить и независимо использовать ряд подбаз, в том числе по минералам (до 2500 соединений), по наиболее распространенным (до 2000—3000), неорганическим, металлическим соединениям и т. д.

Машинная карточка БЭРД [54], как и карточка MINBAS, включает значительно больше информации о соединении, чем базы

данных прочих рассмотренных систем идентификации. В карточке БЭРД содержатся следующие данные: номер стандарта по картотеке JCPDS или другим источникам; название и химическая формула соединения; признак достоверности данных; сингония, пространственная группа и кристаллические параметры; метод получения спектра, используемый фильтр, длина волны, диаметр камеры (приводится при получении спектра фотометодом); признак, указывающий, введены ли в интенсивность поправки на аппаратурные искажения; спектр $\{d_j^s, I_j^s\}$ и значения индексов Миллера hkl . В карточке также содержится список входящих в состав соединения элементов и зарезервирована позиция для задания калибровочного коэффициента интенсивности по корунду (корундового числа). Положения линий спектра задаются в шкале d в плавающем формате, интенсивности — в стандартной столбчатой шкале в целом формате. Карточка БЭРД может содержать до 80 линий спектра.

Непосредственно БЭРД не используется для идентификации, а предназначается для решения трех основных задач: 1) обеспечения информационного поиска; 2) формирования файлов специального формата, используемых для идентификации программами POISF и АКФА; 3) построения выходных таблиц соответствия исследуемого спектра и отобранных стандартов.

Поисковый файл POISF является прямым и состоит, как и БЭРД, из отдельных карточек. Карточка файла POISF включает заголовок (содержательный номер стандарта, название, химическую формулу, признак достоверности) и спектр, представленный в следующих четырех последовательно усложняющихся формах.

Двоичная форма 1 — шкала 32 бита. Наиболее информативная область спектра $0,9-6 \text{ \AA}$ разбивается на 32 интервала равновероятного попадания линий; алгоритм разбиения подробно рассматривается в следующей главе. Наличие единицы в j -м разряде шкалы соответствует попаданию каких-либо линий в j -й интервал диапазона $0,9-6 \text{ \AA}$.

Двоичная форма 2 — более сложная шкала, которая состоит из 32 градаций, а градация — из 5 бит. Данная шкала также находится во взаимно однозначном соответствии с 32 интервалами диапазона $0,9-6 \text{ \AA}$. В двух битах j -й градации в двоичном виде фиксируется число линий, попавших в j -й интервал (00 — отсутствие линий, 01 — одна, 10 — две, 11 — три и более), а в трех оставшихся битах — суммарная интенсивность линий (также двоичным числом от 000 до 111). Интенсивности линий оцениваются в четырехбалльной шкале I^s ,⁴ в соответствии с табл. 4.2.

Форма 3 — положения и интенсивности восьми сильнейших линий стандарта. Положения задаются в шкале H :

$$H^s = \text{Бл. ц.} \left\{ \frac{2000}{d^s} \right\}. \quad (4.42)$$

Таблица 4.2

Оценка значений интенсивности в четырех- и шестнадцатибалльной шкалах в программе POISF

Четырехбалльная		Шестнадцатибалльная			
I^{100}	I^4	I^{100}	I	I^{100}	I
0—19	0	>1	0	11—13	8
20—49	1	1	1	14—18	9
50—74	2	2	2	19—24	10
75—100	3	3	3	25—32	11
		4	4	33—42	12
		5	5	43—56	13
		6—7	6	57—75	14
		8—10	7	76—100	15

Интенсивности задаются в шестнадцатибалльной логарифмической шкале:

$$I^{s,lg} = \text{Бл. ц. } \{8 \lg I^{s,100} - 1\}. \quad (4.43)$$

Соответствие между сто- и шестнадцатибалльными шкалами иллюстрируется табл. 4.2.

Форма 4 — положения и интенсивности сорока сильнейших линий спектра $\{H_j^s, I_j^{s,lg}\}$.

Значения $H^s, I^{s,lg}$ в формах 3 и 4 упакованы в два байта. В целом поисковый файл POISF для 10 000 стандартов имеет объем около 1,6 М байт.

Поисковый файл программы АКФА состоит из двух частей. Первая часть содержит пять записей, в которых хранятся положения пяти сильнейших линий стандартов в шкале H (формула (4.6)). В первой записи собраны положения сильнейших линий всех стандартов $\{H_j^{s,(1)}\}$, во второй — вторых по интенсивности линий $\{H_j^{s,(2)}\}$ и т. д. Линии в записях перечисляются в порядке следования карточек в БЭРД.

Вторая часть поискового файла АКФА подобна инвертированному файлу (4.8) системы Джонсона—Ванда и может содержать информацию о положениях 15—40 линий стандартов. Положения линий также задаются в шкале (4.6). Число линий, включаемых для каждого стандарта в инвертированный файл АКФА, определяется на этапе построения этого файла по полной БЭРД. Обычно информации о положениях первых 15 линий оказывается вполне достаточно для проведения предварительного этапа отбора.

Служебные программы. Подсистема служебных программ обеспечивает пополнение, корректировку и переформирование базы данных. Пополнение БЭРД может осуществляться двумя способами: 1) перфорацией данных на перфокарты или перфолену с последующим вводом в ЕС ЭВМ или мини-ЭВМ и записью на МЛ; 2) подготовкой данных с помощью видеотерминальных устройств

ЭВМ в режиме диалога с непосредственным накоплением данных на МЛ.

Подготавливаемые данные вначале накапливаются на МЛ в предварительном файле, который просматривается с помощью специальной программы контроля, предназначенного для автоматической проверки правильности подготавливаемых данных.

Возможно осуществление корректировки данных непосредственно в БЭРД. Для этого применяются две программы корректировки. Первая программа позволяет исправлять отдельные данные или группы данных в машинной карточке с помощью ввода корректирующей информации с перфокарт, вторая — полностью исключить указанную карточку, ввести новую, исключить карточку с заменой ее на новую.

Специальные программы обеспечивают переформирование базы данных, заключающееся в построении по БЭРД поисковых файлов программ идентификации POISF и АКФА, а также перевод МЛ с базой данных из формата ЕС ЭВМ в форматы мини-ЭВМ M-6000, CM-1 и обратно.

Информационный поиск. Подсистема информационного поиска включает три программы, реализующие первый, третий, четвертый типы поиска, рассмотренные в главе 1. Программа поиска по номерам карточек позволяет выбрать из БЭРД стандарты с заданными номерами. Возможно проведение поиска по порядковым номерам карточек в базе данных и по содержательным номерам, т. е. под которыми данные стандарты фигурируют в картотеке JCPDS и в различных определителях.

Программа поиска по химическим элементам XIMSP позволяет выбрать из БЭРД стандарты определенного состава. Возможно осуществление поиска в трех режимах. В случае первого режима задается список Z_1 присутствующих элементов и разыскиваются соединения, в состав которых входят все элементы Z_1 и любые другие. При втором режиме также задается список Z_1 , но разыскиваются соединения, в состав которых входят только элементы Z_1 . Наконец, в случае третьего режима кроме списка Z_1 задается список отсутствующих элементов Z_2 ; разыскиваются все соединения, в состав которых входят все элементы Z_1 и могут входить любые другие, кроме указанных в Z_2 .

Третья программа RECVIS реализует многопараметрический поиск — четвертый тип, рассмотренный в главе 1. Поиск может производиться по любой совокупности следующих реквизитов: химическому составу, признаку достоверности, сингонии, пространственной группе, значениям параметров кристаллической решетки, методу получения спектра, используемому фильтру, длине волны, наличию в карточке индексов Миллера. Например, возможна следующая формулировка заданий для программы RECVIS. Выбрать все соединения меди, удовлетворяющие условиям: данные высокой и средней достоверности; сингонии — кубическая, гексагональная и тетрагональная; пространственные группы — любые; значение параметра решетки a — в пределах

3—9 Å; метод съемки — дифрактометрический, с никелевым фильтром, на длине волны меди; спектры должны быть проиндицированы.

Результаты поиска, проводимого с помощью программ XIMSP и RECVIS, могут быть выведены на печать в стандартном формате и в виде каталога, а также записаны на МЛ. Методическое назначение этих программ различно. Программа XIMSP в основном применяется для формирования до полной БЭРД специализированных подбаз, включающих химически родственные соединения. Из данных подбаз строятся затем специализированные поисковые файлы для программ идентификации POISF и АКФА. Таким образом, программа XIMSP позволяет наиболее эффективно использовать сведения о химическом составе исследуемых соединений путем выделения из полной базы данных нужных и сравнительно небольших выборок (до 500—1000 стандартов), по которым затем осуществляется идентификация. Подобная методика позволяет значительно повысить как скорость, так и достоверность качественного анализа. Она особенно удобна в случае массового качественного анализа объектов, принадлежащих к нескольким определенным классам соединений (например, карбиды, сульфиды, нитриды металлов и т. д.).

Программа RECVIS может применяться для построения специализированных подбаз, с целью проведения с их помощью идентификации. Так, можно сформировать подбазу данных, содержащую только карточки высокой достоверности, или выбрать соединения указанного химического состава и определенной сингонии. В отличие от XIMSP программа RECVIS позволяет исследователю использовать при формировании специализированных подбаз не только химическую информацию, но и сведения о структуре соединения, а также определять требования к качеству стандартов. Однако основное методическое назначение RECVIS заключается в том, что с помощью этой программы можно изучать структуру базы данных путем анализа результатов поисков при различных наборах поисковых параметров. Этот аспект проблемы качественного анализа ранее практически не отмечался, а в рассмотренных выше системах идентификации отсутствуют программные средства, аналогичные RECVIS.

Программа RECVIS может также использоваться для поиска информации о соединениях, номера которых по картотеке JCPDS неизвестны, но имеется предположительная информация об их химическом составе и структуре. Сформированные с помощью XIMSP и RECVIS подбазы могут являться, в свою очередь, исходным материалом для повторного поиска. Таким образом, можно выделить подбазу с помощью XIMSP (обеспечивающей более сложный поиск по химической информации, чем RECVIS), а затем из нее с помощью RECVIS — меньшую выборку, удовлетворяющую нужным ограничениям в части структуры соединений, качества данных и методов съемки.

Алгоритм идентификации и выходные форматы POISF. Подсистема идентификации включает две программы, POISF и АКФА, которые реализуют значительно различающиеся методики качественного анализа.

Программа POISF разработана на базе ЕС ЭВМ и мини-ЭВМ М-6000, СМ-1, причем второй вариант считается основным. В программу в режиме диалога вводится исследуемый спектр в одной из следующих форм: $\{2\theta^\circ, I^\circ\}$, $\{d^\circ, I^\circ\}$, $\{H^\circ, I^\circ\}$. Кроме этого, вводятся также параметры поиска ΔH , $R_{\text{гпр}}$, $I_{\text{отс}}$. Имеется возможность задания ряда дополнительных условий: число и последовательность этапов идентификации, граничные значения критериев $R_{2\text{гр}}$ и $R_{3\text{гр}}$ на втором и третьем этапах, указания относительно местонахождения поискового файла (на НМЛ или НМД).

Исследуемый спектр переводится в рассмотренные выше формы 1—4. Карточки поискового файла последовательно считываются с НМД или с НМЛ и сравниваются с исследуемым спектром; сравнение осуществляется в четыре этапа. На первом спектры стандарта и образца сравниваются в двоичной форме 1, после чего вычисляется частотный критерий

$$R_1 = \frac{r_{\text{совп}}}{r_s} \quad (4.44)$$

где r_s — число единиц (заполненных разрядов) в двоичной шкале для стандарта, а $r_{\text{совп}}$ — число заполненных единицами разрядов, совпадающих в шкалах стандарта и образца. Если $R_1 \geq 0,5$, то на втором этапе производится сравнение спектров в двоичной форме 2. При этом вычисляется критерий типа (4.17), (4.18):

$$R_2 = \frac{n_{\text{совп}}}{n_s} \left(1 - \frac{\sum_{j=1}^{n_{\text{совп}}} (I_j^{s,4} - I_j^{o,4})}{\sum_{j=1}^{32} I_j^{s,4}} \right) \quad (4.45)$$

причем в случае $(I_j^{s,4} - I_j^{o,4}) < 0$ считается, что отрицательное значение связано с наложением в j -м разряде шкалы образца линий составляющих фаз и что данная разность заменяется нулем. При $R_2 \geq R_{\text{гр}}$ осуществляется переход к третьему этапу сравнения, на котором сопоставляются спектры стандарта и образца в форме 3. Сопоставление производится следующим образом.

1. Определяются совпадающие в смысле выполнения неравенства (4.11) линии образца и стандарта (для восьми сильнейших линий стандарта).

2. По совпадающим линиям вычисляется условная концентрация (формула (4.4)), в которую подставляются логарифмические интенсивности в виде (4.43), после чего осуществляется приведение спектров в единую шкалу отсчета по интенсивности.

3. Приведенные спектры преобразуются в модельное представление (см. рис. 1.1, *е*) путем замены каждой линии штрих-диаграмм треугольным распределением.

4. Модельный спектр стандарта совмещается с модельным спектром образца и определяется площадь перекрывания спектров $\Delta S^{(s)}$.

5. Определяется площадь спектра стандарта $S_s^{(s)}$ (для восьми сильнейших линий стандарта) и критерий соответствия.

$$R_3 = \frac{\Delta S^{(s)}}{S_s^{(s)}} \cdot 100\%. \quad (4.46)$$

При $R_3 \geq R_{3гр}$ осуществляется переход к четвертому этапу сравнения, на котором производится все перечисленные выше в пунктах 1—5 операции сопоставления спектров, но для 40 линий стандарта. При этом вычисляется нормированный критерий соответствия

$$R_4 = \frac{\Delta S}{S_s} \cdot 100\% \quad (4.47)$$

и весовой

$$R_4^w = \frac{\sum_{i=1}^{n_{совп}} \Delta S_i w_i}{S_s}, \quad (4.48)$$

где ΔS_i — площадь перекрывания i -й пары совпадающих линий образца и стандарта, моделированных треугольными распределениями. Для определения веса w_i была построена кривая $N(d)$ плотности распределения спектральных линий по d , которая будет подробно рассмотрена в следующей главе. Вес при максимальном значении кривой N_{max} принимается равным единице. Вес в функции от d вычисляется как

$$w(d) = \frac{N_{max}}{N(d)}, \quad (4.49)$$

т. е. полагается обратно пропорциональным ординате кривой $N(d)$.

Вычисление критериев R_4 R_4^w производится в диапазоне задания спектра образца и по линиям стандарта с интенсивностью $I_s^{100} \geq I_{отс}$. Для стандартов, удовлетворяющих условию $R_4 \geq R_{4гр}$, на печать и дисплей выводится строка первой таблицы результатов, содержащая следующую информацию:

Содержательный номер стандарта	Название	Химическая формула	R_4	R_4^w	$n_{совп.}$
					(4.50)

В некоторых вариантах программы POISF в данную строку добавляются значения R_2 , R_3 . Первая таблица выводится в процес-

се поиска. Строки (4.50) для стандартов с высокими значениями R_4 накапливаются на специальном буфере, где формируется список S_b вероятных фаз. После завершения процесса поиска строится идентифицирующая функция $R_4(m)$ (или просто $R(m)$) путем упорядочения стандартов по убыванию величины R_4 . На печать и дисплей выводится вторая таблица результатов, содержащая строки вида (4.50), но уже упорядоченные по R_4 . Вывод второй таблицы сопровождается построением и выводом графика идентифицирующей функции $R(m)$, где индекс m обозначает порядковые номера вероятных стандартов, расположенных по убыванию R_4 .

Как правило, идентифицирующая функция имеет характерный резкий излом (см. рис. 6.2), отделяющий область вероятных фаз (высокие значения) от области затянутого «хвоста» или «фона», которой соответствуют стандарты с небольшими значениями R . Поведение идентифицирующей функции можно анализировать визуально, используя в этот момент для отбраковки также и сведения о химическом составе образца (поскольку формулы стандартов приводятся в таблицах результатов). Для наиболее интересных стандартов исследователь может в режиме диалога затребовать построения таблиц соответствия (табл. 4.3). В пределах одной таблицы спектр образца может быть сопоставлен с 1—4 спектрами стандартов. Построение таблиц соответствия производится не по поисковому файлу POISF, а по полной базе данных БЭРД, поскольку содержащаяся в поисковом файле информация в виде $\{H^s, I^{s,lg}\}$ недостаточно полна. Обратный перевод H в d и I^{lg} в I^{100} дает значения, которые отличаются от истинных, указанных в первоисточнике. Использование этих значений в таблицах соответствия может приводить к недоразумениям и ошибкам. Таким образом, при корректном построении таблиц соответствия приходится затрачивать машинное время на просмотр исходной, наиболее полной базы, отбирая из нее указанные исследователем вероятные стандарты.

В программе POISF предусмотрен режим работы без задания граничного значения $R_{4гр}$. В этом случае после завершения процедуры поиска производится определение точки перегиба m_n кривой $R(m)$ и вывод таблицы результатов для первых $m_n + m'$ стандартов (где m' обычно порядка 2—5), соответствующих областям высоких значений R , излома кривой $R(m)$ и началу «фона».

Программа POISF позволяет варьировать число и последовательность этапов идентификации. Кроме основного полного четырехэтапного варианта (1, 2, 3, 4), можно обрабатывать одноэтапные (1), (2), (3), (4) (только первый, только второй, только третий, только четвертый), двухэтапные (1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4) и трехэтапные (1, 2, 3), (1, 2, 4), (1, 3, 4), (2, 3, 4). Программа POISF обеспечивает среднее быстродействие идентификации на мини-ЭВМ порядка 1 мин на тысячу анализируемых фаз (при обработке по варианту (1, 2, 3, 4) спектра 3—4-фазного образца, содержащего 30—35 линий).

Пример соответствия спектров образца и вероятных стандартов для смеси галенит-магнетит-селлаит (25 : 25 : 50)

№ п/п	Обр. 1004		Номера по картотеке ICPDS и названия стандартов								
			17-465, магне- зитоферрит		19-629, магнетит		5-592, галенит		6-290, селлаит		
	d	I	d	I	d	I	d	I	d	I	
1	3,4400	39						3,4287	94		
2	3,2800	66								3,2695	100
3	2,9710	100	2,9690	35	2,9670	100	2,9687	100			
4	2,5370	19	2,5322	100	2,5322	30				2,5450	20
5	2,4240	2	2,4240	4	2,4240	8					
6	2,2340	19								2,2314	95
7	2,0990	27	2,0991	25	2,0990	20	2,0991	57			
8	2,0700	17								2,0669	35
9	1,7920	16					1,7900	35			
10	1,7140	21	1,7130	12	1,7150	10	1,7140	16		1,7109	75
11	1,6350	18								1,6350	30
12	1,6170	5	1,6160	30	1,6160	30					
13	1,5270	4								1,5260	20
14	1,4850	15	1,4851	40	1,4851	10	1,4840	10			
15	1,4630	2								1,4620	6
16	1,4430	2								1,4410	4
17	1,3830	3								1,3820	14
18	1,3760	7								1,3750	35
19	1,3720	4									
20	1,3620	2					1,3620	10			
Линии стандарта, не совпавшие с образцом											
			4,8500	8	4,8500	8	1,3270	17	1,3180	8	
			1,2810	8	1,3280	4	1,2120	10	1,2280	6	
			1,2120	4	1,2810	10	1,1424	6	1,1556	4	
			1,1220	4	1,2660	4	1,0034	5	1,1217	6	
			1,0930	10	1,1220	4	0,9893	6	1,1154	10	
			1,0490	4	1,0930	12	0,9386	4	1,0893	6	
			0,9700	6	1,0500	6	0,8952	4	1,0555	4	
			0,9390	6	0,9695	6			1,0524	6	
			0,8800	6	0,9632	4			0,9790	4	
			0,8570	10	0,9388	4					
			0,8120	6	0,8802	6					
					0,8569	8					
					0,8233	4					

Алгоритм идентификации и выходные форматы АКФА. Программа АКФА разработана на базе ЕС ЭВМ. В программу вводится с перфокарт исследуемый спектр, а также параметры поиска (величина окна, химическая информация, размер выводимых таблиц результатов и др.) и некоторые дополнительные условия, определяющие режимы работы программы и формы вывода результатов.

Идентификация осуществляется в два этапа. Первый, предварительный, этап идентификации осуществляется с помощью поискового файла АКФА и состоит из следующих стадий:

1. С помощью первой части поискового файла определяются номера стандартов, у которых $l_{гр}$ сильнейших линий совпадают с линиями исследуемого спектра в смысле выполнения неравенства (4.11). Величина $l_{гр}$ равна минимальному из двух чисел:

$$l_{гр} = \min \left\{ 5; \frac{n_o}{10} + 1 \right\}. \quad (4.51)$$

Поскольку число линий в спектре образца обычно $n_o \sim 30-40$ и более, то $l_{гр}$ равна четырем или пяти. Это условие отбора можно ослабить, задав в исходных данных допустимое значение $l_{гр} < 5$.

2. С помощью второй части поискового файла стандарты, прошедшие первую стадию, сравниваются с исследуемым спектром и для них по 15 сильнейшим линиям вычисляется критерий частотного типа:

$$K = \frac{n_{совп}}{n_o} \cdot \frac{n_{совп}}{n_s}. \quad (4.52)$$

Номера стандартов с наибольшими величинами критериев запоминаются; количество этих вероятных стандартов $N_b = 50$, причем значение N_b можно изменять.

3. Для N_1 вероятных стандартов по полной БЭРД определяются величины $n'_{совп}$. Число совпавших линий $n'_{совп}$ подсчитывается путем сопоставления линий стандарта и исследуемого спектра по (4.11) в порядке убывания интенсивности линий стандарта. Подсчет $n'_{совп}$ ведется только до обнаружения первого несоответствия линий; таким образом, $n'_{совп}$ — число сильнейших линий стандарта, совпавших со спектром образца, до первого пропуска в спектре образца очередной по интенсивности линии стандарта. Номера вероятных стандартов упорядочиваются по убыванию $n'_{совп}$; таким способом формируется список вероятных фаз S_b , который служит исходной информацией для второго этапа идентификации.

На втором, заключительном, этапе из списка S_b отбираются фазы, суперпозиция которых наилучшим образом приближает спектр образца $I^o(d)$. Для отыскания наилучшего приближения для $I^o(\tilde{d})$ (т. е. реконструкции спектра с минимальной вероятностью ошибки) проблема идентификации на заключительном этапе формируется как оптимальная задача линейного программирования. Требуется определить минимум функционала:

$$J = \min_{по C_j} \sum_{j=1}^{N_b} \xi [C_j], \quad (4.53)$$

где

$$\xi(C_j) = \begin{cases} 1, & \text{если } C_j > 0 \\ 0, & \text{если } C_j \leq 0 \end{cases}$$

при следующих ограничениях:

$$C_j \geq 0, \\ \varepsilon_1 \leq I^o(\tilde{d}) - \sum_{i=1}^{N_b} C_j I_j^{s, \text{пер}}(\tilde{d}) \leq \varepsilon_2. \quad (4.54)$$

Выражения (4.53) и (4.54) записаны в обозначениях, введенных в разделе 4.12 (см. формулу (4.33)). Значения ε_1 , ε_2 предельных допустимых отклонений между исследуемым и реконструированным спектрами представляют собой параметры метода.

Минимизация выражения (4.53) при условиях (4.54) — сложная комбинаторная задача, упрощение которой возможно произвести путем линейризации функционала:

$$J = \min_{\text{по } C_j} \sum_{j=1}^{N_b} \frac{C_j}{n'_{\text{совп}, j}}. \quad (4.55)$$

Линейризованный функционал (4.55) строится таким образом, чтобы в определенной мере сохранить стремление к уменьшению числа стандартов, интерпретирующих спектр образца, и отдать предпочтение в процессе минимизации стандартам с большим числом совпавших линий $n'_{\text{совп}}$. Решение задачи (4.55) при условиях (4.54) осуществляется симплексным методом с помощью соответствующей стандартной программы, входящей в пакет прикладных программ для ЕС ЭВМ. Результаты идентификации с помощью программы АКФА выводятся на печать в виде следующих таблиц: 1) таблица результатов, содержащая построчный список номеров отобранных эталонов и значение K или C ; 2) соответствия; 3) сопоставления исследуемого и реконструированного спектров (возможно также сопоставление спектров в графической форме).

Перечисленные таблицы могут быть выведены по указанию исследователя как после первого, так и после второго этапов идентификации.

Программа АКФА обеспечивает среднее быстродействие идентификации на ЭВМ класса ЕС-1033, М-4030 порядка 0,5 мин на тысячу анализируемых фаз.

4.12. ДРУГИЕ СИСТЕМЫ АВТОМАТИЧЕСКОЙ ИДЕНТИФИКАЦИИ

Рассмотренные в разделах 4.2—4.11 системы в основном исчерпывают разнообразие методов и приемов, разработанных для автоматической идентификации объектов неизвестного фазового состава за последние два десятилетия. Перечислим еще ряд работ, посвященных рассматриваемой проблеме.

Одна из первых отечественных систем автоматической идентификации описана в [85]. База данных системы включала около 900 стандартов, собранных в определителе Михеева [49]. Стан-

дарт рассматривался как вероятный, если пять его сильнейших линий присутствовали в исследуемом спектре. Система была реализована для ЭВМ БЭСМ-3М на языке алгол.

Своеобразная система, основанная на синтезе методик Джонсона — Ванда и Фиалы, была рассмотрена в [82]. База данных системы включала 2000 стандартов. Отбор на предварительном этапе идентификации производился с помощью критерия вида (4.18). На втором этапе идентификации для вероятных стандартов методом наименьших квадратов решалась система (4.33), причем линии штрих-диаграмм стандартов и образца заменялись функциями Коши. Система [88] была разработана на алголе на базе ЭВМ БЭСМ-4М.

В работах [60, 61] описана система *, сходная с ФАЗАН; ее особенностью является попытка учета в процессе идентификации статистической зависимости отдельных измерений. Для этого предложена модель спектра, в которой учитывается случайная и систематическая составляющие ошибки измерения. Данная система разработана на языке фортран для ЭВМ БЭСМ-6, кроме программы идентификации, она включает ряд служебных и сервисных программ, но ее широкое использование затрудняется небольшими размерами базы данных (несколько сотен стандартов).

Оригинальная система идентификации, разработанная на базе ЭВМ IBM-360/44, описана в [162]. Основу базы данных составляют совокупность стандартов для минералов. Подготовка стандартных данных осуществляется на перфоленде с помощью ЭВМ PDP-15/10. Перфоленда вводится в ЭВМ IBM-360/44 и стандарты накапливаются на МЛ; при этом имеется возможность корректировки формируемой базы (удаление устаревших данных и замена их новыми). Идентификация основана на сопоставлении сильнейших линий стандартов со спектром исследуемого образца и предусматривает последовательное вычитание найденных фаз. Система позволяет использовать при идентификации дополнительную информацию, задаваемую в виде номеров карточек N_{JCPDS} , названий предполагаемых минералов, их возможного химического состава.

Имеются сообщения о разработке и практическом применении систем автоматической идентификации в Болгарии [59] и КНР [129]. Последняя система, разработанная китайскими специалистами, описана в [129] достаточно подробно. База данных системы формируется по картотеке JCPDS. Для идентификации применяется алгоритм Джонсона — Ванда с использованием рассмотренных в разделе 4.3 прямого и инвертированного файлов. Система описана на языке фортран и ориентирована на мини-ЭВМ с объемом оперативной памяти не менее 32 К двухбайтовых слов.

* Независимо разработанные системы [54—56 и 60, 61] случайно получили одинаковые наименования АРФА. В дальнейшем изложении название АРФА сохраняется за более крупной системой [54—56], поскольку эта система внедрялась под данным названием начиная с 1979 г. во многих научно-исследовательских организациях.

Одна из первых систем автоматической идентификации на основе мини-ЭВМ ICL-190 3A была разработана в 1976 г. О'Коннором и Баглиани [136] — сотрудниками физического факультета Западноавстралийского технологического института. Ван-Трампом и Хауффом разработана оригинальная система [158] на базе PDP-10, ориентированная на исследование минералов. В работе [106] рассмотрена одна из первых систем идентификации, реализованных на микрокомпьютере с накопителями на гибких МЛ.

Проведенный в данной главе обзор показывает, что использование мини- и микроЭВМ для создания интерактивных лабораторных систем идентификации является основным направлением их разработки. Использование малых компьютеров в значительной степени усугубляет сложность проблемы, поскольку особую остроту приобретают такие вопросы, как эффективное сжатие базы данных, выделение наиболее важных при идентификации линий стандартов, быстроедействие процедур идентификации. В следующей главе мы переходим к подробному анализу основных методических моментов автоматической идентификации и к их теоретическому обобщению с учетом возможности эффективного применения рассматриваемых алгоритмов на мини- и микроЭВМ.

Глава 5

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ИДЕНТИФИКАЦИИ

Обзор ряда систем автоматического качественного РФА, проведенный в главе 4, дает обширный материал для сопоставительного анализа различных алгоритмов идентификации. Задачей подобного анализа является выделение наиболее рациональных и эффективных методических элементов процедуры идентификации, реализованных в рассмотренных выше системах. Ниже будет дана также единая математическая и системная трактовка проблемы идентификации, основанная на использовании аппарата векторной алгебры и методов распознавания образов.

5.1. СРАВНИТЕЛЬНЫЙ АНАЛИЗ РАЗЛИЧНЫХ СТРАТЕГИЙ ПОИСКА И ВЫБОРА

Изучение систем автоматической идентификации показывает, что реализованные в них процедуры поиска и выбора содержат достаточно много сходных и совпадающих методических элементов. Некоторые из этих методических элементов могут быть

отнесены к числу основных, другие являются менее важными и носят дополнительный или вспомогательный характер. Помимо этого, в отдельных системах реализованы сравнительно новые и оригинальные методические приемы, которые еще не стали общепринятыми, но представляются очень перспективными. Перечислим методические элементы автоматической идентификации, разбив их совокупность на три группы.

Первая группа — основные методические элементы идентификации:

1. Определенный выбор окна и использование неравенств типа (4.3), (4.11) для определения совпадающих линий спектров образца и стандарта.

2. Построение идентификации в виде двух- или многоэтапной процедуры с целью сокращения времени поиска.

3. Построение баз данных в виде двух или нескольких файлов специального формата и структуры, облегчающих процессы поиска стандартов и сравнения спектров стандартов со спектрами образца.

4. Использование на предварительном этапе идентификации инвертированного или прямого файла сжатого формата.

5. Включение в базу данных не только номера и спектра стандарта, но и соответствующей химической формулы или списка входящих в него химических элементов.

6. Фиксация положений наиболее сильных линий как наиболее информативных признаков спектра монофазы, вследствие чего а) в поисковые файлы многих систем включаются данные только для 3, 5, 8 и т. д. сильнейших линий;

б) необходимым условием предварительного отбора стандарта считается совпадение по положениям нескольких сильнейших линий его спектра (от 2—3 до 4—5) с соответствующими линиями спектра образца.

7. Использование на предварительном этапе идентификации только сведений о положении линий (обычно сильнейших), а на заключительном этапе — информации о положении и интенсивности (всех или большинства линий спектра).

8. Вычисление условной концентрации и приведение спектров образца и стандарта в единую шкалу отсчета по интенсивности с целью расчета идентифицирующей функции (FOM).

9. Использование частотных критериев типа (4.12), (4.27).

10. Построение на заключительном этапе идентифицирующей функции (FOM) в виде упорядоченной по убыванию последовательности значений комбинированного критерия типа (4.18), (4.21), (4.24), (4.26), (4.41), (4.47).

11. Задание граничных значений критериев соответствия для предварительного и окончательного отбора вероятных стандартов и вывода результатов идентификации на печать и дисплей.

12. Использование химической информации.

13. Создание сервисных средств, позволяющих исследователю быстро обозреть и оценить результаты автоматической идентификации с целью окончательного отбора искомого фаз.

14. Осуществление идентификации изучаемого спектра путем нескольких прогонов программы идентификации с варьированием параметров поиска (граничных значений критериев соответствия, химической информации, величины окна и т. д.).

15. Реализация идентификации на ЭВМ в виде интерактивной процедуры.

Вторая группа — дополнительные методические элементы идентификации:

1. Использование шкалы H или же представление d в целом виде; использование логарифмической или другой сглаживающей шкалы интенсивности.

2. Сравнение спектров стандарта и образца в границах спектра образца или же в заданных исследователем границах.

3. Использование параметра отсечки интенсивности $I_{отс}$; линии спектра стандарта с $I < I_{отс}$ не включаются в процедуру сопоставления.

4. Вычитание из спектра образца спектра найденного стандарта.

5. Разделение наложенных линий в спектре образца с целью более точного определения условной концентрации.

6. Использование специализированных подбаз небольшого объема (300—3000 стандартов) для повышения достоверности и увеличения экспрессности идентификации.

Третья группа — новые методические элементы идентификации:

1. Выделение информативных признаков спектра монофазы с помощью функции плотности распределения спектральных линий $N(d)$, вследствие чего

а) отбор в поисковые файлы ограниченного числа линий производится с учетом не только их интенсивностей, но и информационного веса в зависимости от диапазона d ;

б) вычисление критериев соответствия также производится с учетом веса совпадения линий.

2. Определение интервалов равновероятного попадания спектральных линий (интервалов Ханавальта).

3. Автоматическое ограничение списка вероятных фаз путем анализа поведения идентифицирующей функции.

4. Построение иерархической базы данных.

5. Реконструкция исследуемого спектра с помощью спектров вероятных стандартов с использованием соотношения (4.33).

6. Создание информационно-поисковых систем, позволяющих производить не только идентификацию, но и информационный поиск.

Единственная попытка сравнительного анализа стратегий поиска и выбора была сделана Никольсом в конце 70-х годов [135]. Никольс полагал, что существуют четыре различающиеся стратегии, представленные системами Фривела, Никольса, Джонсона — Ванда и PDSM CIS. Различия между этими стратегиями, по мнению Никольса, заключались в следующем:

1) способе выбора окна (элемент 1 первой группы): малое окно, большое окно, малое окно на начальном этапе с переходом к большому окну, если искомые фазы не обнаружались;

2) методе поиска по базе данных (элементы 3 и 4 первой группы);

3) использовании химической информации (элемент 12 первой группы): используется обязательно, не используется (в системе Никольса), может использоваться по желанию исследователя.

Приведенный выше список методических элементов идентификации показывает, что в настоящее время анализ Никольса является далеко не полным. Способ выбора окна не является уже моментом существенного различия, поскольку в современных интерактивных системах размер его варьируется — либо непосредственно исследователем (система АРФА), либо программно (система Снайдера). Возможность изменения исследователем размеров окна является обязательной, поскольку во многих ситуациях только исследователь методом проб и ошибок может подобрать подходящее окно. Рассмотрим этот вопрос несколько шире.

Предположим, необходимо определить состав некоторой группы образцов с помощью системы автоматической идентификации. Исследователь может попытаться провести идентификацию при рекомендованных для данной системы значениях параметров поиска — величине окна, граничном значении числа совпадающих сильных линий $I_{гр}$, уровне отсечки интенсивности $I_{отс}$, граничных значений критериев соответствия, химических и других ограничений. При этом исследователь может убедиться, что рекомендованные авторами системы или другими исследователями значения параметров поиска приводят к неудовлетворительным результатам для изучаемой группы образцов. Тогда необходимо осуществить ряд попыток анализа с вариацией параметров поиска, т. е. определить такую совокупность численных значений параметров, которая позволила бы получить удовлетворительный результат для данной конкретной группы образцов. Фактически это означает параметризацию системы идентификации, создание методики ее использования для конкретной группы объектов.

Возникает естественный вопрос — можно ли избавить исследователя от этой работы и осуществить параметризацию автоматически? Такая попытка делается в системе Снайдера. Действительно, в этой системе, в случае отсутствия успеха при первой попытке идентификации, осуществляются по определенному жесткому алгоритму перебор сильнейших линий исследуемого спектра, расширение окна и сканирование более полных, чем исходные, подбаз. Как указывалось в разделе 4.8, даже при автоматическом варьировании только трех параметров поиска (сильнейшие линии, окно, тип подбазы) алгоритм системы Снайдера оказывается очень сложным. Если же добавить в него автоматическую вариацию $I_{гр}$, $I_{отс}$ и $F_{гр}$, то сложность алгоритма значительно возрастет. Кроме этого, остается открытым вопрос, как варьировать параметры поиска. Можно равномерно перебирать с постоянным шагом

численные значения параметров, например, менять $I_{\text{гр}}$ от $I_{\text{гр}}^{\min}$ до $I_{\text{гр}}^{\max}$ с шагом $\Delta I_{\text{гр}}$, $I_{\text{отс}}$ — от $I_{\text{отс}}^{\min}$ до $I_{\text{отс}}^{\max}$ с шагом $\Delta I_{\text{отс}}$ и т. д. Это означает, что в многомерном фазовом пространстве параметров создается сетка, узлам которой соответствуют определенные наборы численных значений параметров. Для параметризации системы идентификации в автоматическом режиме необходимо обойти все эти узлы сетки, что достаточно длительно; при этом можно пропустить нужный узел, если числа по каким-либо параметрам выбраны неудачно. Можно перебирать совокупности параметров неравномерно, по какому-нибудь сложному и пока пеясному алгоритму, либо сочетать оба способа перебора. В любом случае данная процедура является попыткой подмены исследователя машиной при решении сложной методической задачи, стратегия решения которой полностью не алгоритмизируется. Известно [2, 52], что пока наилучшим выходом является создание человекомашиных интерактивных систем. Таким образом, исследователь должен иметь возможность распоряжаться параметрами поиска, что и реализовано в большинстве современных систем. Данное положение не исключает наличия в системе возможностей для частичной автоматической параметризации — например, изменения величины окна.

Третье положение работы Никольса [135] относительно использования химической информации также не является актуальным. Практически во всех современных системах автоматического качественного РФА химический состав стандартов введен в базы данных, и, следовательно, данная информация может быть использована по желанию исследователя в процессе идентификации.

Перейдем к анализу различий между существующими системами идентификации, которые приводят к различным стратегиям поиска и выбора. Первым таким различием является полнота информации в базе данных, т. е. объем сведения для одного соединения, содержащихся в машинной картотеке. Наиболее полный объем сведений содержится в базах данных MINBAS и БЭРД. Благодаря этому возможно проведение поиска не только по спектру, но и по другим характеристикам соединений.

Второе различие — полнота базы данных, т. е. число представленных в базе стандартов. Наиболее полными являются базы данных, оперирующих с полной картотекой JCPDS (системы Джонсона — Ванда, PDSM CIS, Снайдера, фирм Филипс и IBM). Данное различие приводит к необходимости создания сверхэкспрессных алгоритмов для просмотра большой базы данных, особенно если соответствующая система реализована на мини-ЭВМ.

Третье различие связано с физико-химической структурой баз данных, под которой понимается компоновка баз данных в виде нескольких подбаз, формируемых на основе физико-химических соображений. Примеры такого рода структур демонстрируют, например, системы IBM (подбазы IN, OR, MI, ME, разделенные на секции), Снайдера (подбазы MICRO, MINI, MAXI) и АРФА (возможность построения разнообразных структур с помощью под-

системы программ информационного поиска). Рассматриваемое построение баз данных приводит к стратегии последовательного сканирования ряда подбаз возрастающей полноты.

Четвертое различие — **формат и файловая структура баз данных**, непосредственным образом влияющие на стратегию поиска. Здесь основным моментом является принцип построения файла предварительного отбора, поскольку с помощью этого файла решается задача экспрессного просмотра базы данных большого объема с целью выбора вероятных стандартов для подробного исследования на заключительном этапе. Предложено три способа построения файла предварительного отбора: 1) инвертированный, используемый во многих системах (Джонсона — Ванда, Снайдера, фактически и в системе ZRD-1965); данный принцип построения поискового файла восходит к методике поиска Ханавальта; 2) прямой, в котором в упакованном виде содержится информация только о положениях или о положениях и интенсивностях наиболее информативных линий; подобная структура используется, например, в системах PDSM CIS и ФАЗАН (файл-указатель); 3) прямой с многосекционными карточками (система АРФА); предназначенные для предварительного отбора секции (формы 1 и 2 поискового файла POISF) строятся путем сжатия исходного спектра с выделением его наиболее информативных признаков. Отметим, что на заключительном этапе отбора всегда используется прямой файл.

Пятое различие непосредственно вытекает из четвертого и касается стратегии поиска и сравнения; оно также отмечалось Никольсом [135]. Инвертированный файл позволяет реализовать прямой метод доступа к необходимой информации без просмотра всего файла предварительного отбора. Таким образом, возможен быстрый просмотр большой базы данных за счет выборки только нужной информации и пропуска ненужной. Однако объем инвертированного файла ненамного меньше объема прямого файла заключительного поиска (например, в системе Джонсона—Ванда объемы файлов PS и PSI почти равны). Таким образом, наличие инвертированного файла практически удваивает объем используемой для идентификации информации.

Прямой файл с упаковкой данных необходимо просматривать последовательно. Эта процедура длительная и ее целесообразно реализовывать либо для баз данных малого объема, либо на мощных ЭВМ с малым временем доступа к НМД, НМЛ. Многосекционный прямой файл также необходимо просматривать последовательно, но в этом случае выигрыш во времени достигается за счет сверхэкспрессной процедуры сравнения стандарта с исследуемым спектром. Как правило, во всех поисковых стратегиях с инвертированными и прямыми файлами предварительного отбора, на предварительном этапе проверяется неравенство типа (4.11), после чего рассчитывается частотный критерий соответствия (4.12). Эти операции представляют собой содержание процедуры сравнения, которая в данном случае длительная. Многосекционный файл

позволяет производить предварительное сравнение спектров как сравнение двух двоичных кодов (двух машинных слов), что значительно эффективнее.

Имеется еще одна особенность работы с многосекционным файлом. В обычных стратегиях поиска четко разделяется предварительный этап с поиском по инвертированному или прямому файлу предварительного отбора и формализованием списка вероятных стандартов и заключительный, на котором производится поиск вероятных стандартов по прямому файлу заключительного отбора, их анализ, уточнение списка вероятных стандартов и вывод его на печать или дисплей. Следовательно, в процессе поиска всегда обрабатываются два поисковых файла. Многосекционный прямой файл является единым поисковым, содержащим информацию как для предварительного, так и для заключительного этапов отбора, что позволяет реализовать совершенно иную стратегию поиска. Данная стратегия, как это следует из описания алгоритма POISF в разделе 4.11, основана на полном анализе каждой машинной карточки стандарта на всех этапах идентификации (если стандарт удовлетворяет нужным критериям соответствия).

Шестое и последнее различие связано со стратегией выбора и сводится к конструированию идентифицирующих функций в той или иной конкретной форме. В разделе 5.6 будет показано, что обычно используемые идентифицирующие функции являются некоторыми приближениями универсальной меры близости спектров.

Важное различие стратегии выбора — способ ограничения списка вероятных фаз и его последующая обработка (уточнение). Предложено два способа ограничения: 1) традиционный пороговый, при котором стандарт считается вероятным и выводится на печать и дисплей (или передается на следующий этап анализа), если значение идентифицирующей функции для него равно или превосходит заданное граничное значение $F_{гр}$. Этот способ реализован во всех без исключения системах идентификации; 2) аналитический, при котором ограничение списка вероятных фаз производится путем анализа поведения идентифицирующей функции с целью определения области ее существенного изменения (POISF, АРФА).

Последующая обработка списка вероятных фаз сводится к следующим вариантам: 1) анализ поведения идентифицирующей функции, вывод на печать таблиц соответствия и их изучение. Этот способ в той или иной степени реализован во многих системах идентификации — Фривела, Джонсона — Ванда, PDSM CIS, ФАЗАН, АРФА и др.; 2) уточнение результатов идентификации в режиме диалога с помощью графического дисплея с использованием графического сопоставления спектров и вычитания указанных исследователем стандартов. Данный способ используется в системах фирм Филипс (программа CASSAM), Сименс (IDENT), IBM; 3) реконструкция исследуемого спектра с помощью спектров вероятных стандартов с использованием методики Фиала и ее дальнейшего развития в системах ФАЗАН и АРФА (программа АРФА).

Любой из перечисленных вариантов уточнения списка вероятных фаз может привести к успеху либо не дать удовлетворительных результатов и даже запутать неопытного исследователя. При оценке этих вариантов необходимо отметить, что они являются не конкурирующими, а, скорее, взаимно дополняющими. Однако пока не существует систем идентификации, в которых были бы реализованы все возможные варианты уточнения списка вероятных фаз.

Проведенный выше анализ показывает, что в рамках проблемы автоматической идентификации решаются такие взаимосвязанные задачи, как повышение экспрессности и достоверности, разработка методики окончательного уточнения состава образца. Основной среди трех перечисленных задач является первая — повышение экспрессности идентификации при удовлетворительном уровне достоверности. Ее центральное положение определяется тем, что главный вопрос качественного РФА — быстрый просмотр большого объема информации, который постоянно увеличивается. Кроме этого, вторая и третья задачи (которые имеют, конечно, свои специфические особенности) прямо зависят от экспрессности идентификации. Важный фактор повышения достоверности — хорошая параметризация процедуры поиска и выбора, которая требует, как указывалось выше, многократных просчетов по программе с варьированием поисковых параметров. Таким образом, при прочих равных условиях (при полноте информации и базы данных, качества критериев соответствия, мощности ЭВМ) более высокую достоверность обеспечит та система идентификации, которая позволяет осуществить максимальное число попыток анализа за заданное время. От достоверности идентификации, т. е. от качества подбора вероятных стандартов, непосредственно зависят как результативность, так и время окончательного уточнения состава образца.

Между первой и второй задачами имеются также и обратные зависимости, поскольку более высокая экспрессность может достигаться за счет информационного обеднения базы данных и использования упрощенных критериев соответствия, что ведет к снижению достоверности. Повышение достоверности требует информационной полноты базы данных и расчета сложных критериев соответствия, что ведет, в свою очередь, к снижению экспрессности. Поскольку при низкой достоверности автоматическая идентификация лишается смысла, то задача повышения экспрессности должна решаться при сохранении удовлетворительного уровня достоверности.

Важнейшими факторами повышения экспрессности являются технический — повышение мощности ЭВМ и теоретический — создание высокоэффективных стратегий поиска и выбора. Первое обстоятельство очевидно, поэтому в дальнейшем основное внимание будет уделяться второму. Важнейшим фактором повышения достоверности является улучшение качества стандартов и качества съемки, связанное с получением более точных значений d , I

и созданием новых методик исследования. Таким образом, дальнейшее повышение достоверности анализа может быть достигнуто на пути совершенствования экспериментальной базы и методики эксперимента. Конструирование новых, все более изощренных критериев соответствия является значительно менее эффективным способом повышения достоверности; по-видимому, в этой области уже достигнуто предельное насыщение.

Еще одна особенность применяемых стратегий поиска и выбора. Некоторые приемы автоматической идентификации фактически копируют методику ручного качественного РФА. Использование многих методических элементов ручного анализа необходимо и оправдано, однако вряд ли целесообразно перенести в стратегию автоматической идентификации один из основных ручных приемов — выбор некоторого числа (3, 5, 8) сильнейших линий в качестве наиболее информативных признаков стандартного спектра. Обычно их выбор производится эмпирически. Никольс [135] указывает, что во многих системах автоматического рентгенофазового и масс-спектрометрического анализа на этапе предварительного отбора используется «магическое» число 8 — восемь сильнейших линий стандарта. Возможно, что за проверенным на практике «магическим» характером данного числа стоит некоторая объективная закономерность. Определение ее или указание путей к ее определению — интересная задача, поскольку полученные результаты могут быть использованы как в качественном РФА, так и при идентификации спектральных данных (масс-спектроскопия, инфракрасная и Оже-спектроскопия).

Закончив краткий сравнительный анализ особенностей различных процедур идентификации, перейдем к более детальному исследованию основных этапов автоматического качественного РФА.

5.2. МАТРИЧНОЕ ПРЕДСТАВЛЕНИЕ БАЗ ДАННЫХ

Введем евклидово векторное пространство R_n размерности n . Будем для определенности считать, что пространство R_n получено путем равномерного разбиения диапазона анализа D от нуля до 20 \AA на промежутки по $\delta d = 0,001 \text{ \AA}$ (см. обозн. в разделе 4.9). Тогда размерность R_n предстанет величиной порядка $n = \frac{D}{\delta d} = 2 \cdot 10^4$. Межплоскостное расстояние d обозначим через x ; аргумент x принимает дискретные значения, а в пределе, при $\delta d \rightarrow 0$, становится непрерывным.

Пусть имеется совокупность M спектров (база данных), заданных в диапазоне D . Любой из этих спектров можно представить в виде вектора \vec{y}_j размерности n в пространстве R_n . Индекс j нумерует спектры-вектора и под ними можно понимать как порядковый номер ($j = 1, 2, \dots, M$), так и содержательный (например, N_{JCPDS} по картотеке JCPDS). В зависимости от степени подроб-

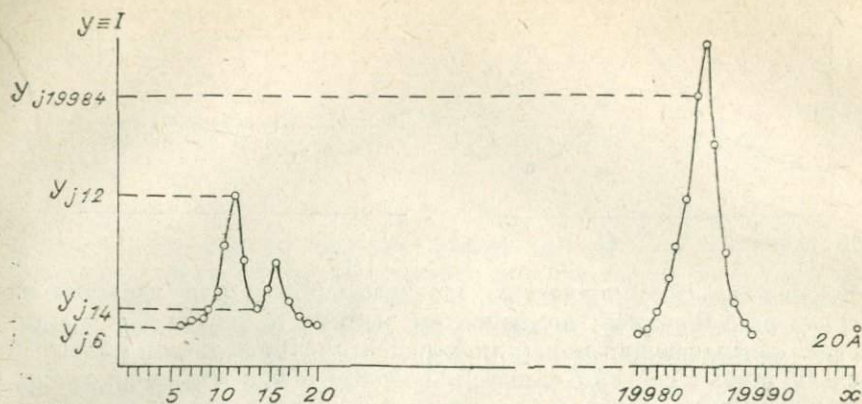


Рис. 5.1. Форма вектора \vec{y}_i при численном задании спектральной функции с шагом $\delta x = 0,001$. Для примера на оси y показаны значения некоторых компонент вектора.

ности задания спектров вектора \vec{y}_j могут принимать конкретные формы. Если спектры заданы в диапазоне D в виде непрерывных кривых (наиболее подробное представление), то аргумент x становится непрерывным, $n \rightarrow \infty$ и R_n переходит в бесконечномерное гильбертово пространство. Тогда \vec{y}_j преобразуется в непрерывные спектральные функции $\vec{y}_j = y_j(x)$, где y обозначает величину интенсивности в точке x . Если спектр задан в виде совокупности численных отсчетов с шагом $\delta d = \delta x = 0,001 \text{ \AA}$ (т. е. спектральная функция $y_j(x)$ задана численно), то i -тая компонента соответствующего вектора \vec{y}_{ji} равна i -тому значению интенсивности (рис. 5.1):

$$\vec{y}_j = (y_{j1}, y_{j2}, \dots, y_{j5}, y_{j6}, \dots, y_{j19990}, y_{j19991}, \dots, y_{j20000}). \quad (5.1)$$

Если некоторые компоненты вектора \vec{y}_j равны нулю (см. рис. 5.1), то можно записать в более простом виде:

$$\vec{y}_j = (0, 0, 0, 0, 0, y_{j6}, y_{j7}, \dots, y_{j19989}, y_{j19990}, 0, \dots, 0, 0, 0, 0). \quad (5.2)$$

Случай задания спектра числом $x > 0,001$ сводится к предыдущему, так как недостающие промежуточные компоненты можно восстановить интерполяцией, как это показано на рис. 5.2 для одной линии спектра, заданной с шагом 0,005.

Если спектр задан в виде штрих-диаграммы, то соответствующий вектор имеет небольшое количество положительных компонент, разделенных нулями (см. рис. 5.1):

$$\vec{y}_j = (0, \dots, 0, y_{j12}^1, 0, \dots, 0, y_{j16}^2, 0, \dots, y_{j19985}^{n_s}, 0, \dots, 0). \quad (5.3)$$

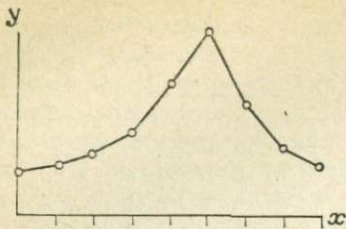


Рис. 5.2. Представления спектра в случае $\delta x = 0,005$.

Выражение (5.3) полностью адекватно (4.30); для удобства их сравнения значимым компонентам вектора приписан вверху индекс, обозначающий номер линии спектра. Поскольку в реальных базах данных спектры стандартов заданы в виде штрих-диаграмм, в дальнейшем будет в основном рассматриваться форма (5.3).

Наконец, если совершен переход от штрих-диаграммы к реконструированному спектру путем замены каждой линии штрих-диаграммы прямоугольным или треугольным распределением, функциями Гаусса, Коши или любой другой функцией размытия $f(x)$ (см. (4.31)), то и такая форма спектра может быть легко представлена в векторном виде. Так, для изображенного на рисунке спектра, получим

$$\vec{y}_j = (0, 0, 0, 0, 0, y_{j6}, y_{j7}, \dots, y_{j20}, 0, \dots, 0, y_{j19978}, \dots, y_{j19990}, 0, \dots, 0), \quad (5.4)$$

где

$$\begin{aligned} y_{j6} &= f_1(6) + f_2(6), & y_{j19978} &= f_{n_s}(19978), \\ y_{j7} &= f_1(7) + f_2(7), & y_{j19979} &= f_{n_s}(19979) \end{aligned} \quad (5.5)$$

и так далее, и так далее.

Выражения (5.5) означают, что шестая компонента y_{j6} равна сумме значений функций размытия для первой и второй налагающихся линий в точке $x=6$, аналогично для точек $x=7, 8, \dots, 20$. Компонента y_{j19978} равна значению функции размытия для последней линии (с номером n_s) в точке $x = 19978$ и т. д. Естественно, что выражения (5.5) справедливы при определенных предположениях об интервалах задания функций размытия; например, эти интервалы могут быть приняты равными удвоенной, утроенной и т. д. полуширине соответствующих аппроксимируемых спектральных линий.

Если считать, что спектры-вектора в базе данных упорядочены по какому-нибудь принципу (например, упорядочены хронологически, как в картотеке JCPDS), то базу данных можно представить

$$Y^{дв} = \begin{pmatrix} 0 & \dots & 010 & \dots & 010 & \dots & 010 & \dots & 0 \\ 0 & \dots & 010 & \dots & & 010 & \dots & 010 & \dots & 0 \\ & \dots & \dots & & & \dots & & \dots & & \\ 0 & \dots & \dots & & & & & & & \\ 0 & \dots & 010 & \dots & & & & & 010 & \dots & 0 \end{pmatrix}. \quad (5.9)$$

При переходе от Y к $Y^{дв}$ можно поставить условие о невключении в $Y^{дв}$ линий, интенсивности которых меньше $I_{отс}$, или о включении в $Y^{дв}$ только 3, 5, 8 и т. д. сильнейших линий. Это приведет к замене нулями некоторых единиц в строках матрицы $Y^{дв}$.

Теперь можно сказать, что матрица Y описывает полный прямой файл, обычно используемый на завершающем этапе идентификации. Матрица $Y^{дв}$ описывает прямой файл положений линий (всех или сильнейших), используемый в некоторых системах на предварительном этапе идентификации. Инвертированный файл Джонсона — Ванда может быть получен путем транспонирования матрицы (5.9):

$$Y^{инв} = (Y^{дв})^T. \quad (5.10)$$

Очевидно, что вместо двух отдельных матриц $Y^{дв}$ размерностью $n \times M$ можно ввести матрицу $Y^{совм}$ размерностью $n \times 2M$, овмещающую свойства (5.7) и (5.9):

$$Y^{совм} = \begin{pmatrix} 0\dots & 01 & 0\dots 01 & 0\dots 01 & 0\dots & 0 \\ 0\dots & 0y_{1,1000}^1 & 0\dots 0y_{1,1500}^2 & 0\dots 0y_{1,4230}^{30} & 0\dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \\ 0\dots 01 & 0\dots & & & & 01 \dots 0\dots 0 \\ 0\dots 0y_{M,895}^1 & 0\dots & & & & 0y_{M,15000}^{85} 0\dots 0 \end{pmatrix}. \quad (5.11)$$

В этой матрице каждому спектру-вектору соответствуют две последовательные строки, представляющие вектор в формах (5.8) и (5.3) (строки для $j = 2$ для краткости в (5.11) опущены). Матрица $Y^{совм}$ представляет двухсекционный прямой файл типа, используемого в программе POISF системы АРФА.

Взяв за основу матрицу Y , можно построить ряд матриц, содержащих меньший объем информации, путем перехода от столбальной шкалы оценки интенсивностей в более грубую, исключения численных значений интенсивности (переход к $Y^{дв}$), учета только сильнейших линий и другими способами сжатия информации. Такая совокупность матриц будет описывать ряд прямых файлов, используемых на последовательно изменяющихся этапах идентификации. Данные матрицы можно совместить в рамках одной $Y^{совм}$, описывающей многосекционный файл, включив в (5.11) соответствующее число строк для каждого спектра-вектора.

В методах распознавания образов [25, 74] совокупность значений x , y называется поисковыми признаками объекта (стандарта),

R_n — фазовым пространством признаков, а y_j — вектором-образом объекта. Мы будем, однако, не слишком часто пользоваться этой терминологией, чтобы не загромождать изложения.

После введения матричного представления базы данных перейдем к изучению ее свойств, важнейшими из которых являются плотность распределения признаков и полнота.

5.3. ПЛОТНОСТЬ РАСПРЕДЕЛЕНИЯ ПОИСКОВЫХ ПРИЗНАКОВ

Введем функцию плотности распределения $N(x)$, показывающую, сколько спектральных линий в базе данных попадает в интервал $(x, x + \delta x)$. Функция $N(x)$ может быть построена численно; для того чтобы получить ее в 20 000 точек (с шагом $\delta x = 0,001$), достаточно подсчитать число единиц в каждом столбце матрицы (5.9). Практически столь *подробного* определения $N(x)$ проводить не требуется; вполне достаточно определить ее с шагом $\delta x \sim 0,1-0,2$ (тем более, что положения многих линий в базе данных известны только с точностью до $0,01 \text{ \AA}$).

В работе [76] функция $N(x)$ представлена в виде гистограммы, вычисленной по совокупности из 297 стандартов с шагом $0,2 \text{ \AA}$ на интервале $0,8-4,0 \text{ \AA}$. Более детальное численное определение, позволяющее представить $N(x)$ практически как гладкую функцию, проведено в [54]. Зависимость $N(x)$ построена в [54] на наиболее информативном интервале значений d : $0,9-6,0 \text{ \AA}$ с шагом $0,1 \text{ \AA}$ по 2000 стандартов (минералы). Данная совокупность стандартных спектров содержит в сумме около 55 000 линий; такой большой объем информации обеспечивает статистически достоверное построение $N(x)$. Полученная в [54] кривая приведена на рис. 5.3. Максимум $N(x)$ лежит в области $1,5-1,6 \text{ \AA}$. Закономерный спад кривой при возрастании x связан с уменьшением числа линий с малыми индексами Миллера. При уменьшении x (т. е.

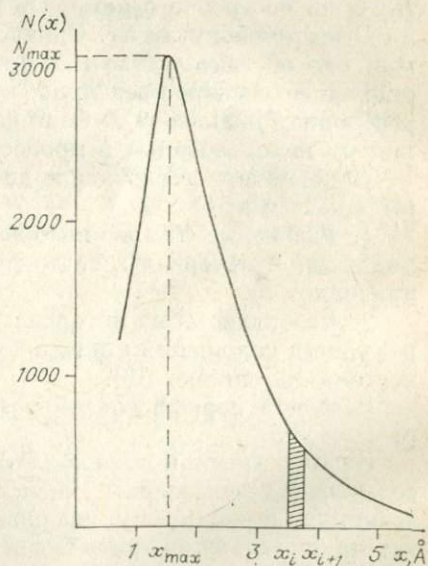


Рис. 5.3. Функция плотности распределения $N(x)$, построенная по совокупности 2000 стандартов минералов.

при увеличении значений индексов Миллера) $N(x)$ должна стремиться к бесконечности; таким образом, функция $N(x)$ при $x \rightarrow 0$ отражает не какую-либо физическую закономерность, а отсутствие методических возможностей для регистрации огромного числа слабых линий с большими h, k, l .

Как отмечалось в [54, 76], анализ $N(x)$ показывает, насколько резко вероятность случайного совпадения линий стандарта и образца зависит от значения x . По данным [54], на интервал $0,1 \text{ \AA}$, взятый в области $1,4-1,8 \text{ \AA}$, приходится около 3000 линий, на такие же интервалы в области $3-4 \text{ \AA}$ — около 700, а в области $5-6 \text{ \AA}$ — около 180 (см. рис. 5.3). Таким образом, вероятность случайных совпадений в указанных областях относится примерно как $20 : 5 : 1$. Данное обстоятельство может быть учтено введением в идентифицирующую функцию веса совпадения $w_1(x) = w(d)$, где $w(d)$ определяется по формуле (4.49).

Поскольку вес совпадения линий в области N_{\max} и при значениях x , больших $5-6 \text{ \AA}$, расходится в десятки раз, линии стандартов с большими x , расположенные в правой части матриц (5.7), (5.9), представляют значительную информационную ценность даже в том случае, когда их интенсивности не превосходят $5-10$ единиц стобальной шкалы. Следовательно, при отборе наиболее информативных признаков спектра нельзя ограничиваться только сильнейшими линиями и отбрасывать средние и слабые линии с большими x , если они имеются.

Рассмотрим связь распределения $N(x)$ с размерностью векторного пространства R_n . В предыдущем разделе пространство R_n было построено с помощью равномерного разбиения диапазона D на промежутки δx . Однако распределение $N(x)$ неравномерное. Это обстоятельство наводит на мысль, что, вероятно, представление спектров-векторов в пространстве с неравномерным разбиением диапазона D было бы более эффективным с точки зрения их использования в процессе идентификации.

Неравномерное разбиение диапазона D можно произвести двумя способами:

1. Разбиение D на интервалы равновероятного попадания линий (т. е. интервалы, содержащие одинаковое число поисковых признаков x).

2. Разбиение D на интервалы, в пределах которых ошибка определения положения линий x (окно Δd) примерно постоянна (при постоянной ошибке $\Delta \theta$).

Разберем первый способ разбиения, второй будет изложен в разделе 5.5.

В рассмотренной нами области задача о построении интервалов, содержащих одинаковое число рентгendifракционных линий, возникла, по-видимому, впервые при создании указателя Ханавальта (см. гл. 1), поэтому будем в дальнейшем называть их интервалами Ханавальта.

Для того чтобы $N(x)$ обладала всеми свойствами функции плотности распределения, ее следует нормировать на единицу:

$$\int_0^D N(x) dx = N_o, \quad (5.12)$$

где N_o — полное число линий, по которому строилась численно $N(x)$. Тогда вместо $N(x)$ можно ввести нормированную функцию

$$\tilde{N}(x) = \frac{1}{N_o} \cdot N(x). \quad (5.13)$$

Вероятность для некоторой линии с положением x попасть в интервал $[x_i, x_{i+1}]$ есть:

$$P(x_i \leq x < x_{i+1}) = \int_{x_i}^{x_{i+1}} \tilde{N}(x) dx. \quad (5.14)$$

Поскольку данная вероятность равна площади под кривой $\tilde{N}(x)$ на промежутке $[x_i, x_{i+1}]$, то для разбиения всего диапазона D или выбранной части этого диапазона на n интервалов равновероятного попадания линий необходимо удовлетворить условие:

$$\int_{x_i}^{x_{i+1}} \tilde{N}(x) dx = \frac{1}{n} = \text{const} \quad (5.15)$$

или же

$$\int_{x_i}^{x_{i+1}} N(x) dx = \frac{N_o}{n} = \text{const}. \quad (5.16)$$

Условие (5.16) означает, что необходимо разыскать все границы n интервалов $x_2, x_3, x_4, \dots, x_{n+1}$; начиная с заданного значения x_1 , причем все интервалы $[x_1, x_2], [x_2, x_3], \dots, [x_n, x_{n+1}]$ должны содержать одинаковое число линий N_o/n . Условие (5.16) является рекуррентным уравнением относительно x_{n+1} . Для его решения произведем аппроксимацию численно заданной функции $N(x)$ аналитическим выражением. Наилучшей аппроксимацией $N(x)$ является ее представление в виде ряда Фурье:

$$N(x) = \frac{a_0}{2} + \sum_{t=1}^T \left(a_t \cos \frac{\pi}{T} xt + b_t \sin \frac{\pi}{T} xt \right). \quad (5.17)$$

Здесь T — число членов ряда Фурье, используемых для аппроксимации $N(x)$. Подставляя (5.17) в (5.16) и интегрируя, получим окончательное уравнение относительно x_{i+1} :

$$x_{i+1} - x_i + \frac{2T}{\pi a_0} \sum_{t=1}^T \left[\frac{a_t}{t} \left(\sin \frac{\pi}{T} t x_{i+1} - \sin \frac{\pi}{T} t x_i \right) - \frac{b_t}{t} \left(\cos \frac{\pi}{T} t x_{i+1} - \cos \frac{\pi}{T} t x_i \right) \right] = 2N_o/n a_0. \quad (5.18)$$

Таблица 5.1

Интервалы Ханавальта, определенные по совокупности 2000 стандартов (минералы)

№ п/п	Граница интервала, А	Число линий в интервале	Вес $w_1(x)$
0	0,9800	1722	1,75
1	1,0830	1679	1,46
2	1,1755	1674	1,37
3	1,2487	1711	1,28
4	1,3148	1719	1,19
5	1,3798	1714	1,12
6	1,4410	1672	1,06
7	1,4985	1748	1,02
8	1,5510	1703	1,00
9	1,6042	1710	1,02
10	1,6590	1767	1,06
11	1,7140	1747	1,10
12	1,7717	1755	1,14
13	1,8355	1743	1,18
14	1,9008	1725	1,28
15	1,9700	1675	1,33
16	2,0395	1793	1,38
17	2,1124	1732	1,48
18	2,1943	1754	1,60
19	2,2790	1742	1,70
20	2,3753	1774	1,85
21	2,4700	1754	2,00
22	2,5730	1739	2,20
23	2,6807	1671	2,35
24	2,8214	1695	2,55

Окончание табл. 5.1

№ п/п	Граница интервала, А	Число линий в интервале	Вес $w_1(x)$
25	2,9552	1735	2,85
26	3,1082	1739	3,40
27	3,2939	1716	4,05
28	3,5151	1716	5,50
29	3,8154	1712	8,00
30	4,2574	1650	11,00
31	4,8100	1706	20,00
32	5,8500		

Таблица 5.2

Распределение $M(y)$

Интервал значений I^{100}	Число линий	$w_2(I)$
до 10	34 004	1
11—20	13 662	2,5
21—30	7 096	5
31—40	5 530	
41—50	3 705	
51—60	2 980	
61—70	1 801	10
71—80	1 774	
81—90	853	
91—100	2 561	
Всего линий	73 966	

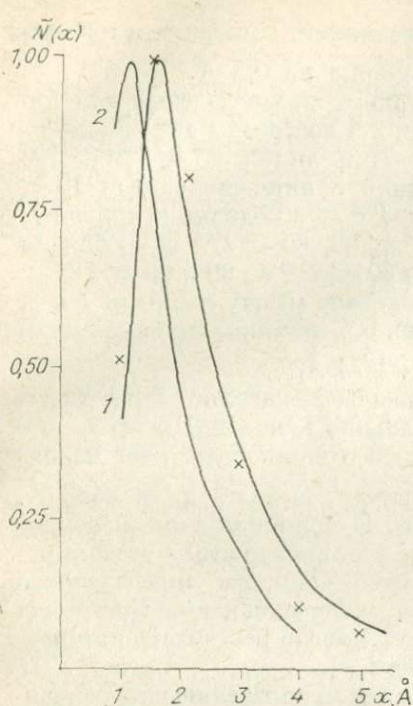


Рис. 5.4. Сопоставление нормированного распределения $\tilde{N}(x)$ (кривая 1) с нормированными распределениями для кубической сингонии (кривая 2) и гексагональной сингонии (крестики).

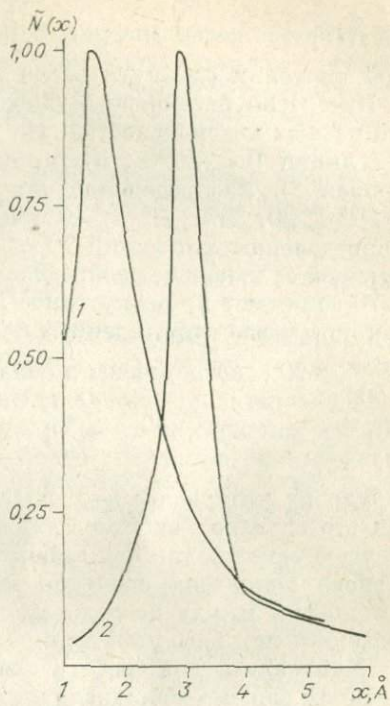


Рис. 5.5. Нормированные распределения для линий с интенсивностями 1—10 (кривая 1) и 90—100 единиц (кривая 2).

В [54] уравнение (5.18) решалось итерационным методом с заданными $x_1 = 0,98 \text{ \AA}$ и $n = 32$. Качество решения проверялось путем прямого подсчета по совокупности из 2000 стандартов числа линий, попадающих в найденные по (5.18) интервалы. Границы интервалов, число попадающих в интервал линий и их относительные веса, вычисленные по формуле (5.11), приведены в табл. 5.1.

Очевидно, что функция $N(x)$ является индивидуальной характеристикой каждой конкретной базы данных и, строго говоря, ее нужно определять отдельно для полной базы (типа картотеки JCPDS) и различных специализированных выборок из нее. В качестве примера на рис. 5.4 нормированная оценка $\tilde{N}(x)$ (см. рис. 5.3) сопоставлена с нормированными распределениями, полученными для минералов кубической и гексагональной сингонии (231 и 388 стандартов, около 4800 и 10 200 линий соответственно). Кривая для гексагональной сингонии почти полностью повторяет

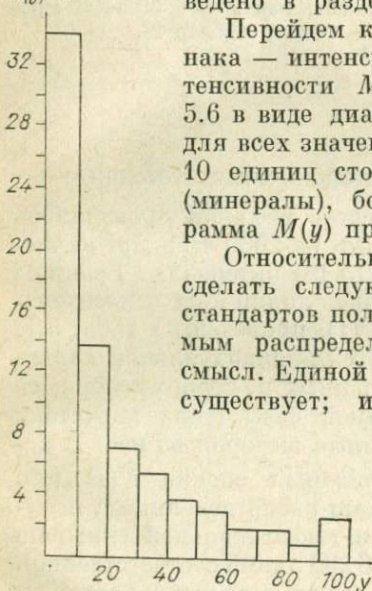
ход $\tilde{N}(x)$ и помечена на рис. 5.4 крестиками. Кривая для кубической сингонии сдвинута влево примерно на $0,5 \text{ \AA}$.

Интересно рассмотреть также кривые распределения для линий разных интенсивностей. На рис. 5.5 построены две кривые — для линий $0 < I^{100} \leq 10$ (кривая 1) и линий $90 < I^{100} \leq 100$ (кривая 2). Распределение для линий с интенсивностями $10 < I^{100} \leq 20$, $30 < I^{100} \leq 40$ и $40 < I^{100} \leq 50$ близки к кривой 1; распределения для линий $70 < I^{100} \leq 80$, $80 < I^{100} \leq 90$ близки к кривой 2; наконец, распределения $50 < I^{100} \leq 60$ и $60 < I^{100} \leq 70$ занимают промежуточное положение между данными 1 и 2. Как показывают приведенные на рис. 5.5 графики, сильные линии с $I^{100} > 70$ сосредоточены в области $2-4 \text{ \AA}$.

Разбиение диапазона D на интервалы Ханавальта неоднозначное, так как зависит от выбора начальной точки x_1 . Выбор $x_1 = 0$ физически не оправдан; целесообразнее остановиться на каком-нибудь из интервала $0-1 \text{ \AA}$ (например, выбрать $x_1 = 0,488 \text{ \AA}$, как это сделано в системе Снайдера). В принципе данный вопрос несущественный, так как выбор того или иного начала отсчета реального диапазона спектров является вопросом определенного соглашения между исследователями; изложенная выше методика позволяет при известной функции $N(x)$ легко рассчитать интервалы Ханавальта для любого значения x_1 .

Более важен выбор числа интервалов n , определяющий размерность векторного пространства с неравномерным разбиением диапазона D . Исследование этого вопроса будет введено в разделе 5.7.

$M(y) \cdot 10^3$



Перейдем к изучению второго поискового признака — интенсивности. Функция распределения интенсивности $M(y)$ или $M(I)$ представлена на рис. 5.6 в виде диаграммы. Зависимость $M(y)$ построена для всех значений интервала $0 < I^{100} \leq 100$ с шагом 10 единиц стобалльной шкалы по 2300 стандартов (минералы), более 70 000 линий. В табл. 5.2 диаграмма $M(y)$ представлена в численном виде.

Относительно распределения $M(y)$ необходимо сделать следующее замечание. Спектры различных стандартов получены в единой шкале по d и тем самым распределение $N(d)$ имеет ясный физический смысл. Единой шкалы отсчета интенсивности пока не существует; интенсивности линий каждого спектра нормированы на интенсивность сильнейшей линии. По этой при-

Рис. 5.6. Диаграмма плотности распределения интенсивности $M(y)$.

чине распределение $M(y)$ имеет минимум при $80 < y \leq 90$, поскольку не во всех спектрах присутствует линия в данной области, но все спектры, по условию формальной нормировки, имеют линию 100 единиц. В последней области $M(91-100) = 2561$, но в нее попадает по крайней мере 2300 линий с $y = 100$ и только порядка 200 линий с $91 < y \leq 99$.

По аналогии с формулой (4.49) можно ввести теперь вес совпадения интенсивностей линий, который будем обозначать $w_2(I)$ или $w_2(y)$:

$$w_2(y) = \frac{M_{\max}}{M(y)}. \quad (5.19)$$

Как неоднократно отмечалось [54, 76], интенсивности измеряются со значительно большей ошибкой, чем d , и являются сравнительно с набором d менее ценным информационным признаком спектра. С учетом этого обстоятельства, а также сделанного выше замечания относительно $M(y)$ можно сделать вывод о невозможности достаточно строгой оценки весов совпадения линий в зависимости от интенсивности. По-видимому, такая оценка должна включать некоторую долю эмпирических соображений.

По распределению $M(y)$ можно выбрать четыре обобщенные градации для оценки $w_2(y)$, приведенные в табл. 5.2. Значение $w_2(y) = 10$ соответствует участку примерно плоского хода $M(y)$ при $y > 40$; $w_2(y) = 5$ — участку начала подъема $M(y)$ при $20 < y < 40$; $w_2(y) = 2,5$ — участку резкого возрастания $M(y)$ при $10 < y \leq 20$ и $w_2(y) = 1$ — максимальному значению $M(y)$ в области 1—10 единиц. Практически в системах автоматической идентификации применяются три оценки $w_2(I)$:

$$w_2(I) \sim I, \quad (5.20)$$

$$w_2(I) \sim \lg I, \quad (5.21)$$

$$w_2(I) \text{ оценивается по эмпирически выбранной шкале.} \quad (5.22)$$

Оценка (5.19) примерно совпадает с оценкой (5.20) и резко расходится с логарифмической оценкой (5.21). Согласно последней оценке, принятой в системе Джонсона — Ванда и АРФА, отношение весов совпадений линий с интенсивностями 10 и 100 единиц равно 1/2, тогда как по (5.19) и (5.20) получается 1/10. Фактически введение логарифмической оценки является попыткой повысить весомость совпадения линий по d (основной признак) сравнительно с весом совпадения по I (дополнительный, менее важный признак). Поскольку d и I являются величинами разной физической природы, оценить теоретически сравнительный вес данных признаков не представляется возможным; вероятно, можно считать, что оценки (5.19), (5.20) и (5.21) в равной степени допустимы.

Изучение распределений $N(x)$, $M(y)$ позволяет дать количественную оценку таких ранее неопределенных понятий, как «сильная», «средняя», «слабая» линии и «наиболее информатив-

ная область диапазона D ». Линии с интенсивностями 100—41 единиц, соответствующие участку примерно плоского хода $M(y)$, можно назвать сильными; с I^{100} в пределах 40—21 — средними, с I^{100} в пределах 20—11 — слабыми и с $I^{100} \leq 10$ — очень слабыми. Если поделить приведенное в табл. 5.2 число линий на число фаз (2300), что будет получено распределение их в среднестатистическом спектре:

Интервал I^{100}	100-61	60-41	40-31	30-21	20-11	10-1	Всего линий
Число линий	3	3	2	3	6	15	32

Теперь можно попытаться интерпретировать часто практикуемый в системах идентификации отбор 3, 5, 8, 10 сильнейших линий спектров стандартов. Отбор трех линий — это отбор наиболее сильных линий, имеющих в среднем интенсивность более 60 единиц, пяти-шести — всех сильнейших линий, восьми — всех сильных линий и наиболее интенсивных средних (несколько менее половины всех средних линий). Наконец, отбор 10—11 линий — это отбор всех сильных и средних линий, которые лежат на диаграмме (см. рис. 5.6) справа от точки начала резкого возрастания $M(y)$.

Опираясь на распределение $N(x)$, $M(y)$, можно ввести понятие полного информационного веса линии $w(x, y)$ или $w(d, I)$. Считая распределения $N(x)$ и $M(y)$ независимыми, определим $w(x, y)$ как

$$w(x, y) = w_1(x)w_2(y) = \frac{1}{N(x)M(y)}. \quad (5.23)$$

Поскольку пределы изменения $w_1(x)$ от 1 до 30 и $w_2(y)$ от 1 до 10, вес $w(x, y)$ меняется в пределах 1—300. Уточненный вес линии $w(x, y)$ может быть использован при расчете критериев соответствия вида (4.37), (4.48).

Как следует из рис. 5.5, распределения $N(x)$ и $M(y)$ не являются, строго говоря, независимыми. Поэтому вместо (5.23) можно использовать более точное выражение

$$w(x, y_1 \leq y \leq y_2) = \frac{1}{N(x/y_1 \leq y \leq y_2) \cdot M(y)}, \quad (5.24)$$

где под $N(x/y_1 \leq y \leq y_2)$ понимается распределение по x линий при условии, что интенсивность линий y попадает в заданные границы y_1, y_2 .

Рассмотрим теперь понятие наиболее информативной области задания спектров. При оценке ее границ необходимо уточнить как вес, так и число попадающих в область линий. В этом смысле наиболее информативная область определяется границами 2—4 Å, в которые попадает около 75% линий с интенсивностями $I^{100} > 70$ и весами $w(x, y)$ от 15 до 80 (см. рис. 5.5). Более широкие границы информативной области — 1—6 Å; в них попадают около 95% линий с $I^{100} > 40$ (с весами 10—200) и около 90% всех ли-

ний. При этом необходимо учитывать, что наиболее информативные линии с весом 200—300 (сильные линии с большим x) расположены в области $x > 5 \text{ \AA}$. Хотя число таких линий мало (не более 3% от полного числа), они являются очень существенными для процедуры идентификации.

5.4. ПОЛНОТА БАЗЫ ДАННЫХ

Будем рассматривать базу данных (5.6) как совокупность векторов \vec{y}_j , $j = 1, 2, \dots, M$ в евклидовом пространстве R_n ; изучим свойства этой совокупности. Поскольку формально уменьшением δx можно получить пространство сколь угодно большой размерности n , будем считать, что $M < n$. В пространстве R_n можно ввести конечный канонический базис $\vec{e}_1 = 1, 0, 0, \dots, 0, \dots, 0$, $\vec{e}_2 = 0, 1, 0, \dots, 0, \dots, 0$, $\dots, \vec{e}_n = 0, 0, 0, \dots, 1$. Тогда любой вектор \vec{y}_j можно представить в виде, аналогичном (5.1) или (5.2):

$$\vec{y}_j = \sum_{i=1}^n y_{ji} \vec{e}_i. \quad (5.25)$$

Из поведения функции $N(x)$ при x , больших 3—4 \AA , ясно, что проекции векторов $\{\vec{y}_{ji}\}$ на многие орты канонического базиса отсутствуют (т. е. в (5.25) соответствующие y_{ji} равны нулю). Физически это означает, что при разбиении диапазона D с шагом $\delta x = 0,001 \text{ \AA}$, в базе данных, например, отсутствуют линии с $d = 6,989 \text{ \AA}$, $d = 8,125 \text{ \AA}$ и т. п. Следовательно, из R_n можно выделить подпространство $R_{n'}$ меньшей размерности $n' < n$ с базисом $\{\vec{e}_i\}$, $i = 1, 2, \dots, n'$, для которого справедливо разложение:

$$\vec{y}_j = \sum_{i=1}^{n'} y_{ji} \vec{e}_i, \quad (5.26)$$

где $y_{ji} > 0$ при $j = 1, 2, \dots, M$; $i = 1, 2, \dots, n'$.

Для построения базиса $\{\vec{e}_i\}$ достаточно просто перечислить все встречающиеся в базе данных значения d , сопоставив каждому орту пространства $R_{n'}$.

Поставим следующий вопрос: является ли совокупность $\{\vec{y}_j\}$, заданная в $R_{n'}$, системой линейно независимых векторов? Поскольку компоненты векторов $\{\vec{y}_j\}$ (интенсивности) и номера ненулевых компонент (положения линий) являются физическими величинами, измеренными с некоторой точностью, то дать строго математический ответ на поставленный выше вопрос нельзя. Физический характер исходных данных, т. е. наличие «шума» как в стандартных, так и в исследуемых спектрах, является основной

причиной некорректности проблемы качественного анализа [12]. Можно, однако, исследовать вопрос о том, являются ли вектора $\{\vec{y}_j\}$ приблизительно линейно зависимыми или независимыми, имеются ли среди них сходные, как много имеется таких векторов и т. д. По-видимому, в реальной базе данных нет векторов, которые могли бы быть представлены как линейная комбинация других векторов. Однако хорошо известно, что в базе данных имеются вектора сходные и даже почти тождественные (см., например, магнезиоферрит и магнетит, табл. 4.3). Сходство векторов может быть чисто случайным, но гораздо чаще оно является следствием сходства структурных и физико-химических характеристик, родственности соединений.

Линейная независимость векторов не является еще условием, достаточным для их хорошей различимости. Например, два вектора на плоскости, соответствующие двум близким точкам, могут быть приняты в качестве базисных ортов; они взаимно линейно независимы, но близкие точки плохо различимы. Следовательно, необходимо выяснить более частный вопрос — образует ли совокупность \vec{y}_j ортонормированную систему или насколько данная совокупность близка к такой системе. Ортогональность векторов означает, что скалярное произведение

$$(\vec{y}_j, \vec{y}_i) = 0 \text{ при } j \neq i \quad (5.27)$$

или для непрерывных спектральных функций (при $\delta x \rightarrow 0$, $n \rightarrow \infty$)

$$\int_0^D y_j(x) y_i(x) dx = 0 \quad \text{при } j \neq i. \quad (5.28)$$

Условие нормировки векторов записывается как

$$(\vec{y}_j, \vec{y}_i) = 1. \quad (5.29)$$

Это условие всегда может быть выполнено за счет соответствующего масштабирования интенсивности линий. Можно использовать другие, чем (5.29), условия нормировки: нормировку по интенсивности сильнейшей линии (традиционная столбчатая шкала) и на суммарную интенсивность всех линий (на площадь спектра).

Если система векторов является линейно независимой, то всегда можно перейти от нее к ортонормированной системе. Однако при таком переходе может быть утеряна физическая суть первоначальной системы. Будем считать, что вектора $\{\vec{y}_j\}$ представлены в виде штрих-диаграмм (5.3), и рассмотрим, какими способами можно было бы добиться их строгой ортогональности. Первый способ заключается в построении по системе $\{\vec{y}_j\}$ новой ортогональной системы $\{\vec{y}_j^{\text{орт}}\}$ с использованием процедуры, сходной с известным

методом ортогонализации Шмидта [70]:

$$\vec{y}_1^{\text{орт}} = \vec{y}_1;$$

$\vec{y}_2^{\text{орт}}$ равен вектору \vec{y}_2 , из которого исключены все компоненты (линии), совпадающие по x с каким-либо компонентом $\vec{y}_1^{\text{орт}}$;

$\vec{y}_3^{\text{орт}}$ равен вектору \vec{y}_3 , из которого исключены все компоненты, совпадающие с какими-либо компонентами $\vec{y}_1^{\text{орт}}$, $\vec{y}_2^{\text{орт}}$ и т. д.

В результате данной процедуры будет получена система из $M^{\text{орт}} < M$ ортогональных векторов $\{\vec{y}_j^{\text{орт}}\}$, $j = 1, 2, \dots, M^{\text{орт}}$, причем очевидно, что вектора при больших значениях j будут содержать немного компонент или даже одну. Однако непосредственно использовать данную систему для идентификации нельзя, так как ортогонализация привела к потере физического смысла векторов $\{\vec{y}_j^{\text{орт}}\}$.

Второй способ ортогонализации заключается в представлении векторов $\{\vec{y}_j\}$ с такой большой точностью, что абсолютно все их компоненты различаются по x (т. е. в базе данных нет совпадающих по положением линий). Будем обозначать такую систему как $\{\vec{y}_j^{\text{точн}}\}$. Предположим, что ее построение можно осуществить экспериментально путем уменьшения δx до $0,0001 - 0,00001 \text{ \AA}$. Легко видеть, что для системы $\{\vec{y}_j^{\text{точн}}\}$ будет выполняться (5.27), так как значимые компоненты одного вектора всегда множатся на нулевые компоненты другого. Предположим, что исследуемый спектр-вектор получен с той же большой точностью δx . Тогда проблема достоверности качественно решена полностью, некорректность отсутствует. Ясно, что данная система идеализирована, однако на первом этапе исследования такая модель может оказаться полезной.

Ортогональная система $\{\vec{y}_j^{\text{точн}}\}$ является линейно независимой и представляет собой базис пространства R_n . Тогда точный исследуемый спектр-вектор $\vec{g}^{\text{точн}}$ можно представить аналогичным (4.32):

$$\vec{g}^{\text{точн}} = \sum_{j=1}^M C_j \vec{y}_j^{\text{точн}}. \quad (5.30)$$

Умножим скалярно левую и правую части уравнения (5.30) на $\vec{y}_1^{\text{точн}}$, $\vec{y}_2^{\text{точн}}$, ..., $\vec{y}_M^{\text{точн}}$. С учетом ортогональности (5.27) и нормировки (5.29) векторов базиса получим, что

$$C_j = (\vec{g}^{\text{точн}}, \vec{y}_j^{\text{точн}}), \quad j = 1, 2, \dots, M. \quad (5.31)$$

Если вектора $\vec{y}_j^{\text{точн}}$ и $\vec{y}_k^{\text{точн}}$ нормированы на единицу, то их скалярное произведение (5.31) дает косинус угла между ними в пространстве R_n [97], что окончательно проясняет смысл коэффициентов C_j (называемых ранее «условными концентрациями»).

$$C_j = \cos(\overrightarrow{g^{\text{точн}}}, \overrightarrow{y_j^{\text{точн}}}). \quad (5.32)$$

Таким образом, в предположении об ортогональности спектр-векторов базы данных можно легко решить исходную систему Филалы (4.32) или (5.30), поскольку вопрос ее решения сводится к расчету M скалярных произведений (5.31). Перейдем теперь к реальному случаю.

В реальной базе данных $\{\overrightarrow{y_j}\}$ ортогональность теряется из-за недостаточной точности задания компонент, что приводит к появлению одинаковых по x компонент у различных векторов. Кроме того, процедура сравнения компонент (линий) исследуемого спектра \overrightarrow{g} с компонентами (линиями) стандарта $\overrightarrow{y_j}$ с помощью неравенств вида (4.3), (4.11) предполагает, что в момент сравнения линии штрих-диаграммы заменяются прямоугольными, треугольными, гауссовыми и другими распределениями. Следовательно, спектры-вектора \overrightarrow{g} и $\overrightarrow{y_j}$ участвуют в процедуре сравнения фактически в виде спектральных функций $g(x)$ и $y_j(x)$, заданных своими численными значениями (форма (5.4)), что также приводит к потере ортогональности.

Пусть база данных достаточно представительная, т. е. содержит совокупность всех возможных стандартных веществ. Тогда коэффициенты C_j будут удовлетворять уравнению замкнутости

$$(\overrightarrow{y_j^{\text{точн}}}, \overrightarrow{g^{\text{точн}}}) = \sum_{j=1}^M C_j^2 \quad (5.33)$$

и система $\{\overrightarrow{y_j^{\text{точн}}}\}$ представится полной в строгом математическом смысле [70]. Систему $\{\overrightarrow{y_j}\}$, в которой ортогональность нарушена, можно считать полной в физическом смысле, имея в виду, что любой исследуемый спектр может быть выражен линейной комбинацией спектров $\{\overrightarrow{y_j}\}$:

$$\overrightarrow{g} = \sum_{j=1}^M C_j \overrightarrow{y_j}. \quad (5.34)$$

Будем использовать далее представление спектров в виде функций и перепишем (5.34) следующим образом:

$$g(x) = \sum_{j=1}^M C_j y_j(x). \quad (5.35)$$

Домножим (5.35) на $y_j(x)$ и проинтегрируем; тогда получим систему уравнения относительно C_1, C_2, \dots, C_M^* :

* Данную систему можно также получить, представив $g(x) = \sum_{j=1}^M C_j y_j(x) + \varepsilon(x)$ и минимизировав $\varepsilon(x)$ по методу наименьших квадратов. Здесь $\varepsilon(x)$ — ошибка или рассогласование между левой и правой частями уравнения (5.35).

$$G_i = \sum_{j=1}^M C_j S_{ji}, \quad i=1, 2, \dots, M, \quad (5.36)$$

где

$$G_i = \int_0^D g(x) y_i(x) dx, \quad (5.37)$$

$$S_{ji} = \int_0^D y_j(x) y_i(x) dx. \quad (5.38)$$

Интегралы G_i являются обобщенными коэффициентами Фурье-функции $g(x)$ [70], а S_{ji} — интегралами перекрытия функций $y_j(x)$, $y_i(x)$. Если бы система $\{y_j(x)\}$ была ортогональной, то, согласно (5.28), $S_{ji} = 0$ при $j \neq i$ и недиагональные члены матрицы системы уравнений (5.36) исчезают, а диагональные $S_{ji} = 1$; тогда $G_j = G_j$.

До сих пор при изложении ряда вопросов данного раздела мы пользовались данными Фиалы [97]. Перейдем теперь к более глубокому анализу свойств системы $\{y_j(x)\}$.

Проведенный выше анализ предполагает, что предельная достоверность идентификации будет достигнута при наилучшей различимости стандартных спектров-векторов, т. е. при их строгой ортогональности. Однако можно получить хорошую различимость, не налагая такого жесткого условия. Если, например, считать, что вектора хорошо различимы, если углы между ними составляют не менее 30° (т. е. $S_{ji} = \cos 30^\circ = \sqrt{3}/2$), то на плоскости можно построить 12 таких векторов. Если не накладывать условие нормировки на единицу и считать, что хорошо различимы вектора с модулями, отличающимися, например, на две единицы масштаба, то на плоскости можно построить бесконечное множество хорошо различимых векторов, удовлетворяющих условиям:

$$\cos(\vec{y}_j, \vec{y}_i) \leq \frac{\sqrt{3}}{2}, \quad (|\vec{y}_j| - |\vec{y}_i|) \geq 2.$$

Таким образом, вектора или функции системы $\{y_j(x)\}$ могут быть достаточно хорошо различимы даже при отсутствии между ними строгой ортогональности; следовательно, вопрос о линейной независимости или ортогональности функций $\{y_j(x)\}$, поставленный в начале раздела, сводится к исследованию их различимости или сходства. В качестве меры различимости можно использовать такую характеристику, как взаимная неортогональность членов системы $\{y_j(x)\}$.

Фактически каждый акт автоматической идентификации — исследование различимости спектров-векторов базы данных. Поскольку во многих случаях идентификация приводит к вполне удовлетворительным и однозначным результатам, можно утверждать, что различимость членов системы $\{y_j(x)\}$ достаточно высока. Было бы желательно, однако, не ограничиваться чисто качественными рассуждениями, а попытаться получить какие-то количе-

ственные характеристики различимости стандартных спектров. Возможно, при этом удастся выделить группы почти тождественных спектров, практически неразличимых рентгеновским методом. Такие группы можно исключить, заменив каждую из них усредненным спектром-вектором, что повысит степень ортогональности в системе $\{y_j(x)\}$ и сократит объем базы данных. Интересно выяснить, сколько имеется таких групп, насколько их исключение уменьшит объем базы данных, какие соединения в них входят и какими методами, отличными от рентгеновских, их все же можно различить. Тем самым, возможно, удастся разграничить идентификацию с помощью только рентгеновских данных и с помощью другой дополнительной информации (например, химической), позволяющей выбирать нужный стандарт из группы почти тождественных по спектру. Возможно также, что анализ базы данных с рассматриваемой точки зрения позволит получить еще какую-нибудь информацию (например, о качестве стандартных спектров).

Численной мерой неортогональности членов системы $\{y_j(x)\}$ являются интегралы перекрывания (5.38). Для их расчета целесообразно воспользоваться представлением спектров в форме (5.4), причем удобнее всего гауссова функция размытия. Тогда спектр будет представлен в виде

$$y_j(x) = \sum_{k=1}^{n_{sj}} a_{jk} \exp[-\alpha_{jk}(x - \beta_{jk})^2]. \quad (5.39)$$

Параметры каждой гауссовой функции a_{jk} , α_{jk} , β_{jk} легко связать с параметрами соответствующей линии — ее интенсивностью, полушириной и положением. Полуширины линий в базе данных не содержатся, поэтому параметры α_{jk} можно вычислить через окно Δd или ΔH . Величину α_{jk} можно выбрать постоянной для всех линий всех спектров и достаточно большой; тогда размытые линии будут узкими, неортогональность за счет перекрытия из хвостов станет незначительной и спектр в гауссовом представлении близким к штрих-диаграмме.

Подставим (5.39) в (5.38); произведение двух гауссовых функций интегрируется аналитически. Таким образом, окончательно получаем [28]:

$$S_{jk} = \sqrt{\pi} \sum_{h=1}^{n_{sj}} \sum_{l=1}^{n_{sl}} a_{jh} a_{il} \exp \left[(-\epsilon_{kl}^{ji}) \cdot \frac{1}{\sqrt{\kappa_{kl}^{ji}}} \right], \quad (5.40)$$

$$\epsilon_{kl}^{ji} = \frac{\alpha_{jh} \cdot \alpha_{il}}{\alpha_{jh} + \alpha_{il}} (\alpha_{jh} - \beta_{il})^2, \quad \kappa_{kl}^{ji} = \alpha_{jh} + \alpha_{il}.$$

Если считать все α_{jh} , α_{il} постоянными и равными α , то (5.40) упростится:

$$S_{ji} = \sqrt{\frac{\pi}{2\alpha}} \sum_{h=1}^{n_{sj}} \sum_{l=1}^{n_{sl}} a_{jh} a_{il} \exp \left[-\frac{\alpha (\beta_{jh} - \beta_{il})^2}{2} \right]. \quad (5.41)$$

Можно провести дальнейшее упрощение (5.41), заменяя экспоненту двумя первыми членами соответствующего степенного ряда.

Легко заметить, что выражение (5.38) для функций $y_j(x)$, $y_i(x)$ адекватно скалярному произведению соответствующих векторов \vec{y}_j , \vec{y}_i . Таким образом, оно характеризует сходство или различие между двумя спектрами-векторами через косинус угла между ними (или через величину проекции одного вектора на направление другого). Однако, хотя эта мера сходства-различия достаточно точно характеризует степень ортогональности спектров-векторов, воспользоваться ею для анализа базы данных затруднительно (даже в рамках модели (5.40), (5.41)). В результате расчета S_{ji} при $j = 1, 2, \dots, M$; $i = 1, 2, \dots, M$, будет получаться симметричная матрица $S_{ji} = S_{ij}$, содержащая $\frac{M(1+M)}{2}$ членов, что при $M \sim 10^4$ составляет $5 \cdot 10^7$. Получение и анализ столь больших объемов численной информации крайне затруднительны даже с учетом мощности современных ЭВМ. Рассмотренная выше расчетная процедура может оказаться полезной для детального исследования степени ортогональности спектров в базах данных небольшого объема (до 300—800 спектров). Анализ крупных баз данных можно осуществить с помощью более быстродействующей, хотя и не столь точной методики, которая будет рассмотрена в разделе 5.7.

5.5. ВЫБОР ОКНА

Выбор окна и процедура сопоставления положений линий с использованием неравенств типа (4.3), (4.11) являются важными элементами процесса идентификации. Рассмотрим, как следует их интерпретировать в рамках введенных выше математических представлений.

Перепишем уравнение Вульфа — Бреггов (1.1)

$$d = \lambda/2 \sin \theta \quad (5.42)$$

и, продифференцировав его по d , получим связь между ошибкой определения θ (окном $\Delta\theta$) и Δd :

$$\Delta d = \frac{\lambda}{2} \frac{\cos \theta}{\sin^2 \theta} \cdot \Delta\theta. \quad (5.43)$$

Выразив $\sin \theta$ и $\cos \theta$ из (5.42) и подставив в (5.43), получим

$$\Delta d = \frac{d}{\lambda} \sqrt{4d^2 - \lambda^2} \cdot \Delta\theta. \quad (5.44)$$

Оценим (5.44) по максимуму для наиболее распространенных анодов с $\lambda > 1 \text{ \AA}$ (Cu, Fe, Co, Cr):

$$\Delta d = \frac{d}{\lambda} \sqrt{4d^2 - \lambda^2} \Delta\theta < \frac{d}{\lambda} 2d\Delta\theta < 2d^2 \cdot \Delta\theta. \quad (5.45)$$

Для обратного межплоскостного расстояния H можно записать так:

$$H = \frac{1}{d} = \frac{2 \sin \theta}{\lambda}, \quad (5.46)$$

отсюда получим

$$\Delta H = \frac{2}{\lambda} \cos \theta \Delta \theta = \frac{1}{\lambda} \sqrt{4 - H^2 \lambda^2} \cdot \Delta \theta. \quad (5.47)$$

Оценка (5.47) по максимуму дает (для $\lambda > 1 \text{ \AA}$)

$$\Delta H = \frac{1}{\lambda} \sqrt{4 - H^2 \lambda^2} \cdot \Delta \theta < \frac{2}{\lambda} \Delta \theta < 2 \Delta \theta. \quad (5.48)$$

Предполагая, что $\Delta \theta = \text{const}$, получим окончательные оценки Δd и ΔH :

$$\Delta d = 2d^2 \cdot \Delta \theta, \quad (5.49)$$

$$\Delta H = 2 \cdot \Delta \theta = \text{const}. \quad (5.50)$$

Если использовать для определения H выражение (4.6), т. е. $H = \frac{1000}{d}$, то вместо (5.50) получим

$$\Delta H = 2000 \Delta \theta = \text{const}. \quad (5.51)$$

При увеличении θ от 2—3 до 90° величина d меняется примерно от 20 до 0,5—0,7 Å. В пределах этого диапазона изменения d окно Δd , согласно (5.43), уменьшается от 800 до 0,5 $\Delta \theta$, а для области 0,8 — 4 Å — в пределах 32—1,3 $\Delta \theta$. Окно ΔH , согласно (5.47), увеличивается от 0,5 до 1,3 $\Delta \theta$ для спектральной области, начинающейся с 0,8 Å, т. е. окно в шкале H значительно меньше зависит от диапазона, чем в шкале d , и может быть принято постоянным.

Перепишем неравенства (4.3), (4.11) в обозначениях, принятых в этой главе.

$$\vec{x}^g - \Delta x \leq \vec{x}^y \leq \vec{x}^g + \Delta x |_{bR_n}. \quad (5.52)$$

Здесь \vec{x}^g и \vec{x}^y — номера соответствующих компонент векторов \vec{g} и \vec{y} , заданных в пространстве R_n . Поскольку величины компонент (интенсивности) при проверке (5.52) не рассматриваются, то можно считать, что \vec{g} и \vec{y} представлены в двоичном виде (5.8). Проверка (5.52) означает, что каждая линия спектров \vec{g} и \vec{y} с высотой 1 в штрих-диаграммах (5.8) заменяется прямоугольником единичной высоты с шириной $\left(\frac{2\Delta x}{\delta x} + 1\right)$; если, например, $\frac{\Delta x}{\delta x} = 2$, то для некоторой значимой компоненты в (5.8) получим

$$\vec{y}^{пв} = (0, \dots, 0, 1, 1, 1, 1, 1, 0, \dots, 0). \quad (5.53)$$

Таким образом, проверка неравенства (5.52) означает переход из пространства R_n в пространство меньшей размерности \tilde{R}_p , где $p = \frac{n}{\left(\frac{2\Delta x}{\delta x} + 1\right)}$. Выполнение неравенства (5.52) в R_n означает выполнение точного равенства

$$\vec{x}^y = x^g |_{\tilde{R}_p} \quad (5.54)$$

в пространстве \tilde{R}_p . Напомним, что положение каждой компоненты в \tilde{R}_p покрывает положение $\left(\frac{2\Delta x}{\delta x} + 1\right)$ компонент в R_n .

Поскольку в шкале d величина $\Delta x = \Delta d$ не является постоянной, то переход из R_n в \tilde{R}_p осуществляется локально при сравнении каждой пары линий образца и стандарта, причем размерность p в пространстве \tilde{R}_p каждый раз иная и определяется величиной Δd . В шкале H с $\Delta x = \Delta H = \text{const}$ переход от R_n к \tilde{R}_p осуществляется глобально для всей процедуры идентификации; фактически, задавая в поисковых параметрах определенное ΔH , исследователь указывает размерность пространства \tilde{R}_p . Если бы было возможно выбрать ΔH постоянным для всех случаев анализа, то можно было бы загрузить шкалу H без риска пропуска каких-либо совпадений. Поясним сказанное на примере.

Пусть $H = \text{Бл. ц.} \left(\frac{1000}{d}\right)$ изменяется от 1 до 2000 ($1000\text{Å} \gg \gg d \gg 0,5 \text{Å}$) и пусть всегда $\Delta H = 2$. Размерность исходного пространства R_n , в данном случае $n = 2000$, а размерность \tilde{R}_p есть $\frac{2000}{5} = 400$. Тогда можно рассчитать H по формуле $H = \text{Бл. ц.} \left(\frac{200}{d}\right)$ и заменить неравенство

$$H^0 - \Delta H \leq H^s \leq H^0 + \Delta H \quad \text{при } H = \text{Бл. ц.} \left(\frac{1000}{d}\right) \quad (5.55)$$

проверкой равенства

$$H^s = H^0 \quad \text{при } H = \text{Бл. ц.} \left(\frac{200}{d}\right). \quad (5.56)$$

Отметим, что переход от шкалы d к шкале H также является переходом от пространства R_n в пространство меньшей размерности R_q . Если переход осуществляется с помощью преобразования

$$H = \text{Бл. ц.} \left(\frac{L}{d}\right), \quad (5.57)$$

где $L = \text{целое число}$ ($L \sim 10^3$ или 10^4), то для диапазона D $0,5 - 10 \text{Å}$ величина H принимает значения $2L, 2L - 1, 2L - 2, \dots, \frac{L}{10}$, и размерность $q = 2L$. Преобразование (5.57) позволяет осу-

цествить второй способ неравномерного разбиения диапазона D , упомянувшийся в разделе 5.3, — разбить D на такие интервалы, в пределах которых окно Δd примерно постоянно (при постоянной ошибке $\Delta\theta$). Следовательно, по заданному $\Delta\theta = \text{const}$ можно определить значение L в (5.57), которое обеспечивает нужную точность по θ в интервалах разбиения D . Выведем формулу связи L с $\Delta\theta$.

Величина интервала Δd_H с номером H следующая:

$$\Delta d_H = \frac{L}{H-0,5} - \frac{L}{H+0,5} = \frac{4L}{4H^2-1} = \frac{L}{H^2}, \quad (5.58)$$

так как $4H^2 \gg 1$ даже для самого минимального значения $H_{\min} = \frac{L}{10}$.

Подставим в (5.58) выражение для H из (5.57) и приравняем Δd_H оценке окна (5.49):

$$\Delta d_H = \frac{d^2}{L} = 2d^2 \cdot \Delta\theta. \quad (5.59)$$

Из (5.59) получим окончательную оценку L :

$$L = \frac{1}{2\Delta\theta}. \quad (5.60)$$

Если взять обычную оценку $\Delta\theta = 0,0125^\circ$ (или в радианах 0,00022) [155], то для L получаем значение порядка 2000, что совпадает с используемым в системах Снайдера и АРФА.

Итак, с помощью соотношения (5.57) можно осуществить разбиение диапазона D на интервалы последовательно увеличивающейся длины. Фактически разбиение D производится за счет вычисления ближайшего целого к $\frac{L}{d}$, благодаря чему в H -й интервал Δd_H попадают все d , соответствующие $H - 0,5$, $H + 0,5$. Отметим, что сравнительно с (5.57), разбиение (5.16) имеет следующие существенные преимущества: 1) с помощью (5.16) можно построить сжатый прямой файл предварительного отбора, который обеспечивает высокую скорость поиска, но, сравнительно с инвертированным файлом, будет иметь значительно меньший объем; 2) использование сжатого прямого файла (т. е. разбиения (5.16)) на этапе предварительного отбора не препятствует использованию на остальных этапах идентификации иных способов разбиения диапазона D . Так, в системе АРФА на этапах 3 и 4 применяется разбиение (5.57) с $L = 2000$.

5.6. МЕРЫ БЛИЗОСТИ

В главе 4 были рассмотрены разнообразные критерии, используемые для численной оценки соответствия стандартного и исследуемого спектров. В рамках развитой в разделе 5.2 модели критерии соответствия интерпретируются как определен-

ные меры близости спектров-векторов. Анализ систем идентификации показывает, что используются два вида мер близости: вероятностная (частотный критерий) и мера, связанная с определением расстояний между сравниваемыми векторами в пространстве R_n (критерий расстояния). Возможна также комбинация этих мер близости.

Оценка качества критериев соответствия должна осуществляться с помощью параметров, характеризующих свойства селективности и стабильности критериев. Введем для этого следующие параметры: 1) селективности P_1 , характеризующий степень разделения истинных фаз и прочих стандартов в базе данных, а также степень присутствия в результирующем списке S_v каких-либо фаз, кроме истинных; 2) стабильности P_2 по числу линий стандарта n_s ; 3) стабильности P_3 по числу фаз в образце или по концентрациям фаз; 4) стабильности P_4 по изменению состава образца, не затрагивающего условия вхождения в образец определяемой фазы.

Возможные численные величины параметров P_1, P_2, P_3, P_4 будут рассмотрены далее. Сейчас отметим, что стабильность по n_s означает, что мера близости не должна сильно изменяться для ряда входящих в образец в одинаковых концентрациях фаз, стандартные спектры которых имеют разное количество линий. Стабильность по числу фаз означает, что мера близости не сильно изменится при варьировании концентрации определяемой фазы. Ясно, что полной стабильности по концентрации достичь нельзя; фазы малой концентрации всегда будут определяться менее достоверно. Стабильность по параметру P_4 означает, что мера близости не будет сильно варьировать, если концентрация соответствующей фазы не изменилась, но остальной состав образца изменен (как в части входящих фаз, так и их относительных концентраций). Из-за возможного случайного наложения линий полной стабильности по параметру P_4 также достичь нельзя. Что касается стабильности по числу линий стандарта, снятых при исследовании образца, то ее в принципе добиться невозможно; пропуск части информативных линий стандарта в пределах съемки спектра образца всегда приведет к резкому понижению меры близости. Отметим также, что при прочих равных условиях удобнее пользоваться мерой близости, которая может быть нормирована на единицу или на 100%.

Перейдем к рассмотрению частотного критерия. Обычно данный критерий характеризует вероятность присутствия фазы в исследуемом образце через относительное число совпадающих линий стандарта:

$$K_1 = \frac{n_{\text{совп}}}{n_s} \quad (5.61)$$

Критерий может быть нормирован на 100% (4.12) или на единицу (5.60). Эффективность критерия (5.63) может быть повышена, если ввести в него веса совпадения линий стандартов $w^*(x, y)$.

Тогда получим (ср. с формулой (4.44), система ФАЗАН):

$$K_1(x, y) = \frac{\sum_{i=1}^{n_{\text{совп}}} w_i^s(x, y)}{\sum_{j=1}^{n_s} w_j^s(x, y)}. \quad (5.62)$$

Иногда вместо критерия (5.60) или наряду с ним используется частотный критерий вида

$$K_2 = \frac{n_{\text{совп}}}{n_0}, \quad (5.63)$$

дающий отношение числа совпавших линий стандарта к числу линий спектра образца. Фактически данный критерий не является относительным; поскольку в процессе идентификации конкретного образца $n_0 = \text{const}$, то $K_2 \sim n_{\text{совп}}$. Кроме этого, критерий K_2 резко нестабилен по параметру P_3 . Чтобы показать это, рассмотрим следующий пример.

Пусть исследуется монофазный образец, спектр которого содержит 20 линий, пусть в процессе идентификации спектр этого образца совпал по всем 20 линиям со спектром соответствующего стандарта; тогда $K_1 = 1$ и $K_2 = 1$. Теперь пусть данная фаза входит в состав трехфазного образца, спектр которого содержит 80 линий, причем из-за наложений линий, ошибок исследователя и т. п. в процессе идентификации удалось опознать только 10 из 20 линий искомой фазы. Тогда $K_1 = 0,5$, а $K_2 = 0,125$, т. е. следовательно, критерий K_2 уменьшился в восемь, а критерий K_1 — только в два раза.

Проведенный выше анализ дает основания утверждать, что использование критерия (5.63) нецелесообразно. Еще более необоснованным является использование произведения критериев (5.61) и (5.63) в системах IDENT, IBM и Снайдера, поскольку нельзя дать разумную физико-математическую интерпретацию подобному или иному способу комбинирования данных критериев.

Рассмотрим критерий расстояния $R(\vec{y}_j, \vec{y}_i)$, который является уникальной мерой близости двух точек в фазовом пространстве R_n , определяемых векторами \vec{y}_j, \vec{y}_i (или функциями $y_j(x), y_i(x)$). Пусть спектры образца и стандарта заданы в самом общем виде функциями $g(x)$ и $y_j(x)$, тогда расстояния между соответствующими точками в евклидовом пространстве определяются как

$$R(g, y_j) = \sqrt{\sum_{l=1}^n (g_l - y_{jl})^2}, \quad (5.64)$$

где g_l, y_{jl} — l -тые компоненты (интенсивности) векторов-функций в пространстве R_n .

Выражение (5.64) легко связать со скалярными произведениями вида (5.31), (5.32), (5.37). Действительно, раскрывая скобки и

почленно суммируя в (5.64), получим

$$R(g, y_j) = \sqrt{\sum_{i=1}^n g_i^2 + \sum_{i=1}^n y_{ji}^2 - 2 \sum_{i=1}^n g_i y_{ji}} = \\ = \sqrt{(g, g) + (y_j, y_j) - 2(g, y_j)}. \quad (5.65)$$

Если считать g и y_j нормированными на единицу, то скалярные произведения (g, g) и (y_j, y_j) равны единице, а $(g, y_j) = \cos(\widehat{g, y_j})$, тогда

$$R(g, y_j) = \sqrt{2[1 - \cos(\widehat{g, y_j})]}. \quad (5.66)$$

Поскольку все компоненты g и y_j положительны, то косинус меняется в пределах $0 \leq \cos(\widehat{g, y_j}) \leq 1$, а $R(g, y_j) — в пределах $\sqrt{2} \geq R(g, y_j) \geq 0$.$

Выражения (5.67), (5.68) показывают, что в основе уникальной меры близости можно использовать как расстояние $R(g, y_j)$, так и скалярное произведение или косинус

$$G_j = (g, y_j) = \cos(\widehat{g, y_j}), \quad (5.67)$$

которые уже были определены ранее в интегральном виде (формулы (5.3), (5.37)). Использование G_j удобнее, так как данная мера имеет естественную нормировку на единицу.

Рассмотрим, насколько введенные универсальные меры близости пригодны в нашем случае.

Если бы $g(x)$ являлась спектром монофазы, соответствующей некоторому стандарту $y_j(x)$, то с точностью до ошибки измерения выполнялись бы следующие равенства:

$$R(g, y_j) = 0, \quad G_j = 1. \quad (5.68)$$

Однако в реальном случае $g(x)$ является суперпозицией m стандартов $y_{j1}, y_{j2}, \dots, y_{jm}$ и определяемая $g(x)$ точка может лежать в пространстве R_n достаточно далеко от точек $y_{j1}, y_{j2}, \dots, y_{jm}$. Следовательно, при прямом использовании мер (5.64) — (5.67) соотношения (5.68) никогда не будут выполняться. Необходимо выделить из функции $g(x)$ те ее компоненты, которые совпали по положению с какими-либо компонентами стандарта $y_j(x)$, т. е. построить приближительную проекцию $g(x)$ на направление $y_j(x)$. Такая проекция g_{y_j} строится путем проверки неравенства (5.52) и отбора соответствующих линий спектров образца и стандарта. Тогда, опуская в (5.64) не нашедшие соответствия компоненты $g(x)$ и $y_j(x)$, получим меру близости между g_{y_j} и y_j :

$$R(g_{y_j}, y_j) = \sqrt{\sum_{k=1}^{n_{\text{совп}}} \sum_{l=1}^{L_k} (g_{kl} - y_{jkl})^2}. \quad (5.69)$$

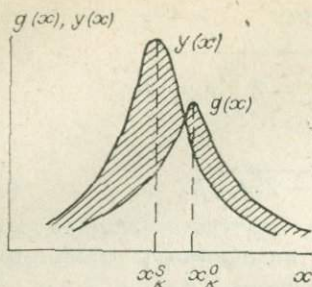


Рис. 5.7. Определение площади неперекрывания $\sum_{l=1}^{L_k} |g_{kl} - y_{jkl}|$ для k -х совпадающих линий спектров стандарта и образца.

Здесь индекс k нумерует совпавшие линии (группы компонент), а l — компоненты в пределах k -й линии, число которых есть $L_k = \left(\frac{2\Delta x_k}{\delta x} + 1\right)$. Для меры (5.69) равенства (5.68) выполняются тем точнее, чем меньше ошибки измерения, связанные со статистикой, сдвигом линий из-за случайных наложений, текстурированности и т. д.

Для (5.69) справедлива оценка сверху.

$$\sqrt{\sum_{h=1}^{n_{\text{совп}}} \sum_{l=1}^{L_h} (g_{hl} - y_{jhl})^2} < \sum_{k=1}^{n_{\text{совп}}} \sum_{l=1}^{L_h} |g_{kl} - y_{jkl}|, \quad (5.70)$$

где правая часть — суммируемая площадь неперекрывания совпавших линий образца и стандарта. На рис. 5.7 показана площадь неперекрывания двух k -х совпадающих линий спектров стандарта и образца при представлении линии гауссовыми функциями; аналогичные построения можно сделать при представлении линий треугольными и прямоугольными распределениями. Если считать, что площади совпадающих линий спектров $g(x)$ и $y_j(x)$ нормированы на единицу (и равны в сумме двум), то в качестве меры близости можно использовать правую часть (5.70) или же, что несколько удобнее, нормированную площадь перекрывания $\Delta S^{\text{норм}}$ спектров стандарта и образца:

$$\Delta S_j^{\text{норм}} = 2 - \sum_{k=1}^{n_{\text{совп}}} \sum_{l=1}^{L_h} |g_{kl} - y_{jkl}|. \quad (5.71)$$

Если не предполагалось заранее нормировки $g(x)$, $y_j(x)$, то площадь перекрывания может быть нормирована путем деления на суммарную площадь совпадающих линий спектров образца и стандарта:

$$\Delta S_j^{\text{норм}} = 1 - \frac{\sum_{k=1}^{n_{\text{совп}}} \sum_{l=1}^{L_h} |g_{kl} - y_{jkl}|}{\sum_{k=1}^{n_{\text{совп}}} \sum_{l=1}^{L_h} (g_{kl} + y_{jkl})}. \quad (5.72)$$

Меру $\Delta S^{\text{норм}}$ можно подвергнуть дальнейшему упрощению. Для

этого будем учитывать в числителе дроби (5.72) только площадь неперекрывания линий стандарта в пределах промежутков $2\Delta x_k$ и относить ее только к суммарной площади совпадающих линий стандарта.

$$\Delta S_j^{\text{норм. I}} = 1 - \frac{\sum_{k=1}^{n_{\text{совп}}} \sum_{l=1}^{L_k} |g_{kl} - y_{jkl}|_{\text{ст}}}{\sum_{k=1}^{n_{\text{совп}}} \sum_{l=1}^{L_k} y_{jkl}}, \quad (5.73)$$

причем

$$|g_{kl} - y_{jkl}|_{\text{ст}} = \begin{cases} (y_{jkl} - g_{kl}) & \text{при } y_{jkl} > g_{kl} \\ 0 & \text{при } y_{jkl} \leq g_{kl} \end{cases}$$

Преобразуем выражение (5.73) к виду

$$\Delta S_j^{\text{норм. I}} = \frac{\sum_{k=1}^{n_{\text{совп}}} \sum_{l=1}^{L_k} (y_{jkl} - |g_{kl} - y_{jkl}|_{\text{ст}})}{\sum_{k=1}^{n_{\text{совп}}} \sum_{l=1}^{L_k} y_{jkl}}, \quad (5.74)$$

причем

$$(y_{jkl} - |g_{kl} - y_{jkl}|_{\text{ст}}) = \begin{cases} g_{kl} & \text{при } y_{jkl} > g_{kl} \\ y_{jkl} & \text{при } y_{jkl} \leq g_{kl} \end{cases}$$

В числителе (5.74) — площадь перекрывания линий стандарта (рис. 5.8).

Можно изменить условие нормировки в (5.74) и относить площадь перекрывания к суммарной площади всех (а не только совпадающих) линий стандарта в пределах задания спектра образца:

$$\Delta S_j^{\text{норм. II}} = \frac{\sum_{k=1}^{n_{\text{совп}}} \sum_{l=1}^{L_k} (y_{jkl} - |g_{kl} - y_{jkl}|_{\text{ст}})}{\sum_{p=1}^{n_s} \sum_{l=1}^{L_p} y_{jkl}}. \quad (5.75)$$

Мера $\overline{\Delta S}_j^{\text{норм. II}} < \Delta S_j^{\text{норм. I}}$, если некоторые линии спектра стандарта пропущены в спектре образца.

Меры близости вида (5.72), (5.74), (5.75), выведенные путем ряда упрощений из основного соотношения (5.64), используются на практике; что касается меры (5.75), то она применяется в таком виде в системах ФАЗАН (критерии R_1 и R_2) и АРФА (критерии R_3 , R_4).

Выведем из (5.73) критерий Джонсона — Ванда (4.18). Для этого используем простейшее представление линий стандарта и об-

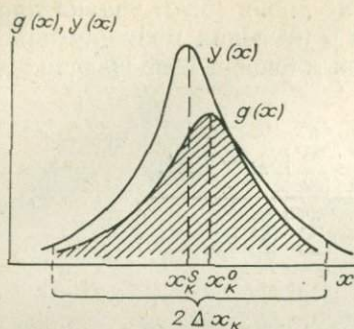


Рис. 5.8. Определение площади перекрытия $\sum_{k=1}^{n_{\text{совп}}} \sum_{l=1}^{L_k} (y_{jkl} - |g_{kl} - y_{jkl}|_{\text{сг}})$ для k -х совпадающих линий спектров стандарта и образца (в пределах окна $2\Delta x_k$).

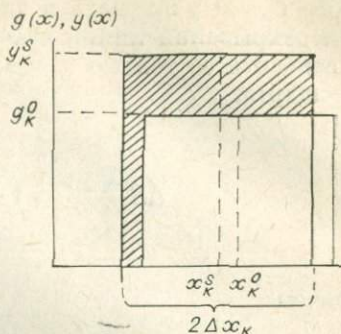


Рис. 5.9. Определение площади неперекрывания для k -й линии стандарта.

разца прямоугольными распределениями (рис. 5.9). Тогда внутренние суммы в (5.73) следующие:

$$\sum_{l=1}^{L_k} |g_{hl} - y_{kl}|_{\text{сг}} = (2\Delta x - |x_k^s - x_k^o|) \cdot (y_k^s - g_k^o) + |x_k^s - x_k^o| y_k^s, \quad (5.76)$$

$$\sum_{l=1}^{L_k} y_{kl} = 2\Delta x \cdot y_k^s.$$

В (5.76) для простоты опущен индекс j , нумерующий функции стандартов; связи используемых в (5.76) стандартных обозначений с обозначениями Джонсона — Ванда в данном случае таковы: $H_k^s = x_k^s$, $H_k^o = x_k^o$, $I_k^s = y_k^s$ (или $I_k^{s,lg} = y_k^s$), $I_k^o = g_k^o$ (или $I_k^{o,lg} = g_k^o$), $\Delta x_k = \Delta x = \Delta H = \text{const}$. Подставив (5.76) в (5.73), получим

$$\Delta S^{\text{норм. I}} \approx \frac{2\Delta x \sum_{k=1}^{n_{\text{совп}}} |y_k^s - g_k^o| + \sum_{k=1}^{n_{\text{совп}}} |x_k^s - x_k^o| \cdot y_k^s - \sum_{k=1}^{n_{\text{совп}}} |x_k^s - x_k^o| \cdot |y_k^s - g_k^o|}{2\Delta x \sum_{k=1}^{n_{\text{совп}}} y_k^s} \approx 1 \quad (5.77)$$

Проведем дальнейшее упрощение (5.77): вынесем из второй и третьей сумм в числителе среднее значение разности

$$\overline{|x_k^s - x_k^o|} = \frac{1}{n_{\text{совп}}} \sum_{k=1}^{n_{\text{совп}}} |x_k^s - x_k^o|. \quad (5.78)$$

Тогда получим

$$\Delta S^{\text{норм. I}} \approx 1 - \frac{\sum_{k=1}^{n_{\text{совп}}} |y_k^s - g_k^o|}{\sum_{k=1}^{n_{\text{совп}}} y_k^s} - \frac{|x_k^s - x_k^o|}{2\Delta x} + \frac{|x_k^s - x_k^o|}{2\Delta x} \cdot \frac{\sum_{k=1}^{n_{\text{совп}}} |y_k^s - g_k^o|}{\sum_{k=1}^{n_{\text{совп}}} y_k^s} =$$

$$= \left(1 - \frac{\sum_{k=1}^{n_{\text{совп}}} |x_k^s - x_k^o|}{n_{\text{совп}} \cdot 2\Delta x} \right) \left(1 - \frac{\sum_{k=1}^{n_{\text{совп}}} |y_k^s - g_k^o|}{\sum_{k=1}^{n_{\text{совп}}} y_k^s} \right). \quad (5.79)$$

Выражение (5.79) полностью соответствует произведению критериев Джонсона — Ванда $F(d) \cdot F(I)$ (формулы (4.16), (4.17)), за исключением того, что в первой скобке в знаменателе стоит $2\Delta H$ ($2\Delta x$ в стандартных обозначениях) вместо ΔH . Для пояснения этого обстоятельства заметим, что, если линия образца попадет чуть дальше левой или правой границы интервала размытия линии стандарта (т. е. x_k^o немного меньше $x_k^s - \Delta x$ или немного больше $x_k^s + \Delta x$, см. рис. 5.9), то перекрывание размытых линий существует, хотя в критерии Джонсона — Ванда такие линии уже исключаются из рассмотрения. В критерии (5.81) несовпадающими считаются полностью неперекрывающиеся линии, т. е. у которых $|x_k^s - x_k^o| > 2\Delta x$. Поскольку Δx — величина относительная и варьируемая, то рассматриваемое обстоятельство является просто вопросом определенного соглашения о том, какие линии считать совпадающими.

Итак, мы показали, что наиболее широко распространенные критерии выводятся из меры близости по расстоянию в пространстве R_n путем ее оценки сверху (5.72) через площадь неперекрывания спектров и использования ряда упрощающих предположений о форме линий. Докажем, что вероятностная мера (частотный критерий (5.63)) представляет собой грубую меру близости по расстоянию. В самой грубой модели будем считать, что все линии спектров стандартов и образца представлены прямоугольными распределениями одинаковой высоты I и ширины $2\Delta x$. Будем также игнорировать неточное совпадение линий стандарта и образца в пределах интервала $2\Delta x$, т. е. если x_k^o попадает в интервал $[x_k^s - \Delta x, x_k^s + \Delta x]$, то прямоугольники линий совмещены полностью и площадь неперекрывания равна нулю. В такой модели внутренние суммы в (5.75) следующие:

$$\sum_{l=1}^{L_h} (y_{hl} - |g_{kl} - y_{kl}|_{\text{ср}}) = \sum_{l=1}^{L_h} y_{hl} = \sum_{l=1}^{L_h} I = 2\Delta x I,$$

$$\sum_{l=1}^{L_p} y_{pl} = \sum_{l=1}^{L_p} I = 2\Delta x I. \quad (5.80)$$

Подставив (5.80) в (5.75), получим определяемый по (5.63) критерий K_1 :

$$\Delta S^{\text{норм. II}} \approx \frac{\sum_{k=1}^{n_{\text{совп.}}} 2\Delta x I}{n_s} = \frac{2\Delta x I \cdot n_{\text{совп.}}}{2\Delta x I \cdot n_s} = \frac{n_{\text{совп.}}}{n_s}. \quad (5.81)$$

Проведенный выше анализ позволяет сделать следующий важный вывод. Критерии соответствия должны иметь разумную физико-математическую (в том числе — вероятностную) интерпретацию. Критерии $K_1, F(d) \cdot F(I), P, R_1, R_2, R_3, R_4$ такую интерпретацию имеют, поскольку отражают в лучшем или худшем приближении расстояние между сравниваемыми спектрами в пространстве R_n . Однако комбинирование данных критериев излишнее и не несет какого-либо физического смысла. Действительно, например, комбинированный критерий Джонсона — Ванда F (4.18), называемый фактором надежности, является просто произведением грубой оценки K_1 площади перекрытия спектров на более точную оценку $F(d) \cdot F(I)$ той же площади перекрытия. Сделанный вывод о необоснованности и искусственности конструирования сложных, якобы более точных, критериев соответствия относится как к критерию (4.18), так и к критериям, определяемым выражениями (4.21), (4.24), (4.26), (4.41). Усложнение и, следовательно, уточнение критериев соответствия должны осуществлять не путем их перемножения или возведения в степень, а путем построения такой модели спектра, которая позволила бы получить более точную оценку универсальной меры близости (5.69). Отметим также, что необоснованное комбинирование грубых и более точных оценок универсальной меры близости разрушает нормировку критериев соответствия.

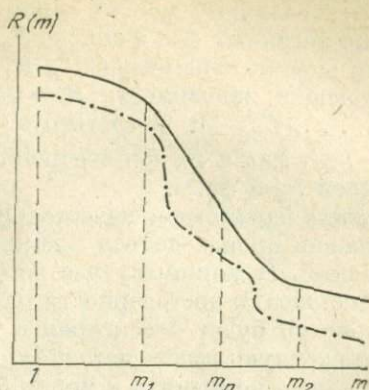
Важным способом улучшения меры близости является учет весовых функций (5.23) или (5.24). В этом случае вместо евклидовой меры расстояния (5.69) в пространстве R_n вводится мера взвешанного.

$$R(g_{vj}, y_j, w) = \sqrt{\sum_{k=1}^{n_{\text{совп.}}} \sum_{l=1}^{L_k} (g_{kl} - y_{jkl})^2 w(x_{jkl}, y_{jkl})}. \quad (5.82)$$

Если исходить из (5.82) и повторить сделанные выше преобразования, то весовая функция войдет во все выводимые формулы для оценки меры близости.

Обратимся к рассмотрению перехода от меры близости к идентифицирующей функции, а также к исследованию свойств последней. Если значения меры близости, вычисленные для $j = 1, 2, \dots, M$ -го стандартов базы данных, упорядочить по убыванию ее величины, то будет получена идентифицирующая функция $R(m)$, в которой m обозначает порядковые номера стандартов, расположенных по убыванию меры близости. В таком виде идентифици-

Рис. 5.10. Обобщенный вид идентифицирующей функции (сплошная линия), возможный реальный ход идентифицирующей функции (штрихпунктирная линия).



рующая функция была введена в разделе 4.11 для критерия (4.47). Обобщенный ход идентифицирующей функции приведен на рис. 5.10.

Идентифицирующая функция позволяет автоматически определить граничное значение критерия соответствия (без задания его исследователем в поисковых параметрах). Для этого необходимо найти точку перегиба m_{Π} или точку m_2 начала области «фона». Точка m_{Π} может быть определена из условия равенства нулю второй производной:

$$\frac{d^2 R}{dm^2} = 0. \quad (5.83)$$

Для определения m_2 можно воспользоваться многочисленными алгоритмами предварительной обработки спектров, позволяющими отделить спектральную линию от фона (т. е. найти начало и конец линии — точки, с которых начинается устойчивое нарастание интенсивности). Определение m_{Π} и m_2 можно проводить численно, с использованием методов скользящего сглаживания или путем аппроксимации $R(m)$ рядом Фурье, полиномом и т. д. В различных ситуациях $R(m)$, спадая от m_1 к m_2 , может образовать ряд уступов; в этом случае имеется несколько точек перегиба, и автоматический анализ по m_{Π} затруднен, вследствие чего целесообразно определять точку m_2 .

Рассмотрим в заключение возможные способы оценки параметров качества P_1, P_2, P_3, P_4 критериев соответствия.

Для оценки P_1 можно использовать две величины:

1) усредненную на участке $[m_1 + 1, m_2]$ первую производную $\left| \frac{dR}{dm} \right|$, характеризующую крутизну спада $R(m)$, т. е. степень разделения между первым и вторым классами; 2) отношение n_{Φ}/m' , где n_{Φ} — число фаз в образце, а m' — порядковый номер найденной фазы; данное отношение характеризует степень присутствия в результирующем списке S_e каких-либо фаз, кроме истин-

ных. Эту величину можно характеризовать суммой порядковых номеров истинных фаз в списке S_B (суммой мест).

Для оценки параметров P_2 , P_3 и P_4 можно построить кривые, выражающие зависимости $R = R(n_s)$, $R = R(C)$ и $R = R(n_{\phi-1}, C_1, C_2, \dots, C_{n_{\phi-1}})$; в третьем случае имеется в виду изменение числа $n_{\phi-1}$ фаз и их концентрации без изменения концентрации последней n_{ϕ} -й фазы.

Оценка параметров качества различных критериев соответствия должна производиться с помощью набора спектров стандартных смесей. В данном случае возникает более широкий вопрос о методике оценки достоверности различных систем идентификации. Подробно он будет рассмотрен в главе 6.

Изложенную выше методику автоматического анализа $R(m)$ можно также применить к мерам близости, построенным с учетом веса. Отметим, однако, что такие меры, сильно зависящие от распределения линий по x и y , резко нестабильны по параметрам P_2 , P_3 и могут давать большой разброс значений $\frac{dR^w}{dm}$ в зависимости от конкретных фаз, составляющих образец.

5.7. СЖАТИЕ РАЗМЕРНОСТИ ВЕКТОРНОГО ПРОСТРАНСТВА И СТРУКТУРИРОВАНИЕ БАЗ ДАННЫХ

Изложенные в разделах 5.3—5.6 вопросы, связанные с распределением поисковых признаков, ортогональностью векторов базы данных, выбором окна и оценок меры близости позволяют продолжить анализ проблемы экспрессной идентификации на более высоком уровне. При этом будут рассмотрены две основные задачи: первая связана со сжатием базы данных с одновременным выделением наиболее информативных признаков спектров, вторая заключается в исследовании базы данных с точки зрения сходства и различия стандартных спектров.

В разделе 5.3 был рассмотрен переход от равномерного разбиения диапазона D , которому соответствует исходное пространство R_n , к неравномерному разбиению диапазона на интервалы Ханавальта. С системой интервалов Ханавальта можно связать новое векторное пространство Q_{n_1} , где n_1 — число интервалов.

Будем рассматривать в дальнейшем для простоты спектры-вектора в двоичном представлении (5.8). Вектор, заданный в пространстве R_n , обозначим \vec{y}_{R_n} , а заданный в Q_{n_1} — $\vec{y}_{Q_{n_1}}$. Переход между этими представлениями одного и того же вектора осуществляется с помощью линейного преобразования L_{RQ} .

$$L_{RQ}\vec{y}_{R_n} = \vec{y}_{Q_{n_1}}. \quad (5.84)$$

Построим представляющую преобразование L_{RQ} матрицу. В качестве строк матрицы используем n_1 векторов $\{\xi_i\}$, $i = 1, 2, \dots, n_1$, размерности n с компонентами:

$$\xi_{ij} = \begin{cases} 1 & \text{при } x \in [x_i, x_{i+1}] \\ 0 & \text{во всех остальных случаях,} \end{cases} \quad (5.85)$$

где $x = j \cdot \Delta x$.

Под $[x_i, x_{i+1}]$ в (5.85) понимается i -й интервал Ханавальта. Применение преобразования

$$L_{RQ} = \begin{pmatrix} \overleftarrow{\xi}_1 \\ \overleftarrow{\xi}_2 \\ \vdots \\ \overleftarrow{\xi}_{n_1} \end{pmatrix} \quad (5.86)$$

к двоичному вектору \overrightarrow{y}_{R_n} приведет к тому, что компоненты вектора $\overrightarrow{y}_{Q_{n_1}}$ будут равны числу линий спектра (значимых компонент \overrightarrow{y}_{R_n} , попадающих в соответствующий интервал Ханавальта). Заменяем после преобразования (5.84) все значимые компоненты $\overrightarrow{y}_{Q_{n_1}}$ единицами; тогда вектор $\overrightarrow{y}_{Q_{n_1}}$ также будет представлен в двоичной форме.

Если размерность $n_1 = n$ и составляет порядка $2 \cdot 10^4$, то точность представления спектра \overrightarrow{y} в пространствах R_n и Q_{n_1} практически одинакова, но в этом случае переход от \overrightarrow{y}_{R_n} к $\overrightarrow{y}_{Q_{n_1}}$ не дает никакой выгоды. Можно, однако, предположить, что представление спектра \overrightarrow{y} в пространстве Q_{n_1} с размерностью $n_1 < n$ не приведет к слишком большому огрублению исходного вектора \overrightarrow{y}_{R_n} (т. е. к потере важных информативных признаков), поскольку пространство Q_{n_1} построено в соответствии с распределением $N(x)$. Можно также предположить, что при равных и малых значениях размерностей n и n_1 представление спектра в пространстве Q_{n_1} всегда будет предпочтительным. Проиллюстрируем изложенные соображения примером.

Рассмотрим спектр магнетита ($N_{\text{JCPDS}} = 19-629$), который содержит 17 линий в области $0,98-6,0 \text{ \AA}$. Данная область разбита на 32 интервала Ханавальта, границы которых приведены в табл. 5.1. Теперь разобьем эту область равномерно на 32 интервала с шагом $\delta x = 0,15 \text{ \AA}$. Построим двоичные представления для спектра магнетита в пространствах Q_{n_1} и R_n при $n_1 = n = 32$:

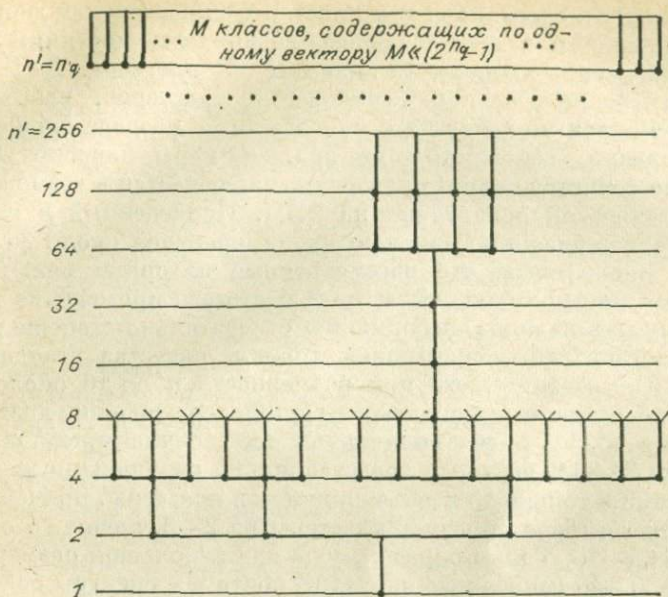


Рис. 5.11. Схема построения кластерной структуры базы данных.

число возможных классов возрастет до 15: $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 0)$, $(0, 0, 0, 1)$, $(1, 1, 0, 0)$ и т. д. до $(1, 1, 1, 1)$. При некотором n' число возможных классов равно $(2^{n'} - 1)$. Отметим, что уже при сравнительно небольших значениях $n' \sim 20$ величина $(2^{n'} - 1) \gg M$ и большинство из возможных классов будут пустыми. Очевидно, что три крупных класса, полученных при $n_1 = 2$, распадаются на 15 классов при $n_2 = 4$, затем на 255 при $n_3 = 8$ и так далее, причем с некоторых размерностей появляются пустые классы. Наконец, при определенной размерности n_q все M исходных векторов попадут в разные классы, т. е. будут формально полностью различимы в двоичном представлении (рис. 5.11).

Под классом понимается некоторое множество объектов, имеющих сходные поисковые признаки. Поскольку в рассматриваемом случае объектами являются спектры-вектора, а сходность признаков означает сходность положений линий спектров, то такие классы можно назвать кластерами* [74]. Изображенная на рис. 5.11 схема представляет, фактически, процесс выделения кластеров или кластеризацию базы данных (структурирование базы данных по сходству спектров). При возрастании n' интерва-

* Наряду с этим термином распознавания образов мы будем использовать и более привычное для физиков название — группа сходных спектров.

лы Ханавальта сжимаются и в кластер с определенным двоичным представлением попадают все более близкие по положениям линий спектры-вектора; происходит локализация кластера.

Рассмотренный алгоритм выделения кластеров, разумеется, не сравнивается по точности с описанным в разделе 5.4 методом кластеризации, основанным на анализе меры близости (5.38); однако достоинство данного алгоритма заключается в возможности его практической реализации на ЭВМ. Приведенный в главе 6 пример кластеризации базы данных, содержащей около 2300 минералов, показывает, что рассмотренный алгоритм может быть аттестован как полукочественный и что его применение позволяет получить важную информацию относительно степени различимости стандартных спектров-векторов и качества стандартных данных. Оказывается, что при размерности $n' = 16$ около 75 % векторов базы данных попадают в группы, содержащие два вектора и более, а 25 % имеют индивидуальное двоичное представление. При $n' = 32$ 85 % векторов представляются различными двоичными формами и только 15 % объединяются в кластеры, причем половина этих векторов образует кластеры из 2—3 членов, а остальные — из 4—10. Таким образом, уже при небольшой размерности порядка 30—40 (много меньше $2 \cdot 10^4$) почти все спектры базы данных могут быть представлены различающимися двоичными векторами. Следовательно, линейная зависимость между стандартными спектрами-векторами сравнительно невелика. При построении файла предварительного отбора можно уменьшить его объем на 7—10 %, заменяя каждую группу сходных векторов (содержащую, как правило, 2—3 члена) обобщенным. Отметим, что среди образующих кластеры векторов (15 % из 2300) имеется некоторое количество почти тождественных (3—4 % из 2000), которые, с учетом влияния ошибок эксперимента, практически не различимы рентгеновским методом. Такие спектры целесообразно заменить обобщенным вектором в файле как предварительного, так и окончательного отборов. Примеры групп сходных и почти тождественных спектров приведены в главе 6.

Сделаем несколько замечаний, связанных с кластеризацией базы данных.

1. Если не считать ситуации случайного сходства спектров, которая встречается редко, кластер образуют спектры родственных соединений, сходных по структурным характеристикам и с близкими значениями параметров кристаллической решетки. Если в базе данных имеются такие группы родственных соединений, то это обстоятельство отразится в кластерной структуре. Таким образом, возможность выделения кластеров связана с составом конкретной базы данных. Совокупность стандартных спектров минералов является очень удобным объектом для кластеризации, поскольку, во-первых, эта выборка часто используется на практике и, следовательно, полезно знать ее свойства, и, во-вторых, использование данной выборки позволяет легко проверить результаты кластеризации, так как известно, что она содержит ряд

групп родственных соединений (например, членов разных изоморфных рядов).

2. Спектры многих родственных соединений различаются из-за неоднозначных условий их получения. Этот вид различия можно трактовать как присутствующую в базе данных систематическую ошибку, из-за которой частично нарушается соответствие между кластерной структурой и совокупностью групп родственных соединений. Если исключить эту ошибку, то при разрядности $n' = 32$ можно ожидать более высокого уровня кластеризации; по-видимому, не менее 25—30 % спектров минералов будут объединяться в кластеры, содержащие от 2 до 10—15 членов.

3. Построенные двоичные представления системы функций $y_j(x)$ очень сильно линейно зависимы. Так, при $n' = 32$ двум тысячам спектров соответствует примерно 1800 различных двоичных 32-разрядных векторов, которые все выражаются как линейные комбинации 32 ортов канонического базиса. Это обстоятельство не следует воспринимать как доказательство значительной линейной зависимости между функциями $y_j(x)$ в R_n или в $R_{n'}$; оно связано с резким понижением размерности пространства (от $2 \cdot 10^4$ до 32), в котором рассматривается система $\{y_j(x)\}$. Понижение размерности векторного пространства, осуществляемое с помощью преобразования (5.86), используется как прием, позволяющий выделить группы сходных спектров и построить файл (файлы) предварительного отбора.

Последнее замечание — хорошая иллюстрация к изложенным в разделе 5.4 рассуждениям о линейной независимости, ортогональности и их связи с различимостью векторов. Несмотря на сильную линейную зависимость двоичных представлений $\{y_j(x)\}$ при $n' = 32$, около 85 % их полностью различимы.

Может показаться, что нами была поставлена первая задача (оптимизация (5.87)), а решается вторая (кластеризация базы данных). Покажем, что они тесно связаны.

Если кластер выделен при некоторой размерности n' (например, при $n' = 4$, см. рис. 5.11), то при повышении размерности (до $n' = 64$) он будет существовать как единое целое, после чего распадается на более мелкие кластеры или на отдельные вектора. Таким образом, по диаграмме на рис. 5.11 можно выделить области устойчивости кластера и указать точки интенсивного ветвления (распада кластера). Размерности векторных пространств, соответствующие точкам интенсивного ветвления, могут быть выбраны в качестве решения задачи (5.87). Построенное таким образом решение, возможно, окажется не самым лучшим в смысле минимизации времени T , однако позволит отбросить множество заведомо худших. Так, на рис. 5.11 показано, что целесообразно выбирать значения $n_1 = 4$ и $n_2 = 64$; выбор любых других n_1 и n_2 в диапазоне 4—64 неоправдан, так как поиск при этих размерностях не приведет к лучшему распознаванию кластера, уже выделенного при $n_1 = 4$. Нецелесообразно также выбирать $4 < n_1 < 64$ и $n_2 > 64$, так как результаты поиска в этом случае совпадут с ре-

результатами поиска при $n_1 = 4$, $n_2 = 64$, но из-за увеличения размерности n_1 и n_2 возрастет затраченное машинное время и увеличится объем файлов предварительного отбора.

Трудно ожидать, что при исследовании реальной базы данных будет получена такая же четкая картина распада кластеров, как изображенная на рис. 5.11. Действительно, в реальном случае наблюдаются два эффекта: некоторого несовпадения точки интенсивного ветвления у разных кластеров и отщепления отдельных членов кластера при повышении размерности (хотя в основном кластер сохраняется). Таким образом, было бы правильнее говорить не о точках, а об областях интенсивного ветвления и относительной устойчивости кластеров. Исследование выборки по минералам позволяет надежно определить по крайней мере одну область интенсивного ветвления при размерности 30—40, соответствующую полному распаду кластеров. Следовательно, при построении идентификации в виде двухэтапной процедуры сжатый файл предварительного отбора целесообразно строить в пространстве указанной размерности.

В настоящем разделе заканчивается обсуждение вопросов, связанных с идентификацией при использовании только спектральных данных. Рассмотрим возможности применения в процедуре идентификации дополнительной информации, прежде всего химической.

5.8. ИСПОЛЬЗОВАНИЕ ХИМИЧЕСКОЙ ИНФОРМАЦИИ

Использование химической информации — важный фактор повышения достоверности результатов автоматического качественного РФА и снижения затрат на его проведение. В работе [130] отмечено, что проще и дешевле получить рентгendifракционный спектр среднего качества плюс данные о химическом составе, чем производить трудоемкие исследования с целью получения спектра с высокой точностью. Как показано в [130] на примере расшифровки 4-фазного образца совместное использование для идентификации дифракционного спектра среднего качества и химической информации позволяет правильно определить все четыре фазы, в то время как при использовании только спектральных данных высокой точности определяются три. Снайдер [155] провел анализ 100 образцов, ряд которых содержал четыре фазы и более. Без задания химической информации программа правильно идентифицировала только 1—2 фазы в наиболее сложных случаях. Сходные выводы сделаны также в работах [54, 55]. В них дается практическое применение системы АРФА.

К настоящему времени разработаны три методики использования химической информации. В первом случае химические дан-

ные непосредственно применяются в процессе автоматической идентификации. Как правило, в поисковых параметрах задаются два списка — обязательно присутствующих и заведомо отсутствующих элементов; данные списки используются для автоматической проверки химического состава вероятных стандартов. Эта методика является достаточно быстросействующей, однако, как отмечает Никольс [135], ее применение может привести иногда к потере полезной информации относительно изоструктурных аналогов разыскиваемых стандартов. Например, возможна ситуация, когда в базе данных отсутствует определяемая фаза, но присутствуют ее изоструктурные аналоги с иным химическим составом. Тогда поиск с автоматическим учетом химических ограничений не даст никаких результатов, в то время как по спектральным данным позволит выявить группу изоструктурных аналогов.

Вторая методика использования химической информации заключается в проведении поиска только по спектральным данным и выводе на печать или дисплей результирующего списка S_b , в котором для каждого вероятного стандарта указывается химическая формула. Отбор по химическому составу осуществляется вручную исследователем путем анализа списка S_b . Как правило, большинство систем идентификации (Джонсона — Ванда, PDSM CIS, Снайдера, ФАЗАН и др.) позволяет включать или не включать по указанию исследователя химические данные в процедуру автоматического анализа. Таким образом, в рамках этих систем можно реализовать обе рассмотренные выше методики.

Третья методика заключается в формировании по полной базе данных специализированной подбазы, которая включает только интересующие исследователя стандарты. Специализированная подбаза может содержать достаточно подробную информацию относительно определенной группы соединений и в то же время иметь сравнительно небольшой объем. При работе с подобной подбазой уменьшается время и повышается достоверность идентификации; при этом нет необходимости проводить автоматическую проверку химических ограничений (т. е. использовать первую методику) и рисковать возможной потерей полезной информации. Если картотека полной базы данных содержит достаточно подробные сведения о стандартном соединении (как в системе АРФА), то при формировании специализированной подбазы можно использовать не только химические данные, но и информацию о структуре соединений, условиях получения спектров и их достоверности.

Рассматриваемая методика реализована в полном объеме в системе АРФА. Системы идентификации Снайдера и фирм Филипс, Сименс и ИВМ также позволяют работать со специализированными файлами, которые формируются либо по вероятностному критерию (файлы MICRO, MINI и MAXI в системе Снайдера), либо с учетом общих физических свойств соединений (минералы, металлы, органика и т. п.), либо по желанию пользователя.

**ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ
ПРОЦЕДУР
ИДЕНТИФИКАЦИИ И БАЗ ДАННЫХ**

Наиболее распространенные системы автоматической идентификации эксплуатируются уже 5—15 лет. Накоплен определенный опыт их использования, который анализируется в этой главе. Будет также рассмотрен комплекс оригинальных исследований, выполненных с помощью системы АРФА и касающихся параметризации системы идентификации и исследования качества стандартов в базе данных.

**6.1. ИССЛЕДОВАНИЯ
И ТЕСТИРОВАНИЕ СИСТЕМ ИДЕНТИФИКАЦИИ**

В конце 70-х годов комитетом JCPDS проведены два крупных испытания, цель которых заключалась в практической оценке автоматизированных методов качественного РФА. Результаты первого испытания рассмотрены в [115], второго — в [119]. В рамках первого испытания 13 специалистов представили 5 контрольных образцов, содержащих 2—4 фазы. Два образца относились к минералам, два — к неорганическим и один — к органическим соединениям. Каждый из участников исследовал образцы на дифрактометре и осуществил попытку идентификации полученных им спектров одним из четырех методов: ручным, с помощью систем Джонсона — Ванда, Никольса или Фривела.

Основные результаты первого испытания сводятся к тому, что полного успеха не добился ни один из исследователей (более того, имеются явные неудачи) и что достоверность результатов, полученных ручным и автоматическими методами, примерно одинакова. Однако необходимо отметить, что это испытание является некорректным относительно оценки способов качественного РФА, поскольку исследователи идентифицировали полученные ими же спектры. В работах Фривела [101] и Ханавальта [104] подробно обсуждаются исследования некоторых из пяти контрольных образцов и приводятся спектры для контрольного трехфазового образца $\text{CuFeS}_2 - \text{BaSO}_4 - \text{Al}_2\text{O}_3$. При этом Фривел приводит в спектре 71 линию, а Ханавальт — почти в 1,5 раза меньше; имеются также расхождения по положениям и интенсивностям линий. Результаты первого испытания, таким образом, в значительной

степени зависели от искусства экспериментатора и точности используемого прибора.

Второе испытание проводилось в иных условиях. Участниками испытания из восьми стран (67 лабораторий) было предложено идентифицировать синтезированные на ЭВМ спектры шести смесей (пять — минералы и неорганика, четырехфазные; один — органика, трехфазный) и один реальный спектр четырехфазной смеси минералов. Синтезированные спектры являются взвешенными суммами стандартных спектров (имитировались концентрации от 10 до 40%), к которым добавлялись распределенные по нормальному закону случайные ошибки. Среднеквадратическое отклонение для ошибок, добавленных к углам 2θ , выбиралось равным 0,02 и 0,05°, а к интенсивностям — 2,5 и 5,0%. Линии, различающиеся по положениям менее чем на 0,1°, совмещались. Для одного из шести спектров добавлялась систематическая ошибка — сдвиг по 2θ на 0,05°. Для первых двух спектров химическая информация не сообщалась; идентификацию остальных четырех можно было проводить с использованием и без использования сведений об элементном составе. Каждый из шести спектров построен в двух вариантах, различающихся параметрами ошибок и возможностями использования химической информации.

Оценка результатов идентификации осуществлялась по следующей системе. В ответах допускалось указание фаз, как безусловно входящих в образец и как возможно входящих. Правильно идентифицированная фаза давала 5 баллов в первом случае и 3 — во втором; неправильно идентифицированная снимала 3 и 1 балл соответственно. Если верно указывалось, что в ответе идентифицированы не все фазы, то дополнительно начислялся 1 балл. Таким образом, максимальная оценка могла достигать 20 баллов.

При идентификации ручными методами Ханавальта и Финка были получены средние баллы 19,1 и 18,1 и среднее время идентификации 2,3 и 2,9 ч соответственно. При идентификации автоматическими методами получены следующие оценки: для системы Джонсона-Ванда (58 актов идентификации) — средний балл 16,8, среднее время 1,7 ч; для прочих систем (17 актов идентификации) — средний балл 20,0 (максимальный) и среднее время 0,7 ч. Было показано, что использование подбазы, содержащей часто встречающиеся стандарты, снижает время анализа в среднем на 40% и увеличивает его достоверность. Отмечалась также целесообразность использования химической информации. В целом второе испытание показало достаточно высокий уровень воспроизводимости результатов идентификации.

Сделаем несколько комментариев по результатам второго испытания:

1) большое время автоматической идентификации (1,7 и 0,7 ч) объясняется, по-видимому, тем, что в него включалось также время ручного анализа списка вероятных фаз, выданного ЭВМ, и время возможных повторных просчетов программы с изменением параметров поиска; 2) высокий средний балл (16,8 — 20,0) связан

с синтетическим характером спектров и отсутствием в них фаз малой концентрации. Этим же объясняется сравнительно небольшое время полного решения задачи в ручном варианте (порядка 2—3 ч).

При исследовании реальных образцов возможны значительно более сложные ситуации, чем предложенные во втором испытании. Отчасти это можно проследить по работам Дисмора [92] и Фиссера [160], в которых рассматривается применение системы Джонсона — Ванда с точки зрения пользователей. Оба автора высоко оценивают как перспективность методики автоматической идентификации, так и систему Джонсона — Ванда. Дисмор отмечает, что в реальных ситуациях использования этой системы необходимо до 35 параметров поиска, поскольку принимаемые в программе «по умолчанию» стандартные значения поисковых параметров часто не позволяют получить верные результаты. Правильная параметризация системы Джонсона — Ванда дается исследователю только долгой практикой и хорошим знанием алгоритмов поиска и выбора. Следовательно, подчеркивает Дисмор, результативность применения методов автоматического РФА зависит от степени владения таким сложным и тонким инструментом, как программа идентификации. Это его мнение можно в той или иной степени отнести ко всем системам, рассмотренным в главе 4.

При экспериментальных исследованиях систем идентификации необходимо различать две задачи — испытание системы и сравнение систем. Вторая задача более общая; для ее объективного решения необходимо сравнивать различные системы идентификации в равных условиях. Это означает, что системы должны испытываться на ЭВМ одного типа, на базах данных равной полноты (включающих одинаковый набор стандартов) и по единому набору контрольных задач. Такого рода сравнительные испытания возможны тогда, когда достаточно четко отработана методика испытания единичной системы.

Синтезированные спектры можно использовать на начальном этапе проверки системы идентификации, однако с их помощью трудно оценить время анализа и обеспечиваемую системой достоверность при исследовании реальных объектов. Результаты второго испытания системы АРФА [18, 54, 55] показали, что идентификация синтезированных спектров практически всегда осуществляется с высокой достоверностью. Таким образом, факт успешной работы системы с тестами, включающими синтезированные спектры, говорит лишь о том, что проверяемая система построена достаточно разумно и можно переходить к более сложным испытаниям. Методика подобных испытаний на реальных объектах была разработана в [18, 54, 55] и использована при исследовании программы POISF системы АРФА (для варианта программы, реализованного на мини-ЭВМ М-6000, СМ-1).

6.2. МЕТОДИКА ИСПЫТАНИЙ СИСТЕМЫ АРФА

Основа методики испытаний — набор контрольных тестов, включающих спектры реальных образцов. В мировой и отечественной практике отсутствует какая-либо совокупность рентгендифракционных спектров, которую можно было бы рассматривать как общепринятые контрольные тесты. Отсутствуют также соглашения о том, спектры каких соединений (минералов) целесообразно использовать для формирования подобного набора тестов. По сути дела, работа [18] является единственной публикацией, в которой приведено значительное число спектров, полученных именно с целью построения набора контрольных тестов для качественного анализа (как ручного, так и автоматического). Разумеется, данная задача выполнена в [18] лишь частично, поскольку полный набор тестов должен явиться плодом работы многих исследователей и определенного соглашения между ними.

Целесообразно разбить набор контрольных тестов на две части. Спектры первой части должны быть получены от реальных образцов ряда широко распространенных монофаз (таких, как кварц, кремний, кальцит, магнетит, флюорит, корунд, анатаз, рутил и т. п.) и от искусственных смесей этих монофаз. Наличие спектров монофаз позволяет сравнить их с соответствующими стандартными спектрами для оценки различия между ними, а также произвести простейшую проверку системы — идентификацию монофаз.

Искусственные смеси целесообразно компоновать в соответствии со следующими правилами: 1) должен быть достаточно широко перекрыт диапазон многофазности, т. е. получены смеси 2-, 3-, 4-, 5-, 6-фазные; 2) для каждой смеси с числом фаз n_{ϕ} должен быть достаточно широко перекрыт диапазон возможных концентраций от 5—10 до 80—90%; 3) спектры некоторых смесей могут быть представлены в нескольких вариантах, различающихся точностью получения набора $\{d, I\}$, числом аналитических линий, наличием или отсутствием систематической ошибки.

Спектры второй части набора тестов должны быть получены от реальных 2-, 6-фазных объектов, являющихся продуктами природного или промышленного происхождения (руд, сплавов, синтетических материалов). Целесообразно подобрать эти продукты таким образом, чтобы они являлись достаточно характерными, широко распространенными и включали не слишком простые объекты, например полевые шпаты, амфиболы и т. п.

Испытание системы идентификации с помощью набора спектров искусственных смесей (первая часть теста) позволяет дать оценку рассмотренных в разделе 5.6 параметров качества P_1, P_2, P_3, P_4 , а также исследовать зависимость времени и достоверности анализа от параметров поиска (размеров окна, граничных значений критериев соответствия, учета химических ограничений и т. п.). Испытания с помощью набора спектров реальных продуктов (вторая часть теста) должны определить работоспособность системы в наиболее жестких условиях.

Рассмотрим результаты испытаний программы POISF системы АРФА, при проведении которых была частично реализована описанная выше методика.

В табл. 6.1 приведены некоторые результаты автоматической идентификации для ряда искусственных смесей (обр. 1—19) и реальных продуктов, имеющих в основном геологическое происхождение (обр. 20—28). Подробные спектры монофаз кварца, кальцита, галенита, магнетита, селлаита, а также спектры соответствующих искусственных смесей (обр. 1—12) приведены в работе [18]. В табл. 6.1 в графе «Определенные фазы» для каждого образца указывается номер найденного по картотеке JCPDS стандарта, а под чертой — соответствующее значение R_m . Анализ обобщенных в табл. 6.1 исследований показал, что, как правило, искомые фазы могут быть достоверно определены при значениях $R_m \geq 80$. При $80 > R_m \geq 50$ искомые фазы можно определить путем анализа таблиц соответствия типа табл. 4.7 без привлечения химической информации, хотя в этом случае наличие сведений о составе образца уже оказывается полезным. При $50 > R_m \geq 30$ для достоверной идентификации необходимы данные о химическом составе образца. Для значений $R_m < 30$ даже при наличии химической информации искомые фазы выделить не удалось.

Для искусственных смесей (обр. 1—19) анализ с помощью системы АРФА практически всегда дает достоверные результаты. Сравнение системы АРФА с системами Снайдера, фирм IBM и Сименс, также разработанными на мини-ЭВМ, показывает, что достоверность получаемых с помощью АРФА результатов не ниже, а быстродействие в 2—3 раза выше, чем у соответствующих зарубежных аналогов. К сожалению, в литературе отсутствуют достаточно подробные данные о применении указанных выше систем, а также систем Джонсона — Ванда, Фривела и других к анализу реальных объектов (обр. 20—28). В этом случае определение искомым фаз часто бывает затруднительным или даже невозможным из-за неудовлетворительного качества спектров — пропуска существенных аналитических линий и значительного сдвига пиков. В обр. 20—28 имеют место оба обстоятельства, причем данные спектры не исправлялись на систематическую ошибку с использованием внешнего или внутреннего стандарта.

При исследовании спектров реальных объектов, как правило, удается либо полностью определить состав исследуемого продукта (иногда с привлечением сведений о химическом составе), либо надежно выделить одну-две основные фазы, вычесть их из исследуемого спектра и путем повторного поиска найти еще одну фазу. Однако не во всех случаях удается добиться полной идентификации. Для некоторых образцов остаются неидентифицированными слабые и очень слабые линии ($I \sim 1-4$, обр. 25, 28), что свидетельствует о присутствии в образцах фаз малой концентрации. В связи с этим становится необходимым исследование концентрации чувствительности метода, т. е. параметра P_3 .

Концентрационная чувствительность метода определяется пу-

Некоторые результаты анализов, проведенных с помощью программы POISF системы АРФА

№ п/п	Известный состав образца	Определенные фазы (номера по PDF—JCPDS и значения R_m)			Примечания
<i>Искусственные смеси — первая часть набора тестов</i>					
	Смеси кварц — кальцит с составом:	5-490		5-586	Для определения кварца и кальцита в 10%-ной концентрации (обр. 1 и 5) привлекались сведения о химическом составе
1	90 : 10	80,5		36,5	
2	75 : 25	88,3		58,5	
3	50 : 50	77,1		70,7	
4	25 : 75	60,9		71,5	
5	10 : 90	35,8		65,6	
	Смеси галенит — магнетит — селлаит с составом:	5-592	19-629	6-290	Вместе с магнетитом устойчиво определяется магнезиоферрит со значениями R_m порядка 90—96
6	40 : 30 : 60	84,5	85,1	83,1	
7	60 : 40 : 30	92,8	83,2	61,7	
8	30 : 60 : 10	91,0	92,4	53,7	
9	25 : 25 : 50	86,8	88,6	79,5	
10	50 : 25 : 25	87,6	92,5	58,5	
11	25 : 50 : 25	87,0	91,1	67,6	
12	1/3 : 1/3 : 1/3	90,1	87,9	63,0	
13	Анализ — рутил	21-1272		21-1276	
		91,1		95,2	
14	Анализ — рутил — флюорит	21-1272	21-1276	4-864	
		96,1	94,8	92,5	
15	Тенорит — куприт	5-662		5-667	
		74,8		97,9	
16	Корунд — бадделит	10-173		13-307	
		76,8		81,3	
17	Корунд — нантоцит	10-173		6-344	
		80,8		75,0	
18	Анализ — рутил — корунд	21-1272	21-1276	10-173	
		94,8	98,1	69,6	
19	Халькопирит — барит — гипсит	9-423	5-448	7-324	
		77,9	84,8	56,6	

Продолжение табл. 6.1

№ п/п	Предполагаемый состав образца	Определенные фазы (название, номера по PDF—JCPDS и значение R_m)			Примечания
<i>Реальные продукты — вторая часть набора тестов</i>					
20	Кварц — флюорит	5-490		4-864	
		76,1		77,1	
21	Каолинит — ортоклаз	Фазы не определены			Низкое качество спектра

№ п/п	Предполагаемый состав образца	Определенные фазы (название, номера по PDF—JCPDS и значение R_m)			Примечания
22	Кварц — гематит	$\frac{5-490}{60,7}$	$\frac{13-534}{56,3}$	Фазы определяются только с привлечением сведений о химическом составе	
23	Муллит — корунд — шпроксен — минерал из группы полевых шпатов	$\frac{15-776}{95,6}$ Микро- клин $\frac{19-932}{56,2}$	$\frac{10-173}{78,9}$	$\frac{19-769}{60,1}$	
24	Диопсид — минерал из группы хлоритов — минерал из группы амфиболов — магнетит	Диоп- сид $\frac{11-654}{75,2}$	Нимит $\frac{22-712}{82,0}$	Тиро- дит $\frac{23-603}{}$	Магнетит отсут- ствует. Тиродит — фаза малой кон- центрации; иден- тифицирован после вычитания из исследуемого спектра стандар- тов 11-654 и 22-712
25	Кварц — эльбаит	$\frac{5-490}{62,0}$	$\frac{21-251}{51,9}$	Неидентифици- рованы 30% ли- ний (слабые); при- сутствует фаза ма- лой концентрации	
26	Хромит — минерал из группы хлоритов	Хромит $\frac{3-873}{70,6}$ Гонье- рит $\frac{10-387}{}$	Магнезиохромит $\frac{9-383}{93,6}$	$\frac{10-351}{82,8}$	Гоньерит иден- тифицирован после вычитания из исследуемого спектра стандар- та 9-383
27	Кальцит — апофил- лит	$\frac{5-586}{99,3}$	$\frac{19-944}{50,5}$	Пропущен ряд линий апофилли- та, его определе- ние возможно только с привле- чением сведений о химическом со- ставе	
28	Минерал из группы слюд — эпидот — мине- рал из группы амфи- болов	Флого- пит $\frac{16-344}{88,6}$	Низкое каче- ство спектра, не идентифицировано 80 % линий		

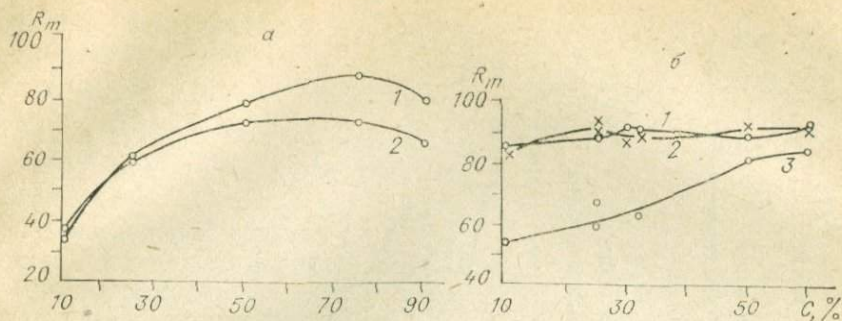


Рис. 6.1. Зависимость $R(C)$ в смесях.

α — кварц (1) — кальцит (2) и β — галенит (1) — магнетит (2) — селлаит (3). Для второй смеси при $C = 25\%$ приведено по два значения R_m в соответствии с табл. 6.1 (№ 9 и 10, 9 и 11).

тем анализа ряда смесей разного состава (кварц — кальцит, галенит — магнетит — селлаит, см. табл. 6.1) с варьированием содержаний фаз от 10 до 90%. На рис. 6.1 представлена зависимость идентифицирующей функции R_m для входящих в исследуемые смеси фаз от концентраций C . Видно, что при $C = 10\%$ значения R_m могут опускаться до 50% и ниже. Таким образом, определение фаз 10-процентной концентрации уже может представлять серьезную проблему и в этом случае полезно привлечение сведений о химическом составе. Надежная идентификация фаз с концентрациями меньше 10% возможна только с привлечением дополнительной информации о химическом составе и структуре исследуемого материала. Сделанные выводы справедливы в том случае, когда идентификация смеси осуществлялась с использованием базы данных, содержащей сведения о нескольких тысячах соединений. Как отмечалось в главах 4 и 5, на практике достаточно часто осуществляется работа с определенной группой соединений и объем соответствующей базы данных не превосходит 500—1000 спектров. Здесь достоверность результатов идентификации повышается, так как уменьшается вероятность случайного совпадения спектров стандартов и исследуемого образца. Уменьшение базы данных, естественно, повышает также возможность определения фазы малой концентрации.

Исследования серии образцов с составом галенит-магнетит-селлаит позволяют сделать также некоторые выводы о стабильности по параметру P_4 . В обр. 6—12 каждая составляющая фаза дважды входит в одинаковых концентрациях (25%) и дважды — в почти одинаковых (30 и 33,3%). При этом, если рассмотреть любую пару смесей, в которых концентрация одной из фаз постоянна, видно, что концентрации двух других фаз меняются значительно — в 2—3 раза (см. обр. 9 и 11, 8 и 12, см. табл. 6.1). Для исследования P_4 среди обр. 6—12 можно выбрать 6 пар смесей: 9, 11 и 8, 12 (для галенита), 9, 10 и 6, 12 (для магнетита), 10, 11 и 7, 12 (для селлаита). В пяти случаях величины R_m исключитель-

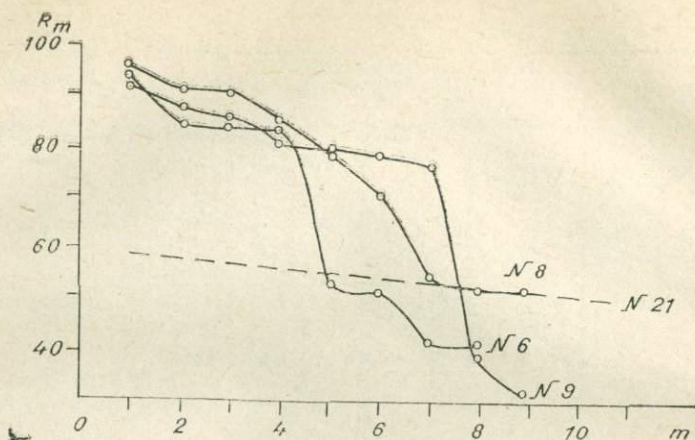


Рис. 6.2. Вид идентифицирующей функции R_m для образцов № 6, 8, 9 и 21 (штриховая линия) из табл. 6.1.

но стабильны. Максимальное относительное изменение наблюдается для магнетита в обр. 9 и 10 и составляет $\frac{92,5 - 88,6}{0,5(92,5 + 88,6)} \approx 0,04$. Однако в шестом случае (обр. 10, 11) наблюдается значительное относительное изменение R_m для селлаита, равное $\frac{67,6 - 58,5}{0,5(67,6 + 58,5)} \approx 0,14$. Рассмотрим, с чем это связано. В смесях

обр. 6—12 сильнейшая линия магнетита ($d = 2,532 \text{ \AA}$, $I = 100$) совпадает с одной из линий селлаита ($d = 2,545 \text{ \AA}$, $I = 22$). При переходе от обр. 10 к обр. 11 концентрация магнетита увеличивается в 2 раза, что ведет к резкому возрастанию интенсивности соответствующей наложенной линии [18] и, как следствие, к увеличению значения критерия соответствия для селлаита. Данный пример демонстрирует чувствительность критерия соответствия к ситуации случайного наложения сильных линий одной фазы на слабые другой.

Рассмотрим теперь, каким образом можно дать оценку параметра селективности. Выполним ее на примере обр. 6, 8, 9, идентифицирующие функции и число совпавших линий для которых приведены в табл. 6.2 (графа R_m/l), а также представлены на рис. 6.2. Крутизну спада функции $R(m)$ оценим по приближенной формуле

$$R'_m = \left| \frac{dR}{dm} \right| \approx \frac{R(m_1) - R(m_2)}{m_2 - m_1}. \quad (6.1)$$

По данным табл. 6.2, величина $R(m_1) \approx 80$, а $R(m_2) \approx 40-50$. Степень присутствия в списке вероятных фаз лишних стандартов оценим по отношению n_ϕ/m' , где m' — номер последней найденной истинной фазы (см. раздел 5.6). Для трех примеров в табл. 6.2

получим следующие характеристики: обр. 6 — $R'_m \approx 14$, $n_{\phi}/m' = 3/4$; обр. 7 — $R'_m \approx 10$, $n_{\phi}/m' = 3/7$; обр. 8 — $R'_m \approx 23$, $n_{\phi}/m' = 3/4$.

Приведенные оценки являются удовлетворительными и позволяют отделить вероятные фазы от фона при $R(m_2) \approx 40-50$, кроме селлаита в обр. 7, для надежного определения которого нужно анализировать таблицы соответствия. Для примера на рис. 6.2 пунктиром показан ход R_m для обр. 21. В этом случае $R'_m \approx 0,5$, т. е. в базе данных не обнаружены спектры, хорошо соответствующие искомому. Данный пример подробно рассмотрен в [54], где показано, что линии спектра обр. 21 расходятся по положениям с линиями стандартов для каолинита и ортоклаза на величины от нескольких сотых до 0,1—0,2 Å.

Интересно отметить, что для всех обр. 6—12 с очень хорошим критерием соответствия определяется магнезиоферрит, спектр которого очень похож на спектр магнетита. В этом случае даже знание химического состава образца не помогает сделать выбор между магнетитом Fe_3O_4 и магнезиоферритом $MgFe_2O_4$, поскольку в состав селлаита MgF_2 также входит магний.

Таблица 6.2

Некоторые результаты анализа смеси галенит — магнетит — селлаит с переменным соотношением концентраций

Обр. 6 (10 : 30 : 60)				Обр. 8 (30 : 60 : 10)				Обр. 9 (25 : 25 : 50)			
Номер	Название	R_m/l	Номер	Название	R_m/l	Номер	Название	Номер	Название	R_m/l	
17-464	Магнезиоферрит	95,2/6	17-465	Магнезиоферрит	96,1/7	17-465	Магнезиоферрит	17-465	Магнезиоферрит	92,7/7	
19-629	Магнетит	85,1/6	19-629	Магнетит	92,4/7	19-629	Магнетит	19-629	Магнетит	88,6/7	
5-592	Галенит	84,5/7	5-592	Галенит	91,0/7	5-592	Галенит	5-592	Галенит	86,8/7	
6-290	Селлаит	83,1/11	4-829	Периклаз	83,7/3	6-290	Селлаит	6-290	Селлаит	79,5/11	
4-755	Магнетит	52,2/4	7-13	Криптогалит	77,6/6	7-13	Криптогалит	7-13	Криптогалит	79,0/6	
5-522	Стиллит	51,3/2	8-322	Уванит	69,7/6	4-829	Периклаз	4-829	Периклаз	77,4/3	
6-266	Циркон	41,9/7	6-290	Селлаит	53,7/6	8-322	Уванит	8-322	Уванит	76,4/5	
13-157	Акаганеит	40,9/9	22-1012	Франклинит	52,2/4	22-1120	Шпротлин	22-1120	Шпротлин	37,5/6	
			5-522	Стиллит	51,6/2	17-452	Метановачекит	17-452	Метановачекит	30,8/5	

Отметим, что кривые идентификации можно использовать для оценки качества экспериментальных спектров (в предположении, что качество стандартных данных достаточно высокое). Если кривые идентификации имеют вид, представленный на рис. 6.2 для обр. 6, 8, 9, то качество экспериментальных спектров удовлетворительное. Если идентифицирующая функция не имеет области резкого спада, т. е. не позволяет выделить спектры стандартов, похожие на исследуемый (см. рис. 6.2, пунктирная кривая), то качество исследуемого спектра неудовлетворительное. Причина неудовлетворительного качества может заключаться как в плохой съемке образца, так и в объективных обстоятельствах — сдвиге линий из-за наличия текстуры, высокой степени наложений линий в многофазном объекте, отсутствии в базе данных нужных стандартов (хотя последняя причина для больших баз данных маловероятна). Величина R'_m может быть принята за численную меру качества идентифицируемого спектра; необходимо сознавать, что эта мера качества основана на предположении о высокой достоверности стандартов.

Поскольку идентифицирующая функция является в конечном счете оценкой меры близости (5,37), (5,40), (5,72), (5,73) или (5,75), то с ее помощью можно осуществить выделение кластеров. Фактически в табл. 6.2 выделен кластер, включающий магнетит и магнезиоферрит. Кластеры будут выделяться при любом акте автоматической идентификации, если для составляющих образец фаз в базе данных имеются родственные соединения или соединения со случайно сходными спектрами. Исследователи, использующие системы автоматической идентификации, встречаются с этим явлением довольно часто. Целенаправленное выделение кластеров заключается в идентификации спектров монофаз, причем в этом случае представляет интерес не идентификация как таковая, а отбор группы стандартов, спектры которых похожи на спектр, предъявленный системе. Такой локальный способ поиска кластеров полностью совпадает с описанным в разделе 5.4, где отмечались также его недостатки — большая трудоемкость и невозможность проведения его в автоматическом режиме.

Перейдем к изучению зависимости достоверности результатов идентификации от значений параметров $I_{отс}$ и ΔH . Для этого с помощью программы POISF проводилось два цикла исследований:

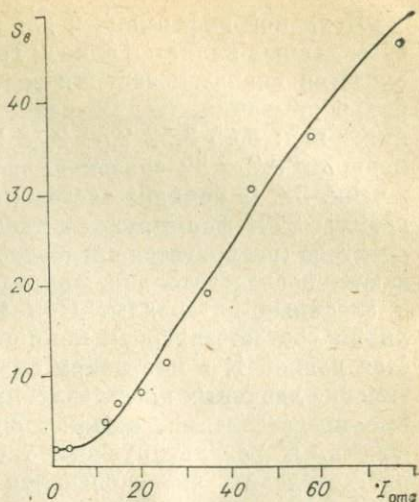
1. При постоянной величине окна $\Delta H = 4$ осуществляется анализ со значениями $I_{отс}$, изменяемыми от 1 до 76. В результаты анализа (список вероятных фаз) включались стандарты с величинами $R_m \geq 70$.

2. При постоянной величине $I_{отс} = 4$ осуществлялся анализ со значениями ΔH , равными 1, 2, 4, 6, 8, 10.

При исследованиях использовалась представительная база данных, включающая около 2300 стандартных спектров (минералы).

Варьирование $I_{отс}$ при постоянной величине окна позволило установить, что достоверность идентификации существенно зави-

Рис. 6.3. Зависимость числа вероятных стандартов S_b при $R_m \geq 70$ от значения $I_{отс}$ для обр. 16 (табл. 6.1). Точки нанесены для $I_{отс} = 1, 4, 11, 14, 19, 25, 33, 43, 57, 76$.



сит от степени учета слабых линий спектров стандартов: вплоть до $I = 4$. На рис. 6.3 приведена зависимость числа стандартов, попадающих в список вероятных фаз, для обр. 16 (корунд — бадделейт). Две истинные фазы выделяются при $I_{отс} = 4$; список вероятных стандартов S_b содержит только две данные фазы, отделенные

от остальных (от фона) областью резкого спада I кривой идентификации. При $I_{отс} = 10-20$ выделяется 5—8 фаз. В этом случае сравнительно несложно определить две истинные фазы из найденных 5—8 вероятных, однако можно исключить дополнительный анализ путем учета линий малой интенсивности. Учет только сильных линий стандартных спектров с $I \geq 30$ приводит к большому числу случайных совпадений. В список вероятных фаз попадает 20—50 спектров стандартов с $R_m \geq 70$ (см. рис. 6.3), причем у 10—20 стандартов значения R_m очень велики и близки к 95—100.

Варьирование величины окна при постоянной величине $I_{отс}$ позволило установить, что наилучшие результаты достигаются при значениях $\Delta H \sim 4-6$ единиц. При узком окне $\Delta H \approx 1-2$, т. е. при очень жестких требованиях на соответствие положений линий стандартных и исследуемого спектров (в смысле неравенства (4.11)), могут выпасть из рассмотрения некоторые линии исследуемого спектра, сдвинутые относительно соответствующих линий стандартов. Причина сдвига заключается не только в наличии статистической ошибки измерений, но и в наложении в спектре образца линий разных фаз, мешающем достаточно точно определять положения отдельных пиков. Широкое окно $\Delta H = 8$ сильно снижает селективную способность фильтра (4.11), и в список вероятных фаз попадают стандарты, случайно совпадающие с исследуемым спектром.

6.3. ИССЛЕДОВАНИЕ БАЗЫ ДАННЫХ

Перейдем к изучению базы данных с помощью рассмотренного в разделе 5.7 метода сжатия размерности исходного векторного пространства. Исследования проводились с базой данных, содержащей около 2300 минералов.

Исходной информацией для программы структурирования STR базы данных является список границ интервалов, на которые разбивается диапазон задания спектров D . Границы задаются в порядке убывания, и для размерности (например, $n' = 8$) список имеет вид: 5,85 3,52 2,82 2,38 2,04 1,77 1,55 1,31 0,98. Для размерности $n' = 32$ список включает все границы, перечисленные в табл. 5.1 (в порядке следования от 5,85 до 0,98). В начале программы STR формируются двоичные представления стандартных спектров (реализуется преобразование (5.84)). Спектры просматриваются последовательно, и определяется попадание линий спектров в заданные интервалы. При наличии в интервале каких-либо линий соответствующая компонента двоичного вектора принимается равной 1, в противном случае — 0. Затем производится сортировка двоичных представлений: все спектры, имеющие одинаковые представления, собираются в одну группу и выводятся на печать. В результирующей таблице указываются вид двоичного представления, число спектров и перечисляются все отобранные в данную группу спектры.

Вид двоичного представления приводится в шестнадцатиричном коде. Преобразование от шестнадцатиричного кода к двоичному осуществляется в соответствии с обычными правилами:

0 = 0000	4 = 0100	8 = 1000	C = 1100 (число 12)
1 = 0001	5 = 0101	9 = 1001	D = 1101 (число 13)
2 = 0010	6 = 0110	A = 1010 (число 10)	E = 1110 (число 14)
3 = 0011	7 = 0111	B = 1011 (число 11)	F = 1111 (число 15).

Таким образом, восьмиразрядное число $FF_{16} = 1111\ 1111_2$ указывает на то, что спектр имеет линии во всех интервалах разбиения от 5,85 до 0,98 Å, а число $9C_{16} = 1001\ 1100$ — на то, что спектр имеет линии в интервалах 5,85—3,52, 2,38—2,04, 2,04—1,77, 1,77—1,55 Å.

После сделанных выше предварительных замечаний рассмотрим результаты, полученные с помощью программы STR.

Структура базы данных, полученная при малой размерности $n' = 8$, является недостаточно детализированной; еще нельзя выделить кластеры, содержащие родственные соединения. Однако результаты этого структурирования позволяют изучить такую интересную характеристику, как обрыв диапазона задания стандартных спектров со стороны малых значений d . При $n' = 8$ спектры базы данных разбиваются на следующие группы:

1. Кластер FF (824 спектра) и близкие к нему 33 кластера вида DF, 7F, BF, F7 и т. д. (вместе объединяют 292 спектра); спектры, попадающие в эти кластеры, имеют линии во всех или почти во всех интервалах разбиения D и обязательно хотя бы одну в последнем интервале 1,31—0,98 Å (т. е. диапазон задания данных спектров простирается со стороны малых d примерно до 0,98 Å).

2. Кластер FE (512 спектров) и близкие к нему 14 кластеров

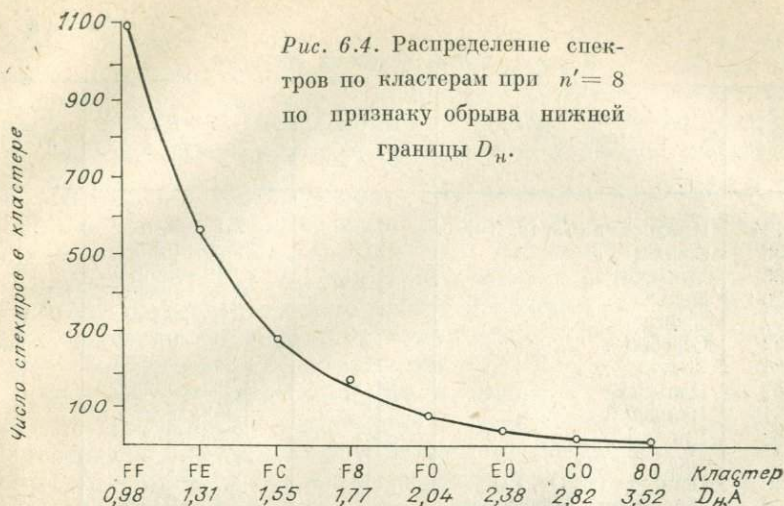


Рис. 6.4. Распределение спектров по кластерам при $n'=8$ по признаку обрыва нижней границы D_n .

(всего 45 спектров); нижняя граница диапазона задания спектров $D_n > 1,31 \text{ \AA}$.

3. Кластер FC (272 спектра) и близкие к нему 5 кластеров (всего 14 спектров); $D_n > 1,55 \text{ \AA}$.

4. Кластер F8 (184 спектра) и близкие к нему 3 кластера (всего 6 спектров); $D_n > 1,77 \text{ \AA}$.

5. Кластер F0 (57 спектров); $D_n > 2,04 \text{ \AA}$.

6. Кластер E0 (36 спектров); $D_n > 2,38 \text{ \AA}$.

7. Кластер C0 (4 спектра); $D_n > 2,82 \text{ \AA}$.

8. Кластер 80 (1 спектр); $D_n > 3,52 \text{ \AA}$.

Приведенное выше распределение спектров дано на рис. 6.4.

В базе данных обнаружено 30—40 стандартов со спектрами, которые условно можно назвать короткими. Как правило, это спектры некоторых чистых элементов кубической сингонии, имеющие малое число линий, сосредоточенных в узком диапазоне. К ним можно также отнести спектры, попадающие в кластеры C0 и 80. Поскольку короткие спектры содержат малое число линий, для них велика вероятность случайного хорошего совпадения с различными исследуемыми спектрами. Данные спектры могут часто попадать в результирующий список вероятных фаз и иметь при этом высокие значения критерия соответствия, т. е. они играют роль помехи в процедуре идентификации и поэтому их желательно знать что называется «по именам» (табл. 6.3).

Обрыв диапазона задания спектров со стороны больших d (верхняя граница D_s) не является ярко выраженным. Только 158 спектров имеют $D_s < 3,52 \text{ \AA}$, 17 — $D_s < 2,82 \text{ \AA}$ и 8 — $D_s <$

Характеристика коротких спектров

Номер карточки	Название	Двойное представление при $n' = 8$	Диапазон задания спектра, Å	Число линий
2-1194	Иридосмин (Os, Ir, Ru)	1F	2,13—0,84	12
4-686	Свинец	27	2,855—0,8251	13
4-835	Бунзенит	13	2,410—0,8040	10
4-836	Медь	19	2,088—0,8083	8
23-297	Тэнит	19	2,08—0,900	6
4-783	Серебро	1B	2,359—0,8341	9
4-784	Золото	1B	2,355—0,8325	9
4-802	Платина	1B	2,265—0,8008	9
5-681	Палладий	1B	2,246—0,8697	8
6-598	Иридий	1B	2,22—0,7838	9
11-65	Карлсбергит	1B	2,394—0,7979	10
17-213	Осмирид (Ir, Os, Pt)	1B	2,23—1,12	5
19-268	Лонсдэлит (модификация графита)	1B	2,19—0,820	10
4-829	Периклаз	33	2,431—0,8109	10
6-615	Вюстит	33	2,49—0,9631	8
6-642	Осборнит	33	2,44—0,816	10
7-230	Манганозит	33	2,568—0,7857	11
4-850	Никель	OD	2,034—0,7880	8
6-675	Алмаз	11	2,06—0,8182	5
6-696	Железо (α -фаза)	OB	2,0268—0,8275	6
8-293	Тантэксенит	70	3,12—2,15	5
15-463	Борнхардит	3A	2,7—1,42	7
15-464	Фребольдит	1D	2,4—0,983	7
18-1634	Рефикит	80	11,2—4,28	7

$< 2,38 \text{ \AA}$ или $D_s < 2,04 \text{ \AA}$, причем все восемь последних спектров относятся к коротким. Необходимо также учитывать, что некоторые спектры с $D_s < 3,52 \text{ \AA}$ или с $D_s < 2,82 \text{ \AA}$ имеют, возможно, линии с $d > 5,85 \text{ \AA}$, которые в настоящем исследовании не учитывались.

Рассмотрим подробнее результаты, касающиеся обрыва диапазона D со стороны малых значений d . Этот обрыв может носить закономерный характер, если соответствующее соединение действительно не имеет линий, например, в области $d < 1,31 \text{ \AA}$ или $d < 1,55 \text{ \AA}$ и т. д. Однако достаточно часто отсутствие линий в рассматриваемых областях связано с тем, что спектр в них просто не снимался, т. е. данные спектры содержат ошибку, обычно называемую «усечение диапазона съемки». Ее можно рассматривать как систематическую; по-видимому, она присутствует даже в некоторых спектрах с представлениями FF, FE и наверняка — в FC, F8 и т. д. Влияние систематической ошибки на результаты идентификации тем значительнее, чем более существенные линии стан-

дарту пропущены. Необходимо также учитывать, что информация в базе данных рассматривается как стандартная и, следовательно, она должна быть достаточно полной. Разумеется, нельзя получить все множество линий при $d \rightarrow 0$, однако одним из минимальных требований к стандартным спектрам является регистрация их, хотя бы до 1 \AA на излучении CuK_α (желательно до $0,6-0,8 \text{ \AA}$). Таким образом, мы приходим к заключению, что спектры, выделенные в кластеры FC, F8, F0, E0, C0, 80, являются подозрительными на наличие систематической ошибки; они должны быть дополнительно исследованы и многие заменены на более корректные. Отметим, что таких подозрительных спектров в совокупности минералов содержится около 25%.

Наличие систематической ошибки приводит к тому, что спектры родственных соединений часто представлены в базе данных в неоднозначных условиях съемки; следовательно, им соответствуют различные двоичные представления, что мешает их закономерному объединению в кластер в рамках рассмотренной в начале этого раздела автоматической процедуры. Это положение будет проиллюстрировано ниже многочисленными примерами, полученными при структурировании базы данных с размерностями $n' = 16$ и 32 .

Приведем характерный пример систематической ошибки «усечение диапазона съемки» в спектрах минерала рихтерита. Этот минерал имеет сложный состав и структуру, наиболее простую формулу, отвечающую его синтетической модификации с R_{JCPDS} 20-982 — $\text{Na}_2\text{CaMg}_5\text{Si}_8\text{O}_{22}(\text{OH})_2$. В выборке по минералам содержатся десять спектров для различных модификаций рихтерита, различающихся присутствием K, Fe, Al, Cr, Ti и F; номера соответствующих карточек перечислены в табл. 6.4 в сопоставлении с двоичными представлениями этих спектров при $n' = 16$. Все модификации рихтерита относятся к моноклинной сингонии, причем разброс параметров элементарной ячейки сравнительно невелик

$(9,677 \text{ \AA} \leq a \leq 9,900 \text{ \AA};$
 $17,83 \text{ \AA} \leq b \leq 17,985 \text{ \AA};$
 $5,259 \text{ \AA} \leq c \leq 5,289 \text{ \AA};$
 $102^\circ 57' \leq \beta \leq 104^\circ 24')$,
 вследствие чего спектры очень похожи. Однако (см. табл. 6.4) граница D_n для спектров рихтерита колеблется от $2,026$ до $0,979 \text{ \AA}$ (при почти неизменном D_s

Таблица 6.4

Некоторые характеристики спектров минерала рихтерита

Номер карточки	Двоичное представление при $n' = 16$	Диапазон задания спектра, \AA	Число линий
20-982		9,03—2,005	26
23-665	FF80	9,01—1,957	23
23-667		8,98—2,026	23
23-310	FFCO	8,93—1,890	23
10-428		8,96—1,609	39
19-1063	FFFO	8,61—1,650	29
14-161	FFDO	8,96—1,571	38
20-484	FFDC	8,85—1,350	40
10-456	FFFE	8,55—1,173	39
17-750	FFFF	8,91—0,979	69

Группа шпинели

№ п/п	Номер карточки	Название	Двоичное представление при $n' = 16$	№ п/п	Номер карточки	Название	Двоичное представление при $n' = 16$
1	3-873	Хромит	953E	6	19-629	Магнетит	953E
2	10-319	Якобсит		7	21-540	Шпинель (Fe^{3+})	
3	10-467	Франклинит		8	10-325	Треворит	95BF
4	17-464	Магнезио-феррит (низкий)			9	10-351	
5	17-465	Магнезио-феррит (высокий)		10	15-122	Коулсонит	
				11	22-1012	Франклинит	
				12	23-1119	Треворит (Fe^{2+})	

порядка 9 \AA), что приводит к пропуску от двух третей до почти половины линий.

При размерности $n' = 16$ из базы данных уже можно выделить кластеры, содержащие сходные спектры. Некоторые из них представлены в табл. 6.5—6.9.

В табл. 6.5 собраны результаты для группы шпинели, содержащей около 20 минеральных видов и разновидностей. Эти соединения относятся к кубической сингонии (пространственная группа $Fd\bar{3}m$); значение параметра решетки a лежит в пределах $8,0848—8,538 \text{ \AA}$. Около 80% спектров минералов этой группы определены в диапазоне $0,8—4,9 \text{ \AA}$. Большинство из этих спектров, полученных в адекватных диапазонах съемки, попадают в два кластера с близкими двоичными представлениями 953E (7 спектров) и 95BF (5 спектров). При размерности $n' = 32$ данные кластеры почти полностью распадаются, т. е. спектры шпинелей становятся раз-

Таблица 6.6

Группа соединений тетрагональной сингонии, сходных с рутилом

№ п/п	Номер карточки	Название	Двоичное представление при $n' = 16$	Диапазон задания спектра, \AA	Число линий	S. G.	a	c
1	6-290	Селлаит	273F	3,27—0,8486	36	$P4_2/mnt$	4,623	3,052
2	16-934	Ильменорутил		3,28—1,038	19	—	4,635	2,986
3	17-543	Стрюверит		3,28—1,038	20	—	4,645	2,999
4	21-1276	Рутил		3,25—0,7877	36	$P4_2/mnt$	4,593	2,959
5	11-396	Ильменорутил	373F	3,23—0,8234	31	$P4_2/mnt$	4,606	2,982
6	11-397	Ильменорутил	F779	5,90—0,8270	25	—	4,623	5,976

Группа хлорита

№ п/п	Номер карточки	Название	Двойное представление при $n' = 16$	№ п/п	Номер карточки	Название	Двойное представление при $n' = 16$	
1	7-76	Ришидолит	DFFE	9	10-183	Пеннин	DFFC	
2	7-77	Шериданит		10	12-185	Кочубейт		
3	7-78	Тюриянгит		11	16-362	Хлорит 1а		
4	7-160	Кочубейт		DEFE	12	7-166	Дафнит	DEFE
5	7-165	Грохаунт			13	19-749	Клинохлор	
6	7-171	Диабантин						
7	12-242	Лейхтенбергит						
8	22-712	Нимит						

личными — за исключением почти тождественных спектров 10-325, 10-351, которые и при $n' = 32$ образуют кластер из двух членов.

В табл. 6.6 сопоставлены спектры соединений тетрагональной сингонии, сходные с рутилом. Первые пять спектров образуют кластер с близкими представлениями 273F и 373F. В шестом спектре обнаружены линии с $d > 3,3 \text{ \AA}$; спектр проиндифицирован при значении $c = 5,976 \text{ \AA}$, вдвое большем, чем у первых пяти спектров. Кроме того, в спектре 11-397 пропущены линии, присутствующие в остальных спектрах. При $n' = 32$ в рассматриваемом кластере остаются только два члена — 16-934 и 17-543.

В табл. 6.7 даны результаты для группы хлорита, объединяющей около 20 минеральных разновидностей. Соединения этой группы относятся в основном к моноклинной сингонии; разброс значений параметров решетки у них сравнительно невелик. Верхняя граница диапазона [съемки составляет порядка 14 \AA , нижняя варьирует в пределах $1-1,4 \text{ \AA}$. Большинство спектров хлоритов имеют близкие представления DFFE, DEFE и DFFC. Отличие спектров соединений № 9—11 от спектров соединений № 1—8 связано с усечением диапазона съемки; для группы DFFE нижняя граница D_n составляет около $1,2 \text{ \AA}$, а для DFFC — около $1,3-1,4 \text{ \AA}$. При размерности $n' = 32$ в кластер объединяются только спектры 7-77, 7-165, 7-171.

В табл. 6.8 и 6.9 представлены несколько групп соединений гексагональной, тетрагональной и моноклинной сингоний со сходными спектрами. Очевидно, что родственные соединения № 1—4 (см. табл. 6.8) должны быть объединены в один кластер; аналогичное замечание следует сделать относительно соединений № 7—14 (см. табл. 6.8) и № 1—5 (см. табл. 6.9). Некоторые различия в двойчных представлениях спектров, входящих в эти группы, связаны главным образом с различием диапазонов съемки. При переходе к размерности $n' = 32$ перечисленные в табл. 6.8 и 6.9 кластеры распадаются, кроме группы, объединяющей почти тождественные спектры 4-843 и 6-662.

Некоторые группы соединений гексагональной и тетрагональной сингонии со сходными спектрами

№ п/п	Номер карточки	Химическая формула	Двоичное представление при $n' = 16$	Диапазон задания спектра, Å	Число линий	Сингония	S. G.	a	c
1	8-449	ZnCO ₃	4BEF	3,55—0,837	39	Гекс.	$R\bar{3}c$	4,653	15,028
2	11-692	CoCO ₃		3,551—0,912	32	»	$R\bar{3}c$	4,659	14,957
3	8-456	CdCO ₃	5F5F	3,78—0,881	38	Гекс.	$R\bar{3}c$	4,930	16,27
4	8-479	MgCO ₃	OFAF	2,742—0,798	16	Гекс.	$B\bar{3}c$	4,633	15,015
5	4-843	BeO	000B	2,337—0,8179	18	Гекс.	$P6_3mc$	2,6979	4,380
6	6-662	Os		2,37—0,8116	18	»	$P6_3/mmc$	2,7341	4,3197
7	8-440	(Th, U)(SiO ₄) _{1-x} (OH) _{4x}		4,69—0,791	36	Тетр.	—	7,068	6,260
8	8-458	CaCrO ₄		4,75—0,905	34	»	$I4_1/amd$	7,242	6,290
9	11-419	ThSiO ₄	DEFF	4,72—0,986	39	»	$I4_1/amd$	7,132	6,322
10	13-158	Ca ₂ BAsO ₄ (OH) ₄		5,02—1,108	39	»	$I\bar{4}$	7,095	6,190
11	17-341	YVO ₄		4,711—1,016	35	»	$I4_1/amd$	7,119	6,290
12	9-377	YPO ₄		4,54—0,853	39	Тетр.	$I4_1/amd$	6,885	5,982
13	11-254	(Y, Er)PO ₄	AFFF	4,55—0,863	39	»	$I4_1/amd$	6,904	6,035
14	11-420	USiO ₄		4,64—0,880	39	»	$I4_1/amd$	6,979	6,252

Подведем итоги рассмотренного выше исследования структуры базы данных.

1. Порядка 15 — 20% карточек минералов могут быть отнесены к группам, объединяющим соединения со сходными спектрами. Основные причины близости спектров некоторой группы соединений заключаются в следующем: соединения имеют сходную структуру (как правило, совпадают их сингонии и пространственные группы); параметры кристаллической решетки колеблются в небольших пределах (линейные—0,1—0,4 Å, угловые—1—2°); спектры получены примерно в одинаковых диапазонах.

2. Около 3—4% карточек минералов образуют группы, содержащие по 2—3 почти тождественных спектра, которые трудно-различимы рентгеновским методом.

3. В базе данных присутствует систематическая ошибка, связанная с усечением диапазона получения многих спектров. Ее влияние затрудняет выявление сходных спектров.

4. Геологическая классификация минералов (амфиболы, цеолиты, полевые шпаты, гранаты, апатиты, хлориты, слюды, шпинели и т. д.) только отчасти отвечает кластерной структуре базы данных. Отсутствие полного соответствия объясняется объективной

Таблица 6.9

Некоторые группы соединений моноклинной сингонии со сходными спектрами

№ п/п	Номер карточки	Химическая формула	Двойное представление при $n' = 16$	Диапазон задания спектра, Å	Число линий	Сингония	S. G.	a	b	c	β
1	11-128	(Zn, Fe, Ca, Mn)WO ₄	D7FF	5,72—0,794	35	Мон.	P2/c	4,680	5,712	4,933	90°18'
2	11-591	(Fe, Mn)WO ₄		5,68—0,904	36	»	P2/c	4,772	5,708	4,976	90°24'
3	21-436	FeWO ₄		5,72—1,098	40	»	P2/c	4,730	5,703	4,952	90°
4	13-434	MnWO ₄	D7F8	5,76—1,475	39	Мон.	P2/c	4,829	5,759	4,998	91°16'
5	15-774	ZnWO ₄		5,72—1,446	39	»	P2/c	4,691	5,720	4,925	89°36'
6	11-140	(Co, Fe)As ₂	5EFF	3,932—0,967	39	Мон.	P2/c	5,232	5,952	2,957	90°
7	11-699	FeAs ₂		3,97—1,082	39	Ромб.	Pnnt	5,300	5,983	2,882	90°
8	14-217	Al ₂ PO ₄ VO ₄ ·6H ₂ O	B080	14,9—1,9	13	Мон.	P2/m	11,4	14,9	9,2	101°
9	14-219	Al ₂ PO ₄ VO ₄ ·8H ₂ O		15,8—1,9	12	»	P2/m	11,4	15,8	9,2	101°

и субъективной причинами. Объективная связана с тем, что в некоторые минералогические группы входят соединения разных сингоний (например, цеолиты — кубическая, гексагональная, тетрагональная, ромбическая, моноклинные сингонии; полевые шпаты — моноклинные и триклинные сингонии) и с большим разбросом значений параметров решетки (например, гранаты — кубическая и тетрагональная сингонии, $11,5 \text{ \AA} \leq a \leq 12,5 \text{ \AA}$). Субъективная причина связана с наличием отмеченной выше систематической ошибки.

Таким образом, если даже устранить систематическую ошибку, нельзя ожидать полного соответствия между геологической классификацией минералов и результатами их кластеризации по сходству рентгенодифракционных спектров. Лишь те группы минералов, структурные свойства которых достаточно четко локализованы (т. е. удовлетворяют ограничениям, указанным выше в пункте 1), будут образовывать кластеры в фазовом пространстве спектральных признаков.

ЛИТЕРАТУРА

1. Азаров Л., Бургер М. Метод порошка в рентгенографии.— М.: ИЛ, 1965.— 364 с.
2. Блэкман М. Проектирование систем реального времени.— М.: Мир, 1977.— 346 с.
3. Бокий Г. Б., Порай-Кошиц М. А. Рентгеноструктурный анализ. Т. 1.— М.: Изд-во МГУ, 1964.— 490 с.
4. Болдырев А. К., Михеев В. И., Ковалев Г. А. и др. Рентгометрический определитель минералов.— Зап. Ленингр. горн. ин-та, 1938, т. XI, вып. 2.
5. Бородкина М. М., Спектор Э. Н. Рентгенографический анализ текстуры металлов и сплавов.— М.: Metallurgia, 1981.— 272 с.
6. Бурова Е. М., Жидков Н. П., Зубенко В. В. и др. Алгоритмизация процесса обработки данных рентгеновского фазового анализа.— Докл. АН СССР, 1977, т. 232, № 5, с. 1066—1068.
7. Бурова Е. М., Жидков Н. П., Зубенко В. В. и др. Архитектура и наполнение системы программ ФАЗАН-78.— В кн.: Математические вопросы структурного анализа и алгоритмы машинных экспериментов в органической химии. М.: Изд-во МГУ, 1979, с. 28—37.
8. Бурова Е. М., Жидков Н. П., Зубенко В. В. и др. Практика определения фазового состава поликристаллических смесей с помощью системы программ ФАЗАН.— В кн.: Математические вопросы структурного анализа. М.: Изд-во МГУ, 1980, с. 55—76.
9. Бурова Е. М., Жидков Н. П., Щедрин Б. М. Задача определения состава поликристаллической смеси по ее рентгенограмме с помощью набора соответствующих рентгенограмм.— В кн.: Математические вопросы структурного анализа. М.: Изд-во МГУ, 1978, с. 53—67.
10. Бурова Е. М., Зильберман А. Г., Зубенко В. В. и др. Информационно-поисковая система для рентгеновского анализа.— Кристаллография, 1977, т. 22, № 6, с. 1182—1190.
11. Бурова Е. М., Уманский М. М. О работе автоматизированной информационно-поисковой системы для рентгеновского качественного анализа.— В кн.: Аппаратура и методы рентгеновского анализа.— Л.: Машиностроение, 1979, № 22, с. 3—10.
12. Бурова Е. М., Щедрин Б. М. О некорректности математической постановки задачи определения фазового состава поликристаллической смеси.— В кн.: Математические вопросы структурного анализа. М.: Изд-во МГУ, 1981, с. 39—44.
13. Васильев Е. К., Васильева Н. П. Рентгенографический определитель оловосодержащих минералов.— Новосибирск: Наука, 1977.— 78 с.
14. Васильев Е. К., Васильева Н. П. Рентгенографический определитель карбонатов.— Новосибирск: Наука, 1980.— 143 с.
15. Васильев Е. К., Васильева Н. П., Ущановская З. Ф. О требованиях к данным, представляемым в рентгенографические определители минера-

- лов.— В кн.: Физические и химические методы исследования горных пород и минералов. Иркутск: Вост.-Сиб. филиал СО АН СССР, 1977, с. 57—65.
16. Васильев Е. К., Кашаева Г. М., Ущановская З. Ф. Рентгенометрический определитель минералов класса фосфатов.— М.: Наука, 1974.— 207 с.
 17. Васильев Е. К., Лиопо В. А. Некоторые вопросы рентгенографического фазового анализа.— Минерал. сб., 1978, вып. 2, № 32, с. 27—33.
 18. Васильев Е. К., Нахмансон М. С. Контроль достоверности результатов качественного рентгенофазового анализа.— Л.: ЛНПО «Буревестник», 1982.— 19 с. (Рукопись деп. в ВИНТИ 13.5.82. № 5376—82. Деп.).
 19. Вишняков Я. Д. Современные методы исследования структуры деформированных кристаллов.— М.: Metallургия, 1978.— 479 с.
 20. Гамарник М. Я. Аналитический метод юстировки гониометра рентгеновского дифрактометра. Препринт ИГФМ-76, 1976.— 56 с.
 21. Гинь А. Рентгенография кристаллов.— М.: Физматгиз, 1961.— 604 с.
 22. Горелик С. С., Расторгуев Л. Н., Скаков Ю. А. Рентгенографический и электронно-микроскопический анализ. Приложения.— М.: Metallургия, 1970.— 107 с.
 23. Гребенщиков Р. Г., Франк-Каменецкий В. А. Щелевой коллиматор рентгеновской камеры ВРС-3 для регистрации дифракционных отражений под малыми углами.— В кн.: Рентгенография минерального сырья. М.: Госгеолтехиздат, 1963, № 3, с. 159—162.
 24. Дерендяев Б. Г., Коптюг В. А., Лебедев К. С. и др. Машинная информационно-поисковая система на базе каталога полных масс-спектров.— Автометрия, 1979, № 4, с. 3—13.
 25. Джуре П., Айзенауэр Т. Распознавание образов в химии.— М.: Мир, 1977.— 230 с.
 26. Долянский В. Г., Ногаллер В. А. Машинное распознавание химических элементов по Оже-спектрограммам.— Электронная техника, 1976, сер. 7, вып. 5, с. 91—95.
 27. Дробышев Ю. П., Нигматуллин Р. С., Одеянко Б. Н. Комплекс программ для обработки химической информации на БЭСМ-6.— В кн.: IV Международная конференция по применению ЭВМ в химии и химическом образовании. (Тезисы докладов). Новосибирск, ИОХ СО АН СССР, 1978, доклад № 2—28.
 28. Евграфов А. А., Нахмансон М. С., Черный Ю. А. Проблема идентификации фаз для качественного анализа поликристаллических смесей.— В кн.: Аппаратура и методы рентгеновского анализа. Л.: Машиностроение, 1979, № 22, с. 42—46.
 29. Зевин Л. С. Об оптимальных соотношениях между параметрами регистрации дифрактограмм поликристаллов.— В кн.: Аппаратура и методы рентгеновского анализа. Л.: Машиностроение, 1975, № 17, с. 49—55.
 30. Зевин Л. С. Влияние щелей Соллера на интенсивность рентгеновских отражений в дифрактометре.— В кн.: Аппаратура и методы рентгеновского анализа. Л.: Машиностроение, 1977, № 18, с. 51—57.
 31. Зевин Л. С., Завьялова Л. Л. Количественный рентгенографический фазовый анализ.— М.: Недра, 1974.— 184 с.
 32. Зевин Л. С., Уманский М. М., Хейкер Д. М. Геометрические aberrации и оптимальные условия дифрактометрических измерений поликристаллов.— Кристаллография, 1963, т. 8, № 4, с. 662—673.
 33. Зевин Л. С., Хейкер Д. М. Рентгеновские методы исследования строительных материалов.— М.: Стройиздат, 1965.— 362 с.
 34. Каменцев И. Е. Использование кварца как эталона при рентгеновских исследованиях и точность определения параметров его элементарной ячейки.— В кн.: Рентгенография минерального сырья. Вып. 7. М., ВИМС, 1970, с. 108—115.
 35. Кан Р., Дермер О. Введение в химическую номенклатуру.— М.: Химия, 1983.— 224 с.
 36. Кане И. А., Консон Е. Д., Нахмансон М. С. и др. Системы автоматической идентификации порошковых дифракционных спектров.— В кн.: Аппаратура и методы рентгеновского анализа. Л.: Машиностроение, 1984, № 32, с. 62—78.

37. **Китайгородский А. И.** Рентгеноструктурный анализ мелкокристаллических и аморфных тел.— М.—Л.: Гостеортехиздат, 1952.— 588 с.
38. **Ковба Л. М., Трунов В. К.** Рентгенофазовый анализ. Изд. 2-е.— М.: Изд-во МГУ, 1976.— 232 с.
39. **Кондратьева В. В.** Рентгенометрический определитель боратов.— Л.: Недра, 1969.— 248 с.
40. **Консон Е. Д., Нахмансон М. С.** Байесовский подход к задаче качественного рентгенофазового анализа.— Автометрия, 1981, № 6, с. 39—44.
41. **Консон Е. Д., Нахмансон М. С., Степанова Н. В.** Система качественного рентгенофазового анализа на основе алгоритмов линейного программирования.— В кн.: Аппаратура и методы рентгенового анализа. Л.: Машиностроение, 1983, № 31, с. 54—61.
42. **Кудряшова В. И.** Транскрипции названий новых минералов, сведения о которых опубликованы в 1980—1981 гг.— Зап. Всесоюз. минерал. о-ва, 1982, ч. III, вып. 3, с. 338—339.
43. **Лидин Р. А., Молочко В. А., Андреева Л. Л. и др.** Основы номенклатуры неорганических веществ.— М.: Химия, 1983.— 224 с.
44. **Лишсон Г., Стинл Г.** Интерпретация порошковых рентгенограмм.— М.: Мир, 1972.— 384 с.
45. **Ложникова О. Н., Яковлева С. Я.** Рентгенометрический справочник — определитель минералов, содержащих редкоземельные элементы.— М.: ОНТИ «Гиредмет», 1961.— 224 с.
46. **Миркин Л. И.** Справочник по рентгеноструктурному анализу поликристаллов.— М.: Физматгиз, 1961.— 683 с.
47. **Миркин Л. И.** Рентгеноструктурный анализ. Индексирование рентгенограмм.— М.: Наука, 1981.— 496 с.
48. **Митчел Р. С.** Название минералов.— М.: Мир, 1982.— 248 с.
49. **Михеев В. И.** Рентгенометрический определитель минералов.— М.: Гостеортехиздат, 1957.— 868 с.
50. **Михеев В. И., Дубинина В. Н.** Рентгенометрический определитель минералов. Ч. II.— Зап. Ленинград. горн. ин-та, т. XIII, вып. 1, 1939, с. 136.
51. **Михеев В. И., Сальдау Э. П.** Рентгенометрический определитель минералов. Т. 2.— Л.: Недра, 1960.— 116 с.
52. **Моисеев Н. П.** Математика ставит эксперимент.— М.: Наука, 1979.— 223 с.
53. **Нахмансон М. С.** Автоматизация исследований фазового состава вещества.— Управляющие системы и машины, 1984, № 2, с. 99—105.
54. **Нахмансон М. С., Черный Ю. А.** Система автоматического рентгенофазового анализа АРФА.— Л.: ЛНПО «Буревестник». 1981.— 63 с. (Рукопись деп. в ВИНТИ 6.11.81. № 497—82. Деп.).
55. **Нахмансон М. С., Черный Ю. А.** Применение системы АРФА к решению задач качественного рентгенофазового анализа. Л., ЛНПО «Буревестник». 1982.— 85 с. (Рукопись деп. в ВИНТИ 13.6.82. № 4150—82. Деп.).
56. **Нахмансон М. С., Черный Ю. А.** Использование системы АРФА и базы данных МИНЕРАЛЫ для определения фазового состава ряда смесей и продуктов геологического происхождения.— В кн.: IX Всесоюзное совещание по рентгенографии минерального сырья. (Тезисы докладов). Казань, 1983, с. 64—66.
57. **Нахмансон М. С., Шульмейстер В. М.** Возможные формы диалога в системах автоматизации научных исследований.— Управляющие системы и машины, 1980, № 5, с. 111—117.
58. **Номенклатурные правила ИЮПАК по химии.** Т. I, п/т. 1.— 287 с.; п/т. 2.— 660 с.; т. II, п/т. 1.— 508 с.; п/т. 2 — 896 с. Пер. с англ.— М., ВИНТИ, 1979.
59. **Петрова В., Ангелова В., Славов С.** Върху възможностите за използване на ЕИМ за сериен качествен фазов анализ.— Металлургия, 1978, т. 33, № 6, с. 12—15.
60. **Пиворюнас В.** Линейная дискриминативная функция для идентификации спектров.— В кн.: Статистические проблемы управления. Вильнюс, Ин-т математики и кибернетики АН ЛитССР, 1979, № 27, с. 72—90.

61. Пиворюнас В. Пакет прикладных программ для автоматизированного рентгенофазового анализа АРФА.— В кн.: Статистические проблемы управления. Вильнюс, Ин-т математики и кибернетики АН ЛитССР, 1981, № 50, с. 109—127.
62. Рентгенография в физическом металловедении/Под ред. Ю. А. Багаряцкого.— М.: Металлургия, 1961.— 368 с.
63. Рентгенография основных типов породообразующих минералов/Под ред. проф. В. А. Франк-Каменецкого.— Л.: Недра, 1983.— 359 с.
64. Руководство по рентгеновскому исследованию минералов/Под ред. В. А. Франк-Каменецкого.— Л.: Недра, 1975.— 399 с.
65. Русаков А. А. Рентгенография металлов.— М.: Атомиздат, 1977.— 480 с.
66. Сакман Г. Решение задач в системе человек — ЭВМ.— М.: Мир, 1973.— 351 с.
67. Сальни А. Л., Дриц В. А. Повышение точности измерения межплоскостных расстояний на дифрактограммах глинистых минералов.— Литол. и полезн. ископ., 1983, № 3, с. 125—133.
68. Сидоренко Г. А. Рентгенографический определитель урановых и урансодержащих минералов.— М.: Недра, 1960.— 416 с.
69. Сидоренко Г. А. Рентгенографический определитель минералов урана.— М.: Энергоиздат, 1981.— 192 с.
70. Смирнов В. И. Курс высшей математики. Т. II.— М.: Физматгиз, 1962.— 628 с.
71. Соколова М. Ф. Об оптимальных условиях регистрации дифрактограмм глинистых минералов.— В кн.: Рентгенография минерального сырья и реальное строение минералов. М.: ВИМС, 1978, с. 28—39.
72. Сухарев Ю. Н., Некрасов Ю. С. Автоматическая обработка и интерпретация масс-спектральной информации.— В кн.: IV Международная конференция по применению ЭВМ в химии и химическом образовании. (Тезисы докладов). Новосибирск, ИОХ СО АН СССР, 1978, доклад № 1—13.
73. Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач.— М.: Наука, 1979.— 285 с.
74. Ту Д., Гонсалес Р. Принципы распознавания образов.— М.: Мир, 1978.— 411 с.
75. Ульянов Г. П., Маслов А. Р., Дерендяев Б. Г. и др. ЭВМ и решение аналитических задач органической химии.— Вестн. АН СССР, 1978, № 3, с. 32—44.
76. Уманский М. М. Некоторые проблемы рентгеновского качественного фазового анализа.— Минералогический сб., 1979, № 33, вып. 1, с. 19—27.
77. Уманский М. М. Рентгеновский качественный анализ и сводки эталонов для него.— В кн.: Аппаратура и методы рентгеновского анализа. Л.: Машиностроение, 1979, № 22, с. 114—122.
78. Уманский М. М. Рентгеновский качественный фазовый анализ.— Заводская лаборатория, 1984, т. 50, № 5, с. 20—26.
79. Уманский Я. С. Рентгенография металлов и полупроводников.— М.: Металлургия, 1969.— 496 с.
80. Уманский Я. С., Скаков Ю. А., Иванов А. П. и др. Кристаллография, рентгенография и электронная микроскопия.— М.: Металлургия, 1982.— 632 с.
81. Утсал К. Р. О технике и методике исследования глинистых минералов рентгеновскими методами.— Уч. зап. Тартуского ун-та, 1971, т. IV, вып. 286, с. 3—50.
82. Федорченко А. А., Поддубный Н. П., Петров С. Е. Об одном алгоритме идентификации фаз неорганических соединений на ЭВМ.— В кн.: IV Международная конференция по применению ЭВМ в химии и химическом образовании. (Тезисы докладов). Новосибирск, ИОХ СО АН СССР, 1978, доклад № 2—35.
83. Фекличев В. Г. Диагностические спектры минералов.— М.: Недра, 1977.— 288 с.
84. Хейкер Д. М., Зевин Л. С. Рентгеновская дифрактометрия.— М.: Физматгиз, 1963.— 380 с.

85. Шиповалов Ю. В., Путилин Ю. М., Матвеев М. А. и др. К вопросу о применении ЭВМ в рентгенографии минерального сырья.— В кн.: Исследования в области химических и физических методов анализа минерального сырья. Вып. 3. Алма-Ата: КазНИИМС, 1973, с. 167—171.
86. An automatic qualitative phase analysis with a computerized X-ray powder diffractometer. Application report. Проспект фирмы Филипс, 1978.— 17 p.
87. APD-10 for automatic qualitative analysis. Application report. Проспект фирмы Филипс, 1978.
88. Bayliss P. X-ray diffraction powder data.— Amer. Mineral., 1976, v. 61, N 3—4, p. 334—336.
89. Blount A. M., Yasilion A. H. A new method of reducing preferred orientation in diffractometer samples.— Amer. Mineral., 1979, v. 64, N 7—8, p. 922—924.
90. Calvert L. D., Flippen—Anderson J. L., Hubbard C. R. e. a. Standards for the publication of powder patterns: The American Crystallographic Association Subcommittee's final report.— Accuracy in Powder Diffraction, NBS Special Publ, N 567, 1980, p. 513—537.
91. Dent Glasser L. S. Crystallography and its application.— Van Nostrand, N. Y.— L., 1977.— 224 p.
92. Dismore P. F. Computer searching of the JCPDS powder diffraction file.— Adv. X-ray Anal., 1977, v. 20, p. 113—123.
93. Donnay G. Suggestions for a quantitative evaluation of powder patterns.— Accuracy in Powder Diffraction, NBS Special Publ., N 567, 1980, p. 537—540.
94. Donnay G., Donnay J. D. H. The symmetry change in the high-temperature alcaly-feldspar series.— Amer. J. Sci., Bowen vol., pt. 1, 1952, p. 115—132.
95. Fiala J. Algebraic conception of powder diffraction identification system.— J. Phys. D: Appl. Phys., 1972, v. 5, N 10, p. 1874—1876.
96. Fiala J. Praskove difrakcni identifikacni systemy a jejich automatizace samocinnymi procitaci.— Ceskosl. Cas. fis., 1974, v. A24, N 3, p. 237—245.
97. Fiala J. Optimisation of powder diffraction identification.— J. Appl. Crystal., 1976, v. 9, N 6, p. 429—432.
98. Fiala J. Optimisierung der quantitativen rontgenographischen Phasenanalyse.— Kristal und Technik, 1977, Bd 12, N 5, S. 505—515.
99. Frevel L. K. Chemical analysis by powder diffraction.— Ind. Eng. Chem., Anal. Ed., 1944, v. 16, N 4, p. 209—218.
100. Frevel L. K. Computational aids for identifying crystalline phases by powder diffraction.— Anal. Chem., 1965, v. 37, N 4, p. 471—482.
101. Frevel L. K. Quantitative matching of powder diffraction patterns.— Adv. X-ray Anal., 1977, v. 20, p. 15—25.
102. Frevel L. K., Adams C. E. Quantitative comparison of powder diffraction patterns.— Anal. Chem., 1968, v. 40, N 8, p. 1335—1340.
103. Frevel L. K., Adams C. E., Ruhberg L. R. A fast search — match program for powder diffraction analysis.— J. Appl. Crystal., 1976, v. 9, N 3, p. 199—204.
104. Hanawalt J. D. Phase identification by X-ray powder diffraction evaluation of various techniques.— Adv. X-ray Anal., 1977, v. 20, p. 63—73.
105. Hanawalt J. D., Rinn H. W., Frevel L. K. Chemical analysis by X-ray diffraction patterns.— Ind. Eng. Chem., Anal. Ed., 1938, v. 10, p. 457—517.
106. Hare T. M., Russ J. C., Lanzo M. J. X-ray diffraction phase analysis using microcomputers.— Adv. X-ray Anal., 1982, v. 25, p. 237—243.
107. Heller S. R., Milne G. W. A., Feldmann R. J. A computer-based chemical information system.— Science, 1977, v. 195, N 4275, p. 253—259.
108. Huang T. C., Parrish W. A new computer algorithm for qualitative X-ray powder diffraction analysis.— Adv. X-ray Anal., 1982, v. 25, p. 213—219.
109. Hubbard C. R. Certification of silicon diffraction standard reference material 640a.— J. Appl. Crystal., 1983, v. 16, N 3, p. 285—288.

110. Hubbard C. R. New standard reference materials for X-ray powder diffraction.— *Adv. X-ray Anal.*, 1983, v. 26, p. 45—51.
111. Hubbard C. R., Mighel A., Stalick J. K. NBS — AIDS 80: A FORTRAN program to evaluate crystallographic data.— *Adv. X-ray Anal.* — 1981, v. 24, p. 99—109.
112. Hubbard C. R., Swanson H. E., Mauer F. A. A silicon powder diffraction standard reference material.— *J. Appl. Crystal.*, 1975, v. 8, N 1, p. 45—48.
113. IBM Series/1 field developed programs X-ray analysis automation products.— *Проспект фирмы Филипс*, 1981.— 25 p.
114. *International Tables for X-ray Crystallography*. Ed. Henry N. F. M., Lonsdale K. V. 1. 2nd edition.— Birmingham, England; Kynoch Press, 1965.— 558 p.
115. Jenkins R. A round robin test to evaluate computer search/match methods for qualitative powder diffractometry.— *Adv. X-ray Anal.*, 1977, v. 20, p. 125—137.
116. Jenkins R., Hahn Y., Pearlman S. The APD 3600, a new dimension in qualitative and quantitative X-ray powder diffractometry.— *Norelco Reporter*, 1979, v. 26, N 1, 15 p.
117. Jenkins R., Hahn Y., Pearlman S. e. a. A computer aided search-match system for qualitative powder diffractometry.— *Adv. X-ray Anal.*, 1980, v. 23, p. 279—285.
118. Jenkins R., Hom T., Willamizar C. e. a. Calibration of the diffractometer at low values of two theta.— *Adv. X-ray Anal.*, 1982, v. 25, p. 289—294.
119. Jenkins R., Hubbard C. R. A preliminary report of the design and results of the second round robin to evaluate search/match method for qualitative powder diffractometry.— *Adv. X-ray Anal.*, 1979, v. 22, p. 133—142.
120. Jobst B. A., Gobel H. E. IDENT — a versatile microfile — based system for interactive XRPD phase analysis.— *Adv. X-ray Anal.*, 1982, v. 25, p. 273—282.
121. Johnson G. G. Resolution of X-ray powder patterns.— *Laboratory system and Spectroscopy*, Chapter 3. N. Y., 1977, p. 45—57.
122. Johnson G. G. User Guide Data base and search program. JCPDS. Swarthmore, Pennsylvania, 1975.
123. Johnson G. G., Vand V. A computerized powder diffraction identification system.— *Ind. Eng. Chem.*, 1967, v. 59, N 8, p. 19—31.
124. Johnson G. G., Vand V. Computerized multiphase X-ray powder diffraction identification system.— *Adv. X-ray Anal.*, 1968, v. 11, p. 376—384.
125. Johnson G. G., Vand V. FORTRAN IV program (version 10) for the identification of multiphase powder diffraction patterns using the Joint Committee's powder diffraction data base.— ASTM, Philadelphia, 1969.
126. Kennard O., Hanawalt J. D., Wilson A. J. C. e. a. Powder Data.— *J. Appl. Crystal.*, 1971, v. 4, p. 81—86. (Русский текст в книге: Кристаллохимия и структурная минералогия. Л.: Наука, 1979, с. 118—128).
127. Klug H. P., Alexander L. E. X-ray diffraction procedures for polycrystalline and amorphous materials. 2nd edn., N. Y.: Wiley, 1974.— 966 p.
128. Lester K. M., Lipson H. The reliability of the interpretation of powder photographs.— *J. Appl. Crystal.*, 1970, v. 3, N 2, p. 92—93.
129. Lin Tian-Hui, Zhang Sai-Zhu, Chen Li-Jun e. a. An improved program for searching and matching of X-ray powder diffraction patterns.— *J. Appl. Crystal.*, 1983, v. 16, N 1, p. 150—154.
130. McCarthy G. J., Johnson G. G. Identification of multiphase unknowns by computer methods: role of chemical information, the quality of X-ray powder data and subfiles.— *Adv. X-ray Anal.*, 1979, v. 22, p. 109—130.
131. McMurdie H. F. Standard data for the identification of phases by X-ray diffraction.— *Adv. X-ray Anal.*, 1974, v. 7, p. 20—31.
132. Marquart R. S., Katsnelson I., Milne G. W. A. e. a. A search-match system for X-ray powder diffraction data.— *J. Appl. Crystal.*, 1979, v. 12, N 6, p. 629—634.
133. Mighel A. D., Stalick J. K. The NBS crystal data file.— *Acta Crystal.*, 1981, v. A37, august, Supplements, p. C—343.

134. Nichols M. C. A FORTRAN II program for the identification of X-ray powder diffraction patterns.— UCRL-70078, Lawrence Livermore Laboratory, 1966.
135. Nichols M. C., Johnson Q. The search — match problem.— *Adv. X-ray Anal.*, 1979, v. 23, p. 273—278.
136. O'Connor B. H., Baglian F. A semi-automated system for identifying crystalline materials with powder diffraction data.— *J. Appl. Crystal.*, 1976, v. 9, N 6, p. 419—423.
137. Parrish W., Ayers G. L., Huang T. C. A versatile microcomputer X-ray search-match system.— *Adv. X-ray Anal.*, 1982, v. 25, p. 221—229.
138. Parrish W., Mack M., Taylor J. Determination of apertures in the focusing plane of X-ray powder diffractometers.— *J. Sci. Instrum.*, 1966, v. 43, N 9, p. 623—628.
139. Petrov S., Kirov G. N. A method for the elimination of the preferred orientation of diffractometric samples.— *Докл. Болг. АН*, 1979, т. 32, № 9, с. 1267—1270.
140. Powder Diffraction File. Book Form. Inorganic Section. Sets 1—5, 6—10, 11—15, 16—18, 19—20, 21—22, 23—24, 25—26. JCPDS. Swarthmore, Pennsylvania, USA, 1960—1984.
141. Powder Diffraction File. Data Cards. Inorganic Section. Sets 1—34. JCPDS. Swarthmore, Pennsylvania, USA, 1948—1984.
142. Powder Diffraction File. Featuring Set 34. JCPDS-International Centre for Diffraction Data. Swarthmore, Pennsylvania, USA, 1984.— 32 p.
143. Powder Diffraction Data from the JCPDS Associateship at the NBS (Sets 1—26): Data Book; Search Manual (Alphabetical, Mineral Name, Organic formula, Hanawalt and Fink Sections). JCPDS. Swarthmore, Pennsylvania, USA, 1977.
144. Powder Diffraction File. Frequently Encountered Phases. (Search Manual, Inorganic Compounds, Alphabetical, Hanawalt, Fink Sections). JCPDS. Swarthmore, Pennsylvania, USA, 1977.
145. Powder Diffraction File. Alphabetical Index. Inorganic Phases. JCPDS. Swarthmore, Pennsylvania, USA, 1982.
146. Powder Diffraction File. Search Manual Hanawalt method. Inorganic. JCPDS. Swarthmore, Pennsylvania, USA, 1982.
147. Powder Diffraction File. Search Manual Fink method. Inorganic. JCPDS. Swarthmore, Pennsylvania, USA, 1982.
148. Schreiner W. N., Jenkins R. Automatically correcting for specimen displacement error during XRD search—match identification.— *Adv. X-ray Anal.*, 1982, v. 25, p. 231—236.
149. Schreiner W. N., Surdukowski C., Jenkins R. Qualitative phase analysis using an X-ray powder diffractometer.— *Adv. X-ray Anal.*, 1981, v. 24, p. 91—96.
150. Selected Powder Diffraction Data for Metals and Alloys: Data Book (Two volumes, sets 1—27), Search Manual (Alphabetical chemical name, Hanawalt and Fink numerical sections). JCPDS. Swarthmore, Pennsylvania, USA, 1978.
151. Selected Powder Diffraction Data for Minerals: Data Book (Sets 1—23), 1974; (Sets 24—29), 1980; Search Manual (Alphabetical chemical name, Hanawalt and Fink numerical sections, Alphabetical mineral name). JCPDS. Swarthmore, Pennsylvania, USA, 1980.
152. Smith D. K., Barrett C. S. Special Handling problems in X-ray diffractometry.— *Adv. X-ray Anal.*, 1979, v. 22, p. 1—12.
153. Smith G. S., Snyder R. L. F_N : a criterion for rating powder diffraction patterns and evaluating the reliability of powder pattern indexing.— *J. Appl. Crystal.*, 1979, v. 12, N 1, p. 60—65.
154. Smith S. T., Snyder R. L., Bhowell W. E. Minimisation of preferred orientation in powders by spray drying.— *Adv. X-ray Anal.*, 1979, v. 22, p. 77—87.
155. Snyder R. L. A Hanawalt type phase identification procedure for a mini-computer.— *Adv. X-ray Anal.*, 1981, v. 24, p. 83—90.

156. Standards for the publication of powder pattern data.— *Acta Crystal.*, 1981, v. A37, N 3, p. 443—444.
157. Standard X-ray diffraction powder patterns: Circular 539, v. 1—10; Monograph 25, sect. 1—20. US Dept. Commers, NBS, Washington, USA, 1953—1984.
158. Van Trump G., Hauff P. L. Computer identification techniques for crystalline compounds using the JCPDS powder diffraction file as data reference.— *Adv. X-ray Anal.*, 1977, v. 20, p. 103—112.
159. Vassamillet L. F., King H. W. Diffractometer technique.— *Handbook of X-rays*. Chapter 9, N. Y.: McGraw — Hill, 1967.
160. Visser J. W. Application of the results of phase identification by computer.— Conference on Applied Crystallography, Katowice, 1978, p. 107—111.
161. Waldo A. W. Identification of the copper ore minerals by means of the X-ray powder diffraction patterns.— *Amer. Mineral.*, 1935, v. 20, N 8, p. 575—597.
162. Weinke H. H., Kracher A., Kluger F. e. a. Programm zur analytischen Auswertung röntgenographischer Daten.— *Mikrochim. Acta*, 1975, Bd 1, H. 5—6, S. 449—456.
163. Wilson A. J. C. A note on peak displacement in X-ray diffractometry.— *Proc. Phys. Soc. (London)*, 1961, v. 78, pt. 2, N 500, p. 249—255.
164. Wilson A. J. C. Mathematical theory of X-ray powder diffractometry.— Endhoven, Netherlands: Philips Techn. Lib., 1963.— 128 p.
165. Wilson A. J. C. Accuracy in methods of lattice — parameter measurement.— Accuracy in powder diffraction. NBS Special Publ., 1980, N 567, p. 325—351.
166. Wolff P. M. de. A simplified criterion for reliability of a powder patterns indexing.— *J. Appl. Crystal.*, 1968, v. 1, N 2, p. 108—113.
167. X-ray diffractometry system.— Проспект фирмы Филипс, 1978.— 25 p.
168. Zevin L. S. Determination of the variable instrumental factors in the X-ray powder diffractometer.— *J. Phys. E: Sci. Inatrum.*, 1981, v. 14, N 6, p. 725—730.
169. Zwell L., Danko A. W. Application of diffraction method to quantitative chemical analysis.— *Appl. Spectr. Rev.*, 1975, v. 9, N 2, p. 167—221.

ПРИЛОЖЕНИЯ

В предлагаемых стандартных форматах приводимые данные должны быть предельно полными, особенно пункты, выделенные прописными буквами и жирным шрифтом. Заполнение остальных пунктов формата крайне желательно.

Основные требования к заполнению формата:

1) таблица, содержащая дифракционный спектр, должна быть по возможности полной и включать как слабые, так и сильные линии до углов $2\theta = 100^\circ$, излучение CuK ; спектры с малым числом линий надо регистрировать до предельных углов 2θ , допускаемых методикой эксперимента;

2) значения $2\theta_{\text{эксп}}$ должны приводиться в градусах и их десятичных долях и быть исправлены на систематические инструментальные ошибки;

3) интенсивности линий спектра необходимо оценивать в численной форме по стобалльной шкале, а интенсивности менее единицы — в десятичных долях; значения интенсивностей не должны быть более точными, чем их можно измерить;

4) необходимо указывать воспроизводимость измеренных 2θ и I из измерений рентгенограмм, получаемых от многократно приготовленных перенабивкой образцов;

5) спектр должен быть проиндцирован и получены соответствующие параметры элементарной ячейки;

6) следует всегда приводить всю доступную исследователю информацию о веществе, от которого получен дифракционный спектр.

Бланк международного формата
«Порошковые дифракционные данные
для характеристики фазы»

(обязательное выделено заглавными буквами)

Характеристика образца

НАЗВАНИЕ (ХИМИЧЕСКОЕ, МИНЕРАЛ, ОБЫДЕННОЕ) _____

ЭМПИРИЧЕСКАЯ ФОРМУЛА _____

ХИМИЧЕСКИЙ АНАЛИЗ НЕТ _____ ЕСТЬ _____

ИСТОЧНИК/СПОСОБ ПРИГОТОВЛЕНИЯ _____

РЕГИСТРАЦИОННЫЙ № по Chem Abstr. _____

ОБОЗНАЧ. ФАЗЫ по ПИРСОНУ _____

Методика

ВИД ИЗЛУЧЕНИЯ, ИСТОЧНИК _____ ДЛИНА ВОЛНЫ _____

ДИСКРИМИНАТОР λ (фильтр, монохроматор и т. п.) _____

ДЕТЕКТОР ИЗЛУЧЕНИЯ (фотопленка, сцинтилляц. и т. п.) _____

ОПИСАНИЕ УСТАНОВКИ (тип и т. п.) _____

ЩЕЛИ: ОГРАНИЧ. _____ ПРИЕМН. _____ СОЛЛЕРА _____

ШИРИНА ИНСТРУМЕНТАЛЬНОГО ПРОФИЛЯ _____ 2 θ

ТЕМПЕРАТУРА _____ °C

ФОРМА ОБРАЗЦА/РАЗМЕР ЧАСТИЦ _____

ОБЛАСТЬ РЕГИСТР. ОТ _____ ° 2 θ ДО _____ ° 2 θ ВИД ДВИЖЕНИЯ ОБР. _____

СТАНДАРТ ДЛЯ 2 θ ВНУТР./ВНЕШ. _____ ПАРАМЕТР ЕГО ЯЧЕЙКИ _____

СПОСОБ КОРРЕКЦИИ 2 θ _{изм} _____

МЕТОД ИЗМЕРЕНИЯ ИНТЕНСИВНОСТИ _____ ПИК. _____ ИНТЕГР. _____

Порог минимальной ин-ти (отн. ед.) _____

Стандарт для ин-ти _____ 2 θ линии стандарта ин-ти _____

Корунд. число I/I_0 _____ Коэфф. перевода, если не использ. корунд _____

Разрешение (полн. ширина на полувысоте максимума) для вещества _____

_____ ° 2 θ .

Воспроизводимость 2 θ для вещества _____ ° 2 θ на _____ ° 2 θ

Данные об элементарной ячейке

Способ определения параметров _____

Способ уточнения параметров _____

Бланк входной информации СКТБ ИПМ АН УССР

- 01 НОМЕР ДОКУМЕНТА _____ 02 СТЕПЕНЬ ДОСТОВЕРНОСТИ _____
- 03 НАЗВАНИЕ ВЕЩЕСТВА (СОЕДИНЕНИЯ) (химич., минералог., синонимы) _____
-
- 04 ФОРМУЛА брутто, химическая, кристаллохимическая _____
-
- 05 КРИСТАЛЛОГРАФИЧЕСКИЕ ДАННЫЕ И ПАРАМЕТРЫ _____
- 06 структурный тип _____ 07 сингония _____ 08 Z _____
- 09 простр. группа _____
- 10 $a =$ _____ 11 $b =$ _____ 12 $c =$ _____
- 13 альфа = _____ 14 бета = _____ 15 гамма = _____
- 16 плотность_{экс} _____ 17 плотность_{расч} _____
- 18 структура исследована в _____
- 19 УСЛОВИЯ ПРОВЕДЕНИЯ ЭКСПЕРИМЕНТА _____
- 20 тип излучения _____ 21 длина волны _____
- 22 способ выделения монохроматического излучения _____
- 23 тип прибора _____
- 24 метод измерения интенсивности _____
- 25 поправка при опр. угла отраж. _____ 26 масс. коэф. погл. обр. _____
- 27 форма и размер образца _____
- 28 корундовое число _____ 29 температура съемки _____
- 30 прочие условия _____
- 31 УСЛОВИЯ ПОЛУЧЕНИЯ (ПРОИСХОЖДЕНИЯ) ОБРАЗЦА _____
-
- 32 ХИМИЧЕСКИЙ СОСТАВ, % _____
-
- 33 ДОПОЛНИТЕЛЬНЫЕ ХАРАКТЕРИСТИКИ ОБРАЗЦА _____
- 34 температура разложения _____ 35 температура сублимации _____
- 36 темпер. фазов. перехода _____ 37 темпер. плавл. _____
- 38 n_g _____ 39 n_m _____ 40 n_p _____ 41 $2V$ _____
- 42 спайность _____
- 43 дефектность _____

Бланк предлагаемого формата
«Стандартная форма порошковых дифракционных данных»

ХАРАКТЕРИСТИКА ОБРАЗЦА

Название вещества _____

Формула _____

Источник _____

Свойства _____

Плотн. _____ (_____) (изм.) _____ (выч.) Осноть, знак _____

 n_g _____ (_____) n_m _____ (_____) n_p _____ (_____) $2V$ _____

Химический анализ _____

Код соединения _____

КРИСТАЛЛОГРАФИЧЕСКИЕ ДАННЫЕ

Сингония _____ Простр. группа _____

Структура _____ Структурный тип _____

 $a =$ _____ (_____) \AA $b =$ _____ (_____) \AA , $c =$ _____ (_____) \AA $Z =$ _____ $\alpha =$ _____ (_____)° $\beta =$ _____ (_____)° $\gamma =$ _____ (_____)° $V =$ _____ \AA^3

Способ определения и уточнения параметров _____

Показатель достоверности _____

УСЛОВИЯ СЪЕМКИ РЕНТГЕНОГРАММЫ

Вид излучения, анод _____ длина волны _____

Способ монохроматизации излучения _____

Тип прибора _____ тип детектора _____

Щели: огранич. _____ приемн. _____ Соллера _____

Область регистрации $2\theta^\circ$ от _____ до _____ температура _____ °C

Форма и размер образца _____

Вид движения образца _____

Внутр./внешн. стандарт 2θ _____ параметр его ячейки _____

ОГЛАВЛЕНИЕ

Предисловие	3
Глава 1	
ОСНОВЫ КАЧЕСТВЕННОГО РЕНТГЕНОФАЗОВОГО АНАЛИЗА	
1.1. Принципы рентгенофазового анализа и их физическая интерпретация	—
1.2. Справочные стандартные данные и предъявляемые к ним требования	10
1.3. Описание форматов для представления порошковых рентгенодифракционных данных справочных стандартов	12
1.4. Структура и состав картотеки JCPDS и указателей к ней	22
1.5. Рентгенодифракционные справочные стандарты Национального бюро стандартов США	32
1.6. Структура и состав отечественных рентгенометрических определителей	33
Глава 2	
МЕТОДЫ РУЧНОГО КАЧЕСТВЕННОГО РЕНТГЕНОФАЗОВОГО АНАЛИЗА	
2.1. Интерпретация экспериментальных данных	37
2.2. Методики рентгенофазового анализа, использующие величины $\{d_i, I_i\}$	38
2.2.1. Методика рентгенофазового анализа по В. И. Михееву	39
2.2.2. Методика рентгенофазового анализа с использованием указателя Ханавальта	43
2.2.3. Методика рентгенофазового анализа с применением указателя Финка	48
Глава 3	
ПОДГОТОВКА ДАННЫХ ДЛЯ РЕНТГЕНОФАЗОВОГО АНАЛИЗА (некоторые вопросы дифрактометрии)	
3.1. Выбор оптимальных условий регистрации дифракционных спектров	52
3.2. Определение положения линий дифракционного спектра	55
3.3. Оценка качества дифракционных спектров эмпирических справочных стандартов	56
Глава 4	
СИСТЕМЫ АВТОМАТИЧЕСКОГО КАЧЕСТВЕННОГО РЕНТГЕНОФАЗОВОГО АНАЛИЗА	
4.1. Применение ЭВМ для решения задач качественного РФА	—
4.2. Система ZRD	63
4.3. Система Джонсона — Ванда	67
4.4. Система PDSM CIS	73

4.5. Система фирмы Филипс	76
4.6. Система фирмы Сименс IDENT	81
4.7. Система фирмы IBM	84
4.8. Система Снайдера	87
4.9. Система Фиалы	91
4.10. Система ФАЗАН	94
4.11. Система АРФА	98
4.12. Другие системы автоматической идентификации	108

Глава 5

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ИДЕНТИФИКАЦИИ 110

5.1. Сравнительный анализ различных стратегий поиска и выбора	—
5.2. Матричное представление баз данных	118
5.3. Плотность распределения поисковых признаков	123
5.4. Полнота базы данных	131
5.5. Выбор окна	137
5.6. Меры близости	140
5.7. Сжатие размерности векторного пространства и структурирование баз данных	150
5.8. Использование химической информации	156

Глава 6

ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ ПРОЦЕДУР ИДЕНТИФИКАЦИИ И БАЗ ДАННЫХ . . . 158

6.1. Исследования и тестирование систем идентификации	—
6.2. Методика испытаний системы АРФА	161
6.3. Исследование базы данных	169
Литература	179
Приложения	187

Евгений Константинович Васильев
Михаил Сергеевич Нахмансон

**КАЧЕСТВЕННЫЙ
РЕНТГЕНОФАЗОВЫЙ АНАЛИЗ**

Утверждено к печати
Институтом земной коры СО АН СССР

Редактор издательства *А. М. Самсоенко*
Художественный редактор *М. Ф. Глазырина*
Художник *С. М. Кудрявцев*
Технический редактор *А. В. Сурганова*
Корректоры *А. А. Надточий, Г. И. Шведкина*

ИБ № 30036

Сдано в набор 04.06.85. Подписано к печати 18.12.85. МН-02657. Формат 60×90^{1/16}.
Бумага типографская № 2. Обыкновенная гарнитура. Высокая печать. Усл. печ.
л. 12,5. Усл. кр.-отт. 12,8. Уч.-изд. л. 13. Тираж 1500 экз. Заказ № 763. Цена 2 руб.

Ордена Трудового Красного Знамени издательство «Наука», Сибирское отделение.
630099, Новосибирск, 89, Советская, 18.

4-я типография издательства «Наука». 630077, Новосибирск, 77, Станиславского, 25.

В СИБИРСКОМ ОТДЕЛЕНИИ ИЗДАТЕЛЬСТВА «НАУКА»

готовятся к выпуску
следующие книги:

Конников Э. Г. Дифференцированные гипербазит-базитовые комплексы докембрия Забайкалья.— 17 л.

Обобщены данные по докембрийским дифференцированным интрузиям основного и ультраосновного составов. Рассмотрены особенности вещественного состава, генезиса и металлогенической специализации комплексов. Исследуется вопрос влияния процессов взаимодействия базитовых расплавов на состав магматических пород и пространственное размещение внутри гипербазит-базитовых плутонов магматического сульфидного и железо-титанового оруденения.

Для специалистов в области региональной геологии докембрия, петрологии расслоенных интрузий, магматического рудообразования.

Проблемы формационного анализа.— 15 л.

Рассматриваются основные задачи и направления развития формационного анализа и его методические аспекты. Применение основных методических разработок хорошо показано на примерах изучения конкретных формаций и их групп. Большое внимание уделено детальному строению и сравнительной характеристике красноцветных, соленосных, карбонатных рифогенных и битуминозно-сланцевых, а также молассовых формаций, развитых в различных регионах Советского Союза.

Для геологов, литологов, специалистов, занимающихся вопросами формационного анализа.

Маршинцев В. К. Вертикальная неоднородность кимберлитовых тел Якутии.— 18 л.

Предложена модель вертикальной неоднородности кимберлитовой колонны, наиболее удовлетворительно объясняющая различный состав и текстурно-структурные особенности кимберлитовых пород трубок. Детально рассмотрен процесс серпентизации оливина, протекающий в позднемагматическую стадию; обсуждаются причины и следствия этого процесса. Установлены закономерности распределения петрогенных и редких элементов в кимберлитовых породах, характеризующих различные степени глубинности, дайках и жилах кимберлитов, альнеитов и кимберлитовых карбонатитах.

Для геологов, геохимиков и петрологов.

Соловьев В. И. Меловой вулканизм Северо-Востока СССР.— 20 л.

Приводится детальная петрографическая, петрохимическая, геохимическая и минералогическая характеристика пород. Впервые для всего Северо-Востока СССР составлена корреляционная схема вулканических свит, комплексов, и формаций мелового возраста. Показана связь вулканизма и тектоники, проявившихся на этой территории в меловое время.

Для геологов, петрологов и металлогенистов, изучающих вулканические породы и их рудоносность.

В IV квартале 1986 г.

готовятся к выпуску следующие книги:

Забродин В. Ю., Оноприенко В. И., Соловьев В. А. Основы геологической картографии.— 15 л.

Коробейников А. Ф. Условия концентрации золота в палеозойских орогенах.— 20 л.

Хаустов А. П. Многомерный анализ гидрогеологических систем горно-складчатых областей.— 10 л.

Базаров Д. Б. Кайнозой Прибайкалья и Западного Забайкалья.— 20 л.

Биостратиграфия мезозоя Сибири и Дальнего Востока.— 19 л.

Афанасьев С. А., Красавчиков В. О. Задачи прогноза и оптимизации при разведке месторождений нефти и газа.— 10 л.

Космическое вещество и Земля.— 20 л.

Шугрин С. М., Обут А. М. Солнечная активность и биосфера.— 9 л.

Ковалев В. П. Устойчивые вариации химизма в петро- и магмогенезе.— 20 л.

Для получения книг почтой
заказы просим направлять по адресам:

- 480091 **Алма-Ата**, ул. Фурманова, 91/97
370005 **Баку**, ул. Джапаридзе, 13
232600 **Вильнюс**, ул. Университето, 4
690088 **Владивосток**, Океанский проспект, 140
320093 **Днепропетровск**, проспект Гагарина, 24
734001 **Душанбе**, проспект Ленина, 95
375002 **Ереван**, ул. Гуманяна, 31
664033 **Иркутск**, ул. Лермонтова, 289
420043 **Казань**, ул. Достоевского, 53
252030 **Киев**, ул. Ленина, 42
252142 **Киев**, проспект Вернадского, 79
252030 **Киев**, ул. Пирогова, 2
252030 **Киев**, ул. Пирогова, 4
277012 **Кишинев**, проспект Ленина, 148
343900 **Краматорск** Донецкой обл., ул. Марата, 1
660049 **Красноярск**, проспект Мира, 84
443002 **Куйбышев**, проспект Ленина, 2
191104 **Ленинград**, Литейный проспект, 57
199164 **Ленинград**, Таможенный пер., 2
196034 **Ленинград**, В/О, 9 линия, 16
197345 **Ленинград**, Петрозаводская ул., 7
220012 **Минск**, Ленинский проспект, 72
103009 **Москва**, ул. Горького, 19а
117312 **Москва**, ул. Вавилова, 55/7
117192 **Москва**, Мичуринский проспект, 51
630076 **Новосибирск**, Красный проспект, 51
630090 **Новосибирск**, Академгородок, Морской проспект, 22
142284 **Протвино** Московской обл., «Академкнига»
142292 **Пуцино** Московской обл., МР, «В», 1
620151 **Свердловск**, ул. Мамина-Сибиряка, 137
700029 **Ташкент**, ул. Ленина, 73
700100 **Ташкент**, ул. Шота Руставели, 43
700187 **Ташкент**, ул. Дружбы народов, 6
634050 **Томск**, наб. реки Ушайки, 18
450059 **Уфа**, ул. Р. Зорге, 10
450025 **Уфа**, ул. Коммунистическая, 49
720001 **Фрунзе**, бульвар Дзержинского, 42
310078 **Харьков**, ул. Чернышевского, 87

2 руб.

4537