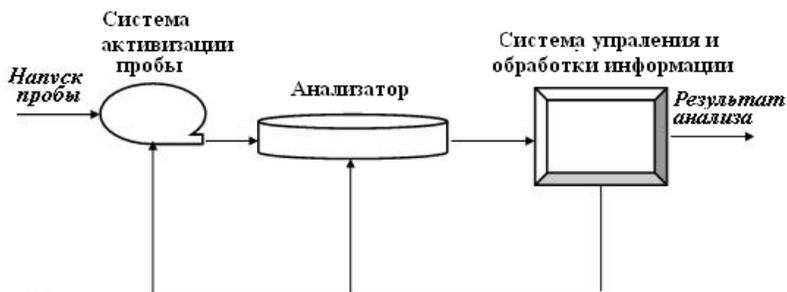


В.В. МАНОЙЛОВ Л. В. НОВИКОВ

ПОЛУЧЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ АНАЛИТИЧЕСКИХ ПРИБОРОВ



Санкт-Петербург
2014

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

УНИВЕРСИТЕТ ИТМО

В.В. МАНОЙЛОВ Л. В. НОВИКОВ

**ПОЛУЧЕНИЕ И ОБРАБОТКА
ИНФОРМАЦИИ АНАЛИТИЧЕСКИХ
ПРИБОРОВ**

 **УНИВЕРСИТЕТ ИТМО**

Санкт-Петербург

2014

Манойлов В.В. Новиков Л.В. Получение и обработка информации аналитических приборов – СПб: Университет ИТМО, 2014. – 176 с.

Аналитические приборы предназначены для определения структуры, состава и концентрации веществ в различных средах, материалах и продуктах. Без аналитических приборов невозможно проведение исследований и решение практических задач в области физики и химии, в создании новых материалов, в том числе на основе нанотехнологий и т.п. Современные аналитические приборы – это сложный комплекс, состоящий из систем напуска пробы, анализатора и системы обработки информации. В пособии рассмотрены основы получения и обработки информации аналитических приборов, которые зависят от характера физико-химических процессов, происходящих в анализаторе.

Учебное пособие адресовано магистрантам, обучающимся по направлению «Приборостроение»- 12.04.01(200100.68) по магистерским программам: «Приборы исследования материалов на микро и наноразмерном уровне» и «Методы диагностики и анализа в бионанотехнологиях». Отдельные разделы пособия адресуются студентам, проходящим подготовку бакалавров по направлению «Приборостроение» 12.03.01(200100.62), по профилю «Приборы исследования материалов на микро и наноразмерном уровне»

Рекомендовано к печати Ученым советом факультета точной механики и технологий, протокол № 8 от 14.10 2014



Университет ИТМО – ведущий вуз России в области информационных и фотонных технологий, один из немногих российских вузов, получивших в 2009 году статус национального исследовательского университета. С 2013 года Университет ИТМО – участник программы повышения конкурентоспособности российских университетов среди ведущих мировых научно-образовательных центров, известной как проект «5 в 100». Цель Университета ИТМО – становление исследовательского университета мирового уровня, предпринимательского по типу, ориентированного на интернационализацию всех направлений деятельности.

© Университет ИТМО, 2014

© Манойлов В.В., Новиков Л.В., 2014

Содержание

Введение	4
1. Физико-химические основы анализа веществ и материалов.	9
1.1 Масс-спектрометрия	10
1.2 Хроматография	16
1.3. Мёссбауэровская спектроскопия	17
1.4. Капиллярный электрофорез	21
2. Анализ наноматериалов	28
2.1 Обзор методов анализа наноматериалов	28
2.2 Оптические микроскопы	35
2.3 Зондовые микроскопы	39
2.4 Электронные микроскопы	54
2.5 Спектроскопия: инфракрасная , рамановская, фотоэлектронная	67
3 Представление данных аналитических приборов	74
3.1 Типы данных	74
3.2 Пространство сигналов - векторов	76
3.3. Представление многомерных данных	89
3.4. Меры близости	92
3.5 Аппаратная функция и модели сигналов	93
4 Проекционные методы обработки данных	95
4.1 Метод главных компонент	96
4.2. Преобразование Фурье	104
4.3 . Вейвлет преобразование	108
5 Повышение разрешающей способности аналитических приборов алгоритмическими средствами	128
5.1 Свертка аппаратной функции и сигнала спектрометров	129
5.2 Методы восстановления спектров	133
5.3 Корректно и некорректно поставленные задачи	134
5.4 Понятие регуляризирующего оператора	138
5.5 Метод регуляризации по Тихонову	141
5.6 Метод регуляризации основанный на Singular Value Decomposition (SVD)	145
6 Программное обеспечение для решения задач восстановления - спектров	154
6.1 Исследование эффективности программ решения задач восстановления масс-спектров методом регуляризации Тихонова	154
6.2 Программы восстановления спектров методом SVD	158
Заключение	165
Список литературы	166
Приложение 1	173

Введение

Аналитические приборы - это то, что необходимо для получения информации о веществах и материалах. В настоящее время аналитическое приборостроение продолжает стремительно развиваться как в направлении создания новых методов и методик анализа, так и в создании современных приборов и приборных комплексов, отвечающих всем признакам и требованиям технического прогресса нашего времени. Аналитические приборы являются основными измерительными средствами в высокоточных исследованиях и измерениях в аналитической химии, в технологиях микроэлектроники, в производстве особо чистых веществ, в нанотехнологиях, в геохронологии (прогнозировании месторождений полезных ископаемых), производстве и контроле ядерного горючего в ядерно-топливном цикле и в большой номенклатуре научных исследований.

Современные методы анализа используют разнообразные явления, вызываемые воздействием на исследуемую среду электрического тока, электромагнитного и оптического излучения в самом широком диапазоне длин волн или проникающей радиации.

Аналитическое приборостроение не стоит на месте - оно развивается вместе с ростом уровня науки, технологий и потребностей общества в самых различных областях его деятельности.

Несмотря на такое разнообразие приборов и методов, можно выделить несколько важных потребительских характеристик, присутствующих всем приборам: *чувствительность, разрешающая способность, воспроизводимость, точность и достоверность анализа*. **Чувствительность** прибора характеризуется минимальным количеством вещества, которое необходимо для оценки его концентрации с заданной погрешностью. **Разрешающая способность** – способность прибора разделить многокомпонентную смесь на составляющие таким образом, чтобы можно было оценить количественный состав каждой компоненты. **Воспроизводимость** определяется стабильностью параметров прибора от анализа к анализу, а **точность** – величиной обратной погрешности, с которой определяется концентрация каждой компоненты. **Достоверность** анализа является, пожалуй, самым важным интегральным показателем качества прибора, который предоставляет пользователю право смело принимать иногда без преувеличения судьбоносные решения. Измерения будут достоверными (истинными, верными), если: а) прибор откалиброван и поверен, т.е. характеризуется необходимой точностью и воспроизводимостью; б) выверены и сертифицированы методики анализа, начиная от пробоподготовки и заканчивая регистрацией результатов. Достоверность анализа иногда

может быть охарактеризована вероятностью оценки измеряемой величины, например, по аналогии с интервальными оценками.

Современные аналитические приборы – это сложный комплекс, состоящий из систем напуска пробы, анализатора и электронной системы управления и обработки данных (см. Рис. В.1).

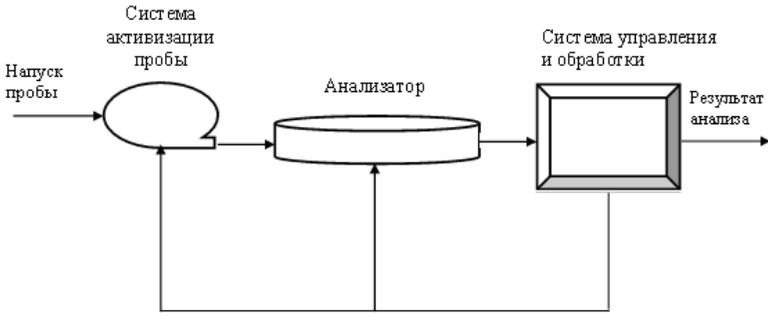


Рис.В.1 Обобщенная схема аналитического прибора

Все операции по отбору и подготовке проб, преобразованию, разделению, дозированию веществ, возбуждению их активности (например, ионизации) в системе ввода пробы, разделению пробы на компоненты в анализаторе, обработке и представлению информации автоматизируются с помощью системы управления и обработки. В результате уровень эксплуатационных характеристик прибора, качество анализа в целом зависит от качества работы как каждой систем, так и совместной работы этих систем, управляемых его электронной частью - интеллектом прибора. Он контролирует последовательность выполнения операций в процессе анализа, считывает данные и производит вычисления, необходимые для получения оценок измеряемых величин. Большое разнообразие применяемых физических принципов и методов анализа, что, естественно, сказывается на его конструкции, порождает большое количество алгоритмов управления и контроля. *Алгоритмы обработки данных должны разрабатываться или выбираться в зависимости от характера физико-химических процессов, происходящих в анализаторе, объёма поступающей с прибора данных, формы сигнала и характера информативных параметров, уровня шума и дрейфа и т.п.* Поэтому применяемые алгоритмы и программное обеспечение в целом для каждого типа приборов может существенно отличаться. Тем не менее, общей задачей системы обработки была и

остаются достижение наилучших потребительских (эксплуатационных) характеристик прибора. Очевидно, что чем выше качество обработки сигнала, обеспечивающее достижение предельных характеристик, тем проще и дешевле может быть физический тракт прибора для достижения одних и тех же результатов. Например, две важнейшие и противоречивые характеристики аналитических приборов – *чувствительность и разрешающую способность*, всегда стремятся улучшить с минимальными потерями для других его показателей: стоимости, надежности, габаритов, веса и др. Естественным и, пожалуй, самым дешевым, способом решения этой задачи является построение систем обработки с использованием соответствующих быстрых, компактных и эффективных вычислительных алгоритмов. *Одним из путей достижения этих целей является внедрение новых программно – алгоритмических средств обработки данных, основанных на последних достижениях информатики, прикладной математики и возможностях элементной базы вычислительной техники.*

Современные измерительные (в том числе аналитические) приборы выполняют, как правило, *косвенные* измерения параметров физических экспериментов, технологических процессов, окружающей среды, а также состава и структуры органических и неорганических веществ в химии, биологии, медицине. Наблюдаемый сигнал на выходе прибора содержит полезную информацию в искаженном виде, поэтому восстановить истинную картину исследуемого явления не всегда представляется возможным. Это можно сделать с некоторой достоверностью только в рамках принятой модели поэтапно. На первом этапе сигнал должен быть очищен от шума и искажений, вызванных измерительным трактом. На втором и последующих этапах, используя калибровочные зависимости и соответствующие базы данных, привлекая статистические методы, решается задача идентификации (интерпретации) измерений.

В зависимости от типа прибора, характера сигналов и шумов математические средства привлекаемые для целей обработки информации отличаются. Если, например, информация содержится в величине тока или напряжения, то производится оценка их среднего значения с применением алгоритмов, учитывающих характер шума и продолжительность измерения (выборочное среднее, метод медианы). Если же информация содержится в интенсивности и временном положении пиков, то применяются, чаще всего, корреляционные алгоритмы оценки этих параметров.

При обработке в реальном времени используется сравнительно небольшой объем данных, размещаемый в так называемом скользящем окне (интервале) данных. К алгоритмам

обработки при этом предъявляются жесткие требования по количеству выполняемых операций, которое должно быть минимальным при достижении заданных критериев качества.

При обработке данных во вторичном времени вся исходная информация в цифровом виде хранится в памяти ЭВМ и исследователь может использовать такие эффективные средства, как прямое и обратное преобразования Фурье, фильтры Винера или Калмана, нейронные сети и др., которые требуют значительного машинного времени.

Можно выделить несколько этапов развития методов и алгоритмов обработки экспериментальных данных, получаемых с помощью приборов, которые в значительной степени были обусловлены возможностями вычислительной техники на каждом этапе:

1. Широкое применение различных методов сглаживания и цифровой фильтрации в текущем окне данных (60-70 -е годы);
2. Внедрение методов, вытекающих из максимизации функционала отношения правдоподобия: корреляционные алгоритмы, метод наименьших квадратов (МНК), фильтры Винера и Калмана (70 -е годы);
3. Применение быстрого преобразования Фурье (БПФ) с различными окнами для сглаживания шума, разделения наложившихся пиков, дифференцирования и др. (70-80-е годы);
4. Внедрение методов обработки многомерных данных (метод главных компонент, вейвлет анализ и др.) (90 - 00-е годы).

В результате, мы имеем достаточно большую палитру алгоритмов обработки данных. Их можно разбить на две большие группы:

- а) алгоритмы, реализуемые во временной области.
- б) алгоритмы, реализуемые в спектральной области;

К первой группе относятся фильтры Савицкого – Голзя, фильтр Калмана, корреляционные алгоритмы. Ко второй – алгоритмы фильтрации, обратной свертки и др., связанные с Фурье, вейвлет или другими преобразованиями сигнала. Каждая из этих групп имеет свои достоинства и недостатки. В частности первая группа алгоритмов в отличие от второй не требует серьезных аппаратных затрат. Вторая по сравнению с первой позволяет достичь более высокой эффективности оценок, что вместе с возможностями современной элементной базы делает эту группу алгоритмов более предпочтительной. *Эта группа алгоритмов получила общее название – проекционные методы обработки.*

Вопросам обработки сигналов различного типа посвящено множество монографий в области математической статистики, обработки сигналов и др. прикладных направлений [1-16].

Для разработки оптимальных методов, позволяющих наиболее точно и быстро оценить параметры реальных сигналов, необходимо изучение существующих и вновь появляющихся подходов к обработке данных и ранжирование их возможностей по отношению именно к задачам конкретного эксперимента.

Это связано, прежде всего, с тем, что методы обработки вновь появляющихся приборов, основанных на новых физико-химических принципах анализа, требуют освоения новых подходов. Эти подходы в области аналитической химии развиваются в рамках *хеометрики* [17,18].

Новые результаты прикладной математики и появление быстродействующей вычислительной техники позволило использовать многие сложные алгоритмы обработки данных, в особенности методы анализа данных многоканальных экспериментов. В свою очередь, это позволило разрабатывать более сложную аналитическую технику, способную выполнять большой объём измерений.

Наиболее ярко эта тенденция проявилась при использовании инфракрасной (ИК) спектроскопии, особенно в ближней области (БИК), прежде считавшейся малополезной из-за высокого и трудно устранимого шума и низкой разрешающей способности. Первые работы по хеометрике были посвящены методам анализа спектроскопических данных, построению для них калибровочных моделей с помощью метода главных компонент и метода проекций на латентные структуры [18].

Хеометрические идеи оказались востребованными в приборах капиллярного электрофореза, полимеразной цепной реакции в реальном времени (ПЦР РВ), в микро- и наноаналитических системах, так называемых «приборов на чипе».

Перевод сепарационных методов химического и биологического анализа (электрофореза, хроматографии и т.п.) на микрочипы привел к изменению формы аналитических пиков от гауссовой к кучочно-линейной (трапециидальной или треугольной, т.е., совокупности трендов нулевого и первого порядков - ЛТ0 и ЛТ1) [19].

Не менее значимыми последствиями миниатюризации приборов (их узлов) стало сокращении времени всех стадий анализа и, как следствие, ужесточение требований к быстродействию преобразования и обработки сигналов и оценивания их параметров. Еще одним следствием миниатюризации может стать многократное уменьшение анализируемых объемов, что при определенных видах детектирования

(напр., амперометрическом или флуориметрическом) приведет к многократному уменьшению информативного сигнала (отношения сигнал/шум). Последнее обстоятельство требует применения помехоустойчивых (робастных) методов оценивания параметров сигналов.

Современный аналитический прибор является гибридным комплексом, аналитические возможности которого определяются тем, насколько оптимально измерительные функции в нем разделены между его аналитической (физической) и аппаратно-программной частями. Функции аппаратно-программного комплекса состоят в сборе, обработке и идентификации информации, поступающей от его аналитической части. При этом задачи программных, предназначенных для обнаружения и оценки параметров полезных сигналов, выявленных из необработанных сигналов, содержащих шумы и наводки, а также для фильтрации сигналов, состоят в обеспечении требуемой точности анализа и в снижении нагрузки на аналитическую часть масспектрометра, обеспечивающего решение прецизионных задач анализа [20].

Таким образом, современное аналитическое приборостроение постоянно ставит новые задачи, требующие развития и совершенствования алгоритмических методов обработки данных, направленных на достижение предельных показателей качества анализа.

Целью данной учебного пособия является раскрытие и объяснение физических принципов анализа для некоторых направлений аналитического приборостроения, а также работы алгоритмического обеспечения различных подсистем, входящих в системы автоматизации аналитических приборов.

1 Физико-химические основы анализа веществ

Современный уровень развития физики, химии и технологии позволяет создавать довольно большое разнообразие аналитических приборов с оригинальным, подчас весьма сложным математическим обеспечением, разработанным с учетом их конструктивных особенностей и методов анализа. Поэтому аналитический прибор следует рассматривать как аппаратный и программно – алгоритмический комплекс, компоненты которого взаимно дополняют друг друга с целью достижения наилучших эксплуатационных характеристик прибора в целом. В главе в самом общем виде дано описание некоторых принципов качественного и количественного анализа веществ и соединений.

1.1 Масс-спектрометрия

Трудно представить область человеческой деятельности, где не нашлось бы места масс-спектрометрии: аналитическая химия, биохимия, клиническая химия, общая химия и органическая химия, фармацевтика, косметика, парфюмерия, пищевая промышленность, химический синтез, нефтехимия и нефтепереработка, контроль окружающей среды, производство полимеров и пластиков, медицина и токсикология, криминалистика, допинговый контроль, контроль наркотических средств, контроль алкогольных напитков, геохимия, геология, гидрология, петрография, минералогия, геохронология, археология, ядерная промышленность и энергетика, полупроводниковая промышленность, металлургия. Масс-спектрометрический анализ основан на взаимодействии ионов анализируемого вещества с электрическим и/или магнитным полем, что приводит к разделению его на компоненты в зависимости от отношения массы к заряду и определения количества заряженных частиц. Масс-спектрометр, таким образом, является в основном прибором многокомпонентного анализа. Существенное отличие масс-спектрометрии от других аналитических методов состоит в том, что оптические, рентгеновские и некоторые другие методы детектируют излучение или поглощение энергии молекулами или атомами, а масс-спектрометрия непосредственно детектирует сами частицы вещества [21].

После напуска пробы, которая выполняется часто достаточно сложными по конструкции устройствами, анализируемое вещество попадает в источник ионов (Рис. 1.1). Источник ионов превращает нейтральные молекулы и атомы в заряженные частицы – ионы.

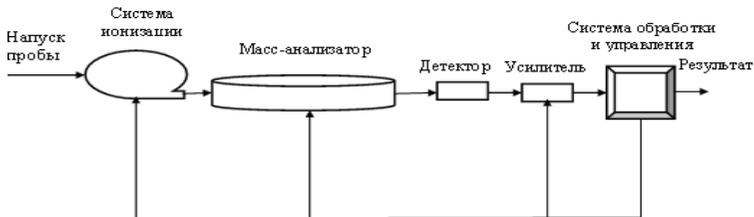


Рис. 1.1 Упрощенная схема масс-спектрометра

Способ ионизации зависит от вида вещества – органическое, неорганическое, и от его состояния - газообразное, жидкое, твердое. Важнейшим узлом источника ионов является его ионно-оптическая

система, который формирует пучки или струйки ионов перед их вводом в анализатор. В результате на выходе источника ионов формируется поток ионов, имеющий определенную пространственно – временную структуру и энергию.

Масс – анализатор определяет основные аналитические характеристики прибора. В нем происходит разделение ионов по массам (точнее, по отношению массы к заряду m/z_e), которые содержат всю информацию о компонентах анализируемого вещества. В зависимости от конструкции различают непрерывные (магнитные и электростатические секторные, квадрупольные) и импульсные (время- пролетные, ионная ловушка, квадрупольная ионная ловушка, ионно – циклотронного резонанса, орбитрэп) масс – анализаторы. Разница между непрерывными и импульсными масс-анализаторами заключается в том, что в первые ионы поступают непрерывным потоком, а во вторые — порциями, через определённые интервалы времени.

Сформированные в масс-анализаторе струйки ионов различных масс последовательно попадают на детектор. Для детектирования используют коллекторы Фарадея, микроканальные пластины, электронные или фото умножители. Выходной ток детектора усиливается, преобразуется в цифровой код и обрабатывается системой обработки и управления с целью извлечения из полученного сигнала необходимой информации об анализируемом веществе.

Так как в источнике ионов молекулы в той или иной степени разрушаются, масс-спектр даже одного вещества содержит несколько спектральных линий. В результате реальный масс-спектр содержит огромное число линий, разобраться в которых должна система обработки информации.

На практике получили распространение “щадающие” методы ионизации, например, электроспрей, когда молекулы подвергаются минимальному разрушению. В элементном и изотопном анализе дело обстоит несколько проще, так как ионизируются не молекулы, а атомы, но существуют другие проблемы, связанные с необходимостью прецизионных измерений интенсивности малых пиков на фоне близкорасположенных пиков большой амплитуды.

Процесс детектирования рассмотрим на примере статического (рис. 1.2а) и времяпролетного (Рис. 1.2б) масс – анализаторов.

В статическом масс-спектрометре ионы, перемещаясь от источника к детектору в магнитном поле, в зависимости от отношения m/z_e расходятся по различным траекториям, одновременно группируясь в струйки на линии фокусов. Радиус такой траектории определяется выражением:

$$r_e = mV/Hz_e, \quad (1.1)$$

где m - масса, V - скорость иона; H - напряженность магнитного поля.

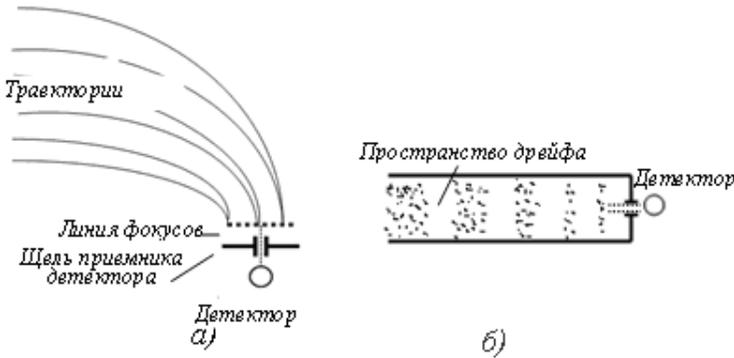


Рис. 1.2 Детектирование ионов в магнитном статическом а) и время-пролетном б) масс-спектрометрах

В масс-анализаторе с двойной фокусировкой (к магнитному добавлен электростатический каскад) достигается более высокий уровень фокусировки, благодаря компенсации разброса ионов по энергии.

Изменением интенсивности, например, магнитного поля осуществляют развертку по массе, в результате которой через щель приемника последовательно проходят ионы разных масс, вызывая в детекторе электрический ток. Величина этого тока пропорциональна плотности ионного тока и площади щели.

В диапазоне ионных токов от 10^{-13} А до 10^{-9} А в качестве детектора используется коллектор, как правило, в форме цилиндра Фарадея с последующим электрометрическим усилителем]. Иногда используют многоканальный детектор, состоящий из нескольких коллекторов. Если же величина ионного тока ниже 10^{-13} А, то в качестве детектора используется электронный умножитель (ВЭУ) или фотоумножитель (ФЭУ). При очень малых величинах ионных токов (меньше 10^{-15} А) после умножителя устанавливают электронный счетчик, регистрирующий число ионов в заданном временном интервале.

Во время-пролётном масс-анализаторе (рис. 1.2б) облако (пакет) ионизированного вещества выталкивается в пространство дрейфа, где из-за различной скорости движения ионы разных масс группиру-

ются в сгустки: чем меньше масса, тем, естественно, больше скорость. В результате детектор последовательно во времени регистрирует эти сгустки, образуя развертку по массе.

В квадрупольном масс-анализаторе разделение по массе происходит при движении облака ионов вдоль оси симметрии четырех стержней, формирующих поперечные постоянное и переменное электрические поля

$u(t) = \pm U_0 + V_0 \cos \omega t$ с гиперболическим распределением потенциала (Рис. 1.3).

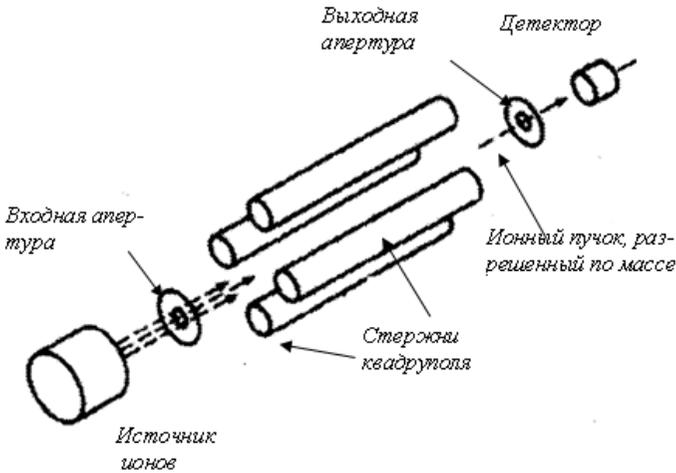


Рис. 1.3 Схема квадрупольного масс-анализатора

При заданных потенциалах U_0, V_0 и частоте ω траектория движения ионов определенной массы будет стабильной и они, пройдя вдоль стержней, попадут на детектор. Развертка по спектру масс производится изменением одного из параметров выражения для $u(t)$. В результате развертки ионы разных масс вызывают в детекторе ток в виде последовательности пиков разной интенсивности.

Таким образом, несмотря на различие физических принципов разделения ионов в упомянутых выше масс-анализаторах, характер сигнала, формируемого детектором, будет один и тот же: последовательность пиков разной интенсивности по оси масс (Рис. 1.4).

Очевидно, что пики близких масс, например m_1 и m_2 , будут располагаться рядом и, если масс-анализатор не обладает достаточ-

ным разрешением, они могут слиться в один. Разрешающую способность масс-спектрометров с такими анализаторами определяют по формуле

$$R = \frac{(m_2 + m_1)/2}{m_2 - m_1} \frac{L}{(d_2 + d_1)/2}, \quad (1.2)$$

где $(m_2 + m_1)/2$ - средняя масса, $(d_2 + d_1)/2$ средняя ширина пика (обычно ширина пика определяется на высоте 10% от основания или полувысоте), L - расстояние между центрами пиков. Из формулы видно, что чем больше разнесены соседние пики по шкале масс и чем они уже, тем больше разрешение.

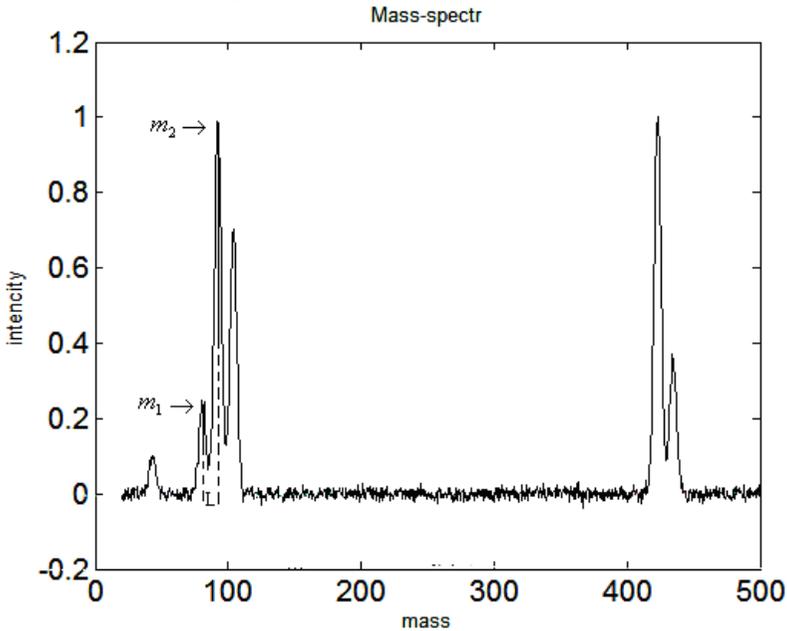


Рис. 1.4 Модельный пример сигнала масс-спектрометра

Ширина пиков определяется качеством фокусировки сгустков ионов одной массы, а их пространственное разделение – длиной пути от источника ионов до детектора. Как следствие, повышение разрешающей способности неизбежно ведет к усложнению источника ионов и ионно-оптической системы и, в итоге – к удорожанию и увеличению габаритов и веса прибора.

В масс-спектрометрах с разверткой форма пика определяется формой ионного пучка в плоскости щели и размерами самой щели (Рис. 1.5).

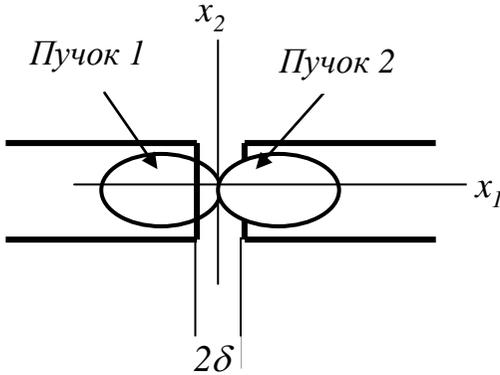


Рис. 1.5. Расположение двух пучков ионов близких масс в плоскости приемной щели

Плотность ионного тока в пучке является функцией двух переменных $i(x_1, x_2)$. При развертке по оси x_1 величина ионного тока за щелью (т.е. в плоскости детектора) в момент времени τ будет зависеть от взаимного расположения щели и пучка в этот момент времени и будет равна

$$I(\tau) = \int_{\tau-\delta}^{\tau+\delta} dx_1 \int_{-\infty}^{+\infty} i(x_1, x_2) dx_2 \quad (1.3)$$

Из анализа рисунка 1.5 следует, что ширина щели влияет на разрешающую способность масс-спектрометра, т.к. два пучка, попавшие в щель, сглаживаются на детекторе и, если щель, например, больше половины ширины пучка, провал между пиками разных масс может исчезнуть. Сужение щели ведет к уменьшению тока детектора и, соответственно, к уменьшению чувствительности – другому важнейшему параметру масс-спектрометра. В время-пролетном и квадрупольном масс-анализаторах форма пика также определяется размерами сфокусированного сгустка ионов и, в меньшей степени – размерами щели.

В современной масс-спектрометрии всё большее распространение получают динамические масс-анализаторы типа ионно-циклотронного резонанса и орбитрэп. Формирование полезного сигнала в них производится наведенным на пластины полем от движущейся

щихся по круговой орбите сгустков ионов. Скорость движения ионов разных масс различна, поэтому выходной сигнал представляет собой смесь гармонических составляющих разной частоты. Система обработки выполняет преобразование Фурье и приведение частотного спектра в спектр масс.

1.2 Хроматография

Хроматография основана на взаимодействии исследуемого вещества в газообразной или жидкой фазе при его движении через слой сорбента или перемещении вдоль него [22]. В качестве сорбента используют твердые (или твердообразные) тела или жидкости. В зависимости от агрегатного состояния анализируемого вещества и сорбента различают газо – твердофазную или газо – сорбционную, газо – жидкостную, жидкостно – твердофазную, жидкостно – жидкостную, жидкостно – гелевую и сверхкритическую флюидную хроматографию [23]. Для проведения хроматографического разделения веществ или определения их физико – химических характеристик служат специальные приборы – хроматографы. Его основные узлы: устройство для ввода пробы, колонка (с сорбентом), детектор и система управления и обработки данных (Рис. 1.6).

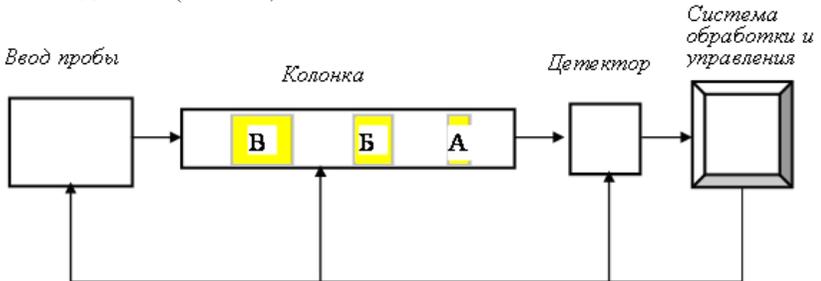


Рис. 1.6 Упрощенная схема хроматографа

После ввода в колонку анализируемого вещества (пробы) вместе с подвижной фазой (элюентом) под действием давления или электростатического поля компоненты смеси (на рисунке – А, Б и В) начинают перемещаться вдоль колонки. Хорошо сорбируемые компоненты (на рисунке - В) двигаются медленнее, чем хуже сорбируемые (А). Поэтому из колонки последовательно выходят и попадают на детектор компоненты А, Б и затем В. При полном разделении на выходе детектора сигнал будет иметь вид, показанный на Рис. 1.7.

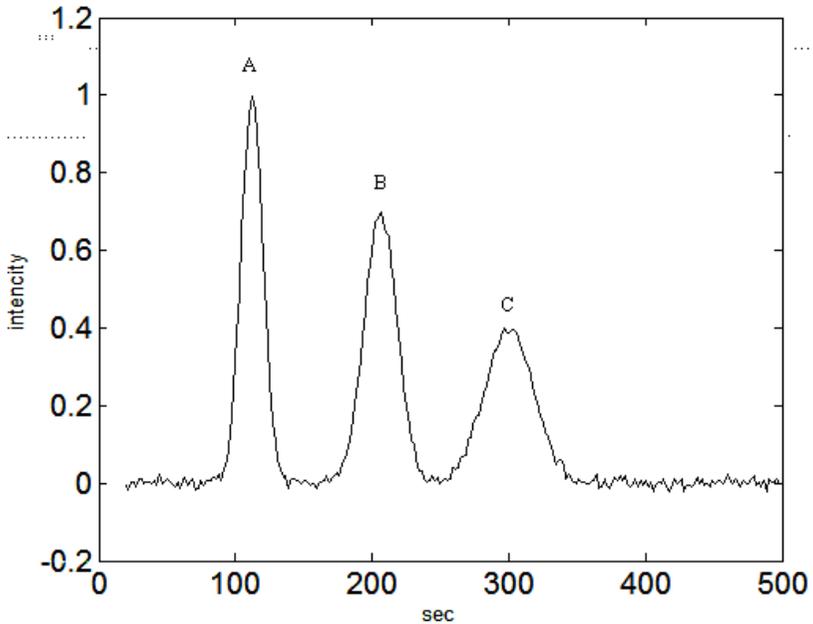


Рис. 1.7 Разделение трехкомпонентной смеси в хроматографической колонке

Сигнал детектора пропорционален концентрации компонентов, преобразуется в цифровой код и обрабатывается системой обработки, которая также осуществляет управление анализом. В процессе обработки определяется время выхода компонентов из колонки (время удерживания) и площадь пиков (или их амплитуда). Так же как и в масс-спектрометрии, многокомпонентная смесь формирует сигнал в виде последовательности пиков разной интенсивности, но отличается тем, что они расширяются со временем. Сами пики могут быть симметричными и ассиметричными, но все – достаточно гладкие.

1.3 Мёссбауэровская спектроскопия

Мёссбауэровская спектроскопия основана на взаимодействии анализируемого вещества с гамма – излучением (γ -излучением) и регистрации его резонансного поглощения ядрами [24-26]. На практике в качестве источника используют чаще всего γ -излучение, возникающее при распаде ядер радиоактивных элементов. Так как энергия γ - кванта весьма значительна – от десятка кэВ до нескольких десятков мэВ – при его излучении одиночное ядро (не связанное в кристал-

лической решетке) получает отдачу, на что тратится часть энергии. При столкновении такого γ -кванта даже с родственным ядром также теряется часть энергии и он уже не способен перевести ядро-поглотитель в возбужденное состояние. В результате резонансное поглощение (или флуоресценция) для таких ядер не наблюдается. Тем не менее, имеются, по крайней мере, два пути для создания условий резонансного поглощения.

Первый – компенсировать энергию на отдачу путем движения источника навстречу ядрам поглотителя. При этом γ -квант источника получает дополнительную энергию, равную кинетической энергии движения.

Второй, более простой, найденный Мёссбауэром (1958 г.), - использование в качестве источника радиоактивных ядер, связанных в кристаллической решетке. Мёссбауэр применил в исследованиях иридий – ^{191}Ir при температуре жидкого гелия и подтвердил, что при этом вероятность отдачи очень мала, так как кристаллическая решетка имеет значительную массу. В результате в своем эксперименте Мёссбауэр наблюдал четкое резонансное поглощение γ -квантов, которое он назвал "упругим ядерным резонансным поглощением γ -излучения" - резонансное ядерное поглощение без отдачи. Вероятность резонансного поглощения велика при сравнительно небольшой энергии γ -квантов (менее 100 кэВ), так как с увеличением этой энергии растет, естественно, вероятность поглощения части энергии кристаллом (рождение фононов). Вероятность мёссбауэровского эффекта также сильно зависит от температуры: часто приходится охлаждать источник и поглотитель до температуры жидкого азота. Однако для γ -переходов очень низких энергий, например, 14,4 кэВ для ядра железа ^{57}Fe или 23,8 кэВ для олова ^{119}Sn , эффект мёссбауэра можно наблюдать вплоть до температуры 1000°K . При прочих равных условиях вероятность этого эффекта тем больше, чем сильнее связь атомов в твердом теле. Резонансное γ -излучение обладает чрезвычайно узкой шириной спектральной линии. Отношение величины энергии к естественной ширине линии (добротность резонансной системы) для некоторых источников равно $0,3 \cdot 10^{13}$ для ^{57}Fe , $0,25 \cdot 10^{11}$ для ^{191}Ir и $0,2 \cdot 10^{16}$ для цинка ^{67}Zn , что недостижимо для известных ныне резонаторов (в том числе - кварцевых). Это обстоятельство позволяет использовать эффект Мёссбауэра для исследования сверхтонких взаимодействий ядер с электронной оболочкой

атома. Важнейшими их типами являются *электрическое монополюное, электрическое квадрупольное и магнитное дипольное взаимодействия*.

В этих взаимодействиях энергия γ -кванта теряется на чрезвычайно малую величину. Это приводит к смещению линий испускания и поглощения друг относительно друга, что легко компенсируется механическим перемещением источника (или приемника). Так, например, удалось измерить предсказанное теорией относительности изменение частоты кванта электромагнитного излучения в гравитационном поле Земли при расстоянии между источником и приемником по вертикали всего 22 метра.

Электрическое *монополюное* взаимодействие приводит к изомерному (или химическому) сдвигу линий резонансного поглощения. Величина сдвига зависит от того, в каком электронном окружении находится ядро, испускающее квант, что позволяет установить природу химических связей и строения химических соединений. Измерение изомерного сдвига широко применяется в физике твердого тела, химии и биологии, благодаря высокой чувствительности электронной плотности в области ядра к особенностям электронной структуры вещества. В частности, величина сдвига зависит от степени окисления железа, которое входит в состав многих биологических объектов.

Электрическое *квадрупольное* взаимодействие – взаимодействие ядра с градиентом электрического поля, создаваемым в области ядра окружающими электронами. Это также приводит к расщеплению ядерных уровней, в результате чего в спектрах резонансного поглощения появляется несколько линий. Например, для ядер Fe^{57} их две – квадрупольный дуплет. Разность энергий между пиками дуплета пропорциональна квадрупольному моменту ядра, который, в свою очередь, характеризует степень отклонения от сферической формы распределения заряда в ядре. Магнитное *дипольное сверхтонкое* взаимодействие – взаимодействие внутреннего магнитного поля ядра с магнитным полем на ядре, создаваемым электронами собственного атома. Величины этих полей очень чувствительны к особенностям электронной структуры твердого тела, к составу магнитных материалов, поэтому исследование магнитной сверхтонкой структуры широко используется для изучения магнитных свойств кристаллов. Магнитное дипольное сверхтонкое взаимодействие также наблюдается в магнитоупорядоченных (ферро-, антиферро-, ферримагнитных) веществах, в которых на ядра атомов действуют сильные магнитные поля H , достигающие величины $\sim 10^6$ э. Энергия магнитного дипольного взаимодействия пропорциональна произведению магнитного момента ядра на H и зависит от ориентации магнитного поля. Поэтому магнитное ди-

полное взаимодействие приводит к расщеплению основного и возбуждённых уровней ядер, в результате чего в спектре поглощения появляется несколько линий, число которых соответствует числу возможных переходов между магнитными подуровнями основного и возбуждённых состояний. Например, для ядра Fe^{57} число таких переходов равно шести. По расстоянию между компонентами магнитной сверхтонкой структуры можно определить напряжённость магнитного поля, действующего на ядро в твёрдом теле. Резонансное поглощение наблюдается только для ядер определенного сорта, поэтому эффект Мёссбауэра используется в тех случаях, когда атомы, на которых наблюдается этот эффект, входят в состав твердых тел и соединений в виде примесей. (Эффект Мёссбауэра наблюдается для 87 изотопов 46 элементов от калия K^{40} до америция Am^{243}). Величина смещения и расщепления уровней энергии ядра, а, следовательно, изменение энергии перехода, сравнительно невелика, поэтому компенсировать эти изменения легко удается за счет эффекта Доплера перемещением источника (или приемника) со скоростью v в пределах от 0.2 до 10 мм/сек. Изменение энергии E γ -квантов будет равно Ev/c (c - скорость света), что при указанных выше скоростях должно быть больше ширины спектральной линии. Измеряя зависимость величины резонансного поглощения (интенсивности γ -квантов после поглотителя) от скорости v получаем мёссбауэровский спектр исследуемого объекта. Упрощенная схема спектрометра Мёссбауэра приведена на Рис.1.8.

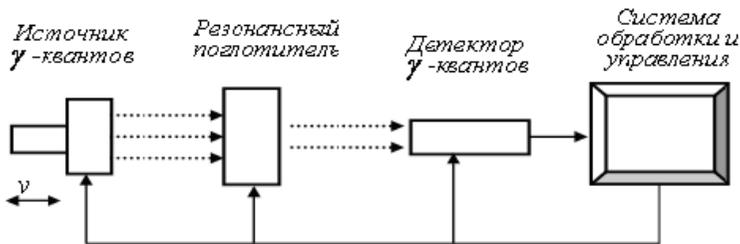


Рис. 1.8 - Упрощенная схема спектрометра Мёссбауэра

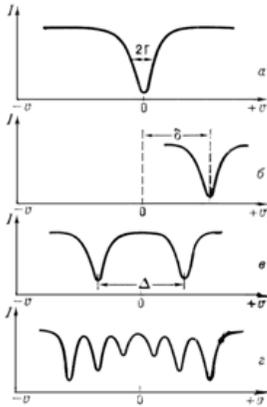


Рис.1.9 Спектры мессбауэровского резонансного поглощения γ -квантов: I — интенсивность потока γ -квантов, прошедших через поглотитель, v — скорость движения источника γ -квантов; а) — одиночные линии испускания и поглощения, не смещенные друг относительно друга при $v = 0$; б) — изомерный или химический сдвиг линии. Сдвиг δ пропорционален электронной плотности в области ядра и меняется в зависимости от особенностей химической связи атомов в твёрдом теле; в) — квадрупольный дублет, наблюдаемый для изотопов Fe^{57} , Sn^{119} , Te^{125} и др. Величина расщепления Δ пропорциональна градиенту электрического поля в области ядра; г) — магнитная сверхтонкая структура, наблюдаемая в спектрах поглощения для магнитоупорядоченных материалов. Расстояние между компонентами структуры пропорционально напряжённости магнитного поля, действующего на ядра атомов в твёрдом теле.

Источник γ -квантов под управлением системы обработки совершает перемещения по заданному закону со скоростью v . Интенсивность потока γ -излучения, прошедшего через исследуемый образец – резонансный поглотитель, регистрируется детектором γ -квантов и анализируется системой обработки и управления. Для регистрации γ -квантов используются сцинтилляционные, газовые пропорциональные или полупроводниковые детекторы. В некоторых случаях применяются детекторы, специально предназначенные для регистрации мессбауэровского излучения (резонансные детекторы).

На Рис. 1.9 приведен пример мессбауэровских спектров при различных типах сверхтонкого взаимодействия.

1.4 Капиллярный электрофорез

Метод капиллярного электрофореза позволяет анализировать ионы металлов, амины, органические и неорганические анионы, сахара, ароматические соединения, нуклеиновые кислоты, производные аминокислот, фрагменты ДНК, производить мониторинг лекарств и др. Метод основан на разделении компонентов смеси в кварцевом капилляре под действием постоянного электрического поля [27]. На внутренней поверхности капилляра диаметром 50-100 мкм (плавленый

кварц SiO_2) в контакте с соответствующим образом подобранным электролитом образуются так называемые силанольные группы, которые диссоциируют с отщеплением ионов H^+ . При pH раствора больше 2.5 на поверхности кварца образуются диссоциированные силанольные группы, которые формируют отрицательный поверхностный заряд. Часть находящихся в растворе катионов, непосредственно прилегающих к поверхности, теряет подвижность (связывается).

Остальные катионы отходят в раствор и образуют заряд, распределенный в объеме жидкости и, в силу меньшей энергии связи с поверхностью, обладают способностью к перемещению. Стенка капилляра, состоящая из отрицательно заряженных остатков силанольных групп, с одной стороны и неподвижный и подвижный слой катионов с другой образуют двойной электрический слой (Рис. 1.10)

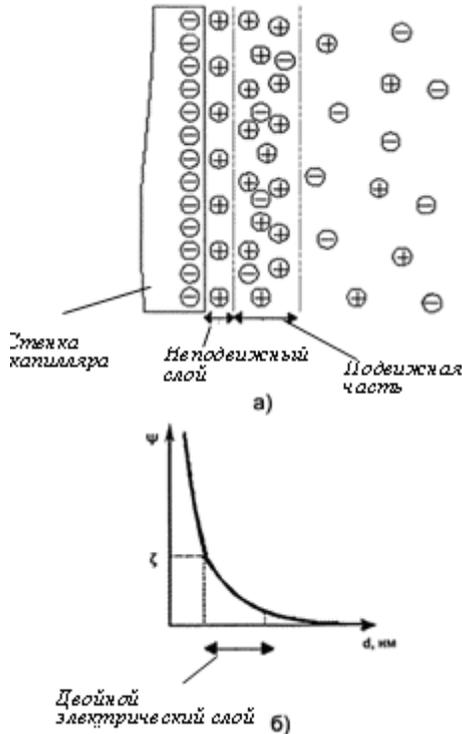


Рис. 1.10 Формирование двойного электрического слоя а) и изменение потенциала на границе раздела: кварц – электролит б).

Одну из обкладок составляют отрицательно заряженные остатки силанольных групп, другая состоит из двух частей - неподвижного слоя катионов, непосредственно примыкающих к поверхности кварца, и диффузного слоя, образованного катионами, находящимися в объеме жидкости. Распределение катионов между неподвижным и диффузным слоями, а, следовательно, и толщина двойного электрического слоя зависит в первую очередь от общей концентрации электролита в растворе. Чем она выше, тем большая часть положительного заряда диффузного слоя перемещается в неподвижный слой и тем меньше становится толщина диффузного слоя (рис. 1.10 б). При концентрации бинарного однозарядного электролита $10^{-3} \dots 10^{-4}$ М толщина двойного электрического слоя составляет в среднем 30...50 мкм.

В капилляре с внутренним диаметром 50...100 мкм оказывается, что практически вся жидкость, заполняющая его, будет представлять собой диффузную часть двойного электрического слоя. Если вдоль оси капилляра приложить электрическое поле, то возникнет продольное движение свободных носителей электрических зарядов (разнополярных ионов) во взаимно противоположных направлениях. В диффузной части двойного электрического слоя присутствует избыточная концентрация катионов, поэтому число ионов, перемещающихся к катоду, будет значительно больше, при этом их движение увлекает за собой всю остальную массу жидкости в капилляре (вследствие молекулярного сцепления и внутреннего трения).

Возникает так называемый электроосмотический поток (ЭОП), направленный к катоду, который будет осуществлять пассивный перенос раствора внутри капилляра (Рис. 1.11).

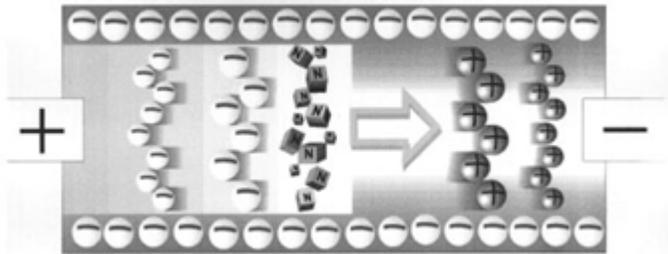


Рис. 1.11 Схема процессов в кварцевом капилляре.

Стрелкой показано направление электроосмотического потока

Вследствие этого процесса в электролите, заполняющем капилляр, возникает направленное перемещение массы жидкости, которое вызвано приложенной разностью потенциалов, при этом вся масса жидкости (за малым исключением

приповерхностного слоя) перемещается с одинаковой скоростью, т.е. формируется плоский профиль скоростей.

Это очень важное обстоятельство, которое позволяет получить чрезвычайно высокую разрешающую способность метода.

Прибор, реализующий метод капиллярного электрофореза, имеет в своём составе кварцевый капилляр, источник высокого напряжения, устройство ввода пробы, детектор и устройство вывода информации. Дополнительные устройства позволяют автоматизировать подачу образцов, термостатировать капилляр и сделать более удобной обработку получаемой информации.

На Рис. 1.12 представлена схема системы капиллярного электрофореза в простейшем случае.

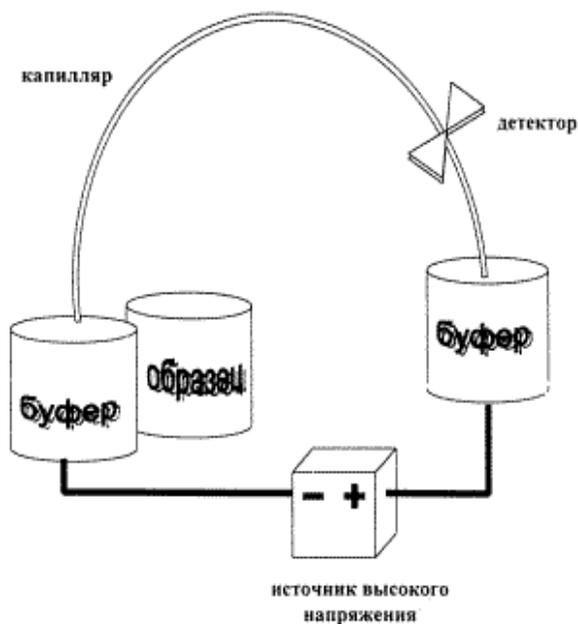


Рис. 1.12 - Схема системы капиллярного электрофореза

Капилляр заполняется раствором электролита, своими концами капилляр опущен в два сосуда, содержащих тот же электролит. Электролит обязательно должен обладать буферными свойствами, чтобы, с одной стороны, воспрепятствовать изменению состава раствора в приэлектродных пространствах, а с другой - стабилизировать состояние компонентов пробы в процессе анализа. В сосуды введены электроды, к которым прикладывается разность потенциалов. Под действием разности потенциалов в капилляре быстро устанавливается стационарное состояние: через него протекает электроосмотический поток, на который будет накладываться электромиграция катионов и анионов во взаимно противоположных направлениях. Как правило, в приборах капиллярного электрофореза ЭОП направлен от входного конца капилляра к детектору поэтому, при использовании кварцевого капилляра, разность потенциалов устанавливают таким образом, что входной конец капилляра имеет положительную полярность (анод), а детектор устанавливается вблизи катода. Если теперь в капилляр со стороны анода ввести небольшой объем раствора пробы, то ЭОП будет переносить зону пробы к катоду, и она некоторое время будет находиться в капилляре под воздействием электрического поля высокой напряженности.

В течение этого времени компоненты пробы, имеющие заряды и отличающиеся от компонентов рабочего буфера, будут перемещаться в соответствии с присущими им электрическими подвижностями, специфичными для каждого компонента. Катионные компоненты пробы, двигаясь к катоду, будут обгонять электроосмотический поток. Скорость их движения будет складываться из скорости ЭОП и скорости электромиграции, поэтому на выходе капилляра катионные компоненты будут появляться первыми и тем раньше, чем больше электрическая подвижность данного иона. Нейтральные компоненты пробы будут перемещаться только под действием ЭОП, и появятся на выходе, когда его достигнет зона пробы. Анионные компоненты перемещаются к аноду с различными скоростями. Некоторые из них, медленно мигрирующие, будут появляться вблизи детектора после выхода ЭОП, а те, чья скорость миграции по абсолютной величине превышает скорость ЭОП, рано или поздно выйдут из капилляра в прианодное пространство.

Если время нахождения пробы в капилляре (которое можно регулировать длиной капилляра, скоростью ЭОП или приложенной разностью потенциалов) достаточно, то на выходе капилляра вблизи катода формируются зоны раствора, в которых находятся индивидуальные компоненты пробы. Происходит, таким образом, разделение исходной смеси. Если теперь с помощью детектора зарегистрировать

появление компонентов на выходе из капилляра, то получим электрофореграмму, которая служит основой для качественного и количественного анализа смеси. Описанный вариант анализа носит название капиллярного зонного электрофореза (КЗЭ). В этом варианте могут определяться катионные компоненты проб и некоторые медленно мигрирующие анионы. Однако главные анионы, определяющие минеральный состав пробы, зарегистрировать таким способом невозможно.

Для того чтобы методом КЗЭ можно было определять анионные компоненты проб, необходимо изменить полярность прикладываемого напряжения. Однако в этом случае изменится не только направление миграции анионов, но также направление ЭОП и он будет препятствовать перемещению в сторону детектора медленно мигрирующих анионов. Для изменения направления ЭОП необходимо модифицировать поверхность кварцевого капилляра таким образом, чтобы знаки зарядов двойного электрического слоя поменялись на обратные, и направление ЭОП совпадало с направлением перемещения анионов. Это достигается введением в рабочий буферный раствор катионного поверхностно-активного вещества, например, бромид цетилтриметиламмония (ЦТАБ). Катион ЦТАБ активно сорбируется на кварцевой поверхности, занимая при достаточной его концентрации, все вакансии в ближайшем к поверхности слое. Поверхность как бы "ошетиливается" длинными цетильными ($C_{16}H_{33}$) цепочками. при дальнейшей промывке рабочим буферным раствором поверхность сорбирует еще один слой поверхностно-активного катиона, ориентированного аммонийным концом наружу (сорбция "щетка в щетку"). В результате первая обкладка двойного электрического слоя становится положительно заряженной, а вторая (в том числе и диффузная её часть) приобретает отрицательный заряд, и теперь ЭОП снова перемещается в направлении от входного конца капилляра к детектору.

В системах капиллярного электрофореза наиболее часто применяется фотометрическое детектирование, в котором используется одна или несколько длин волн, обычно лежащих в ультрафиолетовой области спектра. Соответственно отклик детектора будет наблюдаться только в том случае, когда определяемый компонент имеет заметное поглощение на длине волны детектирования. Это - прямое детектирование. Электрофореграмма будет представлять собой набор положительных пиков, возвышающихся над базовой линией.

Однако, анионы, растворенные в воде, зарегистрировать таким простым способом не удастся, т.к. они не обладают собственным поглощением в указанном спектральном диапазоне. В этом случае применяется косвенное детектирование, суть которого состоит в том, что ведущий электролит готовится с добавкой вещества, поглощаю-

щего свет на длине волны детектирования. В случае определения анионов добавка также должна быть анионом, например, это может быть хромат-ион. Вследствие того, что ионная сила ведущего электролита в процессе разделения остается постоянной, в зоне, где находится непоглощающий ион, эквивалентно уменьшается концентрация поглощающего иона. В этом случае на электрофореграмме будут наблюдаться обратные (отрицательные) пики, площади которых пропорциональны концентрациям определяемых ионов. В дальнейшем, при компьютерной обработке результатов измерений, график "переворачивается" и приобретает вид, удобный для рассмотрения, с положительно расположенными пиками. Таким образом, вариант зонного капиллярного электрофореза с модификацией поверхности капилляра и непрямым детектированием позволяет анализировать компоненты, которые в условиях проведения анализа находятся в форме анионов.

На Рис.1.13 приведен пример электрофореграммы смеси пептидов, из которого следует, что метод отличается высоким разрешением при анализе белков.

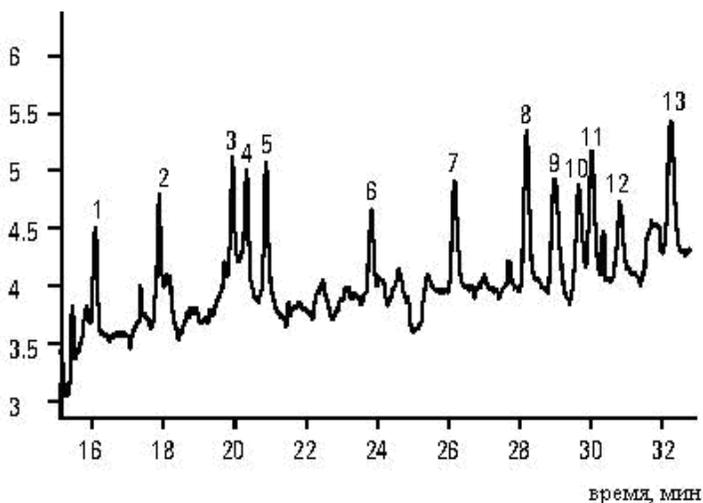


Рис. 1.13 - Электрофореграмма смеси пептидов: 1 – десетилатразин; 2 – гексазион; 3 – метоксурон; 5 – симазин; 6 – цианазин; 7 – метабормурон; 8 - хлортолурон; 9 – изопротурон и атразин; 10 – диурон; 11 – линурон; 12 – метабензтиазурон; 13 – себутилазин; 14 – тетбутилазин; 15 – металахлор. Условия проведения анализа: капилляр длиной 64,5 см и внутренним диаметром 50 мкм; буферный электролит (12 мМ

фосфата, 8 мМ бората, рН 8,9, 100 мМ додецилсульфата натрия);
напряженность поля 440 В/см; детектор УФ, 214 нм

2 Анализ наноматериалов

В главе представлены физические основы экспериментальных методов, являющихся базовыми в исследованиях наноструктур и поверхности твердого тела: оптическая, зондовая, электронная микроскопия и рентгеновская фотоэлектронная спектроскопия, методы структурного анализа по дифракции электронов и рассеянию рентгеновского излучения, инфракрасная, рамановская, оже-электронная и ЯМР спектроскопия, масс-спектрометрия и Мёссбауэровская спектроскопия. Набор этих методов позволяет проводить исследования структурных, электронных и магнитных свойств нанообъектов и поверхности.

2.1 Методы анализа

Анализ свойств, состава и структуры любых материалов предполагает наличие некоторого воздействия на тщательно подготовленный образец и измерение реакции на него с помощью измерительных приборов. В зависимости от типа воздействия и реакции на него со стороны анализируемого объекта, в частности, материала в таблице 2.1 приведён перечень наиболее распространённых видов анализа. [28]. В основу таблицы положена группировка методов по способу внешнего воздействия на исследуемый объект и по типу регистрируемого ответного сигнала: $h\nu$ - электромагнитное излучение; e^- , N^+ - пучки ускоренных электронов и ионов; T-температура. Расшифровка сокращенных названий методов, приведенных в таблице 2.1, дана ниже: РСА - рентгеностурный (или рентгенодифракционный) анализ;

РЭС - рентгеновская эмиссионная спектроскопия;

РАС - рентгеновская абсорбционная спектроскопия;

Таблица 2.1 - Основные физические методы исследования наноструктур и поверхности

		Регистрируемый сигнал			
		$h\nu$	e^-	N^+	Прочие
Воздействие на исследуемый объект	$h\nu$	РСА, РЭС, РАС, ПТСРСП ⁽¹⁾ , ОСРСП ⁽¹⁾	РФЭС, УФЭС, ФЭС, РОЭС		
	e^-	РСМА, ИФЭС	ОЭС, СОЭМ, ДМЭ, ДБЭ, СХПЭЭ, ПЭМ ⁽²⁾ , РЭМ		
	N^+		ИНС	СРМИ, СРБИ, ОРР, ВИМС	
	T, поле, прочее		ПЭМ ⁽³⁾ , СТМ, СТС	ПИМ, ТПД	АСМ, МСМ

Примечания:

1 Ввиду громоздкости русских аббревиатур в литературе обычно используют англоязычные сокращения EXAFS (ПТСРСП) и XANES (ОСРСР).

2 Просвечивающая электронная микроскопия.

3 Полевая эмиссионная микроскопия.

ПТСРСП - протяженная тонкая структура рентгеновского спектра поглощения;

ОСРСР - околопороговая структура рентгеновского спектра поглощения;

РФЭС - рентгеновская фотоэлектронная спектроскопия;

ЭСХА - электронная спектроскопия для химического анализа;

УФЭС - ультрафиолетовая фотоэлектронная спектроскопия;

ФЭС - фотоэлектронная спектроскопия;

РОЭС - возбуждаемая рентгеновским излучением оже-электронная спектроскопия;

РСМА - рентгеноспектральный микроанализ (энергодисперсионный / с волновой дисперсией);

ИФЭС - инверсионная фотоэмиссионная спектроскопия;

ОЭС - оже-электронная спектроскопия;

СОЭМ - сканирующая оже-электронная микроскопия;

ДМЭ - дифракция медленных электронов;

ДБЭ - дифракция быстрых электронов;

СХПЭЭ - спектроскопия характеристических потерь энергии электронами;

ПЭМ - просвечивающая электронная микроскопия.

РЭМ - растровая электронная микроскопия;

ИНС - ионно-нейтрализационная спектроскопия;

СРМИ - спектроскопия рассеяния медленных ионов;

СРБИ - спектроскопия рассеяния быстрых ионов

ОРР - обратное резерфордское рассеяние;

ВИМС - вторично-ионная масс-спектроскопия;

СЗМ - сканирующая зондовая микроскопия;

СТМ - сканирующая туннельная микроскопия;

СТС - сканирующая туннельная спектроскопия;

АСМ - атомно-силовая микроскопия;

МСМ - магнитно-силовая микроскопия;

ПЭМ - полевая эмиссионная микроскопия;

ПИМ - полевая ионная микроскопия;

ТПД - температурно-программируемая десорбция.

В качестве внешнего воздействия на исследуемый объект используется электромагнитное излучение с энергией $h\nu$, пучки

ускоренных до определенной энергии электронов, ионов или нейтральных атомов, температура, электрические и магнитные поля. Сигнал, регистрируемый прибором и несущий информацию о свойствах исследуемого объекта, определяется характером его взаимодействия с выбранным воздействием и лежащими в его основе физическими явлениями. К таким явлениям можно отнести поглощение и эмиссию электромагнитного (в частности, рентгеновского) излучения, фотоэмиссию, эмиссию оже-электронов и ионов, упругое и неупругое рассеяние электронов и ионов, дифракцию рентгеновского излучения и электронов, десорбцию атомов с поверхности, туннелирование электронов через потенциальный барьер между исследуемым объектом и зондом, а также взаимодействие (магнитное, кулоновское, ван-дер-ваальсово) зонда и локального участка поверхности исследуемого объекта.

По способу регистрации ответного сигнала исследуемого объекта на внешнее воздействие физические методы исследования наноструктур и поверхности твердого тела можно разделить на:

- спектроскопические методы, в основе которых лежит регистрация энергетического спектра, поглощенного или испущенного электромагнитного излучения (РЭС, РСМА, ПТСРСП, ОСРСП), электронов (РФЭС, УФЭС, ФЭС, РОЭС, ОЭС, ИОЭС, СХПЭЭ, ИНС) или ионов (СРМИ, СРБИ, ОРР);
- микроскопические методы, в которых наблюдается сформированное определенным образом прямое изображение исследуемого объекта (СТМ, АСМ, МСМ, ПЭМ, РЭМ, СОЭМ, ПИМ, полевая эмиссионная микроскопия);
- дифракционные методы, анализирующие дифракционные картины, образующиеся при прохождении или рассеянии на образце излучения или пучка частиц (РСА, ДМЭ, ДБЭ, ПЭМ в режиме дифракции);
- масс-спектрометрические методы, в которых регистрируется распределение испущенных частиц по массе (ВИМС, МС с бомбардировкой быстрыми атомами, электронно-зондовая МС, ТПД, ЭСД).

Следует отметить, что один и тот же источник воздействия на исследуемый объект может использоваться в различных методах в зависимости от вида регистрируемого сигнала. Например, при облучении образца монохроматическим рентгеновским излучением можно регистрировать картину дифракции рентгеновского излучения на кристаллической решетке образца (метод РСА), спектр поглощения рентгеновского излучения, прошедшего через образец (методы РАС, ПТСРСП, ОСРСП), спектр испущенного образцом

характеристического рентгеновского излучения (РСМА, РЭС), энергетический спектр испущенных образцом фотоэлектронов (РФЭС) и ожс-электронов (РОЭС). Эти спектры несут информацию об элементном и химическом составе образца, его электронной и локальной атомной структуре.

Разные методы позволяют получить различную информацию об исследуемом объекте. Как правило, при анализе наноструктур и поверхности твердого тела наиболее важной является информация об их атомной структуре, элементном составе и электронных свойствах.

Возможности основных методов исследования наноструктур и поверхности с точки зрения получаемой с их помощью информации проиллюстрированы в таблице 2.2.

Таблица 2.2 - Основные методы исследования наноструктур и поверхности

Метод		Атомная структура	Элементный состав	Электронные свойства
Русское название	Английское название			
РСА, ДМЭ/ДБЭ, ПЭМ, РЭМ, ПЭМ, ПИМ, АСМ, МСМ	XRD, LEED/HEED, TEM, SEM, FEM, FIM, AFM, MFM	+	-	-
СТМ/СТС	STM/STS	+	-	+
СРБИ, ОРР	HEIS, RBS	+	+	-
СРМИ	LEIS	(+)	+	(+)
ПТСРСП	EHAFS	+	+	-
ОСРСИ	XANES	(+)	+	+
ОЭС, СОЭМ	AES, SAEM	-/+	+	+
РСМА	EDX/WDX	-	+	-
ВИМС	SIMS	-	+	-
РАС, РЭС	XAS, XES	-	+	(+)
ТПД	TPD	-	+	(+)
РФЭС/ЭСХА ФЭС, РОЭС	XPS/ESCA, PES, XAES	-	+	+
УФЭС, ИФЭС	UPS, IPES	-	-	+
СХПЭЭ	EELS	-	-	+
ИИС	INS	-	-	+

Плюсы в графах таблицы означают, что информацию данного типа можно получить путем стандартного анализа экспериментальных данных. Плюсы в скобках указывают, что экспериментальные данные несут информацию указанного типа, однако ее извлечение либо требует специальной методики анализа полученных данных, либо возможно только для ограниченного набора исследуемых объектов. Отметим, что объем и характер информации, получаемой с помощью того или иного метода, может существенно варьироваться.

Так, метод ДМЭ позволяет определить структуру кристаллической решетки образца, а метод EXAFS чувствителен к локальному атомному окружению; РЭМ используется для исследования морфологии объектов с размерами >10 нм, а СТМ и АСМ позволяют наблюдать отдельные атомы на поверхности твердого тела; метод РСМА используется для определения элементного состава приповерхностной области образца глубиной в единицы микрометров, а метод СРМИ чувствителен к первому атомному слою поверхности.

Кроме того, некоторые методы предъявляют определенные требования к исследуемым образцам. Например, методы СТМ/СТС применимы только для проводящих образцов, МСМ - для исследования магнитных материалов, ПЭМ используется для анализа тонких пленок.

Помимо сокращенных названий методов на русском языке, для справки в таблице 2.2 также приведены названия, используемые в англоязычной литературе.

Из множества разработанных к настоящему времени методов структурного анализа и контроля качества материалов можно выделить несколько наиболее широко используемых: *просвечивающую и растровую электронную микроскопию*, позволяющую получить уникальную визуальную информацию о деталях внутреннего строения различных материалов и их поверхности. Эта информация хорошо воспринимается и является очень удобной для анализа. Методами просвечивающей и растровой электронной микроскопии были выполнены многочисленные работы, связанные с получением новых материалов и разработкой различных технологических процессов. Кроме методов электронной микроскопии хорошо известны *Оже-спектроскопия* (AES: Auger Electron Spectroscopy), *метод дифракции медленных электронов* (LEED: Low Energy Electron Diffraction), *фотоэлектронная спектроскопия - электронная спектроскопия для химического анализа* (ESCA: Electron Spectroscopy for Chemical Analysis),

масс-спектрометрия вторичных ионов (SIMS: Secondary Ion Mass Spectrometry).

При таком многообразии методов важно выбрать адекватный задаче метод и прибор исследования, т.е. такой, который позволяют определить свойства, наиболее интересующие потребителя (конструктора).

Выбор прибора, прежде всего, связан с масштабным уровнем структуры изучаемого объекта, т.е. - с типичными деталями структуры, которые необходимо изучить [29]. Здесь важно использовать прибор, имеющий соответствующую *разрешающую способность*. Она характеризует минимальное расстояние между двумя соседними деталями структуры объекта, которые еще могут быть различимы.

Разрешение прибора определяется длиной волны излучения, а также особенностями взаимодействия излучения с исследуемым материалом. С целью увеличения разрешающей способности метода необходимо уменьшать длину волны и увеличивать угол приема сигнала, т.е. апертуру прибора.

В настоящее время информацию об атомной структуре поверхности позволяют получить *ионно-полевой, сканирующий туннельный и атомно-силовой микроскопы*

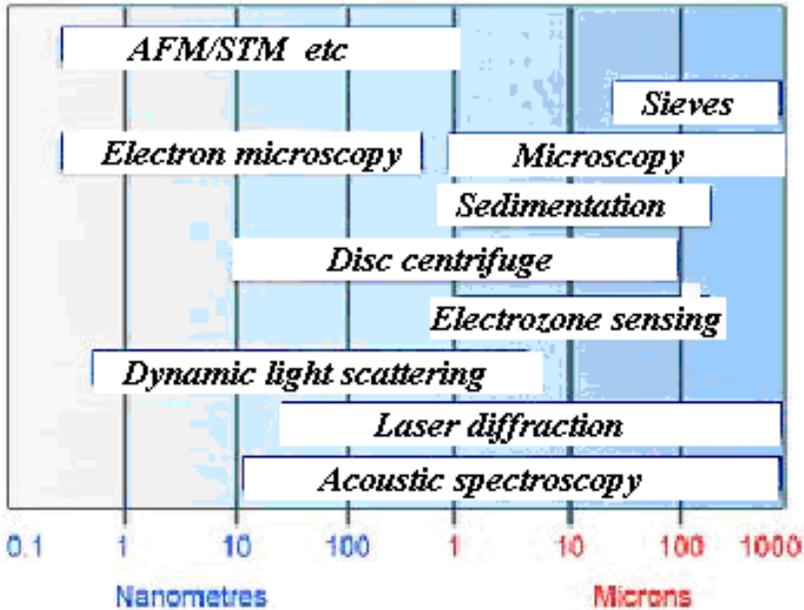


Рис. 2.1 Методы измерения размеров микро- и наночастиц

AFM/STM – Сканирующие зондовые микроскопы (Атомные силовые/
Сканирующие туннельные);
Electron microscopy – Электронные микроскопы;
Dynamic light scattering – Динамическое светорассеяние
Disc centrifuge – Центрифуга (дисковая)
Acoustic spectroscopy – Акустическая спектроскопия
Laser diffraction – Лазерная дифракция
Electrozone sensing – Электронное зондирование
Sedimentation – Седиментация
Microscopy – Микроскопия
Sieves – Фильтрация (просеивание)

В определенном смысле можно полагать, что эти методы позволяют «увидеть» атомы. Поэтому естественно, что каждый из видов анализа предназначен для решения задач измерения в своём диапазоне линейных размеров в зависимости от его разрешающей способности (см. Рис. 2.1).

Все широко применяемые методы анализа и/или диагностики наноструктур удобно разбить на четыре группы:

1. Методы изображения (оптическая микроскопия, зондовая микроскопия, электронная микроскопия, рентгеновская микроскопия)
2. Методы структурного анализа (дифракция электронов, дифракция рентгеновских лучей)
3. Спектроскопия (инфракрасная спектроскопия, рамановская спектроскопия, фотоэмиссионная (ФЭС, Оже, ...), магнитная (ЭПР, ЯМР))
4. Масс-спектрометрия.

Далее рассмотрим более подробно каждую из этих групп

Методы изображения

Группа приборов, формирующих изображение исследуемого объекта, по традиции называют *микроскопами*. Современные микроскопы позволяют не только строить рельеф поверхности анализируемого объекта, но и, при надлежащей калибровке, измерять размеры этого рельефа и определять закономерности расположения наноструктур. На Рис. 2.2 приведены наиболее распространённые типы микроскопов. На этом рисунке введены следующие обозначения, аббревиатуры и сокращения: Optical – оптические микроскопы; TEM – трансмиссионные электронные микроскопы; SEM – сканирующие электронные микроскопы; SPM – сканирующие зондовые микроскопы; AFM – атомный силовой микроскоп; STM – сканирующий туннельный микроскоп; SNOM – сканирующий ближнепольный микроскоп.

Microscopes

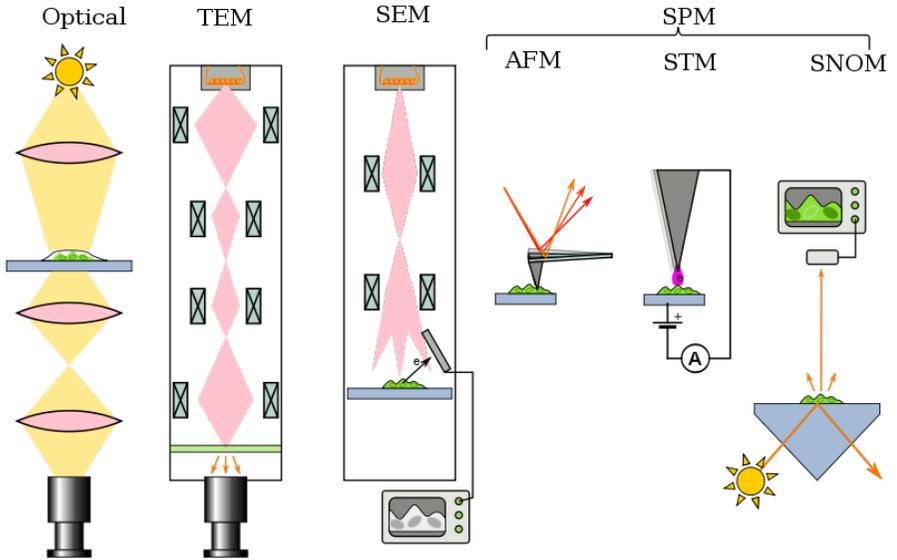


Рис. 2.2 - Разновидности микроскопов.

2.2 Оптические микроскопы

Одним из претендентов на изобретение микроскопа является Галилео Галилей (1609 г.). Этот микроскоп потом усовершенствовали голландцы Корнелиус Дреbbель и Кристиан Гюйгенс. Окуляры Гюйгенса производятся и по сей день.

Важнейшей характеристикой оптического микроскопа является его разрешающая способность (*resolution*) - способность формировать чёткое раздельное изображение двух близких по интенсивности точечных объектов.. Степень проникновения в микромир, возможности его изучения зависят от разрешающей способности прибора. Эта характеристика определяется, прежде всего, длиной волны используемого в микроскопии излучения (видимое, ультрафиолетовое, рентгеновское излучение). Фундаментальное ограничение заключается в невозможности получить при помощи электромагнитного излучения изображение объекта, меньшего по размерам, чем длина волны этого излучения. «Проникнуть глубже» в микромир возможно при применении излучений с меньшими длинами волн. Из вида дифракционной картины распределения интенсивности в фокальной плоскости следует, что разрешение будет определяться степенью перекрытия центральных

пятен распределений двух точечных объектов. Эти пятна называются дисками Эйри. Рэлеем был предложен критерий, согласно которому две точки считаются разрешенными, когда на изображении максимум интенсивности от 1-ой точки попадает на минимум от 2-ой.

При этом величина "провала" в интенсивности по центру между изображениями точек составила 26% от максимума, а расстояние между разрешаемыми точками равно радиусу диска Эйри (Рис. 2.3).

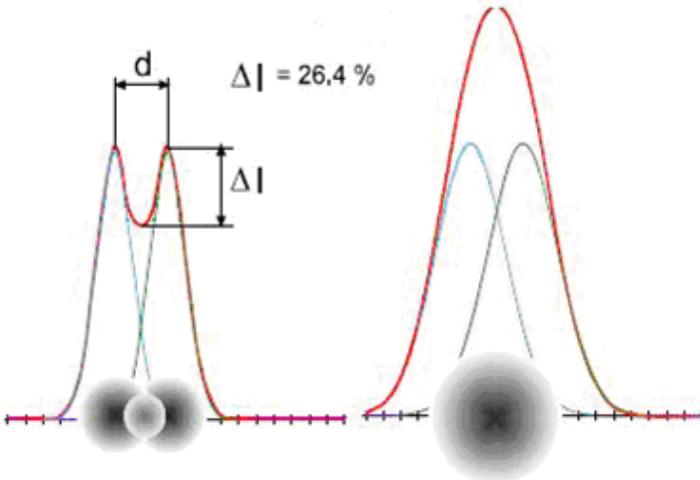


Рис. 2.3 - Предел оптического разрешения микроскопа.

Впервые предел разрешения объектива микроскопа был определен немецким физиком Г. Гельмгольцем на основании критерия Рэля. Формула Гельмгольца имеет вид:

$$d = 0.61 \lambda / NA,$$

где λ – длина волны, величина $NA = n \sin \alpha$ называется числовой апертурой объектива, n – показатель преломления иммерсионной жидкости, α – апертурный угол. У хороших объективов апертурный угол α близок к своему пределу: $\alpha \approx \pi/2$.

Как видно из формулы Гельмгольца, применение иммерсии несколько улучшает предел разрешения. Полагая для оценок $\sin \alpha \approx 1$, $n \approx 1.5$, получим: $d \approx 0.4 \lambda$.

Таким образом, с помощью микроскопа принципиально невозможно рассмотреть какие-либо детали, размер которых значительно меньше длины волны света. Волновые свойства света определяют предел качества изображения объекта, полученного с помощью любой оптической системы.

Один из недостатков обычного светового микроскопа – внефокусные лучи, которые снижают контраст изображения. Если в оптическую схему микроскопа ввести специальную диафрагму, расположенную в плоскости промежуточного изображения, то она пропустит только те световые лучи, которые исходят из очень небольшой области объекта: рассеянный свет, идущий из глубины образца, отсекается апертурами, что обеспечивает высокий контраст изображения. Диафрагма будет играть роль пространственного фильтра. Чем меньше диаметр диафрагмы, тем меньше размеры этой области. Мы получим, таким образом, конфокальный микроскоп (Рис. 2.4)

Полное изображение объекта в конфокальном микроскопе формируется при последовательном просмотре элементарных объемов с применением различных сканирующих систем. Накопление информации происходит либо благодаря свойству инерционности зрения при быстром сканировании, либо посредством использования фотоприемников и электронных запоминающих устройств. Таким образом, конфокальная микроскопия обеспечивает увеличение контраста изображения за счет фильтрации внефокусных лучей (Рис. 2.5).

Конфокальный микроскоп имеет разрешение такое же, как и обычный микроскоп и ограничено оно дифракционным пределом. В видимом диапазоне разрешение составляет ~ 250 нм ($NA=1,45$, $n=1,51$).

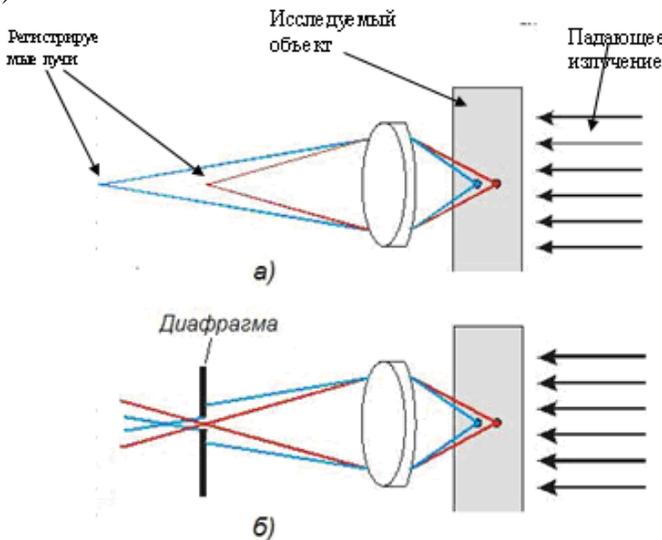
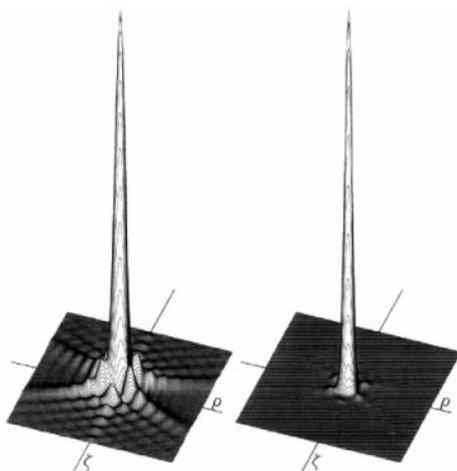


Рис. 2.4 -Оптические микроскопы: а) классический; б) конфокальный (часть лучей, создающая фон, не регистрируется)

Однако, в последние годы успешно развиваются схемы микроскопов, которые используют нелинейные свойства флуоресценции образцов. В этом случае достигается разрешение значительно меньшее дифракционного предела и составляет $\sim 3\text{—}10$ нм.



а)

б)

Рис. 2.5- Сравнение контраста в классическом -а) и конфокальном- б) микроскопах.

Конфокальный микроскоп создаёт чёткое изображение образца, которое при использовании обычного микроскопа представляется размытым.

Конфокальная микроскопия обеспечивает:

- увеличение контраста исследуемого объекта.
- получение «оптических срезов», трехмерная реконструкция.
- мультиспектральные исследования с высокой степенью разделения сигналов от разных флуорохромов.

В конфокальной *лазерной* сканирующей микроскопии несколько улучшается разрешение в латеральной плоскости (~ 1.5 раз в плоскости объекта) и достигается хорошее аксиальное разрешение.

Платой за полученные улучшения является необходимость применения схем сканирования, либо путем перемещения образца, либо путем перестройки оптической системы. Применение сканирования позволяет увеличить поле зрения по сравнению с обычными оптическими микроскопами.

2.3 Зондовые микроскопы

Сканирующие зондовые микроскопы (СЗМ, англ. *SPM — Scanning Probe Microscope*) — класс микроскопов для получения изображения поверхности и её локальных характеристик. Процесс построения изображения основан на сканировании поверхности зондом. В общем случае позволяет получить трёхмерное изображение поверхности (топографию) с высоким разрешением [30, 31, 32, 33].

Основные типы сканирующих зондовых микроскопов (рис. 2.6):

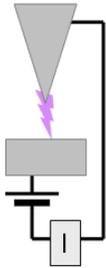
- Сканирующий туннельный микроскоп
- Сканирующий атомно-силовой микроскоп
- Ближнепольный оптический микроскоп

Отличительной особенностью СЗМ является наличие зонда, системы перемещения зонда относительно образца по 2-м (X-Y) или 3-м (X-Y-Z) координатам, регистрирующей системы.

Регистрирующая система фиксирует значение функции, зависящей от расстояния зонд-образец. Обычно регистрируемое значение обрабатывается системой отрицательной обратной связи, которая управляет положением образца или зонда по одной из координат (Z). В качестве системы обратной связи чаще всего используется ПИД-регулятор.

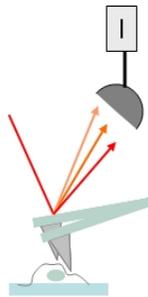
Scanning Probe Microscopy (SPM)

Scanning Tunneling
Microscopy (STM)



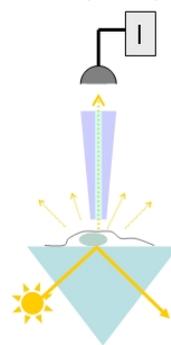
а)

Atomic Force
Microscopy (AFM)



б)

Scanning Near-field
Optical Microscopy
(SNOM)



в)

Рис. 2.6 - Основные типы сканирующих зондовых микроскопов

пов

а). Сканирующий туннельный микроскоп (СТМ) — для получения изображения используется туннельный ток между зондом и об-

разцом, что позволяет получить информацию о топографии и электрических свойствах образца.

б) Атомно-силовой микроскоп (АСМ) — регистрирует различные силы между зондом и образцом. Позволяет получить топографию поверхности и её механические свойства.

в) Сканирующий ближнепольный микроскоп (СБОМ) — для получения изображения используется эффект ближнего поля.

Работа *сканирующего зондового микроскопа* основана на взаимодействии поверхности образца с зондом (кантилевер, игла или оптический зонд). При малом расстоянии между поверхностью и зондом действие сил взаимодействия (отталкивания, притяжения, и других сил) и проявление различных эффектов (например, туннелирование электронов) можно зафиксировать с помощью современных средств регистрации.

Для регистрации используют различные типы сенсоров, чувствительность которых позволяет зафиксировать малые по величине возмущения. Для получения полноценного растрового изображения используют различные устройства развертки по осям X и Y (например, пьезотрубки, плоскопараллельные сканеры).

Следует отметить следующие основные технические сложности при создании сканирующего зондового микроскопа:

- Конец зонда должен иметь размеры сопоставимые с исследуемыми объектами.
- Обеспечение механической (в том числе тепловой и вибрационной) стабильности на уровне лучше 0,1 ангстрема.
- Детекторы должны надежно фиксировать малые по величине возмущения регистрируемого параметра.
- Создание прецизионной системы развертки.
- Обеспечение плавного сближения зонда с поверхностью.

В настоящий момент, в большинстве исследовательских лабораторий сканирующая зондовая и электронная микроскопия используются как дополняющие друг друга методы исследования в силу ряда физических и технических эффектов: измерособенностей. Сканирующие зондовые микроскопы нашли применение практически во всех областях науки. В физике, химии, биологии используют в качестве инструмента исследования СЗМ. В частности, такие междисциплинарные науки, как материаловедение, биохимия, фармацевтика, нанотехнологии, физика и химия поверхности, электрохимия, исследование коррозии, электроника (например, МЭМС), фотохимия и многие другие. Перспективным направлением считается совмещение сканирующих зондовых микроскопов с другими традиционными и современными методами исследованиями, а также создание принципиально

новых приборов. Например, совмещение СЗМ с оптическими микроскопами (традиционными и конфокальными микроскопами), электронными микроскопами, спектрометрами (например, спектрометрами комбинационного (рамановского) рассеяния и флюоресцентными), ультрамикротомами.

Сканирующие туннельные микроскопы (СТМ, англ. STM — scanning tunneling microscope) — вариант сканирующего зондового микроскопа, предназначенный для измерения рельефа проводящих поверхностей с высоким пространственным разрешением (Рис. 2.7).

В СТМ острая металлическая игла подводится к образцу на расстояние нескольких ангстрем. При подаче на иглу относительно образца небольшого потенциала возникает туннельный ток.

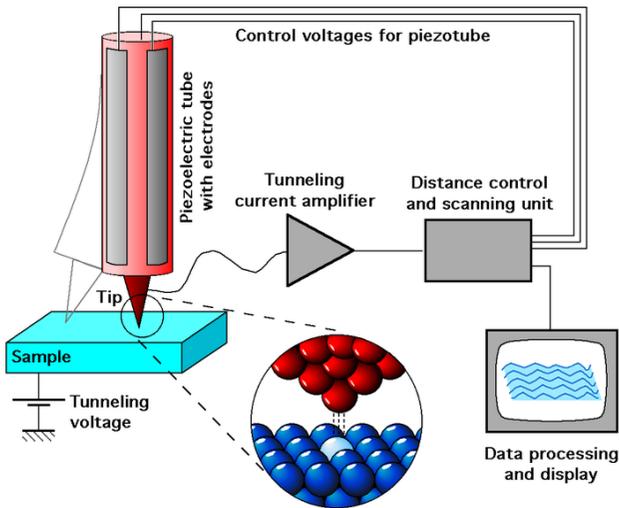


Рис. 2.7 - Сканирующий туннельный микроскоп: схема работы

В процессе сканирования игла движется вдоль поверхности образца, туннельный ток поддерживается стабильным за счёт действия обратной связи, и показания следящей системы меняются в зависимости от топографии поверхности. Такие изменения фиксируются, и на их основе строится карта высот. Другая методика предполагает движение иглы на фиксированной высоте над поверхностью образца. В этом случае фиксируется изменение величины туннельного тока и на основе данной информации идет построение топографии поверхности.

Ограничения на использование метода накладываются, во-первых, условием проводимости образца (поверхностное сопротивление должно быть не больше 20 МОм/см^2), во-вторых, условием «глу-

бина канавки должна быть меньше её ширины», потому что в противном случае может наблюдаться туннелирование с боковых поверхностей. Но это только основные ограничения. На самом деле их намного больше. Например, технология заточки иглы не может гарантировать одного острия на конце иглы, а это может приводить к параллельному сканированию двух разновысотных участков. Кроме ситуации глубокого вакуума, во всех остальных случаях мы имеем на поверхности осаждённые из воздуха частицы, газы и т.д. Технология грубого сближения также оказывает колоссальное влияние на действительность полученных результатов.

Если при подводе иглы к образцу мы не смогли избежать удара иглы о поверхность, то считать иглу состоящей из одного атома на кончике пирамиды будет большим преувеличением.

Атомно-силовой микроскоп (АСМ, англ. *AFM* — *atomic-force microscope*) — сканирующий зондовый микроскоп высокого разрешения. Используется для определения рельефа поверхности с разрешением от десятков ангстрем вплоть до атомарного. В отличие от сканирующего туннельного микроскопа, с помощью атомно-силового микроскопа можно исследовать как проводящие, так и непроводящие поверхности.

Атомно-силовой микроскоп был изобретён в 1986 году Герхардом Биннигом, Калвином Куэйтом и Кристофером Гербером на основе идей, заложенных Герхардом Биннигом и Хайнрихом Рорером (Нобелевская премия по физике, 1986 год). С помощью атомно-силового микроскопа можно получать изображения, как физических объектов, так и биологических и химических объектов (вирусов и бактерий, атомов и молекул). Разрешение таких микроскопов достигает доли нанометров, что позволяет наблюдать атомы! Получением изображений не ограничиваются возможности этого прибора. С помощью атомно-силового микроскопа можно изучать взаимодействие двух объектов, силы трения, упругости, адгезии, а также перемещать отдельные атомы, осаждать и удалять их с какой-либо поверхности

Зонд (остриё, игла) расположен на свободном конце кантилевера (или консоли, Рис. 2.8). Зонды в основном изготавливают из таких материалов, как кремний Si и Si_3N_4 . Чем меньше радиус кривизны и угол схождения острия, тем меньше его влияние на получающееся изображение исследуемого объекта. Консоль – это упругая пластинка, по отклонению которой в принципе можно судить о силе взаимодействия острия с образцом (закон Гука: $F=kz$).

Принцип работы атомно-силового микроскопа основан на регистрации силового взаимодействия между поверхностью исследуемого образца и зондом. В качестве зонда используется **наноразмерное**

остриё, располагающееся на конце упругой консоли, называемой кантилевером.

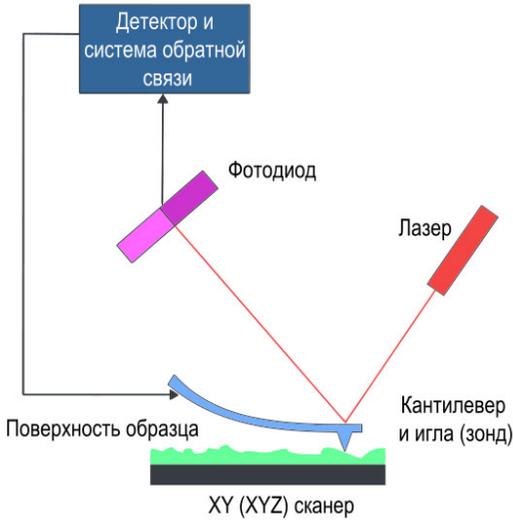


Рис. 2.8 - Атомно-силовой микроскоп: схема работы

Сила, действующая на зонд со стороны поверхности, приводит к изгибу консоли. Появление возвышенностей или впадин под остриём приводит к изменению силы, действующей на зонд, а значит, и изменению величины изгиба кантилевера.

Регистрируя величину изгиба, можно сделать вывод о рельефе поверхности. Под силами, действующими между зондом и образцом, в первую очередь подразумевают дальнедействующие силы Ван-дер-Ваальса (силы межмолекулярного взаимодействия), которые сначала являются силами притяжения, а при дальнейшем сближении переходят в силы отталкивания (Рис. 2.9). В зависимости от характера действия силы между кантилевером и поверхностью образца выделяют три режима работы атомно-силового микроскопа:

1. Контактный (англ. *contact mode*)
2. «Полуконтактный» (англ. *semi-contact mode* или *tapping mode*)
3. Бесконтактный (англ. *non-contact mode*)

Здесь необходимо пояснить, что именно берётся за ноль расстояния во избежание путаницы. На приведённом рисунке ноль соответствует нулевому расстоянию между ядрами атома на поверхности и наиболее выступающего атома кантилевера. Поэтому ноль силы нахо-

дится на конечном расстоянии, соответствующем границе электронных оболочек этих атомов (при перекрытии оболочек возникает отталкивание). Если взять за ноль границы атомов, то сила обратится в ноль в нуле расстояния. При работе в *контактном режиме* атомно-силовой микроскоп является аналогом профилометра. Остриё кантилевера находится в непосредственном контакте между образцом и поверхностью (Рис. 2.9).



Рис. 2.9.-График зависимости силы Ван-дер-Ваальса от расстояния между кантилевером и поверхностью образца

Сканирование осуществляется, как правило, в режиме постоянной силы, когда система обратной связи поддерживает постоянной величину изгиба кантилевера. При исследовании образцов перепадами высот порядка единиц ангстрем возможно применять режим сканирования при постоянном среднем расстоянии между зондом и поверхностью образца. В этом случае кантилевер движется на некоторой средней высоте над образцом. Изгиб консоли ΔZ , пропорциональный силе, действующей на зонд со стороны поверхности записывается для каждой точки. Изображение в таком режиме представляет собой пространственное распределение силы взаимодействия зонда с поверхностью. Достоинства метода:

- Наибольшая, по сравнению с другими методами, помехоустойчивость

Недостатки метода:

- Возможно механическое повреждение, как зонда, так и образца

- Практически не пригоден для изучения объектов с малой механической жёсткостью (органические материалы, биологические объекты)

- Наименьшее латеральное разрешение

При работе в *бесконтактном режиме* пьезовибратором возбуждаются колебания зонда на некоторой частоте (чаще всего, резонансной). Сила, действующая со стороны поверхности, приводит к сдвигу амплитудно-частотной и фазово-частотной характеристик зонда, и амплитуда и фаза изменяют значения. Система обратной связи, как правило, поддерживает постоянной амплитуду колебаний зонда, а изменение частоты и фазы в каждой точке записывается. Однако возможно установление обратной связи -путём поддержания постоянной величины частоты или фазы колебаний.

Достоинства метода:

- Возможность достижения атомарного разрешения
- Обеспечивает наилучшую сохранность зонда и образца

Недостатки метода:

- Крайне чувствителен ко всем внешним шумам
- Попадание на кантилевер во время сканирования частички с поверхности образца меняет его частотные свойства и настройки сканирования "уходят"

Полуконтактный режим работы атомно-силового микроскопа. При работе в полуконтактном режиме также возбуждаются колебания кантилевера. В нижнем полупериоде колебаний кантилевер касается поверхности образца. Такой метод является промежуточным между полным контактом и полным бесконтактом.

Несмотря на то, что при описании работы атомно-силового микроскопа, очень часто упоминаются лишь силы Ван-дер-Ваальса, в реальности со стороны поверхности также действуют упругие силы и силы адгезии. Их вклад особенно очевиден при работе в полуконтактном режиме, когда вследствие "прилипания" кантилевера к поверхности возникает гистерезис которые могут существенно усложнять процесс получения изображения и интерпретацию результатов.

Конструкция атомно-силового микроскопа. Основными конструктивными составляющими атомно-силового микроскопа являются:

- Жёсткий каркас, удерживающий систему
- Держатель образца, на котором образец впоследствии закрепляется
- Устройства манипуляции. В зависимости от конструкции микроскопа возможно движение зонда относительно неподвижного образца или движение образца, относительно закреплённого зонда.

Манипуляторы делятся на две группы. Первая группа предназначена для "грубого" регулирования расстояния между кантилевером и образцом (диапазон движения порядка сантиметров), вторая - для прецизионного перемещения в процессе сканирования (диапазон движения порядка микрон). В качестве прецизионных манипуляторов (или сканеров) используются элементы из пьезокерамики. Они способны осуществлять перемещения на расстояния порядка ангстрем, однако им присущи такие недостатки, как термодрейф, нелинейность, гистерезис, крип.

- Зонд
- Система регистрации отклонения зонда

АСМ включает несколько систем:

- Оптическая (включает лазер и фотодиод, наиболее распространённая)
- Интерферометрическая (состоит из лазера и оптоволоконна)
- Ёмкостная (измеряется изменение ёмкости между кантилевером и расположенной выше неподвижной пластиной)
- Туннельная (исторически первая, регистрирует изменение туннельного тока между проводящим кантилевером и расположенной выше туннельной иглой)
- Система обратной связи
- Управляющий блок с электроникой

Особенности работы. В сравнении с растровым электронным микроскопом (РЭМ) атомно-силовой микроскоп обладает рядом преимуществ. Так, в отличие от РЭМ, который даёт псевдотрёхмерное изображение поверхности образца, АСМ позволяет получить истинно трёхмерный рельеф поверхности. Кроме того, непроводящая поверхность, рассматриваемая с помощью АСМ, не требует нанесения проводящего металлического покрытия, которое часто приводит к заметной деформации поверхности. Для нормальной работы РЭМ требуется вакуум, в то время как большинство режимов АСМ могут быть реализованы на воздухе или даже в жидкости. Данное обстоятельство открывает возможность изучения биомакромолекул и живых клеток. В принципе, АСМ способен дать более высокое разрешение, чем РЭМ. Так было показано, что АСМ в состоянии обеспечить реальное атомное разрешение в условиях сверхвысокого вакуума. Сверхвысоковакуумный АСМ по разрешению сравним со сканирующим туннельным микроскопом и просвечивающим электронным микроскопом.

К недостатку АСМ при его сравнении с РЭМ также следует отнести небольшой размер поля сканирования. РЭМ в состоянии просканировать область поверхности размером в несколько миллиметров в латеральной плоскости с перепадом высот в несколько миллиметров

в вертикальной плоскости. У АСМ максимальный перепад высот составляет несколько микрон, а максимальное поле сканирования в лучшем случае порядка 150×150 микрон². Другая проблема заключается в том, что при высоком разрешении качество изображения определяется радиусом кривизны кончика зонда, что при неправильном выборе зонда приводит к появлению артефактов на получаемом изображении.

Обычный АСМ не в состоянии сканировать поверхность также быстро, как это делает РЭМ. Для получения АСМ-изображения, требуется от нескольких минут до нескольких часов, в то время как РЭМ после откачки способен работать практически в реальном масштабе времени, хотя и с относительно невысоким качеством. Из-за низкой скорости развёртки АСМ получаемые изображения оказываются искажёнными тепловым дрейфом, что уменьшает точность измерения элементов сканируемого рельефа. Для увеличения быстродействия АСМ было предложено несколько конструкций, среди которых можно выделить зондовый микроскоп, названный видео АСМ. Видео АСМ обеспечивает получение удовлетворительного качества изображений поверхности с частотой телевизионной развёртки, что даже быстрее, чем на обычном РЭМ. Для коррекции вносимых термодрейфом искажений было предложено несколько методов.

Кроме термодрейфа АСМ - изображения могут быть также искажены из-за таких свойств пьезокерамики, как нелинейность, крип и гистерезис и перекрёстными паразитными связями, действующими между X, Y, Z-элементами сканера.

Для исправления искажений в реальном масштабе времени современные АСМ используют программное обеспечение (например, особенность-ориентированное сканирование) либо сканеры, снабжённые замкнутыми следящими системами, в состав которых входят линейные датчики положения.

Некоторые АСМ вместо сканера в виде пьезотрубки используют XY и Z-элементы, механически несвязанные друг с другом, что позволяет исключить часть паразитных связей.

АСМ можно использовать для определения типа атома в кристаллической решётке.

Обработка информации и восстановление полученных изображений. Как правило, снятое на сканирующем зондовом микроскопе изображение трудно поддается расшифровке из-за присущих данному методу искажений.

Практически всегда результаты первоначального сканирования подвергаются математической обработке. Для этого используется программное обеспечение непосредственно поставляемое с СЗМ. Суще-

ствует и программное обеспечение распространяемое по GNU лицензии, например, Gwyddion.

Интересные следствия. Манипулятор АСМ и СТМ позволяет при габаритах в несколько сантиметров передвигать иглу с разрешением лучше $0,1 \text{ \AA}$.

Если бы промышленный робот обладал подобной точностью перемещений при габаритах около метра, то иголкой, зажатой в манипуляторах, он мог бы нарисовать окружность диаметром в несколько нанометров.

Температурный коэффициент линейного расширения большинства материалов около 10^{-6} .

При размерах манипулятора в несколько сантиметров изменение температуры на $0,01^\circ$ приводит к перемещению иглы вследствие теплового дрейфа на 1 \AA .

Ближнепольный оптический микроскоп (БОМ) — оптическая микроскопия, обеспечивающая разрешение лучшее, чем у обычного микроскопа.

Повышение разрешения БОМа достигается детектированием рассеяния света от изучаемого объекта на расстояниях меньших, чем длина волны света.

В случае, если зонд (детектор) микроскопа ближнего поля снабжен устройством пространственного сканирования, то такой прибор называют сканирующим оптическим микроскопом ближнего поля.

Такой микроскоп позволяет получать растровые изображения поверхностей и объектов с разрешением ниже дифракционного предела.

Идея БОМа была предложена в 1928 году Сингхом (Е.Н. Syngh), но она намного опередила технические возможности своего времени и осталась практически не замеченной. Ее первое подтверждение было получено Эшем (Е.А. Ash) в опытах с микроволнами в 1972 году. В начале 80-х годов группа исследователей из Цюрихской лаборатории фирмы IBM во главе с Дитером Полем (D.W. Pohl) проникла внутрь дифракционного предела и продемонстрировала разрешение $\lambda/20$ на приборе, работающем в видимом оптическом диапазоне и получившем название сканирующего оптического микроскопа ближнего поля.

Чуть раньше в той же лаборатории был создан первый сканирующий туннельный микроскоп, принесший ей всемирную известность.

Если в качестве зонда взять миниатюрную диафрагму с отверстием в несколько нанометров - апертуру, то в соответствии с законами волновой оптики, видимый свет (с длиной волны несколько сот

нанометров) проникает в такое маленькое отверстие, но не далеко, а на расстояние, сопоставимое с размерами отверстия.

Если в пределах этого расстояния, в так называемом «ближнем поле», поставить образец, рассеянный от него свет будет регистрироваться.

Перемещая диафрагму в непосредственной близости от образца, как в туннельном микроскопе, получим растровое изображение поверхности. Позднее были разработаны ближнепольные микроскопы не использующие апертуру - безапертурный СБОМ.

Уникальность ближнепольной оптической микроскопии по сравнению с другими сканирующими методами состоит в том, что изображение строится непосредственно в оптическом диапазоне, в том числе видимого света, однако разрешение многократно превышает разрешение традиционных оптических систем.

Традиционные методы получения оптических изображений объектов имеют существенные ограничения, связанные с дифракцией света. Одним из основополагающих законов оптики является существование так называемого дифракционного предела, который устанавливает минимальный размер объекта R , изображение которого может быть построено оптической системой при использовании света с длиной волны λ (2.14):

$$d_{\min} \approx \frac{\lambda}{2n}$$

где n – показатель преломления среды.

Для оптического диапазона длин волн предельный размер составляет величину порядка 200–300 нм.

В ближнепольной оптической микроскопии используются другие принципы построения изображения объекта, которые позволяют преодолеть трудности, связанные с дифракцией света, и реализовать пространственное разрешение на уровне 10 нм и лучше.

Ближнепольный оптический микроскоп (БОМ) был изобретен Дитером Полем (лаборатория фирмы IBM, г. Цюрих, Швейцария) в 1982 году сразу вслед за изобретением туннельного микроскопа.

В основе работы данного прибора используется явление прохождения света через **субволновые диафрагмы** (отверстия с диаметром много меньше длины волны падающего излучения) (Рис. 2.10).

Использование дифракции в области ближнего поля позволяет получить с помощью оптической системы разрешение, превышающее дифракционный предел.

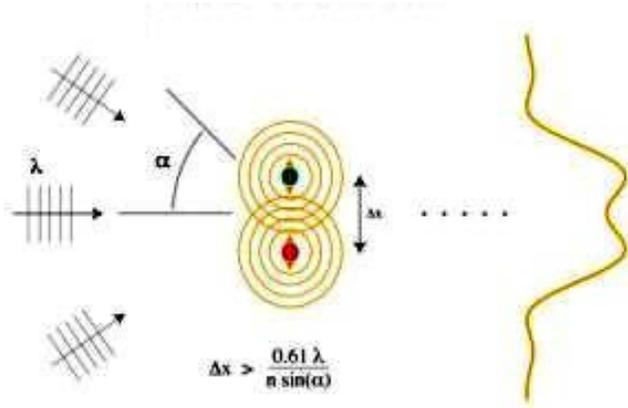


Рис. 2.10 - Два точечных объекта в оптическом микроскопе. Разрешение здесь ограничено дифракцией – правило Рэля

При прохождении света через субволновое отверстие наблюдается ряд особенностей. Электромагнитное поле в области диафрагмы имеет сложную структуру (Рис.2.11).

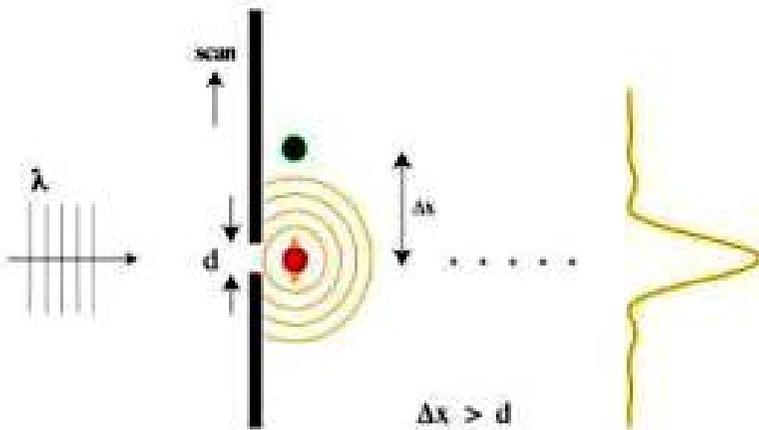


Рис. 2.11 - Два точечных объекта в ближнепольном оптическом микроскопе. В результате дифракции свет проникает через отверстие, которое может иметь диаметр меньше длины волны.

Непосредственно за отверстием на расстояниях $h < 100 \text{ \AA}$ располагается так называемая **ближняя зона**, в которой электромагнитное поле существует, в основном, в виде эванесцентных (не распространяющихся) мод, локализованных вблизи поверхности диафрагмы. В области расстояний $h > 100 \text{ \AA}$ располагается **дальняя зона**, в которой наблюдаются лишь излучательные моды. Мощность излучения за субволновой диафрагмой в дальней зоне может быть оценена по следующей формуле:

$$P = \frac{2}{47} k^4 d^6 W_o,$$

где k – волновой вектор, d – диаметр диафрагмы, W_o – плотность мощности падающего излучения.

Оценки показывают, что для излучения с длиной волны порядка 500 нм и диафрагмы с отверстием ~ 5 нм мощность излучения в дальней зоне составляет по порядку величин 10^{-10} от мощности падающего излучения. Поэтому, на первый взгляд, кажется, что использование малых отверстий для построения растровых оптических изображений исследуемых образцов практически невозможно. Однако, если поместить исследуемый объект непосредственно за отверстием в ближней зоне, то вследствие взаимодействия эванесцентных мод с образцом часть энергии электромагнитного поля переходит в излучательные моды, интенсивность которых может быть зарегистрирована оптическим фотоприемником.

Таким образом, ближнепольное изображение формируется при сканировании исследуемого образца диафрагмой с субволновым отверстием и регистрируется в виде распределения интенсивности оптического излучения в зависимости от положения диафрагмы $I(x, y)$.

Контраст на БОМ изображениях определяется процессами отражения, преломления, поглощения и рассеяния света, которые, в свою очередь, зависят от локальных оптических свойств образца.

Существует два способа локализации электромагнитного поля: апертурный и безапертурный.

Размер апертуры и материал покрытия варьируются в зависимости от методики и требуемого разрешения. Как правило, диаметр апертуры 50-100 нм, и в качестве покрытия используется алюминий или серебро. При этом апертура должна располагаться на расстоянии от поверхности меньшем, чем длина волны ($h \ll \lambda$). Такой способ позволяет использовать различные варианты детектирования сигнала (Рис. 2.12).

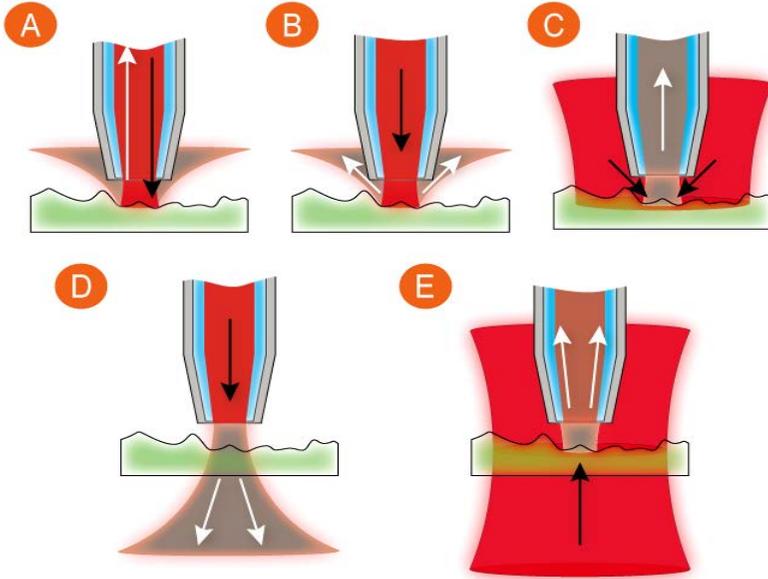


Рис. 2.12- Несколько методик освещения и/или детектирования

A - падающее и отраженное излучение идут по одному и тому же оптическому волокну.

B - падающее излучение идет по оптическому волокну зонда, детектируется отраженное от поверхности образца излучение

C - для освещения образца используется внешний источник лазерного излучения, для сбора отраженного излучения используется зонд.

D - для подвода излучения к образцу используется зонд, детектируется прошедшее через образец излучение. Методика применима только к прозрачным образцам.

E - для подвода излучения к образцу используется внешний источник лазерного излучения, детектируется излучение прошедшее через образец.

Методика применима только к прозрачным образцам.

2) Для локализации излучения *используется иголка зонда*, поднесенная к освещенной поверхности на расстояние меньше длины волны (Рис. 2.13). В таком режиме острие рассеивает (превращает в дальнее) ближнее поле, локализованное у поверхности образца.

Все ближнепольные микроскопы, кроме зонда, включают несколько базовых элементов конструкции (Рис. 2.13).

- система перемещения зонда относительно поверхности образца по 2-м (X-Y) или 3-м (X-Y-Z) координатам (система развертки);
- регистрирующая система;
- оптическая система.

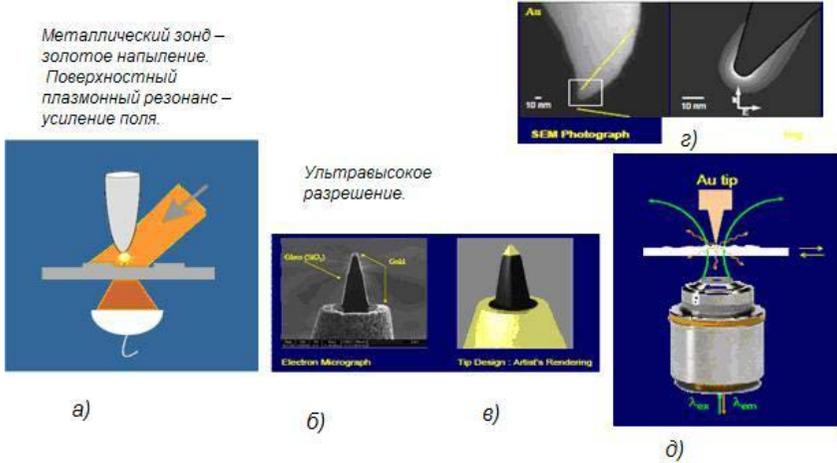


Рис. 2.13 - Безапертурная микроскопия

- | | |
|---------------------------|--|
| а) принцип действия | г) распределение светового поля на зонде |
| б), в) конструкции зондов | д) техническая реализация |

Система развертки. Независимо от конструкции оптический зонд позволяет получить сигнал исключительно в одной точке. Для построения растровых изображений необходима система перемещения зонда относительно поверхности по 2-м (X-Y) или 3-м (X-Y-Z) координатам (система развертки). В качестве системы развертки используют сканеры, аналогичные сканерам туннельных и атомно-силовых микроскопов. Т.е., сканеры на пьезотрубках или плоско-параллельные сканеры.

Регистрирующая система. Для регистрации изменения расстояния между зондом и поверхностью образца (h) используют те же методы, что и в атомно-силовой микроскопии. Наибольшее распространение получили метод “shear-force” и регистрация положения зонда с использованием дефлектометра.

Метод “shear-force” основан на регистрации изменения частоты колебаний зонда, вызванных тангенциальной составляющей сил взаи-

модействия зонд-поверхность. Для определения изменения частоты колебаний используют резонаторы камертонного типа (tuning-fork), колебания с которых детектируют с помощью прямого пьезоэлектрического эффекта кристаллов кварца.

Оптическая система. Для возбуждения и детектирования оптического сигнала используется оптический блок, включающий в себя источник лазерного излучения, поляризаторы и детекторы сигнала (лавинные фотодиоды, ФЭУ или ПЗС матрицы) (Рис.2.14). Конфигурация оптического блока подбирается под задачу и бюджет исследовательских лабораторий.

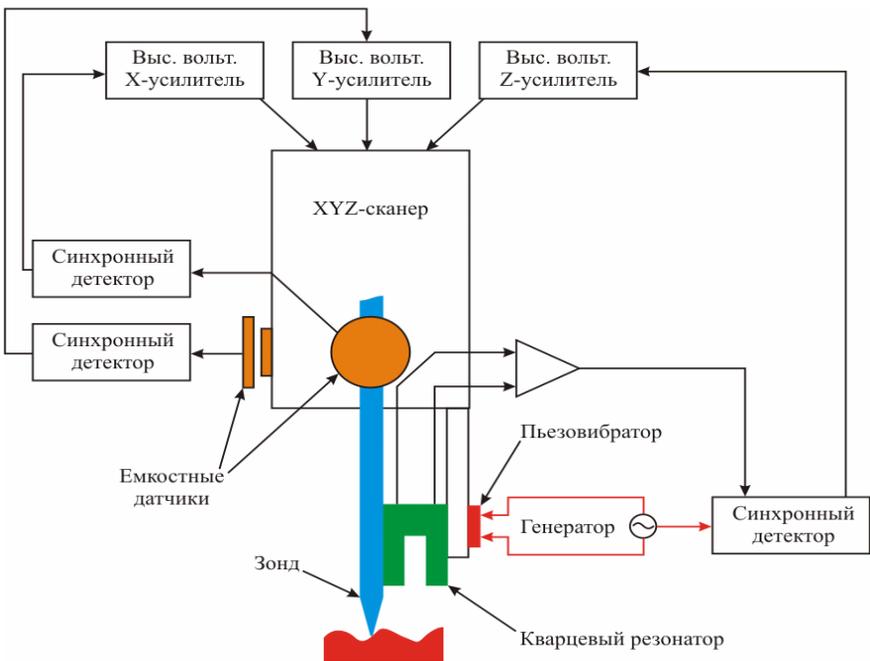


Рис. 2.14 - СБОМ, функциональная схема измерительной головки

2.4 Электронные микроскопы

Электронный микроскоп (ЭМ) — прибор, позволяющий получать изображение объектов с максимальным увеличением до 10^6 раз,

благодаря использованию вместо светового потока пучка электронов с энергиями 30÷200 кэВ и более.

Разрешающая способность электронного микроскопа в 1000÷10000 раз превосходит разрешение светового микроскопа и для лучших современных приборов может составлять несколько ангстрем.

Для получения изображения в электронном микроскопе используются специальные магнитные линзы, управляющие движением электронов в колонне прибора при помощи магнитного поля [34,35].

Теоретически максимально возможное разрешение в оптическом микроскопе ограничено длиной волны фотонов, используемых для освещения образца и угловой апертурой оптической системы.

В начале XX века ученые обсуждали вопрос преодоления ограничений относительно большой длины волны видимого света (длины волн 400—700 нанометров) путём использования электронов.

Электроны эмитируются в электронном микроскопе посредством термоэлектронной эмиссии из нити накаливания (вольфрамовая проволока или монокристалл гексаборида лантана), либо посредством полевой эмиссии.

Затем электроны ускоряются высокой разностью потенциалов и фокусируются на образце электромагнитными (или реже — электростатическими) линзами

Появление электронного микроскопа стало возможным после ряда физических открытий конца XIX — начала XX века. Это открытие в 1897 году электрона (Дж. Томсон) и экспериментальное обнаружение в 1926 году волновых свойств электрона (К. Дэвиссон, Л. Джермер), подтверждающее выдвинутую в 1924 году де Бройлем гипотезу о корпускулярно-волновом дуализме всех видов материи.

В 1926 году немецкий физик Х. Буш создал магнитную линзу, позволяющую фокусировать электронные лучи, что послужило предпосылкой для создания в 1930-х годах первого электронного микроскопа.

В 1931 году Р. Руденберг получил патент на просвечивающий электронный микроскоп, а в 1932 году М. Кнолль и Э. Руска построили первый прототип современного прибора.

Эта работа Э. Руски в 1986 году была отмечена Нобелевской премией по физике, которую присудили ему и изобретателям сканирующего зондового микроскопа Герду Карлу Биннигу и Генриху Рореру.

Использование просвечивающего электронного микроскопа для научных исследований было начато в конце 1930-х годов и тогда же появился первый коммерческий прибор, построенный фирмой Siemens.

В конце 1930-х — начале 1940-х годов появились первые растровые электронные микроскопы, формирующие изображение объекта при последовательном перемещении электронного зонда малого сечения по объекту.

Массовое применение этих приборов в научных исследованиях началось в 1960-х годах, когда они достигли значительного технического совершенства

Взаимодействие электронов с веществом. В классическом микроскопе видимый свет реагирует с образцом и отраженные фотоны анализируются детекторами или глазом человека.

В электронной микроскопии пучок света заменен пучком электронов, взаимодействующих с поверхностью образца и отраженные фотоны заменены целым спектром частиц и излучения: вторичные электроны, обратно отраженные электроны, Оже-электроны, рентгеновское излучение, катодолюминесценция и т. д. (Рис. 2.15).

Эти частицы и излучение являются носителями информации различного типа о веществе, из которого создан образец.

В результате взаимодействия с атомами образца электроны первичного пучка могут передать часть своей энергии электронам из зоны проводимости, то есть слабо связанным с атомами.

В результате такого взаимодействия может произойти отрыв электронов и ионизация атомов.

Такие электроны называются вторичными. Эти электроны обычно обладают небольшой энергией (порядка 50 эВ). Любой электрон первичного пучка обладает энергией, достаточной для появления нескольких вторичных электронов.

Так как энергия вторичных электронов невелика, их выход возможен только с приповерхностных слоев материала (менее 10 нм).

Благодаря небольшой кинетической энергии эти электроны легко отклоняются небольшой разностью потенциалов.

Это делает возможным существенно повысить эффективность детекторов (собрать максимально возможное их количество) и получить высококачественные изображения с хорошим отношением сигнал/шум и разрешением порядка 4 нм при диаметре пучка 3 нм.

Для типичных толщин образцов ПЭМ (100 нм), большинство электронов проходят его не испытав рассеяния (unscattered electrons), либо испытав один акт столкновения (single scattering), кратное число ($1 < n < 20$) столкновений (plural scattering) или многократное ($n > 20$) рассеяние (multiple scattering). Столкновения бывают упругими и неупругими.

Упругое рассеяние (elastically scattered electrons). Упругие столкновения – это такие, при которых энергия не расходуется на воз-

буждение атомов среды. Направление движения электрона может изменяться, но энергия практически не изменяется, т.е. $E \approx E_0$. Мы будем разделять упругое рассеяние на изолированном атоме и на системе атомов. Упругое рассеяние на малые углы обычно вызвано рассеянием на электронах, а на большие углы – на ядрах.

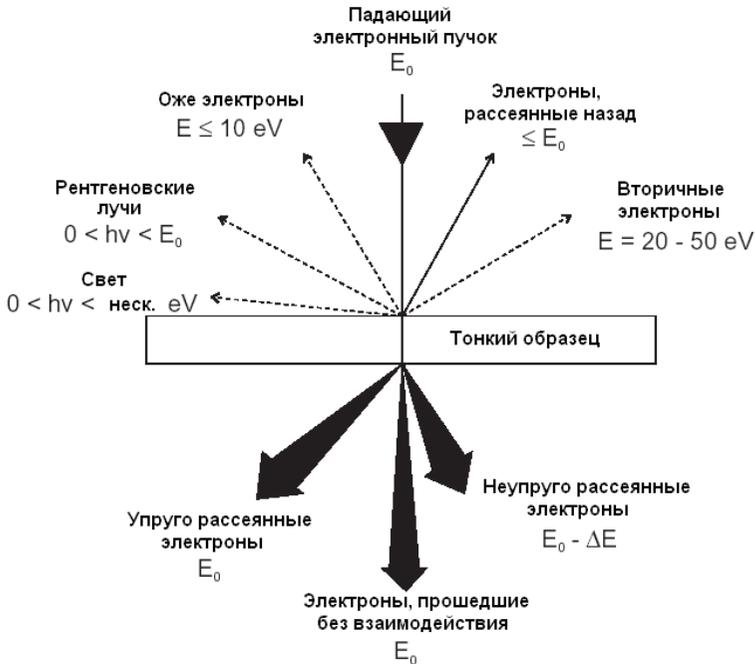


Рис. 2.15-Основные процессы при взаимодействии электронов с веществом

Неупругое рассеяние. Энергия таких электронов (inelastically scattered electrons) $E < E_0$ - (energy loss electrons) теряется на а) коллективное взаимодействие с многими атомами; б) генерацию процессов, приводящих к вылету вторичных электронов; в) генерацию рентгеновских лучей.

Характеристическое рентгеновское излучение (ХРИ или X-ray emission). Генерируемое электронным пучком рентгеновское излучение может иметь, в принципе, энергию γ -квантов до E_0 . Помимо того, что испускание рентгеновского излучения отражается в спектре EELS, на регистрации ХРИ основан аналитический метод XEDS (X-rays energy dispersive spectrometry). Возбуждение ХРИ связано с образова-

нием первичным электроном (пучка) вакансии, скажем, на К-оболочке.

Тормозное рентгеновское излучение (Bremsstrahlung X-rays). Электроны, сталкивающиеся с ядром с малыми прицельными параметрами, испытывают торможение в кулоновском поле ядра, что вызывает рентгеновское излучение, которое называют тормозным (ТРИ). Для оценки выхода ТРИ обычно используют сечение, выведенное Крамерсом (Kramers, 1923).

Генерация вторичных электронов. Вторичные электроны (ВЭ, secondary electrons=SEs) можно разделить на три группы:

- Медленные ВЭ = МВЭ (<50 эВ) (slow SEs), выбитые из валентной зоны или зоны проводимости.
- Быстрые БВЭ (fast SEs=FSEs), выбитые из внутренних оболочек атомов. Вероятность выбивания быстрых ВЭ мала, но если это происходит, то они забирают до 50% энергии исходных электронов.
- Электроны Оже (Auger electrons).

Так как энергия вторичных электронов невелика, их выход возможен только с приповерхностных слоев материала (менее 10 нм). Благодаря небольшой кинетической энергии эти электроны легко отклоняются небольшой разностью потенциалов. Это делает возможным существенно повысить эффективность детекторов (собрать максимально возможное их количество) и получить высококачественные изображения с хорошим отношением сигнал/шум и разрешением порядка 4 нм при диаметре пучка 3 нм.

Медленные вторичные электроны. МВЭ (SE) дают наименьший вклад в неупругие процессы, однако, этот процесс достаточно интенсивен, чтобы его можно было использовать в ПЭМ. Большую часть МВЭ составляют свободные эл-ны, т.е. не связанные с определенным атомом. Поскольку энергия МВЭ мала, то они могут вылетать только из приповерхностных слоев. Поэтому, МВЭ используются в ПЭМ для изображения поверхности образца. Это стандартная методика для СЭМ находит все большее применение и в сканирующей ПЭМ (СПЭМ=STEM), где с помощью МВЭ, получают изображение топографии поверхности с очень высоким изображением.

СПЭМ изображение поверхности с помощью МВЭ имеет гораздо более высокое разрешение, чем СЭМ!

Быстрые вторичные электроны (БВЭ=FSE). БВЭ получают значительную долю электронов пучка. Сечение генерации БВЭ на порядок больше, чем сечение для МВЭ. Для СЭМ, где $E_0 < 30$ кэВ, это не вызывает проблем, однако, в ПЭМ БВЭ могут иметь энергии ~50-200 кэВ, так что БВЭ вылетают из больших глубин из образца. Это приводит к ухудшению пространственного разрешения при микроанализе в

аналитических электронных микроскопах (АЕМ), а также к генерации ХРИ, усложняя интерпретацию данных по ХРИ. Таким образом, БВЭ не представляя проблемы для изображения в ПЭМ, являясь проблемой для химического анализа.

БВЭ в ПЭМ нежелательны и, вместе с тем, неизбежны. Они не используются для формирования изображения или для спектроскопии, но они ухудшают разрешение в обоих случаях.

Оже-электроны (ОЭ=Auger electrons) конкурируют с ХРИ и обычно имеют энергию до 10 эВ.. Процесс генерации ОЭ разделяется на два этапа. На первом электрон пучка выбивает электрон К-оболочки. На втором, образующаяся вакансия заполняется L3-электроном, избыток же энергии $\Delta E = E_K - E_L$ испускается в виде (3-го) L2 ОЭ. Это излучение называется излучением К-L3,L2 ОЭ. Образующаяся вакансия на L-оболочке заполняется электроном из внешней (валентной) оболочки, например М-оболочки (если она есть у атома) и остаток энергии излучается в виде низкоэнергетичного фотона. ОЭ имеют характеристический спектр, зависящий от электронной структуры атома и почти полностью идентичный спектру ХРИ.

Оже-процесс более вероятен в атомах с малой энергией связи электронов, т.е. в легких элементах. Типичные энергии ОЭ находятся в интервале от нескольких сот эВ до нескольких кэВ. ОЭ сильно поглощаются в образце.

Посему, регистрируемые ОЭ вылетают из приповерхностных слоев. Поэтому Оже-электронная спектроскопия (ОЭС, AES-Auger electron spectroscopy) признанный метод химического анализа поверхности. ХРИ не может конкурировать, поскольку слабее поглощается и вылетает со всей толщи образца. ОЭС накладывает особые требования на чистоту поверхности и, как следствие, ультра-высокого вакуума. Пока ОЭС/СПЭМ встречается лишь в разработках отдельных микроскопистов и чаще реализуется в специализированных аналитических СЭМ. С развитием ПЭМ ультра-высокий вакуум перестает быть чем-то особенным, поэтому ОЭС/СПЭМ имеет перспективы.

Генерация плазмонов и фононов. Плазмоны и фононы являются коллективными возбуждениями в твердом теле.

Плазмоны представляют собой продольные коллективные колебания валентных электронов, затухающие в течение фемтосекунд, так что эти волны локализованы в пределах до 10 нм. Генерация плазмонов имеет наибольшее сечение, является наиболее сильным каналом неупругих потерь. Плазмоны возникают в любом материале, имеющем слабо связанные электроны, но доминируют в металлах, особенно в таких металлах как алюминий, имеющих большую по-

верхность Ферми и, след-но, высокую плотность свободных электронов.

Плазмонные осцилляции квантуются, средний пробег плазмона составляет ~ 100 нм. Энергия плазмонов зависит от плотности свободных электронов, и т.о. от химического состава, хотя это и не используется сейчас для микроанализа.

Дифференциальное сечение плазмонного возбуждения получено Феррелем (Ferrel, 1956).

Фононы являются квантами колебаний решетки, т.е., тепла. Механизмы генерации фононов могут быть как прямыми, так и косвенными. Первые обусловлены элементарным актом неупругого рассеяния электронов пучка на атомных ядрах, приводящих к упругим осцилляциям атомов относительно равновесных позиций. Вторые являются следствием других неупругих процессов в конечной стадии. Например, эмиссия ОЭ или ХРИ или межзонные переходы могут конвертироваться в осцилляции решетки. Энергия фонона мала, $< \sim 0.1$ эВ, но электрон, взаимодействующий с атомом с образованием фонона, рассеивается на достаточно большой угол (5-15 мрад). Эти электроны дают вклад в диффузное рассеяние, обуславливающий фон к рефлексам брэгговской дифракции.

Хотя фононное рассеяние не используется для микроанализа ни для анализа изображения, полезно помнить, что сечение фононного рассеяния имеет $Z^{3/2}$ зависимость от атомного номера. Фононное рассеяние возрастает с увеличением температуры. При комнатной температуре средняя длина свободного пробега фонона составляет пару нм для Au и ~ 350 нм для Al, а при температуре жидкого гелия она возрастает в $\sim 2-3$ раза.

Плазмоны и фононы не регистрируются в ПЭМ, однако, регистрируются электроны их породившие.

Принимая во внимание, что вторичные электроны генерируются приповерхностными слоями, они очень чувствительны к состоянию поверхности. Минимальные изменения отражаются на количестве собираемых электронов. Таким образом, этот тип электронов несет в себе информацию о рельефе образца.

Однако, они мало чувствительны в отношении плотности материала, а, следовательно, и фазового контраста.

Виды электронных микроскопов:

Сканирующая (растровая) микроскопия (СЭМ, SEM). В основе лежит телевизионный принцип развертки тонкого пучка электронов по поверхности образца.

- Растровый электронный микроскоп
 - Отражательный растровый электронный микроскоп
- Трансмиссионная* (просвечивающая) микроскопия (ТЭМ, ТЕМ) - пучок электронов, пройдя через образец, фокусируется на экране или пластинке.
- Просвечивающий электронный микроскопи (ПЭМ).
 - Растровый просвечивающий электронный микроскоп (РПЭМ).
 - Сканирующий трансмиссионный электронный микроскоп (СТЭМ, STEM).

В технологии СЭМ изображение формируется сигналами, которые генерируются последовательно по мере того, как тонко сфокусированный пучок электронов сканирует поверхность образца. Разрешение СЭМ ограничено из-за рассеяния электронов.

В технологиях ТЭМ и СТЭМ используются сигналы, генерируемые электронами, проходящими через очень тонкий (как правило, менее 100 нанометров) образец. В ТЭМ прошедшие образец электроны фокусируются в реальное изображение на флуоресцентном экране или на множестве детекторов электронов. В СТЭМ, как и в случае СЭМ, производится сканирование сфокусированного пучка, но на тонком срезе образца, при этом интенсивность прошедших образец электронов или какой-либо другой сигнал отображаются в виртуальное изображение. Поскольку в технологии СТЭМ образцы имеют очень малую толщину, большинство электронов при проходе через образец рассеиваются не более одного раза. Тем самым удается избежать многократного рассеяния, которое сопровождает прохождение электронов через массивные образцы, и расхождения пучка, которое является главным фактором, лимитирующим разрешение в СЭМ.

Системы СТЭМ с скорректированной оптикой (такие, как Titan STEM фирмы FEI) демонстрируют разрешение изображений на суб-ангстремном уровне. Не менее важно, что такое высокое разрешение достигается непосредственно на изображении; отпадает необходимость в сложных и продолжительных процедурах реконструкции, которые требовались для обеспечения максимального разрешения при обработке эффекта *делокализации* в случае некорректированной оптики ТЭМ. Делокализация – явление, при котором информация от некоторой точки объекта «распространяется» на большую площадь изображения из-за аберрации линз. Это мешает интерпретировать тонкие детали в областях неоднородностей образца, т.е. делать ровно то, для чего предназначен электронный микроскоп при исследованиях на атомном уровне – исследовать интерфейсы между материалами, границы частиц или дефекты прозрачности, все то, что так существенно

влияет на свойства изучаемых материалов. Свойства всех материалов, в конечном итоге, определяются взаимодействиями на молекулярном и атомном уровнях. Соответственно, специалисты по материаловедению в течение многих лет ждали таких аналитических инструментов и средств визуализации, которые обеспечили бы пространственное разрешение на уровне индивидуальных атомов – менее одного ангстрема (1 ангстрем = 0.1 нанометра). Первые изображения ТЭМ такого уровня были продемонстрированы много лет назад, но они были доступны только для ограниченного класса образцов и требовали сложнейших процедур сбора данных и их последующей обработки для получения из изображения. Современные достижения, в частности коммерческая доступность технологии корреktированной оптики, делают визуализацию и анализ на атомном уровне в СТЭМ много проще для большого разнообразия образцов.

Сканирующая электронная микроскопия.

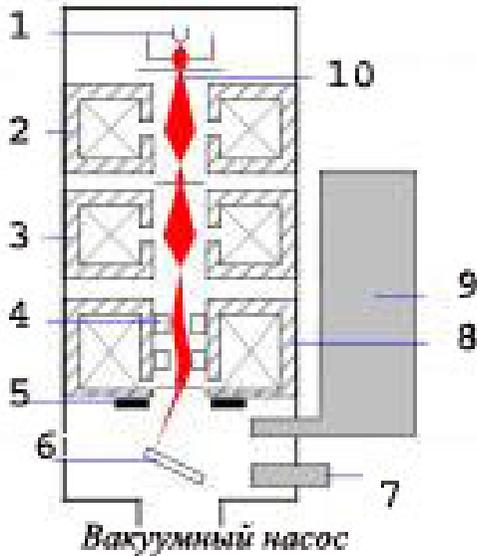
Основа сканирующего электронного микроскопа — электронная пушка и электронная колонна, функция которой состоит в формировании остро-сфокусированного электронного зонда средних энергий (10 — 50 кэВ) на поверхности образца (см. Рис. 2.16, 2.17).



Рис. 2.16 - Схема сканирующего микроскопа.

Прибор обязательно должен быть оснащен вакуумной системой (в современных моделях микроскопов высокий вакуум желателен, но не обязателен). РЭМ есть предметный столик, позволяющий перемещать образец минимум в трех направлениях. При взаимодействии зонда с объектом возникают несколько видов излучений, каждое из

которых может быть преобразовано в электрический сигнал. В зависимости от механизма регистрирования сигнала различают несколько режимов работы сканирующего электронного микроскопа: режим вторичных электронов, режим отражённых электронов, режим катодоллюминесценции и др. Также каждый РЭМ оснащается детекторами, позволяющими отобрать и проанализировать излучение и частицы изменившие энергию в результате взаимодействия электронного зонда с образцом. Разработанные методики позволяют исследовать не только свойства поверхности образца, но и визуализировать информацию о свойствах подповерхностных структур.



- | | |
|------------------------------------|----------------------------------|
| 1. Электронная пушка. | 6. ОБРАЗЕЦ |
| 2. Первый конденсор. | 7. Детектор вторичных электронов |
| 3. Второй конденсор | 8. Объектив. |
| 4. Отклоняющие катушки. | 9. РСМА (микрозонд). |
| 5. Детектор отраженных электронов. | |

Рис. 2.17 -Схема РЭМ, оснащенного детектором рентгеновских лучей— РСМА (микрозонд)

Приведем основные типы сигналов, которые генерируются и детектируются в процессе работы РЭМ:

- вторичные электроны (ВЭ или режим рельефа);
- отражённые электроны (ОЭ или режим фазового контраста);

–прошедшие через образец электроны, в случае установленной STEM-приставки (используется для исследования органических объектов);

–дифракции отражённых электронов (ДОЭ);

–потери тока на образце (ПЭ или детектор поглощенных электронов);

–ток, прошедший через образец (ТЭ или детектор прошедших электронов);

–характеристическое рентгеновское излучение (РСМА или рентгеноспектральный микроанализ);

–ВДА или волнодисперсионный анализ;

–световой сигнал (КЛ или катодолуминесценция).

Всевозможные типы детекторов, установленные на одном приборе встречаются крайне редко. Детекторы вторичных электронов — первый и традиционно устанавливаемый на все РЭМ тип детекторов. В этом режиме разрешающая способность РЭМ максимальна. Разрешение детекторов вторичных электронов в современных приборах уже достаточно для наблюдения субнанометровых объектов. Из-за очень узкого электронного луча РЭМ обладают очень большой глубиной резкости (0,6-0,8 мм), что на два порядка выше, чем у оптического микроскопа и позволяет получать четкие микрофотографии с характерным трехмерным эффектом для объектов со сложным рельефом. Это свойство РЭМ крайне полезно для понимания поверхностной структуры образца. Микрофотография пыльцы демонстрирует возможности режима ВЭ РЭМ.

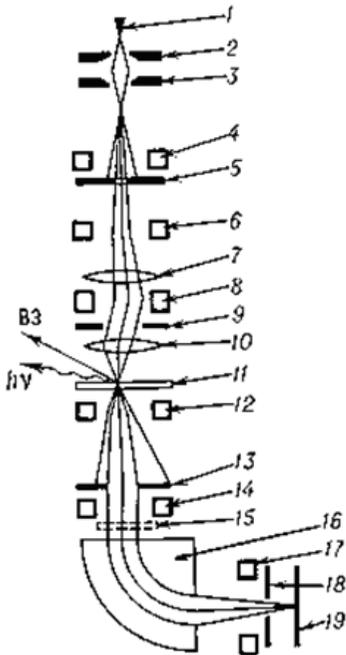
Отражённые электроны (ОЭ) — это электроны пучка, отражённые от образца упругим рассеиванием. ОЭ часто используются в аналитическом РЭМ совместно с анализом характеристических спектров рентгеновского излучения. Поскольку интенсивность сигнала ОЭ напрямую связана со средним атомным номером (Z) засвечиваемой области образца, изображения ОЭ несут в себе информацию о распределении различных элементов в образце. Например, режим ОЭ позволяет обнаружить коллоидные золотые иммунные метки диаметра 5-10 нм, которые очень тяжело или даже невозможно обнаружить в биологических объектах в режиме ВЭ. Микрофотография поверхности шлифа металл-оксидной системы демонстрирует возможности режима ОЭ РЭМ.

Характеристическое рентгеновское излучение генерируется в случае, когда электронный луч выбивает электроны с внутренних оболочек элементов образца, заставляя электрон с более высокого энергетического уровня перейти на нижний уровень энергии с одновременным испусканием кванта рентгеновского излучения. Детектирование

спектра характеристического рентгеновского излучения позволяет идентифицировать состав и измерить количество элементов в образце.

Просвечивающая электронная микроскопия.

Просвечивающий (трансмиссионный) электронный микроскоп — это устройство, в котором изображение от ультратонкого образца (толщиной порядка 0,1 мкм) формируется в результате взаимодействия пучка электронов с веществом образца с последующим увеличением магнитными линзами (объектив) и регистрацией на флуоресцентном экране, фотоплёнке или сенсорном приборе с зарядовой связью (ПЗС - матрице) – см. Рис. 2.18.



1-автоэмиссионный катод; 2 – промежуточный анод; 3 – анод; 4 – отклоняющая система для юстировки пучка; 5 – диафрагма «осветителя»; 6,8 – отклоняющие системы для развёртки электронного зонда; 7 – магнитная длиннофокусная линза; 9 – апертурная диафрагма; 10 – магнитный объектив; 11 – объект; 12, 14 – отклоняющие системы; 13 – кольцевой коллектор рассеянных электронов; 15 – коллектор нерассеянных электронов (убирается при работе со спектрометром); 16 – магнитный спектрометр, в котором электронные пучки поворачиваются магнитным спектрометром на 90° ; 17 – отклоняющая система для отбора электронов с различными потерями энергии; 18 – щель спектрометра; 19 – коллектор; ВЭ – поток вторичных электронов, $h\nu$ - рентгеновское излучение.

Рис. 2.18 –Принципиальная схема просвечивающего электронного микроскопа

Первый ПЭМ создан немецкими инженерами-электронщиками Максом Кноллем и Эрнстом Руской 9 марта 1931 года.

Первый практический просвечивающий (трансмиссионный) электронный микроскоп был построен Альбертом Пребусом и Дж. Хиллиером в университете Торонто (Канада) в 1938 году на основе принципов,

открытых ранее Кноллем и Руской. Эрнсту Руске за его открытие в 1986 году присуждена Нобелевская премия по физике.

ТЭМ состоит из нескольких компонентов:

вакуумная система; источник электронов (электронный прожектор, электронная пушка) для генерирования электронного потока; источник высокого напряжения для ускорения электронов; набор электромагнитных линз и электростатических пластин для управления и контроля электронного луча; экран, на который проецируется увеличенное электронное изображение.

Прошедший через образец луч содержит информацию об электронной плотности, фазе и периодичности; которые используются при формировании изображения.

Принцип действия просвечивающего электронного микроскопа несложен.

Оптическая схема электронного микроскопа просвечивающего типа аналогична схеме обычного светового микроскопа: Конденсорная линза «освещает» узким пучком электронов объект, изображение которого с помощью двух систем электронных линз – объективной и проекционной – в увеличенном масштабе переносится на конечный экран.

Проходя через объект, расположенный вблизи апертурной диафрагмы объективной линзы, электроны взаимодействуют с атомами объекта и отклоняются от первоначального направления падения пучка, т.е. рассеиваются (поглощение электронов в объектах, вследствие их очень малой толщины, в большинстве случаев можно пренебречь).

При этом у части электронов скорость меняется только по направлению, не меняясь по величине, что соответствует упругому рассеиванию.

Скорость другой части электронов меняется и по направлению, и по величине, при этом часть энергии электронов затрачивается на ионизацию и возбуждение атомных электронов в объекте. Вследствие этого электроны, пройдя через объект, после рассеяния в нем имеют вид расходящегося пучка.

При этом электроны, рассеянные на угол, больший апертурного угла объективной линзы, определяемого диаметром апертурной диафрагмы и ее геометрическим положением, поглощаются в толще материала этой диафрагмы, и в дальнейшем в формировании изображения, возникающего на экране ТЭМ, принимает участие только та часть рассеянных электронов, которая прошла через диафрагму.

Таким образом, чем большей рассеивающей способностью обладает некоторый участок исследуемого объекта, тем более темным будет его изображение.

Напряжение, которое используют для ускорения потока электронов в большинстве ТЭМ, достигает 50 000 – 100 000 В.

Длина волны электрона при этих напряжениях составляет от 0,0055 до 0,0039 нм соответственно.

Однако теоретическое максимальное разрешение, равное половине длины волны (приблизительно 0,002 нм) практически не может быть достигнуто по целому ряду причин.

К ним относятся трудности изготовления деталей микроскопа с необходимой точностью, нестабильность высокого напряжения и тока в обмотках линз, а также аберрации линз, которые позволяют использовать линзы только с очень небольшими апертурами.

Разрешение в 0,1 нм было достигнуто для кристаллических (периодических) материалов.

Разрешение в 0,3 – 0,4 нм – для некоторых материалов с непериодической структурой.

Однако для большинства биологических материалов разрешение 1 – 2 нм обычно считается приблизительным пределом; это обусловлено спецификой методов подготовки объекта и самой природой биологических объектов.

2.5 Спектроскопия

Спектроскопия — разделы физики и аналитической химии, посвящённые изучению спектров взаимодействия излучения (в том числе, электромагнитного излучения, акустических волн и др.) с веществом [35].

В физике спектроскопические методы используются для изучения всевозможных свойств этих взаимодействий. В аналитической химии — для обнаружения и определения веществ при помощи измерения их характеристических спектров, то есть методами спектрометрии.

К существенным преимуществам спектроскопии можно отнести возможность диагностики *in situ*, то есть непосредственно в «среде обитания» объекта, бесконтактно, дистанционно, без какой-либо специальной подготовки объекта. Поэтому она получила широкое развитие, например, в астрономии.

Спектроскопия служит для определения характеристик вещества (не являющихся непосредственно наблюдаемыми величинами) по свойствам его спектров (которые наблюдаются непосредственно и

напрямую зависят как от определяемых характеристик, так и от внешних факторов).

По объектам исследования можно выделить следующие виды спектроскопии: атомная спектроскопия, молекулярная спектроскопия, масс-спектроскопия, ядерная спектроскопия и другие [35,37,38].

По типу излучения, которое используется в спектроскопии, её можно разделить на оптическую спектроскопию (в том числе - инфракрасную), фотоэлектронную спектроскопию, Мёссбауэровскую спектроскопию, масс-спектроскопию, спектроскопию с использованием радиоизлучения и т. д.

2.5.1 Инфракрасная спектроскопия

Инфракрасная спектроскопия (ИКС) — раздел спектроскопии, охватывающий длинноволновую область спектра (>730 нм за красной границей видимого света).

Инфракрасная спектроскопия - один из физических методов исследования строения молекул.

Она позволяет получить прямые сведения о состоянии адсорбированных веществ, о природе взаимодействия между реагентами и катализатором, выяснить строение адсорбционных, комплексов и другие вопросы, связанные с катализом.

Начало применению ИК-спектроскопии к исследованиям процессов на твердых поверхностях положил А.Н.Теренин в 1940 г. С 1955 по 1958 г. этот метод был развит американскими исследователями Р. Эйшенсом и В. Плискином и с тех пор широко используется во всех современных лабораториях при изучении поверхностных явлений и каталитических процессов.

В двухатомной молекуле AB атомы A и B удерживаются при данных условиях в определенном положении весьма прочно, однако не совсем жестко: они совершают периодические колебания (отчасти вращательные) с некоторой частотой ν , определяемой массами обоих атомов и упругими силами связей:

$$\nu = \frac{1}{2\pi c} \sqrt{k(m_1 + m_2)/(m_1 * m_2)}$$

где c - скорость света; k - силовая постоянная (сила связи или порядок связи, соответствующая упругой постоянной Гука); m_1 и m_2 -массы атомов. Амплитуда этих колебаний увеличивается при поглощении молекулой энергии.

Для получения ИК-спектра какого-нибудь соединения необходимо подвергнуть его действию ИК-излучения и измерить количество света каждой длины волны, пропущенного этим соедине-

нием. Величине максимального поглощения будет соответствовать некоторая частота колебания, выражаемая в случае двухатомных молекул указанным уравнением, которая является строго определенной характеристикой для данной связи и называется ее полосой поглощения. Совокупность таких полос, характеризующих данную молекулу, называется ее ИК-спектром.

Для изучения поверхностных явлений в пучок ИК-излучения ставят тонкую, приготовленную специальным способом пластинку из данного катализатора или адсорбента, на которой адсорбируются молекулы адсорбатов. В результате адсорбции изменяется прочность тех или иных связей в исходных молекулах, а в некоторых случаях (при необратимой хемосорбции) возможен даже разрыв отдельных связей. Вследствие этого спектр адсорбированного соединения будет отличаться от спектра взятого исходного вещества: в зависимости от силы адсорбционной связи в спектре будет наблюдаться смещение частот нормальных колебаний связей или же будут проявляться новые полосы поглощения, которых в спектре исходного соединения нет.

ИК излучение поглощают многие газы, за исключением таких как O_2 , N_2 , H_2 , Cl_2 и одноатомных газов. Поглощение происходит на длине волны, характерной для каждого определенного газа, для CO , например, таковой является длина волны 4,7 мкм.

Колебательные переходы и соответствующие им колебательные спектры молекул можно получить как при непосредственном поглощении веществом инфракрасного излучения (диапазон волн 2 - 50 мкм), так и при поглощении видимого и ультрафиолетового излучения. В последних случаях лишь часть поглощенной энергии расходуется на возбуждение молекул, большая часть реэмиттируется веществом. В соответствии с этим колебательная спектроскопия разделяется на инфракрасную (ИК) и спектроскопию комбинационного рассеяния (КР).

Вследствие простоты получения спектров и технического совершенства приборов наибольшее распространение получил метод ИК спектроскопии. Однако часто необходимым оказывается совместное применение обоих методов.

Как и всякое движение на молекулярном уровне, колебания молекул необходимо рассматривать через призму квантовой механики. Однако многие частные вопросы, например, вопрос о частоте колебаний молекулы, можно решить, рассматривая молекулу как механическую систему, например, как системы шаров, символизирующих атомы, и связывающих их пружин, символизирующих связи между ато-

мами. Тогда колебания вдоль линии связи, соединяющей два атома, можно описать упомянутым законом Гука.

С помощью приведенного выражения для V можно ориентировочно подсчитать частоту колебаний световой волны, поглощаемой какой-либо связью, оценить направление изменения этой частоты с изменением массы атомов, образующих связь, с увеличением или уменьшением прочности связи оценить влияние таких факторов, как межмолекулярное взаимодействие, водородная связь и т.д.

Энергия в колебательных движениях молекул квантована. Ее величина для так называемого гармонического осциллятора может быть найдена из выражения:

$$E = h\nu \left(V + \frac{1}{2} \right)$$

где V - колебательное квантовое число, принимающее значения натуральных чисел: 0, 1, 2, ...; ν - частота основного колебания, определяемая выражением, приведенным выше.

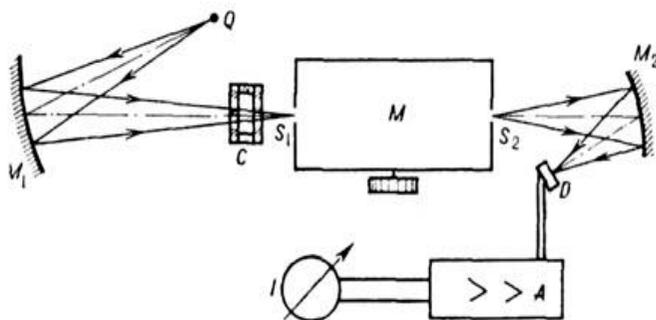
При $V = 0$ $E = 1/2 h\nu$, что указывает на безизлучательные колебательные движения молекул.

При поглощении кванта излучения молекула переходит на более высокий колебательный уровень. Большинство регистрируемых спектрометром полос поглощения обусловлено переходом $V_0 \rightarrow V_1$.

Основные приборы — различного типа инфракрасные спектрометры. Измерение спектра ИК-поглощения сводится к измерению интенсивности ИК-излучения, прошедшего через вещество, в зависимости от частоты излучения ν или длины волны λ . В классической абсорбционной ИКС излучение от источника с непрерывным ИК-спектром (Рис. 2.19) пропускают через кювету с исследуемым веществом; прошедшее через вещество излучение направляют на входную щель монохроматора, а из выходной его щели - на приёмник излучения. Затем сигнал усиливается и измеряется или регистрируется графопостроителем в процессе сканирования.

В лазерной ИКС измеряется зависимость интенсивности прошедшего через вещество излучения узкополосного ИК-лазера (чаще полупроводникового с перестраиваемой частотой) от частоты излучения лазера в процессе её перестройки. Связь между интенсивностью $I(\nu)$ прошедшего через кювету с веществом излучения с длиной волны λ (или волновым числом $1/\lambda$) и величинами, характеризующими поглощающее вещество, даётся обобщённым *Бугера-Ламберта-Бера*

законом. Обычно ИК-спектр поглощения представляют графически в виде зависимости от ν (или λ) величин, характеризующих только поглощающее вещество: коэффициент пропускания, коэффициент поглощения, оптическую плотность, показатель поглощения.



Q - источник непрерывного ИК спектра; M_1 - зеркало осветителя; M_2 - зеркало конденсора; C - кювета с исследуемым веществом; M - монохроматор; S_1 и S_2 - входная и выходная щели монохроматора; D - приёмник излучения; A - усилитель; I - измерительный или регистрирующий прибор.

По инфракрасным спектрам поглощения можно установить строение молекул различных органических (и неорганических) веществ с относительно короткими молекулами: антибиотиков, ферментов, алкалоидов, полимеров, комплексных соединений и др.

Колесательные спектры молекул различных органических (и неорганических) веществ с относительно длинными молекулами находятся в терагерцовом диапазоне, поэтому строение этих молекул можно установить с помощью радиочастотных спектрометров терагерцового диапазона.

Рамановская спектроскопия

Комбинационное рассеяние света (эффект Рамана)— неупругое рассеяние оптического излучения на молекулах вещества (твёрдого, жидкого или газообразного), сопровождающееся заметным изменением частоты излучения. В отличие от рэлеевского рассеяния, в случае комбинационного рассеяния света в спектре рассеянного излу-

чения появляются спектральные линии, которых нет в спектре первичного (возбуждающего) света. Число и расположение появившихся линий определяется молекулярным строением вещества.

Спектроскопия комбинационного рассеяния света - эффективный метод химического анализа, изучения состава и строения веществ. Первым комбинационное рассеяние света предсказал Адольф Сmealь (в 1923 г.), затем последовали теоретические работы Крамерса, Гейзенберга, Дирака, Шрёдингера и других.

В 1928 г. Ч.В. Раман выдвигает предположение, что наблюдаемый эффект - некий аналог эффекта Комптона в оптике, предполагая, что фотон может быть «поглощён частично», причем части не могут быть произвольными и должны соответствовать спектрам инфракрасного поглощения света. Индийские учёные Ч. В. Раман и К. С. Кришнан (Krishnan) предпринимают экспериментальную проверку этой гипотезы и обнаруживают линейчатый спектр у исследуемого ими на протяжении многих лет излучения. Схема эксперимента по наблюдению рассеянного света показана на рисунке 2.26.

Накопленный за многие годы экспериментальный материал позволил им немедленно опубликовать статью, в которой они заявили об обнаружении нового вида свечения, присущего широкому классу веществ. Линии спектра нового излучения были в первый раз наблюдаемы 28 февраля 1928 года. Нобелевская премия по физике 1930 года была присуждена Раману «за его работы по рассеянию света и за открытие эффекта, названного по его имени».

Происхождение эффекта Рамана удобнее всего объяснить в рамках квантовой теории излучения (Рис. 2.20).

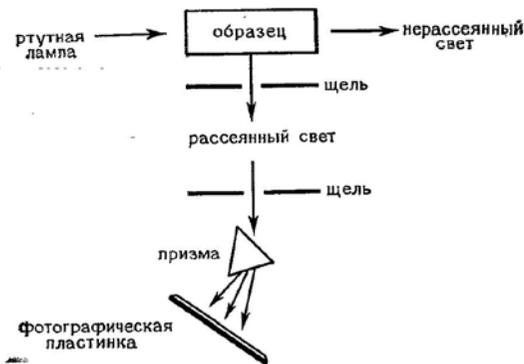


Рис. 2.20 - Схема наблюдения рассеяния света.

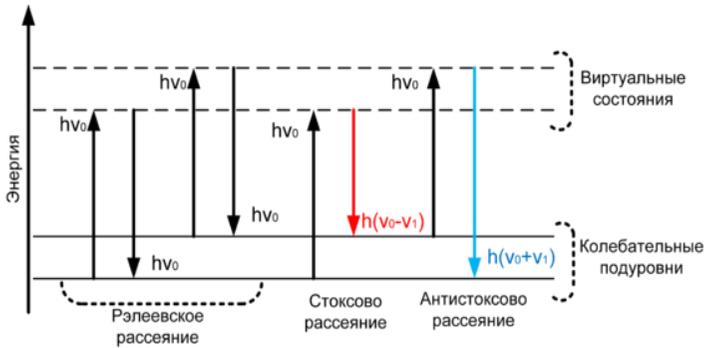


Рис. 2.21 - Взаимодействие фотона с молекулами.

Согласно квантовой теории, излучение частоты ν рассматривается как поток фотонов с энергией $h\nu$, где h — постоянная Планка. При столкновениях с молекулами фотоны рассеиваются. В случае упругого рассеивания, они будут отклоняться от направления своего движения, не изменяя своей энергии (Рэлеевское рассеяние). Но может быть и так, что при столкновении произойдет обмен энергией между фотоном и молекулой. Молекула при этом может как приобрести, так и потерять часть своей энергии в соответствии с правилами квантования - ее энергия может измениться на величину ΔE , соответствующую разности энергий двух разрешенных ее состояний. Иначе говоря, величина ΔE должна быть равна изменению колебательной и/или вращательной энергий молекулы. Если молекула приобретает энергию ΔE , то после рассеяния фотон будет иметь энергию $h\nu - \Delta E$ и соответственно частоту излучения $\nu - \Delta E/h$. А если молекула потеряет энергию ΔE , частота рассеяния излучения будет равна $\nu + \Delta E/h$. Излучение, рассеянное с частотой, меньшей чем у падающего света, называется стоксовым излучением, а излучение с большей частотой называется антистоксовым. При не очень высоких температурах населенность первого колебательного уровня невелика, при комнатной температуре при колебательной частоте 1000 см^{-1} на первом колебательном уровне находится всего 0.7% молекул, поэтому интенсивность антистоксова рассеяния мала. С повышением температуры населенность возбужденного колебательного уровня возрастает и интенсивность антистоксова рассеяния растет.

Преимущества, связанные со специфической природой метода, делают Рамановскую спектроскопию мощным средством анализа и мониторинга химического состава. Каждое соединение имеет свой уникальный Рамановский спектр.

В зависимости от применяемой аппаратуры, эта техника может быть использована для анализа твердых веществ, растворов, а также предоставлять информацию о физических характеристиках исследуемой среды, таких как кристаллическая решетка, ориентация, полиморфные формы и т.д. Одновременно решить проблемы повышения пространственного разрешения вплоть до 20-50 нм и химической идентификации структурных элементов в оптическом диапазоне длин волн возможно путем объединения методов ближнепольной микроскопии, люминесцентного и рамановского анализа [Progress in Nano-Electro-Optics].

3 Представление данных аналитических приборов.

Множество методов анализа, рассмотренных в предыдущей главе, порождает большое количество видов сигналов, поступающих на вход системы обработки. Это, прежде всего, аналоговые или дискретные данные – совокупность достаточно большого числа точек на графике или чисел в компьютере, которые сами по себе сигнал никак не характеризуют, особенно если сам сигнал скрыт в шумах.

Методы обработки экспериментальных данных тесно связаны с математикой и, в особенности, с математической статистикой, откуда они черпают свои идеи. Основная задача обработки состоит в извлечении нужной физической/химической *информации*. Что является информацией, зависит от сути решаемой задачи. В некоторых случаях достаточно знать, что некоторое вещество присутствует в системе, но в других уже необходимо получить и количественные значения. Данные могут содержать нужную информацию, они даже могут быть избыточными, но в некоторых случаях информации в данных может не быть совсем. Во всех случаях данные содержат шум (например, погрешности), который скрывает нужную информацию. *Что считать шумом, а что – информацией*, всегда решается с учетом поставленных целей и методов, используемых для ее достижения. Это второй важнейший принцип. Шум и избыточность в данных обязательно проявляют себя через *корреляционные связи* между переменными.

3.1 Типы данных

Различные типы экспериментальных данных, следуя работе [18], удобно рассматривать в виде векторов и матриц (см. Рис. 3.1): Простейший случай – это $0D$, т.е. просто одно число, например, значение оптической плотности, которое может быть получено на моно-

хроматическом фотометре. Более сложный случай – это 1D-вектор (столбец, или строка). Примерами 1D данных являются спектр или хроматограмма, которые на плоскости изображаются в виде графика. Матрицы данных принято считать многомерными данными, например, 2D-матрица – таблица из чисел, имеющих I строк и J столбцов.

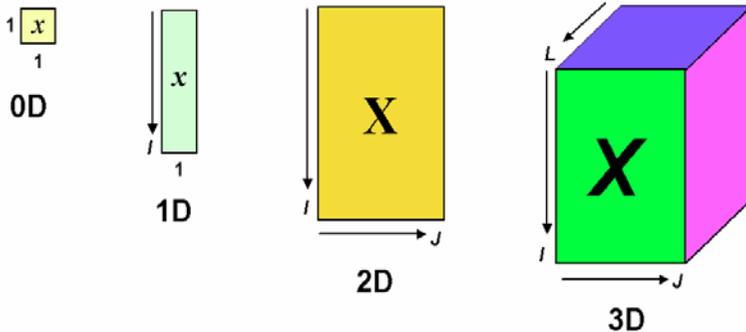


Рис. 3.1 Графическое представление экспериментальных данных

Типичным примером 2D данных является изображение или набор спектров, снятых для I объектов на J длинах волн. Каждая строка в такой матрице представляет объект (в данном случае, образец), а каждый столбец – переменную (длину волны) или признак. Отнесение данных к объектам (образцам) или переменным имеет большое значение для их интерпретации. Например, при анализе данных, полученных методом ВЭЖХ-ДДМ для 30 точек по времени на 28 длинах волн, можно составить матрицу из 28 строк и 30 столбцов, а можно, наоборот, считать длины волн переменными, а времена удержания – объектами. В эволюционных физических экспериментах, объекты соответствуют временам, то есть образец, меняющийся во времени, рассматривается как серия объектов [39].

С развитием гибридных методов [40] появились *трехмерные (и более) матрицы*. Их можно представить в виде параллелепипеда (3D матрицы), в котором каждое ребро соответствует своему типу переменной. Типичный пример - сочетание газовой хроматографии и масс-спектрометрии (ГХ-МС) применяется для количественного определения кленбутерола в биологических матрицах.

Результаты физических экспериментов, или данные для анализа, могут объединяться в блоки. Простейший случай – это один блок X . Такой случай чаще встречается в качественном анализе, например, в задаче разделения спектров и концентраций. Количественный анализ,

основанный на регрессионных зависимостях, использует данные, состоящие из двух и более блоков.

Блок предикторов (например, 2D-матрица спектров X) и блок откликов (например, 1D- вектор концентраций y) составляют набор стандартных данных, по которым строится калибровочная модель $y=Xb$.

3.2 Пространство сигналов - векторов

Рассмотрим некоторые положения теории сигналов на примере 1D данных [41]. Вся совокупность таких сигналов S , встречающихся как на практике, так и в теоретических исследованиях, может быть разбита на ряд групп или *множеств*, каждое из которых объединено каким-либо одним общим и единственным свойством P

. Условно это записывается как: $V_p = \{s; P\}$, т.е. V_p есть множество всех S , для которых справедливо P . Следовательно, определив свойство P , мы определяем тем самым множество сигналов.

Рассматривая элементы, в сущности, сигналы одного и того же множества, необходимо условиться о признаках или признаке, отличающих их друг от друга. Общий и интуитивно ясный подход заключается в том, чтобы эти элементы интерпретировать в виде точек в некоторой системе координат. Оказывается, что наиболее подходящим признаком, отличающим два элемента множества, например x и y , является расстояние между этими точками (обозначим его через $d(x, y)$).

Множество с подходящим образом определенным расстоянием называется *пространством*. Если, кроме того, установлено правило, с помощью которого вычисляется это расстояние, то пространство называется *метрическим*, а само правило называется *метрикой*. Расстояние d обладает следующими аксиоматическими признаками:

- 1) $d(x, y) \geq 0 \Rightarrow d(x, y) = 0$, только если $x = y$
- 2) $d(x, y) = d(y, x)$ (симметрия) (3.1)
- 3) $d(x, z) \leq d(x, y) + d(y, z)$ (неравенство треугольника)

После ввода понятия метрики дальнейшим продуктивным подходом к конструированию пространства для представления сигналов является рассмотрение сигналов как векторов, соединяющих начало координат с точкой (элементом) этого пространства. Тогда каждому элементу x, y, z, \dots пространства соответствует набор чисел (вещественных или комплексных).

$\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \dots, \lambda_1, \lambda_2, \lambda_3, \dots$ и т.д., являющихся проекцией вектора на координатные оси. Кроме того, используя операции векторной алгебры, могут быть определены различные взаимосвязи между элементами пространства. Поскольку эти операции линейны, то и пространство, образованное множеством векторов, является *линейным*. Более точно: множество элементов x, y, z, \dots образует линейное пространство, если в нем определены следующие операции:

1. Каждой паре векторов x и y множества V соответствует вектор $x + y \in V$, называемый суммой x и y ;

2. Каждому элементу $x \in V$ и каждому числу (скаляру) λ поставлен в соответствие определенный элемент $\lambda x \in V$ – произведение на скаляр так, что для любых элементов $x + y \in V$ и любых скаляров λ и μ

а) сложение коммутативно $x + y = y + x$;

б) сложение ассоциативно $x + (y + z) = (x + y) + z$;

в) существует нулевой элемент $0 \in V$ такой, что $x + 0 = x$;

г) для любого x существует единственный вектор $(-x)$ такой, что $x + (-x) = 0$;

д) умножение на скаляр ассоциативно $\lambda(\mu x) = (\lambda\mu)x$;

е) $1x = x$ и $0x = 0$ для любого x ;

ж) $\lambda(x + y) = \lambda x + \lambda y$; $(\lambda + \mu)x = \lambda x + \mu x$ (законы дистрибутивности).

Приведенные выше свойства позволяют ввести достаточные ясные алгебраические взаимосвязи между сигналами.

Если в качестве скаляров $\alpha, \beta, \lambda, \mu$ и т.д. берутся вещественные числа, то такое пространство называется *вещественным (действительным)* линейным пространством.

В противном случае, имеем дело с *комплексным* линейным пространством.

Расстояние от начала координат до какой-либо точки пространства называется *нормой*, а пространство, где она введена – *нормированным*. Норма вектора x , обозначаемая как $\|x\|$, должна удовлетворять следующим аксиомам:

$$а) \|x\| \geq 0 \quad \text{è} \quad \|x\| = 0, \text{ только если } x = 0;$$

$$б) \|\lambda x\| = |\lambda| \|x\|; \quad (3.2)$$

$$в) \|x + y\| \leq \|x\| + \|y\| \text{ - неравенство треугольник}$$

Сопоставляя приведенные аксиомы с требованиями, которым должна удовлетворять метрика $d(x, y)$, легко видеть, что

$$d(x, y) = \|x - y\| \quad (3.3)$$

есть метрика, которая удовлетворяет условиям (3.1). Если такая метрика используется в *нормированном линейном пространстве*, то оно будет *метрическим*.

Пусть $x = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ и $y = \{\beta_1, \beta_2, \dots, \beta_n\}$, тогда следующие функционалы дают примеры метрик:

$$а) d_1(x, y) = \sum_{i=1}^n |\alpha_i - \beta_i|,$$

$$б) d_2(x, y) = \left[\sum_{i=1}^n |\alpha_i - \beta_i|^2 \right]^{1/2},$$

$$в) d_3(x, y) = \max\{|\alpha_i - \beta_i|; i = 1, 2, \dots, n\}.$$

Взаимное отношение двух векторов, т.е. их геометрические свойства, может быть охарактеризовано с помощью числа, называемого *скалярным произведением*. Иными словами, это – отображение упорядоченных пар векторов линейного пространства на вещественную ось R или комплексную плоскость C . Скалярное произведение векторов x и y , обозначаемое как $\langle x, y \rangle$ или (x, y) , часто называют *внутренним произведением*. Линейное вещественное пространство, в котором введено понятие скалярного произведения, называется *евклидовым*. Элементы x и y евклидова пространства удовлетворяют следующим условиям:

$$\begin{aligned} \hat{а}) \quad & \langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle \\ \hat{б}) \quad & \langle \alpha x, y \rangle = \alpha \langle x, y \rangle \\ \hat{в}) \quad & \langle x, y \rangle = \langle y, x \rangle \\ \hat{г}) \quad & \langle x, x \rangle \geq 0, \quad \hat{д}) \quad \langle x, x \rangle = 0, \end{aligned} \quad (3.4)$$

В евклидовом пространстве справедливо очень важное соотношение, известное как неравенство Коши-Буняковского:

$$\left| \langle x, y \rangle \right|^2 \leq \langle x, x \rangle \langle y, y \rangle, \quad (3.5)$$

которое вытекает из свойств (3.4 а, б и в) скалярного произведения. Действительно, для любого вещественного λ имеем, согласно (3.4г),

$$\langle x - \lambda y, x - \lambda y \rangle \geq 0.$$

Раскрывая левую часть неравенства, получим:

$$\langle x, x \rangle - 2\lambda \langle x, y \rangle + \lambda^2 \langle y, y \rangle \geq 0.$$

Видно, что квадратный трехчлен (относительно λ) неотрицателен, следовательно его дискриминант отрицателен (или равен нулю), т.е.

$$\langle x, y \rangle^2 - \langle x, x \rangle \langle y, y \rangle \leq 0.$$

Отсюда вытекает (3.5).

Важным следствием из свойств скалярного произведения является то, что норма вектора x может быть определена как

$$\|x\| = \sqrt{\langle x, x \rangle}. \quad (3.6)$$

Действительно, условия (3.2 а и б) для нормы удовлетворяются. Для доказательства аксиомы треугольника имеем с учетом (3.5):

$$\begin{aligned} \|x + y\|^2 &= \langle x + y, x + y \rangle = \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle \leq \\ &\leq \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| = (\|x\| + \|y\|)^2. \end{aligned}$$

Следовательно, $\|x + y\| \leq \|x\| + \|y\|$ и соотношение (3.6) удовлетворяет условию (3.2в) для нормы.

В соответствие с (3.6) можно утверждать, что скалярное произведение порождает норму, которая, в свою очередь, согласно (3.3), порождает метрику, благодаря чему пространство со скалярным произведением становится метрическим.

Взаимное отношение двух векторов в *комплексном* пространстве также может быть охарактеризовано с помощью скалярного произведения. Комплексное линейное пространство, в котором каждой паре его элементов x и y (векторов) поставлено в соответствие комплексное число $\langle x, y \rangle$ – скалярное произведение, называется *унитарным*. Элементы x и y унитарного пространства удовлетворяют тем

же условиям (3.4), за исключением (3.4в), которое записывается в ви-

$$\langle x, y \rangle = \langle y, x \rangle^*, \quad (3.7)$$

где $*$ – знак комплексного сопряжения.

Заметим, что в унитарном пространстве для любого комплексного числа λ имеем $\langle x, \lambda y \rangle = \lambda^* \langle x, y \rangle$, так как согласно (3.4 в) и (3.4 б):

$$\langle x, \lambda y \rangle = \langle \lambda y, x \rangle^* = \lambda^* \langle x, y \rangle, \text{ где } (*) \text{ – знак ком-}$$

плексного сопряжения. В унитарном пространстве, так же как в евклидовом, можно ввести норму $\|x\| = \sqrt{\langle x, x \rangle}$, удовлетворяющую всем аксиомам о норме.

3.2.1 Представление элементов векторного пространства

Любой вектор линейного пространства x может быть образован суммированием нескольких векторов u_1, u_2, \dots, u_n со скалярными коэффициентами $\alpha_1, \alpha_2, \dots, \alpha_n$

$$x = \sum_{i=1}^n \alpha_i u_i. \quad (3.8)$$

Множество всех линейных комбинаций (3.8) векторов $\{u_1, u_2, \dots, u_n\}$ также образует линейное пространство.

Множество векторов u_1, u_2, \dots, u_n называется *линейно-независимым*, если любой из них не может быть выражен через остальные с помощью линейной комбинации. Иначе говоря, когда равенство

$$\sum_{i=1}^n \alpha_i u_i = 0 \quad (3.9)$$

справедливо только при всех α_i , равных нулю.

Обозначим через V линейное пространство всех комбинаций линейно-независимых векторов $\{u_1, u_2, \dots, u_n\}$. Каждый вектор $x \in V$ соответствует единственному набору скалярных коэффициентов α_i , следовательно, на основе векторов $\{u_i; i = 1, \dots, n\}$ может быть выражен любой вектор из V . Множество векторов $\{u_i\}$, отве-

чающих условию линейной независимости, называется *базисом*, сами векторы имеют смысл осей координат, а совокупность чисел α_i – координаты вектора x .

Упорядоченная последовательность коэффициентов $\{\alpha_i\}$ может рассматриваться как n -мерная векторная строка. Это означает, что набор из n вещественных R^n или комплексных C^n чисел является представлением вектора x (в R^n или C^n) по отношению к базису $\{u_i\}$.

В одном и том же пространстве существует не одно множество линейно-независимых векторов, и соответственно, в нем может быть определен не один базис. Чтобы уточнить, какой же базис используется, говорят о пространстве V , *натянутом* на этот базис.

Одно и то же пространство может быть разбито на подпространства, так же как элементы одного множества могут быть разбиты на ряд подмножеств. Если взять подмножество $\{\alpha_1, \alpha_2, \dots, \alpha_m\}$ множества $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$, где $m < n$, то множество линейных комбинаций векторов подмножества образует линейное пространство, являющееся подпространством исходного линейного пространства.

Более строго, подмножество V_1 из V является *подпространством*, если:

- а) для всех x и y из V_1 $x + y$ также принадлежит V_1 ;
- б) для всех x из V_1 и λ из R или C λx принадлежит V_1 .

В зависимости от величины n вектор x можно рассматривать в двумерном, трехмерном и т.д., n -мерном пространстве. В двумерном пространстве два параллельных вектора u_1 и u_2 линейно-зависимы, т.к. между ними существует очевидная линейная связь $u_2 = \lambda u_1$. Два пересекающихся вектора – линейно-независимы, так как между ними нет такой связи, но любой третий вектор x можно представить в виде линейной комбинации $x = \lambda_1 u_1 + \lambda_2 u_2$ (3.2). Легко видеть, что в трехмерном пространстве будут линейно-независимы только три вектора, в четырехмерном – четыре и т.д. Максимально возможное число линейно-независимых векторов определяет *размерность* пространства. Если углы между базисными векторами равны $\pi/2$, то такая система координат является *декартовой*, а базисная система ортогональной.

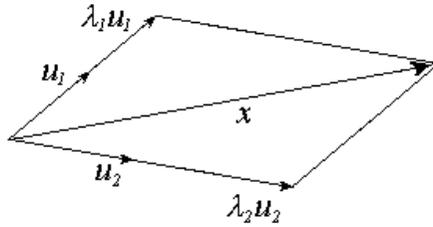


Рис. 3.2 - Представление двумерного вектора x на плоскости (u_1, u_2)

Вектор x , говорят, ортогонален множеству векторов $V_1 = \{y_i\}$, если $\langle x, y_i \rangle = 0$ для всех i . Это обозначают как $x \perp V_1$. Два подпространства V_1 и V_2 называются ортогональными, если все векторы из V_1 ортогональны всем векторам из V_2 , это записывают как $V_1 \perp V_2$.

Пусть векторы u_i в (3.8) образуют линейно-независимую систему. Тогда для определения α_i умножим скалярно левую и правую части (3.8) на векторы v_m попарно ортогональные u_n , т.е.

$$\langle u_i, v_m \rangle = \delta_{im}, \quad (3.10)$$

где δ_{im} – символ Кронекера, $\delta_{im} = 1$ для $i = m$ и $\delta_{im} = 0$ для $i \neq m$

$$\langle x, v_m \rangle = \sum_{i=1}^n \langle u_i, v_m \rangle \lambda_i. \quad (3.11)$$

Для α_i , с учетом (3.10), получаем:

$$\alpha_i = \langle x, v_i \rangle \quad (3.12)$$

Коэффициенты α_i называются коэффициентами Фурье в базисе $\{u_i\}$, а базис $\{v_m\}$, удовлетворяющий (3.10), называется *взаимным* базисом. Нетрудно видеть, что:

$$x = \sum_{i=1}^n \langle x, v_i \rangle u_i = \sum_{i=1}^n \langle x, u_i \rangle v_i \quad (3.13)$$

для любого $x \in V$ и любой пары взаимных базисов для V . Если система координатных векторов $\{u_i; i = 1, 2, \dots, n\}$ ортогональна, то такой базис называется *самовзаимным (самодуальным)* и

$$x = \sum_{i=1}^n \langle x, u_i \rangle u_i \quad (3.14)$$

Ортогональный базис с нормой координатных векторов, равных единице, называется *ортонормальным*.

$$\langle u_i, u_m \rangle = \begin{cases} 1, & i = m, \\ 0, & i \neq m. \end{cases}$$

В случае ортонормального базиса в евклидовом пространстве легко вычисляется скалярное произведение векторов через их проекции на координатные оси

$$\langle x, y \rangle = \left\langle \sum_{i=1}^n \alpha_i u_i, \sum_{m=1}^n \beta_m v_m \right\rangle = \sum_{i=1}^n \alpha_i \beta_i^* \quad (3.15)$$

и норма
$$\langle x, x \rangle = \|x\| = \sqrt{\sum_{i=1}^n |\alpha_i|^2} \quad (3.16)$$

Расстояние между двумя точками в таком пространстве равно:

$$\|x - y\| = \sqrt{\sum_{i=1}^n (\alpha_i - \beta_i)^2}. \quad (3.17a)$$

Неравенство Коши-Буняковского (3.5) имеет вид:

$$\left| \sum_{i=1}^n \alpha_i \beta_i \right|^2 \leq \sum_{i=1}^n \alpha_i^2 \sum_{i=1}^n \beta_i^2 \quad (3.17b)$$

Конечномерное векторное пространство, для которого справедливо соотношение:

$$\|x\| = \sqrt{\sum_{i=1}^n |\alpha_i|^2} < \infty,$$

обозначают $l_2(n)$. Бесконечномерное векторное пространство, для которого справедливо соотношение

$\sum_{i=1}^{\infty} |\alpha_i|^2 < \infty$, обозначают l_2 . Ясно, что если x и y входят в

l_2 , то и $y = \lambda x$ входит в l_2 при любом λ , и сумма векторов $x + y$ входит в l_2 . Для l_2 справедливы соотношения (3.15), (3.16) и (3.17) при замене n на ∞ .

3.2.2 Полнота базиса

Базисная система $\{u_n\}$ называется полной, если размерность соответствующего пространства равна размерности представляемых в этой системе векторов.

Например, если в трехмерном пространстве задан базис u_1 и u_2 , то все векторы, не лежащие в плоскости этого базиса и, следовательно, не принадлежащие двумерному пространству, будут представлены не точно, так как для их точного представления необходим еще третий базис u_3 (3.3).

Однако, иногда для упрощения приходится идти на сокращение размерности пространства, представляя в них векторы с большей размерностью.

Это достигается с помощью процедуры *ортогонального проектирования* вектора, не принадлежащего заданному пространству V_n , на это пространство.

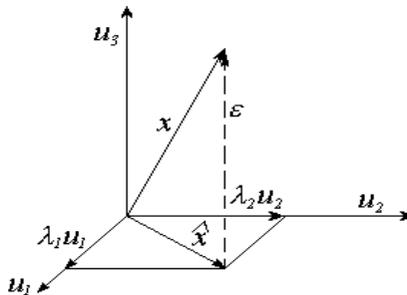


Рис. 3.3 - Представление трехмерного вектора x на плоскости (u_1, u_2) ; ϵ – ошибка представления

Это достигается с помощью процедуры *ортогонального проектирования* вектора, не принадлежащего заданному пространству V_n , на это пространство. *Теорема проектирования* утверждает, что для любого вектора $x \in V$ существует единственный вектор \hat{x} в V_n , задаваемый разложением (см. 3.13)

$$\hat{x} = \sum_{i=1}^n \langle x, v_i \rangle u_i, \quad (3.18)$$

так что разность $(x - \hat{x})$ ортогональна ко всем векторам из V_n , так

$$\begin{aligned} \langle x - \hat{x}, v_i \rangle &= \langle x, v_i \rangle - \sum_{m=1}^n \langle x, v_m \rangle \langle u_m, v_i \rangle = \\ \text{как} \quad &= \langle x, v_i \rangle - \langle x, v_i \rangle = 0; \quad i = 1, 2, \dots, n. \end{aligned}$$

Более того, для любого \tilde{x} из V_n : $\|x - \hat{x}\| < \|x - \tilde{x}\|$.

Вектор \hat{x} называется ортогональной проекцией x на V_n , $\varepsilon = x - \hat{x}$ – погрешность приближения x вектором \hat{x} . Точность приближения ε , или полнота базиса, характеризуется нормой

$$\|\varepsilon\|^2 = \|x - \hat{x}\|^2 = \|x\|^2 - \|\hat{x}\|^2. \quad (3.19)$$

Более точное определение полноты базиса и соответствующего ему пространства заключается в следующем. Пусть мы имеем последовательность векторов $\{x_n\}$ в V_n , которую называют *сходящейся* к x в V , если $\|x - x_n\| \rightarrow 0$ при $n \rightarrow \infty$. Последовательность векторов $\{x_n\}$ называется последовательностью *Коши*, если $\|x_n - x_m\| \rightarrow 0$ при $n, m \rightarrow \infty$. Если каждая последовательность Коши из V_n сходится к вектору из V , то пространство V называется *полным*. Полное бесконечномерное нормированное пространство, в котором норма порождена скалярным произведением, называется *Гильбертовым пространством* H .

3.2.3 Функциональное пространство

Эквивалентность пространств. Рассмотрим множество сигналов с ограниченной энергией, заданных на конечном (или бесконечном) интервале времени, и покажем, что на основе линейного векторного

пространства l_2 может быть сконструировано функциональное пространство, обозначим его L_2 , в котором каждой точке соответствует определенная функция времени. Два таких нормированных пространства называются алгебраически изоморфными и изометрическими, если между их элементами можно установить взаимнооднозначное соответствие, такое что:

а) алгебраическим операциям над элементами из l_2 соответствуют те же операции над их образами в L_2 ;

б) нормы соответствующих друг другу элементов из обоих пространств равны.

Можно показать, что всякое бесконечномерное сепарабельное гильбертово пространство H алгебраически изоморфно и изометрично пространству l_2 . Это означает, что множество сигналов L_2 с ограниченной энергией (или мощностью) образует пространство L_2 на основе его эквивалентности с l_2 .

Скалярное произведение в L_2 определяется как

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f^*(t) g(t) dt \quad (3.20a),$$

норма
$$\|f\| = \sqrt{\langle f, f \rangle} = \sqrt{\int_{-\infty}^{\infty} |f(t)|^2 dt} \quad (3.20б),$$

а неравенство Буняковского-Шварца имеет вид:

$$\left| \int_{-\infty}^{\infty} f(t)g(t)dt \right|^2 \leq \int_{-\infty}^{\infty} |f(t)|^2 dt \cdot \int_{-\infty}^{\infty} |g(t)|^2 dt \quad (3.20в)$$

Проиллюстрируем изоморфизм на примере конечномерных пространств. Если в векторном пространстве $l_2(n)$ существует ортонормированный базис $\{u_k\}$, то в изоморфном ему пространстве сигналов $L_2(T)$ будет существовать ортонормальная базисная система функций, обозначим ее $\{\varphi_k(t)\}$, $k = 0, 1, \dots, n-1$, т.е

$$u_k \leftrightarrow \varphi_k(t).$$

Знак (\leftrightarrow) означает взаимное и однозначное соответствие.

Тогда, в соответствии с (3.8), будет существовать разложение сигнала $s(t) \leftrightarrow x$ по системе функций $\{\varphi_k(t)\}$:

$$s(t) = \sum_{k=0}^{n-1} \lambda_k \varphi_k(t) \quad (3.21a),$$

где при условии ортонормальности $\varphi_k(t)$ аналогично (3.12)

$$\lambda_k = \langle s(t), \varphi_k(t) \rangle \quad (3.21б)$$

Определим норму сигнала как

$$\|s\| = \sqrt{\int_T |s(t)|^2 dt} \quad (3.22)$$

и покажем, что при условии (3.21) она совпадает с (3.16). Действительно,

$$\begin{aligned} \int_T |s(t)|^2 dt &= \int_T \left| \sum_{k=0}^{n-1} \lambda_k \varphi_k(t) \right|^2 dt = \\ &= \sum_{k=0}^n |\lambda_k|^2 \int_T \varphi_k^2(t) dt + \sum_{k=0}^{n-1} \sum_{\substack{m=0 \\ k \neq m}}^{n-1} |\lambda_k|^2 \int_T \varphi_k(t) \varphi_m(t) dt. \end{aligned}$$

С учетом ортонормальности базисной системы, имеем:

$$\int_T |s(t)|^2 dt = \sum_{k=0}^{n-1} |\lambda_k|^2 \int_T |s(t)|^2 dt = \sum_{k=0}^{n-1} |\lambda_k|^2 \quad (3.23)$$

Т.е. норма сигнала, определенная через его энергию, эквивалентна норме евклидоваго векторного пространства:

$$\|s(t)\| \leftrightarrow \|x\|.$$

Аналогично можно показать, что скалярное произведение двух сигналов равно сумме произведения их проекций на координатные оси

$$\langle s_1, s_2 \rangle = \int_T s_1(t) s_2^*(t) dt = \sum_{k=0}^n \lambda_k \beta_k^* \quad (3.24)$$

что эквивалентно (3.15)

$$\langle s_1, s_2 \rangle \leftrightarrow \langle x, y \rangle,$$

а расстояние между двумя сигналами равно

$$\|s_1 - s_2\| = \sqrt{\sum_{k=0}^{n-1} (\lambda_k - \beta_k)^2} \quad (3.25)$$

что эквивалентно (3.17а): $\|s_1 - s_2\| \leftrightarrow \|x - y\|$.

Выражения (3.23) и (3.24) называются равенствами *Парсеваля*. Таким образом, пространства $L_2(T)$ и $l_2(n)$ полностью эквивалентны. Эта эквивалентность пространств порождает функциональное пространство, базисом в котором является система линейно-независимых функций $\{\varphi_k(t)\}$.

Так как $L_2(T)$ образовано множеством $L_{2E}(T)$, то его называют еще пространством квадратично интегрируемых функций:

$$\int_{t \in T} |f(t)|^2 dt < \infty.$$

Условно любой сигнал в таком пространстве графически можно изобразить, как показано на рисунке 3.4.

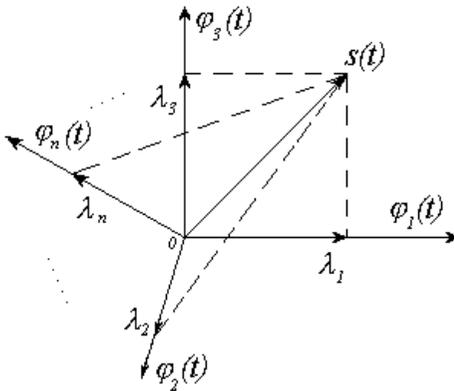


Рис. 3.4 - Представление сигнала $s(t)$ в n -мерном функциональном пространстве

Аналогично, можно геометрически представить сумму сигнала и шума $\xi(t)$ с дисперсией σ^2 (см. Рис. 3.5):

$$x(t) = s(t) + n(t).$$

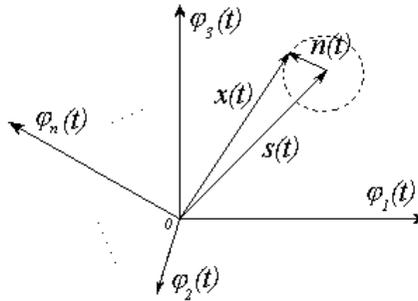


Рис. 3.5 - Представление смеси сигнала $s(t)$ с шумом $n(t)$

Так как $n(t)$ – случайный вектор, то конец вектора $x(t)$ будет лежать на гиперсфере радиуса σ .

3.3 Представление многомерных данных

Рассмотрим, далее, сигналы, состоящие из большого количества векторов, формирующих матрицу данных «объект-признак» [42], строки которой соотнесены с анализируемыми объектами или номером опыта, а столбцы - со значениями изучаемых признаков (параметров, показателей, переменных, критериев, атрибутов, реквизитов, в зависимости от принятой терминологии). Матрицу «объект-признак» обозначим $X = [x_{ij}]$, где $i = 1 \dots I$ - номера объектов; $j = 1 \dots J$ - номера признаков.

Как правило, исследователю известна технология формирования матрицы X , область допустимых значений параметров и тип шкал, по которым производятся измерения. Если в данных отсутствует значение хотя бы одного признака какого-либо объекта, то матрица X называется неполной. Кроме того, в этой матрице могут быть *неправдоподобные данные*, источниками которых являются, например, ошибки, допущенные при измерении; поломка оборудования; сбоях при передаче данных по каналам связи. Отсутствующие и неправдоподобные данные можно идентифицировать как пропуски. Существуют следующие методы восстановления, пропусков основанные на статистических методах [43,44].

1. Метод сглаживания. В этом методе для каждого признака по имеющимся данным вычисляются простейшие статистические характеристики: параметры сдвига, масштаба и коэффициент вариации. Если коэффициент вариации среднего мал (менее 0,1), то пропущен-

ное значение можно заполнить одной из следующих статистик параметра сдвига: среднеарифметическое, медиана, мода, оценки [45].

2. *Метод одномерной регрессии* (линейной, логистической, экспоненциальной, степенной, логарифмической, показательной). Переменная с пропуском берется за отклик, а за регрессор - переменная с наибольшим значением коэффициента корреляции с откликом. Полученное уравнение регрессии используется для прогноза пропущенного значения параметра. К разновидности этого метода относится метод стохастического заполнения по регрессии. К значению, предсказанному по регрессии, добавляют остаток, случайно выбранный из оцененных остатков.

3. *Метод множественной шаговой регрессии*. Метод применим, если набирается достаточно большая выборка комплектных (без пропусков) регрессоров. При применении этого метода необходима проверка соблюдения всех условий регрессионного анализа. Если матрица регрессоров плохо обусловлена, то можно применить алгоритмы регрессионного анализа для мультиколлинеарных систем, таких, как регрессия на главные компоненты, регрессия с использованием ортогональных преобразований на быстрых алгоритмах (например, дискретно-косинусное преобразование), полиномиальная регрессия заданного порядка k .

4. *Методы, основанные на подборе ближайшего соседа* без пропусков к объекту, содержащему пропуск. Здесь можно использовать классические методы распознавания образов и кластерного анализа с различными метриками и методами группировки, такими как - средней связи (*Average*), центроидный (*Centroid*), дальнего (*Furthest*) или близлежащих соседей (*Nearest*), медианный (*Median*). *Методы, основанные на построении модели совместного распределения* значений признака X и пропусков и развитии алгоритмов оценивания параметров распределения на основе метода максимального правдоподобия. Достаточно хорошо разработанные алгоритмы этого метода заполнения пропусков для различных параметрических моделей приведены в [43].

Для заполнения пропусков данных относительно прошлого хотят выждать, пока процессы развернутся во времени и регистрируют значение пропущенного признака. При таком подходе пропущенные данные могут быть заполнены значениями, относящимися к моментам резких изменений в системе. Все это может нарушить *синхронность* данных матрицы X и увеличить долю неправдоподобных наблюдений и тем самым привести к непрезентативности выборки.

Использование для восстановления пропусков *сингулярного разложения* (*Singular Value Decomposition* - SVD) исходной матрицы

данных [46]. Этот метод может быть применен при различных типах пропусков к любой положительно определенной числовой матрице. Преимуществом метода является учет не только сходства между признаками, но также и меры близости между объектами исходного множества признаков.

В задачах многомерного статистического анализа (анализ главных компонент и др.) в матрице «объект-признак» признаки принято считать случайными величинами: $X_j, j = 1, \dots, J$, степень связи между которыми определяется ковариационной и корреляционной матрицей.

. Многие методы многомерного анализа данных, включая и множественную линейную регрессию, анализ главных компонент, факторный анализ, дискриминантный и канонический корреляционный анализы построены на преобразованиях исходной матрицы «объект-признак» в ковариационную или корреляционную матрицу. Ковариационная матрица - это квадратная матрица типа «признак-признак» размера $J \times J$, образованная из попарных ковариаций J случайных признаков X_1, \dots, X_J матрицы «объект-признак» X . Ковариационная матрица K_x для матрицы X определяется как

$$K_x = M[(X - MX)'(X - MX)], \quad (3.26)$$

где M - символ математического ожидания; $MX = (MX_1, \dots, MX_J)$ - вектор средних значений. Компоненты ковариационной матрицы

$$\begin{aligned} k_{ij} &= M[(X_i - MX_i)'(X_i - MX_i)] = \text{cov}(X_i, X_j) = \\ &= \text{cov}(X_i X_j) \quad i, j = 1, \dots, J, \end{aligned} \quad (3.27)$$

причём при $i = j$ величина k_{ij} совпадает с дисперсией X_j .

Ковариационная матрица представляет собой симметричную неотрицательно определенную матрицу. Пусть признаки X_j линейно независимы в том смысле, что ее все нетривиальные линейные комбинации вида $\sum_j a_j X_j$; (a_j - компонента произвольного вектора) являются невырожденными случайными величинами, т. е. не являются константами и имеют положительную дисперсию. Тогда ковариационная матрица матрицы наблюдений X_j положительно определена. Если дисперсии признаков X_1, \dots, X_J равны 1, то ковариационная матрица называется *корреляционной матрицей*

Оценка элементов ковариационной матрицы для матрицы наблюдений X_j размерности $I \times J$ находится по формуле:

$$\hat{k}_{ij} = \frac{1}{I-1} \sum_{s=1}^I (x_{si} - \bar{x}_i)(x_{sj} - \bar{x}_j), \quad i, j = 1, \dots, J, \quad (3.28)$$

где $\bar{x}_i = (1/I) \sum_{s=1}^I x_{si}$ среднее признака X_i .

Оценки элементов r_{ij} корреляционной матрицы вычисляются по формуле

$$\hat{r}_{ij} = \hat{k}_{ij} / S_i S_j, \quad (3.29)$$

где $S_i = (k_{ii})^{1/2}$ и $S_j = (k_{jj})^{1/2}$ - среднеквадратические отклонения признаков X_i и X_j . Если признаки X_1, \dots, X_J имеют нормальное распределение с ковариационной матрицей K_i , то оценка \hat{K}_x , является оценкой максимума правдоподобия.

Все приведенные выше формулы оценок ковариационной матрицы справедливы для матрицы наблюдений «объект-признак» без пропусков.

Представление двумерных данных в многомерном пространстве, синтезируемом на основе знания корреляционных связей между признаками, рассматривается в следующей главе вместе с методами их обработки.

3.4 Меры близости

Во многих задачах классификации важно ввести количественную оценку подобия или схожести объектов. Для этой цели обычно используется расстояние между точками многомерного пространства: большему расстоянию соответствует меньшее сходство [7]. Меры расстояния не ограничены сверху и зависят от выбора масштаба измерений. Если исходные данные заданы матрицей «объект-признак» X_i , то возможно вычисление различных типов расстояний между объектами X_i и X_j . Одним из наиболее часто применяемых расстояний является *расстояние Минковского*.

$$d_{ij} = d(X_i, X_j) = \left(\sum_{k=1}^J w_k |x_{ik} - x_{jk}|^q \right)^{1/q} \quad (3.30)$$

где $w_k > 0$ - весовые коэффициенты при признаках; $q \geq 1$ - параметр степени. Веса выражают степень важности измерения для различении объектов. При сопоставлении двух объектов измерения признака с меньшей значимостью нужно присваивать меньшие веса.

Это расстояние нужно применять тогда, когда все признаки матрицы X взаимонезависимы и однородны, например, измерены в одних и тех же единицах и одинаково важны для решаемой задачи. Частными случаями расстояния Минковского являются:

а) евклидово расстояние

$$d_{ij} = \left(\sum_{k=1}^J (x_{ik} - x_{jk})^2 \right)^{1/2}, w_k = 1, q = 2, \quad (3.31)$$

б) манхеттингское расстояние (расстояние «Сити-блок»)

$$d_{ij} = \sum_{k=1}^J |x_{ik} - x_{jk}|, w_k = 1, q = 1 \quad (3.32)$$

в) доминантное расстояние

$$d_{ij} = \max_k |x_{ik} - x_{jk}|, w_k = 1, q = \infty \quad (3.33)$$

Расстояние Махаланобиса

$$d_{ij} = (X_i - X_j)' K_x^{-1} (X_i - X_j), \quad (3.34)$$

где X_i, X_j - векторы признаков объектов i и j ; K_x^{-1} - обратная корреляционная матрица признаков. Это расстояние применяется при сильной зависимости и физической неоднородности исследуемых признаков. В отличие от евклидовой метрики и метрики Минковского, эта метрика через ковариационную матрицу связана с корреляциями признаков. Если признаки некоррелированы, то расстояние Махаланобиса эквивалентно евклидовому расстоянию.

Несмотря на важность рассмотренных метрик, они имеют серьезные недостатки. Наиболее существенным недостатком является то, что оценка сходства сильно зависит от различий в сдвигах признаков. Если одно-временно велики абсолютные значения и стандартные отклонения признаков, то эти признаки могут подавить влияние переменных с меньшими абсолютными значениями и стандартными отклонениями. Для уменьшения указанных влияний нужно использовать нормированные признаки с нулевыми средними и единичными дисперсиями. Однако нормировка также может сильно повлиять на меру сходства.

3.5 Аппаратная функция и модели сигналов

Как уже было отмечено выше, одной из главных потребительских характеристик аналитического прибора для многокомпонентного анализа веществ является его разрешающая способность. Это – способность прибора различить две и более компонентов вещества с близкими параметрами разделения – массами, молекулярным весом,

размером, зарядом и др. Мысленно подадим на вход прибора несколько очень «узких» или «коротких» входных сигнала с близкими, например, массовыми числами. На выходе прибора будем наблюдать пики с существенно размытыми фронтами, которые приводят к их наложению вплоть до исчезновения седловины между ними. Последнее означает, что анализируемые компоненты разделить невозможно, т.е. такой прибор обладает недостаточной разрешающей способностью.

Отклик прибора на входной бесконечно “узкий” или “короткий” входной сигнал называют аппаратной функцией.

Проиллюстрируем это понятие на примере иона кислорода в масс-спектрометрическом анализаторе. Источник ионов испускает ионы O^+ , масса которых равна 16 а.е.м. (абсолютные единицы массы). У кислорода есть три устойчивых изотопа с массами 16, 17 и 18, но содержание двух последних в атмосфере составляет 0,25% по отношению к первому, и на графике они не видны из-за имеющихся в анализируемой смеси паров воды H_2O ($m = 18$) и гидроксила $(OH)^+$ ($m = 17$). Если бы наш масс-спектрометр был “идеальным” прибором, то зарегистрированный пик кислорода выглядел бы как тонкая вертикальная линия. В действительности он имеет форму, близкую к треугольной. Легко сообразить, что это связано с тем, что ионы вылетают из источника с конечными размерами и регистрируются на коллекторе таких же габаритов (см. рис. 1.5).

В процессе развертки при некотором значении ускоряющего напряжения на коллектор попадает край ионного пучка, и самописец регистрирует ток. Максимум сигнала достигается, когда весь пучок прилетает на коллектор. Далее, по мере изменения ускоряющего напряжения, часть пучка, попадающая на коллектор, снова начнет уменьшаться, а сигнал – падать. Таким образом, аппаратная функция масс-спектрометра должна иметь форму, близкую к пику с острой вершиной. Зная размер щели источника ионов и коллектора, а также ускоряющее напряжение для данного типа ионов, можно рассчитать ширину этого пика на полувысоте в атомных единицах массы.

Ясно, что чем шире аппаратная функция прибора, тем хуже его разрешение по массам, тем труднее отличить или, как говорят, разрешить два близко расположенных пика. Существует, к сожалению, множество факторов, способных внести дополнительное уширение в регистрируемый пик. Если прибор сконструирован правильно, то эти искажения достаточно малы, как в нашем случае. Тем не менее, полезно иметь представление о физических механизмах, влияющих на качество работы прибора. В частности, иногда удается теоретически оценить влияние на аппаратные функции электростатических анализа-

торов флуктуаций потенциалов [47], или вычислить аппаратную функцию аксиальных электростатических анализаторов [48].

В целом ряде случаев исследователю доступна некоторая априорная информация о сути физических процессов, происходящих в исследуемом явлении, на основании которой может быть составлено представление о виде ожидаемых сигналов: форме, продолжительности, количестве пиков и т.п. Чаще всего при этом удается подобрать аналитическое выражение этому сигналу, называемое *моделью*. Моделью является квазидетерминированный сигнал, который можно выразить в виде функции с параметрами — $f(t, \theta)$. Численные значения параметров θ определяют все множество сигналов, описываемых данной моделью. Выбор модели оказывает значительное влияние на достоверность извлекаемой из сигнала информации об исследуемом явлении. Примерами простейших моделей является гауссовая

$$f(t) = A \exp\left\{-\frac{(t-t_0)^2}{2\mu_0^2}\right\} \quad (3.33)$$

или лоренцовая

$$f(t) = A \frac{1}{1 + 4\left(\frac{t-t_0}{\mu_0}\right)^2}. \quad (3.34)$$

Здесь t_0 - положение максимума пиков, μ_0 - их ширина.

Гауссовую модель чаще всего используют в масс-спектрометрии, хроматографии, лоренцова - характерна для мёссбауэровской спектроскопии, в инфракрасной спектроскопии.

4 Проекционные методы обработки данных

Проекционные методы обработки предполагают, что исходные данные на первом этапе представляют в некотором подходящем пространстве, в котором информативные параметры сигнала выделяется более четко и их количественная оценка может быть выполнена с меньшей ошибкой и большей достоверностью. Следует отметить следующие хорошо зарекомендовавшие себя способы таких преобразований: преобразование Фурье, вейвлет-преобразование, спектральный анализ в обобщенном базисе, метод главных компонент.

Если первые три метода представления сигналов хорошо изучены и зарекомендовали себя как в обработке одномерных данных, так и изображений [19,20,49-51], то последний следует рассмотреть более подробно.

4.1 Метод главных компонент

Современные приборы производят огромное количество измерений. Однако, из-за мультиколлинеарности, доля полезной информации в таком массиве может быть относительно невелика. Для выделения полезной информации, при многомерном подходе, используются методы сжатия данных (в отличие от традиционного подхода, когда из данных выделялись только отдельные особо значимые измерения). Идея этих методов состоит в том, чтобы представить исходные данные физического эксперимента, используя новые скрытые переменные. Как правило, число новых переменных бывает существенно меньше, чем число исходных переменных, но потери от такого сжатия данных обычно сопоставимы с шумом. Сжатие данных позволяет представить полезную информацию в более компактном виде, удобном для визуализации и интерпретации.

Метод главных компонент – МГК (Principal Component Analysis, PCA) является наиболее популярным способом сжатия данных [8] и не только. Он лежит в основе других аналогичных проекционных методов, включая эволюционный факторный анализ (EFA)], оконный факторный анализ, итерационный целевой факторный анализ (ITTFA), а также многих методов классификации, например, формального независимого моделирования аналогий классов (SIMCA) [18].

МГК основан на статистических подходах к обработке данных. Многомерные данные представляют в виде матрицы «объект-признак», строки которой соотнесены с анализируемыми объектами, а столбцы – с изучаемыми признаками (см. п. 3). Примерами признаков многомерных данных могут служить: лабораторные анализы продуктов питания или лекарственных препаратов, параметры кардиограммы, в метеорологии – температура воздуха, тип облачности, влажность, скорость ветра и др.

При анализе многомерных данных будем считать, что каждый признак является случайной величиной. Тогда данные типа «объект-признак» размерности $I \times J$, в которых признаки X_j , $j = 1, 2, \dots, J$ являются случайными величинами, удобно представлять в виде I -мерной случайной векторной переменной X .

Обработка многомерных данных включает задачи классификации, создания новой структуры признакового пространства и интер-

претации их, хранения, передачи по каналам связи. Решение этих задач значительно упрощается, если подвергнуть сжатию размерность признакового пространства. Такое сжатие возможно, поскольку в большинстве случаев признаки взаимосвязаны (коррелированы) и, следовательно, данные избыточны с точки зрения информации, и эта избыточность полностью определяется корреляционной матрицей исходных переменных X [6]. Сжатие сводится к преобразованию J -мерного исходного пространства данных X в другое пространство с Y , в котором можно выбрать подмножество латентных (ненаблюдаемых) переменных меньшей размерности $A < J$ без существенной потери информации. В некоторых случаях латентным переменным удается придать ясный физический смысл, например, синдромы болезни в медицине, тип климата в метеорологии, двигательные характеристики в спорте (выносливость и скоростные способности).

Выбор вида преобразования $Y = f(X)$ и число латентных переменных A зависит от конкретной специфики решаемой задачи и должен опираться на критерий, который обеспечивает заданную величину погрешности восстановления заданных свойств X по сжатому образу \hat{Y} . Для осуществления перехода к новым переменным Y используют статистические свойства матрицы X . Если данные имеют многомерное нормальное распределение, то эти свойства определяются ковариационной (корреляционной) матрицей типа «признак-признак».

Если переменные Y находят из условия сохранения большей части суммарной дисперсии, то применяют ортогональные преобразования. Из множества ортогональных преобразований рассматривается оптимальное преобразование - метод главных компонент (МГК), субоптимальное дискретно-косинусное преобразование (ДКП) с быстрым алгоритмом и вейвлет-преобразование (ВП).

МГК был предложен К. Пирсоном в 1901 г. и затем вновь открыт и детально разработан Хоттелингом в 1933 г. МГК посвящено большое количество работ, и он широко представлен в статистической литературе, обратившись к которой можно получить информацию об МГК с различной степенью детализации [2, 52].

Ограничимся рассмотрением классического метода нахождения главных компонент. В этом методе ГК определяются собственными векторами, которые соответствуют наибольшим собственным числам ковариационной (корреляционной) матрицы исходных данных.

Пусть задана $(I \times J)$ -матрица наблюдений случайной векторной переменной $X = [X_1 \dots X_J]$ (здесь X_j - I -мерный вектор - столбец) с

вектором средних $\mu_x = [\mu_1, \dots, \mu_J]$ и ковариационной матрицей K_x , определяющей структуру зависимости между переменными X_j ,

$j = 1, \dots, J$. Нужно найти такое линейное преобразование, которое позволило бы получить представление исходных данных меньшим числом латентных переменных без существенной потери информации, содержащейся в исходной матрице. Преобразуем эти наблюдения $(J \times J)$ -ортогональной матрицей вида

$$\Phi' = [\varphi_1 \varphi_2 \dots \varphi_J] \quad (4.1)$$

где $\varphi_j = [\varphi_{1,j}, \dots, \varphi_{I,j}]'$ - векторы-столбцы, система J -мерных ортонормированных векторов и для скалярного произведения $(*, *)$ справедливо:

$$(\varphi_i, \varphi_j) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (4.2)$$

В результате преобразования получаем случайную векторную переменную - матрицу Y с некоррелированными компонентами

$$Y' = \Phi X', \quad (4.3)$$

или

$$Y' = X \Phi'. \quad (4.4)$$

Из условия (4.1) и (4.2) следует, что $\Phi' \Phi = I$ (I - единичная матрица), и, следовательно,

$$X = Y \Phi = Y [\varphi_1' \varphi_2' \dots \varphi_J'] = Y_1 \varphi_1' + Y_2 \varphi_2' + \dots + Y_J \varphi_J' \quad (4.5)$$

Ковариационная матрица данных X (по определению) равна

$$K_x = M \left\{ (X - \mu_x)' (X - \mu_x) \right\} \quad (4.6)$$

Определитель $|K_x|$ ковариационной матрицы K_x называется обобщенной дисперсией матрицы данных X .

Ковариационная матрица K_y случайной векторной переменной Y определяется выражением:

$$\begin{aligned} K_y &= M \left\{ (Y - \mu_y)' (Y - \mu_y) \right\} = M \left\{ \Phi (X - \mu_x)' (X - \mu_x) \Phi' \right\} = \\ &= \Phi M \left\{ (X - \mu_x)' (X - \mu_x) \right\} \Phi' = \Phi K_x \Phi' \end{aligned} \quad (4.7)$$

Так как K_x и Φ являются квадратными матрицами и определитель произведения квадратных матриц равен произведению определителей

перемножаемых матриц, то определитель ковариационной матрицы \mathbf{K}_y равен

$$|\mathbf{K}_y| = |\Phi \mathbf{K}_x \Phi'| = |\Phi \Phi'| |\mathbf{K}_x| = |\mathbf{K}_x|, \quad (4.8)$$

следовательно, обобщенные дисперсии матриц \mathbf{X} и \mathbf{Y} равны.

Наилучшее ортогональное преобразование должно обеспечить наименьшую избыточность. Результатом такого преобразования является векторная переменная \mathbf{Y} с некоррелированными компонентами Y_j , $j = 1, \dots, J$ и диагональной матрицей \mathbf{K}_y .

$$\mathbf{K}_y = \text{diag}[\sigma_{y_1}^2, \dots, \sigma_{y_J}^2] \quad (4.9)$$

где $\sigma_{y_j}^2$ - дисперсия j -й компоненты случайной векторной переменной \mathbf{Y} . Обозначим $\lambda_j = \sigma_{y_j}^2$, $j = 1, \dots, J$. Тогда

$$|\mathbf{K}_y| = \prod_{j=1}^J \lambda_j \quad (4.10)$$

Предположим, что дисперсии упорядочены: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_J \geq 0$.

Если не все λ_j равны между собой, то матрицу \mathbf{Y} можно сжать отбрасыванием компоненты с пренебрежимо малыми дисперсиями.

Компонента Y_1 называется первой ГК матрицы \mathbf{X}

$$Y_1 = \mathbf{X} \varphi_1$$

Используя формулу (4.7), найдем дисперсию этой ГК

$$\sigma_{y_1}^2 = \varphi_1' \mathbf{K}_x \varphi_1.$$

Потребуем, чтобы компонента Y_1 имела наибольшую дисперсию при условии сохранения ортогональности векторов φ_j . Тогда задача поиска наилучшего преобразования φ_1 , сводится к нахождению максимума функции $\varphi_1' \mathbf{K}_x \varphi_1$ при условии

$$(\varphi_1, \varphi_1) = \sum_{j=1}^J \varphi_{1,j}^2 = 1$$

Чтобы решить эту задачу оптимизации, введем функцию Лагранжа

$$L(\varphi) = \varphi_1' \mathbf{K}_x \varphi_1 - \lambda_1 (\varphi_1' \varphi_1 - 1)$$

где λ_1 - множитель Лагранжа.

Необходимое условие экстремума получим, приравняв нулю частные производные $\partial L / \partial \varphi_1$:

$$\partial L / \partial \varphi_1 = 2(\mathbf{K}_x \varphi_1 - \lambda_1 \varphi_1) = 2(\mathbf{K}_x - \lambda_1 \mathbf{I}) \varphi_1 = 0 \quad (4.11)$$

где \mathbf{I} - единичная матрица. Поскольку нас интересуют только решения, при которых $\varphi_1 \neq 0$, то должно удовлетворяться условие на определитель

$$|\mathbf{K}_x - \lambda_1 \mathbf{I}| = 0.$$

Отсюда следует, что λ_1 есть собственное число ковариационной матрицы \mathbf{K}_x , а φ_1 - соответствующий этому числу собственный вектор. Выражение (4.11) можно переписать в виде:

$$\mathbf{K}_x \varphi_1 = \lambda_1 \varphi_1$$

умножая которое слева на φ_1' и, учитывая выражение (4.2), получаем

$$\varphi_1' \mathbf{K}_x \varphi_1 = \lambda_1 \varphi_1' \varphi_1 = \lambda_1 \quad (4.12)$$

Левая часть равенства (4.12) есть $\sigma_{y_1}^2$, а поскольку решалась задача максимизации $\sigma_{y_1}^2$, то λ_1 есть λ_{\max} матрицы \mathbf{K}_x . Чтобы найти вторую ГК $Y_2 = \mathbf{X} \varphi_2$, потребуем выполнения двух условий: нормировки - $(\varphi_2, \varphi_2) = \sum_{j=1}^p \varphi_{2,j}^2 = 1$ и ортогональности - $\varphi_1' \varphi_2 = 0$. Вектор

φ_2 определяется так, чтобы дисперсия $\sigma_{y_2}^2$ была максимальна при выполнении двух указанных условий. Эта задача требует использования двух множителей Лагранжа λ_2 и β . Определяем максимум выражения

$$\varphi_2' \mathbf{K}_x \varphi_2 - \lambda_2 (\varphi_2' \varphi_2 - 1) - \beta (\varphi_1' \varphi_2 - 0).$$

Взяв производную по φ_2 от этого выражения и приравняв ее к 0, находим в соответствии с условием (4.2), что $\beta = 0$. Учитывая условия нормировки, получаем, что есть второе по величине собственное число матрицы \mathbf{K}_x , равное дисперсии второй ГК $\lambda_2 = \sigma_{y_2}^2$, а φ_2 - соответствующий ему собственный вектор. Процесс повторяется до тех пор, пока не будут найдены все собственные числа и ассоциированные с ними собственные векторы, которые являются дисперсиями и коэффициентами линейных комбинаций ГК.

Таким образом, столбцы ортогонального преобразования матрицы Φ являются собственными векторами ковариационной матрицы K_x и они удовлетворяют соотношению

$$K_x \varphi_r = \lambda_r \varphi_r, \quad (4.13)$$

где λ_r - соответствующие собственные числа.

С точки зрения геометрической интерпретации это ортогональное преобразование означает вращение системы координат в J -мерном пространстве.

Относительный вклад компоненты Y_j , в общую дисперсию случайной векторной переменной Y равен

$$\sigma_{y_j}^2 / \sum_{i=1}^J \sigma_{y_i}^2 = \lambda_j / \sum_{i=1}^J \lambda_i = \lambda_j \text{tr} K_x, \text{ где } \text{tr} K_x - \text{след матрицы } K_x$$

Полученное преобразование максимизирует дисперсию первых ГК Y_j , что обеспечивает наилучшее сжатие. Это преобразование в системах сжатия данных называют преобразованием Карунена—Лоэва (ПКЛ) [44].

Если в пространстве Y ограничиться A первыми ГК, а остальные $J - A$ компонент положить равными нулю, то получим сжатую матрицу случайной векторной переменной \hat{Y} с A ГК. Из A ($A < J$) собственных векторов формируется ($A \times J$)-матрица ортогонального преобразования $\Phi' = [\varphi_1, \dots, \varphi_A]$ и определяется матрица $\hat{Y} = X\Phi'$.

При этом обобщенная дисперсия уменьшится на величину $\prod_{j=A+1}^J \lambda_j$

Анализ ГК является методом сокращения числа переменных. Возникает вопрос: сколько ГК следует выделять? Для последовательного выделения числа компонент можно воспользоваться дисперсионным критерием. Решение об остановке процедуры выделения ГК зависит от того, что считать малой долей дисперсии. Это решение достаточно произвольно, однако имеются два критерия: *критерий Кайзера (Kaiser)* и *критерий «каменистой осыпи» Кэттелла (Cattell)*, которые в большинстве случаев помогают рационально выбрать число компонент.

При использовании критерия Кайзера отбираются только компоненты с $\lambda_j > 1$. Критерий удобен при нахождении ГК по корреляционной матрице исходных данных. Критерий «каменистой осыпи» является графическим методом. В этом случае строится график соб-

ственных чисел в порядке убывания их величин. Выделение компонент заканчивается на той компоненте, после которой исследуемая зависимость близка к горизонтальной и похожа на «каменистую осыпь» обломков горных пород, скапливающихся в нижней части скалистого склона.

Критерий Кайзера часто сохраняет слишком много факторов, в то время как критерий каменной осыпи иногда выделяет слишком мало факторов; однако оба критерия дают хорошие результаты, когда имеется относительно небольшое число компонент и много переменных. На практике принимается тот критерий, для которого полученное число ГК может быть содержательно интерпретировано. Поэтому обычно исследуется несколько решений с большим или меньшим числом компонент, и затем выбирается одно наиболее «осмысленное».

С учетом сокращения числа главных компонент (сжатия данных) выражение (4.5) можно записать в виде:

$$X = Y\Phi + E, \quad (4.14)$$

где E - матрица *ошибки* размерности $I \times J$.

В приложениях к аналитической химии – *хеометрике*, выражение (4.14) записывают в других обозначениях и матрицы, входящие в это выражение, получили другое название [8,18]:

$$X = TP + E = \sum_{j=1}^A \mathbf{t}_j \mathbf{p}_j' + E, \quad (4.15)$$

где матрица T (в предыдущем обозначении - \hat{Y}) называется матрицей *счетов* (scores), ее размерность - $(I \times A)$; матрица P (в предыдущем обозначении - Φ) называется матрицей *нагрузок* (loadings), ее размерность $(J \times A)$; E - это матрица *остатков*, размерностью $(I \times J)$; \mathbf{t}_j - называются *главными компонентами* (Principal Components), A - число главных компонент. Число столбцов $A - \mathbf{t}_j$ в матрице T и \mathbf{p}_j в матрице P – равно эффективному рангу матрицы X . Величина A заведомо меньше числа переменных J и числа образцов I . Выражение (4.13) в новых обозначениях можно записать в виде

$$K_x \mathbf{p}_r = \lambda_r \mathbf{p}_r$$

или для центрированных исходных данных X

$$X X \mathbf{p}_r = \lambda_r \mathbf{p}_r$$

При этом векторы счетов \mathbf{t}_j вычисляется как $\mathbf{t}_j = X \mathbf{p}_j$, т.е. являются линейной комбинацией исходных значений матрицы X . Столбцы

\mathbf{t}_j ортогональны и $\mathbf{t}'_j \mathbf{t}_j = \lambda_j$, а столбцы матрицы \mathbf{P} ортонормированы, т.е. $\mathbf{P}'\mathbf{P} = \mathbf{I}$ – единичной матрице.

Важным свойством МГК является *ортогональность* (независимость) главных компонент. Поэтому при увеличении числа компонент в матрице счетов \mathbf{T} прибавляется еще один столбец, соответствующий новому направлению. То же происходит и с матрицей нагрузок \mathbf{P} .

Метод главных компонент можно интерпретировать как проецирование данных на подпространство меньшей размерности. При этом матрицу исходных данных $X(I \times J)$ рассматривают как набор из I точек (строк в матрице X , изучаемых объектов) в пространстве размерности J (столбцов матрицы X), т.е. $X_j \in \mathbf{R}^J$. Метод главных компонент проецирует точки X_j на подпространство меньшей размерности A , переходя от матрицы $\mathbf{X}(I \times J)$ к матрице счетов $\mathbf{T}(I \times A)$, где $\mathbf{t}_j \in \mathbf{R}^A$ (обычно $A < J$). Всё это означает, что МГК проецирует точки из пространства \mathbf{R}^J в пространство \mathbf{R}^A . Возникающие при этом остатки \mathbf{E} рассматриваются как шум, не содержащий значимой физической информации. В этом подпространстве можно ввести меру близости объектов (см. п. 3.4), с помощью которой удастся решить многие проблемы качественного анализа. Другим мощным инструментом анализа данных в проекционном подпространстве является *прокрустово* (Procrustes) вращение [53].

Одним из важнейших преимуществ проекционных методов, в том числе и МГК, является возможность представить сложные данные физического эксперимента в более простом виде, так чтобы исследователь смог «увидеть» результаты экспериментов в простой графической интерпретации. В этом контексте МГК можно рассматривать как прием, при котором данные заключаются в эллипсоид, и вычисляются главные оси этого эллипсоида. В результате, первые две главные оси (две первые главные компоненты) образуют плоскость, при проекции на которую видны наибольшая вариация в данных.

Каждая следующая ось эллипсоида будет меньше или равна предыдущей, и таким образом, в этом направлении вариации в данных будут все менее заметны. В общем случае, если структура данных такова, что в них имеются какие-то группы или кластеры, обычно это видно при исследовании проекций, построенных для первых нескольких компонент. Поэтому, при исследовании данных методом МГК, особое внимание уделяется графикам счетов и нагрузок. Они несут в себе информацию, полезную для понимания того, как устроены дан-

ные. На графике счетов каждый объект изображается в координатах (t_i, t_j) , чаще всего – (t_1, t_2) .

Близость двух точек означает их схожесть, т.е. положительную корреляцию. Точки, расположенные на линиях под прямым углом, являются некоррелированными, а расположенные диаметрально противоположно – имеют отрицательную корреляцию.

Применяя этот подход в задачах хроматографического анализа [54] можно, например, установить, что линейные участки на графике счетов соответствуют областям чистых компонентов на хроматограмме, искривленные участки представляют области наложения пиков, а число таких участков соответствует числу различных компонентов в сложном кластере. Если график счетов используется для анализа взаимоотношений объектов, то график нагрузок применяется для исследования роли переменных. На графике нагрузок каждая переменная отображается точкой в координатах (p_i, p_j) , например (p_1, p_2) .

Анализируя его аналогично графику счетов, можно понять, какие переменные связаны, а какие независимы. Совместное исследование парных графиков счетов и нагрузок, также может дать много полезной информации о данных: [8].

4.2 Преобразование Фурье

Разложение сигнала на элементарные составляющие для их рассмотрения по отдельности либо для сравнения свойств различных сигналов. Такое разложение производится с использованием рядов и интегральных преобразований, важнейшими среди которых являются ряд Фурье и преобразование Фурье.

Существуют два способа описания произвольного сигнала. Первый способ основан на математическом представлении сигнала в виде $y=f(t)$, где независимая переменная t —время. Математическое представление сигнала по второму способу имеет вид $Y=F(v)$, где независимая переменная v —частота (размерность обратная размерности t). Эти два представления сигнала связаны друг с другом преобразованием Фурье. Следовательно, использование методов обработки сигнала требует знания свойств преобразования Фурье. Преобразование Фурье над функциями времени может быть применено и для функций других переменных. Например, если независимая переменная—длина, то переменной, аналогичной частоте, будет величина, обратно пропорциональная длине и т. д. Различают понятия “преобразование Фурье” и “ряд Фурье”. Преобразование Фурье предполагает непрерывное распределение частот, ряд Фурье задается на дискретном наборе частот.

Сигналы также могут быть заданы в наборе временных отсчетов или как непрерывная функция времени.

Наиболее практична с точки зрения цифровой обработки сигналов дискретизация и во временной, и в частотной области, но не следует забывать, что она является аппроксимацией непрерывного преобразования. Непрерывное преобразование Фурье позволяет точно представлять любые явления. Сигнал, представленный рядом Фурье, может быть только периодичен. Сигналы произвольной формы могут быть представлены рядом Фурье только приближенно, т.к. при этом предполагается периодическое повторение рассматриваемого интервала сигнала за пределами его задания. На стыках периодов при этом могут возникать разрывы и изломы сигнала, и возникать ошибки обработки, вызванные явлением Гиббса, для минимизации которых применяют определенные методы (весовые окна, продление интервалов задания сигналов, и т.п.).

При дискретизации и во временной, и в частотной области, обычно говорят о дискретном преобразовании Фурье (ДПФ):

$$S(n) = \sum_k s(k) \exp(-j2\pi nk / N)$$

где N - количество отсчетов сигнала. Применяется оно для вычисления спектров мощности, оценивания передаточных функций и импульсных откликов, быстрого вычисления сверток при фильтрации, расчете корреляции, расчете преобразований Гильберта, и т.п. Расчет ДПФ по приведенной формуле требует вычисления n коэффициентов, каждый из которых зависит от k элементов исходного отрезка, так что число операций не может быть меньше nk . Существует целое семейство алгоритмов, известное, как “Быстрое Преобразование Фурье” - БПФ, сокращающее число операций для вычисления коэффициентов до $n \log(k)$. “Быстрое” не следует трактовать, как “упрощенное” или “неточное”. При точной арифметике результаты расчетов ДПФ и по алгоритмам БПФ совпадают.

Преобразование Фурье позволяет вычислить спектральную плотность сигнала, представляющего собой функцию (как правило, времени либо пространственных координат). Дискретный же сигнал является последовательностью чисел. Традиционно он представляется последовательностью отсчетов в виде дельта-функций с соответствующими множителями и задержками. Для последовательности отсчетов $\{x(k)\}$ получится следующий сигнал:

$$s(t) = \sum_{k=-\infty}^{k=\infty} x(k) \delta(t - k)$$

Преобразование Фурье линейно, спектр дельта-функции равен единице, а задержка сигнала во времени приводит к умножению спектра на комплексную экспоненту. Все это позволяет сразу же записать спектр дискретного сигнала:

$$\dot{S}(\omega) = \sum_{k=-\infty}^{k=\infty} x(k)e^{-j\omega k}$$

Из этой формулы видно главное свойство спектра любого дискретного сигнала спектр является периодическим, и его период в данном случае равен 2π (то есть круговой частоте дискретизации, поскольку, составляя сигнал из дельта-функций, мы выбрали единственный интервал между ними, что дает: $\omega_s = 2\pi$

$$\dot{S}(\omega \pm 2\pi) = \dot{S}(\omega)$$

Пусть значения $\{x(k)\}$ являются отсчетами аналогового сигнала $s(t)$, взятыми с периодом T : $x(k) = s(kT)$

$$s_s(t) = \sum_{k=-\infty}^{k=\infty} s(kT)\delta(t - kT), \quad s_s(t) = s(t) \sum_{k=-\infty}^{k=\infty} \delta(t - kT)$$

В связи с периодичностью дискретной функции. Коэффициенты ряда Фурье могут быть вычислены по формуле:

$$\dot{C}_n = \frac{1}{T} \int_{-T/2}^{T/2} \delta(t)e^{-j\omega_n t} dt = \frac{1}{T}$$

В формуле было учтено, что в интервал интегрирования $(-T/2, T/2)$ попадает только одна дельта-функция, соответствующая $k=0$. Таким образом, периодическая последовательность дельта-функций может быть представлена в виде комплексного ряда Фурье:

$$\sum_{k=-\infty}^{\infty} \delta(t - kT) = \frac{1}{T} \sum_{n=-\infty}^{\infty} e^{j\omega_n t}, \quad \omega_s = 2\pi/T$$

После преобразования предыдущих формул:

$$s_s(t) = \frac{s(t)}{T} \sum_{k=-\infty}^{\infty} e^{j\omega_s t} = \frac{1}{T} \sum_{k=-\infty}^{\infty} s(t)e^{j\omega_s t}$$

$$\dot{S}(\omega) = \frac{1}{T} \sum_{n=-\infty}^{\infty} \dot{S}\left(\omega - \frac{2\pi n}{T}\right)$$

Таким образом, спектр дискретизированного сигнала представляет собой бесконечный ряд сдвинутых копий спектра исходного непрерывного сигнала. Расстояние по частоте между соседними копиями спектра равно частоте дискретизации $\omega_s = \frac{2\pi}{T}$.

Характер спектра дискретизированного сигнала еще раз демонстрирует частотно-временную дуальность преобразования Фурье: периодический сигнал — дискретный спектр; периодический спектр — дискретный сигнал.

Периодичность спектра дискретного сигнала иллюстрируется на Рис. 4.1

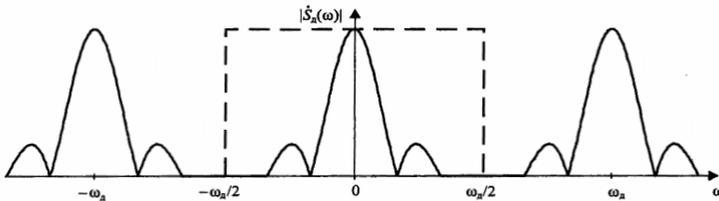


Рис. 4.1 Периодичность спектра дискретного сигнала

Для восстановления непрерывного сигнала по дискретным отсчетам его необходимо пропустить через идеальный фильтр нижних частот (ФНЧ) с частотой среза, равной половине частоты дискретизации. АЧХ такого фильтра показана на рисунке 4.31.

Точное восстановление сигнала возможно, если сдвинутые копии спектра не перекрываются. Для этого необходимо, чтобы частота дискретизации как минимум в два раза превышала верхнюю граничную частоту в спектре сигнала: $\omega_s > 2\omega_m$

Известное применение находят и варианты преобразования Фурье: косинусное для четных и синусное для нечетных сигналов, а также преобразование Хартли, где базисными функциями являются суммы синусов и косинусов, что позволяет повысить производительность вычислений и избавиться от комплексной арифметики. Вместо косинусных и синусных функций используются также меандровые функции Уолша, принимающие значения только +1 и -1. И, наконец, в последнее время в задачах спектрально-временного анализа нестационарных сигналов, изучения нестационарностей и локальных особен-

ностей сигналов "под микроскопом", очистки от шумов и сжатия сигналов начинают получать в качестве базисов разложения вейвлеты ("короткие волны"), локализованные как во временной, так и в частотной области. Понятие о вейвлетах и возможностях обработки сигналов с помощью вейвлет-преобразований рассматривается далее в данной главе.

Традиционные методы анализа данных предназначены, как правило, для линейных и стационарных сигналов и систем, и только в последние десятилетия начали активно развиваться методы анализа нелинейных, но стационарных и детерминированных систем, и линейных, но нестационарных данных.

Между тем, большинство естественных материальных процессов, реальных физических систем и соответствующих этим процессам и системам данных в той или иной мере являются нелинейными и нестационарными, и при анализе данных используются определенные упрощения, особенно в отношении априорно устанавливаемого базиса разложения данных.

4.4 Основы вейвлет-преобразования сигналов

Вейвлетное преобразование сигналов является обобщением спектрального анализа, типичный представитель которого – классическое преобразование Фурье. Термин "вейвлет" (wavelet) в переводе с английского означает "маленькая (короткая) волна". Вейвлеты - это обобщенное название семейств математических функций определенной формы, которые локальны во времени и по частоте, и в которых все функции получаются из одной базовой (порождающей) посредством ее сдвигов и растяжений по оси времени. Вейвлет-преобразования рассматривают анализируемые временные функции в терминах колебаний, локализованных по времени и частоте. Как правило, вейвлет-преобразования (WT) подразделяют на дискретное (DWT) и непрерывное (CWT).

DWT используется для преобразований и кодирования сигналов, CWT - для анализа сигналов. Вейвлет-преобразования в настоящее время принимаются на вооружение для огромного числа разнообразных применений, нередко заменяя обычное преобразование Фурье. Это наблюдается во многих областях, включая молекулярную динамику, квантовую механику, астрофизику, геофизику, оптику, компьютерную графику и обработку изображений, анализ ДНК, исследования белков, исследования климата, общую обработку сигналов и распознавание речи.

Вейвлетный анализ представляет собой особый тип линейного

преобразования сигналов и физических данных. Базис собственных функций, по которому проводится вейвлетное разложение сигналов, обладает многими специфическими свойствами и возможностями. Вейвлетные функции базиса позволяют сконцентрировать внимание на тех или иных локальных особенностях анализируемых процессов, которые не могут быть выявлены с помощью традиционных преобразований Фурье и Лапласа. К таким процессам в геофизике относятся поля различных физических параметров природных сред. В первую очередь это касается полей температуры, давления, профилей сейсмических трасс и других физических величин.

Вейвлеты имеют вид коротких волновых пакетов с нулевым средним значением, локализованных по оси аргументов (независимых переменных), инвариантных к сдвигу и линейных к операции масштабирования (сжатия/растяжения). По локализации во временном и частотном представлении вейвлеты занимают промежуточное положение между гармоническими функциями, локализованными по частоте, и функцией Дирака, локализованной во времени.

Теория вейвлетов не является фундаментальной физической теорией, но она дает удобный и эффективный инструмент для решения многих практических задач. Основная область применения вейвлетных преобразований – анализ и обработка сигналов и функций, нестационарных во времени или неоднородных в пространстве, когда результаты анализа должны содержать не только частотную характеристику сигнала (распределение энергии сигнала по частотным составляющим), но и сведения о локальных координатах, на которых проявляют себя те или иные группы частотных составляющих или на которых происходят быстрые изменения частотных составляющих сигнала. По сравнению с разложением сигналов на ряды Фурье вейвлеты способны с гораздо более высокой точностью представлять локальные особенности сигналов, вплоть до разрывов 1-го рода (скачков). В отличие от преобразований Фурье, вейвлет-преобразование одномерных сигналов обеспечивает двумерную развертку, при этом частота и координата рассматриваются как независимые переменные, что дает возможность анализа сигналов сразу в двух пространствах.

Одна из главных и особенно плодотворных идей вейвлетного представления сигналов на различных уровнях декомпозиции (разложения) заключается в разделении функций приближения к сигналу на две группы: аппроксимирующую - грубую, с достаточно медленной временной динамикой изменений, и детализирующую - с локальной и быстрой динамикой изменений на фоне плавной динамики, с последующим их дроблением и детализацией на других уровнях декомпозиции сигналов.

Это возможно как во временной, так и в частотной областях представления сигналов вейвлетами.

4.1.1. Истоки Вейвлет - преобразования

История спектрального анализа восходит к И. Бернулли, Эйлеру и Фурье, который впервые построил теорию разложения функций в тригонометрические ряды. Однако это разложение долгое время применялось как математический прием и не связывалось с какими-либо физическими понятиями. Спектральные представления применялись и развивались сравнительно узким кругом физиков–теоретиков. Однако, начиная с 20-х годов прошлого века, в связи с бурным развитием радиотехники и акустики, спектральные разложения приобрели физический смысл и практическое применение. Основным средством анализа реальных физических процессов стал гармонический анализ, а математической основой анализа - преобразование Фурье. Преобразование Фурье разлагает произвольный процесс на элементарные гармонические колебания с различными частотами, а все необходимые свойства и формулы выражаются с помощью одной базисной функции $\exp(j\omega t)$ или двух действительных функций $\sin(\omega t)$ и $\cos(\omega t)$. Гармонические колебания имеют широкое распространение в природе, и поэтому смысл преобразования Фурье интуитивно понятен независимо от математической аналитики.

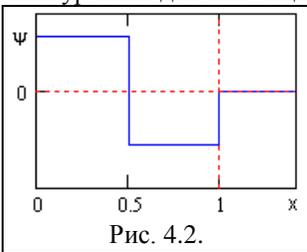
Преобразование Фурье обладает рядом замечательных свойств. Областью определения преобразования является пространство L^2 интегрируемых с квадратом функций, и многие физические процессы в природе можно считать функциями, принадлежащими этому пространству. Для применения преобразования разработаны эффективные вычислительные процедуры типа быстрого преобразования Фурье (БПФ). Эти процедуры входят в состав всех пакетов прикладных математических программ и реализованы аппаратно в процессорах обработки сигналов.

Было также установлено, что функции можно разложить не только по синусам и косинусам, но и по другим ортогональным базисным системам, например, полиномам Лежандра и Чебышева, функциям Лагерра и Эрмита. Однако практическое применение они получили только в последние десятилетия XX века благодаря развитию вычислительной техники и методов синтеза цифровых линейных систем обработки данных. Непосредственно для целей спектрального анализа подобные ортогональные функции не нашли широкого применения из-за трудностей интерпретации получаемых результатов. По тем же причинам не получили развития в спектральном анализе функции ти-

па "прямоугольной волны" Уолша, Радемахера, и пр.

Теоретические исследования базисных систем привели к созданию теории обобщенного спектрального анализа, которая позволила оценить пределы практического применения спектрального анализа Фурье, создала методы и критерии синтеза ортогональных базисных систем. Иллюстрацией этому является активно развивающаяся с начала 80-х годов прошлого столетия теория базисных функций типа вейвлет. Благодаря прозрачности физической интерпретации результатов анализа, сходной с "частотным" подходом в преобразовании Фурье, ортогональный базис вейвлетов стал популярным и эффективным средством анализа сигналов и изображений в акустике, сейсмике, медицине и других областях науки и техники.

Вейвлет-анализ является разновидностью спектрального анализа, в котором роль простых колебаний играют функции особого рода, называемые вейвлетами. Базисная функция вейвлет – это некоторое "короткое" колебание, но не только. Понятие частоты спектрального анализа здесь заменено масштабом, и, чтобы перекрыть "короткими волнами" всю временную ось, введен сдвиг функций во времени. Базис вейвлетов – это функции типа $\psi((t-b)/a)$, где b - сдвиг, a – масштаб. Функция $\psi(t)$ должна иметь нулевую площадь и, еще лучше, равными нулю первый, второй и прочие моменты. Фурье-преобразование таких функций равно нулю при $\omega=0$ и имеет вид полосового фильтра. При различных значениях масштабного параметра 'a' это будет набор полосовых фильтров. Семейства вейвлетов во временной или частотной области используются для представления сигналов и функций в виде суперпозиций вейвлетов на разных масштабных уровнях декомпозиции (разложения) сигналов.



Первое упоминание о подобных функциях (которые вейвлетами не назывались) появилось в работах Хаара (Haar) еще в начале прошлого века. Вейвлет Хаара - это короткое прямоугольное колебание на интервале $[0,1]$, показанное на 1.1.1. Однако он интересен больше теоретически, так как не является непрерывно дифференцируемой функцией и имеет

длинные "хвосты" в частотной области. В 30-е годы физик Paul Levy, исследуя броуновское движение, обнаружил, что базис Хаара лучше, чем базис Фурье, подходит для изучения деталей броуновского движения.

Сам термин "вейвлет", как понятие, ввели в своей статье J. Morlet и A. Grossman, опубликованной в 1984 г. Они занимались ис-

следованиями сейсмических сигналов с помощью базиса, который и назвали вейвлетом. Весомый вклад в теорию вейвлетов внесли Гуппилауд, Гроссман и Морлет, сформулировавшие основы CWT, Ингрид Добеши, разработавшая ортогональные вейвлеты (1988), Натали Делпрат, создавшая время-частотную интерпретацию CWT (1991), и многие другие. Математическая формализация в работах Mallat и Meyer привела к созданию теоретических основ вейвлет-анализа, названного мультиразрешающим (кратномасштабным) анализом.

В настоящее время специальные пакеты расширений по вейвлетам присутствуют в основных системах компьютерной математики (Matlab, Mathematica, Mathcad, и др.), а вейвлет-преобразования и вейвлетный анализ используются во многих областях науки и техники для самых различных задач. Многие исследователи называют вейвлет-анализ "математическим микроскопом" для точного изучения внутреннего состава и структур неоднородных сигналов и функций.

Не следует рассматривать вейвлет-методы обработки и анализа сигналов в качестве новой универсальной технологии решения любых задач. Возможности вейвлетов еще не раскрыты полностью, однако это не означает, что их развитие приведет к полной замене традиционных средств обработки и анализа информации, хорошо отработанных и проверенных временем. Вейвлеты позволяют расширить инструментальную базу информационных технологий обработки данных.

Оконное преобразование Фурье.

У анализа сигналов с использованием преобразования Фурье есть серьезный недостаток. В преобразовании к области частоты теряется информация времени. Чтобы исправить этот недостаток, Деннис Гэбор (1946) предложил анализировать только маленькую секцию сигнала. Такая техника, называемая Фурье анализ в окне или windowing анализ. Адаптация Гэбора, названная Short-Time Fourier Analysis представляет сигнал в двумерном виде, отображая функцию времени и частоты см. Рис. 4.3



Рис. 4.2 Фурье анализ в окне .

Оконное преобразование Фурье(ОПФ) с движущейся по сиг-

налу оконной функции, имеющей компактный носитель. Временной интервал сигнала разделяется на подинтервалы и преобразование выполняется последовательно для каждого подинтервала в отдельности. Тем самым осуществляется переход к частотно-временному (частотно-координатному) представлению сигналов, при этом в пределах каждого подинтервала сигнал "считается" стационарным. Результатом оконного преобразования является семейство спектров, которым отображается изменение спектра сигнала по интервалам сдвига окна преобразования. Это позволяет выделять на координатной оси и анализировать особенности нестационарных сигналов. Размер носителя оконной функции $w(t)$ обычно устанавливается соизмеримым с интервалом стационарности сигнала. По существу, таким преобразованием один нелокализованный базис разбивается на определенное количество базисов, локализованных в пределах функции $w(t)$, что позволяет представлять результат преобразования в виде функции двух переменных - частоты и временного положения окна. Оконное преобразование выполняется в соответствии с выражением:

$$S(\omega, b_k) = \int_{-\infty}^{\infty} s(t)w^*(t - b_k)dt \quad (4.16)$$

Функция $w^*(t-b)$ представляет собой функцию окна сдвига преобразования по координате t , где параметром b задаются фиксированные значения сдвига. При сдвиге окон с равномерным шагом значения b_k принимаются равными $k\Delta b$. В качестве окна преобразования может использоваться как простейшее прямоугольное окно, так и специальные весовые окна (Бартлетта, Гаусса, и пр.), обеспечивающие малые искажения спектра при вырезке оконных отрезков сигналов (нейтрализация явления Гиббса).



Рис. 4.4 Пример оконного преобразования для нестационарного сигнала

По спектру сигнала можно судить о наличии в его составе гармонических колебаний на трех частотах, определять соотношение между амплитудами этих колебаний и конкретизировать локальность колебаний по интервалу сигнала. Координатная разрешающая способность оконного преобразования определяется шириной оконной функции и обратно пропорциональна частотной разрешающей способности. При ширине оконной функции, равной b , частотная разрешающая способность определяется значением $\Delta\omega = 2\pi/b$. При требуемой величине частотного разрешения $\Delta\omega$ соответственно ширина оконной функции должна быть равна $b = 2\pi/\Delta\omega$. Для оконного преобразования Фурье эти ограничения являются принципиальными. Так, для рисунке 4.3 при размере массива данных $N = 300$ и ширине оконной функции $\Delta b = 100$ частотная разрешающая способность результатов преобразования уменьшается в $N/\Delta b = 3$ раза по сравнению с исходными данными, и графики $S_w(n\Delta\omega_{sw})$ по координате n для наглядного сопоставления с графиком $S(n\Delta\omega_s)$ построены с шагом по частоте $\Delta\omega_{sw} = 3\Delta\omega_s$, т.е. по точкам $n = 0, 3, 6, \dots, N$.

Частотно-временное оконное преобразование

ЧПрименяется для анализа нестационарных сигналов, если их частотный состав изменяется во времени. Функция оконного преобразования (4.17) может быть переведена в вариант с независимыми переменными и по времени, и по частоте:

$$S(t, \omega_k) = \int_{-\infty}^{\infty} w(\tau) \exp(-j\omega\tau) d\tau \quad (4.17)$$

Координатная разрешающая способность оконного преобразования определяется шириной оконной функции и, в силу принципа неопределенности Гейзенберга, обратно пропорциональна частотной разрешающей способности. Хорошая разрешающая способность по времени подразумевает небольшое окно времени, которому соответствует плохая частотная разрешающая способность и наоборот. Оптимальным считается ОПФ с гауссовым окном, которое получило название преобразование Габора (Gabor). Пример преобразования приведен на Рис. 4.5. На Рис. 4.5 пример вычисления и представления (модуль правой части главного диапазона спектра) результатов спектрограммы при дискретном задании зашумленного входного сигнала $sq(n)$. Сигнал представляет собой сумму трех последовательных радиоимпульсов с разными частотами без пауз, с отношением сигнал/шум, близким к 1. Оконная функция w_i задана в одностороннем варианте с эффективной шириной окна $b \cong 34$ и полным размером $M = 50$.

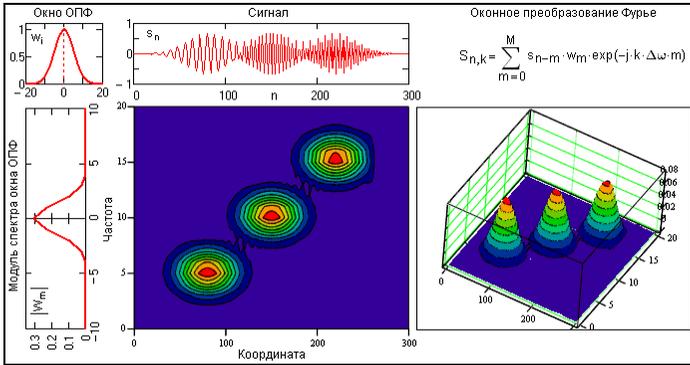


Рис.. 4.5. Пример вычисления ОПФ дискретного сигнала с шумом

Установленный для результатов шаг по частоте $\Delta\omega = 0.1$ несколько выше фактической разрешающей способности $2\pi/M = 0.126$. Для обеспечения работы оконной функции по всему интервалу сигнала задавались начальные и конечные условия вычислений (продление на M точек обоих концов сигнала нулевыми значениями).

Вейвлет-анализ

Вейвлет-анализ –это следующий шаг : windowing техники с областями переменного размера.

Вейвлет-анализ позволяет использовать большие временные интервалы времени, где мы хотим получить более точную низкочастотную информацию, и более короткие области, где мы хотим высокочастотную информацию.

На Рис 4.6 и частотной областях (правый рис.). Представление в частной области выполняется с помощью обычного преобразования Фурье (не оконного).



Рис. 4.6 Схемы представления сигнала во временной (левый рис.) и частотной областях (правый рис.)

На Рис. 4.7 представлены схемы представления сигнала с помощью ОПФ (левый рис.) и с помощью вейвлет-преобразования (правый рис.). Понятие о шкалах (scale) при вейвлет-преобразованиях будет описано дальше



Рис. 4.7 Схемы представления сигнала после оконного преобразования Фурье (левый рис.) и после вейвлет-преобразования (правый рис.)

Вейвлет-преобразование позволяет выявлять неоднородности в анализируемых сигналах и оценивать временное положение этих неоднородностей. На Рис. 4.8 показан синусоидальный сигнал с неоднородностью. Показанные на рисунке 4.8 коэффициенты преобразования Фурье и Вейвлет коэффициенты позволяют оценить преимущества вейвлет-преобразования для оценки неоднородности.



Рис. 4.8а Синусоидальный сигнал с маленькой неоднородностью

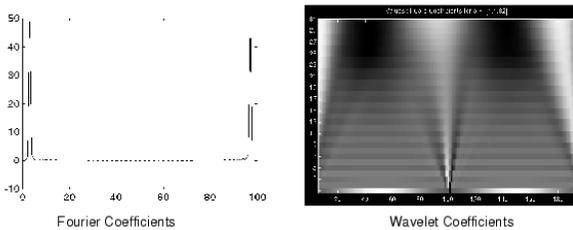


Рис. 4.8б Фурье коэффициенты и Вейвлет коэффициенты для сигнала на Рис. 4.8а

Синусы и вейвлеты Для пояснения преобразований с помощью вейвлет функций рассмотрим простые примеры. На рисунке 4.9 а представлен синусоидальный сигнал определенной частоты. Такие сигналы являются базисом преобразования Фурье, с помощью которого проявляется частотная информация о преобразуемом сигнале.

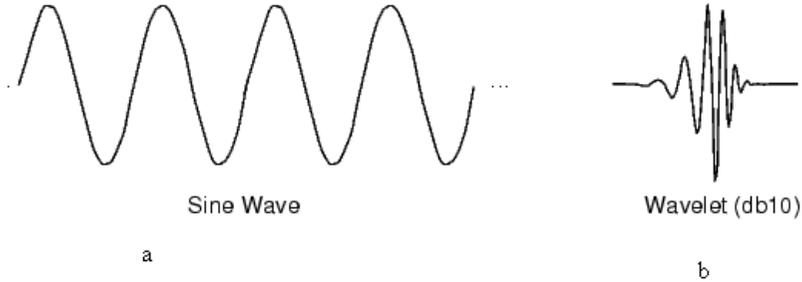


Рис.. 4.9 Синусоидальный сигнал определенной частоты(а) и материнский вейвлет (b)

Как известно, *у синусоид нет ограниченной продолжительности – они простираются от минус до плюс бесконечность.*

На рисунке 4.9 b представлен один из наиболее известных материнский в, который является базисом для вейвлет преобразований. **Вейвлеты имеют тенденцию быть нерегулярными и асимметричными.**

Анализ Fourier разбивает сигнал на *синусы различных частот.* Вейвлет анализ - разбивает сигнал *в смещенные версии материнского вейвлета.*

Местные особенности могут *быть описаны лучше вейвлетами.*

Непрерывное вейвлет-преобразование.

В результате преобразования Фурье непрерывной функции $f(t)$ выполняемого по формуле (4.18)

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \exp(-j\omega t) dt \quad (4.18)$$

получается «набор» синусоид разных частот и разных амплитуд как показано на рисунке 4.10. Функция $F(\omega)$, которая получается в

результате преобразования Фурье по формуле (4.18) является функцией одной переменной ω .

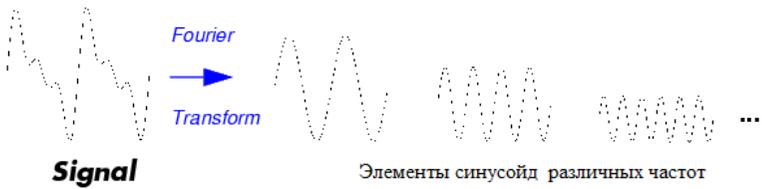


Рис. 4.10 Элементы синусоид различных частот и амплитуд в результате преобразования Фурье непрерывной функции

Рассмотрим теперь непрерывное вейвлет преобразование, которое выполняется по формуле (4.19).

$$C(\text{scale}, \text{position}) = \int_{-\infty}^{\infty} f(t)\psi(\text{scale}, \text{position}, t)dt \quad (4.19)$$

Непрерывное вейвлет преобразование, выполняемое по формуле (4.19) является функцией двух переменных: шкалы (*scale*) и положения (*position*) при этом форма самой материнской вейвлет функции не меняется. Такая материнская вейвлет функция смещается и сжимается, что показано на Рис. 4.11



Рис. 4.11 Элементы материнской вейвлет функций различных шкал и положений в результате непрерывного вейвлет преобразования

Для пояснения понятия шкалы (*scale*) рассмотрим синусоиды кратных частот, которые представлены на Рис. 4.12. Чем выше частота тем коэффициент шкалы a меньше.

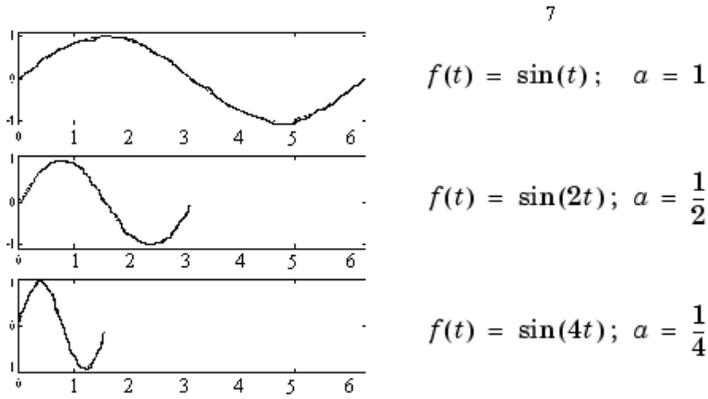


Рис. 4.12 Синусоиды с различными шкалами

Теперь рассмотрим материнские вейвлет функции с различными шкалами, которые представлены на рисунке 4.13. Для этих функции также как и для синусоид, чем выше частота тем коэффициент шкалы a меньше.

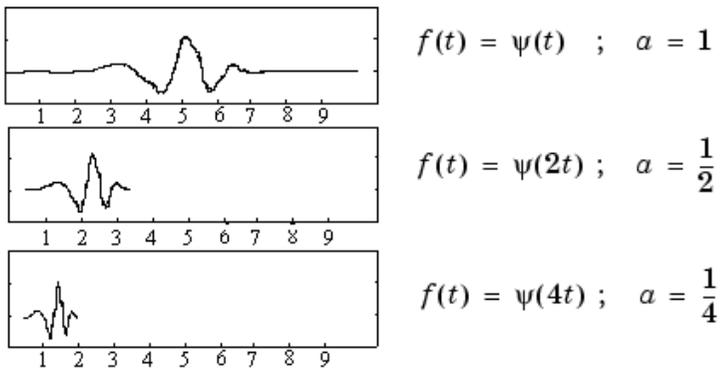


Рис. 4.13 Материнские вейвлет функции с различными шкалами

Поясним теперь понятие сдвига (*position*) для выполнения непрерывного вейвлет преобразования. На рисунке 4.14 представлена материнская вейвлет функция при различных значениях коэффициента сдвига k . На левом рисунке коэффициент сдвига равен нулю, а

на правом материнская вейвлет функция сдвинута на заданную величину.

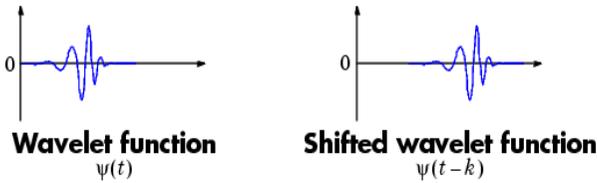


Рис. 4.14 Материнские вейвлет функции с различными коэффициентами сдвига

Пять шагов для выполнения непрерывного вейвлет-преобразования Рассмотрим основные шаги выполнения алгоритма непрерывного вейвлет преобразования.

Шаг 1: Возьмем материнскую вейвлет функцию сравним ее с началом сигнала, как показано на рисунке 4.15

Шаг 2: Вычислим вейвлет- коэффициент, C , который представляет корреляцию вейвлет функции с этой секцией сигнала. Чем выше C , тем больше подобие. C может интерпретироваться как коэффициент корреляции.

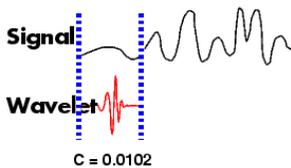


Рис. 4.15. Первый и второй шаги непрерывного вейвлет преобразования.

Шаг 3: Переместите вейвлет вправо и повторим шаги 1 и 2, пока не дойдем до конца обрабатываемого сигнала. (Рис. 4.16)

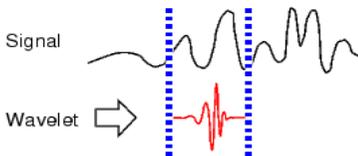


Рис. 4.16. Третий шаг непрерывного вейвлет преобразования
Шаг 4 : Растянем шкалу и повторим шаги 1...3. (Рис. 4.17)

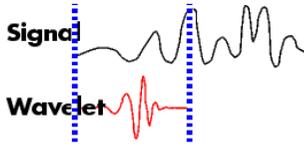


Рис. 4.17. Четвертый шаг- растяжка шкалы для непрерывного вейвлет преобразования

Шаг 5: Повторяем шаги 1 - 4 для всех шкал.

Таким образом выполнение вычисления коэффициентов непрерывного вейвлет преобразования сводится к вычислению корреляции между анализируемым сигналом и сдвинутым и растянутым копиям материнской вейвлет функции.

Представление вейвлет преобразования Для графического представления результатов вейвлет преобразований существуют две формы: 2D представление вейвлет-преобразования (рис. 4.18) и 3D представление вейвлет-преобразования (рис. 4.19).

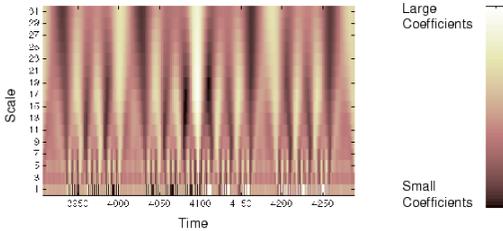


Рис. 4.18 2D представление вейвлет-преобразования

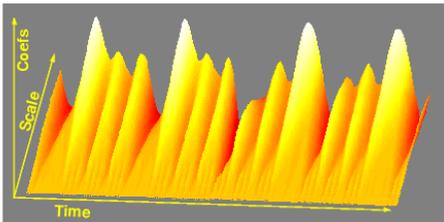


Рис. 4.19 3D представление вейвлет-преобразования

Шкала и частота_ На Рис. 4.20 поясняются возможности использования низкой шкалы (low scale) и высокой шкалы (high scale) для выявления различных особенностей анализируемого сигнала.

Левый рис. -низкая шкала $a \Rightarrow$ сжатый вейвлет \Rightarrow быстрые изменения (мелкие детали) сигнала \Rightarrow высокая частота ω

Правый рис. -высокая шкала $a \Rightarrow$ растянутый вейвлет \Rightarrow медленные изменения (грубые детали) сигнала \Rightarrow низкая частота ω



Рис. 4.20 Низкая и высокая шкалы для выявления особенностей анализируемого сигнала

Что непрерывного в непрерывном вейвлет преобразованиях

В отличие от дискретного вейвлет преобразования, которое будет рассмотрено дальше, непрерывное вейвлет преобразование (CWT) может работать в каждой шкале.

Мы выбираем между нашей потребностью в подробном анализе и доступной вычислительной мощностью (Рис. 4.21)

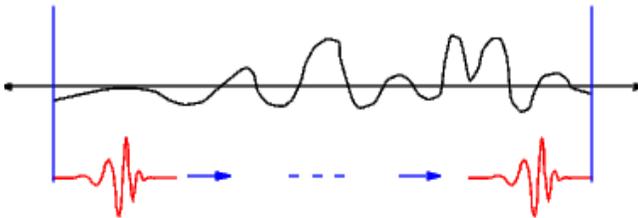


Рис. 4.21 Непрерывное вейвлет преобразование для выбранной шкалы на временном интервале полной области анализируемого сигнала

Непрерывное вейвлет преобразование непрерывно с точки зрения сдвига. Во время вычисления вейвлет функция перемещается гладко по полной области анализируемой функции.

Дискретное вейвлет-преобразование

Для многих сигналов низкочастотное содержание - самая важная часть. Это - то, что дает сигналу его идентичность. Высокочастотное содержание, с другой стороны, передает аромат или нюанс. Рассмотрите человеческий голос. Если Вы удаляете высокочастотные компоненты, голос кажется различным, но Вы можете все еще сказать то, что говорится. Однако, если Вы удаляете достаточно многие из низкочастотных компонентов вы слышите тарабарщину. В вейвлет анализе мы часто говорим об аппроксимации и деталях. Аппроксимация - высокий масштаб, низкочастотные компоненты сигнала. Детали - низкий масштаб, высокочастотные компоненты. На Рис. 4.22 представлено разделение сигнала на высокочастотную и низкочастотную части с помощью низкочастотных (lowpass) и высокочастотных (highpass) вейвлет-фильтров. С помощью таких фильтров мы разделили наш сигнал на аппроксимирующую (A) и детализирующую части (D).

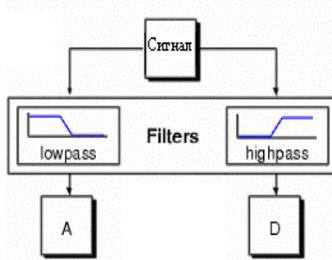


Рис. 4.22 Разделение сигнала на аппроксимирующую(A) и детализирующую(D) части с помощью дискретного вейвлет-преобразования

Представим дискретный сигнал состоящий из $N=1000$ точек (1000 data points). С помощью вейвлет фильтров мы получили 500 точек (вейвлет коэффициентов Cd), описывающих высокочастотную составляющую сигнала и 500 точек (вейвлет коэффициентов CA) описывающих низкочастотную составляющую сигнала. Такие составляющие сигнала представлены на Рис. 4.23а. Дискретное вейвлет преобразование обозначается DWT. Стрелка острием вниз обозначает уменьшение размерности сигнала. В данном случае с 1000 до 500 точек. Возможно дальнейшее разделение сигнала на аппроксимирующую и детализирующую части, то есть многоуровневое преобразование. В нашем случае показано одноуровневое преобразование, при котором мы разделили сигнал на две части.

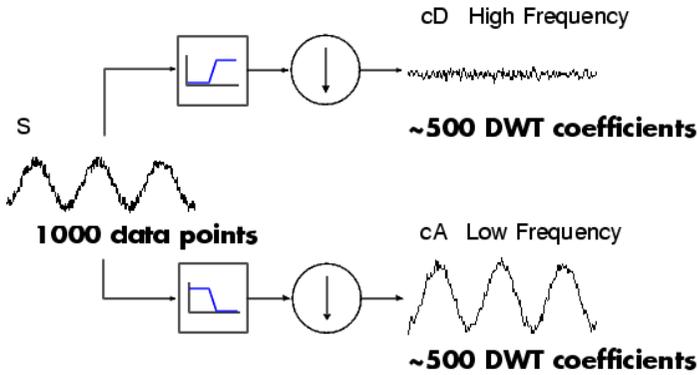


Рис. 4.23а Одноуровневое дискретное вейвлет- преобразование
 Разделим аппроксимирующую часть сигнала еще на две части, затем еще на две части и т.д. В результате получается многоуровневое вейвлет –преобразование, схема которого показана на Рис. 4.24б

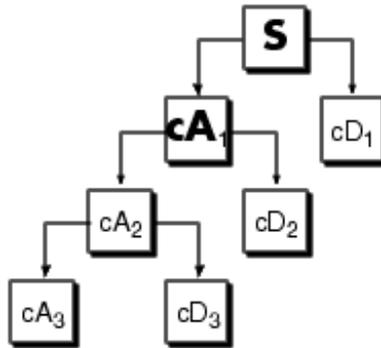


Рис. 4.23б Многоуровневое дискретное вейвлет- преобразование

Поясним процесс многоуровневого разделение сигнала на аппроксимирующую и детализирующую части, то есть многоуровневое на примере фильтрации шумов в зашумленном масс-спектрометрическом сигнале.

На Рис. 4.24 представлен исходный сигнал, содержащий 6000 точек.

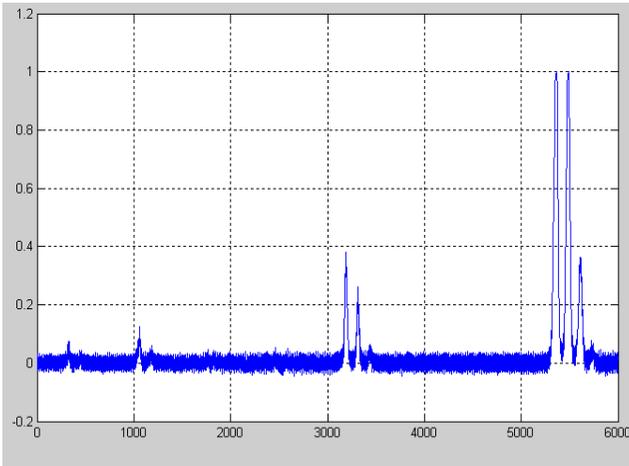


Рис. 4.24 Исходный масс-спектрометрический сигнал

При выполнении многоуровневого DWT количество точек N уменьшается и отношение полезного сигнала к шуму увеличивается за счет низкочастотной фильтрации. На рисунках 4.25, 4.26, 4.27 представлен сигнал, обработанный с помощью DWT разных уровней.

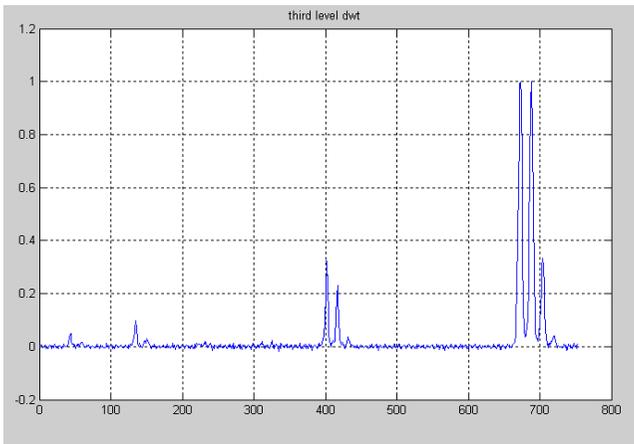


Рис. 4.25. Третий уровень ДВП сигнала ($N/8$ точек, повысилось отношение сигнал/шум)

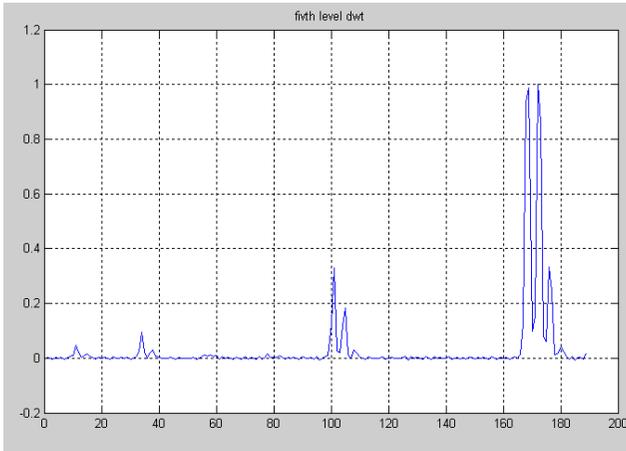


Рис. 4.26. Пятый уровень ДВП сигнала в нормализованном виде ($N/32$ точек, повысилось отношение сигнал/шум, ухудшилось разрешение)

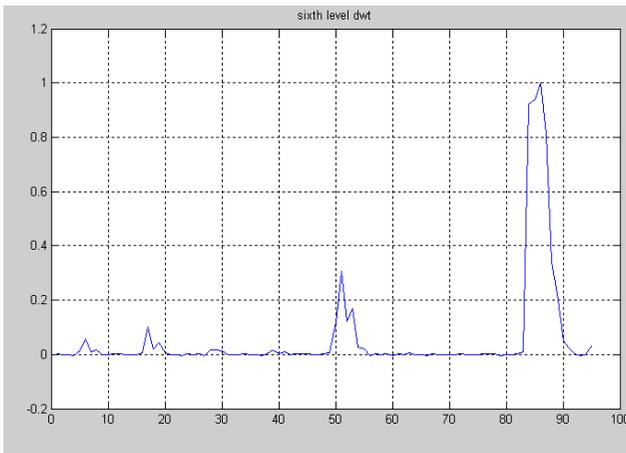


Рис. 4.26. Шестой уровень ДВП сигнала в нормализованном виде ($N/64$ точек, повысилось отношение сигнал/шум, сильно ухудшилось разрешение)

Для определения количества уровней декомпозиции DWT для повышения отношения сигнал/шум можно вычисляют значения энтропии для каждого уровня, которые представлены в 4.1.

Таблица 4.1. Значения значений типов энтропии при декомпозиции масс-спектра.

Уровень декомпозиции	Вейвлет функция db1		
	Энтропия	S/N	R
Нулевой –(исходный сигнал)	73	35	2
Первый	36	60	2
Второй	17	100	2,2
Третий	7.2	150	2
Четвертый	3.6	230	2
Пятый	1.6	300	1.6
Шестой	1.5	350	*

Таблица составлена для масс-спектра, изображенного на рисунке 4.24.

В таблице: db1 – вейвлет - функции Добеши, * на 10 % уровне пики не разрешены, S/N – отношение сигнала к шуму. R – разрешение.

Величины отношения сигнал шум (S/N) и разрешение (R) подсчитаны после проведения эксперимента и приведены в таблице для количественной оценки информации, представленных на приведенных выше графиках.

Наиболее приемлемыми уровнями вейвлет декомпозиции являются уровни 4 и 5, т.к. они дают максимальное отношение сигнала к шуму и не снижают разрешающую способность.

Этот вывод можно сделать из характера уменьшения энтропии, представленного в таблице 4.1 и на рисунке. 4.27.

Исходя из данных, представленных в таблице и на рисунке видно, что после 5 уровня энтропия падает значительно медленнее, чем в промежутке от 1 до 5 уровней.

Предложенный подход к фильтрации не требует априорных знаний о полосе фильтруемого сигнала и не требует выполнения синтеза передаточной функции фильтра по заданной полосе среза.

Кроме того, данный метод одновременно с фильтрацией производит сжатие сигнала.

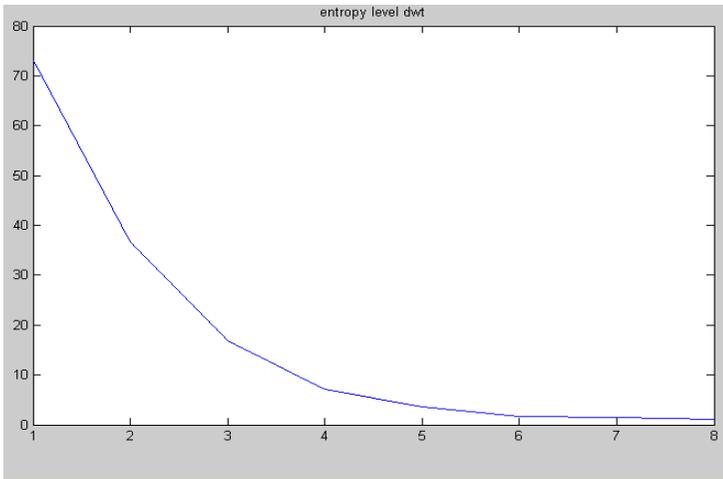


Рис. 4.27 Уменьшение энтропии по Шеннону в зависимости от уровня вейвлет-декомпозиции.

Вейвлет-реконструкция. Комплекс математических операций, который выполняет процесс синтеза называется обратным дискретным вейвлет-преобразованием. Этот синтез выполняется на основе реконструкционных фильтров. Правильный выбор реконструкционного фильтра является решающим моментом в достижении наименьшей разности между исходным и синтезируемым сигналами. Низкочастотный и высокочастотные декомпозиционные фильтры, а также связанные с ними реконструкционные фильтры (образуют систему, которая называется квадратурными зеркальными фильтрами).

Кроме обнаружения локальных особенностей сигналов фильтрации, сжатия информации вейвлет-преобразования в задачах обработки информации аналитических приборов еще целесообразно применять для сравнения анализируемых сигналов с «эталонными» в решении задач идентификации.

5 Повышение разрешающей способности аналитических приборов алгоритмическими средствами

Для разработки оптимальных методов, позволяющих наиболее точно и быстро оценить параметры реальных спектрометрических сигналов, особенно следует выделить такую характерную особенность спектрометрических сигналов, как наложение друг на друга спектральных пиков, возникающее из-за недостаточной разрешающей способности аналитической части спектрометра. Для понимания таких

методов необходима информация об операции свертки, которая приведена ниже.

5.1 Свертка аппаратной функции и сигнала спектрометров

Предположим, что выполнены два условия:

1. Линейность регистрирующей системы, т.е. реакция системы на сумму сигналов равна сумме реакций на каждый из них в отдельности. Для каждой системы существует диапазон входных сигналов, в котором это условие выполняется. Исследователь должен определить этот диапазон и не использовать сигналы вне этого диапазона для каких-либо количественных оценок сигналов.

2. Инвариантность прибора, т. е. аппаратные искажения не зависят от абсолютных значений переменной по горизонтальной оси спектральной линии. В спектроскопии достаточно, чтобы это условие выполнялось в пределах исследуемой спектральной линии, что практически всегда имеет место.

Тогда сигнал на выходе системы есть свертка сигнала на входе и некоторой характеризующей свойства прибора функции $g(x)$, называемой *аппаратной функцией* или *импульсным откликом прибора*:

$$f_{out} = \int_{-\infty}^{\infty} f_{in}(x') * g(x - x') dx \quad (5.1)$$

В общем случае аппаратные искажения можно исключить путем решения интегрального уравнения (5.1).

Об этом пойдет речь в специальном параграфе данного раздела.

Здесь же отметим, что аппаратная функция - основная теоретическая характеристика прибора, определяемая физическим принципом его функционирования и параметрами, к тому же она содержит более полную информацию о приборе, чем разрешающая способность.

Свертка

Свертка – это операция, которая производится измерительными приборами, например, оптическими и в результате которой получается размытое (неясное) изображение изучаемого объекта. Изображение точки в любом аналитическом приборе никогда не бывает точкой, а представляет собой пятно.

Размеры этого пятна определяются качеством прибора.

В оптическом приборе изображения двух различных точек будут разделены только при определенном условии.

Этим условием является то, что расстояние между точками превышает некоторую минимальную величину, определяющую возможность разрешения [55, 61].

На выходе спектрометра вместо абсолютно острого «пика» всегда наблюдается некоторая кривая, степень размытия которой определяется качеством спектрометра. Поэтому трудно, а иногда и невозможно разделить два соседних острых пика. Аналогичная ситуация наблюдается и в электронике, где поступающий на вход амплитудного анализатора импульс бесконечно малой продолжительности на выходе анализатора приобретает форму сигнала конечной ненулевой продолжительности, то есть продолжительность выходного сигнала определяется шириной полосы пропускания прибора (рис. 5.1).

Сигнал на выходе, соответствующий импульсу бесконечно малой продолжительности на входе, называется импульсным откликом [57]. Поэтому любой входной сигнал изменяет свою форму на выходе. Зная импульсный отклик $h(t)$ системы, предполагаемой линейной (система, параметры которой не зависят от времени – в этом случае применима теорема сложения сигналов), можно по входному сигналу $e(t)$ рассчитать выходной сигнал $S(t)$.

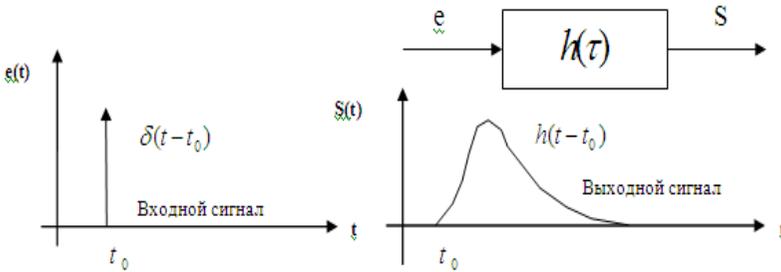


Рис. 5.1 Изменение входного сигнала на выходе

Решение этой задачи осуществляется с помощью свертки [6]. Импульсному отклику $h(t)$ соответствует входной сигнал $\delta(t - t_0)$, то сигнал на выходе имеет форму $h(t-t_0)$ с тем же смещением t_0 .

5.1.2 Уравнение свертки

Рассмотрим произвольный входной сигнал $e(t)$. Мы можем всегда разложить сигнал $e(t)$ в последовательность импульсов продолжи-

Так как система линейна, можно применить теорему сложения, согласно которой сигнал на выходе равен сумме всех откликов

$$s(t) = \sum_{k=0}^l e(k\Delta t)h_1(t - k\Delta t)\Delta t \quad (5.3)$$

где $l = [t/\Delta t]$ - целая часть числа $t/\Delta t$.

Необходимо отметить, что в силу принципа причинности отклик физической системы $h_1(t)$ равен нулю для $t < 0$ и поэтому $h_1(t - k\Delta t) = 0$ для $t < k\Delta t$, следовательно, если справедливо равенство (5.3), то справедливы и равенства (5.4а) и (5.4б) :

$$s(t) = \sum_{k=0}^{\infty} e(k\Delta t)h_1(t - k\Delta t)\Delta t \quad (5.4 \text{ а, б})$$

$$s(t) = \sum_{k=-\infty}^{\infty} e(k\Delta t)h_1(t - k\Delta t)\Delta t$$

Последовательность $e(k\Delta t)$ стремится к функции $e(t)$, а $h_1(t)$ стремится к импульсному отклику системы $h(t)$ при $\Delta t \rightarrow 0$.

Интегральные суммы правых частей равенств (5.3) и (5.4 а, б) стремятся к соответствующим интегралам при $\Delta t \rightarrow 0$.

Поэтому, переходя в равенствах (5.3) и (5.4 а, б) к пределу при $\Delta t \rightarrow 0$, получаем формулы:

$$s(t) = \int_0^t e(\tau)h(t - \tau)d\tau, \quad (5.5)$$

$$s(t) = \int_0^{\infty} e(\tau)h(t - \tau)d\tau, \quad (5.6 \text{ а})$$

$$s(t) = \int_{-\infty}^{\infty} e(\tau)h(t - \tau)d\tau, \quad (5.6 \text{ б})$$

Последние три эквивалентных выражения обычно записывают символически в виде свертки:

$$s(t) = e(t) * h(t) \quad (5.7)$$

Свертка имеет следующие свойства:

1. Дистрибутивность: $a(t) * [b(t) + c(t)] = a(t) * b(t) + a(t) * c(t)$.
2. Коммутативность: $a(t) * b(t) * c(t) = a(t) * c(t) * b(t) = c(t) * b(t) * a(t)$.

3. Ассоциативность: $[a(t)*b(t)]*c(t)=a(t)*[b(t)*c(t)]=a(t)*b(t)*c(t)$.

5.1.2 Замечания относительно свертки

На Рис. 5.3а дано неправильное представление операции свертки, так как значение выходного сигнала в момент t зависит не только от значений $h(t)$ и $e(t)$ в момент t , но и от значений этих функций для всех предыдущих моментов времени.

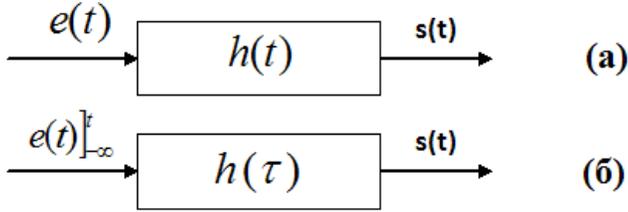


Рис.. 5.3. Представление операции свертки

Корректное представление операции свертки дано на рисунке 5.3 б. В действительности импульсный отклик $h(\tau)$ отличен от нуля только на ограниченном промежутке (Θ_1, Θ_2) .

Поэтому уравнение свертки можно записать в виде :

$$s(t) = \int_{t-\Theta_2}^{t-\Theta_1} e(\tau)h(t-\tau)d\tau \quad (5.8)$$

Поскольку для физических фильтров импульсный отклик $h(\tau)$ равен нулю при $\tau < 0$, то уравнение свертки принимает форму:

$$s(t) = \int_{t-\Theta}^t e(\tau)h(t-\tau)d\tau \quad (5.9)$$

5.2 Методы восстановления спектров

Для оценки параметров сложных пиков (состоящих из двух или более спектральных линий) в спектрах аналитических приборов различных типов широко применяется математическая обработка экспериментальных данных, основанная на методах решения некорректных задач [58,60].

Информация о таких задачах приведена в следующем подразделе.

5.3 Корректно и некорректно поставленные задачи

Корректные и некорректные задачи это классы математических задач, которые различаются степенью определённости их решений.

Многие математические задачи состоят в том, что по исходным данным u ищется решение z . При этом считается, что u и z связаны функциональной зависимостью $z = R(u)$ [58].

Задача определения решения z из пространства F по "исходным данным" u из пространства U называется корректно поставленной задачей на паре метрических пространств $\{F, U\}$ если выполнены следующие условия (условия корректности):

1) Существование решения - задача имеет решение при любых допустимых исходных данных, то есть, среди исходных данных нет противоречащих друг другу условий, что исключало бы возможность решения задачи (для всякого элемента u из пространства U существует решение z из пространства F);

2) Однозначность задачи - каждым исходным данным u соответствует только одно решение, то есть исходных данных достаточно для однозначной определённости решения задачи;

3) Решение устойчиво на пространствах (F, U) . Это условие заключается в следующем. Если u_1 и u_2 - два различных набора исходных данных, мера уклонения которых друг от друга достаточно мала, то мера уклонения решений $z_1 = R(u_1)$ и $z_2 = R(u_2)$ меньше любой наперёд заданной меры точности. При этом предполагается, что в многообразии $U = \{u\}$ допустимых исходных данных и в многообразии возможных решений $Z = \{z\}$ установлено понятие меры уклонения (или меры близости) $r(u_1, u_2)$ и $r^*(z_1, z_2)$. Это объясняется тем, что исходные данные физической задачи, как правило, задаются с некоторой погрешностью; при нарушении же устойчивости решения как угодно малые возмущения исходных данных могут вызывать большие отклонения в решении. Задачи, не удовлетворяющие хотя бы одному условию корректности, называются некорректными задачами (или некорректно поставленными).

Широко распространёнными методами решения некорректных задач является метод регуляризации по Тихонову и алгебраический метод. Задача восстановления сигналов, искажённых прибором с конечной разрешающей способностью, при наличии шумов измерений в

простейшем случае сводится к решению линейного интегрального уравнения Фредгольма 1 - го рода:

$$\int_a^b K(x, s)z_r(s)ds = u_r(x) \quad , \quad \begin{matrix} a \leq x \leq b \\ a \leq s \leq b \end{matrix} \quad (5.10)$$

В этом уравнении $z_r(s)$ - искомая функция (сигнал) из пространства F , $u_r(x)$ - заданная функция из пространства U , то есть результат преобразования (искажения) сигнала измерительным прибором, характеризуемым аппаратной функцией $K(x, s)$ - ядром интегрального уравнения, причем будем полагать, что ядро $K(x, s)$ по переменной x является непрерывной функцией с непрерывной частной производной $\frac{\partial K}{\partial x}$ [65].

Оператор $\int_a^b K(x, s)z(s)ds$ для краткости обозначим через Az .

Целью задачи восстановления спектра является более точное определение функции $z_r(s)$ по результатам измерения случайной функции $u(x)$, однако, результатом измерения $u(x)$ в реальности является сумма $u_r(x)$ со случайным шумом $v(t)$, статистические характеристики которого известны:

$$u(x) = u_r(x) + v(x)$$

Подобного рода задачи, часто называемые так же обратными, встречаются практически в каждой области экспериментальной физики.

Если измерительная аппаратура линейна, то функциональная связь между $z_r(s)$ и $u(x)$ дается формулой:

$$Az = \int_a^b K(x, s)z_r(s)ds = u(x)$$

где $K(x, s_0)$ - аппаратная функция, предполагаемая известной, которая представляет экспериментальный спектр (как функция x).

Основная трудность решения некорректных задач заключается в том, что правая часть уравнения $u(x)$ является, как правило, результа-

том измерения и известна лишь приближенно. Поэтому решение $z(s)$ ищутся в классе непрерывных на отрезке $[a, b]$ функций. Уклонения правых частей друг от друга оцениваются в квадратичной метрике (метрике L_2), то есть по формуле:

$$\rho_u(u_1, u_2) = \left[\int_c^d [u_1 - u_2]^2 \right]^{1/2}$$

А уклонения решений $z(s)$ оцениваются в равномерной метрике (метрике C), то есть по формуле:

$$\rho_F(z_1, z_2) = \max_{s \in [a, b]} |z_1(s) - z_2(s)|$$

Пусть для некоторой правой части $u = u_1(x)$ функция $z_1(s)$ является решением уравнения (5.10), то есть

$$\int_a^b K(x, s) z_r(s) ds = u(x)$$

Если вместо функции $u_1(x)$ нам известно лишь ее приближение $u(x)$, мало отличающееся в метрике L_2 от $u_1(x)$, то речь может идти лишь о нахождении приближенного к $z_1(x)$ "решения" уравнения (5.10). При этом правая часть $u(x)$ может быть получена в эксперименте и иметь "угловые" точки или выбросы, в которых функция $u(x)$ не имеет производной. При такой правой части $u(x)$ уравнение (5.10) не имеет решения, понимаемого в классическом смысле, то есть определяемого по формуле:

$$z = A^{-1}u$$

где A^{-1} - оператор, обратный оператору A в уравнении (1), так как ядро $K(x, s)$ имеет непрерывную производную по x и, следовательно, правая часть также должна иметь непрерывную производную по x . Значит, в качестве приближенного к $z_1(s)$ «решения» уравнения (5.10) нельзя брать точное решение этого уравнения с приближенно известной правой частью $u(x) \neq u_1(x)$, так как такого решения может и не существовать.

Очевидно, уравнение (5.10) имеет решение, понимаемое в классическом смысле, только для таких правых частей $u(x)$, которые принадлежат образу AF множества F функций $z(s)$ при отображении:

$$u = Az \equiv \int_a^b K(x, s)z(s)ds, \quad z(s) \in F$$

Задача нахождения приближенного решения некорректно поставленной задачи вида:

$$Az = u \quad (5.11)$$

в естественном классе элементов F является практически недоопределенной, то есть эта задача является некорректной, например, в случаях, когда A – вполне непрерывный оператор [65]. Тогда обратный ему оператор A^{-1} не будет непрерывным на U и решение уравнения (5.11) не будет устойчивым к малым изменениям правой части u (в метрике пространства U). Предположим оператор A нам известен точно, а правая часть уравнения (5.11) известна с точностью δ , то есть вместо ее точного значения u_t нам известны элемент \tilde{u} и число δ такие, что

$$\rho_u(u_t, \tilde{u}) \leq \delta$$

По этим данным, то есть по (\tilde{u}, δ) , требуется найти такой элемент $z_\delta \in F$, который стремился бы (в метрике F) к $z_t \rightarrow 0$ при $\delta \rightarrow 0$. Такой элемент будет приближенным (к z_t) решением уравнения $Az = \tilde{u}$. Элементы $z \in F$, удовлетворяющие условию $\rho_u(u_t, \tilde{u}) \leq \delta$, будут сопоставимыми по точности с исходными данными (\tilde{u}, δ) . Пусть Q_δ – совокупность всех таких элементов $z \in F$. Тогда приближенные решения уравнения $Az = \tilde{u}$ будут в классе Q_δ элементов z , сопоставимых по точности с исходными данными \tilde{u}, δ . Однако, в ряде случаев этот класс элементов достаточно широк, что приводит к значительным отличиям этих элементов между собой.

Поэтому не все из них будут приближенными решениями уравнения (5.11), а значит, возможные решения должны отбираться, основываясь на использовании априорной информации.

5.4 Понятие регуляризирующего оператора

Пусть z_t есть решение уравнения $Az = u_t$, то есть,

$$Az_t = u_t \quad (5.12)$$

тогда вместо u_t будет некоторый элемент u_δ и известное число $\delta > 0$ такие, что

$$\rho(u_\delta, u_t) \leq \delta.$$

Это означает, что вместо точных исходных данных (u_t, A) мы имеем приближенные исходные данные (u_δ, A) и оценку их погрешности δ .

Задача состоит в том, чтобы по известным исходным данным (u_δ, A, δ) найти приближение z_δ к элементу z_t , обладающее свойством устойчивости к малым изменениям u_δ .

Очевидно, что в качестве приближенного решения z_δ уравнения (5.12) нельзя брать точное решение этого уравнения с приближенной правой частью $u = u_\delta$.

Это означает, что элемент z_δ определяемый по формуле:

$$z_\delta = A^{-1}u_\delta$$

существует не для всякого элемента $u \in U$ и не обладает свойством устойчивости к малым изменениям правой части u .

Числовой параметр δ характеризует погрешность правой части уравнения (5.11).

Поэтому представляется естественным определить такое решение z_δ с помощью оператора, зависящего от параметра. Значения такого оператора надо брать согласованными с погрешностью δ исходных данных u_δ .

Эта согласованность должна быть такой, чтобы при $\delta \rightarrow 0$, то есть при приближении (в метрике пространства U) правой части u_δ уравнения (5.11) к точному значению u_t , приближенное решение z_δ стремилось бы (в метриках пространства F) к искомому точному решению z_t уравнения $Az = u_t$.

Пусть элементы $z_t \in F$ и $u_t \in U$ связаны соотношением

$$Az_t = u_t.$$

Оператор $R(u, \delta)$, действующий из пространства U в пространство F , называется *регуляризирующим* для уравнения $Az = u$ (относительно элемента u_t), если он обладает свойствами:

1) Существует такое число $\delta_1 > 0$, что оператор $R(u, \delta)$ определен для всякого $0 \leq \delta \leq \delta_1$, и любого $u_\delta \in U$ такого что

$$\rho_u(u_\delta, u_t) \leq \delta;$$

2) Для всякого $\varepsilon > 0$ существует такое $\delta_0 = \delta_0(\varepsilon, u_\delta) \leq \delta_1$, что из неравенства: $\rho_u(u_\delta, u_t) \leq \delta \leq \delta_0$ следует неравенство:

$$\rho_F(z_\delta, z_t) \leq \varepsilon,$$

где $z_\delta = R(u_\delta, \delta)$. Здесь не предполагается однозначность оператора $R(u, \delta)$. Через z_0 обозначается произвольный элемент из множества $\{R(u_\delta, \delta)\}$ значений оператора $R(u_\delta, \delta)$.

Теперь рассмотрим более общий случай. Оператор $R(u, \alpha)$, зависящий от параметра α и действующий из U в F , называется *регуляризирующим* для уравнения $Az = u$ (относительно элемента u_t), если он обладает свойствами:

1) Существует такие числа $\delta_1 > 0, \alpha_1 > 0$, что оператор $R(u, \alpha)$ определен для всякого α , принадлежащего промежутку $(0, \alpha)$, и любого $u \in U$, для которого $\rho_u(u, u_t) \leq \delta_1$.

2) Существует такой функционал $\alpha = F(u, \delta)$, определенный на множестве $U_{\delta_1} = \{u : \rho(u, u_t) \leq \delta_1\}$ элементов $u \in U$ так, что для любого $\varepsilon > 0$ найдется число $\delta(\varepsilon) \leq \delta_1$ такое, что, если $\tilde{u} \in U$ и $\rho_u(\tilde{u}, u_t) \leq \delta \leq \delta(\varepsilon)$, то

$$\rho_F(z_t, z_\alpha) \leq \varepsilon, \text{ где } z_\alpha = R(\tilde{u}, \alpha(\tilde{u}, \delta)).$$

В этом определении не предполагается однозначность оператора $R(\tilde{u}, \alpha(\tilde{u}, \delta))$. При $\alpha = \delta$ получаем определение 1.

Если $\rho_u(u_t, u_\delta) \leq \delta$, то в качестве приближенного решения уравнения (5.11) с приближенно известной правой частью u_δ можно брать элемент $z_\alpha = R(u_\delta, \alpha)$, полученный с помощью регуляризирующего оператора $R(u, \alpha)$, где $\alpha = \alpha(u_\delta) = \alpha_1(\delta)$ согласовано с погрешностью исходных данных u_δ .

Это решение называется регуляризованным решение уравнения (5.11). Числовой параметр α называется параметром регуляризации.

Всякий регуляризирующий оператор вместе с выбором параметра регуляризации α , согласованного с погрешностью исходных данных u_δ , $\alpha = \alpha(u_\delta)$, определяет устойчивый к малым изменениям правой части u метод построения приближенных решений уравнения (5.11).

Таким образом, задача нахождения приближенного решения уравнения (5.12), устойчивого к малым изменениям правой части, сводится:

- 1) К нахождению регуляризирующих операторов;
- 2) К определению параметра регуляризации a по дополнительной информации о задаче, например, по величине погрешности, с которой задается правая часть u_δ . Описанный метод построения приближенных решений называется методом регуляризации.

5.5 Метод регуляризации по Тихонову

Говоря о некорректных задачах, нельзя не отметить, что для их решения советским математиком Тихоновым был предложен чрезвычайно эффективный метод, называемый регуляризацией и основанный на привлечении дополнительной априорной информации о решении, которая может быть как качественной, так и количественной [58].

Рассмотрим уравнения типа свертки:

$$z(t) * K(t) = u(t) \quad (5.13),$$

в котором $K(t)$ и $u(t)$ - заданные функции, а $z(t)$ - искомая; $u \in U$, $z \in F$, F и U - метрические пространства.

Будем полагать, что при $u(t) = u_t(t)$ это уравнение имеет единственное решение $z_t(t)$, принадлежащее F , то есть

$$z_t(t) * K(t) = u_t(t)$$

Задача заключается в нахождении функции $z_t(t)$. Если правая часть известна с погрешностью, то есть вместо $u_t(t)$ имеем функцию $u(t)$ такую, что

$$\rho_u(u_t, u) \geq \delta,$$

Вместо нахождения z_t можно ставить лишь задачу о нахождении приближенного решения. В качестве приближенного решения будем брать регуляризирующее решение:

$$z_\alpha(t) = R(u, \alpha)$$

где $R(u, \alpha)$ - регуляризирующий оператор.

Рассмотрим способ построения широкого класса регуляризирующих операторов (предложенный впервые А.Н.Тихоновым), получаемых с помощью классических интегральных преобразований.

Пусть F и U - некоторые множества функций ($F \subset L_1, U \subset L_2$) и A - линейный непрерывный оператор с областью определения $D_A \subset F$. Рассмотрим уравнение:

$$Az = u, u \in U \quad (5.14)$$

Будем полагать, что оно имеет единственное решение на F . Применяя к соотношению (5.14) линейное интегральное преобразование \mathfrak{Z} , например, преобразование Фурье (Лапласа, Меллина и др.), получим

$$\mathfrak{Z}[Az] = \mathfrak{Z}[u] = u(\omega) \quad (5.15)$$

Пусть оператор A таков, что из соотношения (5.15) можно определить $\mathfrak{F}[z] = z(\omega)$ в виде $z(\omega) = \psi(u(\omega), \omega)$.

Если Az есть свертка $z(t)*K(t)$ функций $z(t)$ с некоторой заданной функцией $K(t)$ (ядром) и при применении преобразования (5.14) к этой свертке справедлива теорема умножения, то есть, то $\mathfrak{F}[z * K] = \mathfrak{F}[z] \cdot \mathfrak{F}[K]$, то

$$\psi(u(\omega), \omega) = \frac{u(\omega)}{K(\omega)} \quad (5.16)$$

Рассмотрим для определенности уравнение вида

$$Az \equiv \int_{-\infty}^{+\infty} K(t - \tau)z(\tau)d\tau = u(t) \quad (5.17)$$

И применим преобразование Фурье. Здесь,

$$u(t) \in L_2(-\infty, +\infty), K(t) \in K \in L_1(-\infty, +\infty)$$

Если правая часть уравнения (5.17) известна приближенно, то есть

$$u(t) = u_r(t) + v(t)$$

где $v(t)$ - помеха (шум), то

$$z(\omega) = \frac{u(\omega)}{K(\omega)} = \frac{u_r(\omega)}{K(\omega)} + \frac{v(\omega)}{K(\omega)}$$

Так как $u_r(\omega) = K(\omega) + z_r(\omega)$, то

$$u_r(\omega) = K(\omega) + z_r(\omega) \text{ и } z(\omega) = z_r(\omega) + \frac{v(\omega)}{K(\omega)}$$

Эта формула дает нам преобразование Фурье точного решения уравнения (5.17) с приближенной правой частью $u(t)$. Казалось бы естественным в качестве приближенного решения уравнения (5.17) с приближенной правой частью $u(t)$ брать функцию, полученную с помощью обратного преобразования Фурье, то есть функцию:

$$z(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} z(\omega) \exp(j\omega t) d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} z_r(\omega) \exp(j\omega t) d\omega + \\ + \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{v(\omega)}{K(\omega)} \exp(j\omega t) d\omega = z_r(t) + \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{v(t)}{K(\omega)} \exp(j\omega t) d\omega$$

Однако такая функция может не существовать, так как, последний интеграл может быть расходящимся.

В самом деле, по свойству преобразования Фурье функции $K(\omega)$ и $v(\omega)$ стремятся к нулю ∞ при $\omega \rightarrow \infty$, но это есть стремление к нулю “несогласованное” поскольку функция $v(t)$ (а, следовательно, и $v(\omega)$) обычно носит случайный характер. Поэтому отношение:

$$\frac{v(\omega)}{K(\omega)}$$

может не иметь обратного преобразования Фурье из-за влияния высоких частот от случайной функции $v(\omega)$.

Но, если даже функция

$$\frac{v(\omega)}{K(\omega)}$$

и имеет обратное преобразование Фурье $\omega(t)$, то уклонение функции $\omega(t)$ от нуля (в метрике C или L_2) может быть сколь угодно большим.

Таким образом, в качестве приближенного решения уравнения (5.12) с приближенной правой частью нельзя брать точное решение этого уравнения.

Такого решения может не существовать, а если оно и существует, то не обладает свойством устойчивости к малым отклонениям правой части $u(t)$.

Причиной неустойчивости такого алгоритма построения “решений” является влияние высоких частот ω преобразования Фурье $v(\omega)$ помехи $v(t)$.

Поэтому, если мы хотим строить приближенные решения уравнения (5.17), устойчивые к малым отклонениям правой части и, с помощью обратного преобразования Фурье, то надо “подавить” влияние

высоких частот ω , умножая, например, функцию $\frac{v(\omega)}{K(\omega)}$ на соответствующий множитель (зависящий от параметра α).

Возвращаясь к уравнению (5.2), рассмотрим оператор вида:

$$R_f(u, \alpha) = \mathfrak{T}^{-1}[\psi(u(\omega), \omega) f(\omega, \alpha)],$$

где \mathfrak{T}^{-1} - преобразование, обратное преобразованию \mathfrak{T} , а $f(\omega, \alpha)$ некоторая заданная функция, определенная для всех неотрицательных значений параметра α и любых ω , по которым берется оператор \mathfrak{T}^{-1} .

Если функцию подчинить соответствующим условиям, то оператор $R_f(u, \alpha)$ будет регуляризирующим для уравнения (5.14). Для определенности рассмотрим уравнение типа свертки вида:

$$Az \equiv \int_{-\infty}^{\infty} K(t - \tau) z(\tau) d\tau = u(t)$$

и в качестве \mathfrak{T} возьмем преобразование Фурье. В этом случае

$$R_f(u, \alpha) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\frac{f(\omega, \alpha)}{K(\omega)} u(\omega) \exp(-j\omega t) \right) d\omega \quad (5.18)$$

Пусть функция $f(\omega, \alpha)$ удовлетворяет следующим условиям:

- 1) $f(\omega, \alpha)$ определена в области $(\alpha \geq 0, -\infty \leq \omega \leq \infty)$
- 2) $0 \leq f(\omega, \alpha) \leq 1$ для всех значений $\alpha \geq 0$ и ω ;
- 3) $f(\omega, 0) \equiv 1$;
- 4) для всякого $\alpha > 0$ $f(\omega, \alpha)$ четная по ω и $f(\omega, \alpha) \in L_2(-\infty, \infty)$;
- 5) для всякого $\alpha > 0$ $f(\omega, \alpha) \rightarrow 0$ при $\omega \rightarrow \pm\infty$;

6) при $\alpha \rightarrow 1$ $f(\omega, \alpha) \rightarrow 1$, не убывая, причем на всяком отрезке $|\omega| \leq \omega_1$ эта сходимость равномерная;

7) для всякого $\alpha > 0$ $\frac{f(\omega, \alpha)}{K(\omega)} \in L_2(-\infty, \infty)$;

8) для всякого $\omega \neq 0$ $f(\omega, \alpha) \rightarrow \infty$ и эта сходимость равномерная на всяком отрезке $[\omega_1, \omega_2]$, $0 < \omega_1 < \omega_2$.

Таким образом, заданием функции $f(\omega, \alpha)$, удовлетворяющей требованиям 1-8, определяется однопараметрический оператор $R_f(u, \alpha)$ вида (5.14).

Если уклонение правой части уравнения (5.15) оценивать в метрике $L_2(-\infty, \infty)$, а уклонение решения $z(t)$ - в метрике C , и полагать, что $z_T(t) \in L_2(-\infty, \infty)$, то справедлива Теорема:

Если функция удовлетворяет условиям 1-8, то определенный с ее помощью оператор $R_f(u, \alpha)$ вида (5.16) является регуляризирующим оператором для уравнения (5.14).

Пусть $M(\omega)$ - заданная четная функция, причем:

- 1) она кусочно-непрерывна на любом конечном отрезке;
- 2) неотрицательна; $M(0) \geq 0$ и $M(\omega) > 0$ при $\omega \neq 0$;
- 3) для достаточно больших $|\omega|$ $M(\omega) \geq c > 0$;
- 4) для всякого $\alpha > 0$

$$\frac{K(-\omega)}{L(\omega) + \alpha M(\omega)} \in L_2(-\infty, +\infty),$$

где

$$L(\omega) = K(\omega)K(-\omega) = |K(\omega)|^2$$

Полагая, что

$$f(\omega, \alpha) = \frac{L(\omega)}{L(\omega) + \alpha M(\omega)}$$

получим классы регуляризирующих операторов для уравнения (5.14).

Каждый такой класс определяется заданием функции $M(\omega)$. При фиксированной функции $M(\omega)$ параметр регуляризации α можно находить по невязке. Если уклонение правой части $u(t)$ оценивать в метрике L_2 , то квадрат невязки регуляризованного решения $z_\alpha(t)$ будет вычисляться по формуле:

$$\begin{aligned} \rho^2_{L_2}(Az_\alpha, u) &= \int_{-\infty}^{\infty} [Az_\alpha - u(t)]^2 = \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} |K(\omega)z_\alpha(\omega) - u(\omega)|^2 d\omega = \Phi(\alpha) \end{aligned}$$

Так как

$$\begin{aligned} z_\alpha(\omega) &= \frac{K(-\omega)u(\omega)}{L(\omega) + \alpha M(\omega)}, \text{ то} \\ \Phi(\alpha) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\alpha^2 M^2(\omega) |u(\omega)|^2}{[L(\omega) + \alpha M(\omega)]^2} d\omega. \end{aligned}$$

Очевидно, что $\Phi(0) = 0$,

$$\Phi(\alpha) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} |u(\omega)|^2 d\omega = \|u(t)\|_{L_2}^2$$

и $\Phi(\alpha)$ стремится к $\|u(t)\|_{L_2}^2$ при $\alpha \rightarrow \infty$.

Кроме того,

$$\Phi'_\alpha(\alpha) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{\alpha L(\omega) M^2(\omega) |u(\omega)|^2 d\omega}{[L(\omega) + \alpha M(\omega)]^3} > 0$$

Таким образом, невязка регуляризованного решения – строго возрастающая функция переменного α , изменяющаяся от 0 до $\|u(t)\|_{L_2}^2$.

Следовательно, если погрешность правой части u_δ уравнения (5.12) равна δ и $\delta < \|u(t)\|^2$, то существует единственное число $\bar{\alpha}$, для которого $\Phi(\bar{\alpha}) = \delta^2$. Если $\delta \geq \|u_\delta(t)\|_{L_2}$, то уравнение $\Phi(\alpha) = \delta^2$ не имеет решения. Очевидно, что $\Phi(\alpha)$ зависит от выбора функции $M(\omega)$, то есть:

$$\Phi(\alpha) = \Phi_M(\alpha).$$

При малых значениях α основной вклад в $\Phi_M(\alpha)$ дают большие частоты. Поэтому если $M_1(\omega) \geq M_2(\omega)$ для достаточно больших ω и для достаточно малых α имеем:

$$\Phi_{M_1}(\alpha) \geq \Phi_{M_2}(\alpha).$$

В частности, если брать $M_1(\omega) = \omega^{2p_1}$ и $M_2(\omega) = \omega^{2p_2}$ то при $p_1 > p_2$:

$$\Phi_{p_1}(\alpha) > \Phi_{p_2}(\alpha).$$

Решение $\bar{\alpha}$ уравнения $\Phi_p(\alpha) = \delta^2$ очевидно, зависит от p , $\bar{\alpha} = \bar{\alpha}(p)$ и при $p_1 > p_2$ - $\bar{\alpha}(p_1) < \bar{\alpha}(p_2)$. Таким образом, с увеличением порядка регуляризации p , $\bar{\alpha}(p)$ убывает.

5.6 Метод регуляризации основанный на Singular Value Decomposition (SVD)

Кроме метода Тихонова для решения задачи по оценке параметров «наложившихся» пиков может быть применен метод регуляризации основанный на Singular Value Decomposition (SVD) - разложении. Перед его представлением опишем некоторые особенности решения систем линейных алгебраических уравнений.(СЛАУ).

Переопределенные СЛАУ

Для «хороших» систем линейных алгебраических уравнений (СЛАУ), то есть описываемых квадратной хорошо обусловленной матрицей A , решение существует, единственно и может быть с легкостью найдено численно, например методом Гаусса (это метод после-

довательного исключения переменных, когда с помощью элементарных преобразований система уравнений приводится к равносильной системе ступенчатого (или треугольного) вида, из которого последовательно, начиная с последних (по номеру) переменных, находятся все остальные переменные) [66]. Однако, при обработке эксперимента чаще встречаются задачи, для которых матрица A либо не является квадратной, либо плохо обусловлена. Рассмотрим переопределенные системы, в которых число уравнений больше числа неизвестных.

При наличии шума, СЛАУ с прямоугольной матрицей размера $M \times N$ (при $M > N$, то есть при числе уравнений большем числа неизвестных) вовсе не имеет решения, то есть является несовместимой или переопределенной. Конечно, в крайне редких случаях (например, при нулевом шуме) система с прямоугольной матрицей может оказаться совместимой (если выбран соответствующий вектор b).

Подход к общей проблеме переопределенных СЛАУ заключается в том, чтобы вместо точного решения системы уравнений организовать поиск такого вектора u , который будет наилучшим образом удовлетворять всем уравнениям, то есть минимизировать их невязку (расхождение между вектором $A \cdot x$ и вектором b правой части СЛАУ). Чтобы получить более точную и надежную меру близости к вырожденности, нам потребуется ввести понятие нормы вектора. Норма – это число, которое измеряет общий уровень элементов вектора. Наиболее употребительной векторной нормой является евклидова длина:

$$\left(\sum_{i=1}^n |z_i|^2 \right)^{\frac{1}{2}}$$

Поскольку невязка $A \cdot x - b$ является векторной величиной, то минимизации надо подвергать ее норму.

Таким образом, при интерпретации переопределенных СЛАУ принято искать не точное решение, которого нет, а псевдорешение – вектор, минимизирующий норму невязки системы уравнений. В подобном случае, задача решения линейной системы уравнений заменяется задачей отыскания глобального минимума функции $f(x) = |Ax - b|$, где x – это неизвестный вектор, а символ модуля означает операцию вычисления нормы (евклидовой длины вектора). Поскольку эта минимизируемая норма зависит от суммы квадратов компонент неизвестного вектора, то процедура поиска псевдорешения

является ничем иным, как реализацией метода наименьших квадратов (МНК).

Метод наименьших квадратов — один из методов регрессионного анализа (статистический метод исследования зависимости между зависимой переменной Y и одной или несколькими независимыми переменными X_1, X_2, \dots, X_p) для оценки неизвестных величин по результатам измерений, содержащим случайные ошибки.

Метод заключается в минимизации евклидова расстояния $|A \cdot x - b|$ между двумя векторами: вектором восстановленных значений зависимой переменной и вектором фактических значений зависимой переменной.

Подход к решению переопределенных систем может быть и несколько иным, если имеется дополнительная априорная информация (некоторая оценка неизвестного вектора x).

Недоопределенные СЛАУ.

Альтернативным методом переопределенных СЛАУ с прямоугольной матрицей $M \times N$ (при $M > N$) является случай, когда количество уравнений меньше количества неизвестных [67]. Такие системы либо имеют бесконечное число решений, либо не имеют решений вовсе. Для того чтобы получить разумное единственное решение задачи, необходимо «доопределить» ее, добавив некоторые априорные соображения о значении неизвестного вектора x . Если априорной информации о примерной величине вектора x нет, единственным образом решить СЛАУ невозможно. Однако если о нем имеется априорная информация, она позволит дополнить систему уравнений и получить решение. Проще всего ввести в задачу определенные ожидания о величине вектора x .

Математически можно полагать ожидаемое значение вектора x нулевым, поскольку переход от любого x к нуль-вектору осуществляется линейным преобразованием переменных, которое изменит только вектор правой части b . Таким образом, решением недоопределенной СЛАУ будет то решение, которое ближе всего находится к нулевому вектору, то есть обладает минимальной нормой:

$$\|x\| \approx \min.$$

Это решение называется нормальным псевдорешением СЛАУ и искать его следует, минимизируя норму вектора x на последовательном полученном семействе решений СЛАУ, то есть решение недоопреде-

ленной СЛАУ сводится к условной минимизации функции $f(x) = |x|$, при условии выполнения самой СЛАУ.

Метод регуляризации, основанный на SVD - разложении для оценки амплитуд «наложившихся» пиков

Метод регуляризации, основанный на SVD разложении в последующих подразделах данной главы будем называть алгебраическим. Сингулярное разложение (Singular Value Decomposition, SVD) - декомпозиция вещественной матрицы с целью ее приведения к каноническому виду [65]. Сингулярное разложение является удобным методом при работе с матрицами. Оно показывает геометрическую структуру матрицы и позволяет наглядно представить имеющиеся данные. Сингулярное разложение используется при решении самых разных задач - от приближения МНК и решения систем уравнений до сжатия изображений. При этом используются разные свойства сингулярного разложения, например, способность показывать ранг матрицы, приближать матрицы данного ранга (ранг матрицы равен наибольшему числу линейно независимых строк (или столбцов) матрицы). SVD позволяет вычислять обратные и псевдообратные матрицы большого размера, что делает его полезным инструментом при решении задач регрессионного анализа. Регрессионный анализ — метод моделирования измеряемых данных и исследования их свойств. Данные состоят из пар значений зависимой переменной (переменной отклика) и независимой переменной (объясняющей переменной). Регрессионная модель есть функция независимой переменной и параметров с добавленной случайной переменной. Параметры модели настраиваются таким способом, что модель наилучшим образом приближает данные [65]. Критерием качества приближения (целевой функцией) обычно является среднеквадратичная ошибка: сумма квадратов разности значений модели и зависимой переменной для всех значений независимой переменной в качестве аргумента. Предполагается, что зависимая переменная есть сумма значений некоторой модели и случайной величины. Относительно характера распределения этой величины делают предположения, называемые гипотезой порождения данных. Для подтверждения или опровержения этой гипотезы выполняются статистические тесты, называемые анализом остатков. При этом предполагается, что независимая переменная не содержит ошибок.

Рассмотрим систему уравнений вида:

$$Az = u \quad (5.18)$$

где A - матрица размера $m \times n$, а z и u - вектора. Будем полагать, что A и u ($u \neq 0$) заданы, а z подлежит определению. Условие решения неоднородного уравнения является, как известно, условие

$$\det(A) \neq 0 \quad (5.19)$$

то есть, матрица A должна быть невырожденной. В этом случае решение уравнения (1) будет определяться по формуле:

$$z = A^{-1}u$$

где A^{-1} оператор, обратный оператору A в уравнении (5.18).

Коэффициенты матрицы и правой части системы линейных уравнений (5.18) редко бывают известны точно. Некоторые системы возникают из эксперимента и тогда коэффициенты подвержены ошибками наблюдения.

Коэффициенты других систем записываются формулами, что влечет ошибки округлений при их вычислении. Даже если систему можно точно записать в память машины в ходе ее решения почти неизбежно будут сделаны ошибки округлений. Вследствие этого мы подходим к фундаментальному вопросу: если в коэффициентах системы линейных уравнений делаются ошибки, то как сильно при этом меняется решение. Или, другими словами, если $Az = u$, то как можно измерить чувствительность z по отношению к изменениям в A и u . Если A - вырожденная матрица, то для некоторых u решение не существует, тогда как для других u оно будет неединственным. Таким образом, если A почти вырождена, то можно ожидать, что малые изменения в A и u вызовут очень большие изменения в z . С другой стороны, если A - единичная матрица, то и u и z - один и тот же вектор. Следовательно, если A близка к единичной матрице, то малые изменения в A и u должны влечь за собой соответственно малые изменения в z . С помощью нормы вектора можно получить более точную и надежную меру близости к вырожденности.

Умножение вектора z на матрицу A приводит к новому вектору Az , норма которого может очень отличаться от нормы вектора z . Область возможных изменений может быть задана двумя числами

$$M = \max_z \frac{\|Az\|}{\|z\|} \quad \text{и} \quad m = \min_z \frac{\|Az\|}{\|z\|}$$

где максимумы и минимумы берутся по всем ненулевым векторам. Заметим, что если A вырождена, то $m=0$. Отношение M/m называется числом обусловленности матрицы A :

$$\text{cond}(A) = \frac{\max_z \frac{\|Az\|}{\|z\|}}{\min_z \frac{\|Az\|}{\|z\|}} .$$

Рассмотрим систему уравнений: $Az = u$ и другую систему, полученную изменением правой части:

$$A(z + \Delta z) = u + \Delta u .$$

Будем считать Δu ошибкой в u , а Δz - соответствующей ошибкой в z . Поскольку $A(\Delta z) = \Delta u$, то определения M и m медленно ведут к неравенствам: $\|u\| \leq M\|z\|$ и $\|\Delta u\| \leq m\|\Delta z\|$.

$$\text{Следовательно, при } m \neq 0: \frac{\|\Delta z\|}{\|z\|} \leq \text{cond}(A) \frac{\|\Delta u\|}{\|u\|} .$$

Полученное неравенство показывает, что число обусловленности выполняет роль множителя в увеличении относительной ошибки. Таким образом, изменения в правой части могут повлечь за собой изменения в решении, большие в $\text{cond}(A)$ раз. Оказывается, что то же самое справедливо в отношении изменений в коэффициентах матрицы. Описываемый ниже метод регуляризации решения уравнения (5.12) основан на использовании SVD - разложения матрицы A и анализе ее сингулярных чисел.

Среди матричных разложений особую роль играют ортогональные, обладающие свойством сохранения нормы вектора. Матрица Q называется ортогональной, если она удовлетворяет условию:

$$Q^T \cdot Q = I$$

где I - единичная матрица. Свойство сохранения нормы при ортогональных преобразованиях:

$$\|Q \cdot x\| = \|x\|$$

дает рецепт поиска псевдорешения вырожденных СЛАУ, а именно, замену исходной задачи минимизации невязки с «плохой» матрицей $|Ax - b|$ задачей $|Q^T \cdot (Ax - b)|$, в которой матрица $Q^T \cdot A$ уже будет «хорошей» благодаря специальному построению матрицы Q . Таким образом ортогональные разложения используются

при решении любых систем (даже с прямоугольной матрицей A , причем как переопределенных, так и недоопределенных). Одним из важнейших вариантов ортогональных разложений некоторой матрицы A является QR -разложение вида:

$$A = QR,$$

Где Q -ортогональная матрица, а R -верхняя треугольная матрица.

Для невырожденной СЛАУ $A \cdot x = b$ можно сразу записать:

$$Q^T \cdot Q \cdot R \cdot x = Q^T b,$$

откуда следует, благодаря ортогональности матрицы Q , решение «хорошей» системы:

$$R \cdot x = Q^T \cdot b$$

Так как матрица R – треугольная, то решение данной системы получается по формулам прямого хода.

Для решения СЛАУ с сингулярной квадратной $N \times N$ или прямоугольной $N \times M$ (с рангом k) матрицей A известно, что получающаяся треугольная матрица R имеет следующую структуру:

$$R = \begin{pmatrix} R_1 & R_2 \\ 0 & 0 \end{pmatrix} \quad (5.20),$$

где R_1 - верхняя треугольная матрица, а R_2 - просто некая матрица. Нули в формуле обозначают нулевые матрицы соответствующих размеров.

Если система вырожденная, то она имеет бесконечное множество псевдорешений (векторов, минимизирующих норму невязки). При помощи QR -разложения можно сразу выписать одно из них. Согласно (5.20) одна или несколько последних строк матрицы R содержат одни нули, поэтому одна или несколько компонент вектора псевдорешения x может быть произвольной. В связи с этим, целесообразно разбить вектор x на два вектора меньшего размера:

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Если положить $x_2 = 0$, то остальные компоненты x определяются из треугольной СЛАУ:

$$R_1 \cdot x_1 = Q^T \cdot b.$$

Для того, чтобы выбрать из всего множества псевдорешений, минимизирующих невязку исходной СЛАУ, нормальное псевдорешение (то есть обладающее минимальной нормой), необходимо решить соответствующую задачу минимизации [63]. Если построено QR -разложение, сделать это намного легче. Для этого произвольную компоненту решения x_2 следует определить из условия минимума функции $f(x_2)$:

$$f(x_2) = \left\| \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right\| \approx \min \Leftrightarrow R_1 \cdot x_1 = b_1 - R_2 \cdot x_2 \quad (5.21)$$

В формуле (5.21) вертикальными чертами обозначена операция вычисления нормы вектора и символически записано, что вектор x_1 , от которого также зависит функция $f(x_2)$, есть решение выписанной линейной системы с треугольной матрицей R_1 . Остальные составляющие решения x (то есть вектор x_1) определятся из треугольника СЛАУ:

$$R_1 \cdot x_1 = Q^T \cdot b - R_2 \cdot x_2.$$

Алгоритмы решения СЛАУ на основе QR -разложения практически одинаковы, как для хорошо обусловленных, так и для сингулярных систем [66]. Наиболее эффективным (в то же время очень ресурсоемким) способом решения СЛАУ посредством матричных разложений являются полные ортогональные разложения, имеющие вид:

$$A = U \cdot K \cdot V^T.$$

Здесь матрица A имеет размер $M \times N$ и ранг k , а U и V – это ортогональные матрицы размером $N \times N$ и $M \times M$ соответственно. Еще одна матрица K размера $N \times M$ имеет следующую структуру:

$$K = \begin{pmatrix} W & 0 \\ 0 & 0 \end{pmatrix} \quad (5.21),$$

где W – матрица размера $k \times k$.

Однако самым лучшим современным алгоритмом решения произвольных СЛАУ является SVD-разложение или сингулярное разложение, которое имеет вид:

$$A = U \cdot S \cdot V^T,$$

где S - диагональная матрица, состоящая из нулей и расположенных на диагонали сингулярных чисел матрицы A . Согласно фор-

муле (5.10), из матрицы S можно выделить подматрицу W с ненулевыми диагональными элементами. После того, как сингулярное разложение найдено, дальнейшая последовательность действий по построению нормального псевдорешения выглядит так:

- Нахождение единственного решения вспомогательной СЛАУ:

$$W \cdot y = U^T \cdot b.$$

- Дополнение вектора нулевыми элементами до размера искомого вектора x .
- Вычисление x простым умножением $x = V \cdot y$.

Например, матрица

$$A = \begin{pmatrix} 0.96 & 1.72 \\ 2.28 & 0.96 \end{pmatrix}$$

имеет сингулярное разложение

$$A = USV^T = \begin{pmatrix} 0.6 & 0.8 \\ 0.8 & -0.6 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0.8 & -0.6 \\ 0.6 & 0.8 \end{pmatrix}^T$$

Легко увидеть, что матрицы U и V ортогональны,

$$U^T U = U U^T = I, \text{ также } V^T V = V V^T = I$$

и сумма квадратов значений их столбцов равна единице.

6 Исследование эффективности программ решения задач восстановления масс-спектров методом регуляризации Тихонова и алгебраическим методом

Реальные спектральные приборы не являются идеальными гармоническими анализаторами излучения, то есть распределение энергии по спектру, получаемое с помощью реального спектрального прибора, отличается от того «идеального» или «истинного» распределения. Поэтому, сигнал, искаженный аппаратной функцией называется реальным, а сигнал, который был до свертки – «идеальным» [69].

6.1. Перечень элементов программы и функций, решаемых при реализации метода Тихонова

Моделирование исходных «идеальных» сигналов;

- ❖ Моделирование аппаратной функции;
- ❖ Свертка аппаратной функции с исходным сигналом - получение «реального сигнала»;
- ❖ Фурье преобразование «реального» сигнала;
- ❖ Метод регуляризации Тихонова;
- ❖ Восстановление сигнала с помощью обратного преобразования Фурье;
- ❖ Отображение восстановленного сигнала;
- ❖ Проектирование фильтра для удаления ложных частот;
- ❖ Фильтрация ложных частот;
- ❖ Получение отфильтрованного восстановленного сигнала.

На основе составленного алгоритма для решения некорректных задач, была написана программа на язык программирования Mat Lab [69] для решения задач восстановления масс-спектров методом регуляризации Тихонова. Текст программы приведен в Приложении 1

Исходные данные для проверки программы восстановления масс-спектров, искаженных аппаратной функцией методом Тихонова для идеально сформированных пиков

На Рис. 6.1. изображен сформированный идеальный сигнал с двумя одинаковыми пиками (соотношение амплитуд пиков 1/1).

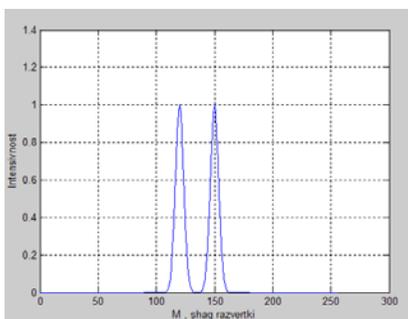


Рис.6.1. Исходный сигнал

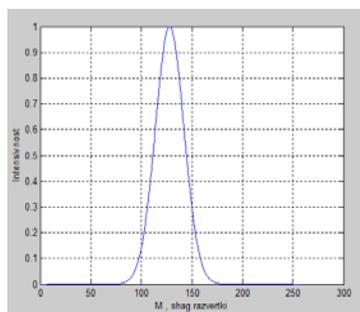


Рис. 6.2. Аппаратная функция

На Рис.6.2. изображен график аппаратной функции, с помощью которой будет произведена конволюция. Конволюция – это свертка данной аппаратной функции с исходным сигналом.

На Рис. 6.3. представлен уширенный сигнал, то есть исходный сигнал, искаженный аппаратной функцией. На Рис.6.4. представлен график восстановленного сигнала после обратного преобразования Фурье

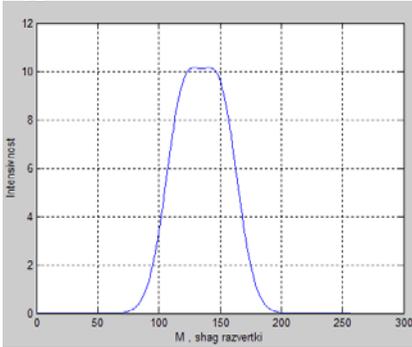


Рис. 6.3. Реальный сигнал

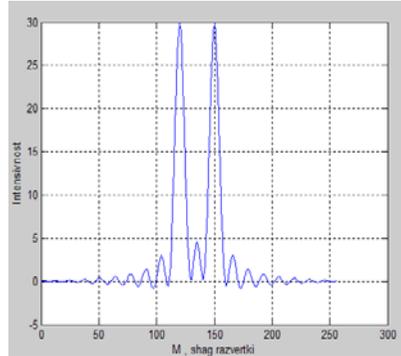


Рис. 6.5. График восстановленного сигнала

В восстановленном сигнале присутствуют ложные частоты, поэтому на Рис. 6.6. представлен восстановленный сигнал, полученный после применения фильтра.

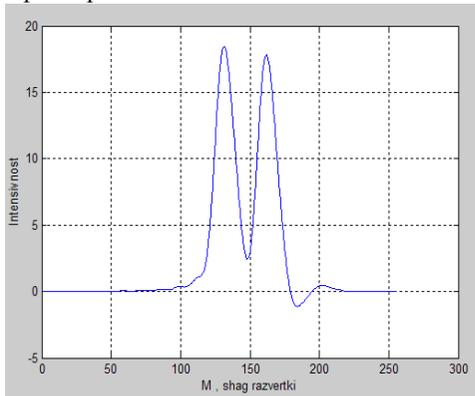


Рис.6.10 Полученный сигнал.

При анализе полученных данных, можно сделать вывод о том, что с помощью метода регуляризации Тихонова пики хорошо разделяются без смещения положения вершин пиков и без изменения их амплитуд. Однако появляются ложные частоты, которые после применения фильтра «ложных частот» сглаживаются. Недостаток фильтра в том, что после его применения происходит небольшое искажение амплитуд сформированных пиков.

6.2 Программы восстановления спектров методом SVD

- ❖ Моделирование исходных идеальных сигналов;
- ❖ Формирование аппаратной матрицы;
- ❖ Свертка аппаратной функции с исходным сигналом, путем произведения с аппаратной матрицы;
- ❖ Создание обратной матрицы
- ❖ Метод SVD;
- ❖ Получение новой преобразованной сингулярной матрицы;
- ❖ Решение СЛАУ с преобразованной матрицей;
- ❖ Применение фильтра Савицкого-Голея;
- ❖ Сравнение аппаратной функции и аппаратной матрицы;
- ❖ Восстановленный и отфильтрованный сигнал.

На основе составленного алгоритма для решения некорректных задач, была написана программа на язык программирования Mat Lab для решения задач восстановления масс-спектров алгебраическим методом. Текст программы приведен в Приложении 1 .

Исходные данные для проверки программы восстановления масс-спектров, искаженных аппаратной функцией алгебраическим методом для идеально сформированных пиков

На Рис.6.11. изображен сформированный идеальный сигнал с двумя пиками.

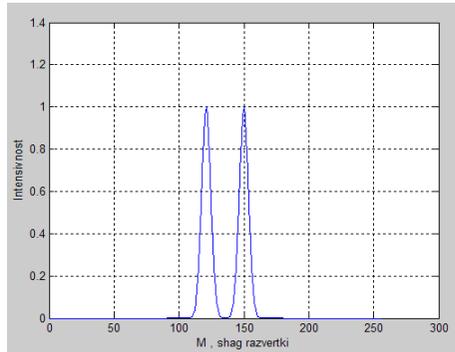


Рис.6.11. Идеальный сигнал

Чтобы получить реальный сигнал (Рис. 6.12.) в алгебраическом методе свертки с аппаратной функцией выполняется не как интеграл свертки, а как произведение аппаратной матрицы на вектор исходного сигнала.

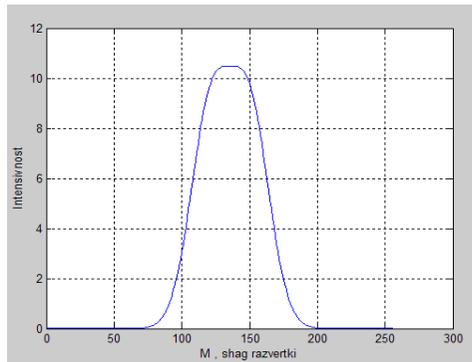


Рис. 6.12. Реальный сигнал

На Рис.6.13. представлен реальный сигнал, полученный с помощью операции *CONV*.

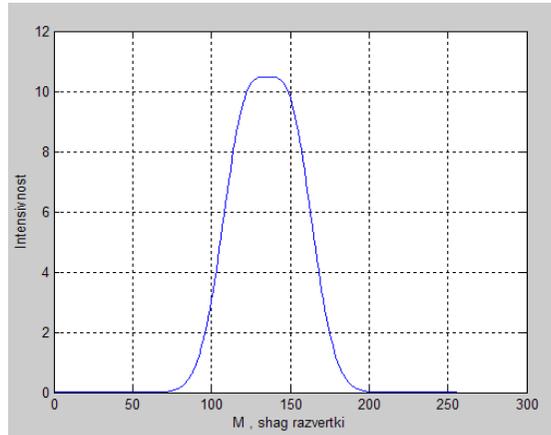


Рис.6.13. Реальный сигнал

Сравнивая Рис 6.12. и Рис 6.13., можно убедиться, что операция CONV и умножение на аппаратную матрицу дают один и тот же результат. На Рис. 6.14. представлен восстановленный сигнал после применения сингулярного разложения.

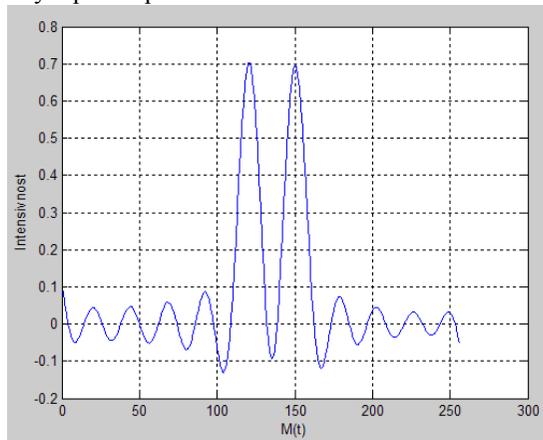


Рис. 6.14. Восстановленный сигнал

В восстановленном сигнале присутствуют ложные частоты, поэтому на Рис. 6.15. изображен восстановленный сигнал после применения фильтра Совицкого-Голея

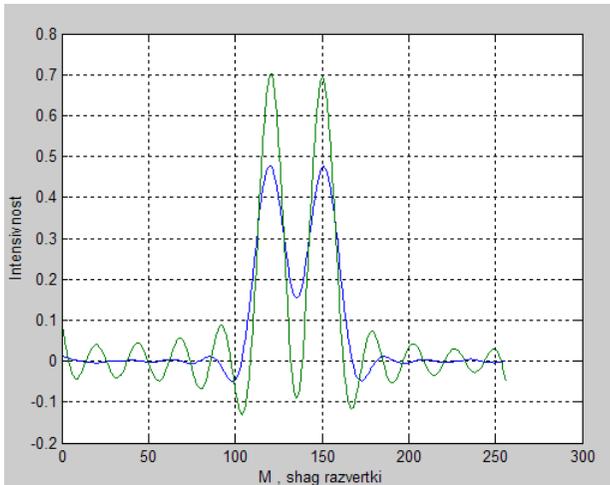


Рис.6.15. Восстановленный сигнал

При анализе полученных данных, можно сделать вывод о том, что с помощью алгебраического метода пики хорошо разделяются, без смещения положения вершин пиков и без изменения их амплитуд.

Однако появляются ложные частоты с большими амплитудами, которые после применения фильтра «ложных частот» сглаживаются.

После применения фильтра происходит искажение амплитуд сформированных пиков.

Однако это не является недостатком, так как зная параметры фильтра можно восстановить первоначальные амплитуды.

Если сравнить полученные данные по двум методам, то можно сделать вывод, что в ситуации идеально сформированных пиков, метод Тихонова дает лучший результат разделения «наложившихся» пиков нежели алгебраический метод.

Формирование исходного сигнала с добавлением шума

Формирование исходного сигнала с добавлением шума может происходить несколькими способами, как показано на Рис.6.16.

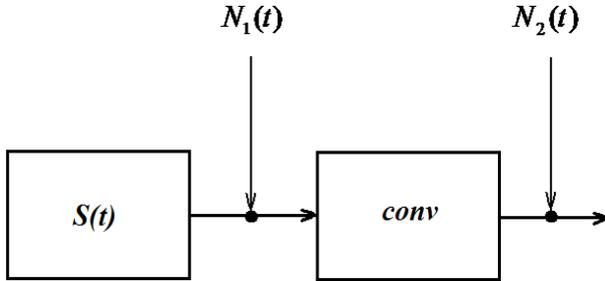


Рис.6.16. Возможные варианты наложения шума,

$S(t)$ -исходный сигнал, $N_1(t)$ и $N_2(t)$ - шумы

$N_1(t)$ - это шум, который можно добавлять в идеальный спектр перед операцией «конволюция», а $N_2(t)$ - это шум, который можно добавлять в исходный сигнал после операции «свертки».

Проверим работоспособность исследуемых нами методов при наличии шума в исходном сигнале. На Рис.6.17. представлен входной сигнал с наложенным шумом $N_1(t)$ и $N_2(t)$ и восстановленный сигнал после применения метода регуляризации Тихонова (зеленый цвет).

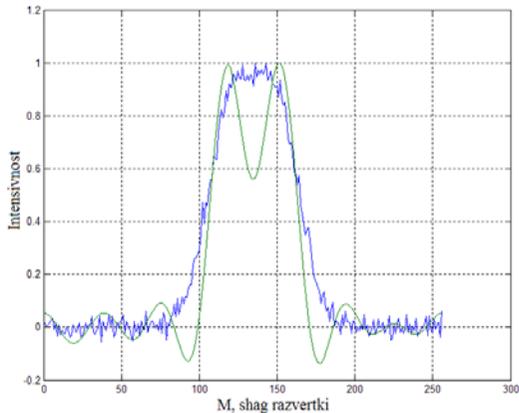


Рис.6.17. Исходный сигнал с добавлением шума и восстановленный сигнал методом регуляризации Тихонова

На Рис.6.18. представлен входной сигнал без шума (красный цвет), сигнал с наложенным шумом $N_1(t)$ и $N_2(t)$ (зеленый цвет) и восстановленный сигнал после применения алгебраического метода (синий цвет).

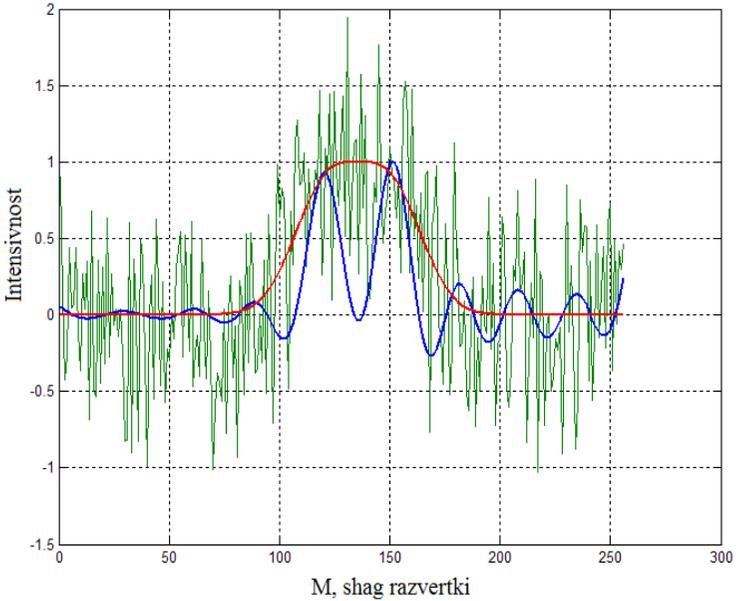


Рис.6.18. Исходный сигнал, сигнал с добавлением шума и восстановленный сигнал алгебраическим методом

В ходе эксперимента было выявлено, что алгебраический метод наиболее эффективен при сильном зашумлении входного сигнала.

Чтобы наглядно представить эффективность методов разделения наложившихся пиков, были получены результаты работоспособности метода регуляризации Тихонова и алгебраического метода при различных отношениях сигнала к шуму (S/N) и представлены в виде Таблицы 6.1.

В данной таблице обозначение «+» соответствует «работоспособности», а обозначение «-» соответствует «не работоспособности».

Таблица 6.1
Работоспособность методов Тихонова и алгебраического в зависимости от S/N

S/N	Метод Тихонова	Алгебраический метод (SVD)
100	+	+
50	+	+
20	+	+
13	+	+
10	-	+
5	-	+
2	-	+

Из Таблицы :6.1 видно, что метод регуляризации Тихонова работает лучше при отношении сигнала к шуму равном 13 и выше, а алгебраический метод работоспособен при небольшом отношении сигнала к шуму равном 2 и выше.

Так же проанализирована работоспособность метода регуляризации Тихонова и алгебраического метода при различном соотношении амплитуд наложившихся пиков в мультиплете A_1 / A_2 и результаты представлены в виде Таблицы 6.2.

В данной таблице обозначение «+» соответствует «работоспособности», а обозначение «-» соответствует «не работоспособности».

Таблица 6.2
Работоспособность методов Тихонова и алгебраического в зависимости A_1 / A_2

A_1/A_2	Метод Тихонова	Алгебраический метод (SVD)
5	+	+
10	+	+
15	+	+
20	+	-
30	-	-

Из Таблицы 6.2 видно, что метод регуляризации Тихонова лучше работает с большим отношением амплитуд наложившихся пиков, нежели алгебраический метод.

Так как в методе регуляризации Тихонова важным параметром является параметр регуляризации α , то был проведен анализ эффективности повышения разрешения спектров в зависимости от параметра регуляризации α при разных соотношениях сигнал/шум и представлен в виде Таблицы 6.3.

В данной таблице R означает новое разрешение, а R_1 - исходное разрешение.

Таблица 6.3 Зависимость повышения разрешения от α и S/N

α	S/N	R/R_1
10^{-30}	107	5
10^{-25}	107	4.3
10^{-15}	104	3.1
10^{-10}	103	2.1
10^{-6}	50	1.8

Из Таблицы 6.3 видно, что разрешение улучшается при увеличении параметра регуляризации α и уменьшении шума.

Заключение

Итогом проведенного выше изложения является раскрытие физических принципов приборов, которые относятся к важнейшим направлениям химического анализа таких как масс-спектрометрия, хроматография, капиллярный электрофорез, мессбаурской спектроскопии. Отдельное внимание было уделено описанию принципов действия приборов, предназначенных анализа наноматериалов в их числе оптические и зондовые микроскопы, электронные и рентгеновские спектрометры, приборы на основе явлений ядерного и электронного парамагнитного резонанса. Несмотря на большое разнообразие, эти приборы имеют много общих методов по представлению и обработке информации. Наиболее важными из этих методов обработки являются

использование прямого и обратного преобразования Фурье, цифровая фильтрация, анализ и учет аппаратной функции, вейвлет-анализ и проекционные методы. Достаточно сложными, разнообразными, но в настоящее время еще недостаточно изученным являются методы обработки информации, предназначенные для повышения разрешающей способности приборов математическими методами. В данном пособии раскрыты классические алгоритмы повышения разрешения, основанными на решениях обратных задач методами регуляризации Тихонова и SVD. Описание других методов повышения разрешения с помощью программного обеспечения не вошло в данное пособие не вошло в данное пособие в связи с ограниченным его объемом и эти методы следует описать в отдельном пособии. Для практического использования в приложениях имеются исходные тексты на языке системы Матлаб для решения обратных задач методами Тихонова и SVD.

Таким образом, из данного пособия должно быть ясно, современное аналитическое приборостроение постоянно ставит новые задачи, требующие развития и совершенствования алгоритмических методов обработки данных, направленных на достижение предельных показателей качества анализа.

Список литературы

1. Zeev B. Alfassi. Statistical Treatment of Analytical Data/ Zeev B. Alfassi, Zvi Boger, Yigal Ronen. - CRC Press, 2005. - 266 p.
2. Ахмед Н. Ортогональные преобразования при обработке цифровых сигналов/ Н. Ахмед, К.Р. Пао.- М.: Связь, 1980.- 248 с.
3. Howard Mark. Statistics in Spectroscopy/ Howard Mark, Jerry Workman – Academic Press, Elseiver, 2003.-328 p.
4. Howard Mark. Chemometrics in Spectroscopy/ Howard Mark, Jerry Workman – Academic Press, Elseiver, 2007.-559 p.
5. Adams M. J. Chemometrics in Analytical Spectroscopy/ Mike J. Adams. – The Royal Society of Chemistry, 1995.- 216 p.
6. Foo-tim Chau. Chemometrics From Basics to Wavelet Transform/ Foo-tim Chau, Yi Zeng Liang, Junbin Gao, Xue-Guang Shao. – John Wiley&Sons, Inc., 2004.- 316 p.
7. Большаков, А.А. Методы обработки многомерных данных и временных рядов/ А.А.Большаков, Р.Н. Каримов.- М.: Горячая линия-Телеком, 2007. – 522 с.

8. Эсбенсен К. Анализ многомерных данных/ Ким Эсбенсен. – Черноголовка: Издательство ИПХФ, 2005.- 158 с.
9. Hyvarinen A. Independent Component Analysis/ Aaro Hyvarinen, Juha Karhunen, Erkki Oja. – John Wiley&Sons, Inc., 2001.- 505 p.
10. Левин Б.Р. Теоретические основы статистической радиотехники/ Б.Р. Левин. – 3-е изд, - М.: Радио и связь, 1989. – 656 с.
11. Гонсалес, Р. Цифровая обработка изображений в среде МАТЛАБ/ Р Гонсалес, Р. Вудс, С. Эддинс. – М.: Техносфера, 2006. – 618 с.
12. Ефимов, В.М. Анализ и прогноз временных рядов методом главных компонент/ В.М. Ефимов, Ю.К. Галактионов, Н.Ф. Шушпанова. – Новосибирск: Наука, Сибирское отделение, 1988. – 74 с.
13. Гонсалес, Р. Цифровая обработка изображений/ Р Гонсалес, Р. Вудс. – М.: Техносфера, 2005. – 1070 с.
14. Советов, Б.Я. Моделирование систем: учеб. пособие для вузов/ Б.Я. Советов, С.А. Яковлев. – М.: Высшая школа, 2001. – 343 с.
15. Косарев Е.Л. Методы обработки экспериментальных данных: учеб. пособие/ Е.Л. Косарев. – М.: МФТИ, 2003. – 258 с.
16. Добеши И. Десять лекций по вевлетам/ И. Добеши. – Москва – Ижевск: РИЦ РХД, 2001. – 418 с.
17. Шараф, М.А. Хемометрика/ М.А. Шараф, Д.Л. Иллмэн, Б.Р. Ковальски. – Л.: Химия, Ленинградское отделение, 1989. - 269 с.
18. Родионова, О. Е. Интервальный метод обработки результатов многоканальных экспериментов: дис. ... д-ра ф.-м. наук: 01.04.01/ О. Е. Родионова; ИХФ им Н.Н.Семёнова РАН.- М., 2008.-272 с.
19. Буляница, А.Л. Линеаризация информационных сигналов в микроаналитических приборах и методы их обработки: дис. ... д-ра ф.-м. наук: 01.04.01/А.Л.Буляница; ИАП РАН. – СПб, 2008. – 278 с.
20. Манойлов, В.В. Развитие методов обработки информации в масс-спектрометрии для изотопного и элементного анализа: дис. ... д-ра техн. наук: 01.04.01/В.В. Манойлов; ИАП РАН. – СПб, 2007. – 263 с.
21. Галь, Л.Н. Физические основы масс-спектрометрии и ее применение в аналитике и биофизике/ Галь Л.Н. . — Санкт-Петербург: Издательство Политехнического университета 2010. —164 с.
22. Рудаков, О.Б. Спутник хроматографиста/ О.Б. Рудаков, И.А. Востров. - Воронеж: Водолей, 2004. — 528 с.
23. Яшин, Я. И. Газовая хроматография/ Яшин Я. И., Яшин Е. Я., Яшин А. Я. — М.: Транслит, 2009. — 512 с.
24. Вертхейм, Г. Эффект Мессбауэра/ Г. Вертхейм. – М.: Мир, 1966. – 169с.
25. Шпинель, В.С. Резонанс гамма-лучей в кристаллах/ В.С. Шпинель. – М.: Наука, 1969. – 407 с.

26. Сверхтонкие взаимодействия и ядерные излучения/ Н.Н. Делягин, Б.А. Комиссарова, Л.Н. Крюкова, В.П. Парфёнова, А.А. Сорокин, М.: Изд. МГУ, 1985. – 240 с.
27. Беленький, Б. Г. Высокоэффективный капиллярный электрофорез/ Б.Г. Беленький .- СПб.: Наука, 2009.- 320 с.
28. Троян, В.И. Физические основы методов исследования наноструктур и поверхности твердого тела: учебное пособие./ В.И. Троян, М.А. Пушкин, В.Д. Борман, В.Н. Тронин; под общ. ред. В.Д. Бормана. - М.: МИФИ, 2008. - 260 с.
29. Батаев, В.А. Методы структурного анализа материалов и контроля качества деталей: учебное пособие / В.А. Батаев, А.А. Батаев, А.П. Алхимов. - 2-е изд. - М.: Флинта. Наука, 2007. - 224 с.
30. Оура, К. Введение в физику поверхности/ К. Оура, В. Г. Лифшиц, А. А. Саранин [и др.], под ред. В. И. Сергиенко. — М.: Наука, 2006. — 490 с.
31. Поздняков, В. А. Физическое материаловедение наноструктурных материалов / В.А. Поздняков. – М.: МГИУ, 2007. – 424 с.
32. Абрамчук, Н.С. Нанотехнология - азбука для всех / Н.С. Абрамчук, С.М. Авдошенко, А.Н. Баранов [и др.] - М.: Физматлит, 2008. - 368с.
33. Брандон, Д. Микроструктура материалов. Методы исследования и контроля/ Д. Брандон, У. Каплан М. – М: Техносфера, 2004. - 384 с.
34. Синдо, Д. Аналитическая просвечивающая электронная микроскопия/ Д. Синдо, Т. Оикава.- М.: Техносфера, 2006. - 256 с.
35. Ельяшевич, М. А. Атомная и молекулярная спектроскопия: общие вопросы спектроскопии/ М. А. Ельяшевич. – М.: КОМКНИГА, 2006 – 896 с.
36. Лиопо, В.А. Рентгеновская дифрактометрия: учебное пособие/ В.А. Лиопо, В.В. Война.- Гродно: ГрГУ, 2003. - 171 с.
37. Большаков, А. А. Перспективы аналитической атомной спектроскопии/ А. А. Большаков, А. А. Ганеев, В. М. Немец // Успехи химии. - 2006, - Т. 75, № 4. – С. 322-338
38. Пентин, Ю.А. Основы молекулярной спектроскопии/ Ю.А. Пентин, , Г. М. Курамшина - М.: Мир; БИНОМ, 2008. - 398 с
39. Brereton, R.G. Chemometrics: Data analysis for the laboratory and chemical plant/ R.G. Brereton. - Wiley, Chichester, UK, 2003.- 489 pp.
40. Аналитическая химия. Проблемы и подходы (в 2-х т.), под. ред. Кельнер Р., Мерме Ж.-М., Отто М., Видмер Г.М., пер. с англ., М., Мир АСТ, 2004
41. Френкс, Л. Теория сигналов/ Л. Френкс. - М.: Сов.радио, 1974. - 343с.

42. Миркин, Б. Г. Анализ качественных признаков и структур/ Б. Г. Миркин. - М.: Статистика, 1980.- 319 с.
43. Литтл, Р. Дж.А. Статистический анализ данных с пропусками / Р. Дж.А. Литтл, Д. А. Рубин, пер. с англ. - М.: Финансы и статистика, 1990. - 366 с.
44. Ахмед Н. Ортогональные преобразования при обработке цифровых сигналов/ Н. Ахмед, К.Р.Рао. Пер. с англ.- М.: Связь, 1980. – 248 с.
45. Литвак, Б. Г. Экспертная информация: Методы получения и анализа/ Б. Г. Литвак. - М.: Радио и связь, 1982. - 184 с.
46. Каримов, Р. Н. Восстановление пропусков в многомерных данных методом сингулярного разложения/ Р.Н. Каримов, О. В. Червякова // Вопросы преобразовательной техники и частотного электропривода и управления: Межвуз. научн. сб. Саратов. госуд. техн. ун-т. - 1999. - С. 19-26.
47. Курнаев, В.А. Влияние флуктуаций потенциалов на аппаратные функции электростатических анализаторов/ В.А. Курнаев, В.А. Урусов // Журнал технической физики. – 1997, - т. 67, №6. – С.92-95.
48. Шевченко, С.И. Метод вычисления аппаратной функции аксиальных электростатических энергоанализаторов/ С.И. Шевченко// Научное приборостроение. – 2010, т.20, №2. – С.73-81.
49. Малла С. Вейвлеты в обработке сигналов/ С. Мала, пер: с англ. - М.Мир, 2005. – 338 с.
50. Трахтман А.М. Введение в обобщенную спектральную теорию сигналов/ А.М. Трахтман. - М.: Сов. Радио, 1972. – 352 с.
51. Солодовников А. И. Основы теории и методы спектральной обработки информации: учеб. пособие / А. И. Солодовников, А. М. Спиваковский. - Л.: Изд-во ЛГУ, 1986.- 269 с.
52. Прикладная статистика: Классификация и снижение размерности/ Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. – М.: Финансы и статистика, 1989.- 607 с.
53. Andrade, J.M. Procrustes rotation in analytical chemistry, a tutorial/ J.M. Andrade, M. P. Gomez-Carracedo, W. Krzanowski, M. Kubista// Chemom. Intell. Lab. Syst.- 2004, v.72 (2). - pp.123 -132
54. Brereton, R.G. Chemometrics: Data analysis for the laboratory and chemical plant/ R.G.Brereton/ - Wiley, Chichester, UK, 2003. – 478 p.
55. Сирвидас С.И., Заруцкий И.В., Ларионов А.М., Манойлов В.В. Использование метода сверток с произвольными базовых функций для обнаружения и разделения пиков в экспериментальных данных // Dspa'99. Доклады. Т. 1. С. 105.
56. Карасек Ф., Клемент Р. Введение в хромато-масс-спектрометрию. – М.: Мир, 1993. 237 с

57. Разников В.В., Разникова М.О. Информационно-аналитическая масс-спектрометрия. – М.: Наука, 1991. 248 с.
58. Тихонов А.Н., Арсение В.Я. Методы решения некорректных задач. – М.: Наука, 1986.- 275 с.
59. Вайнштейн Л. А., Собельман И. И., Юков Е. А. Возбуждение атомов и уширение спектральных линий. – М.: Наука, 1979. 319 с.
60. Антохин Ю. Т. Некорректные задачи для уравнения типа свертки // Дифференциальные уравнения. 1968. Т. 4. №9 С. 129–178.
61. Сирвидас С.И., Заруцкий И.В., Ларионов А.М., Манойлов В.В. Обнаружение, разделение и оценка параметров масс-спектрометрических пиков методом свертки экспериментальных данных с производными гауссовых шумов функций // Научное приборостроение. 1999. Т. 9. №4. С. 84.
62. Жданов А.И. Введение в методы решения некорректных задач. – Самара: Изд. СГАУ, 2007. 79с.
63. Василенко Г.И., Тараторин А.М. Восстановление изображений. – М.: Радио и связь, 1986. 303 с.
64. Бардин Б.В., Чубинский-Надеждин И.В. Обнаружение локальных объектов на цифровых микроскопических изображениях // Научное приборостроение. 2009. Том 19. № 4. С. 102.
65. Вержбецкий В.М. Численные методы: линейная алгебра и нелинейные уравнения). – М.: Высш. шк., 2000. 266с.
66. Деммель Дж. Вычислительная линейная алгебра. – М.: Мир, 2001. 430 с.
67. Гуревич А.Л., Могильницкий А.М., Русинов Л.А. Автоматизация обработки масс-спектрометрической информации. – М.: Энергия, 1978. 182 с.
68. Раутиан С. Г. Реальные спектральные приборы // Успехи физических наук. 1958. Т. 66. №. 3. С. 475 – 517.
69. Дьяконов В.П., Абраменкова И. В. Справочник по Matlab. Обработка сигналов и изображений. Специальный справочник. – СПб.: Питер, 2002. 608 с.
70. V.I. Gelfgat, E.L. Kosarev, E.R. Podolyak. Program for signal recovery from noisy using the maximum likelihood principle //Comput. Phys. Commun. 1993. V.74. P.335.
71. .А.В Погуляй., Ю.В Васильев., Г.М Туймедов., В.А.Мазунов Повышение разрешения спектрометра методами деконволюции // Химия и компьютерное моделирование. Бутлеровские сообщения. 1999, № 2, приложение Материалы VI Всероссийской конференции “Структура и динамика молекулярных систем”.
72. .А.В Погуляй, Р.Р. Аблазимов, Г.М Туймедов., В.А.Мазунов Сравнительный анализ методов деконволюции // Химия и компьютер-

ное моделирование. Бутлеровские сообщения. 2001, № 6, приложение Материалы VIII Всероссийской конференции “Структура и динамика молекулярных систем”.

73. Б.В. Бардин Б.В., В.Д. Белов В.Д., Л.В. Новиков, Ю.В. Чижов Использование оптимального фильтра Винера для деконволюции электронных спектров.// Научное приборостроение. Т.9. №1. 1999:

74. .R.I. Shrager Chemical transition measured by spectra resolved using singular value decomposition // Chemom. Intell. Lab. Syst., vol 1, Issue 1 (November 1986) P 59-70

75. .B.V. Bardin, V.D. Belov V.D., N.V.Mamro ,I.R. Nikolaev, Yu. V. Chizov A Computerized Control and Data Handling System of a Multiunctional Electron Spectrometer and Software for Processing Complex electron Spectra // Instrumnts and Experimentl Techniques vol 42 No 2 1999 pp 205-211 translated from Pribory I Tekhnika Experimenta No 2 , 1999 pp 68-74.

76. С.И. Сирвидас, И.В.Заруцкий, А.М. Ларионов, В.В. Манойлов Обнаружение, разделение и оценка параметров масс-спектрометрических пиков методом свертки экспериментальных данных с производными гауссовых шумов функций // Научное приборостроение. 1999. Т. 9. №4. С. 84

77. В.В. Манойлов, И.В.Заруцкий Возможности алгоритма сверток с производными для оценки параметров масс-спектров, содержащих наложившиеся пики// Научное приборостроение. 2009. Т. 19. №4. С. 103-109

78. Нувärinen Ааро, Survey on Independent Component Analysis, Neural Computing Surveys 2, 94-128, 1999, <http://www.icsi.berkeley.edu>

79. Т. Н. Князева, Л. В. Новиков, Н. И. Орешко. Удаление нестационарного шума из экспериментальных данных// Научное приборостроение_2008. Т.18. №2. С. 61-65

80. Косарев Е.Л. О пределе сверхразрешения при восстановлении сигналов // Радиотехника и электроника . 1990. Т. 35. № 1. С. 68–87;

81. Коротких В.Л., Косарев Е.Л., Ормонт А.Б., Коротких А.В. Улучшение энергетического разрешения фотоэлектронных спектров программной корреляцией на аппаратную функцию // Приборы и техника эксперимента. 1994. №6. С. 88– 94

82. Баркер Ф . Компьютеры в аналитической химии: Перевод с английского. М., Мир.1987 г 520 с.

**Программа решения задач восстановления масс-спектров
методом регуляризации Тихонова**

```

clear all;
delta_t_ms=10; %interval diskredizacii v ms;
delta_t_sec=delta_t_ms/1000; %interval diskredizacii v sec;
t=1:256; t1=0:delta_t_sec:delta_t_sec*256;
b=5;a=128;a1=120;
h1=exp(-((t-a1)/b).^2);
a2=150; h2=1.*exp(-((t-a2)/b).^2);
h=h1+h2+0.000*randn(1,length(h1));
figure,plot(h); grid;
title('ishodniy signal'); b1=b*4;
z=exp(-((t-a)/b1).^2);
u1=conv(h,z);
mx=length(t); u2=u1(128:length(u1)-128);
Fs=100; [b,a] = butter(3,43*2/Fs);
[H,w]=freqz(b,a,128); figure,plot(w*Fs/(2*pi),abs(H));
xlabel('Frequency (Hz)'); ylabel('Mag. of frequency response');
grid;
length(u2); Df=fft(u2); b2=4;
fK(1:128)=exp(-((t(1:128)-1)/b2).^2);
fK(129:256)=exp(-((t(129:256)-256)/b2).^2);
b3=40;
fbG(1:128)=exp(-((t(1:128)-1)/b3).^2);
fbG(129:256)=exp(-((t(129:256)-256)/b3).^2);
alfa_reg=1e-25;
N=1;
for(i=1:mx)
    if(i<mx/2)
        iw = i;
    else
        iw = mx - i;
    end
    w(i) = (iw*iw).^2;
    Flt(i) = fK(i) / (fK(i)*fK(i) + alfa_reg*w(i));
    Bt(i)=sqrt(1./(1+(w(i)./10000).^2.*N));
    end
    Df_N=Df.*Flt.*1;
df=ifft(Df_N);
figure,plot(t,u2/max(u2),t,real(df)/max(real(df)));

```

```

grid;
title('real df after restore & u2');
Pdf = df.* conj(df) / 256;
Ffdf=fft(real(df));
Ffh=fft(h);
Y=Ffh;
Pyy = Y.* conj(Y) / 512;
Y1=Ffdf;
Pyy1 = Y1.* conj(Y1) / 512;
figure,plot(Pyy1);
title('fft vost signal');
grid;
figure, plot( real(df));
grid;
title('real df after restore');
f1=0:0.5/128:0.5;
f1=0:1/256:1;% shcala chastoty dla Fs=1;
figure,plot(f1(1:129),Pyy1(1:129)); % grafic dla opredelenia chastoty
sreza
grid;
title('fft 0.5 points of vost signal');
Frec_cut = 0.03;
Fs=1;
[b,a] = butter(3,Frec_cut*2/Fs);
[H,w]=freqz(b,a,128); % postroenie ACH
figure,plot(w*Fs/(2*pi),abs(H));
grid;
ylabel('Mag. of frequency response');
xlabel('Frequency (Hz)');
sf=filter(b,a,real(df));
figure,plot(sf);
grid;
title('real df after restore and filter');

```

**Программа решения задач восстановления масс-спектров
алгебраическим методом на основе SVD**

```

clear all;
t=1:256;
b=5;
a=128;
a1=121;
format long;
h1=1.*exp(-((t-a1)/b).^2);
a2=150;
h2=1.*exp(-((t-a2)/b).^2);
h=h1+h2+0.00*randn(1,length(h1));
figure,plot(h);
grid;
title('ishodniy signal')
b1=b*4;
z=exp(-((t-a)/b1).^2);
z1=exp(-((t-1)/b1).^2);
for i=1:length(t)
Rnd1(i,:)=0.05*randn(1,length(t));
end
fid =fopen('Noise1','r');
F = fread(fid,'float');
fclose(fid);
for i=1:length(t)
Rnd1(i,:)=(F(1+ length(t)*(i-1): length(t)*i));
end
for i=1:length(z1)
z2=exp(-((t-i)/b1).^2)+0.00*randn(1,length(h1));
A(i,:)=z2;
end
A=A+0.0*Rnd1;
C=A*h';%(+0.01*randn(1,length(h1)))';
figure,plot(C);
grid;
title('after conv');
a11=118;
for i=1:length(z1)
z2=exp(-((t-i)/b1).^2)+0.5*randn(1,length(h1));
A2(i,:)=z2;

```

```

end
C11=1.*exp(-((t-a1)/(4.0*b)).^2)+1.*exp(-((t-a2)/(4.0*b)).^2)+
0.05*randn(1,length(h1));;
u1=conv(h,z);
u2=u1(128:length(u1)-128)+ 0.001*randn(1,length(h1));
figure,plot(u2);
grid;
title('after conv2');
b=inv(A)*C;
[U,S,V] = svd(A);
b1=inv(U*S*V)*C;
sigma=svd(A);
S2=zeros(length(z1));
for i=1:length(z1)
S2(i,i)=sigma(i);
end
tau1=max(sigma);
tau2=min(sigma);
for i=1:length(sigma)
z(i)=C(i)/sigma(i);
if sigma(i)<100*tau2
z(i)=1;
end
end
for i=1:length(sigma)
S2(i,i)=sigma(i);
if sigma(i)<7.5e+16*tau2
S2(i,i)=0;
end
end
k=rank(S2);
A0=(U*S2*V');
k=rank(A0);
b2=u2';
b22=U'*b2;
for(i=1:k)
W(i,1:k)=S(i,1:k);
b3(i)=b22(i);
end
b3=b3';
y=inv(W)*b3;
for(i=1:length(C11))

```

```
        if i<=k
            y(i)=y(i);
        else
            y(i)=0;
        end
    end
    x=V*y;
    figure,plot(x);
    grid;
    title('after separate C');
    SG_F=sgolayfilt(x,3,41);
    figure,plot(t,SG_F,t,x);
    grid;
    title('after separate C and Fir');
    tau1=max(sigma);
    tau2=min(sigma);
    cond=tau1/tau2;
```

Миссия университета – генерация передовых знаний, внедрение инновационных разработок и подготовка элитных кадров, способных действовать в условиях быстро меняющегося мира и обеспечивать опережающее развитие науки, технологий и других областей для содействия решению актуальных задач.

КАФЕДРА НАНОТЕХНОЛОГИЙ И МАТЕРИАЛОВЕДЕНИЯ

Свое название кафедра получила в 2006 году в связи с расширением исследований в области нанотехнологий. За последние несколько лет на кафедре выполнено ряд работ по проектированию новых алгоритмов обработки информации аналитических приборов, которые расширяют возможности современных аналитических приборов, в том числе для анализа материалов. В числе таких работ алгоритмы оценки параметров неразрешенных спектральных пиков, прецизионного управления изотопного масс-спектрометра и алгоритмы цифрового управления зондового микроскопа.

Манойлов Владимир Владимирович
Новиков Лев Васильевич

**ПОЛУЧЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ
АНАЛИТИЧЕСКИХ ПРИБОРОВ**

Учебное пособие

В авторской редакции

Редакционно-издательский отдел Университета ИТМО

Зав. РИО

Н.Ф. Гусарова

Подписано к печати

Заказ №

Тираж

Отпечатано на ризографе

**Редакционно-издательский отдел
Университета ИТМО
197101, Санкт-Петербург, Кронверкский пр., 49**