

ФИЗИКО-ТЕХНОЛОГИЧЕСКИЕ ОСНОВЫ МАКРО-, МИКРО- И НАНОЭЛЕКТРОНИКИ

**Под общей редакцией
профессора А.А. Барыбина**

Рекомендовано учебно-методическим объединением вузов
Российской Федерации по образованию в области радиотехники,
электроники, биомедицинской техники и автоматизации в качестве
учебного пособия для студентов высших учебных заведений,
обучающихся по направлениям

210100 «Электроника и нанoeлектроника»,
211000 «Конструирование и технология электронных средств»,
222900 «Нанотехнологии и микросистемная техника»



**МОСКВА
ФИЗМАТЛИТ®**

2011

ОГЛАВЛЕНИЕ

Предисловие	8
Введение	17
Глава 1. Общие принципы термодинамического управления равновесными и неравновесными процессами	73
1.1. Статистическое определение энтропии	76
1.2. Свойства энтропии. Статистический смысл температуры . . .	81
1.3. Закон сохранения энергии в открытых системах	85
1.4. Типы термодинамических процессов	88
1.5. Тепловой эффект изохорных и изобарных процессов	90
1.6. Тепловой эффект и энтропия необратимых процессов.	95
1.7. Второе начало термодинамики и производство энтропии . . .	97
1.8. Равновесие, устойчивость и эволюция замкнутых и закры- тых систем	100
1.9. Стационарность, устойчивость и эволюция открытых систем	106
1.10. Термодинамические потенциалы и соотношения между ними	119
1.11. Химический потенциал и парциальные молярные величины	122
1.12. Термодинамические функции идеального газа	127
1.13. Понятие о стандартном состоянии веществ и таблицах стан- дартных термодинамических величин	132
1.14. Условия фазового и химического равновесия	135
1.15. Правило фаз Гиббса	138
1.16. Условия фазового и электрохимического равновесия в систе- мах с заряженными частицами	140
1.17. Флуктуационная устойчивость равновесного состояния . . .	143
1.18. Нарушение условий фазового и химического равновесия . . .	148
1.19. Формализм теории Онсагера	155
1.20. Теорема Пригожина о минимуме производства энтропии . . .	160
1.21. Локальное квазиравновесие и локальные термодинамические соотношения	167
1.22. Локальные уравнения баланса числа частиц, внутренней энергии и энтропии.	175

1.23. Плотность потока энтропии и локальное производство энтропии	184
1.24. Локальная устойчивость неравновесного состояния и универсальный принцип эволюции	197
1.25. Контрольные вопросы	201
 Глава 2. Управление фазовыми и химическими равновесиями в технологических задачах электроники	205
2.1. Фазовые превращения в однокомпонентных системах.	206
2.2. Давление насыщенного пара чистого вещества.	211
2.3. Термодинамика образования твердых и жидких растворов . .	215
2.4. Модели и термодинамические свойства растворов	222
2.5. Давление насыщенных паров над растворами и смесями веществ в конденсированном состоянии	230
2.6. Молекулярная и атомарная растворимость газов в конденсированных средах	237
2.7. Равновесие жидкой и твердой фаз в однокомпонентных и бинарных системах	242
2.8. Диаграмма плавкости бинарных систем без твердых растворов	249
2.9. Диаграммы плавкости бинарных систем с неограниченным твердым раствором	252
2.10. Диаграммы плавкости бинарных систем с ограниченными твердыми растворами	256
2.11. Диаграммы плавкости бинарных систем с химическими соединениями в твердой фазе	262
2.12. Закон действующих масс и константы химического равновесия	269
2.13. Направление протекания химической реакции. Уравнение изотермы Вант-Гоффа	275
2.14. Температурная зависимость констант химического равновесия. Уравнения изобары и изохоры Вант-Гоффа	280
2.15. Управление химическими реакциями. Принцип Ле Шателье	284
2.16. Газотранспортные химические реакции	288
2.17. Ионное равновесие и термодинамические свойства растворов электролитов.	293
2.18. Термодинамика электрохимических процессов	298
2.19. Электродные процессы в электрохимических системах.	307
2.20. Управление ионными процессами в водных растворах электролитов	317

2.21. Анализ процессов в гальваническом и электролитическом режимах работы элемента	322
2.22. Контрольные вопросы	329

Глава 3. Управление структурными равновесиями и дефектообразованием в кристаллах 333

3.1. Химические связи в твердых телах	334
3.2. Классификация пространственных решеток кристаллов	342
3.3. Пространственная упаковка и координация частиц в кристаллических структурах.	349
3.4. Классификация и свойства атомных дефектов кристаллической структуры	369
3.5. Точечные дефекты и физические свойства кристаллов	377
3.6. Квазихимический метод описания дефектов.	399
3.7. Электронно-дырочное равновесие в полупроводниках.	402
3.8. Растворимость примесей в полупроводниках с учетом ионизации примесных атомов	406
3.9. Взаимодействие заряженных дефектов: компенсация и ассоциация	411
3.10. Внутреннее равновесие собственных и примесных дефектов	418
3.11. Управление собственными дефектами путем отжига кристаллов в парогазовой среде.	424
3.12. Распределение амфотерной примеси в кристаллической решетке полупроводников.	428
3.13. Контрольные вопросы	431

Глава 4. Управление диффузионными и кинетическими процессами в технологических задачах электроники. . . . 434

4.1. Механизм кинетики химических реакций. Уравнение Аррениуса	435
4.2. Механизмы диффузии атомов в твердом теле.	439
4.3. Законы Фика. Начальные и граничные условия в задачах диффузии	442
4.4. Диффузионные задачи на удаление вещества из твердого тела	445
4.5. Принципы вакуумного обезгаживания материалов	454
4.6. Роль диффузии в газопроницаемости вакуумных оболочек. .	457
4.7. Диффузионные задачи на введение вещества в твердое тело	462
4.8. Принципы диффузионного легирования полупроводников. . .	468

4.9. Диффузионная и химическая кинетика гетерогенных процессов	475
4.10. Маскирующие свойства слоев двуоксида кремния	483
4.11. Кинетика термического окисления кремния	487
4.12. Кинетика химического травления полупроводников	495
4.13. Принципы выращивания монокристаллических слоев методами эпитаксии	501
4.14. Кинетика химического транспорта веществ в проточных системах	514
4.15. Кинетика химического транспорта веществ в сэндвич-системах	519
4.16. Кинетика химического транспорта веществ при реактивном распылении	527
4.17. Кинетика процессов в электрохимических системах	540
4.18. Контрольные вопросы	549
 Глава 5. Управление свойствами поверхности, межфазными взаимодействиями и формированием нанообъектов. . .	552
5.1. Квантовые особенности нанообъектов пониженной размерности	555
5.2. Структурно-геометрические особенности и квазикристаллическое состояние нанокластеров	560
5.3. Атомная структура, релаксация и реконструкция поверхностей	568
5.4. Структурные дефекты реальных поверхностей	577
5.5. Термодинамика поверхностных слоев и межфазных границ	581
5.6. Поверхностно-активные вещества, термодинамические свойства и технологии. Пленки Ленгмюра–Блоджетт	589
5.7. Основы золь-гель технологии наноструктурированных материалов	605
5.8. Термодинамика искривленных поверхностей	621
5.9. Адсорбционные явления на поверхности твердых тел	644
5.10. Кинетика мономолекулярной адсорбции и двумерная конденсация	655
5.11. Полимолекулярная адсорбция и капиллярная конденсация	671
5.12. Роль адсорбции, растворения и диффузии в газопоглощении материалов	682
5.13. Движущие силы и термодинамика гомогенного зародышеобразования	685

5.14. Термодинамика и кинетика гетерогенного зародышеобразования	695
5.15. Механизмы формирования и роста пленок на подложках . . .	713
5.16. Механизмы формирования герметичных соединений материалов	737
5.17. Контрольные вопросы	743
Приложение А. Термодинамические расчеты в технологических задачах электроники	747
А.1. Состав газовой фазы и окисляемость металлов при термообработке оксидного катода	747
А.2. Химическое осаждение арсенида галлия из газовой фазы в системе GaAs–H ₂ O–H ₂	756
Приложение Б. Справочный материал	763
Б.1. Важнейшие физические константы	763
Б.2. Некоторые внесистемные единицы	763
Б.3. Основные обозначения	764
Литература, рекомендуемая для углубленного изучения	771
Предметный указатель	773

Предисловие

Настоящая книга является существенно переработанным и дополненным вариантом ранее изданного учебного пособия¹⁾, написанного по материалам курса лекций, читаемого двумя соавторами (А.А.Б. и В.И.Ш.) в Санкт-Петербургском государственном электротехническом университете (бывшем ЛЭТИ) для студентов факультета электроники. В книге также частично использованы материалы лекционного курса, читаемого третьим соавтором (В.И.Т.) в Сибирском федеральном университете (бывшем КГТУ) и изданного как учебное пособие²⁾.

Переработка и дополнения вызваны настоящей необходимостью включения в лекционный курс, во-первых, базовых положений термодинамики неравновесных процессов (глава 1) и, во-вторых, физико-технологических основ современной наноэлектроники (глава 5). Кроме того, в главу 3 по дефектообразованию в кристаллах добавлены элементы кристаллохимии и кристаллографии.

Изложение вопросов современной нанотехнологии отразилось и в названии книги, которое теперь, наряду с уже привычными *микроэлектроникой* и *наноэлектроникой*, содержит новый термин «**макроэлектроника**». Сюда мы относим все многообразие дискретных приборов вакуумной, газоразрядной и твердотельной электроники с размерами от миллиметров (навесные элементы гибридных интегральных схем) до дециметров и вплоть до метров (рентгеновские трубки, мощные СВЧ-приборы, генераторные и модуляторные лампы). Эти приборы, принципиально разработанные еще в давние годы, успешно функционируют и ныне. На наш взгляд, введенная нами триада «**макро – микро – нано**», наиболее полно и естественно отражает все направления современной электроники в ее широком понимании.

¹⁾ Барыбин А.А. Электроника и микроэлектроника. Физико-технологические основы. — М.: Физматлит, 2006.

²⁾ Томилин В.И., Томилина Н.П. Физико-химические и термодинамические основы материаловедения и технологий электронных средств. — Красноярск: ИПЦ КГТУ, 2004.

Название книги «Физико-технологические основы... электроники», как и ранее, отражает технологическую направленность в изучении физических и физико-химических явлений и закономерностей, лежащих в основе принципов, методов и процессов, используемых в производстве современных электронных приборов и устройств. Именно этим данное учебное пособие отличается от учебных дисциплин, близких по названию к «Физическим основам электроники», где изучаются физические закономерности в вакууме, плазме и твердом теле, положенные в основу принципа действия того или иного прибора.

Повседневный быт современного человека заполнен разнообразными электронными устройствами — телевизорами и плазменными панелями, микроволновыми печами и холодильниками, персональными компьютерами, сотовыми телефонами, спутниковыми навигаторами и другой техникой, обеспечивающей высокий комфорт в жизни населения планеты. «Большой взрыв» в развитии техники, в частности радиоэлектроники, возникший в последние десятилетия прошлого века, в первую очередь обязан прогрессу в разработке элементной базы электроники.

Высокотехнологичная электронная индустрия, являясь одной из важнейших базовых отраслей, во многом определяет сегодня место развитого государства в постиндустриальном информационном обществе. Электронная промышленность, взявшая старт в начале прошлого века с выпуска вакуумного диода и триода, почти за столетие прошла громадный путь. Примером современных достижений в массовом электронном производстве можно назвать кремниевые процессоры, имеющие на одном квадратном сантиметре до 300 млн транзисторов при минимальном размере элементов 32 нм. Но этот результат, хотя и является впечатляющим, не исчерпывает всех достижений ученых и инженеров, работающих над поиском и созданием новых наноэлементов.

Еще около двадцати лет назад ключевое направление развития электроники связывали с планарно-интегральной технологией, на основе которой изготавливали кремниевые интегральные микросхемы, приборы оптоэлектроники, акустоэлектроники, криоэлектроники и спинволновой электроники. Но на рубеже веков возникла «нанотехнологическая революция», охватившая практически все важные сферы жизнедеятельности человека. По мнению большинства экспертов в области научно-технической политики, последствия внедрения нанотехнологических разработок будут обширнее и глубже, чем результаты «компьютерной революции» последней трети прошлого столетия. В настоящее время среди элементов наноэлектроники уже имеются лазеры на

квантовых ямах и одноэлектронные транзисторы, изготовленные методами современной нанотехнологии.

Элементная база электроники включает приборы, перекрывающие огромный частотный диапазон от сверхнизких и низких звуковых частот до высоких и сверхвысоких радиочастот, кончая оптическим и рентгеновским излучением. Они выполняют самые разнообразные функции в системах передачи, приема, хранения и обработки информации. Сюда относятся такие функции, как:

- генерация и усиление сигналов;
- различные виды преобразования спектра сигналов (модуляция, детектирование, смещение, фильтрация, сжатие, расширение или фурье-трансформация сигналов, временная задержка и свертка сигналов);
- коммутационные преобразования в устройствах хранения и цифровой обработки сигналов;
- электрооптические и фотоэлектрические преобразования в устройствах индикации и визуализации сигналов.

Одно только перечисление функциональных назначений приборов свидетельствует об их огромном многообразии. Более того, такое многообразие расширяется за счет различия в принципе действия, частотно-геометрических факторах и конструктивно-типовых признаках приборов. Эти различия естественно порождают многообразие технологических процессов, методов, операций и приемов изготовления электронных приборов и устройств.

Номенклатура современной элементной базы электроники (включая электровакуумные, газоразрядные, полупроводниковые и другие приборы дискретного и интегрального микро- и наноисполнения) обширна и разнообразна по своему функциональному назначению, принципу работы, конструктивным и технологическим особенностям. Поэтому не представляется возможным детальное и углубленное изучение технологии всех типов приборов в рамках часов, предусмотренных учебными планами технологических дисциплин.

В этих условиях при построении учебного курса единственно возможным и методически оправданным остается *дедуктивный подход* к отбору учебного материала (от общего к частному). Такой подход должен базироваться на изложении фундаментальных физических и физико-химических закономерностей, присутствующих различным технологическим процессам, используемым при изготовлении приборов современной электроники. Эти общие закономерности характерны практически для всех типов приборов, принадлежащих к одному из трех направлений электроники:

- **макроэлектроника**, к которой отнесены все типы электровакуумных, газоразрядных, дискретных полупроводниковых приборов и пленочных интегральных схем, изготавливаемых по *индивидуальным технологиям* для каждого типа приборов;

- **микроэлектроника**, в которую объединены твердотельные приборы и интегральные микросхемы, для производства которых используются *интегрально-групповые технологии*;

- **наноэлектроника**, которая включает современные приборы и наноструктуры, изготавливаемые методами *нанотехнологии*, в том числе с применением новых наноматериалов.

Характерной особенностью современного этапа развития технологии элементной базы электроники является быстрое техническое освоение последних достижений в различных областях науки. Сюда относятся: физика вакуума, физика плазмы, физика твердого тела и тонких пленок, физика поверхности, физика взаимодействия с твердым телом концентрированных потоков энергии (пучков заряженных частиц и фотонов), химическая термодинамика и кинетика, электрохимия, кристаллохимия, материаловедение и многое другое. Многообразие технологических методов, приемов и процессов, используемых в современной электронике, делает методически оправданным вышеупомянутый дедуктивный подход при лекционном изложении курса технологии электронных приборов в ее широком понимании.

В основе любого технологического процесса лежит определенное физическое, химическое, электрохимическое или иное воздействие на материал с целью управляемого изменения его состояния, структуры или состава. Такой взгляд на технологические процессы позволяет выделить их базовые физико-технологические черты и показать общие физические закономерности, управляющие этими процессами вне зависимости от принадлежности прибора к той или иной группе. Именно подобный методологический подход был положен в основу построения настоящего учебного пособия. Авторский отбор материала, необходимого для изложения физических и физико-химических основ технологии, определялся исключительно его применимостью к физико-технологическим проблемам электроники (в ее макро-, микро- и нано-реализациях). Бесспорно, это ограничивает круг рассматриваемых физических явлений, но позволяет в достаточно компактной форме охватить наиболее значимые технологические процессы.

Введение открывает учебное пособие с целью дать читателю общее представление о современном состоянии элементной базы электроники в ее исторической ретроспективе. Кратко изложены

основные технологические особенности изготовления электронных приборов макроэлектроники и описаны пути перехода кремниевой микропроцессорной технологии к наноразмерному диапазону. Нанотехнологическое направление развития элементной базы электроники связано с перспективами в области создания новых (в том числе углеродных) наноматериалов и с разработкой новых методов и приемов нанотехнологии и нанодиагностики.

Первая глава является базовой и содержит фундаментальные вопросы равновесной и неравновесной термодинамики, важные для понимания тех физических закономерностей, которые составляют основу многообразных процессов, используемых в технологии современной электроники. Традиционное изложение равновесной термодинамики базируется на трех постулатах, называемых началами термодинамики. Они феноменологически выражают общие закономерности, наблюдаемые в природе, без выяснения их микроскопической основы. Благодаря общности базовых положений, термодинамический подход дает результаты, применимые к разнообразным макроскопическим системам.

В отличие от традиционного изложения, в первой главе за основу взят статистический подход, основанный на больцмановском определении энтропии. В таком изложении известные начала равновесной термодинамики, выражающие закон сохранения энергии в открытых системах, критерий самопроизвольности процессов в закрытых системах и принцип недостижимости абсолютного нуля температуры, появляются в обратном порядке. Полученные в этой главе общие соотношения обратимой термодинамики, в частности условия фазового и химического равновесия, активно работают в последующих главах.

В первую главу включены также концептуальные вопросы неравновесной термодинамики необратимых процессов, базирующиеся на фундаментальных результатах Онсагера и Пригожина. Сюда относятся элементы линейной теории Онсагера, обоснование понятий локального квазиравновесия, потока и производства энтропии, теорема Пригожина о минимуме производства энтропии и универсальный принцип эволюции неравновесных систем. Эти вопросы существенно расширяют физический кругозор читателя и формируют новый взгляд на необратимость неравновесных процессов в открытых системах как источник самоорганизации системы, приводящий к формированию устойчивых диссипативных структур по терминологии Пригожина. Параграфы, посвященные изложению этих вопросов (пп. 1.19–1.24), при первом чтении книги могут быть опущены.

Вторая глава рассматривает основные физические закономерности, которые описывают фазовые равновесия и управляют изменением фазового состояния веществ, а также основные положения химической термодинамики для систем с реактивными компонентами. Сюда также включены те вопросы, связанные с ионными и электрохимическими процессами, которые имеют непосредственное отношение к технологическим проблемам электроники. Как уже отмечалось, несмотря на многообразие процессов технологии, они могут быть объединены общностью физических явлений, лежащих в основе их действия. Например, технологические операции, проводимые в вакууме и газовых средах (такие как вакуумное обезгаживание и очистка материалов, термовакуумное испарение и конденсация, газофазная эпитаксия, легирование полупроводников из газовой фазы и др.), управляются законами испарения и сублимации конденсированных сред, растворения и диффузии газов в этих средах и т. п.

Закономерности взаимодействия между жидкой и твердой фазами, описываемые при помощи диаграмм плавкости, лежат в основе таких технологических операций, как пайка припоями, формирование омических контактов к полупроводникам, жидкофазная эпитаксия, кристаллизационные методы выращивания и очистки веществ и др. Законы химического и электрохимического взаимодействия управляют процессами очистки, травления и окисления поверхности, химическим транспортом веществ при газофазной эпитаксии, формированием термоэмиттирующего и люминофорного покрытий в электровакуумных приборах и т. п. Большинство из вышеперечисленных вопросов и составляет основное содержание второй главы.

Третья глава содержит изложение основных физических закономерностей, которые определяют поведение дефектов в кристаллах и позволяют управлять ими для целенаправленного изменения электрофизических свойств путем легирования, то есть искусственного создания точечных дефектов в кристаллических полупроводниковых структурах.

Некоторые свойства твердых тел (плотность, упругость, решеточная теплоемкость, диэлектрическая проницаемость) мало чувствительны к дефектности кристалла и адекватно описываются в рамках модели идеальной кристаллической решетки. Однако большинство физических свойств (электрические, магнитные, оптические, механические, тепловые) являются дефекточувствительными, то есть зависят от степени несовершенства кристаллической решетки. Поскольку реальные твердые тела

имеют в большей или меньшей степени нарушения периодической структуры, то управление дефектообразованием в технологических процессах приобретает важное практическое значение.

Как уже отмечалось, материал этой главы изначально дополнен краткими сведениями из кристаллохимии и кристаллографии (пп. 3.1–3.4), необходимыми для последующего изложения как в этой главе, так и в главе 5.

Четвертая глава рассматривает неравновесные физические явления, возникающие при нарушении условий фазового и химического равновесия, в применении к технологическим системам. В предыдущих главах были рассмотрены равновесные закономерности, лежащие в основе управления фазовыми, химическими и электрохимическими превращениями веществ и процессами дефектообразования в кристаллах полупроводников. Принципиальная возможность (или невозможность) смещения термодинамического равновесия в нужном направлении устанавливалась на основании знака приращения изобарного (или изохорного) потенциала ΔG (или ΔF) для рассматриваемого процесса. Однако практическая осуществимость процесса определяется не только термодинамическими, но и кинетическими факторами.

Важную роль при этом, в частности при формировании скорости протекания процесса, играют существенно неравновесные закономерности, лежащие в основе химической и диффузионной кинетики. Именно эти закономерности управляют процессами вакуумного обезгаживания и газопроницаемости материалов, диффузионного легирования полупроводников, термического окисления кремния, химического травления полупроводников и эпитаксиального выращивания монокристаллических слоев, рассмотрение которых и составляет содержание четвертой главы.

Наряду с химическим транспортом веществ в проточных и сэндвич-системах, в эту главу добавлено рассмотрение кинетики химического транспорта при реактивном распылении металлов.

Пятая глава подверглась наибольшему изменению в сравнении с предыдущим изданием учебного пособия. Эти изменения, в первую очередь, связаны с дополнительным включением вопросов, относящихся к современным проблемам нанотехнологии и наноматериалов. Сюда относятся квантовые особенности нанообъектов пониженной размерности, структурно-геометрические особенности и квазикристаллическое состояние нанокластеров, атомная структура и дефекты поверхности, поверхностно-активные вещества в технологии пленок Ленгмюра–Блоджетт, основы золь-гель технологии, полимолекулярная адсорбция, двумерная

и капиллярная конденсация. Остальной материал главы, посвященный изложению основных положений термодинамики поверхностных явлений и межфазных взаимодействий, претерпел некоторые изменения за счет внесения дополнений. В первую очередь, они имеют отношение к таким физико-технологическим проблемам, как сорбционные явления на поверхности твердых тел, модели зародышеобразования и механизмы эпитаксиального роста пленок на совершенных подложках (механизмы Фольмера–Вебера, Франка–Ван-дер-Мерве и Странского–Крастанова).

Приложение А содержит изложение методики термодинамических расчетов в применении к двум избранным технологическим задачам электроники: *во-первых*, анализ состава газовой фазы и окисляемости конструкционных материалов в вакуумном приборе при термообработке оксидного катода и, *во-вторых*, расчет процесса химического осаждения арсенида галлия из газовой фазы в системе GaAs–H₂O–H₂. Справочный материал, включая обозначения используемых величин, приведен в **приложении Б**.

В учебное пособие, как и ранее, не включены вопросы, относящиеся к процессам электронного, ионного и лазерного воздействия на материалы, которые сегодня находят применение в микроэлектронной и нанoeлектронной технологиях. Это сделано сознательно, так как специфика физических явлений при взаимодействии корпускулярных потоков с поверхностью твердых тел не вписывается в избранную нами «термодинамическую схему» изложения физических и физико-химических основ технологии.

В настоящем учебном пособии изложение материала построено так, чтобы дать читателю возможность самостоятельно сформировать общие физико-технологические представления путем изучения основных физических, химических и электрохимических закономерностей, как правило, без обращения к другой литературе. Для этого рассмотрение большинства вопросов начинается с обсуждения исходных модельных представлений, на базе которых приводится детальный вывод необходимых математических соотношений с последующей интерпретацией их физического содержания. Каждая глава заканчивается перечнем контрольных вопросов. С целью расширения кругозора студенческой аудитории в книге даны подстрочные сноски, содержащие краткие научно-биографические сведения об известных ученых.

Для углубленного изучения отдельных разделов в конце книги приведен список рекомендуемой литературы, который дополнен перечнем избранных публикаций по отдельным вопросам наноматериалов и нанотехнологий, приведенным в конце введения.

Книга предназначена главным образом для студентов в качестве учебного пособия по физико-технологическим основам микроэлектроники, микроэлектроники и наноэлектроники. Она может оказаться полезной также и специалистам в этих областях, поскольку содержит некоторые вопросы, выходящие за рамки учебной дисциплины, которые могут быть опущены студентами (например, пп. 1.19–1.24, 2.20, 2.21, 4.14–4.17).

Авторы благодарят своих коллег, деловое общение с которыми оказало полезное воздействие на методический отбор материала и структуру книги. Искреннюю признательность авторы адресуют В. В. Кузнецову и О. Ф. Луцкой за замечания при чтении отдельных глав рукописи книги, которые содействовали ее улучшению.

Любые критические замечания и пожелания, высланные по нижеуказанному адресу, будут приняты с благодарностью.

Санкт-Петербург
сентябрь 2010 г.

А. А. Барыбин
barybin@mail.ru

Введение

ЭТАПЫ РАЗВИТИЯ И СОВРЕМЕННОЕ СОСТОЯНИЕ ТЕХНОЛОГИИ МАТЕРИАЛОВ И ПРИБОРОВ МАКРО-, МИКРО- И НАНОЭЛЕКТРОНИКИ

Зарождение электроники как научно-технического направления связывают с именами Джона Флеминга (изобретение диода, 1904 г.) и Ли де Фореста (изобретение триода, 1906 г.)¹⁾. С тех пор за сто с небольшим лет электроника (в широком ее понимании, включающем все современные направления) прошла огромный путь, определивший на сегодняшний день по сути дела весь научно-технический прогресс человечества. За эти годы можно четко проследить три стимулирующие тенденции, которые формировали различные направления в развитии электроники:

- повышение рабочей мощности устройств,
- освоение все более высоких рабочих частот,
- стремление к микроминиатюризации аппаратуры.

Позднее, с появлением интегральных микросхем и микропроцессоров, наметилась четвертая тенденция — переход от аналоговой обработки сигналов к цифровой обработке с целью повышения быстродействия устройств.

¹⁾ Флеминг (Fleming) Джон Амброз (1849–1945) — английский ученый в области электротехники и радиотехники, изобрел (1904) двухэлектродную вакуумную лампу с термокатодом (диод) и выпрямитель на ее основе (назвав их, соответственно, «кенотроном» и «осцилляторным вентилем»), известен работами в области электрических трансформаторов, электроизмерений, фотометрии, электроники и беспроводной (радио) связи. Удостоен (1933) Медали почета IRE (Института радиоинженеров) за «заметную роль, которую он сыграл в деле внедрения физических и инженерных принципов в радиотехнике».

Ли де Форест (Lee de Forest) (1873–1961) — американский изобретатель в области радиоэлектроники, изобрел (1906) вакуумный триод (названный им «аудионом») и создал на его основе приемник для беспроводного телеграфа, дав ему впервые название «радио», считается основателем общественного радиовещания (1910) и автором первого патента на озвучивание фильмов (1919). Удостоен (1922) Медали почета IRE (Института радиоинженеров) за «изобретение трехэлектродной лампы и признание его большого вклада в радио».

Перечисленные выше тенденции развития электроники привели в настоящее время к огромному многообразию конкретных электронных приборов и устройств, предназначенных для решения различных естественно-научных, инженерно-технических, эколого-космических, медико-биологических и повседневных бытовых проблем жизнеобеспечения людей.

В историческом плане можно выделить следующие основные этапы, которые прошла электроника до современного состояния.

1. Эра вакуумной электроники начинается с появления первых вакуумных приборов (диода и триода) в начале прошлого столетия и успешно продолжается вплоть до наших дней в классе мощных генераторных и модуляторных ламп, приборов СВЧ-диапазона (клистроны, магнетроны, лампы бегущей и обратной волны и др.), рентгеновских и электронно-лучевых приборов (осциллографические, радиолокационные, передающие и приемные телевизионные трубки, мониторы компьютеров и др.) и разнообразных газоразрядных приборов силовой и бытовой электроники.

2. Эра полупроводниковой электроники начинается с изобретения транзистора в 1947 году Д. Бардиным, У. Браттейном и У. Шокли, получившими впоследствии Нобелевскую премию¹⁾.

¹⁾ Бардин (Bardeen) Джон (1908–1991) — крупнейший американский физик, специалист в области физики твердого тела и физики низких температур, один из изобретателей полупроводниковых транзисторов (точечного, 1947 г., и плоскостного, 1950 г.) и один из соавторов микроскопической теории сверхпроводимости; исследуя проблемы электропроводности металлов, первым предсказал притяжение между электронами благодаря обмену виртуальными фононами («куперовские пары»). Вторая Нобелевская премия по физике (совместно с Л. Купером и Дж. Шриффером, 1972) «за совместное создание теории сверхпроводимости, обычно называемой теорией БКШ».

Браттейн (Brattain) Уолтер Хаузер (1902–1987) — американский физик-экспериментатор, специалист в области физики полупроводников, один из создателей первых полупроводниковых транзисторов; изучал влияние адсорбционных пленок на электронную эмиссию с поверхности твердых тел, исследовал механизмы рекомбинации зарядов и поверхностные свойства германия и кремния, чрезвычайно важные для работы полевого транзистора, предложенного позже У. Шокли.

Шокли (Shockley) Уильям Брэдфорд (1910–1989) — американский физик, специалист по физике твердого тела (ферромагнетизм, пластичность металлов, теория дислокаций), изучал роль дефектов и поверхностных явлений в полупроводниках (поверхностные состояния Шокли), первым наблюдал дырочную проводимость, исследовал эффекты инъекции носителей заряда, обнаружил «эффект поля» и построил теорию p - n -перехода, на основе которой предложил плоскостной p - n - p -транзистор.

Совместная Нобелевская премия по физике (1956) «за исследования полупроводников и открытие транзисторного эффекта».

Изначально полупроводниковые диоды и триоды (транзисторы) исполнялись в виде дискретных (корпусных) элементов на основе германия или кремния, которые заменяли вакуумные приборы в электрических цепях, составленных из резисторов, конденсаторов и индуктивностей. Это продолжалось около десятка лет, пока накопленный технологический опыт не позволил создавать то или иное функциональное устройство в виде единой *интегральной схемы* — *гибридной* (ГИС с навесными полупроводниковыми элементами в бескорпусном исполнении) или *монокристаллической* (МИС с пассивными и активными элементами схемы, изготовленными на одной подложке в едином технологическом цикле). Так зародилась новая эра интегральной электроники.

3. Эра квантовой электроники ведет отсчет с изобретения принципа действия мазера/лазера¹⁾ в 1954 году Н. Г. Басовым, А. М. Прохоровым и независимо от них Ч. Таунсом, получившими впоследствии Нобелевскую премию²⁾. Первые демонстрации мазерного эффекта на пучке молекул аммиака инициировали широкий фронт теоретических и экспериментальных исследований, главным образом, в радиочастотном диапазоне. Лишь в 1960 году

¹⁾ Сокращение от англ. *Microwave/Light Amplification Stimulated by Emission of Radiation*, предложенное Ч. Таунсом.

²⁾ Басов Николай Геннадиевич (1922–2001) — советский и российский физик, специалист по квантовой электронике, установил (вместе с А. М. Прохоровым) принцип создания квантовых усилителей и генераторов, предложил различные типы полупроводниковых, газовых, химических и эксимерных лазеров, разработал физические основы создания квантовых стандартов частоты и выдвинул идею использования лазеров для управления термоядерным синтезом.

Прохоров Александр Михайлович (1916–2002) — советский и российский физик, специалист в области радиофизики, радиоспектроскопии, квантовой электроники и нелинейной оптики, сформулировал (вместе с Н. Г. Басовым) основные принципы квантового усиления и генерации, исследовал структуру волновых пучков в нелинейной среде, распространение оптических солитонов в световодах, лазерную генерацию ультразвука и управление свойствами твердого тела и лазерной плазмы при воздействии световыми пучками.

Таунс (Townes) Чарльз Хард (род. 1915) — американский физик, специалист в области радиоспектроскопии, нелинейной оптики, радиоастрономии и квантовой электроники, выдвинул идею генерации и усиления электромагнитных волн и на ее основе создал первый квантовый генератор (мазер на аммиаке), позже применил лазеры для высокоточной проверки специальной теории относительности и для исследований в астрофизике, открыл мазерный эффект в космосе, ввел новые современные методы инфракрасного детектирования на основе лазеров в астрономическую спектроскопию и интерферометрию.

Совместная Нобелевская премия по физике (1964) «за фундаментальную работу в области квантовой электроники, которая привела к созданию генераторов и усилителей, основанных на лазерно-мазерном принципе».

был создан оптический квантовый генератор (лазер) на кристалле рубина, а через два года появились первые инжекционные лазеры на основе p – n -перехода в полупроводнике GaAs. Однако работали они в импульсном режиме при пониженных температурах и потому представляли малый практический интерес, несмотря на высокий коэффициент полезного действия (до 70%).

Основной задачей было получение непрерывной генерации при малых пороговых токах и комнатной температуре с сохранением высокого к.п.д. Решение этой задачи оказалось возможным путем замены гомоперехода в GaAs двойной гетероструктурой (ДГС) — узкозонный GaAs между двумя широкозонными полупроводниками на основе твердого раствора $\text{Al}_x\text{Ga}_{1-x}\text{As}$. Первыми это удалось реализовать в 1970 году сотрудникам Физико-технического института имени А. Ф. Иоффе (Ленинград) под руководством Ж. И. Алферова, получившего впоследствии Нобелевскую премию¹⁾. Последующие работы этого коллектива и конкурирующих с ним зарубежных фирм над созданием гетеролазеров со сверхрешетками и размерным квантованием заложили физико-технологические основы современной оптоэлектроники. Именно появление ДГС-лазеров открыло широкую дорогу для применения интегрально-оптических систем и систем волоконно-оптической связи во всемирной сети Интернет.

4. Эра интегральной электроники начинается с промышленного внедрения *кремниевой планарной технологии*, разработанной к началу 60-х годов на базе освоенных к тому времени технологических процессов. Сюда в первую очередь следует отнести: эпитаксиальное выращивание монокристаллических пленок, диффузионное легирование, термическое окисление кремния, вакуумное напыление алюминия, фотолитографию. Именно эти процессы сформировали надежный фундамент для перехода к планарно-эпитаксиальной технологии интегральных схем (ИС). У истоков этой технологии стояли два человека — Дж. Килби из Texas Instruments, получивший впоследствии, через сорок лет,

¹⁾ Алферов Жорес Иванович (род. 1930) — советский и российский ученый, специалист в области полупроводниковой оптоэлектроники и физики наноструктур пониженной размерности, разработал (с участием сотрудников ФТИ им. А. Ф. Иоффе) новые физико-технологические принципы конструирования полупроводниковых гетероструктур на основе уникальных материалов, включая объекты пониженной размерности (квантовые ямы, проволоки/нити, точки) для разнообразных приборных применений (лазеры, светодиоды, солнечные батареи и транзисторы). Нобелевская премия по физике (совместно с Г. Кремером, 2000) «за разработку полупроводниковых гетероструктур для быстродействующей электроники и оптоэлектроники».

Нобелевскую премию¹⁾, и Р. Нойс из Fairchild Semiconductor²⁾. Их плодотворные идеи, сформулированные независимо друг от друга (1959 г. — патент Килби: ГИС на германии; патент Нойса: МИС на кремнии), дали весомый инженерно-практический вклад, определив (в большей степени технологическими идеями Р. Нойса) направление развития интегральных схем как элементной базы будущей *цифровой микроэлектроники*.

Наряду с этим, в те же годы возникли новые принципы аналоговой обработки сигналов, положенные в основу так называемой *функциональной электроники*. Здесь привычные схемотехнические решения заменены нетрадиционным использованием особого рода динамических неоднородностей в твердом теле (например, в форме движущихся доменов или линейных и нелинейных волн различной природы) для выполнения тех или иных функций. Именно это породило такие направления функциональной электроники, как акустоэлектроника, спин-волновая электроника, плазменно-волновая электроника и криоэлектроника.

Несмотря на ряд технически привлекательных и многообещающих эффектов, открытых при разработке этих направлений функциональной электроники, наибольшие перспективы в настоящее время связывают все-таки с наноэлектроникой.

5. Эра наноэлектроники в ее сегодняшнем широко популярном виде, как и другие нанотехнологии, имеет тем не менее известные физические истоки и технологические предпосылки. Для подтверждения этого приведем в хронологическом порядке наиболее значимые научно-технические достижения, на которых сегодня базируются нанотехнология и физика наноструктур.

¹⁾ Килби (Kilby) Джек Сен-Клер (1923–2005) — американский инженер-схемотехник, специалист в прикладной электронике, первым выдвинул идею интеграции множества элементов электрической цепи в одном кристалле, запатентовал около 60 изобретений, в том числе карманный калькулятор и термопринтер; его имя занесено в Национальный зал славы изобретателей США, среди которых А. Нобель, Г. Форд, Т. Эдисон и другие известные имена. Нобелевская премия по физике (2000) «за вклад в создание интегральных схем».

²⁾ Нойс (Noyce) Роберт Нортон (1927–1990) — американский инженер-электронщик, работал в лаборатории У. Шокли, изобретателя транзисторов, позже совместно с Гордоном Муром (см. сноску на с. 32) основал компании по производству кремниевых микропроцессоров — Fairchild Semiconductor (1957) и Intel (1968). Общительного Роберта Нойса по-дружески называли «мэром Силиконовой долины» (всемирно известного района в Северной Калифорнии, где сосредоточены предприятия компьютерной электроники). Практически во всех современных интегральных микросхемах используется разработанная Нойсом производственная технология, и только преждевременная кончина не позволила ему разделить Нобелевскую премию совместно с Дж. Килби.

- 1933 г. — создание электронного микроскопа просвечивающего типа с разрешением 50 нм (*Ernest Ruska*, Нобелевская премия по физике, 1986).
- 1939 г. — коммерческий выпуск электронного микроскопа с разрешением 10 нм (компания *Siemens*).
- 1966 г. — изобретение пьезодвигателя, примененного позже в сканирующем туннельном микроскопе для позиционирования с точностью до 10^{-3} нм (*Russel Yang*).
- 1968 г. — разработка метода газофазной эпитаксии из металлоорганических соединений (*Harold Manasevit*).
- 1971 г. — изобретение и разработка принципов молекулярно-лучевой эпитаксии (МЛЭ, *Alfred Cho*).
- 1973 г. — реализация методом МЛЭ полупроводниковой сверхрешетки GaAs–Al_{0,5}Ga_{0,5}As с периодом 7–10 нм (*Leo Esaki*, Нобелевская премия по физике, 1973).
- 1974 г. — появление термина «нанотехнология» (*nanos* по гречески «карлик»), означающего процесс модифицирования материала путем воздействия одиночными атомами или молекулами (*Norio Taniguchi*).
- 1982 г. — создание сканирующего туннельного микроскопа с межатомным разрешением (*Gerd Binnig, Heinrich Rohrer*, Нобелевская премия по физике, 1986).
- 1985 г. — открытие фуллеренов (*Richard Smalley, Robert Curl, Harold Croto*, Нобелевская премия по физике, 1996).
- 1986 г. — создание атомно-силового микроскопа с межатомным разрешением (*Gerd Binnig, Christophe Gerber*).
- 1988 г. — создание ДГС-лазера на *квантовой яме*, ограниченной короткопериодными сверхрешетками, с рекордно низким пороговым током (лаборатория Ж. И. Алферова, ФТИ имени А. Ф. Иоффе, Ленинград).
- 1991 г. — открытие углеродных нанотрубок (*Sumio Iijima*).
- 1999 г. — разработка общих принципов манипулирования атомами или молекулами (*Mark Reed, James Tour*).
- 2000 г. — создание гетеролазера на *квантовых точках* с выходной мощностью 3,5–4 Вт в непрерывном режиме и к.п.д. до 50% (лаборатория Ж. И. Алферова, ФТИ имени А. Ф. Иоффе, Санкт-Петербург).
- 2004 г. — получение графена, однослойной аллотропной модификации углерода (*Andre Geim, Konstantin Novoselov*, Нобелевская премия по физике, 2010).

Таким образом, в течение нескольких предшествующих десятилетий был достигнут серьезный технологический задел в создании и контроле элементов нанометровых размеров. Это открывает обнадеживающую перспективу для дальнейшего развития электроники нанометрового диапазона в направлении воспроизводимости и надежности технологических результатов.

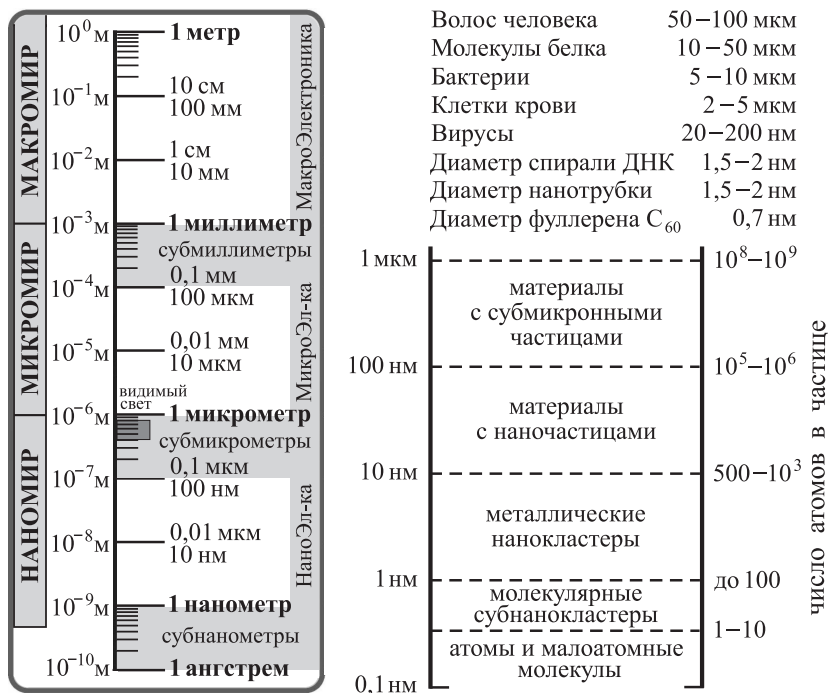


Рис. 1. Шкала размеров, характерных для объектов макромира (макроэлектроники), микромира (микроэлектроники) и наномира (наноэлектроники)

На рис. 1 приведены диапазоны размеров физических объектов, обосновывающие введение триады «**макро – микро – нано**» как для окружающего нас мира (макромир – микромир – наномир), так и для электроники (макроэлектроника – микроэлектроника – наноэлектроника), что ранее было отмечено в предисловии. Здесь также даны характерные размеры естественных (биологических) и искусственно созданных объектов. Из их сравнения видно, что технология сегодняшнего уровня уже позволяет создавать предельно малые объекты с размерами нанометрового и субнанометрового диапазона, такие как углеродные нанотрубки (УНТ), фуллерены, графены и им подобные.

Объектам макроэлектроники условно выделен диапазон размеров более 1 мм, объектам микроэлектроники — от 1 мкм до 100 мкм, а объектам наноэлектроники — от 1 нм до 100 нм. Между ними располагаются промежуточные области (выделенные серым цветом на рис. 1) *субмиллиметрового диапазона* (от 1 мм до 0,1 мм) и *субмикрометрового* (как часто говорят, *субмикронного*) *диапазона* (от 1 мкм до 0,1 мкм). В этих областях проявляются одновременно физико-технологические особенности объектов, принадлежащих соседним примыкающим диапазонам.

«Кирпичиками» нанометрового диапазона являются так называемые *кластеры* (от англ. *cluster*) — скопления взаимодействующих друг с другом атомных частиц, не превышающие 10 нм. *Наночастицами* называют частицы с размером в несколько десятков нанометров, но, как правило, не более 100 нм. Они содержат вплоть до 10^6 атомов, а их свойства, как и свойства кластеров, существенно отличаются от свойств объемного вещества, состоящего из таких же атомов.

Объектами *субнанометровой* области (затемненной на рис. 1) с размерами от 1 нм до 0,1 нм (1 Å) являются атомы, обычные (малоатомные) молекулы, более сложные (многоатомные) молекулы (типа молекул ДНК, фуллерена C_{60} и др.) и супрамолекулярные структуры (субнанокластеры), образованные межмолекулярными связями (типа дендритов, мицелл и т. п.). Примерно на пять порядков ниже лежит область внутриядерных частиц (не показанная на рисунке).

Следует обратить внимание на некоторую противоречивость вышеизложенной терминологии с той, которая принята в физике. Квантовая (микроскопическая) физика имеет дело с микрочастицами (атомами, ионами, молекулами) размером в десятки доли и единицы ангстрема, в отличие от классической (макроскопической) физики, изучающей свойства объемных материалов и макротел. Размерную область между ними (от долей миллиметра до единиц нанометра) принято называть *мезофизикой* («мезо» от греч. *средний*). Следовательно, мезофизика перекрывает диапазон размеров, характерных для микро- и наноэлектроники. Поэтому *физическая триада* «макро–мезо–микро» до некоторой степени эквивалентна *электронной триаде* «макро–микро–нано» и по сути соответствует ее смыслу (см. вводную часть гл. 5).

Задача последующего изложения — краткое описание основных этапов изготовления дискретных приборов макроэлектроники, технологических особенностей современной микроэлектроники и направлений развития и путей перехода к наноэлектронике.

МАКРОЭЛЕКТРОНИКА:

технологические особенности изготовления приборов

К приборам микроэлектроники будем относить те приборы, которые по своим массо-габаритным параметрам выходят за общепринятые рамки изделий микроэлектроники. Сюда, кроме приборов вакуумного и полупроводникового класса, можно также отнести разнообразные диэлектрические и ферритовые элементы, которые находят широкое применение в современной радиоэлектронной аппаратуре. Как правило, подобного рода приборы изготавливают как дискретные элементы устройств (тем более, мощные приборы) и размещают индивидуально внутри аппаратуры.

Ниже остановим свое внимание только на технологии вакуумных приборов и дискретных полупроводниковых приборов (ДПП), современный внешний вид которых показан на рис. 2.

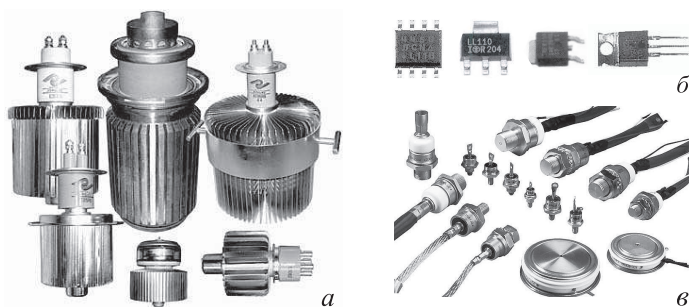


Рис. 2. Элементная база современной макроэлектроники: *а* — вакуумные генераторные и модуляторные лампы; *б* — полупроводниковые мощные транзисторы; *в* — полупроводниковые силовые диоды и тиристоры

Вакуумные приборы имеют два общих признака, характерных для прибора любого типа и функционального назначения:

- *дискретный характер конструкции* прибора, состоящего из отдельных деталей и узлов — вакуумного корпуса, катода, подогревателя, анода, сеток, катодной ножки и других элементов, собранных в единую конструкцию;

- *дискретно-индивидуальный характер технологии*, порожденный дискретностью конструкции и проявляющийся в форме строгой последовательности технологических операций, которая сохраняется для всех без исключения приборов, но применяется, как правило, индивидуально к каждому экземпляру прибора.

Технологический маршрут изготовления вакуумного прибора любого типа (в том числе и газоразрядного) в обязательном порядке включает следующие операции (см. рис. 3).



Рис. 3. Технологический маршрут изготовления вакуумных приборов

Операции формообразования — для получения необходимой формы деталей при помощи механических, термических, электрофизических и электрохимических методов обработки материалов, таких как обработка резанием и давлением, порошковая металлургия и отливка деталей под вакуумом, ультразвуковая, электроискровая, электрохимическая, электронно-лучевая и лазерная обработка.

Операции по получению вакуумно-чистых деталей — для очистки поверхности и обезгаживания деталей с помощью таких методов, как обезжиривание (в трихлорэтилене и четыреххлористом углероде), химическое и электрохимическое травление и полировка (в водных растворах кислот и щелочей), промывка (в дистиллированной воде и этиловом спирте, в том числе с применением ультразвука) и рафинирующий отжиг (в вакууме или в атмосфере водорода) — см. пп. 4.4, 4.5 и 5.6.

Операции по нанесению поверхностных покрытий и защитных пленок — для обеспечения активных функций (термоэмиссионные, фотоэмиссионные, антиэмиссионные, люминесцентные и газопоглощающие покрытия) и вспомогательных функций (электроизоляционные, электропроводящие, теплоизлучающие и антикоррозийные покрытия) с помощью таких методов, как пульверизация, гальваническое и электрофоретическое осаждение, термовакuumное и электронно-лучевое испарение, катодное и ионно-плазменное распыление — см. п. 3.5.

Монтажно-сборочные операции и вакуумная герметизация прибора — для сборки отдельных узлов и конструкции прибора в целом при помощи таких методов, как пайка припоями, сварка плавлением (газовая, аргоно-дуговая, плазменно-лучевая, электронно-лучевая, лазерная), сварка давлением (электроконтактная, термодиффузионная, ультразвуковая, холодная), спаи металлов со стеклом и керамикой — см. п. 5.16.

Операции при откачке — для обезгаживания деталей, активировки катода и наполнения газом с применением термического прогрева в печах, нагрева деталей высокочастотным индуктором или электронной бомбардировкой и напуска рабочей газовой смеси (если необходимо) — см. пп. 3.5 и 4.4–4.6.

Операции на отпаянном приборе — для стабилизации газовой среды, формирования рабочих параметров прибора и придания ему «товарного вида» при помощи распыления газопоглотителя (если необходимо) и длительной тренировки прибора в жестком электрическом режиме, а также цоколевки, окраски и выборочных испытаний приборов.

В зависимости от класса и типа прибора отдельные технологические этапы имеют свои отличительные особенности. Так, маломощные приборы откачивают на автоматах и полуавтоматах с кратким циклом откачки (измеряемым часами и даже минутами), на заключительном этапе которого лишь распыляется газопоглотитель, но активировка катода проводится на этапе тренировки. Наоборот, мощные приборы, откачиваемые индивидуально или малыми партиями, проходят при этом полную электрическую обработку, включая активировку катода и первичную тренировку под рабочими напряжениями.

Дискретные полупроводниковые приборы имеют два общих признака, характерных для прибора любого типа и любого функционального назначения:

- *монолитный характер конструкции* прибора, делающий невозможным разделение его на отдельные элементы; например, эмиттер, база и коллектор транзистора являются неразделимыми областями одного и того же полупроводникового кристалла;
- *интегрально-групповой характер технологии*, порожденный монолитностью конструкции прибора, позволяет проведение определенной группы технологических воздействий интегрально (по всей поверхности пластины) для совместного изготовления совокупности однотипных элементов, например, создание области эмиттера сразу для всех приборов на пластине.

Дело в том, что активные элементы дискретных полупроводниковых приборов (ДПП) изготавливают *не индивидуально* для каждого прибора по отдельности, а одновременно для всей партии приборов в виде совокупности однотипных элементов, размещенных на одной пластине (подложке). В этом смысле процесс их изготовления по сути не отличается от изготовления интегральных микросхем (ИМС), о которых речь пойдет ниже.

Различие между производством дискретных приборов и интегральных схем заключается лишь в том, что идентично повторенным элементом на пластине в первом случае является активный элемент дискретного прибора, а во втором случае — полностью интегральная микросхема, называемая «чипом» (от англ. *chip*). По этой причине, описывая ниже технологические особенности

полупроводниковой электроники, будем применять их с равным правом как для ДПП, так и для ИМС.

В обоих случаях технологическую основу производства составляет так называемый **планарный процесс**¹⁾, под которым понимается совокупность технологических воздействий на материал (чаще всего кремний), осуществляемых с одной стороны исходной полупроводниковой пластины (называемой подложкой). Планарный процесс дает возможность одновременного изготовления в едином технологическом цикле большого числа идентичных ДПП или ИМС (до нескольких сотен и даже тысяч на одной пластине). Групповая обработка пластин (до 100–200 штук в одной партии) обеспечивает хорошую воспроизводимость параметров и высокую производительность при существенном снижении себестоимости изделий. Благодаря этим особенностям планарная технология занимает доминирующее положение в современном полупроводниковом производстве.

По результату воздействия на материал все методы полупроводниковой технологии можно разделить на четыре группы.

I группа — методы удаления материала с использованием не только обычных жидкостных травителей (кислот и щелочей), но и «сухого травления», основанного на применении реактивных парогазовых смесей (газовое травление) и низкотемпературной плазмы (ионно-плазменное травление) — см. пп. 4.12 и 4.17.

II группа — методы осаждения материала на поверхность полупроводниковой пластины, такие как:

1) *нанесение металлических пленок* методами резистивного, электронно-лучевого и лазерного испарения, либо катодным и ионно-плазменным распылением металлов, выполняющих функции: а) *контактного сплава* для омических контактов, б) *высокоомного сплава* для резисторов, в) *низкоомного сплава* для токоведущих дорожек (межсоединений), обкладок конденсаторов и затворов МДП- или МОП-структур;

2) *создание диэлектрических (окисных) пленок* различными методами (например, окислением кремния), выполняющих следующие функции: а) *активная функция* подзатворного диэлектрика в МДП-приборах и изолятора элементов схемы; б) *пассивная функция* защиты поверхности приборов от внешних воздействий; в) *технологическая функция* маскирующего покрытия, селективно защищающего поверхность полупроводника от тех или

¹⁾ Первая интегральная микросхема на кремнии была создана Робертом Нойсом (см. сноску на с. 21) по планарной технологии, идею которой предложил в 1958 году Джин Хорни (Jean Hoerni), сотрудник Fairchild Semiconductor.

иных технологических воздействий, например, при локальном диффузионном или ионном легировании, а также при формировании контактных площадок и межсоединений — см. п. 4.10;

3) *осаждение монокристаллических пленок* методами жидкофазной, газофазной и молекулярно-лучевой эпитаксии¹⁾ — см. пп. 4.13–4.15 и 5.15.

III группа — методы модификации свойств материала в результате диффузионного легирования или ионной имплантации — см. пп. 4.7 и 4.8.

IV группа — методы локальной микрообработки, называемые *методами литографии*, такие как фотолитография (ФЛ), электронная литография (ЭЛ), ионная литография (ИЛ) и рентгеновская литография (РЛ).

Сущность методов литографии состоит в формировании на поверхности пластины элементов прибора или рисунка схемы с помощью специального химически стойкого органического материала — *резиста*, чувствительного к облучению ультрафиолетом (фоторезист при ФЛ), пучком ускоренных электронов (электронорезист при ЭЛ), пучком ускоренных ионов (ионорезист при ИЛ) или рентгеновским излучением (рентгенорезист при РЛ). Экспонирование выполняется через специальную маску — *фотошаблон* (ФС) с изображением элементов прибора или рисунка схемы, селективно пропускающим облучающий поток. При фотолитографии используют обычный ультрафиолет (УФ, $\lambda = 300$ – 400 нм) или глубокий ультрафиолет (ГУФ, $\lambda = 150$ – 300 нм).

Облученные участки резиста либо разрушаются (*позитивные фоторезисты*, повторяющие рисунок ФС), либо полимеризуются (*негативные фоторезисты*, создающие негатив рисунка ФС), формируя скрытое изображение. После удаления соответствующих участков резиста защищенная им поверхность оказывается устойчивой к воздействию химических травителей. Это позволяет сформировать позитивное или негативное изображение элементов прибора или рисунка схемы на поверхности пластины.

Интегрально-групповая планарная технология в применении к изготовлению ДПП (например, биполярных транзисторов) позволяет установить последовательность технологических операций, приведенную на рис. 4 и отчасти применимую к ИМС.

¹⁾ Термин «эпитаксия», введенный в 1928 году Л. Руайе (L. Royer), происходит от греческих слов — *эпи* «на» и *таксис* «расположение в порядке» — что означает ориентированное наращивание на поверхности монокристаллической подложки нового слоя, который при своем росте продолжает кристаллографическую ориентацию подложки.

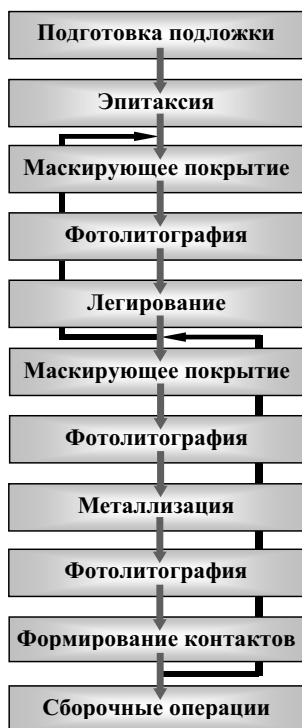


Рис. 4. Технологический маршрут изготовления ДПП

Операции механической и химической обработки — для подготовки поверхности подложки к эпитаксиальному наращиванию монокристаллической пленки или к планарному процессу изготовления приборов и интегральных схем — см. пп. 4.12 и 4.17.

Операции эпитаксиального роста — для формирования на поверхности подложки полупроводниковых слоев с необходимыми электрофизическими свойствами — см. пп. 4.13–4.15 и 5.15.

Операции формирования маскирующего покрытия (диоксида кремния) — для проведения локальных технологических воздействий на поверхность полупроводника — см. п. 4.10.

Фотолитографические операции — для формирования последовательного набора совмещенных изображений на поверхности пластины, создающих совокупность повторяющихся дискретных приборов или топологию интегральной микросхемы.

Операции диффузионного или ионного легирования — для создания различных областей структуры, например, p – n -переходов, диффузионных резисторов, конденсаторов и изолирующих участков в ИМС — см. пп. 4.7 и 4.8.

Операции формирования контактов — для подвода напряжения к активным областям прибора, а также для создания разветвленной цепи межсоединений в интегральных микросхемах.

Сборочные операции — для разделения пластины на чипы путем скрайбирования (от англ. *scribe*) алмазной иглой или лазерным лучом, монтажа чипа на кристаллодержатель с помощью пайки или ультразвуковой сварки, присоединения выводов к контактам кристалла и траверсам корпуса, защитного покрытия кристалла компаундом с окончательной герметизацией корпуса прибора.

Двумя черными стрелками на рис. 4 отражены циклы повторных операций. Так, левой стрелкой обозначено двойное проведение цикла, состоящего из трех последовательных операций:

а) термическое окисление кремния (для создания слоя SiO_2); б) фотолитография (для создания окон в SiO_2); в) локальное введение легирующих примесей в окна для формирования сначала базовой области, а затем внутри нее эмиттерной области (коллекторной областью транзистора является изначальная пленка кремния, эпитаксиально выращенная на поверхности подложки).

После третьего цикла окисление–фотолитография во вновь образованном слое SiO_2 вскрывают окна к областям коллектора, эмиттера, базы и на всю поверхность пластины наносят металлизацию (обычно вакуумным напылением алюминия, имеющего хорошую адгезию как к Si, так и к SiO_2). Четвертой фотолитографией по слою Al (для удаления лишнего металла) и последующим отжигом (при температуре 500–550°C) формируют омические контакты к рабочим областям транзистора, проводящие дорожки и контактные площадки. Следом за этим идут сборочные операции, завершающие производство дискретного транзистора.

Правая стрелка цикла повторных операций на рис. 4 характерна для интегральных микросхем, в которых разветвленная система металлизации выполняется в несколько слоев, изолированных друг от друга чаще диэлектрической пленкой SiO_2 . Кроме диэлектрической изоляции, в ИМС на биполярных транзисторах для их разделения также используют обратно-смещенные p – n -переходы. В этом случае в подложке, например, из кремния с проводимостью p -типа, создают n -области («карманы»), внутри которых формируют необходимые активные или пассивные элементы схемы. Электрический переход между «карманом» и подложкой поддерживается в работающей ИМС под обратным напряжением, что обеспечивает высокое сопротивление изоляции.

Фотолитография как один из этапов, показанных на рис. 4, включает следующую последовательность операций: нанесение фоторезиста, его сушка и экспонирование через фотошаблон с удалением ненужных мест и термическим задубливанием оставшегося резиста (что создает окна, через которые локально удаляются химическим травлением диэлектрик SiO_2 или металл Al) и, наконец, полное удаление фоторезиста с пластины. В результате этого на поверхностном слое пластины оказывается сформированный рисунок, перенесенный (позитивно или негативно) с фотошаблона, — либо маска из SiO_2 , либо электрические контакты и проводящие дорожки из Al.

При многократном применении фотолитографии важную роль играет операция совмещения отдельных фотошаблонов, соответствующих активным элементам транзистора, участкам изоляции, контактными площадкам, рисункам металлизации и т. д.

МИКРОЭЛЕКТРОНИКА: кремниевая технология процессоров на пути от «микро» к «нано»

Технологическое продвижение микроэлектроники наиболее наглядно демонстрирует процессорная техника, достижения которой определяют прогресс компьютерной индустрии в целом.

Ныне первенство в производстве процессоров держат три компьютерных гиганта — Intel, AMD и IBM, но лидирующие позиции занимает Intel, чья продукция на рынке процессоров составляет около 75 %. Поэтому ниже будем ориентироваться на достижения этой корпорации, которая в настоящее время является законодателем «технологической моды» в процессорной технике.

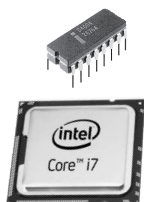


Рис. 5. Процессоры Intel

Днем рождения микропроцессора принято считать 15 ноября 1971 года, когда фирма Intel представила на суд общественности первый в мире процессор Intel 4004¹⁾, названный Гордоном Муром²⁾ «одним из самых революционных продуктов в истории человечества». Внешний вид этого микропроцессора показан на рис. 5, где для сравнения ниже приведен современный четырехъядерный процессор Intel Core i7 Extreme, выпущенный в 2008 году.

За прошедшие годы процессорная технология совершила гигантское продвижение: если первый процессор имел 2300 транзисторов с тактовой частотой 740 кГц и около 70 инструкций, то последний содержит 731 млн транзисторов на чипе при тактовой частоте 3,33 ГГц и свыше 700 инструкций. В основу технологического прогресса положено решение трех задач:

- увеличение степени интеграции (числа транзисторов на чипе),
- повышение быстродействия (тактовой частоты) процессоров,
- снижение стоимости производства.

Главным конструкторско-технологическим направлением для *одновременного* решения всех трех задач явилось уменьшение

¹⁾ Изобретателем первого микропроцессора является Тед Хофф (Ted Hoff), идею которого реализовали инженеры фирмы Intel Федерико Феджин (Federico Faggin) и Стан Мейзор (Stan Mazor), чьи имена в 1996 году были занесены в Национальный зал славы изобретателей США.

²⁾ Мур (Moore) Гордон Эрл (род. 1929) — американский инженер, совместно с Робертом Нойсом (см. сноску на с. 21) основал (1968) компанию Intel (сокращение от англ. *Integrated Electronics*), автор «закона Мура» (см. рис. 9), эмпирического правила, ставшего основополагающим принципом развития процессорной индустрии. Ныне Г. Мур является почетным председателем совета директоров корпорации Intel и занимает, по данным Форбс за 2008 год, 288-ю позицию в списке самых богатых людей планеты с состоянием \$ 3,7 млрд.

размеров активных элементов интегральных микросхем, составляющих схемотехническую базу цифровых устройств. Действительно, с уменьшением размеров возрастает скорость переключения нуль/единица (тактовая частота) электронного ключа, а также возрастает плотность размещения элементов, что увеличивает их общее количество на пластине и снижает стоимость производства в расчете на один элемент микросхемы¹⁾.

В качестве базового активного элемента современных процессоров на протяжении уже многих лет традиционно используют комплементарную МОП-ячейку (КМОП или англ. CMOS, сокращение от *Complementary Metal-Oxide-Semiconductor*). Такая ячейка состоит из двух полевых транзисторов с индуцированным (инверсионным) каналом — n -типа (n -МОП или англ. NMOS) и p -типа (p -МОП или англ. PMOS), изготовленных на кремнии методами планарной полупроводниковой технологии.

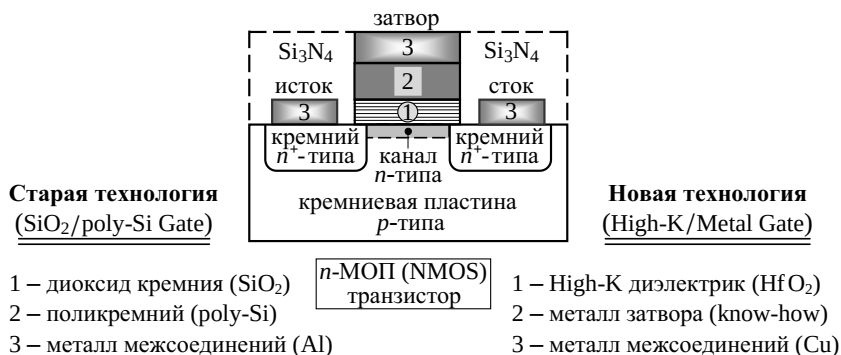


Рис. 6. Схематическое изображение структуры n -МОП транзистора с указанием принципиальных особенностей, присущих старой и новой технологиям: 1 — подзатворный диэлектрик, 2 — электрод затвора, 3 — контактный металл межэлементных соединений

Структура n -МОП транзистора, созданного на пластине кремния p -типа, изображена на рис. 6. Граница токового канала между истоком и стоком (в форме сильно легированных n^+ -областей)

¹⁾ Занимательные цифры: 1. Стоимость одного транзистора в составе современного процессора примерно в миллион раз ниже аналогичной стоимости для ИМС в 1968 году, так что при подобных темпах снижения цен сегодня автомобиль стоил бы около одного цента. 2. Размер 45-нанометрового транзистора (используемого в процессоре Intel Core i7) примерно в 2000 раз меньше толщины человеческого волоса, так что около сотни таких приборов размещается на диаметре красной кровяной клетки. 3. Ежегодно все фирмы выпускают такое количество процессоров, которое содержит около 10^{19} транзисторов, что примерно в 100 раз превышает всю популяцию муравьев на планете.

показана штриховой линией, поскольку проводящий канал n -типа возникает только в режиме инверсии при положительном напряжении на затворе. Комплементарный p -МОП транзистор с каналом, истоком и стоком p -типа (не показанный на рис. 6) создают внутри n -Si кармана, имплантированного в исходную p -Si пластину. Ионная имплантация применяется также при локальном формировании областей истока и стока для n -МОП и p -МОП транзисторов в составе КМОП-ячейки.

На рис. 6 указаны принципиальные особенности старой (традиционной) технологии и новой (модифицированной за последние годы) технологии. В кремниевой КМОП-технологии в качестве диэлектрика затвора (1) и металла межсоединений (3) традиционно (со времен Роберта Нойса, см. сноску на с. 21) использовали SiO_2 и Al. Роль проводящего электрода затвора (2) с конца 60-х годов до последнего времени исполнял сильно легированный поликристаллический кремний (poly-Si). При этом для n -канала применяли n^+ -поликремний с работой выхода около 4,15 эВ, а для p -канала — p^+ -поликремний с работой выхода около 5,15 эВ (напоминаем, что ширина запрещенной зоны и электронное сродство кремния при 300 К, соответственно, равны $\Delta E_g = 1,12$ эВ и $\chi = 4,05$ эВ, см. п. 3.5). Приведенные численные значения обеспечивают минимальное пороговое напряжение переключения нуль/единица для КМОП-ячейки (определяемое началом перехода от режима обеднения к режиму инверсии в n -МОП и p -МОП транзисторах).

Необходимость модификации традиционной кремниевой технологии проявилась около десяти лет назад в ходе уменьшения размеров КМОП-ячейки с целью решения вышеуказанных задач.

В основу уменьшения размеров положен *принцип масштабирования*, сформулированный еще в 1972 году: с уменьшением линейных размеров в α раз во столько же раз уменьшаются напряжение и ток, но увеличивается концентрация легирующей примеси при сохранении неизменным распределения электрического поля в приборе. Как следствие этого, быстродействие повышается в α раз (в результате снижения емкости межсоединений и приборных структур при неизменной крутизне прибора), а мощность, рассеиваемая одной ячейкой, уменьшается в α^2 раз (из-за одновременного уменьшения напряжения и тока в α раз) при возрастании плотности размещения ячеек пропорционально α^2 (в результате уменьшения размеров межсоединений и топологических размеров приборных структур). Два последних фактора обеспечивают неизменное энергопотребление одного чипа при росте числа транзисторов на кристалле постоянной площади.

Однако с уменьшением длины токового канала (между истоком и стоком) и толщины подзатворного диэлектрика (между электродом затвора и каналом) возрастают токи утечки (продольный и поперечный). Именно это явилось принципиальным ограничением масштабированного уменьшения размеров КМОП-транзисторов, с чем и столкнулись в 90-х годах прошлого века.

Основные изменения конструктивно-технологических параметров, вводимые в производство процессоров компанией Intel, приведены на рис. 7 в форме таблицы, показывающей характерные особенности производственного техпроцесса с временным сдвигом в два года, начиная с 1997 г.

Кодовый номер техпроцесса	P856	P858	Px60	P1262	P1264	P1266	P1268	P1270
год ввода в производство	1997	1999	2001	2003	2005	2007	2009	2011
проектная норма техпроцесса, нм	250	180	130	90	65	45	32	22
длина волны УФ лазера, нм	248	248	193	193	193 ¹⁾	193 ¹⁾	193 ²⁾	193 ³⁾
материал межсоединений	Al	Al	Cu	Cu	Cu	Cu	Cu	Cu
диаметр пластины, мм	200	200	200/300	300	300	300	300	300
материал канала	Si	Si	Si	str-Si	str-Si	str-Si	str-Si	str-Si
диэлектрик затвора	SiO ₂	SiO ₂	SiO ₂	SiO ₂	SiO ₂	High-K	High-K	High-K
электрод затвора	poly-Si	poly-Si	poly-Si	poly-Si	poly-Si	metal	metal	metal

1) сухая DUV-литография с фазовым сдвигом,

2) иммерсионная DUV-литография с фазовым сдвигом,

3) иммерсионная DUV-литография с фазовым сдвигом и двойной экспозицией.

Рис. 7. Конструктивно-технологические особенности производственного изготовления процессоров корпорацией Intel

Базовой характеристикой технологического процесса является *проектная норма*, под которой понимается минимальный топологический размер микросхемы, формируемый с помощью ультрафиолетового (УФ) освещения фоторезиста через фотошаблон при проекционной литографии. Из третьей строки таблицы видно, что за десять лет проектная норма уменьшилась более, чем в 5 раз (с 250 нм до 45 нм). В конце 2009 г. выпущены 32-нанометровые процессоры с переходом на технологию 22-нм в 2011 г. При этом, как показывают эмпирические данные, длина канала

составляет около 0,7–0,8 от проектной нормы, а толщина подзатворного диэлектрика — 0,020–0,025 от длины канала.

Среди технологических достижений, которые были внедрены компанией Intel в производственный процесс (в том числе с целью снижения проектной нормы) следует отметить принципиальные нововведения, выделенные в таблице на рис. 7 полужирным шрифтом на сером фоне. Отсюда видно, что, начиная с 2001 года, изменения в производственный процесс вносились каждые два года. Ниже кратко рассмотрим особенности новой технологии.

1. Использование лазерной литографии на основе глубокого УФ-излучения (DUV-литография, от англ. *Deep Ultra Violet*). Со времени изобретения фотолитографической технологии в начале 70-х годов прошлого столетия ее усовершенствование шло в направлении повышения разрешающей способности, в первую очередь, путем уменьшения длины волны излучения. С конца 80-х годов для освещения фотошаблонов стали применять эксимерные УФ-лазеры — сначала на основе молекул KrF с длиной волны $\lambda = 248$ нм (для 350-нм, 250-нм и 180-нм техпроцессов), а с 2000 года на основе молекул ArF с длиной волны $\lambda = 193$ нм (для 130-нм, 90-нм, 65-нм и 45-нм техпроцессов) (см. четвертую строку таблицы).

Разрешение линий на фоторезисте из-за оптической дифракции света обычно ограничивается величиной $\lambda/2$. Следовательно, ArF-лазер ($\lambda = 193$ нм) еще подходит для 90-нм техпроцесса, но уже не может обеспечить проектную норму 65 нм и менее.

Поэтому для повышения разрешающей способности были использованы *фотошаблоны с фазовым сдвигом*, принцип действия которых поясняет рис. 8. В таком шаблоне на одно из двух соседних окон дополнительно накладывается фазовый фильтр, сдвигающий фазу проходящей волны на 180° . В результате интерференции соседних противофазных полей происходит их взаимное ослабление в области между окнами, как показано на рис. 8 б, в сравнении с рис. 8 а. Фазовый сдвиг позволяет повысить разрешение в полтора раза, т. е. делает различимыми линии шириною до $\lambda/3$. Следовательно, DUV-литография с $\lambda = 193$ нм становится применимой и к 65-нм техпроцессу.

До сих пор речь шла о «сухой» проекционной литографии, переносящей топологический рисунок с фотошаблона через воздушную среду путем фокусирования его на поверхности фоторезиста (подложки). Для этой цели были разработаны специальные проекционные установки шагового мультиплицирования (так называемые степперы, от англ. *step*). Повышение разрешения таких

установок оказалось возможным при использовании *иммерсионной литографии*. В этом случае рисунок шаблона переносится не на сухую подложку, а на ее поверхность, покрытую слоем специальной иммерсионной жидкости (чаще воды, очищенной от загрязнений) с высоким коэффициентом преломления n . Длина электромагнитной волны в жидкости укорачивается в n раз, т. е. $\lambda_{\text{ж}} = \lambda/n$. Теперь оптическое разрешение при использовании фотомаски с фазовым сдвигом ограничивается величиной $\lambda_{\text{ж}}/3$. Таким образом, иммерсионная литография с фазовым сдвигом делает 193-нм лазер применимым даже к 32-нм техпроцессу.

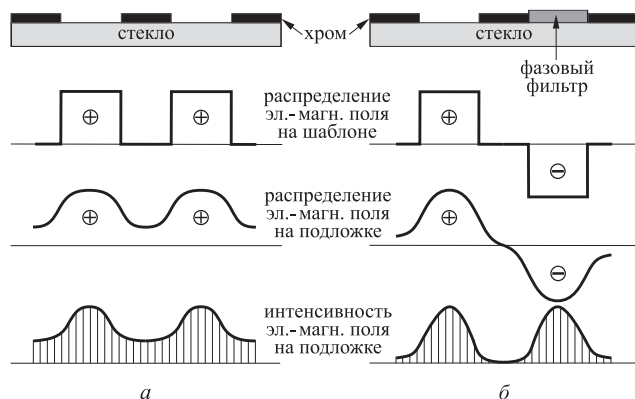


Рис. 8. Качественные картины распределения электромагнитного поля на поверхности фотомаски, подложки и его интенсивности, воздействующей на фоторезист: *а* — обычный фотомаска, *б* — фотомаска с фазовым сдвигом

Сокращение топологической проектной нормы до 22 нм, запланированное на 2011 год (см. третью строку таблицы), оказалось возможным при использовании иммерсионной 193-нм литографии с фазовым сдвигом в сочетании с технологией *двойной экспозиции*. Здесь вместо одного шаблона применяют два, тем самым удваивая число экспозиций фоторезиста. При этом усложняется процедура совмещения шаблонов, однако решаемая современными степперами, обеспечивающими точность до 1 нм.

Следовательно, двойная экспозиция, повышая разрешающую способность литографии, увеличивает стоимость техпроцесса за счет удвоения числа шаблонов. Однако при этом сохраняется в качестве источника УФ-излучения 193-нм лазер. Это существенно снижает затраты производства, возникающие при переходе на более коротковолновое излучение. В будущем намечен переход с DUV-литографии на EUV-литографию (сокращение от англ. *Extreme Ultra Violet* — сверхжесткий УФ) с длиной волны 13 нм.

2. Замена алюминия медью в качестве металла межсоединений (пятая строка таблицы). При формировании многоуровневой архитектуры (до 6–7 слоев металлизации) использование Си вместо Al снижает электрическое сопротивление проводящих дорожек и время задержки на межсоединениях. Это не только уменьшает тепловыделение, но и увеличивает быстродействие. С той же целью для изоляции проводящих слоев применяют диэлектрики с малой диэлектрической проницаемостью (типа фтор-силикатного стекла), что снижает емкость между слоями и повышает скорость переключения.

3. Увеличение диаметра кремниевых пластин с 200 мм до 300 мм (шестая строка таблицы) позволило в два с лишним раза увеличить число процессоров, изготавливаемых на одной пластине, и тем самым снизить стоимость одного чипа.

4. Использование технологии «напряженного кремния» (англ. *Strained Silicon* или сокращение str-Si на рис. 7, седьмая строка таблицы) увеличивает подвижность носителей заряда в токовом канале с целью повышения тактовой частоты процессора. Задача увеличения подвижности дырок (в p -канале) и электронов (в n -канале) конструктивно решается по-разному для p -МОП и n -МОП транзисторов.

Поскольку движение дырки (как электронной вакансии в связях Si–Si) реализуется путем перескока электрона из одной соседней связи в другую (вакантную), то этот процесс облегчается при уменьшении расстояния между атомами путем сжатия кристаллической решетки в области между истоком и стоком. Такое сжатие канала в p -МОП транзисторе осуществляется путем замены кремния в истоке и стоке на твердый раствор $\text{Si}_{1-x}\text{Ge}_x$. При изменении величины x от 0 до 1 период решетки линейно возрастает от $a = 5,43 \text{ \AA}$ (Si) до $a = 5,66 \text{ \AA}$ (Ge). С увеличением объема $\text{Si}_{1-x}\text{Ge}_x$, занимаемого истоком и стоком, расположенный между ними p -канал испытывает напряжение сжатия и сокращается (на доли процента) в направлении протекания тока.

В n -МОП транзисторе, где ток в канале создается электронами проводимости, их движение ускоряется при увеличении расстояния между атомами кремния. Это достигается в результате внешнего покрытия транзистора нитридом кремния (Si_3N_4 , гексагональная решетка с параметрами $a = 7,76 \text{ \AA}$ и $c = 5,62 \text{ \AA}$ — см. п. 3.2), что схематически изображено верхней штриховой линией на рис. 6. При этом n -канал за счет механических напряжений, создаваемых нитридом кремния, растягивается (на доли процента) вдоль направления протекания тока.

По данным корпорации Intel, применение технологии «напряженного кремния» в обоих типах прибора (со сжатым каналом в p -МОП и с растянутым каналом в n -МОП) на 20–30% увеличивает скорость переключения КМОП-ячейки (за счет возрастания тока в канале) при повышении стоимости производства процессора всего на 2%.

5. Использование технологии High-K/Metal-Gate подразумевает замену SiO_2 материалом с более высокой диэлектрической проницаемостью (называемым в англоязычной литературе *High-K*, или *high-k*¹⁾ материалом) при одновременной замене поликремния (*poly-Si*) металлическим затвором (англ. *Metal-Gate*) (см. восьмую и девятую строки в таблице на рис. 7).

Быстрота переключения МОП-транзистора из непроводящего состояния в проводящее (нуль/единица) определяется скоростью индуцирования инверсионного заряда в канале: $dQ/dt = I$, т. е. величиной тока в канале. Этот заряд создается напряжением V на затворе и равняется $Q = CV$, где $C = \epsilon\epsilon_0/d$ — емкость затвора (на единицу площади) для подзатворного диэлектрика с проницаемостью ϵ и толщиной d . Следовательно, для повышения быстродействия путем увеличения тока канала, $I = CdV/dt$, надо в подзатворном диэлектрике увеличивать ϵ и уменьшать d . Однако толщина диэлектрика имеет нижний физический предел, связанный с экспоненциальным ростом туннельного тока утечки, что приводит к недопустимо большому энергопотреблению и тепловыделению. Поэтому единственной возможностью остается переход к материалам с большей диэлектрической проницаемостью.

С подобной ситуацией фирма Intel столкнулась при реализации техпроцесса P1264 с топологической нормой 65 нм, где диэлектрик затвора достиг предельно малой толщины (около 1,2 нм, что составляет примерно пять моноатомных слоев кремния). Понижение проектной нормы до 45 нм оказалось возможным лишь при замене диоксида кремния High-K диэлектриками, среди которых наиболее подходящим оказался диоксид гафния (HfO_2). Такая замена позволила в 45-нм транзисторе увеличить толщину подзатворного диэлектрика до 3 нм, что снизило ток утечки почти в десять раз, а общее энергопотребление на 30%.

Однако сохранение поликремния в качестве электрода затвора привело к неприемлемо высокому пороговому напряжению для

¹⁾ Следует отметить терминологическую путаницу в этих обозначениях. На самом деле, следовало бы писать High- ϵ или High- κ , где ϵ и κ характеризуют диэлектрическую проницаемость и электрическую восприимчивость материала, при этом $\epsilon = 1 + \kappa$ в системе СИ.

пары $\text{HfO}_2/\text{poly-Si}$. Поэтому большие усилия были направлены на поиск металлов для замены поликристаллического кремния. Такой поиск увенчался успехом, однако корпорация Intel не раскрывает секреты «know-how». В литературе упоминается силицид никеля (NiSi), нитрид тантала (TaN) и другие металлические соединения с работой выхода, близкой к 4,2 эВ (для n -МОП транзисторов) и к 5,1 эВ (для p -МОП транзисторов).

Использование технологии High-K/Metal-Gate в 45-нм процессорах сократило время переключения транзисторов более, чем на 20%, повысив суммарную производительность компьютеров. По мнению специалистов, производственное внедрение технологии High-K/Metal-Gate по значимости для компьютерной индустрии сравнимо в хирургии с операцией по пересадке сердца.

Сегодня по 45-нм производственной технологии выпускают одноядерные процессоры Intel Atom, двухъядерные процессоры Intel Core 2 Duo, четырехъядерные процессоры Intel Core i7 и шестиядерные процессоры Intel Xeon серии 7500. В ближайшем будущем — переход на 32-нм и 22-нм проектные производственные нормы, в основу которых, как видно из таблицы рис. 7, положены описанные выше технологические нововведения последних лет.

Следует отметить, что именно эти нововведения смогли существенным образом продлить жизнь известному **закону Мура** — эмпирическому правилу, сформулированному еще в 1965 году (через шесть лет после изобретения интегральной схемы) Гордоном Муром (см. сноску на с. 32), согласно которому *число транзисторов на чипе будет удваиваться через каждые два года*.

Закон Мура и его практическая реализация отображены графически на рис. 9 (левая вертикальная ось) сплошной линией и кружками с указанием числа транзисторов на чипе для Intel процессоров конкретного типа. Одновременно с ростом степени интеграции происходило увеличение быстродействия процессоров: Intel 4004 — 740 кГц, Intel 8086 — (4,7–10) МГц, Intel 486 — (25–100) МГц, Pentium 4 — (1,4–2,0) ГГц, современные многоядерные процессоры типа Intel Core — разгон до 4,0 ГГц.

На рис. 9 (правая вертикальная ось) также показано, как в процессе производственных разработок фирма Intel изменяла проектную норму изготовления транзисторов, что отмечено треугольниками с аппроксимацией штриховой линией. Отсюда видно, что темп снижения проектной нормы, заложенный в начальные годы, был продолжен вплоть до настоящего времени, правда, со сдвигом на 5–6 лет (разрыв штриховой линии с 1985 г.). Этот временной сдвиг был связан с трудностями перехода с микронной фотолитографии на субмикронную. Аналогично, возникшая

в конце 90-х годов задержка на 2–3 года с производственным выпуском процессоров Pentium II и Pentium III, по-видимому, вызвана промышленным освоением DUV-литографии ($\lambda = 365$ нм).

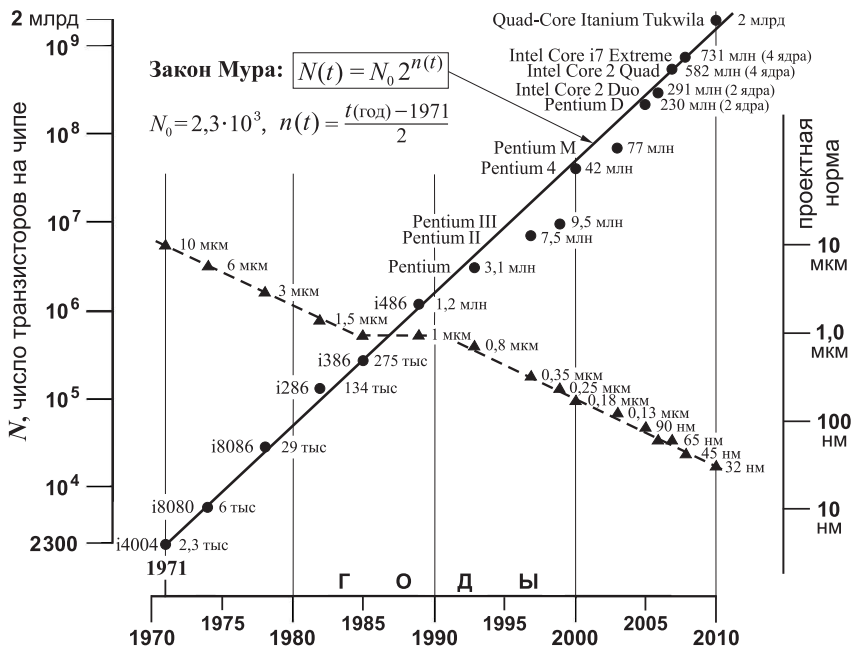


Рис. 9. Эмпирический закон Мура (сплошная линия и кружки) и снижение топологической проектной нормы (штриховая линия и треугольники) в производстве процессоров корпорацией Intel

По современным оценкам экспертов, действие закона Мура может быть продлено еще примерно на десять лет. Предполагается, что МОП-транзисторы сохраняют традиционную конструкцию (в виде истока и стока с затвором между ними) при постепенном снижении проектной нормы до 16 нм (2013 г.), 11 нм (2015 г.) и 8 нм (2017 г.) с переходом на EUV-литографию ($\lambda = 13$ нм).

При этом базовым материалом останется кремний с некоторыми инновационными изменениями, такими как (ограничимся лишь перечислением): модернизация High-K и Metal-Gate материалов, использование технологии SOI (Silicon-On-Insulator), применение так называемых Tri-Gate (трехзатворных) и Multi-Channel Tri-Gate (многоканальных трехзатворных) конструкций, а также кремниевых нанонитей. Дальнейшее продвижение процессорной техники связывают с использованием новых наноматериалов типа фуллеренов, графенов и углеродных нанотрубок.

НАНОЭЛЕКТРОНИКА:

новые материалы, технологии и диагностика

Современная нанoeлектроника базируется на «трех китах»:
наноматериалы, нанотехнологии, нанодиагностика.

I. Краткая характеристика наноматериалов

Наноматериалы, синтезированные и изученные к настоящему времени многочисленными исследователями и сведения о которых опубликованы в современной литературе, довольно многообразны не только по размерам наночастиц и количеству атомно-молекулярных составляющих (см. рис. 1), но и по назначению, способам получения и применения материалов. Согласно рекомендации 7-й Международной конференции по нанотехнологиям (Висбаден, 2004 г.), выделяют следующие типы наноматериалов (см. также текст после рис. 1 и вводную часть гл. 5):

- *нанокластеры и наночастицы* – нульмерные (0D) нанообъекты с размерами по трем направлениям, не превышающими 10 нм и 100 нм, соответственно;
- *нанотрубки, нановолокна, нанопроволоки, нанонити* — одномерные (1D) нанообъекты с поперечными размерами по двум направлениям, не превышающими 100 нм;
- *нанопленки, нанослои, приповерхностные структуры* — двумерные (2D) нанообъекты с одним размером (по толщине), не превышающим 100 нм;
- *нанодисперсии* — коллоидные системы, в которых частицы дисперсной фазы имеют размеры менее 100 нм;
- *молекулярные наноструктуры (или супрамолекулярные структуры)* — нанообъекты, получаемые в результате так называемого нековалентного синтеза с образованием слабых связей (типа ван-дер-ваальсовой и водородной) между молекулами (или их ансамблями);
- *консолидированные наноматериалы* — наноструктурированные материалы (нанокомпозиты) со структурными элементами нанометрового размера, разупорядоченно расположенными в композиционной матрице;
- *нанокристаллические материалы* — наноструктурированные материалы с упорядоченным расположением структурных элементов (нанокристаллитов);
- *нанопористые структуры* — пористые вещества с нанометровым размером пор.

Особое место среди пористых веществ занимают так называемые **фрактальные структуры** (от лат. *fractus* «ломаный», «дробный»). По определению Б. Мандельброта, основоположника фрактальной геометрии, фрактал — «это самоподобная структура, чье изображение не зависит от масштаба». Свойство самоподобия фрактальной структуры означает, что разные равнообъемные части структуры содержат в среднем одинаковое число первичных частиц и одинаково выглядят на разных масштабах. Небольшая часть фрактала содержит информацию о фрактале в целом, который может включать в себя повторяющиеся исходные кластеры (с уменьшением) бесконечное число раз. Образование фрактальных кластеров сопровождается уменьшением средней плотности вещества за счет увеличения объема пустот (пор).

Не вдаваясь в детали фрактальной геометрии, отметим лишь, что геометрическая размерность (которая для компактных сред равняется евклидовой размерности: $D = 1, 2, 3$ для одномерных, двумерных и трехмерных объектов, соответственно) в случае фракталов является дробной. Например, фрактальные агрегаты в форме самоподобной ветвистой структуры, полученные путем конденсации твердого вещества в газовой среде, имеют фрактальную размерность $D = 1,7-1,9$ и достигают размеров в единицы микрон при максимальном числе частиц около 10^4 . Однако объемная доля твердого вещества составляет $10^{-2} - 10^{-4}$, что свидетельствует об очень высокой пористости.

Другим примером фрактальных структур являются пористые твердые тела (например, пористый кремний). Здесь распределение пор по размерам описывается функцией $f(r) = \text{const}/r^{(3-D)}$, при этом 3 означает евклидову размерность трехмерного пространства, а $D < 3$ — фрактальную размерность пористости, подлежащую экспериментальному определению. Последняя характеризует адсорбционную способность пористого тела, так как влияет на суммарную площадь его внутренней поверхности. Для пористого кремния со средним размером пор $\bar{r} \approx 2$ нм, удельная площадь поверхности становится огромной, достигая $1000 \text{ м}^2/\text{см}^3$.

Среди разнообразных типов наноматериалов и наноструктур, перечисленных в вышеприведенной классификации, ограничимся ниже описанием только **новых углеродных материалов**, полученных за последние десятилетия, которые представляют наибольший интерес для электроники. Сюда относятся **фуллерены, фуллериты, фуллериды, графены и углеродные нанотрубки** (см. рис. 3.11), даты открытия которых с именами первооткрывателей были указаны в хронологическом списке на с. 22.

1. Фуллерены были открыты в 1985 году первыми среди новых аллотропных (от греческих слов — *аллос* «другой» и *тропос* «поворот», «изменение», см. сноску на с. 365) модификаций углерода. До этого времени были широко известны только две основные кристаллические структуры углерода (не считая аморфного углерода и менее распространенного *карбина*, линейного углеродного полимера) — *алмаз* и *графит*.

Структура алмаза представляет собой двояную гранецентрированную кубическую решетку (ГЦК), показанную ниже на рис. 3.9 а. Здесь две плотноупакованные ГЦК-решетки смещены по отношению друг к другу на четверть пространственной диагонали. Поэтому каждый атом углерода окружен четырьмя соседями по углам тетраэдра и связан с ними сильными ковалентными связями, что и определяет высокую твердость алмаза. Период решетки равен $3,56 \text{ \AA}$, а расстояние между атомами $1,54 \text{ \AA}$.

Структура графита (см. ниже рис. 3.10) состоит из отдельных слоев, выложенных из шестиугольников, образующих гексагональную сетку со стороной $d_{\text{C-C}} = 1,42 \text{ \AA}$ (расстояние между атомами углерода, связанными ковалентной связью, более сильной, чем в алмазе, где $d_{\text{C-C}} = 1,54 \text{ \AA}$). Соседние слои удерживаются на расстоянии $3,35 \text{ \AA}$ более слабыми ван-дер-ваальсовыми силами (с периодом $6,7 \text{ \AA}$). Это обеспечивает взаимное скольжение слоев и, как результат, пластичные свойства графита.

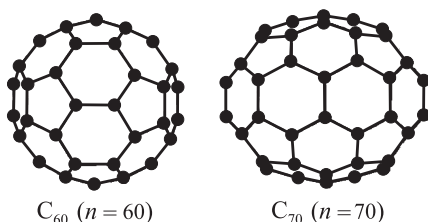
Возможность существования высокосимметричной молекулы углерода C_{60} (состоящей из 60 атомов и по форме напоминающей футбольный мяч) была теоретически предсказана за 15 лет до ее экспериментального обнаружения (Р. Смолли, Г. Крото, Р. Керл) в результате воздействия мощного лазерного импульса на графитовую мишень. Из возникающей при этом углеродной плазмы с температурой вплоть до 10^4 K путем продувки струей гелия был получен конденсат. В составе этого конденсата методом масс-спектрометрии были обнаружены два типа молекул фуллерена — C_{60} (основной пик с массой $60 \cdot 12 = 720 \text{ а.е.м.}$) и C_{70} (менее интенсивный пик с массой $70 \cdot 12 = 840 \text{ а.е.м.}^1$).

Структура молекул C_{60} и C_{70} показана на рис. 10 в форме замкнутых сфероидных поверхностей, названных *фуллеренами*²).

¹) Атомная единица массы (а.е.м.) равна $1,66 \cdot 10^{-27} \text{ кг}$.

²) Молекулы фуллерена получили название от фамилии американского архитектора Р.Б. Фуллера (R. B. Fuller), построившего на выставке в Монреале (1967 г.) купол павильона США в форме сочлененных пентагонов (правильных пятиугольников) и гексагонов (правильных шестиугольников).

Здесь также приведены две формулы Эйлера¹⁾, позволяющие выяснить геометрические особенности выпуклых многогранников с вершинами (числом n), расположенными на поверхности сферы радиуса R , в том числе для плоской поверхности при $R \rightarrow \infty$.



Формулы Эйлера
для многогранников с n вершинами

- ① $\sum_m N_m (6-m) = 12s$
 $s = 0$ — на плоскости ($N_5 = 0$)
 $s = 1$ — на сфере ($N_5 = 12$)
- ② $n = 20 + 2N_6$ — на сфере

Рис. 10. Структура молекул фуллерена C_{60} и C_{70} . Формулы Эйлера для выпуклого многогранника связывают число m -угольников (N_m) — пентагонов ($m = 5$) и гексагонов ($m = 6$), образующих его грани, — с числом его вершин (n)

Прежде всего, из формулы ① при $s = 0$ следует, что плоская поверхность может быть выложена плотно прилегающими гексагонами ($m = 6$), что полностью исключено для пентагонов ($N_m = 0$ при $m = 5$). Последнее объясняет отсутствие оси симметрии 5-го порядка в кристаллических структурах (см. с. 343).

Иная ситуация возникает на сферической поверхности, для которой из формулы ① при $s = 1$ следует, что все многогранники, вписанные своими вершинами внутрь сферы, имеют одинаковое число поверхностных пентагонов ($N_5 = 12 \equiv \text{const}$), при этом число гексагонов изменяется, начиная с $N_6 = 0$. Тогда формула ② дает число вершин многогранника (n), соответствующих данному значению N_6 , что определяет формулу молекулы фуллерена C_n . Это положено в основу классификации фуллеренов:

низшие фуллерены : $C_{20} (N_6=0)$, $C_{24} (N_6=2)$, $C_{28} (N_6=4)$,
 $C_{32} (N_6=6)$, $C_{36} (N_6=8)$;

средние фуллерены : $C_{50} (N_6=15)$, $C_{60} (N_6=20)$, $C_{70} (N_6=25)$;

высшие фуллерены : $C_{76} (N_6=28)$, $C_{84} (N_6=32)$, $C_{90} (N_6=35)$;
 $C_{94} (N_6=37)$, $C_{110} (N_6=45)$;

гиперфуллерены : C_{240} , C_{540} , C_{960} и далее.

¹⁾ Эйлер (Euler) Леонард (1707–1783) — российско-швейцарский математик, внесший значительный вклад в развитие математики, механики, физики, астрономии, теории музыки и ряда прикладных наук, включая картографию, баллистику, кораблестроение и др. Прожив почти половину жизни в России, Эйлер во многом определил становление и развитие российской науки.

Простейший фуллерен C_{20} имеет молекулу в форме *додекаэдра* (от греческих слов — *додека* «двенадцать» и *эдра* «грань»). Додекаэдр, состоящий из 12 пентагонов (без гексагонов), имеет форму, изображенную на рис. 11 а. Во всех других фуллеренах с номером $n > 20$ появляются гексагональные грани, при этом молекулы с нечетным числом таких граней, например C_{50} , C_{70} и др., имеют сфероидальную форму. Так, молекула C_{70} получается из C_{60} (см. рис. 10) введением пояса из десяти атомов углерода в экваториальную область сферы с последующим ее растяжением.

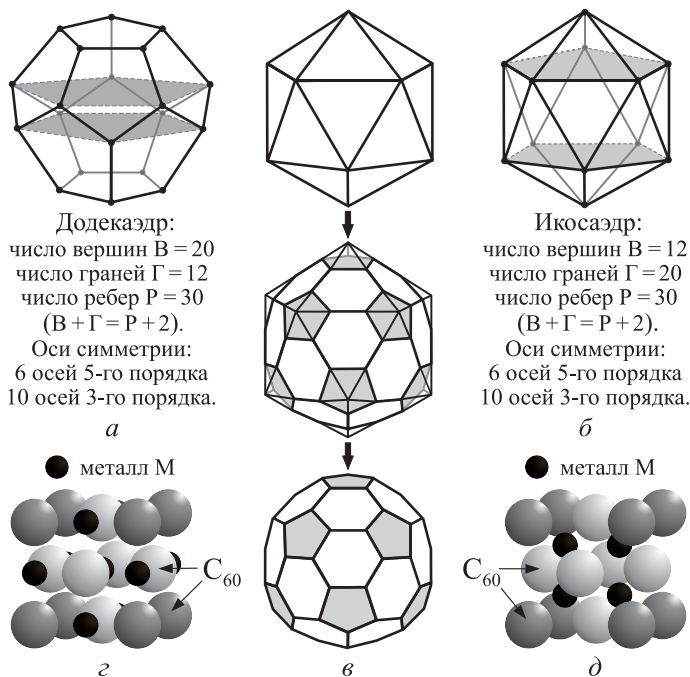


Рис. 11. Форма и геометрические особенности додекаэдра (а) и икосаэдра (б); трансформация икосаэдра в усеченный икосаэдр, моделирующий молекулу фуллерена C_{60} (в); строение фуллеридов щелочных металлов (металлофуллеренов), описываемых химической формулой MC_{60} (г) и M_2C_{60} (д)

Среди правильных многогранников (называемых телами Платона), к которым относятся тетраэдр, гексаэдр (куб), октаэдр, додекаэдр и икосаэдр (см. сноску на с. 354 в отношении греческой терминологии), додекаэдр имеет самую высокую вращательную симметрию. Такой же симметрией обладает и *икосаэдр* (от греческих слов — *икоса* «двадцать» и *эдра* «грань»), составленный из 20 правильных треугольников, как изображено на рис. 11 б.

Специфической особенностью обоих многогранников (додекаэдра и икосаэдра) является наличие осей симметрии 5-го порядка, которые запрещены правилами кристаллографии (см. п. 3.2). Вращательную симметрию 5-го порядка, характерную в том числе и для квазикристаллов, принято называть *икосаэдрической симметрией* (более детальное обсуждение см. в п. 5.2).

Как видно из рис. 11 а и б, и додекаэдр, и икосаэдр построены на основе двух базовых пятиугольников (выделенных серым цветом), которые расположены параллельно друг другу и взаимно развернуты на угол 36 градусов, так что вершины одного находятся посередине сторон другого. Вдоль сторон этих базовых пятиугольников пристраиваются пентагоны для додекаэдра (рис. 11 а) и правильные треугольники для икосаэдра (рис. 11 б). В первом случае каждая пятерка пентагонов замкнута таким же пентагоном сверху и снизу, завершая конструкцию додекаэдра. Во втором случае вершины каждой пятерки треугольников сведены вместе наверху и внизу, а для завершения конструкции икосаэдра вершины верхнего и нижнего базовых пентагонов соединены десятью линиями, создающими десять треугольников.

На рис. 11 в изображен переход от икосаэдра к усеченному икосаэдру, который моделирует молекулу фуллерена C_{60} . Для этого ребро исходного икосаэдра делим на три равные части и через точки, ближайšie к каждой из 12 вершин, проводим плоскости, отсекающие пятиугольные пирамиды (основания которых в форме пентагона затемнены как на среднем, так и на нижнем рисунке в). Таким способом 20 правильных треугольников икосаэдра превращаются в 20 гексагонов усеченного икосаэдра, имеющего 60 вершин ($V = 60$), 32 грани ($F = 12 + 20$) и 90 ребер (по формуле Эйлера: $P = V + F - 2$). Радиус сферы, на поверхности которой лежат вершины усеченного икосаэдра, равен $R \approx 2,478 d$, где d — длина ребра (межатомное расстояние), втрое меньшая первоначальной длины ребра исходного икосаэдра.

Усеченный икосаэдр (фуллерен C_{60}) сохраняет вращательную симметрию исходного икосаэдра, то есть также, как и додекаэдр (фуллерен C_{20}), имеет 6 осей пятого порядка и 10 осей третьего порядка, но в дополнение к ним приобретает еще 60 осей второго порядка. Можно ожидать, что в силу высокой симметрии молекулы C_{20} и C_{60} , обладая наибольшей энергией межатомной связи, являются более стабильными по сравнению с другими фуллеренами. Поэтому они вероятнее всего должны были бы наблюдаться в условиях эксперимента.

На самом деле это справедливо только для фуллерена C_{60} , в то время как молекулы C_{20} до сих пор не были экспериментально

наблюдаемы. Дело в том, что фуллерены чаще всего технологически получают путем термического разрушения графита, состоящего из слоев с правильными шестиугольниками (см. ниже рис. 3.10). Однако в структуре молекулы C_{20} (рис. 11 а) гексагоны отсутствуют. Поэтому сборку молекул фуллерена C_{20} осуществить таким способом (из слоев графита) нельзя, а синтезировать их можно только из отдельных атомов углерода.

В случае молекулы C_{60} имеются предпосылки для ее сборки из слоев графита путем перестройки межатомной связи с превращением части исходных гексагонов в пентагоны и скручивания их всех в сферическую поверхность усеченного икосаэдра (см. нижнюю фигуру на рис. 11 в). Полагая, что в состав C_{60} входят те же шестиугольники, что и в графите, со стороной $d = 1,42 \text{ \AA}$, получаем радиус этой молекулы $R = 2,478 d = 3,52 \text{ \AA}$. В фуллерене C_{60} имеются два типа межуглеродных связей: одна (двойная) связь вдоль границы двух гексагонов длиной $1,39 \text{ \AA}$ и другая (одинарная) связь вдоль общей стороны гексагона и пентагона длиной $1,44 \text{ \AA}$, то есть гексагоны не являются правильными. Результаты рентгеноструктурного анализа показывают, что радиус молекулы C_{60} составляет $3,57 \text{ \AA}$.

2. Фуллериты, фуллериды и эндоэдраны представляют собой кристаллические соединения, построенные на основе фуллеренов, чаще всего, на основе фуллерена C_{60} .

Фуллериты являются молекулярными кристаллами, образованными из молекул фуллерена, например C_{60} , взаимодействующих между собой при помощи сил Ван-дер-Ваальса (см. с. 340). Симметричное расположение атомов углерода по сферической поверхности усеченного икосаэдра позволяет сравнить молекулу C_{60} со сферически-симметричными атомами инертных газов. В обоих случаях при кристаллизации возникает гранецентрированная кубическая (ГЦК) решетка (см. п. 3.2).

Экспериментально измерен период ГЦК-решетки фуллерита, $a = 14,15 \text{ \AA}$. Тогда число молекул в единице объема $N_{C_{60}} = 4/a^3 = 1,41 \cdot 10^{21} \text{ см}^{-3}$, с учетом того, что элементарная ГЦК-ячейка содержит четыре молекулы. Зная массу одной молекулы C_{60} ($m = 720 \text{ а.е.м.} = 1,195 \cdot 10^{-21} \text{ г}$), получаем плотность фуллерита $\rho = N_{C_{60}} m = 1,69 \text{ г/см}^3$. Это свидетельствует о достаточно рыхлой структуре фуллерита по сравнению не только с алмазом ($\rho = 3,5 \text{ г/см}^3$), но и с графитом ($\rho = 2,3 \text{ г/см}^3$).

Это заключение подтверждается также оценкой расстояния d между центрами соседних молекул, расположенных вдоль диагонали грани в ГЦК-ячейке (см. четвертую строку в табл. 3.2):

$d = 2r = a/\sqrt{2} = 10 \text{ \AA}$, что согласуется с измерениями. Согласно вышесказанному, диаметр молекулы C_{60} ($2R = 7,14 \text{ \AA}$) меньше расстояния d между центрами молекул на $2,86 \text{ \AA}$, т. е. имеется зазор между сферическими поверхностями соседних фуллеренов.

Следовательно, молекулы фуллерена в ГЦК-решетке фуллерита, будучи сферически-симметричными и неплотно прижатыми друг к другу, могут, как показал эксперимент, вращаться с частотой около 10^{11} с^{-1} при 300 К (что втрое ниже такой же частоты для свободной молекулы C_{60} в газовой фазе). При температуре 257 К происходит фазовый переход первого рода (см. п. 2.1) с перестройкой ГЦК-решетки в простую кубическую с очень малым уменьшением периода (на $0,04 \text{ \AA}$). Однако вращение фуллеренов затормаживается до $5 \cdot 10^8$ оборотов в секунду вследствие возрастания энергии межмолекулярного взаимодействия.

Фуллериты являются полупроводниками, имеющими ширину запрещенной зоны около 1,5 эВ для C_{60} . Поскольку фуллериты, как отмечалось, имеют достаточно рыхлую структуру, то их электрические свойства заметно изменяются при изменении не только температуры, но и давления. Например, при увеличении давления от 10^4 до 10^5 атм удельное электрическое сопротивление уменьшается почти в десять раз. Сжатие кристалла при высоких температурах приводит сначала к уменьшению периода ГЦК-решетки на 5% при $T = 600 - 700 \text{ К}$. При дальнейшем нагреве до $800 - 1100 \text{ К}$ наблюдается переход к ромбоэдрической структуре за счет формирования химической связи в результате уменьшения расстояния между молекулами фуллерена в кристалле. При комнатной температуре наблюдалось превращение поликристаллического фуллерена C_{60} в алмаз под давлением около $2 \cdot 10^5$ атм (аналогичная величина для превращения графита в алмаз почти на полтора порядка больше).

Как и любой полупроводник, фуллериты допускают легирование путем введения инородных примесей, включая металлы. Это существенно изменяет их свойства вплоть до перехода к металлической проводимости. Чтобы отличить такие материалы от чистых фуллеритов, им присвоено общее название «фуллериды» или «металлофуллериты» для случая легирования металлами.

Фуллериды (металлофуллериты) наиболее часто получают легированием фуллерита (твердого фуллерена) C_{60} щелочными металлами ($M = K, Rb, Cs$) с образованием химического соединения M_3C_{60} , имеющего ГЦК-решетку, как и чистый фуллерит. Атом металла M связан с молекулой C_{60} посредством ионной связи, которая гораздо сильнее ван-дер-ваальсовой (см. п. 3.1) и

поэтому ограничивает возможные движения молекул фуллерена. Как результат этого, в металлофуллеритах отсутствует как фазовый переход, так и вращение молекул C_{60} при высоких температурах, что было характерно для чистого фуллерита.

Как известно (см. ниже рис. 3.6 и относящийся к нему текст), элементарная ГЦК-ячейка фуллерита, составленная из плотноупакованных молекул C_{60} , имеет 4 октаэдрические пустоты и 8 тетраэдрических пустот, внутри которых могут размещаться атомы щелочных металлов М. Поскольку на одну ГЦК-ячейку приходится 4 молекулы C_{60} , то при заполнении октаэдрических и тетраэдрических пустот возникают, соответственно, фуллериды MC_{60} и M_2C_{60} , показанные на рис. 11 з и рис. 11 д (где молекулы, находящиеся в центре граней, изображены более светлыми). После полного заполнения всех пустот (октаэдрических и тетраэдрических) возникает соединение M_3C_{60} .

Легирование фуллеритов щелочными металлами приводит к появлению металлической проводимости, которая при низких температурах проявляется в форме сверхпроводимости. Температура сверхпроводящего перехода, как показал эксперимент, составляет 19 К для K_3C_{60} , 29 К для Rb_3C_{60} и 33 К для $RbCs_2C_{60}$.

Эндоэдралы (от англ. *endohedral*) представляют собой специфическую форму легированных фуллеренов, полученную путем интеркалирования (от англ. *intercalation*) частиц (атомов или молекул) *внутри* фуллереновой сферы. Такая возможность основана на том факте, что диаметр внутренней полости в фуллеренах (от 7 Å для C_{60} до 1,5 нм для высших фуллеренов) заметно превышает эффективные размеры атомов и простейших молекул (от 1 Å до 4 Å).

Эндоэдральные фуллерены принято обозначать символом $M_m@C_n$, для того чтобы отличить их от экзоэдральных структур (обычных металлофуллеритов), традиционно обозначаемых M_mC_n . Из-за большего размера внутренней полости наиболее распространены эндоэдралы на основе высших фуллеренов, например для скандия $Sc_2@C_{82,84,86}$, $Sc_3@C_{82}$ и т. п.

К настоящему времени список веществ, которые удалось интеркалировать (в атомарной или молекулярной форме) внутрь сферической (сфероидальной) оболочки фуллеренов, включает в себя значительную часть элементов периодической таблицы. Получены экспериментальные данные, характеризующие зарядовое состояние атомов, тип их химической связи с атомами углерода, геометрическое положение и характер движения (колебательного и вращательного) внутри углеродной оболочки фуллерена. Так,

исследование эндоэдрала $\text{Li}_2@\text{C}_{60}$ показало, что димер лития (Li-Li длиной $3 \text{ \AA}$) размещается в середине сферы, координируясь к противоположным вершинам шестиугольников, и при 300 K вращается с частотой $2,5 \cdot 10^{12} \text{ с}^{-1}$, при этом собственная частота вращения сферы C_{60} примерно равна $3,4 \cdot 10^9 \text{ с}^{-1}$.

Наряду с одиночными эндоэдрами, ныне синтезированы и исследованы *эндоэдральные фуллериты*, то есть кристаллические структуры, составленные из эндоэдральных фуллеренов. Так как последние по размерам мало отличаются от полых фуллеренов, то параметры кристаллических решеток также близки.

Различие может наблюдаться в том случае, когда инкапсулированные частицы смещены относительно центра сферы. Благодаря этому возникает электрический дипольный момент, влияющий на характер межмолекулярного взаимодействия в кристалле. В частности, возникает упорядочение в расположении эндоэдральных молекул, что обуславливает спонтанную электрическую поляризацию и проявляется в виде сегнетоэлектрических свойств кристаллов эндоэдральных фуллеритов.

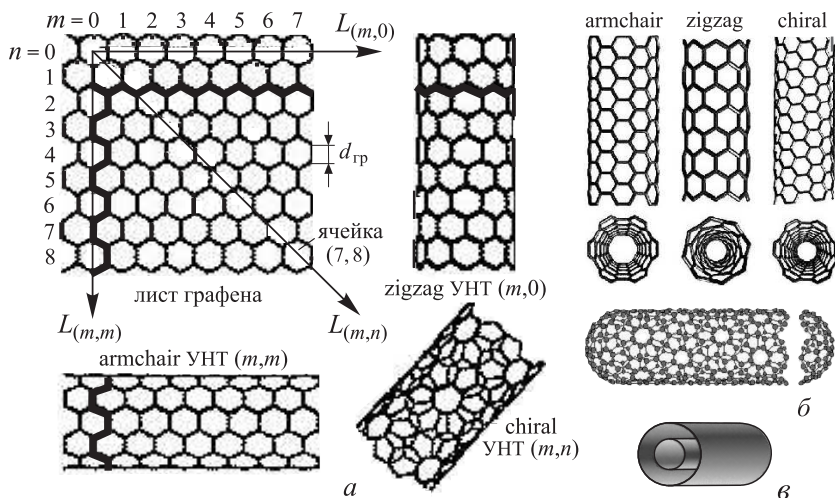
Кроме того, эндоэдральная специфика проявляется в том, что валентные электроны внутреннего атома могут переходить на внешнюю поверхность молекулы фуллерена. Это изменяет ее химические свойства и отражается на возможности взаимодействия с другими частицами, приводящего к возникновению новых эндоэдральных химических соединений. Специфическая особенность таких соединений в форме появления изомеров связана с дополнительной асимметрией, порожденной смещением внутренней частицы относительно центра сферы.

Практическую перспективность имеют эндоэдральные комплексы с молекулярным водородом типа $\text{H}_2@\text{C}_{70}$. Они позволяют запасать водород для использования его в качестве высокоэффективного энергоносителя (как топливо для двигателей внутреннего сгорания и как основа для производства аккумуляторных батарей большой емкости). Аналогичным образом открывается перспектива для хранения ядовитых и радиоактивных веществ. Эндоэдральные фуллерены с интеркалированными радиоактивными изотопами позволят избирательно воздействовать на вирусы и раковые клетки.

3. Углеродные нанотрубки (УНТ) представляют собой специфическую форму аллотропной модификации углерода в виде скрученной в цилиндр графитовой плоскости (*графена*) (см. рис. 3.11 б и в). Впервые углеродные нанотрубки были наблюдаемы в 1991 году, то есть намного раньше, чем экспериментально

был получен графен (2004 г.), хотя задолго до этого слои графита теоретически изучались как модель двумерного кристалла.

Цилиндрическая поверхность идеальной трубки, в отличие от сфероидальной поверхности фуллеренов, составлена только из правильных шестиугольников (без пентагонов, проявляющихся как дефект поверхности), в вершинах которых расположены атомы углерода. Ориентация гексагонов по отношению к оси является важной структурной характеристикой, показывающей скрученность трубки, называемой *хиральностью* (от англ. *chirality*).



$$L_{(m,n)} = \pi D_{(m,n)} = \sqrt{3(m^2 + n^2 + mn)} d_{\text{ГР}}, \quad L_{(m,m)} = 3m d_{\text{ГР}} = \sqrt{3} L_{(m,0)}$$

Рис. 12. а — схема, иллюстрирующая образование из листа графена однослойной нанотрубки хиральной конфигурации (chiral УНТ при $n \neq m$) и двух специфических конфигураций: зигзагообразной (zigzag УНТ при $n = 0$) и креслообразной (armchair УНТ при $n = m$); продольный и поперечный вид трех конфигураций показан в правом верхнем углу; б — однослойная нанотрубка с закрытыми концами; в — модель двухслойной нанотрубки. Формула внизу связывает длину окружности трубки ($L_{(m,n)}$) или ее диаметр $D_{(m,n)} = L_{(m,n)}/\pi$) с длиной связи С–С в графене ($d_{\text{ГР}} = 1,44 \text{ \AA}$) и индексами хиральности (m, n)

Хиральность нанотрубок принято обозначать набором индексов (m, n) , нумерующих шестиугольные ячейки по горизонтали ($m = 0, 1, 2, \dots$) и по вертикали ($n = 0, 1, 2, \dots$), как показано на рис. 12 а. При сворачивании листа графена в трубку вдоль произвольного направления, перпендикулярного стрелке $L_{(m,n)}$, вершины (m, n) -гексагона должны совпасть с эквивалентными вершинами $(0, 0)$ -гексагона, находящегося в начале координат.

Расстояние $L_{(m,n)}$, измеренное вдоль направления стрелки, дает развертку окружности поперечного сечения трубки с диаметром $D_{(m,n)} = L_{(m,n)} / \pi$. Величина $L_{(m,n)}$ определяется хиральным индексом (m, n) и вычисляется по формуле, приведенной на рис. 12.

Два частных случая, соответствующие $n = 0$ и $n = m$, дают специфические конфигурации, названные *зигзагообразной нанотрубкой* (zigzag УНТ при $n = 0$) и *креслообразной нанотрубкой* (armchair УНТ при $n = m$). Специфика этих названий пояснена соответствующими жирными линиями на рис. 12 а. В правом верхнем углу рисунка показаны три конфигурации нанотрубок с их поперечными сечениями.

Согласно формуле, приведенной внизу рис. 12, диаметры armchair и zigzag УНТ связаны соотношением: $D_{(m,m)} = \sqrt{3} D_{(m,0)} = (3/\pi) m d_{\text{гp}}$, где $d_{\text{гp}} = 1,44 \text{ \AA}$ — длина углеродной связи в графене, что немного больше, чем в слоях графита ($d_{\text{C-C}} = 1,42 \text{ \AA}$).

Интересно отметить, что хиральность углеродных нанотрубок определяет их электронные свойства, в частности тип электрической проводимости (металлический или полупроводниковый). Теоретически установлено и подтверждено экспериментами, что металлической проводимостью обладают нанотрубки, для которых $m - n = 3i$, где i — целые числа, включая $i = 0$.

Следовательно, креслообразные нанотрубки ($m = n$) — всегда металлические, а зигзагообразные ($n = 0$) и хиральные ($m \neq n$) нанотрубки могут быть как металлическими (при $m - n = 0, 3, 6, 9, \dots$), так и полупроводниковыми (при $m - n \neq 0, 3, 6, 9, \dots$). В последнем случае ширина запрещенной зоны обратно пропорциональна диаметру УНТ, изменяясь примерно от 0,8 эВ до 0,4 эВ при изменении $D_{(m,n)}$ от 1 нм до 2 нм.

Все вышеизложенное относится к **однослойным нанотрубкам**, которые обычно получают закрытыми, то есть с концами, замкнутыми фуллереновыми полусферами, как показано на рис. 12 б. В идеальном случае каждая из двух концевых полусфер содержит по шесть правильных пятиугольников, общим числом 12, как и в случае фуллеренов. Согласно формуле Эйлера, приведенной на рис. 10, правильный выпуклый многогранник всегда содержит 12 пентагонов ($N_5 = 12 \equiv \text{const}$) и произвольное число гексагонов ($N_6 = \text{var}$), превращающих сферу в вытянутый сфероид. Поэтому идеальную однослойную нанотрубку можно рассматривать как сильно вытянутый сфероид с огромным количеством шестиугольников на цилиндрической поверхности. Появление здесь пятиугольников и семиугольников создает дефект, который изгибает и закручивает цилиндрическую поверхность.

Среди однослойных нанотрубок преимущественно наблюдаются креслообразные (armchair) трубки с хиральностью (10,10), обладающие металлической проводимостью и повышенной термодинамической устойчивостью по сравнению с УНТ других хиральностей. Такие нанотрубки имеют диаметр 1,38 нм (что соответствует формуле на рис. 12) и длину до сотен микрометров.

Однако на практике чаще всего при синтезе УНТ возникают не однослойные, а **многослойные нанотрубки** в форме коаксиально вложенных друг в друга однослойных цилиндров с расстоянием между соседними слоями, как и в графите, равным 0,34 нм (см. рис. 12 в для двухслойной УНТ). Кроме концентрических многослойных нанотрубок, называемых «русской матрешкой», наблюдаются также «свитки» (не показанные на рис. 12) в форме единой графеновой плоскости, скрученной в спираль с периодом, близким к 0,34 нм.

В реальности многослойные нанотрубки имеют геометрические дефекты и отличаются широким разнообразием форм и конфигураций как в продольном, так и в поперечных направлениях. Экспериментально наблюдаются многослойные УНТ (с числом слоев до 50–100 и внешним диаметром в десятки нанометров), которые плотно скручены в витые жгуты (нитеподобные структуры) диаметром около 0,1 мм и длиной до 3 мм. Такие нанонити обладают чрезвычайно высокой механической прочностью с модулем упругости в аксиальном направлении вплоть до 7000 ГПа, что в 35–40 раз превышает значения, характерные для стали.

Закрытые углеродные нанотрубки, подобно эндоэдральным комплексам фуллеренов, перспективны как хранилища газообразного водорода (для двигателей транспортных средств) и радиоактивных отходов (для изоляции их от окружающей среды), которые инкапсулированы внутри нанотрубки.

В применении к электронике представляют практический интерес, во-первых, управляемая электрическая проводимость УНТ (от металлической до полупроводниковой) и, во-вторых, автоэмиссионная способность, обеспечивающая плотность тока эмиссии до $0,1 \text{ А/см}^2$ при напряжении примерно 500 В (что соответствует напряженности поля около $1,5 \cdot 10^5 \text{ В/см}$ и работе выхода, равной 1 эВ). Как показывают расчеты, создание в однослойной нанотрубке с хиральностью (8,0) парного дефекта пятиугольник–семиугольник изгибает ее и изменяет хиральность трубки после изгиба на (7,1). Поскольку исходная (8,0)-трубка является полупроводником (с шириной запрещенной зоны 1,2 эВ), а полученная (7,1)-трубка обладает металлической проводимостью (так как $m - n = 6$), то возникает контакт металл–полупроводник.

Капиллярные свойства открытых нанотрубок позволяют заполнять их внутренность проводящими (Pb, Fe, Ni, Co) и даже сверхпроводящими (TaC, TiC) материалами. Такие нанотрубки могут найти применение как сверхминиатюрные и супернизкоомные элементы межсоединений в интегральных микросхемах, в том числе для компьютерных процессоров и устройств памяти. Электропроводящие и упруго-механические свойства УНТ в сочетании с супермалыми размерами закрытого конца нанотрубок делают их привлекательными для использования в качестве измерительного зонда сканирующего туннельного микроскопа.

В последнее время всерьез обсуждается проблема электронного применения графеновых нанолент для формирования транзисторных структур. С учетом особенностей металлической и полупроводниковой природы нанолент различной ширины, графен рассматривается как серьезный претендент на замену кремния в нанопроцессорных технологиях будущего.

Графеновые наноленты (ГНЛ), имея электронные характеристики, во многом аналогичные таковым для углеродных нанотрубок, обладают преимуществом перед ними из-за отсутствия контроля степени хиральности, присущей УНТ. В применении к графеновым приборным структурам сохраняются преимущества планарной технологии, в частности литографическое формирование рисунков схемы на одном графеновом слое. Матрица, содержащая полевые транзисторы на ГНЛ и межсоединения, может быть сформирована как монолитная конструкция на кремниевой пластине, покрытой термическим окислом с медной пленкой переменной толщины, на которой выращен однослойный графен. На этом пути открываются также перспективы для использования *bottom-up* технологий (см. ниже).

II. Краткое описание современных нанотехнологий

В современных условиях технологические процессы нанoeлектроники принято строить на основе двух стратегий:

- технологическая стратегия «сверху вниз» (англ. *top-down*),
- технологическая стратегия «снизу вверх» (англ. *bottom-up*).

1. Стратегия «сверху вниз» в применении к наноструктурам подразумевает создание малых элементов (наноразмерного масштаба) путем разделения больших структурных единиц (макро- и микроразмерного масштаба) на части, последовательно (а возможно и параллельно, т. е. одновременно) воспроизводимые в ходе технологического процесса. Принципиальные вопросы такого подхода были решены в рамках интегральной микроэлектроники

и заложены в основу планарно-эпитаксиальной технологии (см. рис. 4). Естественно, что переход к наноразмерным технологическим объектам (с размерами менее 100 нм) требует модификации и усовершенствования существующих микроэлектронных технологий. Но важно, что их принципиальные особенности и способы реализации при этом сохраняются, а это существенно снижает материальные и энергетические затраты.

Среди разнообразных методов планарно-эпитаксиальной технологии в нанoeлектронике нашли применение нижеследующие.

- **Молекулярно-лучевая эпитаксия** (МЛЭ) применяется, в первую очередь, для создания полупроводниковых сверхрешеток на основе двойных, тройных и четверных твердых растворов типа $A^{III}B^V$, а также полупроводниковых гетероструктур с использованием квантовых ям, квантовых проволок и квантовых точек как базовых элементов оптической нанoeлектроники (см. рис. 4.22).

- **МОС-гидридная эпитаксия** применяется, в первую очередь, как и МЛЭ, для создания сверхрешеток и многослойных структур с толщиной слоев в несколько межатомных расстояний и резкими гетерограницами, используемых в полупроводниковых лазерах на квантовых ямах, нитях и точках (см. п. 4.13).

- **Реактивное распыление металлов в плазме инертного газа** применяется для пленочного осаждения химических соединений металлов, таких как оксиды, нитриды и т. п. (см. п. 4.16).

- **Методы фотолитографии**, усовершенствованные применением лазерного УФ-излучения, фотошаблонов с фазовым сдвигом и технологии двойной экспозиции (сухая и иммерсионная DUV-литография) позволили при создании процессоров реализовать 32-нм техпроцесс без перехода на EUV-литографию с длиной волны $\lambda = 13$ нм (см. рис. 8).

- **Химическая сборка** поверхностных наноструктур методами молекулярного наслаивания и атомно-слоевой эпитаксии применяется для управляемого послойного наращивания химических соединений, в частности диоксида гафния HfO_2 при формировании High-K/Metal-Gate структур в современной процессорной технологии МОП-транзисторов (см. с. 39, а также рис. 4.20 и 4.21). Послойная молекулярная и атомарная химическая сборка поверхностных наноструктур является первым шагом к реализации стратегии «снизу вверх».

2. Стратегия «снизу вверх» подразумевает формирование наноструктур из элементарных атомно-молекулярных блоков путем их *самосборки*. В основу такой самосборки разумно закладывать

принципы, присущие живой природе¹⁾. Для создания сложных молекулярных конструкций природа использует принцип *самораспознавания*: молекулы “узнают” друг друга и, как результат этого, собираются по принципу «ключ–замок» на основе структурного соответствия (вспомним структуру молекул ДНК). Именно принципы самораспознавания и структурного соответствия запускают природный механизм молекулярной самосборки.

На сегодняшний день перед исследователями (физиками и химиками) стоит фундаментальная задача — понять, каким образом по принципу «снизу вверх» искусственно собирать из отдельных атомов и молекул самоорганизующиеся системы для того, чтобы создавать электронные функциональные устройства наноразмерного диапазона. При этом подобная самосборка должна осуществляться по наперед заданной схеме с особо высокой точностью, присущей природным объектам. В настоящее время единственную возможность реализации принципа «снизу вверх» в материаловедении дают *зондовые нанотехнологии*.

Зондовые нанотехнологии появились лишь после изобретения *сканирующего туннельного микроскопа* — СТМ (Г. Бинниг и Г. Рорер, 1982 г.) и разработки на его основе принципов манипулирования атомами и молекулами (М. Рид и Дж. Тур, 1999 г.). Простота общей идеи функционирования СТМ усугубляется, однако, невероятной сложностью ее технического воплощения.

В основу работы СТМ положен принцип квантово-механического туннелирования электронов через «ангстремный» зазор между исследуемой поверхностью и вольфрамовым зондом с радиусом острия моноатомных размеров. Движение зонда по нормали к поверхности (z координата) и сканирование в плоскости (x, y) управляется пьезоэлектрическими датчиками, выполненные из титаната бария. Чувствительность таких датчиков примерно равна $1 \text{ \AA/V}$, что позволяет перемещать зонд на расстояния от долей ангстрема до долей микрона. Для надежной работы СТМ система виброизоляции и демпфирования должна обеспечивать виброперемещения зонда не более $0,01 \text{ \AA}$.

¹⁾ Природа демонстрирует уникальный пример самоорганизации и экономии ресурсов: живой организм на 99% состоит всего из 4 элементов — углерод, кислород, азот, водород, плюс еще 12 элементов, которые входят в оставшийся один процент. Исходя из этих составляющих, природа построила систему, которая, начиная с растений и заканчивая человеком, слаженно функционирует и воспроизводит себе подобные виды. Главная суть такой самоорганизации, построенной по принципу «снизу вверх», заключается в том, что молекулы собираются в коллективы не в результате ковалентных связей, а за счет сил межмолекулярного взаимодействия.

Сигналы пьезодатчиков через электронную цепь обратной связи обрабатываются компьютерной системой сбора и преобразования электрических данных в топографическое изображение рельефа исследуемой поверхности. Кроме трех координат (x, y, z) положения зонда, варьируемыми параметрами также являются напряжение V и туннельный ток I . В зависимости от того, какие из этих параметров фиксированы, а какие варьируются, различают следующие три режима работы СТМ.

1. *Режим постоянного тока*: при $I = \text{const}$ и $V = \text{const}$ измеряют z -перемещения в точках $(x, y) = \text{var}$ (сканирование). Постоянство тока означает неизменность зазора d между атомом на острие зонда и атомом на поверхности в точке (x, y) , т. е. z -перемещения зонда отражают *топографическую картину рельефа*.

2. *Режим постоянной высоты*: при $z = \text{const}$ и $V = \text{const}$ измеряют ток I в точках $(x, y) = \text{var}$ (сканирование). Постоянство z -положения зонда создает переменную величину зазора $d(x, y)$, проявляющуюся в величине тока $I(x, y)$, тем самым получается *токовое изображение рельефа*.

3. *Туннельная спектроскопия токовых изображений*: при $z = \text{const}$ измеряют ток I как функцию $V = \text{var}$ в разных точках $(x, y) = \text{var}$ (сканирование). Так как туннельный ток определяется всеми электронными состояниями в интервале энергий, определяемых напряжением V , то, варьируя V , получают спектры локальной плотности состояний как функции энергии.

Дальнейшее продвижение зондовых технологий возникло благодаря изобретению *атомно-силового микроскопа* — АСМ (Г. Бинниг и К. Гербер, 1986 г.) В отличие от сканирующей туннельной микроскопии, применимой лишь для исследования проводящих поверхностей, принцип АСМ распространяется также на полупроводниковые и диэлектрические материалы.

В основе работы АСМ лежит проявление межатомных сил, действующих между атомами на острие зонда и на исследуемой поверхности. Для этого зонд размещается на конце относительно длинного (100–200 мкм) и упругого кронштейна, называемого *кантилевером* (от англ. *cantilever*), другой конец которого закреплен в подвижном (x, y) -сканере с пьезодатчиками горизонтальных смещений. Вертикальное смещение кантилевера регистрируется при помощи: а) туннельного тока, б) оптической интерферометрии, в) отражения лазерного луча, г) измерения емкости между кантилевером и близко расположенным электродом. Датчики регистрируют отклонение кантилевера на $0,01 \text{ \AA}$. Сканирование зонда дает топографическую картину рельефа.

На практике для получения топографического изображения поверхности обычно используют следующие три АСМ-режима.

1. *Контактный режим при постоянной высоте сканера.* Малое расстояние (в несколько ангстрем) между зондом и поверхностью, благодаря действию межатомных сил отталкивания, обеспечивает мягкий физический контакт между ними. Отклонения кантилевера в ходе сканирования показывают топографический рельеф поверхности.

2. *Контактный режим при постоянной силе,* действующей на кантилевер. Постоянство силы в ходе сканирования обеспечивается поддержанием неизменного отклонения кантилевера путем непрерывной подстройки высоты сканера с помощью системы слежения. Топографическое изображение строится на основе вертикальных перемещений кантилевера.

3. *Бесконтактный режим* обеспечивается при расстояниях от зонда до поверхности в десятки или сотни ангстрем. В этом режиме между атомами проявляются силы притяжения, под действием которых кантилевер колеблется с частотой, близкой к своей резонансной частоте, обычно лежащей в диапазоне от 100 до 400 кГц. В ходе сканирования контролируется либо смещение частоты, либо амплитуда колебаний кантилевера, типичные значения которой составляют десятки ангстрем.

До сих пор речь шла об использовании зондовых методов для атомарной визуализации поверхности твердых тел в виде топологической картины рельефа. Однако зондовые методы с успехом могут также применяться для решения технологических задач с целью формирования и модификации поверхностных наноструктур путем позиционирования атомов на поверхности подложек. Эти две возможности совместно позволяют осмотреть поверхность, выбрать там место или интересующий атомарный объект и провести необходимые атомные манипуляции.

Атомные манипуляции при помощи зонда СТМ (рис. 13) сводятся к следующим трем элементарным процессам.

1. *Перемещение атома вдоль поверхности* осуществляется двумя способами (см. рис. 13 а):

- перемещение под действием межатомных сил возникает при приближении острой зонда к поверхностному (адсорбированному) атому на расстояние, при котором перекрываются волновые функции атомов, что обеспечивает их взаимное притяжение;
- перемещение под действием электростатического поля возникает при приложении напряжения между образцом и зондом,

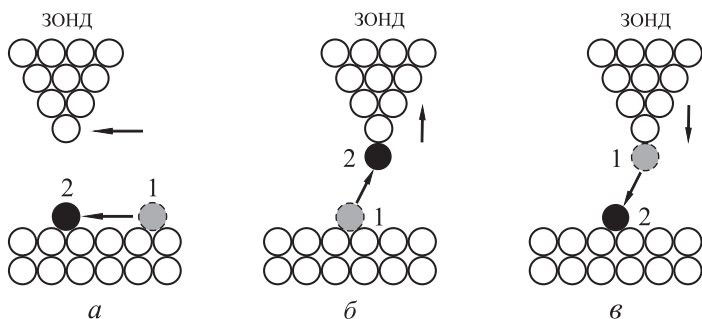


Рис. 13. Элементарные атомные манипуляции при помощи зонда СТМ: а — перемещение атома вдоль поверхности; б — удаление атома с поверхности на зонд; в — осаждение атома с зонда на поверхность. Начальное и конечное положения одного и того же атома обозначены цифрами 1 и 2, а направления перемещения атома и зонда показаны стрелками

так что поле поляризует атом, индуцируя в нем диполь, который втягивается в область максимального поля под острием зонда.

2. *Удаление атома с поверхности на зонд* осуществляется за счет двух механизмов (см. рис. 13 б):

- межатомное взаимодействие, возникающее при перекрытии волновых функций атомов на зонде и на поверхности, обеспечивает перемещение атома не только вдоль поверхности, но и по нормали, способствуя его удалению;
- испарение атомов полем происходит как результат их ионизации при подаче импульса напряжения на зонд, который своим полем вытягивает эти ионы на острие.

3. *Осаждение атома с зонда на поверхность* (см. рис. 13 в) является менее надежным процессом, чем удаление, при этом чаще происходит осаждение не отдельных атомов, а их кластеров. После контакта зонда с поверхностью он плавно отводится с образованием между ними тонкой «шейки», разрыв которой приводит к появлению бугорка из материала зонда.

В литературе появились публикации по технологическому решению проблемы создания базовой ячейки квантового компьютера, называемой *кубитом* (сокращение от англ. *quantum bit*). Кубит, как квантовую систему с двумя состояниями спина $\pm 1/2$ в постоянном магнитном поле, предложено реализовать на ядерных спинах донорных атомов фосфора (P) в чистом кремнии (Si). Квантовый регистр, составленный из последовательной цепочки кубитов, может быть построен на нанометровых транзисторных структурах с расстоянием между соседними кубитами в несколько десятков нанометров.

Последовательность технологических операций по введению атомов фосфора Р (темные кружки) в матрицу монокристаллического кремния Si показана на рис. 14.

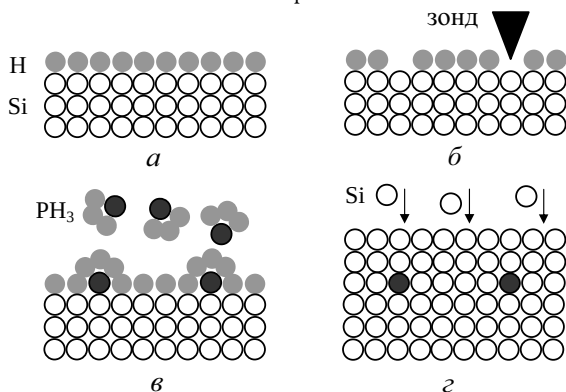


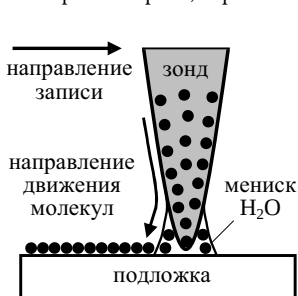
Рис. 14. Технологическая схема формирования кубита путем точного позиционирования атомов фосфора в матрице монокристаллического кремния

Первоначально поверхность кремния очищают в сверхвысоком вакууме и атомарно чистую поверхность пассивируют монослоем водорода Н (рис. 14 а). Затем по специальной программе с помощью зонда сканирующего туннельного микроскопа удаляют отдельные атомы водорода в заданных местах поверхности (рис. 14 б). После этого в камеру вводятся пары фосфина PH_3 . Адсорбированные на поверхности структуры молекулы фосфина (рис. 14 в) диссоциируют при 500°C , оставляя в местах адсорбции атомы фосфора, связанные с кремнием. После этого удаляется монослой водорода и производится низкотемпературное наращивание кремнием полученной структуры (рис. 14 г).

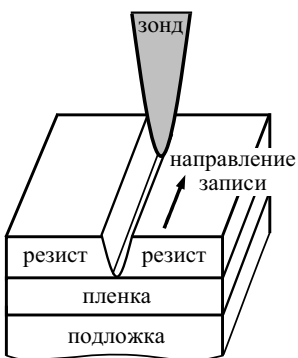
Описанные выше зондовые способы манипуляции атомами требуют тонкой и кропотливой работы, возможной лишь в лабораторных условиях, оснащенных необходимой (сложной и дорогостоящей) аппаратурой, что вряд ли доступно в производстве. Поэтому особую важность приобретает альтернативный подход, основанный на стимулировании *процессов самоорганизации* путем создания таких условий, в которых система сама стремится формировать нужные наноструктуры спонтанным образом. Именно в этом направлении работают сегодня научно-исследовательские коллективы ведущих научных центров, институтов и фирм.

Зондовые методы нанолитографии используют сканирующий туннельный микроскоп или атомно-силовой микроскоп в качестве источника электронов для локального экспонирования

электронорезиста в методе электронной литографии. При этом применяют различные режимы эмиссии электронов: режим постоянного туннельного тока СТМ, бесконтактный режим АСМ с электропроводящим зондом и др. При использовании в качестве резиста полиметилметакрилата удается формировать изображения с разрешением менее 100 нм. Толщина слоя резиста, требуемая для экспонирования, обычно составляет 35–100 нм. Экспонирование проводят в режиме сканирования зондом, на который подают отрицательное смещение в несколько десятков вольт. Проявление экспонированной области осуществляют в стандартных растворах, при этом достигается разрешение порядка 50 нм.



а



б

Рис. 15. Зондовые методы нанолитографии: а — метод «погруженного пера» на основе СТМ, б — метод механической зондовой литографии на основе АСМ

К новым методам зондовой нанолитографии можно отнести так называемый метод «погруженного пера» (рис. 15 а). Этот метод основан на переносе молекул с зонда на подложку посредством их диффузии через мениск воды, соединяющий зонд и подложку. Молекулы с необходимыми химическими свойствами наносятся на поверхность зонда путем окунания его в соответствующий разбавленный раствор реагента с последующим испарением растворителя (часто осуществляемым в потоке дифторэтана). Данный метод позволяет формировать линии шириной до 12 нм на расстоянии 5 нм одна от другой.

Существуют также другие методы формирования нанометровых рисунков с помощью зонда АСМ. Например, возможно механическое модифицирование поверхности, которое проводят в режиме постоянного или импульсного давления на нее. Второй способ предпочтительнее, поскольку меньше подвержен влиянию шероховатости подложки. Принципиальная схема механической зондовой литографии представлена на рис. 15 б.

Для осуществления локального воздействия в маскирующем покрытии, например в слое резиста толщиной 5–6 нм, с помощью зонда механически прорезают окна. Для получения тонких мас-

кирующих слоев обычный раствор резиста разбавляют до 50 раз и проводят центрифугирование со скоростью около 5000 об/мин. Таким способом через маску резиста можно вытравить канавки шириной менее 100 нм в эпитаксиальных пленках полупроводников и других материалов. Механически модифицированные структуры могут быть использованы в качестве маски для взрывной литографии или для последующего локального электрохимического осаждения.

Значительные успехи достигнуты в настоящее время при комбинировании механического и термического воздействий (термомеханическая литография). Данный метод позволяет формировать углубления в пленках органических материалов шириной менее 100 нм за счет термомеханического стимулирования фазовых переходов типа полимер–стекло. Термомеханическая литография является базовой идеей при создании терабитных запоминающих устройств, выдвинутой компанией IBM. Их концептуальный подход строится на создании многозондового картриджа.

В заключение кратко опишем технологические этапы синтеза новых углеродных материалов (фуллеренов, эндоэдралов и УНТ).

3. Технология получения фуллеренов и углеродных нанотрубок использует несколько способов: сжигание углеводородов, резистивный, высокочастотный и лазерный нагрев графита в атмосфере инертных газов, испарение графитовых микрочастиц в термической плазме при атмосферном давлении и др.

Чаще всего применяют дуговой разряд с двумя графитовыми электродами (плоский диск — катод и заостренный стержень — анод) в атмосфере инертного газа, обычно гелия, под давлением 50–150 Торр для фуллеренов и около 500 Торр для УНТ. Разряд с температурой около 3000 К горит при напряжении 10–30 В, токе в десятки и сотни ампер (вплоть до 500 А) и межэлектродном расстоянии в единицы миллиметров. Материал анода термически распыляется со скоростью примерно 10 г/час.

Продукты распыления осаждаются, во-первых, на охлаждаемых стенках камеры в виде фуллеренсодержащей сажи и, во-вторых, на катоде (более холодном, чем анод) в виде сажи с нанотрубками, преимущественно ориентированными по нормали к поверхности катода. В оптимальных условиях содержание фуллеренов в осадке на стенках составляет 20–30%, а выход УНТ из катодного осадка может достигать 60%.

Для получения эндоэдральных комплексов фуллеренов и нанотрубок, заполненных металлами, анод изготавливают из композитного материала, содержащего порошок металла либо его

соединения (карбид или оксид). Однако для эндодральных металлофуллеренов типа $M@C_{60}$ и $M@C_{70}$, где $M = Li, K, Rb$, наиболее эффективным оказался метод ионной имплантации, обеспечивающий выход металлофуллерита до 50% при облучении фуллеритовой пленки ионами с энергией около 30 эВ.

Наиболее технологически сложной процедурой является выделение (экстракция) полезного продукта из сажи и разделение (сепарация) его по типам частиц.

Фуллерены представляют собой, по существу, единственную модификацию углерода, растворимую в большинстве органических растворителей, таких как толуол, бензол, ксилол и другие. Поэтому обычно фуллереносодержащую сажу в течение трех часов выдерживают в кипящем толуоле, после чего жидкость выпаривают с остатком в виде черной пудры, содержащей фуллерены C_{60} и C_{70} в отношении примерно 10:1. Разделение фуллеренов достигается методом жидкостной хроматографии при пропускании раствора пудры через пористый сорбент, избирательно поглощающий фуллерены разного типа. Последующая обработка сорбента чистым органическим растворителем приводит к экстракции молекул фуллеренов из сорбента в раствор.

В отличие от фуллеренов, нанотрубки практически не растворяются в органических растворителях. Поэтому для их экстракции из катодного осадка применяют ультразвуковое диспергирование. С этой целью осадок заливают метанолом и обрабатывают ультразвуком, что создает суспензию, которая подвергается центрифугированию. Крупные частицы сажи прилипают к стенкам центрифуги, а нанотрубки остаются плавающими в суспензии. После промывки и просушки таким способом получают преимущественно многослойные УНТ с закрытыми концами.

Для превращения многослойных трубок в однослойные применяют обработку в кипящей азотной кислоте. Такой же процедурой достигается открытие концов, так как находящиеся здесь пятиугольные звенья содержат более активные атомы углерода по сравнению с атомами в узлах гексагонов, образующих цилиндрическую поверхность нанотрубок.

На основании вышеизложенного можно сделать заключение, что на сегодняшний день разработанная в научно-исследовательских лабораториях технология управляемого синтеза фуллеренов и углеродных нанотрубок достаточно сложна и далека от промышленного освоения. В применении к массовой микроэлектронной аппаратуре еще предстоит длительная технологическая работа инженеров-исследователей.

III. Краткое описание методов нанодиагностики

С развитием нанoeлектроники особый, даже первостепенный, интерес приобретает исследование свойств поверхности материалов. В настоящее время отсутствие надежных физических и физико-химических моделей служит препятствием для разработки теории поверхности. Поэтому теория заметно отстает от современных экспериментальных достижений, особенно в области методов туннельной и атомно-силовой микроскопии.

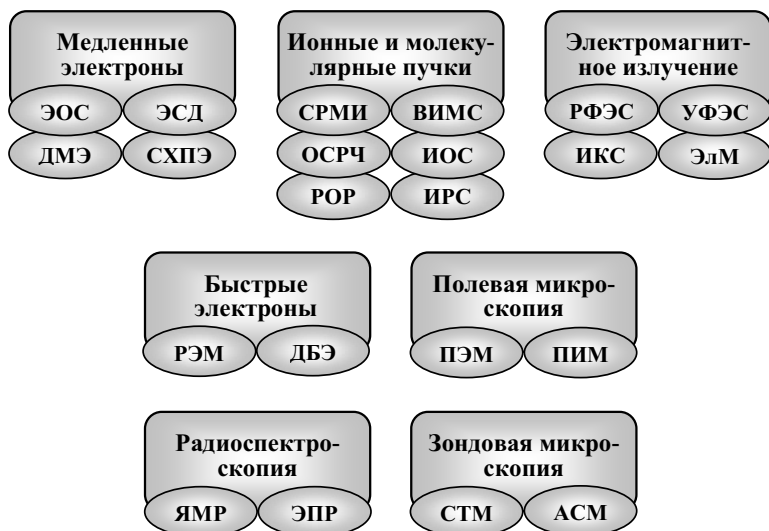


Рис. 16. Современные методы нанодиагностики: ЭОС — электронная оже-спектроскопия; ЭСД — электронно-стимулированная десорбция; ДМЭ — дифракция медленных электронов; СХПЭ — спектроскопия характеристических потерь энергии электронов; СРМИ — спектроскопия рассеяния медленных ионов; ВИМС — вторично-ионная масс-спектрометрия; ОСРЧ — оптическая спектроскопия распыленных частиц; ИОС — ионная оже-спектроскопия; РОР — резерфордское обратное рассеяние; ИРС — ионно-рентгеновская спектроскопия; РФЭС — рентгеновская фотоэлектронная спектроскопия; УФЭС — ультрафиолетовая фотоэлектронная спектроскопия; ИКС — инфракрасная спектроскопия; ЭлМ — эллипсометрия; РЭМ — растровая электронная микроскопия; ДБЭ — дифракция быстрых электронов; ПЭМ — полевая электронная микроскопия; ПИМ — полевая ионная микроскопия; ЯМР — ядерный магнитный резонанс; ЭПР — электронный парамагнитный резонанс; СТМ — сканирующая туннельная микроскопия; АСМ — атомно-силовая микроскопия

На сегодняшний день из всего многообразия методов исследования поверхности, число которых перевалило за семьдесят, основными являются методы электронной спектроскопии,

дифракции электронов (медленных и быстрых) и разнообразные методы микроскопии, которые собраны на рис. 16. Ниже приводим краткое описание этих методов нанодиагностики.

Диагностирование медленными электронами

При взаимодействии с поверхностью медленных электронов (с энергией от 10 эВ до 10 кэВ) возникают их поглощение, вторичная электронная эмиссия и электромагнитное (в том числе рентгеновское) излучение. В энергетический спектр электронов вносят вклад собственно вторичные электроны (с максимумом при 5–10 эВ), а также неупруго и упруго отраженные первичные электроны. Первые занимают широкий диапазон энергий с пиками, созданными ожэ-электронами или возникшими за счет характеристических потерь энергии. Упруго отраженные электроны создают в энергетическом спектре узкий максимум при энергии, равной энергии падающих электронов.

Электронная ожэ-спектроскопия (ЭОС) основана на измерении энергии и количества ожэ-электронов. Выход таких электронов в составе неупруго отраженных появляется вследствие ионизации внутренних оболочек атома. Возникшая электронная вакансия заполняется (за времена $10^{-12} - 10^{-14}$ с) электроном, перешедшим с внешней оболочки. Такой переход может быть как излучательным с испусканием кванта рентгеновского излучения, так и безызлучательным, когда избыток энергии расходуется на выход ожэ-электрона с вышележащего уровня в вакуум.

Энергетические спектры ожэ-электронов измерены для всех элементов периодической системы (водород и гелий не образуют ожэ-электронов). Поэтому измерение спектров вторичных электронов позволяет определить элементный состав поверхности и оценить концентрацию и поверхностное распределение элементов. Для изучения элементного состава по глубине регистрируют спектры ожэ-электронов по мере удаления материала в процессе ионного распыления. ЭОС применяют также для изучения процессов адсорбции и диффузии примесей из объема к поверхности. Абсолютная чувствительность метода около 10^{-14} г при относительной чувствительности по составу порядка $10^{-3}\%$.

Электронно-стимулированная десорбция (ЭСД) основана на определении состава и количества частиц, десорбированных с поверхности в результате ее облучения медленными электронами. Энергия электронов должна быть достаточной для разрыва связей адсорбированных атомов с поверхностью и ионизации десорбированных частиц. Измеряют ток положительных или отрицательных ионов в составе десорбированного потока.

С помощью ЭСД изучают энергетическую структуру адсорбированных частиц, химический состав поверхности и др. Чувствительность и точность метода определяется параметрами масс-спектрометрического анализатора, применяемого в приборе.

Дифракция медленных электронов (ДМЭ) лежит в основе измерения тока, созданного упруго рассеянными первичными электронами с начальной энергией 10–500 эВ. Относительная доля таких электронов изменяется от 20–30% при начальной энергии около 10 эВ до примерно 1% при 100 эВ.

Сила тока и угол отражения обусловлены кристаллической структурой одного–двух атомных слоев. При когерентном рассеянии от кристаллической решетки электроны формируют дифракционную картину, наблюдаемую на дисплее измерительной установки. Из анализа дифракционной картины получают информацию о структурных параметрах поверхности. Рекордная точность определения геометрического положения атомов методом ДМЭ достигает $3 \cdot 10^{-3}$ нм.

Спектроскопия характеристических потерь электронов (СХПЭ) основана на измерении энергии и количества тех электронов, которые в результате рассеяния потеряли некоторую долю энергии. Отличительной чертой является то, что при изменении начальной энергии падающих электронов на величину ΔE пики в энергетическом спектре вторичных электронов, соответствующие ХПЭ, сдвигаются на ту же величину ΔE . В то же время для ожэ-электронов такой сдвиг отсутствует. ХПЭ могут быть вызваны ионизацией атомов, одночастичными возбуждениями, коллективными возбуждениями электронного газа и возбуждением фононов. Для изучения каждого из этих механизмов применяют соответствующую разновидность СХПЭ.

С помощью метода СХПЭ изучают элементный состав поверхности, природу химического взаимодействия между поверхностными атомами и процессы адсорбции–десорбции, а также измеряют оптические константы материала.

Диагностирование ионными и молекулярными пучками

Спектроскопия рассеяния медленных ионов (СРМИ) основана на явлении рассеяния моноэнергетического пучка медленных ионов с энергией 0,1–10 кэВ. При исследовании поверхности измеряют энергетический спектр отраженных ионов, обусловленный однократными или последовательными парными упругими столкновениями первичных ионов с отдельными атомами в поверхностном слое. В качестве первичных ионов обычно применяют ионы инертных газов.

На основании анализа энергетических спектров определяют элементный состав поверхностного слоя. Разрешающая способность и чувствительность метода определяется массой ионов.

Вторично-ионная масс-спектрометрия (ВИМС) основана на измерении характеристик вторичных ионов, возникающих при распылении поверхности под воздействием первичных ионов (обычно аргона) с энергией 1–20 кэВ. Распыление происходит в результате упругого и неупругого взаимодействия первичных ионов с поверхностными атомами. Часть атомов в распыленном потоке находится в ионизованном состоянии. Измерение энергетических и угловых распределений вторичных ионов позволяет найти элементный состав поверхностного слоя. Распределение атомов по глубине определяют при длительном воздействии первичных ионов, когда происходит послойное удаление материала. Чувствительность ВИМС лежит в пределах 10^{-7} – 10^{-3} ат. %.

Оптическая спектроскопия распыленных частиц (ОСРЧ). Кроме вторичных ионов, в потоке, распыленном под воздействием первичных ионов, присутствуют также возбужденные атомы из поверхностных слоев материала. Переход возбужденных атомов в стационарное состояние приводит к возникновению электромагнитного излучения, спектр которого простирается от инфракрасной до ультрафиолетовой области.

Из измеренных спектров излучения определяют элементный состав поверхности и концентрационные профили примесей по глубине. По чувствительности метод ОСРЧ близок к ВИМС.

Ионная оже-спектроскопия (ИОС) основана на изучении энергетического спектра вторичных электронов, возникающих при бомбардировке поверхности твердого тела ускоренными ионами. Спектр вторичных электронов простирается до 100 эВ и в области 10–100 эВ содержит тонкую структуру, обусловленную эмиссией оже-электронов. Для проведения анализа обычно используют легкие ионы H^+ или He^+ с энергией до 10 кэВ.

Как и ЭОС, данный метод применяют для определения элементного состава поверхности. Он менее чувствителен и более трудоемок, чем ЭОС, но имеет преимущество при изучении профилей элементов по глубине, поскольку ионная бомбардировка обеспечивает непрерывное распыление материала поверхности.

Резерфордское обратное рассеяние (РОР) основано на обратном рассеянии потока быстрых ионов, взаимодействующих с поверхностью твердого тела и имеющих энергию в интервале 0,1–5 МэВ. Рассеяние первичных ионов происходит вследствие их упругого взаимодействия с атомами тела. В энергетическом

спектре рассеянных ионов содержится информация о массе атомов исследуемого твердого тела.

Метод РОР при анализе тонких пленок позволяет определить как их состав, так и толщину. Наиболее высокой чувствительностью метод обладает в применении к анализу тяжелых элементов, достигая значений $10^{13} - 10^{14} \text{ см}^{-2}$.

Ионно-рентгеновская спектроскопия (ИРС) основана на возбуждении характеристического рентгеновского излучения при бомбардировке поверхности твердого тела ускоренными ионами с энергией 0,1–10 МэВ (либо легкими ионами H^+ , He^+ , либо более тяжелыми ионами Ne^+ , Ar^+). Полученные энергетические спектры рентгеновского излучения позволяют определить элементный состав поверхности с чувствительностью около 0,01%.

Диагностирование электромагнитным излучением

Метод рентгеновской фотоэлектронной спектроскопии (РФЭС) основан на возбуждении электронов внутренних оболочек атома под воздействием рентгеновского излучения с энергией фотонов 1,2–1,5 кэВ. Релаксация возбужденного состояния приводит к эмиссии электрона в результате ожэ-перехода.

По фотоэлектронным спектрам находят элементный состав поверхности с чувствительностью $10^{-8} - 10^{-6} \text{ г}$ при точности до 5–10%. Химические сдвиги РФЭС-пиков свидетельствуют о валентных состояниях поверхностных атомов соединений. Метод позволяет изучать адсорбированные частицы на поверхности.

Метод ультрафиолетовой фотоэлектронной спектроскопии (УФЭС), в отличие от РФЭС, использует возбуждающее излучение с меньшей энергией фотона (до 40–50 эВ). Этой энергии достаточно только для инициирования внешнего фотоэффекта вследствие возбуждения валентных электронов.

По фотоэлектронным спектрам для прямых межзонных переходов изучают зонную структуру твердого тела. Адсорбированные частицы изменяют электронный спектр поверхности, что проявляется в виде тонкой структуры, анализ которой позволяет изучать процесс адсорбции.

Метод УФЭС более чувствителен к состоянию поверхности и имеет более высокое энергетическое разрешение, чем РФЭС, но не позволяет определять химический состав твердого тела.

Метод инфракрасной спектроскопии (ИКС) основан на измерении молекулярных спектров поглощения и отражения ИК-излучения. При этом ИК-область условно разделяют на ближнюю ($\lambda = 0,74 - 2,5 \text{ мкм}$), среднюю ($\lambda = 2,5 - 50 \text{ мкм}$) и дальнюю

($\lambda = 50 - 2000$ мкм). В ИК-диапазоне расположены колебательные и вращательные спектры молекул.

Поглощение ИК-излучения происходит на частотах, совпадающих с частотами колебания атомов в молекуле и частотами вращения молекул. Поэтому по ИК-спектрам поглощения и отражения определяют структуру молекул, их химический состав и влияние адсорбции. Чувствительность метода, определяемая измерительной аппаратурой, может быть заметно увеличена при использовании многократного отражения.

Метод эллипсометрии (ЭМ) основан на изменении поляризации монохроматического излучения при его отражении от границ раздела сред. Этим методом определяют оптические константы и толщину полупроводниковых и диэлектрических пленок. Чувствительность метода достигает единиц нанометра.

Диагностирование быстрыми электронами

Метод растровой электронной микроскопии (РЭМ) основан на вторичной электронной эмиссии, порожденной взаимодействием быстрых электронов (с энергией более 10 кэВ) с поверхностью твердого тела. Регистрируются вторичные электроны с энергией не более 50 эВ, которые созданы в поверхностном слое толщиной 1–5 нм. Ток вторичных электронов зависит от угла падения первичных электронов на поверхность тела. Зондируя участок поверхности остросфокусированным пучком электронов и разворачивая его в растр, получают в каждой точке ток вторичных электронов. Этим сигналом модулируется яркость монитора, давая изображение поверхности в черно-белом контрасте. С помощью РЭМ наблюдают структуры с разрешением до 3 нм.

Метод дифракции быстрых электронов (ДБЭ) использует явление дифракции электронов с энергией 30–1000 эВ на атомах твердого тела. Метод применяется для исследования строения поверхностных слоев вещества и называется электронографией. По характеру получаемых электронограмм обычно изучают кристаллическую структуру материала.

Методы полевой микроскопии

Полевая электронная микроскопия (ПЭМ) основана на явлении эмиссии электронов с холодного металлического катода под действием электрического поля. Из исследуемого металла изготавливают тонкое острие с радиусом кривизны 50–300 нм. При потенциале на острие $10^3 - 10^4$ В возникает эмиссия электронов, которые создают на флуоресцентном экране атомное изображение структуры поверхности катода. Этот метод позволяет оце-

нить работу выхода электрона из металла и распределение электронов вблизи поверхности Ферми с точностью до 15–75 мэВ.

Полевая ионная микроскопия (ПИМ), в отличие от ПЭМ, использует тонкое острие, изготовленное из исследуемого металла, в качестве анода. В основе метода лежит ионизация молекул инертного газа (обычно гелия) на поверхности металлического анода в сильном электрическом поле за счет туннелирования валентных электронов в металл. Ионы газа, ускоряясь в поле напряженностью 10^8 – 10^9 В/см, создают на флуоресцентном экране атомное изображение структуры поверхности анода.

Методы радиоспектроскопии

Метод ядерного магнитного резонанса (ЯМР) основан на резонансном поглощении веществом, помещенным в магнитное поле, энергии радиоизлучения в диапазоне 5–50 МГц. Поглощение происходит на частоте генератора, равной частоте прецессии ядерного спина атома в постоянном магнитном поле. При резонансном поглощении спин ядра изменяется на противоположный.

Метод электронного парамагнитного резонанса (ЭПР) основан на резонансном поглощении энергии СВЧ (в диапазоне частот 10–40 ГГц) атомами, молекулами или кластерами, имеющими неспаренный электронный спин, который изменяет свою ориентацию в результате поглощения СВЧ-излучения.

Методы радиоспектроскопии используются для анализа химического состава веществ.

Методы зондовой микроскопии

Сканирующая туннельная микроскопия (СТМ) основана на измерении туннельного тока между металлическим зондом с острием атомных размеров и исследуемой поверхностью металла. Принцип и режимы работы СТМ кратко изложены на с. 57–58.

Атомно-силовая микроскопия (АСМ), в отличие от СТМ, пригодна для исследования не только металлических, но и непроводящих поверхностей. Принцип и режимы работы АСМ кратко изложены на с. 58–59.

Методы зондовой микроскопии применяются для получения топографического изображения поверхности твердых тел с атомным разрешением. Их практическая ценность в том, что они дают непосредственную картину расположения атомов (правда, без идентификации их типа) с точностью около $0,1 \text{ \AA}$, что примерно на порядок ниже, чем при ДМЭ. Однако интерпретация дифракционных результатов требует привлечения модели структуры.

Завершая раздел введения, посвященный краткому описанию наноматериалов, нанотехнологий и нанодиагностики, приведем список литературы, в которой можно найти более подробное изложение затронутых вопросов.

- *Елецкий А.В., Смирнов Б.М.* Фуллерены и структуры углерода. — *Успехи физических наук*, т. 165, №9, с. 977, 1995.
- *Елецкий А.В.* Углеродные нанотрубки. — *Успехи физических наук*, т. 167, №9, с. 945, 1997.
- *Елецкий А.В.* Эндоэдральные структуры. — *Успехи физических наук*, т. 170, №2, с. 113, 2000.
- *Суздаев И.П.* Нанотехнология: физико-химия нанокластеров, наноструктур и наноматериалов. — М.: КомКнига, 2005.
- *Гусев А.И.* Наноматериалы, наноструктуры, нанотехнологии. — М.: Физматлит, 2005.
- Нанотехнологии в электронике. Под ред. *Чаплыгина Ю.А.* — М.: Техносфера, 2005.
- Нанотехнология: физика, процессы, диагностика, приборы. Под ред. *Лучинина В.В., Таирова Ю.М.* — М.: Физматлит, 2006.
- *Пул Ч., Оуэнс Ф.* Нанотехнологии. — М.: Техносфера, 2006.
- *Щука А.А.* Наноэлектроника. — М.: Физматкнига, 2007.
- *Рыжонков Д.И., Левина В.В., Дзидзигури Э.Л.* Наноматериалы. — М.: Бином. Лаборатория знаний, 2008.
- *Старостин В.В.* Материалы и методы нанотехнологии. — М.: Бином. Лаборатория знаний, 2008.
- *Кобаяси Н.* Введение в нанотехнологию. — М.: Бином. Лаборатория знаний, 2008.

В дополнение к этому списку литература, имеющая общую физическую, физико-химическую и физико-технологическую направленность, отчасти включающая и вопросы нанотехнологии, приведена в списке рекомендуемой литературы в конце книги.

Глава 1

ОБЩИЕ ПРИНЦИПЫ ТЕРМОДИНАМИЧЕСКОГО УПРАВЛЕНИЯ РАВНОВЕСНЫМИ И НЕРАВНОВЕСНЫМИ ПРОЦЕССАМИ

Термодинамика зародилась в конце 17 века на основе интуитивных представлений о теплоте как некоей неразрушимой субстанции — калорической жидкости под названием «теплород» («флюид тепла» Роберта Бойля, 1673 г.) или «флогистон» («флюид горения» Георга Шталя, 1697 г.). Научный подход к термодинамике начинается лишь в 19 веке с работ Карно¹⁾ и Клаузиуса²⁾. В течение последующих ста лет термодинамика развивалась трудами многочисленных исследователей как аксиоматическое направление в изучении равновесных (обратимых) процессов, базирующееся на трех постулатах, называемых *началами термодинамики*. Они феноменологически выражают общие

¹⁾ Карно (Carnot) Николя Леонар Сади (1796–1832) — французский военный инженер, автор знаменитой теоремы Карно, опубликовал результаты изучения теплообмена, расширения газов и механического применения пара в единственной работе «Размышление о движущей силе огня и о машинах, способных развивать эту силу» (1824). Здесь впервые введены основные понятия термодинамики, такие как идеальная тепловая машина, идеальный цикл, обратимость и необратимость термодинамических процессов. Умер Карно от холеры, поэтому все его имущество было сожжено; уцелела только одна записная книжка, где сформулировано первое начало термодинамики.

²⁾ Клаузиус (Clausius — псевдоним, настоящая фамилия — Gottlieb) Рудольф Юлиус Эммануэль (1822–1888) — немецкий физик, один из основателей термодинамики и молекулярно-кинетической теории теплоты, дал одновременно с У. Томсоном (лордом Кельвином) (1824–1907) первую формулировку второго начала термодинамики (1850), ввел понятия энтропии (1865), идеального газа и длины свободного пробега молекул. Обосновал так называемое уравнение Клапейрона–Клаузиуса, разработал теорию поляризации диэлектриков (формула Клаузиуса–Моссотти) и сформулировал гипотезу о «тепловой смерти Вселенной», развенчанную позднее трудами Больцмана.

равновесные закономерности, наблюдаемые в природе, без выяснения их микроскопической основы.

Статистическое обоснование равновесной термодинамики появилось впервые в работах Больцмана¹⁾, заложившего основы статистического подхода в классической физике. Именно он впервые дал статистическую интерпретацию термодинамического определения энтропии, введенной Клаузиусом, и вывел кинетическое уравнение, носящее его имя, которое описывает установление равновесного состояния. Окончательное построение статистической механики в ее современном виде было выполнено в трудах Гиббса²⁾, который ввел в рассмотрение статистические ансамбли и канонические распределения.

Термодинамика равновесных процессов, которая сформировалась к началу 20 века, своим названием, в отличие от гидродинамики и электродинамики, не отражала общепринятого смысла динамики как науки о движении материальной среды и полей под действием внешних сил. Такое название терминологически означает не более, чем связь между теплотой («термо») и механической работой («динамика») в ходе обратимых процессов. Поскольку при этом не рассматривается развитие процессов во времени, то правильнее было бы называть такое описание «термостатикой», что и предлагалось многими авторами. Однако за традиционным направлением, сформированным в окончательном

¹⁾ Больцман (Boltzmann) Людвиг (1844–1906) — австрийский физик, основатель молекулярно-кинетической теории газов и классической статистической механики, на основе идей Дж. Максвелла (1831–1879) вывел распределение газовых молекул по скоростям (статистика Максвелла–Больцмана) (1862) и получил кинетическое уравнение (1872), являющееся основой современной физической кинетики. Доказал статистический характер второго начала термодинамики (*H*-теорема Больцмана), связав энтропию физической системы с вероятностью ее макросостояния (1872), чем был нанесен удар по гипотезе Клаузиуса о «тепловой смерти Вселенной». Впервые применив к излучению принципы термодинамики, теоретически открыл закон теплового излучения (1884), экспериментально установленный ранее (1879) Стефаном (закон Стефана–Больцмана).

²⁾ Гиббс (Gibbs) Джозайя Уиллард (1839–1903) — американский физик и физико-химик, один из основоположников химической термодинамики и статистической физики, разработал теорию термодинамических потенциалов и их флуктуаций (1875–1878), введя понятие химического потенциала, изучил условия равновесия и термодинамической устойчивости гетерогенных систем и сформулировал правило расчета степеней свободы гетерогенной системы (правило фаз Гиббса). Заложил основы термодинамики поверхностных явлений (адсорбционное уравнение Гиббса) и электрохимических процессов (1874–1878). Публикацией книги «Основные принципы статистической механики» Гиббс завершил (1902) создание классической статистической физики.

виде Больцманом и Гиббсом, сохранилось название равновесной термодинамики, в отличие от термодинамики неравновесных процессов, признанными создателями которой являются нобелевские лауреаты Ларс Онсагер¹⁾ и Илья Пригожин²⁾.

Работы Онсагера явились первым шагом на пути расширения термодинамических представлений на неравновесные физические ситуации. Несмотря на то, что знаменитые соотношения взаимности Онсагера имеют достаточно общий характер (не зависящий от микроскопических гипотез), их применимость ограничена линейной областью отклонения от состояния равновесия. Однако именно работы Онсагера ознаменовали поворотный момент в смещении научного интереса от равновесия к неравновесию.

Ключевая роль здесь принадлежит Илье Пригожину, нашему соотечественнику, главе брюссельской школы неравновесной термодинамики. Благодаря его фундаментальным работам коренным образом изменился взгляд на физическую роль энтропии, порождаемой необратимыми процессами и поэтому рассматривавшейся ранее как инструмент для описания беспорядка в равновесных системах. Из-за потока энтропии в открытых системах даже вблизи равновесного состояния необратимость не может быть отождествлена с тенденцией к хаосу. Такая точка зрения явилась для Пригожина исходной при анализе сильно неравновесных систем, в которых необратимость начинает играть конструктивную роль. Необратимые процессы порождают самоорганизацию в сильно неравновесных ситуациях, приводящую к возникновению устойчивых пространственно-временных структур, названных Пригожиным *диссипативными структурами*.

Современная термодинамика различает три типа состояний, процессов и структур:

¹⁾ Онсагер (Onsager) Ларс (1905–1976) — норвежско-американский химик, создатель феноменологической теории необратимых процессов, которые происходят в неравновесных условиях, на основе статистической механики доказал принцип симметрии кинетических коэффициентов (1931). Нобелевская премия по химии (1968) «за открытие соотношений взаимности в необратимых процессах, названных его именем, которые имеют принципиально важное значение для термодинамики необратимых процессов».

²⁾ Пригожин (Prigogine) Илья Романович (1917–2003) — бельгийский физик и физико-химик русского происхождения, специалист в области статистической механики и термодинамики неравновесных процессов, ввел понятия потока энтропии и производства энтропии, предложил принцип локального равновесия, дал локальную формулировку второго начала термодинамики и доказал критерий стационарности в форме минимума производства энтропии. Нобелевская премия по химии (1977) «за работы по термодинамике необратимых процессов, особенно за теорию диссипативных структур».

- *равновесные состояния и процессы* обратимой термодинамики, управляемые **принципом экстремумов** (максимума энтропии в замкнутой системе и минимума термодинамического потенциала типа свободной энергии Гиббса или Гельмгольца в закрытой системе), который формирует равновесные структуры;

- *квазиравновесные состояния и процессы* необратимой линейной термодинамики Онсагера, управляемые **принципом минимума производства энтропии**, который формирует стационарные квазиравновесные структуры;

- *неравновесные состояния и процессы* необратимой нелинейной термодинамики Пригожина, управляемые **универсальным принципом (критерием) эволюции**, который формирует стационарные диссипативные структуры.

Благодаря общности базовых положений, неравновесный термодинамический подход к описанию наблюдаемых явлений дает результаты, применимые к разнообразным макроскопическим системам (физическим, химическим, биологическим и др.).

Задача настоящей главы — рассмотрение фундаментальных положений равновесной и неравновесной термодинамики с единых концептуальных позиций. Это важно для понимания разнообразных физико-технологических проблем, возникающих в современной электронике в применении к широкому классу различных приборов как дискретного, так и интегрального исполнения. Методический отбор материала для этой главы, как, впрочем, и для последующих глав, определяется применимостью его к технологическим задачам электроники, что исключает из рассмотрения ряд вопросов, например циклические процессы в термодинамике тепловых машин и т. п.

В отличие от классического изложения термодинамики, нами за основу взят статистический подход, основанный на больцмановском определении энтропии. В таком изложении известные начала термодинамики, выражающие закон сохранения энергии в открытых системах, критерий самопроизвольности процессов в замкнутых системах и принцип недостижимости абсолютного нуля температуры, появляются в обратном порядке.

1.1. Статистическое определение энтропии

1. Определение энтропии по Больцману. Любая материальная система состоит из большого числа частиц (молекул, атомов, ионов, электронов), движущихся и взаимодействующих между собой по законам классической или квантовой механики.

Количественной мерой числа частиц может служить постоянная Авогадро ¹⁾ $\mathcal{N}_A = 6,022 \cdot 10^{23}$ моль⁻¹, равная числу частиц в одном моле ²⁾ любого вещества. Следовательно, макроскопические системы обладают огромным числом степеней свободы s (так, для системы из N классических бесструктурных частиц, совершающих поступательное движение, имеем $s = 3N$). Это делает практически невозможным чисто механическое описание системы путем задания в любой момент времени обобщенных координат и импульсов в количестве, равном удвоенному числу ее степеней свободы ($2s = 6N$).

Неполнота механического описания приводит к тому, что координаты и импульсы частиц становятся случайными величинами. Поэтому единственно возможным остается лишь вероятностное описание макросистемы с помощью функции статистического распределения $w(\varepsilon)$, выражающей вероятность реализации состояния с энергией ε .

При статистическом описании принято различать микроскопическое и макроскопическое состояния системы.

Микросостояние системы классических частиц определяется набором $3N$ координат и $3N$ импульсов частиц, что соответствует чисто механическому описанию. Для системы квантовых объектов микросостояние описывается с помощью волновых функций и энергетического спектра квантовых состояний.

Макросостояние системы определяется набором ограниченного числа внешних параметров (температуры, объема, внешних полей и др.), задание которых однозначно определяет все макроскопические (усредненные) характеристики системы (внутреннюю энергию, давление, концентрации компонентов и др.), т.е. соответствует неполному механическому описанию. При этом неконтролируемые перемещения частиц, возникающие как

¹⁾ Авогадро (Avogadro) Амедео (1776–1856) — итальянский физик, заложивший основы молекулярной теории, установил атомный состав ряда молекул и открыл закон, согласно которому в равных объемах газов при одинаковых температурах и давлениях содержится одно и то же число молекул.

²⁾ Напоминаем, что *молем* называется количество вещества, масса которого, *выраженная в граммах*, численно равняется его молекулярному весу. Для атомов и ионов величину, аналогичную молю, называют *грамм-атомом* (г-атом) и *грамм-ионом* (г-ион). В соответствии с этой терминологией, моль является *грамм-молекулой* сложного вещества. Количество ионов в растворах электролитов часто измеряют в *грамм-эквивалентах* (г-экв): 1 г-экв равен 1 г-иону, деленному на число z элементарных зарядов иона. Для ионов любого вида 1 г-экв имеет одинаковый заряд $\mathcal{F} = e\mathcal{N}_A = 96\,485$ Кл, называемый постоянной (числом) Фарадея, в то время как 1 г-ион несет заряд $z\mathcal{F} = ze\mathcal{N}_A$.

результат их взаимодействия друг с другом и с внешними рассеивателями, проявляются в виде хаотического теплового движения.

В стационарных условиях макросостояние системы сохраняется, а микросостояния непрерывно изменяются в результате теплового движения частиц. Иными словами, макросостояние реализуется путем перехода системы из одного микросостояния в другое. Это позволяет утверждать, что одному макросостоянию соответствует большое, но конечное число микросостояний (из-за конечности числа степеней свободы). Среднее число микросостояний, посредством которых реализуется данное макросостояние, принято называть **статистическим весом** W (или **термодинамической вероятностью**) этого состояния, а величину

$$S = k_B \ln W \quad (1.1.1)$$

называют **энтропией** системы ($k_B = 1,38 \cdot 10^{-23}$ Дж/К — постоянная Больцмана).

Покажем, что энтропия S , введенная Больцманом в форме (1.1.1), однозначно определяется функцией статистического распределения $w(\varepsilon)$, описывающей распределение квантовых состояний системы по энергии.

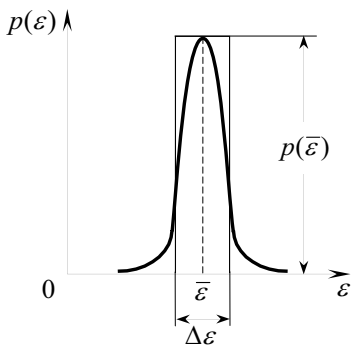


Рис. 1.1. К определению понятия статистического веса и энтропии

Для любой макросистемы энергетический спектр квантовых микросостояний настолько густой, что позволяет ввести в рассмотрение так называемую плотность энергетического спектра $\gamma(\varepsilon) = d\Gamma(\varepsilon)/d\varepsilon$, где $d\Gamma(\varepsilon)$ — число квантовых состояний с энергией, лежащей в интервале от ε до $\varepsilon + d\varepsilon$. Для нахождения вероятности $dP(\varepsilon)$ того, что макросистема обладает энергией, лежащей в указанном интервале, необходимо вероятность $w(\varepsilon)$ реализации микросостояния умножить на число квантовых состояний $d\Gamma(\varepsilon)$, доступных системе в интервале энергий $d\varepsilon$: $dP(\varepsilon) = w(\varepsilon)d\Gamma(\varepsilon) = w(\varepsilon)\gamma(\varepsilon)d\varepsilon$. Отсюда получаем плотность вероятности,

$$p(\varepsilon) \equiv \frac{dP(\varepsilon)}{d\varepsilon} = w(\varepsilon)\gamma(\varepsilon),$$

удовлетворяющую условию нормировки

$$\int p(\varepsilon) d\varepsilon = \int w(\varepsilon)\gamma(\varepsilon) d\varepsilon = 1. \quad (1.1.2)$$

Пусть рассматриваемая система, содержащая N частиц, имеет в определенных условиях среднюю энергию $\bar{\varepsilon}$, а ее мгновенные значения ε флуктуируют вследствие теплового движения около этой величины. Однако вероятность заметных отклонений от $\bar{\varepsilon}$ чрезвычайно мала ($\propto 1/N^{1/2} \leq 10^{-12}$ при $N \geq N_A$). Это означает, что функция $p(\varepsilon)$ имеет узкий максимум вблизи значения $\varepsilon = \bar{\varepsilon}$, показанный на рис. 1.1. Если построить равновеликий по площади прямоугольник высотой $p(\bar{\varepsilon})$, то его ширина $\Delta\varepsilon$ определит средний интервал энергий, доступных данной системе.

Из условия нормировки (1.1.2), переписанного с учетом рис. 1.1 в форме $p(\bar{\varepsilon})\Delta\varepsilon = w(\bar{\varepsilon})\gamma(\bar{\varepsilon})\Delta\varepsilon = 1$, находим среднее число $\gamma(\bar{\varepsilon})\Delta\varepsilon$ микросостояний, которое соответствует данному макросостоянию со средней энергией $\bar{\varepsilon}$. Тогда статистический вес (или термодинамическая вероятность) макросостояния равен

$$W(\bar{\varepsilon}) \equiv \Delta\Gamma(\bar{\varepsilon}) = \left. \frac{d\Gamma(\varepsilon)}{d\varepsilon} \right|_{\varepsilon=\bar{\varepsilon}} \Delta\varepsilon = \gamma(\bar{\varepsilon}) \Delta\varepsilon = \frac{1}{w(\bar{\varepsilon})}. \quad (1.1.3)$$

В дальнейшем среднюю энергию $\bar{\varepsilon}$, называемую *внутренней энергией* системы, будем общепринято обозначать буквой U .

Подставляя равенство (1.1.3) в формулу (1.1.1), окончательно получаем следующее выражение для энтропии по Больцману:

$$k_B \ln W(\bar{\varepsilon}) \equiv k_B \ln W(U) = S(U), \quad (1.1.4, a)$$

или

$$-k_B \ln w(\bar{\varepsilon}) \equiv -k_B \ln w(U) = S(U). \quad (1.1.4, б)$$

Таким образом, знание функции распределения микросостояний по энергии позволяет с помощью формулы (1.1.4, б) найти энтропию системы. В равновесных условиях функция распределения имеет вид распределения Гиббса: $w(\varepsilon) = A \exp(-\varepsilon/k_B T)$.

2. Сравнение с информационной энтропией. Для систем с конечным дискретным числом (W) состояний Шеннон¹⁾ определил информационную энтропию выражением

$$H = - \sum_{k=1}^W p_k \log_b p_k, \quad (1.1.5)$$

где p_k — вероятность k -го состояния, при этом $\sum_k p_k = 1$.

¹⁾ Шеннон (Shannon) Клод Элвуд (1916–2001) — американский инженер и математик, один из основоположников современной кибернетики и математической теории информации и связи, впервые определил количество информации через энтропию, ввел единицу информации, впоследствии названную «битом», и доказал теорему, устанавливающую предельную скорость передачи информации через зашумленный канал связи, называемую пределом Шеннона.

Величина H трактуется как *мера информации*, получаемой при выяснении того, какое именно состояние реализовалось; при этом обычно выбирают $b = 2$ и выражают информацию в битах. Для частного случая $p_k = 1$ (достоверность появления k -го состояния) и $p_l = 0$ ($l \neq k$) получаем $H = 0$, т. е. информация равна нулю, если результат заранее известен.

Другой частный случай соответствует системе с W равновероятными состояниями, когда для каждого состояния имеем $p_k = 1/W$. В этом случае количество информации, получаемой при выяснении состояния системы, максимально и равно

$$H = \log_b W. \quad (1.1.6)$$

Из сравнения выражений (1.1.1) и (1.1.6) видно, что больцмановская и информационная энтропии пропорциональны друг другу:

$$S = cH, \quad \text{где } c = k_B \ln b. \quad (1.1.7)$$

Соотношение (1.1.7) было получено для системы с равновероятными состояниями, что соответствует, как увидим ниже, частному случаю *замкнутой* (полностью изолированной) физической системы, в которой отсутствуют всяческие контакты с внешним окружением. Однако смысловое совпадение энтропии по Больцману с энтропией по Шеннону имеет место и при более общих условиях *квазизамкнутости*. С целью обоснования этого утверждения приведем больцмановскую энтропию (1.1.4, б) к форме, аналогичной (1.1.5).

Любая *замкнутая* макросистема, будучи изначально возмущенной, после снятия возмущения самопроизвольно возвращается в равновесие (называемое *глобальным* или *полным равновесием*) за некоторое время релаксации $\tau_{\text{гл}}$. Отдельные части этой системы, рассматриваемые как макроскопические подсистемы, слабо взаимодействуют между собой через посредство «приграничных» частиц, число которых много меньше числа «объемных» частиц. Поэтому энергией взаимодействия между подсистемами можно пренебречь по сравнению с полной внутренней энергией, аддитивно составленной из внутренних энергий подсистем, т. е. для системы, состоящей из двух подсистем, имеем $\varepsilon = \varepsilon_1 + \varepsilon_2$.

Слабое взаимодействие позволяет считать отдельные подсистемы *квазизамкнутыми* в том смысле, что каждая из них восстанавливает свое равновесие (называемое *локальным* или *частичным квазиравновесием*) за времена $\tau_{\text{лок}}$, существенно меньшие $\tau_{\text{гл}}$. На этом основании подсистемы можно считать статистически независимыми на интервалах времени, меньших $\tau_{\text{гл}}$, но

бóльших $\tau_{\text{лок}}$. Следовательно, состояние одной подсистемы не влияет на вероятность различных состояний другой подсистемы.

Тогда для функции распределения имеем $w(\varepsilon) = w(\varepsilon_1)w(\varepsilon_2)$ и $\ln w(\varepsilon) = \ln w(\varepsilon_1) + \ln w(\varepsilon_2)$. Одновременная аддитивность двух величин, ε и $\ln w$, позволяет представить $\ln w(\varepsilon)$ в линейной форме: $\ln w(\varepsilon) = a + b\varepsilon$. В этом случае выражение (1.1.4, б) для больцмановской энтропии принимает следующий вид:

$$S = -k_B \ln w(\bar{\varepsilon}) = -k_B \overline{\ln w(\varepsilon)}. \quad (1.1.8)$$

Если вероятность реализации k -го состояния (из W возможных) равна w_k , то за *меру неопределенности* этого состояния принимают величину $-\ln w_k$. Действительно, для достоверного состояния (при $w_k = 1$) неопределенность равняется нулю, а при $w_k = 0$ превращается в бесконечность.

Следовательно, выражение (1.1.8) может быть статистически истолковано как математическое ожидание неопределенности:

$$S = -k_B \overline{\ln w(\varepsilon)} \equiv k_B M\{-\ln w\} = -k_B \sum_{k=1}^W w_k \ln w_k. \quad (1.1.9)$$

Структура выражения (1.1.5) для информационной энтропии по Шеннону совпадает со статистической энтропией (1.1.1) по Больцману, но представленной в форме (1.1.9). Последняя интерпретируется как *мера неупорядоченности* и вместе с тем как *мера структурной организованности* системы. При этом максимум статистической энтропии (при равновероятности состояний) соответствует низшей степени организации в системе.

1.2. Свойства энтропии. Статистический смысл температуры

1. Свойство положительности энтропии. Из определения статистического веса как среднего числа микросостояний, соответствующих данному макросостоянию, имеем $W \geq 1$. Следовательно, согласно больцмановскому определению (1.1.1), энтропия является неотрицательной величиной ($S \geq 0$). Обращение энтропии в нуль возможно лишь при $W = 1$. Физически такую ситуацию можно представить лишь при стремлении температуры к абсолютному нулю, когда «вымораживаются» все степени свободы системы и она переходит в единственно возможное основное квантовое состояние. Это утверждение, основанное на

анализе экспериментальных данных, принято записывать в форме **третьего начала термодинамики**,

$$S \rightarrow 0 \quad \text{при} \quad T \rightarrow 0; \quad (1.2.1)$$

впервые сформулировано Нернстом¹⁾ и известного как **тепловая теорема Нернста** (называемая также *постулатом Планка*).

2. Свойство аддитивности энтропии. Разделим рассматриваемую систему на две подсистемы со своими статистическими весами W_1 и W_2 . Так как комбинация микросостояния одной подсистемы с любым микросостоянием другой представляет собой микросостояние общей системы, то ее статистический вес и энтропия, соответственно, равны $W = W_1 W_2$ и $S = S_1 + S_2$. На этом основании в общем случае записываем суммарную энтропию системы в виде

$$S(U) = \sum_n S_n(U_n), \quad \text{где} \quad U = \sum_n U_n. \quad (1.2.2)$$

Следовательно, энтропия сложной системы, как совокупности входящих в нее подсистем, равна сумме энтропий отдельных подсистем, каждая из которых зависит от собственной внутренней энергии, которая, в свою очередь, также является аддитивной величиной для сложной системы.

3. Закон возрастания энтропии замкнутой системы. Определим **замкнутую систему** как полностью изолированную систему, лишенную каких-либо взаимодействий с окружающей средой, то есть не обменивающуюся с ней ни веществом, ни энергией (в форме тепла или работы). В результате внутренняя энергия U замкнутой системы остается неизменной.

Согласно определению (1.1.1), энтропия возрастает с увеличением числа микросостояний, с помощью которых может быть реализовано данное макросостояние системы. Увеличение числа микросостояний означает рост хаоса в системе, при котором ее энергия (в условиях сохранения) перераспределяется между степенями свободы системы равновероятным образом. Иными словами, полностью изолированная система стремится охватить все доступные ей микросостояния, и ее энтропия подчиняется

¹⁾ Нернст (Nernst) Вальтер (1864–1941) — немецкий физико-химик, работавший в области теории растворов, химической кинетики, электрохимии и общих вопросов термодинамики, где сформулировал третий постулат, известный как теорема Нернста (1906), и принцип недостижимости абсолютного нуля температуры. Нобелевская премия по химии (1920) «в признание его работ по термодинамике».

так называемому *закону возрастания энтропии*, записанному, с учетом свойства аддитивности (1.2.2), в виде

$$dS(U) = \sum_n dS_n(U_n) > 0 \quad \text{при} \quad U = \sum_n U_n = \text{const.} \quad (1.2.3)$$

Закон возрастания энтропии гласит: *в замкнутой системе неравновесные процессы самопроизвольно протекают в направлении возрастания энтропии, стремясь перевести систему в равновесное состояние, которому соответствует максимальное значение энтропии.*

Формула (1.2.3) выражает **второе начало термодинамики**, имеющее несколько различных формулировок, отличных от приведенной выше, в частности, как обоснование невозможности создания вечного двигателя второго рода.

Из соотношения (1.2.3) следует, что в неравновесных условиях требование возрастания суммарной энтропии S позволяет отдельным ее частям S_n возрастать, убывать или оставаться неизменными. Нетрудно понять, что любой *неравновесный процесс* является **необратимым**, поскольку для обратного его протекания суммарная энтропия замкнутой системы должна убывать, что запрещается законом возрастания энтропии. Иными словами, при неравновесных процессах в замкнутой системе происходят такие необратимые изменения, что систему нельзя вернуть в исходное состояние, не приложив к ней внешних воздействий, то есть не сделав ее незамкнутой. Тогда в новых изменившихся условиях старая система может быть возвращена в исходное для нее состояние с уменьшенной энтропией, являясь при этом лишь частью новой системы, расширенной за счет источника внешнего воздействия. Суммарная энтропия новой замкнутой системы будет при этом неизбежно возрастать.

В состоянии *термодинамического равновесия* энтропия замкнутой системы максимальна и неизменна, тогда

$$dS(U) = \sum_n dS_n(U_n) = 0 \quad \text{при} \quad U = \sum_n U_n = \text{const.} \quad (1.2.4)$$

Это означает, что *равновесные процессы* в системе являются **обратимыми**, поскольку их протекание как в прямом, так и в обратном направлениях не приводит к изменению суммарной энтропии S (хотя отдельные ее части S_n могут изменяться).

4. Статистический смысл температуры. Дадим статистическое определение температуры, доказав, что

$$\frac{1}{T} = \frac{dS(U)}{dU}, \quad (1.2.5)$$

то есть обратная температура определена как скорость изменения энтропии при изменении внутренней энергии системы.

Рассмотрим замкнутую систему, состоящую из двух частей, между которыми происходит обмен энергией. В таком случае $U = U_1 + U_2 = \text{const}$, так что $dU_1 = -dU_2$. При этом энтропию системы $S(U) = S_1(U_1) + S_2(U_2)$ можно считать зависящей лишь от U_1 , так как при $U = \text{const}$ имеем $U_2 = U - U_1$.

Для **равновесных условий**, используя выражения (1.2.4) и (1.2.5), записываем

$$\frac{dS}{dU_1} = \frac{dS_1}{dU_1} + \frac{dS_2}{dU_2} \frac{dU_2}{dU_1} = \frac{dS_1}{dU_1} - \frac{dS_2}{dU_2} = 0$$

или $\frac{1}{T_1} = \frac{1}{T_2}$.

(1.2.6)

Для **неравновесных условий**, когда процесс изменяется во времени, согласно закону возрастания энтропии (1.2.3) и определению (1.2.5), имеем

$$\frac{dS}{dt} = \frac{dS_1}{dU_1} \frac{dU_1}{dt} + \frac{dS_2}{dU_2} \frac{dU_2}{dt} = \left(\frac{dS_1}{dU_1} - \frac{dS_2}{dU_2} \right) \frac{dU_1}{dt} > 0$$

или $\frac{1}{T_1} > \frac{1}{T_2}$ при $\frac{dU_1}{dt} > 0$ и $\frac{dU_2}{dt} < 0$.

(1.2.7)

Как видно из формул (1.2.6) и (1.2.7), величина T , введенная равенством (1.2.5), по физическому смыслу соответствует абсолютной температуре: при равновесии, согласно (1.2.6), температуры двух тел равны ($T_1 = T_2$), а в его отсутствие энергия (в форме тепла) самопроизвольно переходит от тела с высокой температурой к телу с низкой температурой. Действительно, при $T_2 > T_1$ из неравенств (1.2.7) получаем $dU_2/dt < 0$ и $dU_1/dt > 0$, т. е. энергия нагретого тела уменьшается, а охлажденного — увеличивается.

Следовательно, абсолютная температура термодинамической системы, введенная равенством (1.2.5), определяется обратной производной от энтропии по внутренней энергии. Это означает, что обратная температура тела с энергией U на кривой $S(U)$ графически соответствует тангенсу угла наклона касательной к этой кривой в точке $U = \text{const}$ для рассматриваемого тела. Следует иметь в виду, что температура, как и энтропия, имеет статистический смысл, справедливый только для макроскопических систем, состоящих из большого числа частиц, и неприменима к отдельным микроскопическим объектам.

1.3. Закон сохранения энергии в открытых системах

Из предыдущего рассмотрения следует, что замкнутая система характеризуется внутренней энергией U , статистическим весом $W(U)$ (или энтропией $S(U) = k_B \ln W(U)$) и температурой $T = (dS/dU)^{-1}$. Из-за отсутствия контактов с внешней средой внутренняя энергия в этом случае сохраняется ($dU = 0$), а энтропия может либо возрастать ($dS > 0$) для неравновесных процессов, либо оставаться неизменной ($dS = 0$) при равновесии.

Модель замкнутой (полностью изолированной) системы является в определенной степени абстракцией, нередко лишенной смысла для реальных систем. В обычных условиях любое тело, обладая тем или иным запасом внутренней энергии и количеством материальных частиц, может обмениваться ими с окружающей средой, являясь в этом смысле **открытой системой**.

Пусть рассматриваемая система с энергией U имеет энтропию S , объем V и, будучи многокомпонентной, содержит частицы разного сорта в количествах $N_1, N_2, \dots, N_K$. Тогда ее состав может быть охарактеризован числом молей $n_1 = N_1/N_A$, $n_2 = N_2/N_A$, $\dots$, $n_K = N_K/N_A$, где K — число компонентов в системе, N_A — постоянная Авогадро. Эти величины в результате взаимодействия открытой системы с внешним окружением получают приращения dS , dV и dn_i ($i = 1, 2, \dots, K$), вносящие аддитивные вклады в изменение dU внутренней энергии системы, что показано на рис. 1.2.

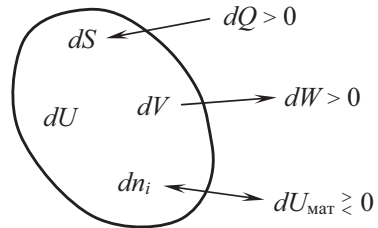


Рис. 1.2. К выводу закона сохранения энергии в открытой системе

Существуют следующие типы контактов, приводящие к трем видам обмена энергией:

- **тепловой контакт**, осуществляемый в результате обмена тепловой энергией: $dU_{\text{теп}} \equiv dQ$;
- **механический контакт**, осуществляемый путем совершения механической работы за счет внешних силовых воздействий: $dU_{\text{мех}} \equiv -dW$ (здесь W , от англ. *Work*, не следует путать с ранее введенным статистическим весом макросостояния, определяющим его энтропию в виде $S = k_B \ln W$);
- **материальный контакт**, осуществляемый в форме обмена материальными частицами (молекулами, атомами, ионами, электронами) с изменением энергии на величину $dU_{\text{мат}}$.

Следовательно, закон сохранения энергии для открытой системы может быть записан в дифференциальной форме:

$$dU = dU_{\text{теп}} + dU_{\text{мех}} + dU_{\text{мат}} \equiv dQ - dW + dU_{\text{мат}}. \quad (1.3.1)$$

Согласно равенству (1.3.1), положительной теплотой ($dQ > 0$) принято считать тепло, получаемое системой от посторонних источников, а положительной работой ($dW > 0$) — работу, производимую системой над внешней средой. При этом первое (dQ) дает положительный вклад в изменение внутренней энергии системы, а второе (dW) — отрицательный вклад.

Естественно, в любой системе, кроме процессов обмена энергией и веществом с внешней средой, протекают также характерные для нее внутренние процессы, которые, согласно законам сохранения энергии и массы, не вносят вклада в выражение (1.3.1). Их роль будет учтена ниже (см. п. 1.22) при рассмотрении диссипативных процессов в неравновесных условиях.

В зависимости от характера изоляции системы от внешней среды принята следующая классификация систем:

$$\begin{array}{ll} dQ = 0, & dW = 0, & dU_{\text{мат}} = 0 & \left\{ \begin{array}{l} \text{замкнутая система} \\ \text{(полная изоляция);} \end{array} \right. \\ dQ = 0, & dW \neq 0, & dU_{\text{мат}} \neq 0 & \left\{ \begin{array}{l} \text{адиабатическая система} \\ \text{(тепловая изоляция);} \end{array} \right. \\ dQ \neq 0, & dW \neq 0, & dU_{\text{мат}} = 0 & \left\{ \begin{array}{l} \text{закрытая система} \\ \text{(материальная изоляция);} \end{array} \right. \\ dQ \neq 0, & dW \neq 0, & dU_{\text{мат}} \neq 0 & \left\{ \begin{array}{l} \text{открытая система} \\ \text{(отсутствие изоляции).} \end{array} \right. \end{array}$$

Открытые системы обеспечивают обмен между рассматриваемой системой и внешней средой как энергией (в форме передачи тепла, $dQ \neq 0$, и совершения работы, $dW \neq 0$), так и веществом (в форме потоков частиц, $dU_{\text{мат}} \neq 0$).

Раскроем физическое содержание каждого слагаемого, характеризующего тип контакта в равенстве (1.3.1).

В предыдущем параграфе был рассмотрен теплообмен между двумя телами в отсутствие механического и материального контактов, т. е. при $dU_{\text{мех}} = dU_{\text{мат}} = 0$. На основании статистического определения обратной температуры в форме (1.2.5) для теплового контакта можно записать

$$dU_{\text{теп}} \equiv dQ = TdS. \quad (1.3.2)$$

Внешние источники (включая в общем случае электромагнитные поля) воздействуют на систему и совершают над ней ра-

боту, равную $-dW$. Для последующего физико-технологического рассмотрения потребуется, в основном, механическая работа,

$$dU_{\text{мех}} \equiv -dW = -PdV, \quad (1.3.3)$$

производимая внешней средой против давления P в системе при ее сжатии ($dV < 0$), где V — объем системы.

Учет электрической работы, $dW_{\text{эл}} = -dU_{\text{эл}} = -\varphi dq$, по переносу элементарного заряда dq через разность потенциалов φ , совершаемой в электрохимических системах, выполнен в п. 2.18. Вклад во внутреннюю энергию за счет работы по изменению на величину dA площади границы фаз с поверхностным натяжением σ ($dU_{\text{пов}} = -dW_{\text{пов}} = \sigma dA$) учтен в гл. 5. При изменении поляризуемости диэлектрика на величину $d\mathbf{P}$ в электрическом поле $\mathbf{E}$ или намагниченности магнетика на величину $d\mathbf{M}$ в поле магнитной индукции $\mathbf{B}$ следует учесть соответствующие вклады: $dU_{\text{диэл}} = -dW_{\text{диэл}} = -\mathbf{E} \cdot d\mathbf{P}$ или $dU_{\text{магн}} = -dW_{\text{магн}} = -\mathbf{B} \cdot d\mathbf{M}$.

Изменение состава системы за счет изменения числа молей каждого сорта i на величину dn_i вносит соответствующий вклад в величину $dU_{\text{мат}}$, так что в общем случае

$$dU_{\text{мат}} = \sum_{i=1}^K \mu_i dn_i. \quad (1.3.4)$$

Здесь коэффициент пропорциональности μ_i называется *химическим потенциалом* частиц i -го сорта, физический смысл которого будет раскрыт ниже (см. п. 1.11).

Подстановка выражений (1.3.2)–(1.3.4) в закон сохранения энергии, записанный в виде (1.3.1), приводит его к форме **фундаментального уравнения Гиббса**,

$$dU = TdS - PdV + \sum_{i=1}^K \mu_i dn_i, \quad (1.3.5)$$

выражающего **первое начало термодинамики**.

Следует иметь в виду, что такие термодинамические величины, как U , S и другие, вводимые ниже (H , F , G — см. п. 1.10), изменяются пропорционально количеству вещества в системе и поэтому называются *экстенсивными величинами*, в отличие от *интенсивных величин* типа T , P , μ_i и т. п., этим свойством не обладающих. Кроме того, экстенсивные величины являются так называемыми **функциями состояния**. Это означает, что их значения однозначно определяются только данным состоянием системы вне зависимости от того, в результате какого процесса (равновесного или неравновесного) система оказалась в данном

состоянии. Дифференциальные равенства типа (1.3.5) всегда содержат слагаемые в виде произведения интенсивной величины на дифференциал экстенсивной переменной.

В отличие от этого, теплота Q и работа W являются **функциями процесса**, поскольку их значения определяются характером процесса, в результате которого был осуществлен переход системы из начального состояния в конечное. Более того, нельзя говорить о количестве энергии, запасенной телом в виде тепла или механической работы, так как она запасается во внутренних степенях свободы тела, состоящего из частиц, совершающих поступательное, вращательное или колебательное движение.

Разделение энергии на тепловую и механическую возможно лишь, когда речь идет об ее изменении: $dU = dQ - dW$ (при $dU_{\text{мат}} = 0$). При этом элементарное приращение внутренней энергии dU (а также и других функций состояния dS , dH , dF , dG) является полным дифференциалом, в отличие от dQ и dW , не являющихся таковыми. На этом основании для элементарного количества теплоты и работы часто вводят иные обозначения, например, δQ и δW . Но мы не будем этого делать и сохраним для них те же обозначения dQ и dW , что и для полных дифференциалов функций состояния, помня, однако, о различии их физического содержания.

1.4. Типы термодинамических процессов

Изменение внешних параметров системы, характеризующих ее связь с внешней средой, нарушает равновесное состояние системы и инициирует протекание термодинамического процесса, стремящегося привести систему к новому равновесию в изменившихся условиях. Процесс самопроизвольного возвращения системы в равновесное состояние называется *релаксацией*. Любой релаксационный процесс характеризуется постоянной времени τ , называемой временем релаксации (см. с. 80).

Если релаксационные процессы в системе настолько быстрые, что для временного интервала Δt , характеризующего скорость изменения внешних условий, имеет место неравенство $\tau \ll \Delta t$, то возникающий при этом термодинамический процесс является медленным и называется *квазистатическим*. Этот термин означает, что при достаточно медленном изменении внешних условий по сравнению с процессами установления равновесия в системе она практически успевает в каждый момент времени отрелаксировать к равновесию и тем самым последовательно проходит через ряд равновесных состояний.

Рассматривая тело и внешнюю среду как в целом единую замкнутую систему, можно применить к ней рассуждения, приведенные в п. 1.2, и сделать заключение о том, что любой квазистатический процесс является *обратимым*. Нарушение условия квазистатичности процесса, когда $\Delta t \leq \tau$, порождает его *необратимость*.

Для осуществления квазистатического процесса термодинамические параметры внешней среды должны изменяться на бесконечно малую величину при изменении условий за время Δt . На этом основании можно утверждать, что любой термодинамический процесс (даже протекающий неравновесно по отношению к рассматриваемому телу) с точки зрения внешней среды протекает квазистатическим образом.

Действительно, внешняя среда (то есть окружающая нас Вселенная) обладает настолько огромным запасом энергии, объемом и количеством разнообразных частиц, что происходящие в ней изменения, вызванные ее взаимодействием с любой реальной системой, такие как

$$\begin{aligned} dU_{\text{теп}}^{\text{cp}} &\equiv dQ^{\text{cp}} = -dQ, \\ dU_{\text{мех}}^{\text{cp}} &\equiv -dW^{\text{cp}} = dW, \quad dU_{\text{мат}}^{\text{cp}} = -dU_{\text{мат}}, \end{aligned} \quad (1.4.1)$$

являются с точки зрения среды бесконечно малыми при любых скоростях протекания процесса.

Следовательно, изменения типа (1.4.1) для внешней среды не приводят к сколько-нибудь заметному изменению ее температуры, давления и химических потенциалов частиц, поэтому можно положить $T^{\text{cp}} = \text{const}$, $P^{\text{cp}} = \text{const}$ и $\mu_i^{\text{cp}} = \text{const}$. Именно это свойство квазистатичности для внешней среды будет использовано ниже в п. 1.6 при рассмотрении энтропии необратимых процессов.

Среди разнообразных термодинамических процессов обычно выделяют процессы, протекающие в условиях постоянства того или иного параметра системы.

При рассмотрении **теплового контакта** исследуемой системы с внешней средой различают:

- а) *адиабатические процессы*, протекающие в условиях идеальной тепловой изоляции системы, когда $dQ = TdS = 0$, в результате чего энтропия системы сохраняется неизменной ($S = \text{const}$);
- б) *изотермические процессы*, протекающие в условиях идеального теплового контакта тела с термостатом, когда тело принимает его температуру $T = T^{\text{cp}} = \text{const}$, при этом $dQ = TdS \neq 0$.

При рассмотрении **механического контакта** исследуемой системы с внешней средой различают:

а) *изохорные процессы*, протекающие в условиях постоянства объема системы $V = \text{const}$, так что $dW = PdV = 0$;

б) *изобарные процессы*, протекающие в условиях постоянства давления в системе, равного давлению внешней среды $P = P^{\text{ср}} = \text{const}$, при этом $dW = PdV \neq 0$.

В реальных технологических процессах трудно обеспечить идеальную тепловую изоляцию системы от внешней среды. Поэтому адиабатические (изэнтропийные) процессы представляют меньший практический интерес по сравнению с изотермическими процессами, происходящими при постоянной температуре, равной температуре термостата. На этом основании в дальнейшем исключим из рассмотрения изэнтропийные процессы.

Изохорные условия обычно осуществляются при проведении тех или иных технологических операций, например в запаянных кварцевых ампулах. Однако на практике технологические процессы более часто проводят в так называемых открытых проточных системах, в которых реализуется изобарное условие путем уравнивания суммарного давления P газообразных компонентов внешним атмосферным давлением ($P = P_{\text{вн}} = 1 \text{ атм}$).

Нередко технологические операции выполняют на вакуумных установках при непрерывной откачке выделяющихся газов, так что могут быть осуществлены как изохорные, так и изобарные условия. В том случае, когда скорость откачки много меньше скорости газовыделения, технологический процесс можно считать протекающим в условиях $V = \text{const}$, где V — объем вакуумной камеры. При обратном соотношении между скоростями откачки и газовыделения имеет место условие $P = \text{const}$, где P — предельное давление вакуумного насоса.

1.5. Тепловой эффект изохорных и изобарных процессов

Изохорные процессы, протекающие при $V = \text{const}$, по определению не сопровождаются механической работой, поскольку $dW = PdV = 0$. Поэтому, согласно (1.3.1), вся теплота dQ_V , подведенная в изохорных условиях к закрытой системе с неизменным составом ($n_i = \text{const}$, $i = 1, 2, \dots, K$), идет на увеличение внутренней энергии системы:

$$dU = TdS \equiv dQ_V \quad \text{при} \quad V = \text{const} \text{ и } n_i = \text{const}. \quad (1.5.1)$$

Для изобарных условий (при $P = \text{const}$) удобно ввести новую термодинамическую функцию $H = U + PV$, называемую **энтальпией**. На основе равенства (1.3.5) получаем соотношение для дифференциала энтальпии:

$$dH = TdS + VdP + \sum_{i=1}^K \mu_i dn_i. \quad (1.5.2)$$

Отсюда следует, что теплота dQ_P , подведенная к телу в изобарном процессе, идет на увеличение его энтальпии, так как

$$dH = TdS \equiv dQ_P \quad \text{при} \quad P = \text{const} \quad \text{и} \quad n_i = \text{const}. \quad (1.5.3)$$

Количество тепла, необходимое для увеличения температуры тела на один градус, называется *теплоемкостью* (которая может быть отнесена к одной частице и одному молю или к единице объема и единице массы; последнюю называют удельной теплоемкостью). Соответственно рассмотренному выше, различают:

- а) *теплоемкость при постоянном объеме* и неизменном составе,

$$C_V = \left(\frac{dQ}{dT} \right)_{V, n_i} = T \left(\frac{\partial S}{\partial T} \right)_{V, n_i} = \left(\frac{\partial U}{\partial T} \right)_{V, n_i}; \quad (1.5.4)$$

- б) *теплоемкость при постоянном давлении* и неизменном составе,

$$C_P = \left(\frac{dQ}{dT} \right)_{P, n_i} = T \left(\frac{\partial S}{\partial T} \right)_{P, n_i} = \left(\frac{\partial H}{\partial T} \right)_{P, n_i}. \quad (1.5.5)$$

Здесь и далее нижние индексы соответствуют величинам, остающимся постоянными при взятии производных.

Установим связь между теплоемкостями C_P и C_V . Для этого запишем уравнение Гиббса (1.3.5) при $n_i = \text{const}$ в виде $dQ \equiv TdS = dU + PdV$, полагая энтропию S и внутреннюю энергию U зависящими от температуры T и объема V :

$$\begin{aligned} dQ &\equiv TdS(T, V) = dU(T, V) + PdV = \\ &= \left(\frac{\partial U}{\partial T} \right)_{V, n_i} dT + \left[\left(\frac{\partial U}{\partial V} \right)_{T, n_i} + P \right] dV. \end{aligned}$$

Отсюда, с учетом определения теплоемкостей в форме (1.5.4) и (1.5.5), вытекают следующие равенства:

$$\left(\frac{\partial S}{\partial V} \right)_{T, n_i} = \frac{1}{T} \left[\left(\frac{\partial U}{\partial V} \right)_{T, n_i} + P \right], \quad (1.5.6)$$

$$\begin{aligned}
 C_P - C_V &= \left(\frac{dQ}{dT} \right)_{P, n_i} - \left(\frac{\partial U}{\partial T} \right)_{V, n_i} = \left[\left(\frac{\partial U}{\partial V} \right)_{T, n_i} + P \right] \left(\frac{\partial V}{\partial T} \right)_{P, n_i} = \\
 &= T \left(\frac{\partial S}{\partial V} \right)_{T, n_i} \left(\frac{\partial V}{\partial T} \right)_{P, n_i} = T \left(\frac{\partial P}{\partial T} \right)_{V, n_i} \left(\frac{\partial V}{\partial T} \right)_{P, n_i}, \quad (1.5.7)
 \end{aligned}$$

где в последнем равенстве использовано полученное ниже уравнение (1.10.10), а именно $(\partial S / \partial V)_{T, n_i} = (\partial P / \partial T)_{V, n_i}$.

Из равенств (1.5.7) следует соотношение, известное как **уравнение Гельмгольца**¹⁾ (ср. уравнение (1.10.13)),

$$\left(\frac{\partial U}{\partial V} \right)_{T, n_i} = T \left(\frac{\partial P}{\partial T} \right)_{V, n_i} - P \equiv T^2 \frac{\partial}{\partial T} \left(\frac{P}{T} \right)_{V, n_i}, \quad (1.5.8)$$

из которого находят изменение внутренней энергии системы при изменении объема, если известно уравнение состояния среды.

В частности, уравнение Гельмгольца (1.5.8) можно использовать для того, чтобы для *идеального газа* из его уравнения состояния, $PV = nRT$ (см. ниже уравнение (1.12.1)), сделать вывод о независимости внутренней энергии U от объема V при постоянной температуре ($(\partial U / \partial V)_{T, n_i} = 0$). В этом случае из формулы (1.5.7) для разности теплоемкостей получаем

$$C_P - C_V = nR. \quad (1.5.9)$$

Знание зависимостей $C_V(T)$ и $C_P(T)$ позволяет рассчитать температурный ход кривых $U(T)$, $H(T)$ и $S(T)$:

а) при $V = \text{const}$ и $n_i = \text{const}$

$$U(T) = U(T_0) + \int_{T_0}^T C_V(T) dT, \quad S(T) = S(T_0) + \int_{T_0}^T \frac{C_V(T)}{T} dT; \quad (1.5.10)$$

б) при $P = \text{const}$ и $n_i = \text{const}$

$$H(T) = H(T_0) + \int_{T_0}^T C_P(T) dT, \quad S(T) = S(T_0) + \int_{T_0}^T \frac{C_P(T)}{T} dT. \quad (1.5.11)$$

¹⁾ Гельмгольц (Helmholtz) Герман Людвиг Фердинанд (1821–1894) — немецкий физик, математик, физиолог и психолог, впервые (1847) математически обосновал закон сохранения энергии, показав его всеобщий характер, разработал термодинамическую теорию химических процессов, ввел понятия свободной и связанной энергии, занимался проблемами электродинамики и заложил основы теории аномальной дисперсии и вихревого движения жидкости.

Здесь $U(T_0)$, $H(T_0)$ и $S(T_0)$ — известные значения соответствующих функций при температуре T_0 . Если в интервале $T - T_0$ лежит температура фазового превращения вещества ($T_{\text{фп}}$), то равенства (1.5.10) и (1.5.11) должны содержать дополнительные члены $\Delta U_{\text{фп}}$, $\Delta S_{\text{фп}} = \Delta U_{\text{фп}}/T_{\text{фп}}$ и $\Delta H_{\text{фп}}$, $\Delta S_{\text{фп}} = \Delta H_{\text{фп}}/T_{\text{фп}}$, соответствующие тепловому эффекту и изменению энтропии при фазовом переходе первого рода (см. ниже п. 2.1 и рис. 2.3).

Основное термодинамическое содержание как внутренней энергии U , так и энтальпии H состоит в том, что их изменение $\Delta U = U_{\text{к}} - U_{\text{н}}$ или $\Delta H = H_{\text{к}} - H_{\text{н}}$, произошедшее в результате того или иного перехода системы из начального состояния (н) в конечное состояние (к), равняется тепловому эффекту процесса фазового или химического превращения,

$$Q_V = \Delta U = T\Delta S \quad \text{или} \quad Q_P = \Delta H = T\Delta S, \quad (1.5.12)$$

соответственно, для изохорных или изобарных условий.

Только для этих двух случаев теплота процесса (Q_V или Q_P) оказывается равной изменению функций состояния (ΔU или ΔH) и поэтому не зависит от пути процесса, а определяется лишь начальным и конечным состояниями системы.

Эта особенность изохорных и изобарных условий лежит в основе так называемого **теплого закона Гесса**, который гласит: *тепловой эффект реакции, происходящей при постоянном объеме или при постоянном давлении, зависит лишь от начального и конечного состояний системы, но не зависит от промежуточных состояний и путей перехода между ними.*

На основании закона Гесса¹⁾ теплота ΔH процесса растворения твердых и газообразных веществ в жидкостях представляется в виде суммы теплоты фазового перехода $\Delta H_{\text{фп}}$ (связанного с изменением агрегатного состояния вещества при плавлении твердого тела или при конденсации газа в жидкость) и теплоты $\Delta H_{\text{вз}}$, вызванной взаимодействием частиц двух жидкостей (жидкого растворителя и растворяемого вещества, находящегося в жидком состоянии, возникающем после плавления твердого тела или при конденсации газа в жидкость):

$$\Delta H = \Delta H_{\text{фп}} + \Delta H_{\text{вз}}. \quad (1.5.13)$$

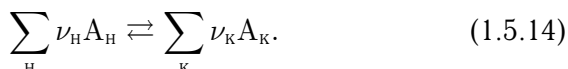
¹⁾ Гесс (Hess) Герман Иванович (1802–1850) — русский химик, один из основоположников термохимии, открыл основной ее закон (1840), носящий его имя, изучал химический состав минеральных вод, обнаружил поглощение газов платиной, описал многие минералы и впервые (1849) предложил систематику химических элементов, использованную позже (1870) Д.И. Менделеевым при построении периодической системы элементов.

Фазовые переходы первого рода и термодинамические соотношения, характерные для процесса образования растворов, будут рассмотрены в пп. 2.1, 2.3 и 2.4.

В применении к химическим реакциям закон Гесса позволяет выразить теплоту любой реакции в виде алгебраической суммы теплот образования реагентов из простых веществ, для которых теплота образования принимается равной нулю. *Простыми веществами* принято считать элементарные вещества, находящиеся в том агрегатном состоянии, которое устойчиво при атмосферном давлении ($P = 1$ атм) и рассматриваемой температуре T . Например, при комнатной температуре $T = 298 \text{ К} \approx 25^\circ\text{С}$ простыми веществами являются атмосферные газы (N_2 , O_2 , H_2), только две жидкости (бром Br и ртуть Hg), а также ряд твердых веществ, таких как C , Si , Fe , I_2 (йод) и др.

Теплота образования сложного вещества из простых веществ называется *стандартной теплотой образования* и обозначается в виде ΔH_{298}° ¹⁾. Здесь нижний индекс соответствует комнатной температуре $T = 298 \text{ К}$, а верхний кружок отмечает так называемое **стандартное состояние**, введенное ниже (см. п. 1.13). Значения стандартной теплоты образования ΔH_{298}° для химических соединений, наряду с другими стандартными значениями, приведены в специальных таблицах стандартных термодинамических величин. Пересчет значений ΔH_{298}° на другую температуру (ΔH_T°) производится с помощью формулы (1.5.11).

Любая обратимая химическая реакция может быть записана в следующей обобщенной форме:



Здесь $\text{A}_{\text{н}}$ и $\text{A}_{\text{к}}$ — химические символы (типа H_2 , O_2 , H_2O и т. п.) начальных (индекс «н») и конечных (индекс «к») участников реакции, $\nu_{\text{н}}$ и $\nu_{\text{к}}$ — их стехиометрические коэффициенты, уравнивающие количество молей каждого химического элемента по обе стороны реакции.

По закону Гесса теплота химической реакции (1.5.14) вычисляется как разница между конечным ($H_{\text{к}}$) и начальным ($H_{\text{н}}$) значениями энтальпии системы:

$$\Delta H \equiv H_{\text{к}} - H_{\text{н}} = \sum_{\text{к}} \nu_{\text{к}} \Delta H_T^\circ(\text{A}_{\text{к}}) - \sum_{\text{н}} \nu_{\text{н}} \Delta H_T^\circ(\text{A}_{\text{н}}). \quad (1.5.15)$$

¹⁾ Нередко в справочной литературе энтальпию образования ΔH_{298}° обозначают в виде ΔH_{f298}° с добавлением индекса f от англ. *formation*.

Здесь $\Delta H_T^\circ(A_n)$ и $\Delta H_T^\circ(A_k)$ — стандартные молярные (на один моль) теплоты образования начальных (A_n) и конечных (A_k) реагентов при температуре T . Химические компоненты, стоящие в обратимой реакции (1.5.14) слева, условно названы начальными, а стоящие справа — конечными. При обратной записи реакции начальные и конечные реагенты меняются местами, а знак теплоты (1.5.15) изменяется на противоположный.

1.6. Тепловой эффект и энтропия необратимых процессов

Любое тело вместе с внешней средой можно рассматривать как замкнутую (полностью изолированную) систему, для которой в неравновесных условиях применим закон возрастания энтропии, сформулированный в виде (1.2.3)–(1.2.4):

$$dS + dS^{\text{cp}} \geq 0, \quad (1.6.1)$$

где S — энтропия рассматриваемого тела, а S^{cp} — энтропия окружающей его внешней среды.

Знак равенства в формуле (1.6.1) соответствует равновесному состоянию термодинамической системы.

Неравенство в формуле (1.6.1) показывает направление самопроизвольного протекания неравновесных процессов внутри данного тела, рассматриваемого как открытая система, взаимодействующая с внешней средой, часто называемой *термостатом* из-за ее огромной теплоемкости. Свойства внешней среды чрезвычайно многообразны и недоступны в полной мере нашему описанию. Поэтому задача состоит в том, чтобы исключить свойства среды из неравенства (1.6.1).

Это возможно сделать, если воспользоваться тем обстоятельством, что любой процесс теплообмена можно считать с точки зрения термостата *квазистатическим* (см. п. 1.4). В этом случае, согласно (1.4.1), все тепло dQ^{cp} , подведенное к термостату, равняется теплу $-dQ$, отобранному от тела, которое идет на *обратимое изменение* энтропии термостата, тогда

$$dS^{\text{cp}} = \frac{dQ^{\text{cp}}}{T} = -\frac{dQ}{T}. \quad (1.6.2)$$

Подстановка (1.6.2) в (1.6.1) дает общее **условие равновесия и самопроизвольности** термодинамического процесса,

$$dS \geq \frac{dQ}{T}, \quad \text{или} \quad dQ \leq T dS, \quad (1.6.3)$$

которое связывает между собой величины dS и dQ только для рассматриваемого тела и не зависит от свойств внешней среды.

Итак, *неравновесные процессы*, самопроизвольно протекающие в открытой системе в направлении, указанном неравенствами (1.6.3), являются *необратимыми* с точки зрения рассматриваемой системы, несмотря на то, что для термостата любой процесс теплообмена происходит квазистатически.

1. Если *одно и то же* конечное количество тепла Q подводить к телу в обратимом и необратимом процессах, то приращение энтропии для этих двух процессов будет различным:

- для **обратимого процесса**, согласно (1.5.12),

$$\Delta S_{\text{обр}} = \frac{Q}{T}; \quad (1.6.4)$$

• для **необратимого процесса**, согласно (1.6.3), возрастание энтропии $\Delta S_{\text{необр}}$ превышает значение (1.6.4) на некоторую положительную величину $\Delta S' > 0$, вызванную необратимостью процесса и связанную с *производством энтропии* (см. п. 1.7), тогда

$$\Delta S_{\text{необр}} = \Delta S_{\text{обр}} + \Delta S' = \frac{Q}{T} + \Delta S' > \Delta S_{\text{обр}}. \quad (1.6.5)$$

Положительная величина $Q' = T\Delta S' > 0$ названа Клаузиусом *некомпенсированной теплотой*.

Поскольку энтропия является функцией состояния, то из выражения (1.6.5) следует, что в результате обратимого и необратимого процессов с одинаковым тепловым эффектом Q тело из начального состояния n переходит в разные конечные состояния k_1 и k_2 , как показано на рис. 1.3 а. При этом $\Delta S_{\text{необр}} > \Delta S_{\text{обр}}$, поскольку в неравновесном процессе энтропия возрастает как за счет *подвода тепла* к системе ($Q/T > 0$), так и в результате *необратимости процесса* ($\Delta S' > 0$). Если неравновесный процесс протекает даже в адиабатических условиях (без подвода или отвода тепла), то он все равно сопровождается возрастанием энтропии на величину $\Delta S'$. Примерами этого являются адиабатическое расширение газа в пустоту, изобарно-изотермическое смешение идеальных газов (см. п. 1.12) и образование атермальных растворов (см. п. 2.4).

2. Если, в противоположность рассмотренному случаю 1, переводить тело с помощью обратимого и необратимого процессов из начального состояния n в *одно и то же* конечное состояние k (см. рис. 1.3 б), то изменение энтропии для этих двух процессов будет одинаковым ($\Delta S_{\text{необр}} = \Delta S_{\text{обр}} \equiv \Delta S = S_k - S_n$), в то время как тепловой эффект в этом случае будет различным:

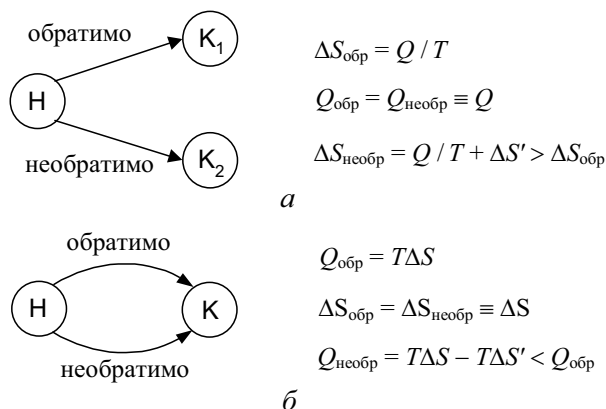


Рис. 1.3. Изменение энтропии в обратимом ($\Delta S_{\text{обр}}$) и необратимом ($\Delta S_{\text{необр}}$) процессах при одинаковом тепловом эффекте $Q \equiv Q_{\text{обр}} = Q_{\text{необр}}$ (а) и теплота обратимого ($Q_{\text{обр}}$) и необратимого ($Q_{\text{необр}}$) процессов при одинаковом изменении энтропии $\Delta S \equiv \Delta S_{\text{обр}} = \Delta S_{\text{необр}}$ (б)

- для **обратимого процесса**, по аналогии с (1.6.4),

$$\Delta S_{\text{обр}} = \Delta S = \frac{Q_{\text{обр}}}{T}; \quad (1.6.6)$$

- для **необратимого процесса**, по аналогии с (1.6.5),

$$\Delta S_{\text{необр}} = \Delta S = \frac{Q_{\text{обр}}}{T} = \frac{Q_{\text{необр}}}{T} + \Delta S'. \quad (1.6.7)$$

Из сравнения формул (1.6.6) и (1.6.7) получаем, что $Q_{\text{необр}} = Q_{\text{обр}} - T\Delta S' < Q_{\text{обр}}$, т. е. тепловые эффекты для необратимого и обратимого процессов отличаются на величину некомпенсированной теплоты Клаузиуса $Q' = T\Delta S' > 0$. Поскольку теплота может быть превращена в работу в некоем циклическом процессе (с возвратом системы в исходное состояние), то в силу неравенства $Q_{\text{обр}} > Q_{\text{необр}}$, максимально возможная работа, совершаемая системой, достижима только в обратимом циклическом процессе ($W_{\text{обр}} > W_{\text{необр}}$).

1.7. Второе начало термодинамики и производство энтропии

Возвращаемся к общему дифференциальному условию (1.6.3) для самопроизвольного протекания неравновесных процессов, которое принято записывать в следующем виде (см. рис. 1.4):

$$dS \equiv d_e S + d_i S > d_e S. \quad (1.7.1)$$

Здесь $d_e S$ — изменение энтропии системы в результате ее контактов с *внешней* средой (индекс «е» от англ. «external»); $d_i S$ — прирост энтропии, порожденный *внутренними* неравновесными процессами в системе (индекс «i» от англ. «internal» не путать с курсивным индексом i , обозначающим номер компонента); при этом $T d_i S = dQ'$ есть элементарная некомпенсированная теплота, введенная Клаузиусом.

В случае открытых систем, обменивающихся с внешней средой как теплотой ($(d_e S)_{\text{теп}} = dQ/T$ за счет теплового контакта), так и веществом ($(d_e S)_{\text{мат}}$ за счет материального контакта), внешнее приращение энтропии содержит оба эти вклада:

$$d_e S = (d_e S)_{\text{теп}} + (d_e S)_{\text{мат}} = \frac{dQ}{T} + (d_e S)_{\text{мат}}. \quad (1.7.2)$$

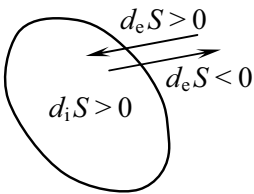


Рис. 1.4. Обмен энтропией между системой и внешней средой

Легко убедиться в том, что вычитанием материального вклада $(d_e S)_{\text{мат}}$ из обеих сторон неравенства (1.7.1) получаем условие (1.6.3) предыдущего параграфа, где рассматривался лишь теплообмен системы с внешней средой, что характерно для *закрытых систем*, когда $(d_e S)_{\text{мат}} = 0$.

Согласно второму началу термодинамики, *внутреннее приращение энтропии всегда положительно* ($d_i S > 0$ за счет неравновесности процессов), в то время как *внешнее приращение энтропии* может быть как *положительным* ($d_e S > 0$ за счет притока тепла и частиц), так и *отрицательным* ($d_e S < 0$ за счет оттока тепла и частиц) (см. рис. 1.4).

Как будет показано далее (см. пп. 1.8 и 1.9), независимо от того, идет ли речь о замкнутых (полностью изолированных), закрытых (частично изолированных, т. е. без обмена веществом) или открытых системах, самая **общая формулировка второго начала термодинамики** имеет вид

$$d_i S \geq 0. \quad (1.7.3)$$

Более того, формулировка (1.7.3) применима не только ко всей открытой системе, но и, в силу аддитивности энтропии, к отдельным макроскопическим подсистемам, входящим в состав системы, т. е. для каждой n -й подсистемы также $d_i S_n \geq 0$. Иными словами, невозможна физическая ситуация, при которой для всей *открытой системы* $d_i S = d_i \sum_n S_n > 0$, как результат компенсации вкладов *разных знаков* от подсистем, взаимодействующих с внешней средой.

Условие $d_i S_n \geq 0$ положено в основу разработки теории необратимых процессов, использующей базовую гипотезу о локальном (частичном) квазиравновесии внутри элементарных подсистем, из которых состоит сложная открытая система, не находящаяся в условиях глобального (полного) равновесия (см. п. 1.21).

Формулировка второго начала термодинамики в форме (1.7.3) является более строгой и отличается большей общностью, чем классическая форма (1.2.3) для изолированной системы, согласно которой энтропия всегда возрастает в неравновесных условиях.

Согласно Пригожину, протекание неравновесных процессов характеризуется **производством энтропии**, введенным как скорость временного изменения энтропии только в результате протекания внутренних процессов в системе:

$$\mathcal{P}(t) = \frac{d_i S(t)}{dt}. \quad (1.7.4)$$

Здесь производство энтропии обозначено буквой $\mathcal{P}$, в отличие от введенного Пригожиным обозначения в виде P . Это сделано для того, чтобы производство энтропии не перепутать с полным давлением, обозначенным в нашем случае как P (у Пригожина оно обозначено малой буквой p).

В соответствии со вторым началом термодинамики, записанным в форме (1.7.3), производство энтропии в любой момент времени положительно и обращается в нуль лишь при равновесии:

$$\mathcal{P}(t) \geq 0. \quad (1.7.5)$$

Необратимые процессы допускают описание на языке **термодинамических сил** X_α , вызванных пространственной неоднородностью интенсивных величин (типа температуры, давления, химических потенциалов веществ и т. п.) и **термодинамических потоков** J_α , появляющихся как следствие этих сил. В общем случае необратимое изменение $d_i S$ связано с некоторой элементарной величиной dx_α (например, теплотой или веществом), такой что $J_\alpha = dx_\alpha/dt$. Тогда необратимое изменение энтропии $d_i S$ и ее производство $\mathcal{P} = d_i S/dt$, вызванные всеми необратимыми процессами ($\alpha = 1, 2, \dots$), протекающими в системе, записывают как

$$d_i S = \sum_{\alpha} X_{\alpha} dx_{\alpha} \geq 0 \quad \text{и} \quad \mathcal{P} = \sum_{\alpha} J_{\alpha} X_{\alpha} \geq 0, \quad (1.7.6)$$

где неравенства отражают второй закон термодинамики.

Таким образом, производство энтропии каждым необратимым процессом α равно произведению термодинамической силы X_α и сопряженного с ней потока J_α . Именно это составляет основу

формализма теории Онсагера, более подробно рассмотренной далее (см. п. 1.19).

Ниже последовательно рассматриваем три физические ситуации, соответствующие (см. формулу (1.7.2) и рис. 1.4):

- **замкнутым системам**, в которых $(d_e S)_{\text{теп}} = 0$, $(d_e S)_{\text{мат}} = 0$;
- **закрытым системам**, в которых $(d_e S)_{\text{теп}} \neq 0$, $(d_e S)_{\text{мат}} = 0$;
- **открытым системам**, в которых $(d_e S)_{\text{теп}} \neq 0$, $(d_e S)_{\text{мат}} \neq 0$.

1.8. Равновесие, устойчивость и эволюция замкнутых и закрытых систем

Замкнутые системы в силу своей полной изолированности от внешней среды характеризуются отсутствием внешнего потока (оттока) энтропии ($d_e S = 0$). Поэтому, с учетом (1.7.4), для них справедливы следующие соотношения:

$$dS = d_i S \geq 0 \quad \text{и} \quad \mathcal{P} = \frac{dS}{dt} \geq 0, \quad (1.8.1)$$

где неравенства соответствуют неравновесным процессам, а равенства — равновесному состоянию системы.

Следовательно, в неравновесных условиях энтропия изолированной системы как функция термодинамической силы X_α для процесса α , приведенная на рис. 1.5 *a*, растет ($dS > 0$) до тех пор, пока система не достигнет равновесного состояния при значении $X_\alpha^{(0)}$, когда энтропия становится максимальной:

$$S = S_{\max} \quad \text{и} \quad dS = 0. \quad (1.8.2)$$

Как видно из рис. 1.5 *a*, вблизи точки равновесия при $X_\alpha^{(0)}$ энтропия S как функция параметра X_α имеет выпуклую форму. Поэтому аналогичная кривая для производства энтропии, $\mathcal{P} = f(X_\alpha)$, является вогнутой в окрестности точки $X_\alpha^{(0)}$ и обращается в нуль в этой точке. Следовательно, для процесса эволюции замкнутой системы должно выполняться следующее требование:

$$d\mathcal{P} = \frac{d}{dt} dS \leq 0. \quad (1.8.3)$$

Экстремальные значения энтропии ($S = S_{\max}$) и производства энтропии ($\mathcal{P}_{\min} = 0$), реализуемые при равновесии в замкнутой системе, являются *флуктуационно устойчивыми*.

Действительно, в окрестности равновесного значения $X_\alpha^{(0)}$ *двусторонние флуктуации* термодинамической силы ($\delta X_\alpha^\pm \geq 0$) приводят, как следует из рис. 1.5 *a*, к *односторонним флуктуациям*: отрицательным для энтропии ($\delta S^- < 0$) и положитель-

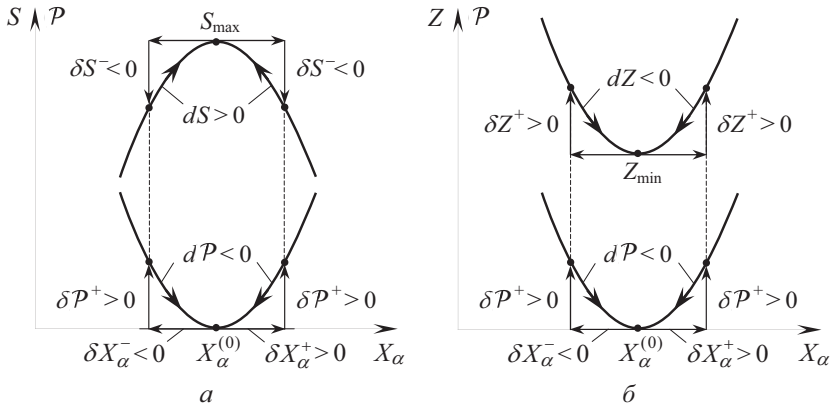


Рис. 1.5. Экстремальные свойства замкнутой системы (а) и закрытой системы (б) в виде зависимостей S , P и Z от термодинамической силы X_α для неравновесного процесса в окрестности точки термодинамического равновесия при $X_\alpha^{(0)}$ и релаксация системы к равновесию как отклик на двусторонние флуктуации термодинамической силы ($\delta X_\alpha^\pm \geq 0$)

ным для ее производства ($\delta P^+ > 0$). Это создает неравновесную ситуацию в системе, которая самопроизвольно запускает релаксационный механизм восстановления термодинамического равновесия. В соответствии с формулами (1.8.1) и (1.8.3), требования увеличения энтропии ($dS > 0$) и уменьшения ее производства ($dP < 0$), показанные стрелками вдоль кривых на рис. 1.5 а, обязательно обеспечивают затухание начальных флуктуаций $\delta S(0) \equiv \delta S^- < 0$ и $\delta P(0) \equiv \delta P^+ = -\delta S(0)/\tau > 0$ в виде следующих выражений релаксационного типа:

$$\delta S(t) = \delta S(0) e^{-t/\tau} < 0, \quad (1.8.4)$$

$$\delta P(t) = \frac{d}{dt} \delta S(t) = \delta P(0) e^{-t/\tau} > 0. \quad (1.8.5)$$

Здесь τ — время релаксации процесса, равное $\tau_{\text{гл}}$ для глобального (полного) равновесия системы или $\tau_{\text{лок}}$ для локального (частичного) равновесия квазизамкнутой ее части (см. с. 80).

Более строгое математическое обоснование флуктуационной устойчивости замкнутой системы будет дано ниже (см. п. 1.17).

Закрытые системы сохраняют неизменным свой состав из-за отсутствия обмена веществом с внешней средой ($(d_e S)_{\text{мат}} = 0$), так что тепловой эффект изохорных и изобарных процессов описывается соотношениями (1.5.1) и (1.5.3), соответственно.

Изотермическое условие ($T = \text{const}$) позволяет переписать неравенство (1.6.3), указывающее направление самопроизвольно

протекающих неравновесных процессов, в виде

$$d(Q - TS) \leq 0. \quad (1.8.6)$$

С учетом выражений (1.5.1) и (1.5.3) для теплового эффекта, дополнительное требование в форме $V = \text{const}$ или $P = \text{const}$, наложенное на систему, приводит условие самопроизвольности неравновесных процессов (1.8.6) к следующему виду:

а) для **изохорно-изотермических процессов**, при которых $dQ_V = dU$,

$$d(U - TS) \leq 0 \quad \text{или} \quad dF \leq 0, \quad (1.8.7)$$

где $F = U - TS$ — **свободная энергия Гельмгольца**, называемая также изохорно-изотермическим потенциалом или просто *изохорным потенциалом*;

б) для **изобарно-изотермических процессов**, при которых $dQ_P = dH$,

$$d(H - TS) \leq 0 \quad \text{или} \quad dG \leq 0, \quad (1.8.8)$$

где $G = H - TS$ — **свободная энергия Гиббса**, называемая также изобарно-изотермическим потенциалом или просто *изобарным потенциалом*.

Для нахождения величин dF и dG , соответственно, в изохорно-изотермических и изобарно-изотермических условиях воспользуемся тем обстоятельством, что, согласно (1.7.1) и (1.7.2), для неравновесных процессов всегда $dS = d_e S + d_i S$ и $d_i S \geq 0$, а в закрытых системах $dQ = T(d_e S)_{\text{теп}} \equiv T d_e S$. Тогда получаем

- для *изохорных условий* (см. формулу (1.5.1) при $V = \text{const}$):

$$dU = dQ_V = T d_e S = T dS - T d_i S = -T d_i S \leq 0 \quad \text{при} \quad S = \text{const},$$

$$dF = dU - d(TS) = -S dT - T d_i S = -T d_i S \leq 0 \quad \text{при} \quad T = \text{const};$$

- для *изобарных условий* (см. формулу (1.5.3) при $P = \text{const}$):

$$dH = dQ_P = T d_e S = T dS - T d_i S = -T d_i S \leq 0 \quad \text{при} \quad S = \text{const},$$

$$dG = dH - d(TS) = -S dT - T d_i S = -T d_i S \leq 0 \quad \text{при} \quad T = \text{const}.$$

Полученные выше дифференциалы четырех функций состояния, часто называемых *термодинамическими потенциалами*, можно записать в общей форме

$$dZ = -T d_i S \leq 0. \quad (1.8.9)$$

Здесь введены следующие обозначения потенциалов:

- при *изоэнтروпийных условиях* ($S = \text{const}$)

$$Z = \begin{cases} U & \text{при } V = \text{const}, \\ H & \text{при } P = \text{const}; \end{cases} \quad (1.8.10)$$

- при *изотермических условиях* ($T = \text{const}$)

$$Z = \begin{cases} F \equiv U - TS & \text{при } V = \text{const}, \\ G \equiv H - TS & \text{при } P = \text{const}. \end{cases} \quad (1.8.11)$$

Следовательно, любой термодинамический потенциал Z (из четырех вышеуказанных в форме (1.8.10) и (1.8.11)) достигает минимума при равновесии системы в соответствующих условиях. Однако, как уже отмечалось ранее (см. п. 1.4), практический интерес представляют лишь *изотермические условия*, при которых уменьшение свободной энергии Гельмгольца ($dF < 0$ при $V = \text{const}$) или Гиббса ($dG < 0$ при $P = \text{const}$) характеризует направление эволюции закрытой системы.

Из (1.8.9) получаем выражение для производства энтропии:

$$\mathcal{P} \equiv \frac{d_i S}{dt} = -\frac{1}{T} \frac{dZ}{dt} \geq 0, \quad (1.8.12)$$

откуда следует неравенство, характеризующее процесс самопроизвольной эволюции в закрытой системе (ср. формулу (1.8.3)):

$$d\mathcal{P} = \frac{d}{dt} d_i S = -\frac{1}{T} \frac{d}{dt} dZ \leq 0. \quad (1.8.13)$$

Соотношения (1.8.9), (1.8.12) и (1.8.13) позволяют качественно изобразить на рис. 1.5 б зависимость термодинамического потенциала Z и производства энтропии $\mathcal{P}$ от термодинамической силы X_α для неравновесного процесса α . Из сравнения рис. 1.5 а и рис. 1.5 б видно, что, несмотря на различный характер функций $S = f_S(X_\alpha)$ (с максимумом) и $Z = f_Z(X_\alpha)$ (с минимумом), соответственно, для замкнутых и закрытых систем (что и выражает *принцип экстремума*), производство энтропии в обоих случаях имеет *нулевой минимум* при равновесии.

Это и есть общий **принцип минимума производства энтропии** в равновесном состоянии для замкнутых и закрытых систем.

Таким образом, как замкнутые, так и закрытые системы, выведенные из состояния равновесия, всегда *самопроизвольно эволюционируют* в направлении снижения производства энтропии ($d\mathcal{P} < 0$) до тех пор, пока оно не станет равным нулю в точке минимума кривой $\mathcal{P} = f(X_\alpha)$ на рис. 1.5 а и б, где $d\mathcal{P} = 0$. Более того, состояние термодинамического равновесия *всегда флуктуационно устойчиво* (см. п. 1.17).

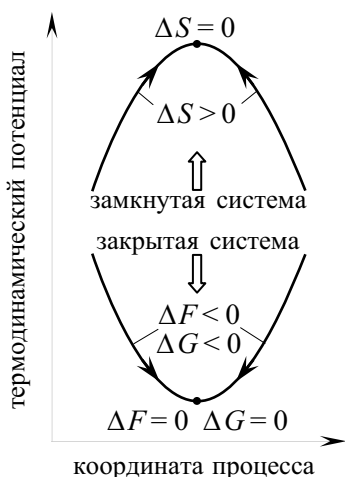


Рис. 1.6. К условиям равновесия и самопроизвольной эволюции процессов в замкнутых и открытых системах

В дальнейшем будем рассматривать только изохорно-изотермические и изобарно-изотермические условия, к которым применимы неравенства (1.8.7) и (1.8.8) для потенциалов в форме (1.8.11).

Подобные неравенства вкладывают вполне определенное физическое содержание в свободные энергии F и G : они являются такими функциями состояния, уменьшение которых показывает направление *самопроизвольного протекания* процессов, соответственно, в изохорно-изотермических и изобарно-изотермических условиях. Состояние равновесия в этих условиях характеризуется минимальным значением свободной энергии Гельмгольца, $F = U - TS$, или сво-

бодной энергии Гиббса, $G = H - TS$.

По знаку *конечных приращений* этих величин,

$$\Delta F = \Delta U - T\Delta S \quad \text{или} \quad \Delta G = \Delta H - T\Delta S, \quad (1.8.14)$$

обычно судят о возможности реализации тех или иных процессов в закрытых системах при соответствующих условиях:

а) при $V = \text{const}$ и $T = \text{const}$:

$$\begin{aligned} \Delta F < 0 \quad \text{или} \quad \Delta U < T\Delta S & \text{— условие самопроизвольности,} \\ \Delta F = 0 \quad \text{или} \quad \Delta U = T\Delta S & \text{— условие равновесия;} \end{aligned} \quad (1.8.15)$$

б) при $P = \text{const}$ и $T = \text{const}$:

$$\begin{aligned} \Delta G < 0 \quad \text{или} \quad \Delta H < T\Delta S & \text{— условие самопроизвольности,} \\ \Delta G = 0 \quad \text{или} \quad \Delta H = T\Delta S & \text{— условие равновесия.} \end{aligned} \quad (1.8.16)$$

Критерии (1.8.15) и (1.8.16) равновесия и самопроизвольности процессов качественно изображены на рис. 1.6, где также показан аналогичный критерий для замкнутой (полностью изолированной) системы, определяемый законом возрастания энтропии в форме соотношений (1.8.1) и (1.8.2). Условию равновесия соответствует точка экстремума на кривой. По оси абсцисс отложена так называемая «координата процесса», под которой понимается любая физическая величина (например, концентрация одного из

реагентов), которая меняется в ходе процесса и доступна опытному наблюдению.

В технологической практике чаще приходится иметь дело с процессами, протекающими при постоянном давлении ($P = 1$ атм). Поэтому в дальнейшем для оценки возможности осуществления процесса будем использовать изобарный критерий:

$$\Delta G = \Delta H - T\Delta S \leq 0, \quad (1.8.17)$$

где знак равенства соответствует термодинамическому равновесию процесса.

Как видно из выражения (1.8.17), возможность протекания любого процесса определяется конкуренцией двух факторов — энтальпийного (энергетического) ΔH и энтропийного (статистического) $T\Delta S$. Направление процесса зависит от соотношения между этими факторами. Безусловное выполнение неравенства в формуле (1.8.17) имеет место лишь при

$$\Delta H < 0 \quad \text{и} \quad \Delta S > 0, \quad (1.8.18)$$

т. е. в тех условиях, когда теплота выделяется, а энтропия возрастает. В этом случае повышение температуры делает более весомым вклад энтропийного фактора в выражении (1.8.17) и тем самым увеличивает возможность протекания процесса.

В условиях равновесия (при $\Delta G = 0$) энтальпийный и энтропийный факторы компенсируют друг друга, так что

$$\Delta H = T\Delta S. \quad (1.8.19)$$

Нетрудно убедиться в том, что одновременное выполнение обоих неравенств в формуле (1.8.18) противоречиво для любого процесса, сопровождаемого изменением степени упорядоченности или разупорядоченности системы. Действительно, повышение разупорядоченности, происходящее при увеличении температуры, объема или числа частиц в системе, увеличивает ее статистический вес и повышает энтропию ($\Delta S > 0$). Наоборот, упорядочение сопровождается ростом сил притяжения между частицами, приводящим к снижению потенциальной энергии системы (см. ниже рис. 2.5) с выделением избыточной энергии в виде тепла ($\Delta H < 0$).

Следовательно, переход к более упорядоченным структурам всегда сопровождается выделением тепла и уменьшением энтропии:

$$\Delta H < 0 \quad \text{и} \quad \Delta S < 0. \quad (1.8.20)$$

Сюда относятся процессы, усложняющие структуру вещества и повышающие его агрегацию, такие как необратимое сжатие

газов и тел, конденсация паров и газов, кристаллизация веществ из расплавов и растворов, образование молекул из атомов и ассоциация молекул в комплексы, экзотермические химические реакции с уменьшением числа молей и т. п.

Наоборот, все процессы, связанные с понижением степени упорядоченности, всегда сопровождаются поглощением тепла и увеличением энтропии:

$$\Delta H > 0 \quad \text{и} \quad \Delta S > 0. \quad (1.8.21)$$

Сюда относятся процессы, связанные с дезагрегацией вещества, такие как необратимое расширение и смешение газов, испарение и кипение жидкостей, сублимация и плавление твердых тел, растворение веществ, диссоциация молекул на атомы, эндотермические химические реакции с увеличением числа молей в системе и т. п.

Следуя неравенствам (1.8.20) и (1.8.21), повторно подчеркнем, что установление равновесия в системе определяется противоборством двух конкурирующих тенденций к упорядочению и разупорядочению структуры вещества.

1.9. Стационарность, устойчивость и эволюция открытых систем

Открытые системы обмениваются с внешней средой как энергией в форме теплопередачи и совершения работы (что имеет место и в закрытых системах), так и материальными частицами. Это приводит к трем дополнительным вкладам в изменения:

- а) энтропии на величину $(d_e S)_{\text{мат}}$ (см. уравнение (1.7.2)),
- б) числа молей частиц i -го сорта на величину $d_e n_i$,
- в) внутренней энергии системы на величину $d_e U$.

С учетом сказанного, полные приращения вышеуказанных величин, включая вклады $d_i S$, $d_i n_i$, $d_i U$ от процессов, протекающих *внутри системы*, принимают следующий вид (см. рис. 1.4):

$$dS = (d_e S)_{\text{теп}} + (d_e S)_{\text{мат}} + d_i S, \quad (1.9.1)$$

$$dn_i = d_e n_i + d_i n_i, \quad (1.9.2)$$

$$dU = d_e U + d_i U. \quad (1.9.3)$$

В соответствии с законом сохранения энергии, внутренние процессы, протекающие без участия внешней среды, не в состоянии изменить энергию системы, которая может лишь превра-

щаться из одной формы в другую. В результате этого в равенстве (1.9.3) всегда имеем

$$d_i U = 0 \quad \text{и} \quad dU \equiv d_e U. \quad (1.9.4)$$

Число молей n_i любого вещества может изменяться *внутри системы* на величину $d_i n_i$ только за счет химических реакций.

Поведение энтропии в открытых системах принципиально отличается от ее поведения в замкнутых системах. Эволюция последних в неравновесных условиях всегда самопроизвольно происходит по направлению к термодинамическому равновесию за счет роста энтропии системы ($dS = d_i S \geq 0$).

Открытая система, будучи способной *отдавать* внешней среде энтропию ($d_e S \equiv (d_e S)_{\text{теп}} + (d_e S)_{\text{мат}} \leq 0$) или, как часто говорят, принимать отрицательную энтропию (называемую *негэнтропией*), тем самым уменьшает свою полную энтропию ($dS = d_e S + d_i S \leq 0$). В этом случае может возникать без нарушения второго начала термодинамики (т. е. при $d_i S > 0$) *стационарное неравновесное состояние* открытой системы, устойчивое по отношению к малым флуктуациям, которое нередко называют **текущим равновесием**.

В состоянии *текущего равновесия* термодинамические характеристики системы становятся стационарными, что позволяет записать выражения (1.9.1)–(1.9.4) в виде

$$dS = 0 \quad \text{или} \quad d_e S = -d_i S \leq 0, \quad (1.9.5)$$

$$dn_i = 0 \quad \text{или} \quad d_e n_i = -d_i n_i \geq 0, \quad (1.9.6)$$

$$d_e U = 0. \quad (1.9.7)$$

Для поддержания текущего равновесия в открытой системе должен происходить *приток негэнтропии* (т. е. отток энтропии), компенсирующий производство энтропии внутри системы, а также приток/отток вещества, компенсирующий изменение состава системы, вызванное химическими реакциями. Следовательно, при текущем равновесии, в отличие от термодинамического равновесия, обязательно должно быть $d_i S > 0$, а это значит, что в формуле (1.9.5) остается только знак *строгого неравенства*.

В соответствии с равенством (1.9.7), при текущем равновесии всегда строго сбалансированы процессы притока и оттока энергии, что и обеспечивает $d_e U = 0$. Поскольку при этом $d_i S > 0$, то протекающие в системе энергетические процессы всегда диссипативны (т. е. сопровождаются рассеянием энергии), что и дает возникающим при этом структурам название *диссипативных*, по терминологии Пригожина.

Поскольку открытые системы находятся в полном контакте с окружающей средой, то их стационарные состояния возможны лишь в том случае, если внешняя среда сама стационарна. Это обеспечивает неизменные во времени условия на границе системы. Они, с одной стороны, препятствуют достижению системой термодинамического равновесия, а с другой стороны, реализуют новое текущее равновесие при наличии в системе ненулевого производства энтропии и стационарного потока вещества через границу системы.

Рассмотрим изохорно-изотермические и изобарно-изотермические процессы в открытых системах. Для этого используем уравнение (1.3.5) для dU и уравнение (1.5.2) для dH , а также выражения (1.5.1) и (1.5.3) для теплового эффекта изохорных и изобарных процессов, согласно которым

$$(dU)_{V, n_i} = dQ_V \quad \text{и} \quad (dH)_{P, n_i} = dQ_P. \quad (1.9.8)$$

Разрешая уравнения (1.3.5) и (1.5.2) относительно дифференциала энтропии, записываем:

$$dS = \begin{cases} \frac{1}{T} dU - \frac{1}{T} \sum_i \mu_i dn_i & \text{при } V = \text{const}, \\ \frac{1}{T} dH - \frac{1}{T} \sum_i \mu_i dn_i & \text{при } P = \text{const}. \end{cases} \quad (1.9.9)$$

Приравнявая dS в (1.9.9) выражению (1.9.1) и подставляя равенства (1.9.2) и (1.9.8) в правую часть (1.9.9), получаем следующие три вклада в полный дифференциал энтропии dS :

$$(d_e S)_{\text{теп}} = \begin{cases} dQ_V/T & \text{при } V = \text{const}, \\ dQ_P/T & \text{при } P = \text{const}; \end{cases} \quad (1.9.10)$$

$$(d_e S)_{\text{мат}} = -\frac{1}{T} \sum_i \mu_i dn_i; \quad (1.9.11)$$

$$d_i S = -\frac{1}{T} \sum_i \mu_i d_i n_i. \quad (1.9.12)$$

Выясним, как влияет материальный контакт системы с внешней средой, осуществляемый путем взаимного обмена частицами, на изохорно-изотермический потенциал $F = U - TS$ (свободную энергию Гельмгольца) и изобарно-изотермический потенциал $G = H - TS$ (свободную энергию Гиббса).

Для этого используем формулы (1.9.1), (1.9.4), (1.9.8) и (1.9.10), из которых следует:

- для *изохорно-изотермических условий* ($V = \text{const}$ и $T = \text{const}$)

$$dF = d(U - TS) = dQ_V - Td_e S - Td_i S,$$

$$\text{или} \quad dF + T(d_e S)_{\text{мат}} = -Td_i S < 0; \quad (1.9.13)$$

- для *изобарно-изотермических условий* ($P = \text{const}$ и $T = \text{const}$)

$$dG = d(H - TS) = dQ_P - Td_e S - Td_i S,$$

$$\text{или} \quad dG + T(d_e S)_{\text{мат}} = -Td_i S < 0. \quad (1.9.14)$$

Чисто формально, обозначая левую часть уравнений (1.9.13) и (1.9.14) в виде dZ , получаем формулу,

$$dZ = -Td_i S < 0, \quad (1.9.15)$$

по виду аналогичную формуле (1.8.9), полученной для закрытых систем. Разница между ними заключается в том, что в данном случае функция Z не является термодинамическим потенциалом, так как зависит от пути процесса подпитки системы веществом из внешней среды.

Кроме того, справа в формуле (1.9.15) стоит строгое неравенство, что характерно для открытых систем, внутри которых производится ненулевая энтропия ($d_i S > 0$), компенсируемая при текущем равновесии притоком негэнтропии ($-d_e S < 0$), так что $dS = d_e S + d_i S = 0$ (см. формулу (1.9.5)).

Исходя из формулы (1.9.15), записываем выражение для производства энтропии в открытых системах,

$$\mathcal{P} \equiv \frac{d_i S}{dt} = -\frac{1}{T} \frac{dZ}{dt} > 0, \quad (1.9.16)$$

которое по виду совпадает с аналогичным выражением (1.8.12) для закрытых систем и отличается, согласно вышесказанному, лишь знаком строгого неравенства.

Сравнение формул (1.8.1), (1.8.12) и (1.9.16) для производства энтропии в замкнутой, закрытой и открытой системах показывает, что независимо от типа системы существует универсальное условие формирования равновесия (термодинамического или текущего). Такое условие формулируется в виде **теоремы Пригожина о минимуме производстве энтропии**: *эволюция неравновесной системы с неизменными граничными условиями происходит самопроизвольно в направлении уменьшения производства энтропии, достигающего в стационарном состоянии минимальной величины: $\mathcal{P}_{\min} = 0$ для замкнутых и закрытых систем и $\mathcal{P}_{\min} > 0$ для открытых линейных систем (см. доказательство теоремы в п. 1.20).*

Теорема Пригожина устанавливает следующие условия:

$$dP < 0 \quad \text{— условие самопроизвольной эволюции,} \quad (1.9.17)$$

$$dP = 0 \quad \text{— условие равновесия/стационарности.} \quad (1.9.18)$$

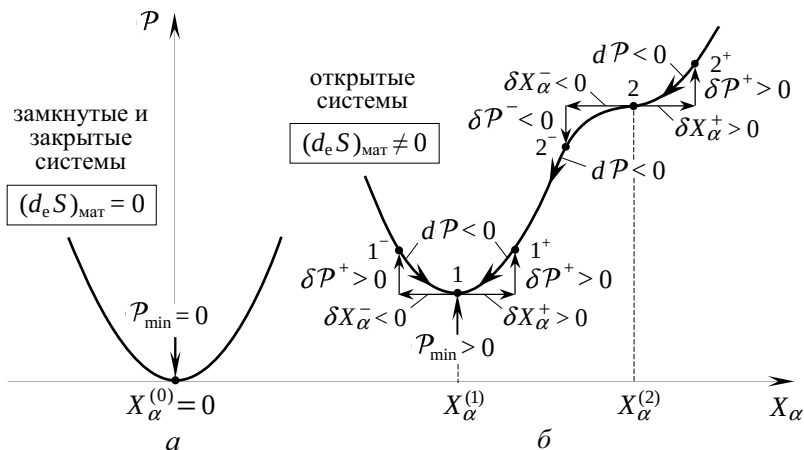


Рис. 1.7. Поведение производства энтропии как функции термодинамической силы X_α для неравновесного процесса: a — в окрестности точки равновесия при $X_\alpha^{(0)} = 0$ для замкнутых и закрытых систем, b — в окрестности стационарных точек 1 и 2 при $X_\alpha^{(1)}$ и $X_\alpha^{(2)}$ для открытой системы, а также отклик последней на двусторонние флуктуации термодинамической силы ($\delta X_\alpha^\pm \geq 0$)

Графическая интерпретация теоремы Пригожина изображена на рис. 1.7. Здесь приведены две кривые, качественно показывающие зависимость производства энтропии P от термодинамической силы X_α : одна (рис. 1.7 a) — для замкнутых и закрытых систем без материального контакта со средой ($(d_e S)_{\text{mat}} = 0$), другая (рис. 1.7 b) — для открытых систем, обменивающихся с внешней средой веществом ($(d_e S)_{\text{mat}} \neq 0$). При этом первая кривая соответствует аналогичным кривым, изображенным на рис. 1.5. Однако кривая на рис. 1.7 a правильнее отображает физическую реальность. Дело в том, что сила X_α , исполняя роль возвращающей силы, отлична от нуля лишь в неравновесных условиях и исчезает при равновесии. Поэтому нулевой минимум ($P_{\min} = 0$) для обеих кривых на рис. 1.5 a и b следует расположить так же, как на рис. 1.7 a , то есть при $X_\alpha^{(0)} = 0$.

Включение материального контакта (начиная с нулевого значения $(d_e S)_{\text{mat}}$, характерного для замкнутых и закрытых систем) плавно смещает кривую на рис. 1.7 a вправо и вверх, превращая ее в часть кривой на рис. 1.7 b , расположенную в районе ста-

онарной точки 1 при $X_\alpha^{(1)}$. Этот участок соответствует открытым системам при малом отклонении от состояния равновесия, которые описываются линейной неравновесной термодинамикой. На этой кривой приведена и другая стационарная точка 2 при $X_\alpha^{(2)}$, в которой, как и в первой точке, для производства энтропии $dP = 0$. Этот участок кривой соответствует открытым системам при сильном возмущении, далеком от состояния равновесия, которые описываются нелинейной неравновесной термодинамикой.

В точке 1 реализуется текущее равновесие, устойчивое к *двусторонним флуктуациям* термодинамической силы, $\delta X_\alpha^\pm \geq 0$. Откликом системы в точке 1 являются *односторонние флуктуации* производства энтропии, $\delta P^+ > 0$, смещающие систему в точки 1^+ и 1^- , обе лежащие выше минимума $P_{\min} > 0$. Именно это и обеспечивает *флуктуационную устойчивость* системы.

Действительно, согласно условию (1.9.17), система самопроизвольно эволюционирует в состояние текущего равновесия с минимальным ненулевым производством энтропии ($P_{\min} > 0$), что показано вдоль кривой стрелками $dP < 0$, направленными из точек 1^+ и 1^- к точке 1 на рис. 1.7 б.

Иное поведение системы наблюдается в точке 2, где в ответ на двусторонние флуктуации $\delta X_\alpha^\pm \geq 0$ возникают также *двусторонние флуктуации* производства энтропии $\delta P^\pm \geq 0$, смещающие систему в точки 2^+ и 2^- , которые лежат по разные стороны от горизонтальной касательной к кривой в точке 2.

Здесь положительные флуктуации $\delta P^+ > 0$, как и в точке 1, запускают релаксационный механизм, возвращающий систему из точки 2^+ в точку 2. Однако отрицательные флуктуации $\delta P^- < 0$ в точке 2^- , подчиняясь тому же условию (1.9.17), уводят систему из точки 2 вниз по стрелке вдоль кривой в направлении $dP < 0$.

Другими словами, если неравновесное состояние в точке 2^- , возникшее в результате флуктуации, лежит ниже (по оси P) стационарного состояния в точке 2, то условие (1.9.17) самопроизвольной эволюции ($dP < 0$) вызывает движение системы вниз вдоль кривой $P = f(X_\alpha)$. Это способствует развитию исходной флуктуации, что приводит ко все большему отклонению от стационарного состояния. Следовательно, точка 2 (являющаяся точкой перегиба кривой или седловой точкой двухкоординатной поверхности, в которой имеет место смена знака кривизны кривой в точке перегиба или кривизны поверхности в седловой точке), соответствует *неустойчивому стационарному состоянию*.

Таким образом, устойчивость любого состояния (равновесного или стационарного) обеспечивается лишь положительными

флуктуациями производства энтропии, $\delta\mathcal{P}^+ > 0$. Они затухают, согласно условию (1.9.17) теоремы Пригожина, по уравнению (1.8.5), возвращая систему в состояние с минимумом производства энтропии ($\mathcal{P}_{\min} \geq 0$). Поскольку малые отклонения от этого состояния могут лишь увеличить энтропию ($\delta\mathcal{P}^+ > 0$ в точке 1 на рис. 1.7 б), то в *линейной области* отклонений от текущего равновесия (описываемых теорией Онсагера) *принципиально невозможны неустойчивые состояния* открытых систем.

В *нелинейной области* при достаточно больших отклонениях устойчивость стационарных состояний *не обеспечивается автоматически*: они устойчивы, только если все возможные отклонения, возмущения и флуктуации имеют характер «подъема в гору», как показано на рис. 1.7 б в точке 2 для случая $\delta\mathcal{P}^+ > 0$. Именно это неравенство обеспечивает устойчивость стационарного состояния нелинейных открытых систем (см. п. 1.24).

Условия флуктуационной устойчивости системы (замкнутой, закрытой, открытой) могут быть путем получены простейшим способом.

Пусть кроме равновесного состояния системе доступны стационарные состояния, среди которых она под действием внешних условий выбирает состояние с минимальным производством энтропии. Будем отмечать эти состояния верхним индексом (s), при этом $s = 0$ соответствует равновесному состоянию, а $s = 1, 2, \dots$ — неравновесным стационарным состояниям (см. рис. 1.7).

Согласно формуле (1.7.6), производство энтропии выражается как сумма произведений термодинамического потока J_α и силы X_α для каждого неравновесного процесса α :

$$\mathcal{P} = \sum_{\alpha} J_{\alpha} X_{\alpha}. \quad (1.9.19)$$

Для s -го стационарного состояния эти силы и потоки представляем в виде суммы стационарных значений ($X_{\alpha}^{(s)}, J_{\alpha}^{(s)}$) и малых добавок к ним ($\delta X_{\alpha}, \delta J_{\alpha}$), порожденных флуктуациями:

$$X_{\alpha} = X_{\alpha}^{(s)} + \delta X_{\alpha} \quad \text{и} \quad J_{\alpha} = J_{\alpha}^{(s)} + \delta J_{\alpha}, \quad (1.9.20)$$

при этом в равновесном состоянии $X_{\alpha}^{(0)} = 0$ и $J_{\alpha}^{(0)} = 0$.

Подстановкой выражений (1.9.20) в формулу (1.9.19) получаем производство энтропии в виде

$$\mathcal{P} = \mathcal{P}_{\min} + \mathcal{P}_{\text{ex}}, \quad (1.9.21)$$

где

$$\mathcal{P}_{\min} = \sum_{\alpha} J_{\alpha}^{(s)} X_{\alpha}^{(s)} \quad \text{и} \quad \mathcal{P}_{\text{ex}} = \sum_{\alpha} \delta J_{\alpha} \delta X_{\alpha}. \quad (1.9.22)$$

Здесь $\mathcal{P}_{\min}$ — *производство энтропии* в s -м стационарном состоянии, а квадратичная по флуктуациям величина $\mathcal{P}_{\text{ex}}$ (индекс ex от англ. *excess*) называется **избыточным производством энтропии** (обозначенным у Пригожина $\delta_X P$). Ясно, что избыток введен по отношению к стационарному значению $\mathcal{P}_{\min}$.

При переходе от формулы (1.9.19) к окончательному выражению (1.9.21) для линейной по флуктуациям суммы было принято

$$\sum_{\alpha} \left(J_{\alpha}^{(s)} \delta X_{\alpha} + X_{\alpha}^{(s)} \delta J_{\alpha} \right) = 0. \quad (1.9.23)$$

Именно этого требует условие (1.9.18) теоремы Пригожина ($d\mathcal{P} = 0$) для минимизации функции $\mathcal{P} = f(X_{\alpha})$ в стационарном (равновесном) состоянии системы.

По аналогии с (1.9.20), разложим энтропию S неравновесной системы при наличии флуктуаций в ряд по малым вариациям, только удерживая теперь, кроме первой вариации δS , также и вторую вариацию $\delta^2 S$ (ср. формулу (1.9.21)):

$$S = S_{\text{st}} + \delta S_{\text{st}} + \frac{1}{2} \delta^2 S = S_{\text{st}} + \frac{1}{2} \delta^2 S \equiv S_{\text{st}} + S_{\text{ex}}, \quad (1.9.24)$$

где S_{st} — стационарная энтропия (индекс st от англ. *stationary*). Добавка $\frac{1}{2} \delta^2 S$ может быть названа **избыточной энтропией** S_{ex} , по аналогии с $\mathcal{P}_{\text{ex}}$. Во втором равенстве (1.9.24) опущено линейное по флуктуациям слагаемое ($\delta S_{\text{st}} = 0$) из условия экстремальности S в стационарной точке, что аналогично (1.9.23) для $\mathcal{P}$.

Как следует из условия (1.9.5), на стационарное состояние накладывается следующее требование:

$$\frac{dS_{\text{st}}}{dt} = 0, \quad \text{откуда} \quad \frac{d_e S_{\text{st}}}{dt} = -\frac{d_i S_{\text{st}}}{dt} \equiv -\mathcal{P}_{\min}. \quad (1.9.25)$$

Ранее уже было отмечено, что стационарное состояние открытой системы реализуется только в условиях стационарности внешней среды и неизменности во времени условий на границе системы, которые сохраняются и при наличии флуктуаций. Поэтому из двух слагаемых $\mathcal{P}_{\min}$ и $\mathcal{P}_{\text{ex}}$ в выражении (1.9.21) для производства энтропии первое задействовано в стационарном состоянии (см. формулу (1.9.25)), а в неравновесных условиях остается работающим лишь $\mathcal{P}_{\text{ex}}$. Следовательно, избыточное производство энтропии $\mathcal{P}_{\text{ex}}$ ответственно за изменение энтропии во времени, обеспечивая $dS/dt \neq 0$ при наличии флуктуаций.

Дифференцируя энтропию (1.9.24) по времени с учетом (1.9.25), на основании вышесказанного получаем искомое уравнение:

$$\frac{d}{dt} S_{\text{ex}}(t) \equiv \frac{d}{dt} \frac{\delta^2 S(t)}{2} = \mathcal{P}_{\text{ex}}(t). \quad (1.9.26)$$

Интегрированием уравнения (1.9.26) находим временное поведение избыточной энтропии (второй вариации $\delta^2 S/2$):

$$S_{\text{ex}}(t) = S_{\text{ex}}(0) + \int_0^t \mathcal{P}_{\text{ex}}(t) dt. \quad (1.9.27)$$

Здесь начальный отсчет $t = 0$ соответствует появлению флуктуации производства энтропии — положительной ($\delta \mathcal{P}^+ > 0$) или отрицательной ($\delta \mathcal{P}^- < 0$) (см. рис. 1.7 б). Проанализируем выражение (1.9.27) для этих двух случаев.

Положительная начальная флуктуация $\delta \mathcal{P}^+$, как следует из вышесказанного, обязательно затухает во времени. Тогда, по аналогии с формулой (1.8.5), записываем (см. рис. 1.8 а)

$$\mathcal{P}_{\text{ex}}(t) = \delta \mathcal{P}^+ e^{-t/\tau} > 0. \quad (1.9.28)$$

Согласно формуле (1.9.26), при $\mathcal{P}_{\text{ex}}(t) > 0$ скорость изменения избыточной энтропии $S_{\text{ex}}(t)$ также положительна:

$$\frac{d}{dt} S_{\text{ex}}(t) = \mathcal{P}_{\text{ex}}(t) > 0. \quad (1.9.29)$$

Это означает, что полная энтропия флуктуирующей системы, $S(t) = S_{\text{st}} + S_{\text{ex}}(t)$, возрастает в процессе релаксации, увеличивая беспорядок в системе по мере приближения к стационарному состоянию, в котором $S = S_{\text{st}} = \text{const}$.

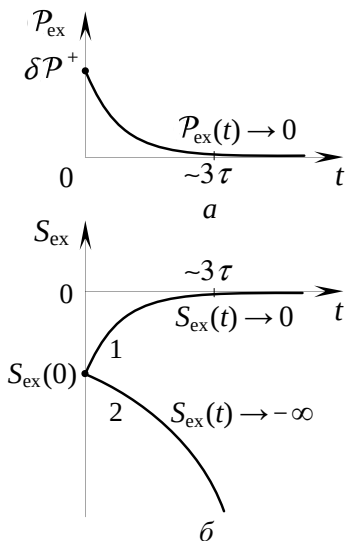
Подстановка (1.9.28) под интеграл в формуле (1.9.27) приводит к искомой временной зависимости избыточной энтропии:

$$S_{\text{ex}}(t) = (S_{\text{ex}}(0) + \tau \delta \mathcal{P}^+) - \tau \delta \mathcal{P}^+ e^{-t/\tau}. \quad (1.9.30)$$

Рис. 1.8. Изменение во времени избыточного производства энтропии $\mathcal{P}_{\text{ex}}(t)$ (а) и избыточной энтропии $S_{\text{ex}}(t) < 0$ (б)

Требование затухания положительных флуктуаций по мере приближения системы к стационарному состоянию обеспечивает асимптотическое поведение избыточной энтропии:

$$S_{\text{ex}}(t) \rightarrow 0 \quad \text{при} \quad t \rightarrow \infty. \quad (1.9.31)$$



Использование (1.9.31) позволяет из выражения (1.9.30) найти начальное значение избыточной энтропии,

$$S_{\text{ex}}(0) = -\tau \delta \mathcal{P}^+ < 0, \quad (1.9.32)$$

которое оказалось отрицательным.

Таким образом, положительная флуктуация, выводящая открытую систему из стационарного состояния, всегда уменьшает энтропию. Этот результат полностью коррелирует с флуктуационным поведением замкнутой системы, для которой всегда имеем $\delta S \equiv \delta S^- < 0$ (см. рис. 1.5 а).

С учетом начального условия (1.9.32), выражение (1.9.30) принимает окончательный вид:

$$S_{\text{ex}}(t) = S_{\text{ex}}(0) e^{-t/\tau} < 0. \quad (1.9.33)$$

Следовательно, в течение всего времени флуктуационной релаксации (восстановления стационарного состояния) величина $S_{\text{ex}}(t)$ остается отрицательной и, в соответствии с (1.9.26), возрастает (уменьшается по модулю), асимптотически приближаясь к нулю (см. формулу (1.9.31)), как показано кривой 1 на рис. 1.8 б.

Таким образом, **флуктуационная устойчивость** замкнутых, закрытых и открытых систем, работающих в линейном режиме, гарантируется отрицательным знаком кривизны поверхности энтропии ($\delta^2 S < 0$) как функции параметров, флуктуирующих около равновесного или стационарного состояния. На основании равенств (1.9.29) и (1.9.33) достаточные условия устойчивости для таких систем записываются через избыточные величины как

$$S_{\text{ex}} \equiv \frac{\delta^2 S}{2} < 0 \quad \text{и} \quad \mathcal{P}_{\text{ex}} \equiv \frac{d}{dt} S_{\text{ex}} > 0. \quad (1.9.34)$$

В перечисленных выше системах флуктуации производства энтропии могут быть только положительными ($\delta \mathcal{P}^+ > 0$) из-за минимального производства энтропии в точках равновесия (см. рис. 1.5) или стационарности (см. точку 1 на рис. 1.7 б). Поэтому для них условия устойчивости (1.9.34) выполняются автоматически, то есть *линейные системы* в равновесном/стационарном состоянии *не могут быть неустойчивыми*.

В отличие от этого, для открытых систем в нелинейном режиме, возникающем при больших отклонениях от равновесия, условия устойчивости (1.9.34) работают только для положительных флуктуаций в точке 2 на рис. 1.7 б. Кроме них, в этой точке возможны и отрицательные флуктуации ($\delta \mathcal{P}^- < 0$), которые следует анализировать отдельно.

Отрицательная начальная флуктуация $\delta\mathcal{P}^-$, как следует из ранее сказанного в связи с рис. 1.7 б, по модулю нарастает во времени, оставаясь отрицательной, тогда

$$\mathcal{P}_{\text{ex}}(t) = \delta\mathcal{P}^- e^{t/\tau} < 0. \quad (1.9.35)$$

Отрицательный знак избыточного производства энтропии, даваемого формулой (1.9.35), не противоречит тем не менее второму закону термодинамики, поскольку, согласно (1.9.21), полная величина $\mathcal{P}(t) = \mathcal{P}_{\text{min}} - |\mathcal{P}_{\text{ex}}(t)|$ всегда остается положительной. Однако по формуле (1.9.26) при $\mathcal{P}_{\text{ex}}(t) < 0$ скорость изменения $S_{\text{ex}}(t)$ во времени становится также отрицательной:

$$\frac{d}{dt} S_{\text{ex}}(t) = -|\mathcal{P}_{\text{ex}}(t)| < 0. \quad (1.9.36)$$

Подстановка выражения (1.9.35) под интеграл в (1.9.27) приводит к искомой временной зависимости избыточной энтропии:

$$S_{\text{ex}}(t) = (S_{\text{ex}}(0) + \tau|\delta\mathcal{P}^-|) - \tau|\delta\mathcal{P}^-| e^{t/\tau}. \quad (1.9.37)$$

Из формулы (1.9.37) видно, что избыточная энтропия $S_{\text{ex}}(t)$, независимо от модуля и знака ее начального значения $S_{\text{ex}}(0)$, за достаточно большое время становится отрицательной и экспоненциально нарастает по модулю, что в пределе дает

$$S_{\text{ex}}(t) \rightarrow -\infty \quad \text{при} \quad t \rightarrow \infty. \quad (1.9.38)$$

Это характеризует неустойчивость нелинейной открытой системы по отношению к отрицательным флуктуациям производства энтропии. Зависимость (1.9.37) качественно изображена при $S_{\text{ex}}(0) < 0$ на рис. 1.8 б в форме кривой 2.

Понятно, что предельная ситуация (1.9.38) на практике не реализуется, так как вступают в действие нелинейные эффекты в системе, которые рано или поздно остановят бесконечный рост и в конце концов сформируют новое стационарное состояние. По формуле (1.9.24) при $S_{\text{ex}}(t) < 0$ энтропия $S(t) = S_{\text{st}} - |S_{\text{ex}}(t)|$ уменьшается. Поэтому новому состоянию обязательно должны соответствовать структуры с более высокой степенью упорядоченности. Именно в этом и заключается *принцип самоорганизации* неравновесных систем. В таких условиях необратимость, в отличие от ее дезорганизующей функции в равновесных системах, начинает играть конструктивную роль, организуя при помощи флуктуаций порядок из хаоса.

Структурная упорядоченность/разупорядоченность системы непосредственным образом связана с ее флуктуационной неустойчивостью/устойчивостью. Дело в том, что термодинами-

ческие системы проявляют два типа поведения под действием флуктуаций и возмущений. *Вблизи равновесия* при малых воздействиях система самопроизвольно стремится к менее упорядоченным или неупорядоченным структурам, характерным для стационарного или равновесного состояния. *Вдали от равновесия*, при условии подчинения системы нелинейным законам, наблюдается спонтанное возникновение упорядоченности, порожденное неустойчивостью состояний, обладающих обычным термодинамическим (т. е. неупорядоченным) поведением.

Следовательно, в природе существуют два основных класса необратимых (неравновесных) процессов:

- *Разрушение* структуры *вблизи* положения равновесия, что является *универсальным* термодинамическим свойством всех систем. При этом выполняется второй закон термодинамики о положительном производстве энтропии внутри системы при произвольных внешних условиях.

- *Возникновение* структуры *вдали* от положения равновесия при *особых* внешних и внутренних условиях, что характерно для открытых систем с нелинейной внутренней динамикой. При этом внешние параметры превышают некоторые критические значения (ниже которых реализуются универсальные термодинамические свойства системы), так что отток энтропии (негэнтропия) преобладает над ее внутренним производством, уменьшая полную энтропию системы.

Из рассмотренного выше следует, что устойчивые и неустойчивые нелинейные системы отличаются друг от друга знаком избыточного производства энтропии. Согласно (1.9.22), (1.9.29) и (1.9.36), для этих систем можно записать следующие критерии:

$$\mathcal{P}_{\text{ex}} = \begin{cases} \sum_{\alpha} \delta J_{\alpha} \delta X_{\alpha} > 0 & \text{— критерий устойчивости,} \\ \sum_{\alpha} \delta J_{\alpha} \delta X_{\alpha} < 0 & \text{— критерий неустойчивости.} \end{cases} \quad (1.9.39)$$

Так как устойчивость в большинстве случаев, особенно для линейных систем, выполняется автоматически, то зачастую интереснее выяснить условия неустойчивости под действием внешних возмущений, существенно отклоняющих систему от равновесия. При $\mathcal{P}_{\text{ex}} < 0$ можно ожидать вступления в действие механизмов самоорганизации, приводящих к появлению наиболее упорядоченных структур, называемых **диссипативными структурами**. Важнейшим физическим условием их возникновения в неравновесной системе является взаимная согласованность (кооперативность) в поведении ее структурных элементов.

Кривая, изображенная на рис. 1.7 б, показывает, что в точке 2 отрицательные флуктуации производства энтропии ($\delta\mathcal{P}^- < 0$), подчиняясь требованию теоремы Пригожина ($d\mathcal{P} < 0$), могут, на первый взгляд, заставить систему скатиться из точки 2 вниз и оказаться в устойчивой стационарной точке 1. Однако такое заключение физически не может быть реализовано.

Дело в том, что такие флуктуации, как было показано выше, спонтанно нарастают (см. уравнение (1.9.35)), переводя систему в нелинейный режим. Дифференциальные уравнения, описывающие динамику процессов в системе, становятся нелинейными. Они отличаются от линейных уравнений (всегда имеющих единственное решение) многообразием возможных решений, называемых *ветвями*. Эти ветви зависят от внешних параметров λ , характеризующих степень отклонения системы от равновесия, где $\lambda = \lambda_{\text{eq}}$ (индекс eq от англ. *equilibrium*). При изменении λ ветви могут начать пересекаться в точках $\lambda = \lambda_{\text{bi}}$ (индекс bi от англ. *bifurcation*), называемых **точками бифуркации** (точками ветвления). Бифуркация в этой точке совершается путем перехода с одной (неустойчивой) ветви на другую (устойчивую).

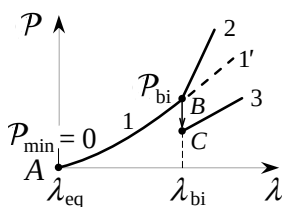


Рис. 1.9. Ветвление решений в точке бифуркации B при удалении системы от равновесия в точке A

Поведение ветвей решения нелинейного уравнения вблизи точки бифуркации B в форме зависимости производства энтропии $\mathcal{P}$ от параметра λ качественно показано на рис. 1.9. При малых отклонениях от равновесия (точка A , где минимальное производство энтропии равно нулю, $\mathcal{P}_{\min} = 0$) возникают стационарные структуры с $\mathcal{P}_{\min} > 0$ (см. кривую на рис. 1.7 б в районе точки 1). Такие структуры могут быть получены из равновесной путем непрерывной де-

формации и поэтому не имеют принципиальных отличий от нее.

Все эти состояния соответствуют так называемой *термодинамической ветви*, изображенной на рис. 1.9 в форме линии 1. При достижении параметром λ критического значения в точке бифуркации B , равного λ_{bi} , термодинамическая ветвь (продолжаемая при $\lambda > \lambda_{\text{bi}}$ в виде штриховой линии 1') становится неустойчивой. Система переходит в новый устойчивый режим, либо начинающийся в точке B с тем же значением $\mathcal{P}_{\text{bi}}$ (кривая 2), либо со скачкообразным *понижением* производства энтропии из точки B в точку C (кривая 3). При этом возникает более упорядоченная структура, существующая за счет диссипативных процессов с отличным от нуля минимальным производством энтропии.

Принципиальную роль в таких структурах, наряду с термодинамическими процессами, приобретают также гидродинамические, диффузионные и кинетические процессы, которые будут частично рассмотрены в последующих главах.

В общем случае возникновение диссипативных структур возможно при выполнении следующих четырех условий.

1. Термодинамическая система является открытой, т. е. обменивается с внешней средой как энергией, так и веществом.

2. Отклонение системы от равновесного состояния превышает некоторое критическое значение в точке бифуркации.

3. Макроскопические процессы в системе управляются законами нелинейной термодинамики и гидродинамики.

4. Микроскопические процессы в среде являются кооперативными, т. е. происходят согласованно во времени и в пространстве.

Математическое обоснование теоремы Пригожина и критерии устойчивости равновесных и стационарных состояний, включая формализм теории Онсагера, приведены ниже (см. пп. 1.17, 1.19, 1.20 и 1.24). Обсуждаются также физические свойства системы, гарантированно обеспечивающие отрицательную кривизну поверхности энтропии ($\delta^2 S < 0$) в зависимости от параметров, флуктуирующих около равновесного значения. Именно это условие, согласно неравенствам (1.9.34), необходимо для того, чтобы равновесное состояние любой системы было всегда устойчивым в силу неизбежного затухания флуктуаций. В п. 1.24 показано, что в линейной области при малых отклонениях от равновесия (в рамках применимости теории Онсагера) локальные значения тех же самых физических параметров обеспечивают флуктуационную устойчивость неравновесной системы. Кроме того, рассмотрены особенности анализа неустойчивости нелинейных систем, приводящей к повышению упорядоченности структуры.

1.10. Термодинамические потенциалы и соотношения между ними

Четыре термодинамических потенциала — внутренняя энергия U , энтальпия $H = U + PV$, свободная энергия Гельмгольца $F = U - TS$ и свободная энергия Гиббса $G = H - TS$, являются *функциями состояния*, т. е. каждая из них полностью описывает термодинамическое состояние системы при соответствующем выборе независимых переменных. Это означает, что, зная одну из величин U , H , F , G и ее частные производные, можно найти все остальные.

Выбор того или иного потенциала в качестве основного определяется физическими условиями рассматриваемой задачи. В качестве таковых обычно выступают *изохорно-изотермические* или *изобарно-изотермические* условия, при которых попарно работают следующие термодинамические функции:

а) U и F при $V = \text{const}$, $T = \text{const}$ или

б) H и G при $P = \text{const}$, $T = \text{const}$.

При этом приращения ΔU и ΔH , в соответствии с формулами (1.5.1) и (1.5.3), характеризуют тепловой эффект соответствующих процессов, а величины ΔF и ΔG , согласно (1.8.15) и (1.8.16), служат критерием для определения условий самопроизвольности неравновесных процессов и условий равновесия.

Используя выражения (1.3.5) и (1.5.2) для получения полных дифференциалов $dF = dU - d(TS)$ и $dG = dH - d(TS)$, запишем их совместно для всех четырех термодинамических потенциалов:

$$dU = TdS - PdV + \sum_i \mu_i dn_i, \quad (1.10.1)$$

$$dH = TdS + VdP + \sum_i \mu_i dn_i, \quad (1.10.2)$$

$$dF = -SdT - PdV + \sum_i \mu_i dn_i, \quad (1.10.3)$$

$$dG = -SdT + VdP + \sum_i \mu_i dn_i. \quad (1.10.4)$$

Отсюда очевидно, что независимыми переменными являются: S, V, n_i — для внутренней энергии U ; S, P, n_i — для энтальпии H ; T, V, n_i — для свободной энергии Гельмгольца (изохорного потенциала) F ; T, P, n_i — для свободной энергии Гиббса (изобарного потенциала) G . Отсюда следуют вполне понятные дифференциальные соотношения:

$$T = \left(\frac{\partial U}{\partial S} \right)_{V, n_i} = \left(\frac{\partial H}{\partial S} \right)_{P, n_i}, \quad (1.10.5)$$

$$S = - \left(\frac{\partial F}{\partial T} \right)_{V, n_i} = - \left(\frac{\partial G}{\partial T} \right)_{P, n_i}, \quad (1.10.6)$$

$$P = - \left(\frac{\partial U}{\partial V} \right)_{S, n_i} = - \left(\frac{\partial F}{\partial V} \right)_{T, n_i}, \quad (1.10.7)$$

$$V = \left(\frac{\partial H}{\partial P} \right)_{S, n_i} = \left(\frac{\partial G}{\partial P} \right)_{T, n_i}. \quad (1.10.8)$$

В этих равенствах, как и ранее, нижние индексы показывают, какие физические величины сохраняются постоянными при вычислении частных производных, например индекс n_i означает: $n_1 = \text{const}$, $n_2 = \text{const}$, ..., $n_K = \text{const}$ (для всех компонентов).

Взятием частных производных от четырех групп уравнений (1.10.5)–(1.10.8), с учетом независимости смешанной второй производной от порядка дифференцирования, получаем четыре новых уравнения:

$$\left(\frac{\partial T}{\partial V}\right)_{S, n_i} = -\left(\frac{\partial P}{\partial S}\right)_{V, n_i}, \quad \left(\frac{\partial T}{\partial P}\right)_{S, n_i} = \left(\frac{\partial V}{\partial S}\right)_{P, n_i}; \quad (1.10.9)$$

$$\left(\frac{\partial S}{\partial V}\right)_{T, n_i} = \left(\frac{\partial P}{\partial T}\right)_{V, n_i}, \quad \left(\frac{\partial S}{\partial P}\right)_{T, n_i} = -\left(\frac{\partial V}{\partial T}\right)_{P, n_i}. \quad (1.10.10)$$

Если в выражениях $U = F + TS$ и $H = G + TS$ исключить энтропию S с помощью равенств (1.10.6), то получим так называемые **уравнения Гиббса–Гельмгольца**:

$$U = F - T\left(\frac{\partial F}{\partial T}\right)_{V, n_i} \equiv -T^2 \frac{\partial}{\partial T} \left(\frac{F}{T}\right)_{V, n_i}, \quad (1.10.11)$$

$$H = G - T\left(\frac{\partial G}{\partial T}\right)_{P, n_i} \equiv -T^2 \frac{\partial}{\partial T} \left(\frac{G}{T}\right)_{P, n_i}. \quad (1.10.12)$$

При $T = \text{const}$ и $n_i = \text{const}$ дифференцированием уравнений (1.10.11) и (1.10.12) по V и P , соответственно, получаем следующие дифференциальные соотношения, называемые **уравнениями Гельмгольца**:

$$\left(\frac{\partial U}{\partial V}\right)_{T, n_i} = T^2 \frac{\partial}{\partial T} \left(\frac{P}{T}\right)_{V, n_i} \equiv -P + T\left(\frac{\partial P}{\partial T}\right)_{V, n_i}, \quad (1.10.13)$$

$$\left(\frac{\partial H}{\partial P}\right)_{T, n_i} = -T^2 \frac{\partial}{\partial T} \left(\frac{V}{T}\right)_{P, n_i} \equiv V - T\left(\frac{\partial V}{\partial T}\right)_{P, n_i}. \quad (1.10.14)$$

Уравнения (1.10.11) и (1.10.12) содержат функции состояния (см. п. 1.3). Поэтому записывая их дважды для начального состояния и конечного состояния, а затем вычитая первое из второго, получаем соответствующие уравнения для теплового эффекта изохорного и изобарного процессов в следующем виде (опуская вполне очевидные нижние индексы):

$$\Delta U = \Delta F - T \frac{\partial \Delta F}{\partial T} \equiv -T^2 \frac{\partial}{\partial T} \frac{\Delta F}{T}, \quad (1.10.15)$$

$$\Delta H = \Delta G - T \frac{\partial \Delta G}{\partial T} \equiv -T^2 \frac{\partial}{\partial T} \frac{\Delta G}{T}. \quad (1.10.16)$$

Уравнения Гиббса–Гельмгольца в форме (1.10.15) и (1.10.16) будут использованы ниже (см. п. 2.14) для вывода уравнений изохоры и изобары Вант-Гоффа.

1.11. Химический потенциал и парциальные молярные величины

1. Свойства химического потенциала. Химический потенциал i -го компонента (μ_i) входит в выражения (1.10.1)–(1.10.4) как коэффициент пропорциональности между полным дифференциалом каждой термодинамической функции и приращением dn_i числа молей i -го сорта при постоянстве других независимых переменных, включая $n_j = \text{const}$ при $j \neq i$. На этом основании записываем следующие равенства:

$$\begin{aligned} \mu_i &= \left(\frac{\partial U}{\partial n_i} \right)_{S, V, n_j \neq n_i} = \left(\frac{\partial H}{\partial n_i} \right)_{S, P, n_j \neq n_i} = \\ &= \left(\frac{\partial F}{\partial n_i} \right)_{T, V, n_j \neq n_i} = \left(\frac{\partial G}{\partial n_i} \right)_{T, P, n_j \neq n_i}. \end{aligned} \quad (1.11.1)$$

Отсюда следует физический смысл химического потенциала: *химический потенциал μ_i компонента i выражает скорость изменения любого термодинамического потенциала U, H, F, G при увеличении числа n_i частиц (молей) данного сорта в условиях постоянства других независимых переменных.*

Покажем, что химический потенциал (точнее, его пространственная неоднородность) является движущей силой процесса массопереноса в условиях неоднородного распределения частиц.

Пусть химический потенциал i -го компонента в одной части системы равен μ'_i , а в другой — μ''_i . Если число частиц в этих областях изменяется, соответственно, на dn'_i и dn''_i , так что $n'_i + n''_i = \text{const}$, то, согласно (1.10.4), изменение изобарного потенциала системы, происходящее в результате переноса частиц между двумя областями при $T = \text{const}$ и $P = \text{const}$, равно

$$dG = \mu'_i dn'_i + \mu''_i dn''_i = (\mu'_i - \mu''_i) dn'_i = -(\mu'_i - \mu''_i) dn''_i < 0.$$

Здесь неравенство записано на основании формулы (1.8.8), выражающей условие самопроизвольного протекания изобарно-

изотермического процесса, откуда немедленно следует, что

$$dn'_i < 0 \quad \text{и} \quad dn''_i > 0 \quad \text{при} \quad \mu'_i > \mu''_i.$$

Таким образом, частицы любого компонента *самопроизвольно* переходят из области с большим химическим потенциалом в область с меньшим его значением, стремясь обеспечить однородное распределение химического потенциала в пространстве. Как увидим ниже (см. п. 1.14), пространственно однородное распределение химического потенциала μ_i в системе является условием одновременного сосуществования нескольких равновесных фаз, содержащих i -й компонент.

Из четырех термодинамических функций U , H , F , G особое место занимает изобарно-изотермический потенциал G . Он является функцией одновременно температуры T и давления P , которые относятся к числу так называемых **интенсивных величин**, не зависящих от количества вещества в системе.

Кроме T и P , к интенсивным величинам также относится концентрация компонентов, выраженная в виде *мольно-объемной концентрации* $c_i = n_i/V$ или *мольной доли* $x_i = n_i/n$, где $n = n_1 + n_2 + \dots + n_K$ — суммарное число молей в системе.

В отличие от этого, все энергетические величины, такие как U , H , F , G , а также S и V , являясь аддитивными, пропорциональны количеству вещества в системе и по этой причине называются **экстенсивными величинами**.

Воспользуемся указанным выше свойством экстенсивности свободной энергии Гиббса и увеличим в λ раз число частиц каждого сорта в рассматриваемой системе. Тогда во столько же раз возрастет и величина G при сохранении интенсивных величин T и P неизменными:

$$\lambda G(T, P, n_1, n_2, \dots, n_K) = G(T, P, \lambda n_1, \lambda n_2, \dots, \lambda n_K). \quad (1.11.2)$$

Дифференцируя левую и правую части выражения (1.11.2) по λ , приходим к следующему равенству:

$$G(T, P, n_i) = \sum_{i=1}^K \frac{\partial G(T, P, \lambda n_i)}{\partial (\lambda n_i)} \frac{\partial (\lambda n_i)}{\partial \lambda} = \sum_{i=1}^K \frac{\partial G(T, P, n_i)}{\partial n_i} n_i.$$

Отсюда, используя определение химического потенциала в форме (1.11.1), получаем следующее важное соотношение:

$$G = \sum_i n_i \mu_i \quad \text{при} \quad \mu_i = \left(\frac{\partial G}{\partial n_i} \right)_{T, P, n_j \neq n_i}. \quad (1.11.3)$$

Используем равенство (1.11.3) для получения соотношений, которым подчиняются химические потенциалы. На его основании записываем полный дифференциал изобарного потенциала как

$$dG = \sum_i n_i d\mu_i + \sum_i \mu_i dn_i. \quad (1.11.4)$$

Сравнение выражений (1.10.4) и (1.11.4) приводит к равенству, известному как **уравнение Гиббса–Дюгема**¹⁾:

$$SdT - VdP + \sum_i n_i d\mu_i = 0. \quad (1.11.5)$$

Это уравнение связывает между собой вариации интенсивных величин T , P и μ_i . В условиях $T = \text{const}$ и $P = \text{const}$ из равенства (1.11.5) следует упрощенная форма уравнения Гиббса–Дюгема,

$$\sum_i n_i (d\mu_i)_{T,P} = 0, \quad (1.11.6)$$

которая устанавливает связь между изменениями химических потенциалов компонентов в изобарно-изотермических процессах.

Таким образом, химические потенциалы компонентов в системе не являются независимыми, а связаны между собой уравнением Гиббса–Дюгема (1.11.5) или (1.11.6). Для каждого компонента химический потенциал зависит от температуры, давления и состава многокомпонентной системы, т. е.

$$\mu_i = \mu_i(T, P, n_1, n_2, \dots, n_K). \quad (1.11.7)$$

2. Парциальные молярные величины. Следует подчеркнуть, что соотношение (1.11.3) имеет место только для изобарного потенциала G и не выполняется для других термодинамических функций, так как среди своих независимых переменных они содержат экстенсивные величины, в частности $F \neq \sum_i n_i \mu_i$. Однако при постоянных T и P все без исключения экстенсивные величины U , H , F , G , S , V , подобно изобарному потенциалу G , обладают свойством аддитивности по i , т. е. для каждой из них можно, по аналогии с равенством (1.11.3), записать:

$$U = \sum_i n_i \bar{U}_i, \quad \text{где} \quad \bar{U}_i = \left(\frac{\partial U}{\partial n_i} \right)_{T,P,n_j \neq n_i}; \quad (1.11.8)$$

¹⁾ Дюгем (Duhem) Пьер (1861–1916) — французский физик-теоретик, философ и историк науки, работал в области термодинамики, гидродинамики, теории упругости и магнетизма, наибольшую известность получил за философскую интерпретацию физических законов.

$$H = \sum_i n_i \bar{H}_i, \quad \text{где} \quad \bar{H}_i = \left(\frac{\partial H}{\partial n_i} \right)_{T, P, n_j \neq n_i}; \quad (1.11.9)$$

$$F = \sum_i n_i \bar{F}_i, \quad \text{где} \quad \bar{F}_i = \left(\frac{\partial F}{\partial n_i} \right)_{T, P, n_j \neq n_i}; \quad (1.11.10)$$

$$S = \sum_i n_i \bar{S}_i, \quad \text{где} \quad \bar{S}_i = \left(\frac{\partial S}{\partial n_i} \right)_{T, P, n_j \neq n_i}; \quad (1.11.11)$$

$$V = \sum_i n_i \bar{V}_i, \quad \text{где} \quad \bar{V}_i = \left(\frac{\partial V}{\partial n_i} \right)_{T, P, n_j \neq n_i}. \quad (1.11.12)$$

Величины $\bar{U}_i$, $\bar{H}_i$, $\bar{F}_i$, $\bar{S}_i$, $\bar{V}_i$ (отмеченные чертой сверху) называют *парциальными молярными* (или *мольными*) *величинами*. Они характеризуют *скорость изменения* соответствующих величин (без черты) при изменении числа молей i -го сорта *в изобарно-изотермических условиях*. Из сравнения (1.11.3) с формулами (1.11.8)–(1.11.12) следует, что химический потенциал μ_i представляет собой парциальный изобарный потенциал $\bar{G}_i$. Парциальные мольные величины играют важную роль в термодинамике растворов (см. п. 2.4). Их использование удобно по той причине, что условия термодинамического равновесия (1.14.1) и (1.14.2) требуют именно $T = \text{const}$ и $P = \text{const}$ (см. п. 1.14).

Нетрудно увидеть, что парциальные мольные величины связаны между собой теми же самыми соотношениями, которые имеют место для аналогичных полных величин:

$$\bar{H}_i = \bar{U}_i + P\bar{V}_i, \quad (1.11.13)$$

$$\bar{F}_i = \bar{U}_i - T\bar{S}_i, \quad (1.11.14)$$

$$\bar{G}_i = \bar{H}_i - T\bar{S}_i \equiv \mu_i. \quad (1.11.15)$$

Частные производные, вычисленные для функции (1.11.7) с учетом формул (1.10.6), (1.10.8), (1.11.1), (1.11.3), (1.11.11) и (1.11.12), равны (индексы n_j , n_k означают $n_j \neq n_i$, $n_k \neq n_i(j)$):

$$\left(\frac{\partial \mu_i}{\partial T} \right)_{P, n_i} = \left(\frac{\partial^2 G}{\partial n_i \partial T} \right)_{P, n_j} = - \left(\frac{\partial S}{\partial n_i} \right)_{T, P, n_j} \equiv -\bar{S}_i, \quad (1.11.16)$$

$$\left(\frac{\partial \mu_i}{\partial P} \right)_{T, n_i} = \left(\frac{\partial^2 G}{\partial n_i \partial P} \right)_{T, n_j} = \left(\frac{\partial V}{\partial n_i} \right)_{T, P, n_j} \equiv \bar{V}_i, \quad (1.11.17)$$

$$\left(\frac{\partial \mu_i}{\partial n_j} \right)_{T, P, n_i} = \left(\frac{\partial^2 G}{\partial n_i \partial n_j} \right)_{T, P, n_k} = \left(\frac{\partial \mu_j}{\partial n_i} \right)_{T, P, n_j}. \quad (1.11.18)$$

С учетом того, что дифференциал функции (1.11.7) при $T = \text{const}$ и $P = \text{const}$ равен

$$(d\mu_i)_{T,P} = \sum_j \left(\frac{\partial \mu_i}{\partial n_j} \right)_{T,P} dn_j,$$

уравнение Гиббса–Дюгема в форме (1.11.6) можно записать как

$$\sum_i \sum_j n_i \left(\frac{\partial \mu_i}{\partial n_j} \right)_{T,P} dn_j = \sum_j \left(\sum_i n_i \left(\frac{\partial \mu_i}{\partial n_j} \right)_{T,P} \right) dn_j = 0.$$

Поскольку здесь dn_j суть произвольные приращения независимых переменных, то общее нулевое равенство выполняется, только если множитель при каждом dn_j (в виде суммы по i , стоящей в круглых скобках) по отдельности равен нулю. На этом основании записываем следующие равенства:

$$\sum_i n_i \left(\frac{\partial \mu_i}{\partial n_j} \right)_{T,P} = 0 \quad \text{или} \quad \sum_i n_i \left(\frac{\partial \mu_j}{\partial n_i} \right)_{T,P} = 0, \quad (1.11.19)$$

где использована формула (1.11.18).

Поскольку химический потенциал является парциальной молярной свободной энергией Гиббса ($\mu_i \equiv \bar{G}_i$), то равенства, аналогичные (1.11.19), могут быть записаны и для других парциальных молярных величин (1.11.8)–(1.11.12). Для этого в каждой производной, входящей в равенства (1.11.19), нужно заменить $\mu_i(\mu_j)$ на $\bar{U}_i(\bar{U}_j)$, $\bar{H}_i(\bar{H}_j)$, $\bar{F}_i(\bar{F}_j)$, $\bar{S}_i(\bar{S}_j)$, $\bar{V}_i(\bar{V}_j)$ (см. формулы (1.11.8)–(1.11.12)).

Все парциальные молярные величины, введенные равенствами (1.11.3) и (1.11.8)–(1.11.12), в предельном случае $x_i \rightarrow 1$ должны давать молярные величины для *чистого* вещества i (что отмечено верхним нулевым индексом):

$$\begin{aligned} \bar{G}_i &\equiv \mu_i \rightarrow \mu_i^0, & \bar{U}_i &\rightarrow U_i^0, & \bar{H}_i &\rightarrow H_i^0, \\ \bar{F}_i &\rightarrow F_i^0, & \bar{S}_i &\rightarrow S_i^0, & \bar{V}_i &\rightarrow V_i^0. \end{aligned} \quad (1.11.20)$$

В частности, для чистого i -го вещества вместо (1.11.16) и (1.11.17) имеем

$$\left(\frac{\partial \mu_i}{\partial T} \right)_P = -S_i^0 < 0 \quad \text{и} \quad \left(\frac{\partial \mu_i}{\partial P} \right)_T = V_i^0 > 0. \quad (1.11.21)$$

Разумеется, для чистого вещества молярные величины (отмеченные верхним нулевым индексом) всегда положительны, что и определяет неравенства в формуле (1.11.21). В то же время в выражениях (1.11.8)–(1.11.12) соответствующие парциальные

величины (отмеченные чертой сверху) могут быть любого знака для многокомпонентных систем (жидких и твердых растворов).

Конкретный вид функциональных зависимостей (1.11.7) в общем случае установить достаточно сложно, так как это требует знания деталей взаимодействия между частицами. Однако для *идеальных систем*, в которых роль таких взаимодействий пренебрежимо мала, можно получить фундаментальные соотношения, которые далее будут обобщены на реальные растворы.

1.12. Термодинамические функции идеального газа

Идеальным газом принято называть газ, в котором потенциальная энергия взаимодействия между молекулами (проявляющаяся в моменты их столкновения) пренебрежимо мала по сравнению с кинетической энергией теплового хаотического движения (совершаемого на длине свободного пробега молекул).

Пренебрежение взаимодействием между газовыми молекулами в рамках модели идеального газа позволяет записать универсальное уравнение газового состояния, называемое **уравнением Клапейрона–Менделеева**¹⁾, в одной из следующих форм:

$$PV = Nk_B T, \quad PV = nRT, \quad P = cRT. \quad (1.12.1)$$

Здесь $n = N/N_A$ — число газовых молей в объеме V , $c = N/V$ — мольно-объемная концентрация (моль/литр), $R = k_B N_A = 8,314$ Дж/(моль · К) ≈ 2 кал/(моль · К) — универсальная газовая постоянная. В силу пренебрежения взаимодействием между молекулами, каждое из уравнений (1.12.1) не зависит от сорта газа и поэтому применимо к смесям идеальных газов.

Величина P , входящая в равенства (1.12.1), представляет собой суммарное давление, создаваемое всеми газообразными

¹⁾ Клапейрон (Clapeyron) Бенуа Поль Эмиль (1799–1864) — французский физик, работавший в области теплоты, пластичности и равновесия твердых тел, придал математическую форму идеям Карно, объединил законы Бойля–Мариотта, Гей–Люссака и Авогадро в единое уравнение состояния идеального газа (1834), обобщенное позднее (1874) Д.И. Менделеевым, вывел уравнение связи температуры плавления и кипения вещества с давлением, термодинамически обоснованное Р. Клаузиусом (1851) (уравнение Клапейрона–Клаузиуса).

Менделеев Дмитрий Иванович (1834–1907) — русский ученый-энциклопедист, автор фундаментальных исследований по химии, химической технологии, физике, метрологии, метеорологии, геологии, экономике и воздухоплаванию. Помимо открытия периодического закона химических элементов (1869), Д.И. Менделеев заложил основы теории растворов, предложил промышленный способ фракционного разделения нефти и изобрел вид бездымного пороха.

компонентами в объеме V и подчиняющееся **закону Дальтона**¹⁾:

$$P = \sum_i p_i = \sum_i c_i RT. \quad (1.12.2)$$

Здесь $p_i = c_i RT$ и $c_i = n_i/V$ — парциальное давление и мольно-объемная концентрация i -го компонента. Следует отметить, что парциальное давление не относится к числу парциальных молярных величин, ранее введенных соотношениями (1.11.8)–(1.11.12); совпадение названий чисто терминологическое.

Кроме *мольно-объемных концентраций* c_i , для характеристики состава многокомпонентной системы (газовой смеси, жидкого или твердого раствора) часто используют *мольные доли* x_i , равные долевого вкладу i -го вещества в один моль системы:

$$x_i = \frac{n_i}{n} = \frac{c_i}{c} = \frac{n_i}{\sum_i n_i} = \frac{N_i}{\sum_i N_i}, \quad \text{при этом} \quad \sum_{i=1}^K x_i = 1. \quad (1.12.3)$$

Для смеси идеальных газов переписываем уравнение $p_i = c_i RT$ с помощью выражений (1.12.1) и (1.12.3) в следующем виде:

$$p_i = x_i P. \quad (1.12.4)$$

1. Однокомпонентный газ. Начнем анализ с однокомпонентного газа в пренебрежении внутренними (колебательными и вращательными) степенями свободы газовой молекулы. В силу классического закона равнораспределения энергии по степеням свободы, каждая молекула, совершающая поступательное движение, имеет среднюю кинетическую энергию, равную $3k_B T/2$.

Это означает, что внутренняя энергия U и энтальпия H идеального газа не зависят от объема V и давления P , а определяются лишь температурой T :

$$U(T) = \frac{3}{2} nRT \quad \text{и} \quad H(T) = U + PV = \frac{5}{2} nRT. \quad (1.12.5)$$

Для нахождения других термодинамических величин используем уравнение (1.10.2), переписанное в виде

$$dS = \frac{1}{T} dH - \frac{V}{T} dP = \frac{C_P}{T} dT - \frac{nR}{P} dP, \quad (1.12.6)$$

где последнее равенство записано с учетом (1.5.5) и (1.12.1).

¹⁾ Дальтон (Dalton) Джон (1766–1844) — английский физик и химик, сформулировал закон парциальных давлений газов, обнаружил зависимость их растворимости от парциального давления, ввел понятие атомной массы и, приняв массу атома водорода за единицу, рассчитал атомные веса ряда элементов и составил первую таблицу относительных атомных масс (1803).

Интегрирование уравнения (1.12.6) дает соотношение

$$\begin{aligned} S(T, P) &= \int \frac{C_P(T)}{T} dT - nR \int \frac{dP}{P} = \\ &= S^\circ(T) - nR \ln \frac{P}{P^\circ}. \end{aligned} \quad (1.12.7)$$

Здесь слагаемое $S^\circ(T)$ учитывает температурную зависимость энтропии, а величина P° , появившаяся в составе постоянной интегрирования $C = nR \ln P^\circ$, носит название *стандартного давления*, введенного таким образом, чтобы $S(T, P^\circ) = S^\circ(T)$. При этом верхний индекс $^\circ$ обозначает так называемое *стандартное состояние*, определенное ниже (см. п. 1.13).

Интегрирование уравнения $dH = C_P dT$ дает для идеального газа температурную зависимость энтальпии в форме (1.12.5):

$$H(T) = \int C_P(T) dT = \frac{5}{2} nRT. \quad (1.12.8)$$

Из формул (1.12.7) и (1.12.8) получаем следующие выражения для однокомпонентного газа:

- свободная энергия Гиббса

$$G(T, P) = H(T) - TS(T, P) = G^\circ(T) + nRT \ln \frac{P}{P^\circ}, \quad (1.12.9)$$

- химический потенциал

$$\mu(T, P) = \frac{G(T, P)}{n} = \mu^\circ(T) + RT \ln \frac{P}{P^\circ}, \quad (1.12.10)$$

где $G^\circ(T) = H(T) - TS^\circ(T)$ и $\mu^\circ(T) = G^\circ(T)/n$ — изобарный потенциал и химический потенциал газа в стандартном состоянии.

2. Многокомпонентный газ. Поскольку в смеси идеальных газов каждый газообразный компонент ведет себя независимо от других, то выражение (1.12.10) может быть обобщено для i -го компонента в следующей форме:

$$\mu_i(T, p_i) = \mu_i^\circ(T) + RT \ln \frac{p_i}{p_i^\circ}. \quad (1.12.11)$$

Отсюда с помощью равенств $p_i = c_i RT = x_i P$ получаем две другие формы записи химического потенциала для i -го компонента в газовой смеси:

$$\mu_i(T, c_i) = \mu_i^\circ(T) + RT \ln \frac{c_i}{c_i^\circ}, \quad (1.12.12)$$

$$\mu_i(T, P, x_i) = \mu_i^\circ(T, P) + RT \ln x_i, \quad (1.12.13)$$

где введены следующие обозначения: $c_i^\circ = p_i^\circ/RT$ и $\mu_i^\circ(T, P) = \mu_i^\circ(T) + RT \ln(P/p_i^\circ)$.

В формулах (1.12.11)–(1.12.13) первые слагаемые соответствуют химическому потенциалу в стандартном состоянии (отмеченном верхним кружком), которое рассмотрено в следующем параграфе. Особенность модели идеального газа отражена в этих формулах в виде логарифмической зависимости химического потенциала от парциального давления p_i и концентрации c_i или x_i для i -го компонента.

3. Изобарно-изотермическое смешение газов. Рассмотрим процесс образования смеси идеальных газов при условии сохранения неизменными полного давления ($P = \text{const}$) и температуры ($T = \text{const}$). *Чистые компоненты* (до смешения) будем характеризовать, подобно (1.11.20), следующими молярными величинами:

$$V_i^0, \quad S_i^0, \quad U_i^0, \quad H_i^0 = U_i^0 + P_i^0 V_i^0, \quad G_i^0 = H_i^0 - T S_i^0, \quad (1.12.14)$$

отмечая их *верхним нулевым* индексом.

Пусть до смешения каждый газовый компонент имел n_i молей в объеме $n_i V_i^0$ при температуре T и одинаковых давлениях $P_i^0 = P$. В процессе смешения температура T и общее давление P в результирующем суммарном объеме,

$$V = \sum_i n_i V_i^0 \equiv V_0, \quad (1.12.15)$$

остаются неизменными, т. е. $\Delta V_{\text{см}} \equiv V - V_0 = 0$. Здесь объем V_0 , а также использованные ниже суммарные величины U_0 , H_0 , G_0 соответствуют исходному состоянию до смешения газов, отмеченному теперь *нижним нулевым* индексом.

В то же время парциальные давления p_i для каждого компонента в силу увеличения объема при смешении снижаются от исходной величины $P_i^0 = P$ до $p_i = x_i P < P$, где $x_i = n_i/n$ — молярная доля i -го компонента после смешения. Из-за идеальности газов их внутренняя энергия после смешения равна сумме энергий до смешения:

$$U = \sum_i n_i U_i^0 \equiv U_0. \quad (1.12.16)$$

Здесь $U_i^0 = 3RT/2$ — внутренняя энергия одного моля газа в отсутствие теплового возбуждения внутренних (колебательных и вращательных) степеней свободы молекулы.

На основании формул (1.12.14)–(1.12.16) для энтальпии газа после смешения получаем (с учетом того, что $P = P_i^0$)

$$H = U + PV = \sum_i n_i (U_i^0 + P V_i^0) = \sum_i n_i H_i^0 \equiv H_0. \quad (1.12.17)$$

Следовательно, тепловой эффект изобарно-изотермического смешения идеальных газов отсутствует, поскольку

$$\Delta H_{\text{см}} \equiv H - H_0 = H - \sum_i n_i H_i^0 = 0. \quad (1.12.18)$$

Для нахождения изобарного потенциала G газовой смеси используем формулы (1.11.3), (1.12.14), а также выражение для химического потенциала в форме (1.12.13). Тогда получаем искомым результат в виде

$$\begin{aligned} G &= \sum_i n_i \mu_i = \sum_i n_i \mu_i^\circ + RT \sum_i n_i \ln x_i \equiv \\ &\equiv G_0 + nRT \sum_i x_i \ln x_i, \end{aligned} \quad (1.12.19)$$

где $G_0 = \sum_i n_i \mu_i^\circ \equiv \sum_i n_i G_i^0$ — суммарный изобарный потенциал чистых газообразных компонентов до смешения.

Из выражения (1.12.19) следует *изобарный потенциал смешения идеальных газов*:

$$\Delta G_{\text{см}} \equiv G - G_0 = nRT \sum_i x_i \ln x_i. \quad (1.12.20)$$

Исходя из общего соотношения (1.8.14), записываем $\Delta G_{\text{см}} = \Delta H_{\text{см}} - T \Delta S_{\text{см}}$. Следовательно, в отсутствие теплоты смешения ($\Delta H_{\text{см}} = 0$) из уравнения (1.12.20) получаем так называемую **энтропию смешения**:

$$\Delta S_{\text{см}} = -\frac{1}{T} \Delta G_{\text{см}} = -nR \sum_i x_i \ln x_i. \quad (1.12.21)$$

Поскольку в результате смешения газов для каждого компонента его мольная доля уменьшается ($x_i < 1$), то имеют место неравенства $\Delta G_{\text{см}} < 0$ и $\Delta S_{\text{см}} > 0$.

Полученный результат физически означает, что процесс изобарно-изотермического смешения газов всегда происходит *самопроизвольно* (из-за $\Delta G_{\text{см}} < 0$) путем *необратимого* (в силу $\Delta S_{\text{см}} > 0$) *расширения* каждого компонента из начального объема $n_i V_i^0$ в конечный суммарный объем $V = \sum_i n_i V_i^0$. При этом парциальные давления газов уменьшаются от исходного значения $P_i^0 = P$ до $p_i = x_i P < P$, и процесс смешения не сопровождается тепловым эффектом ($\Delta H_{\text{см}} = 0$).

Отмеченная выше особенность изобарно-изотермического смешения идеальных газов обобщается на процесс образования идеальных растворов. В отличие от идеальных газов, потенциальная энергия взаимодействия между частицами идеального раствора

сравнима с кинетической энергией их теплового движения, но практически одинакова для однородных и разнородных частиц. Как и для идеальных газов, образование идеального раствора происходит самопроизвольно ($\Delta G < 0$), необратимо ($\Delta S > 0$) и не сопровождается объемным эффектом ($\Delta V = 0$) и тепловым эффектом ($\Delta H = 0$) (см. п. 2.4).

1.13. Понятие о стандартном состоянии веществ и таблицах стандартных термодинамических величин

При любых термодинамических расчетах практический интерес представляют не абсолютные значения величин, а их относительные изменения, происходящие в результате того или иного процесса. Отсюда возникает необходимость отсчета реального состояния многообразных веществ от некоторого условного состояния, единого для всех веществ и называемого **стандартным состоянием**. В этом смысле химический потенциал аналогичен электрическому, который принято отсчитывать от потенциала заземления, условно принимаемого за нуль (о связи химического потенциала μ_i и электрического потенциала φ в форме электрохимического потенциала $\tilde{\mu}_i$ см. в п. 1.16).

Химический потенциал идеального газа был получен в трех формах (1.12.11)–(1.12.13), которые применимы и для идеальных растворов (их обобщение на реальные растворы выполнено ниже, см. формулу (1.13.4)). Для газов обычно используют выражение (1.12.11), а для жидких и твердых растворов — либо (1.12.12), либо (1.12.13). Отсюда видно, что μ_i равняется стандартному химическому потенциалу μ_i° , принимаемому за точку отсчета, при равенстве единице величин, стоящих под знаком логарифма.

Таким образом, в соответствии с тремя формами записи химического потенциала (формулы (1.12.11)–(1.12.13)) возможны три варианта выбора стандартного состояния:

1) для химического потенциала, записанного в форме выражения (1.12.11), условились полагать $p_i^\circ = 1$ атм, так что стандартное состояние соответствует газовой смеси с **единичным парциальным давлением** ($p_i = 1$ атм) для *каждого* компонента;

2) для химического потенциала, записанного в форме выражения (1.12.12), условились полагать $c_i^\circ = 1$ моль/л, так что стандартное состояние соответствует раствору с **единичной мольно-объемной концентрацией** ($c_i = 1$ моль/л) для *каждого* компонента;

3) для химического потенциала, записанного в форме выражения (1.12.13), стандартное состояние жидкого или твердого раствора соответствует совокупности входящих в раствор независимых компонентов, взятых **в чистом виде**, при этом $x_i = 1$ (для реальных растворов молярные доли x_i заменяют активностями $a_i = \gamma_i x_i$ — см. формулу (1.13.4) и п. 2.4).

Для многокомпонентных конденсированных сред чаще используют третий вариант стандартного состояния в виде набора компонентов раствора в чистом виде, хотя применяется и второй вариант, например при анализе жидких растворов и для квази-химического описания дефектов в кристаллах (см. гл. 3).

Итак, в стандартном состоянии все вещества берутся в той агрегатной форме, которая для них устойчива при рассматриваемой температуре (обычно $T = 298 \text{ К}$) и общем давлении $P = 1 \text{ атм}$, при этом парциальное давление p_i каждого газообразного компонента принимается равным одной атмосфере, а конденсированные вещества (жидкие и твердые), входящие в состав растворов, берутся в чистом виде.

На основании вышесказанного, выбор стандартных значений в форме $p_i^\circ = 1 \text{ атм}$ и $c_i^\circ = 1 \text{ моль/л}$ позволяет записать выражения (1.12.11)–(1.12.13) для химического потенциала в виде

$$\mu_i(T, p_i) = \mu_i^\circ(T) + RT \ln p_i, \quad (1.13.1)$$

$$\mu_i(T, c_i) = \mu_i^\circ(T) + RT \ln c_i, \quad (1.13.2)$$

$$\mu_i(T, P, x_i) = \mu_i^\circ(T, P) + RT \ln x_i. \quad (1.13.3)$$

Здесь под логарифмами стоят *безразмерные величины*, численно равные количеству атмосфер, количеству молей в единице объема и молярной доле i -го компонента. Это же относится и ко всем последующим формулам, при выводе которых использованы равенства (1.13.1)–(1.13.3), в частности, к выведенному ниже закону действующих масс (см. п. 2.12).

Выражение (1.13.1) вполне применимо для описания реальных газов. Однако в применении к реальным растворам формулы (1.13.2) и (1.13.3) обычно модифицируют следующим образом:

$$\mu_i = \mu_i^\circ + RT \ln a_i. \quad (1.13.4)$$

Здесь вместо c_i или x_i введена так называемая *активность* $a_i = \gamma_i c_i$ или $a_i = \gamma_i x_i$, характеризующаяся *коэффициентом активности* γ_i , который учитывает взаимодействие между частицами раствора и в общем случае зависит от его состава и температуры (см. п. 2.4).

Следует отметить, что в выражениях (1.13.1)–(1.13.4) химический потенциал соответствует одному молю вещества. Переход к величинам для одной частицы осуществляется делением на N_A , то есть везде надо заменить R на k_B .

Введение стандартного состояния в указанных выше формах диктуется практической необходимостью, возникающей в связи с выполнением термодинамического анализа многообразных физико-химических систем.

Действительно, при одинаковых для всех веществ значениях температуры (обычно $T = 298$ К) и давления ($P = 1$ атм), а также при фиксированном составе системы (заданном единичными концентрациями и парциальными давлениями) стандартные значения различных термодинамических величин отражают только внутреннюю специфику конкретных веществ. Это позволяет сравнивать вещества по тем или иным физико-химическим признакам и проводить термодинамические расчеты фазовых и химических равновесий с целью выяснения принципиальной возможности фазовых и химических превращений.

Для этой цели созданы таблицы термодинамических свойств веществ в стандартных условиях. Таблицы стандартных термодинамических величин обычно соответствуют температуре $T = 298$ К и обязательно содержат следующие *молярные величины* (см. табл. А.1 и А.2 в приложении А.1):

- во-первых, стандартные теплоты ΔH_{298}° (кал/моль) образования¹⁾ химических соединений из простых веществ (для последних $\Delta H_{298}^\circ = 0$ — см. п. 1.5);
- во-вторых, стандартные энтропии S_{298}° (кал/(моль·К)), которые, в отличие от энтальпии, на основании теоремы Нернста (1.2.1) имеют абсолютные значения, отсчитанные от общей нулевой величины при $T = 0$ К, так что для всех без исключения веществ, включая простые вещества, имеем $S_{298}^\circ > 0$;
- в-третьих, температурные зависимости теплоемкости $C_P(T)$ (кал/(моль·К)), необходимые для пересчета величин ΔH_{298}° и S_{298}° на другие температуры при вычислении ΔH_T° и S_T° с помощью формул (1.5.11).

Методика пользования таблицами стандартных термодинамических величин и выполнения термодинамических расчетов изложена в приложениях А.1 и А.2.

¹⁾ Нередко в справочной литературе энтальпию образования ΔH_{298}° обозначают в виде ΔH_{f298}° с добавлением индекса f от англ. *formation*.

1.14. Условия фазового и химического равновесия

Фазой называется совокупность однородных частей системы, одинаковых во всех точках по составу, структуре, физическим и химическим свойствам и имеющих видимые границы раздела.

По составу фазы могут быть однокомпонентными и многокомпонентными. При этом под *компонентом* понимается любое составляющее вещество (элементарное или в виде химического соединения), которое может быть выделено из системы и существовать вне нее. Такие многокомпонентные системы, как смеси газов, жидкие и твердые растворы, а также химически однородные вещества в одном агрегатном состоянии, являются однофазными или *гомогенными системами*. В многофазных системах, называемых *гетерогенными системами*, существует единственная газовая (или парогазовая) фаза, в то время как конденсированных фаз (жидких и твердых) может быть несколько.

1. Фазовое равновесие подразумевает одновременное сосуществование двух или более фаз. Оно осуществляется путем непрерывного обмена частицами между фазами, то есть фазовое равновесие имеет *динамический характер*: число частиц, перешедших из одной фазы в другую за единицу времени, в среднем равно числу частиц, совершивших обратный переход.

Условия фазового равновесия представляют собой математическую формулировку соотношений между физическими величинами, обеспечивающих существование нескольких фаз в многокомпонентной системе *в отсутствие* химических реакций.

Для их вывода будем рассматривать интересующую нас многофазную систему вместе с внешней средой как замкнутую систему, условием равновесия которой является максимум суммарной энтропии, что записано в виде соотношения (1.8.2).

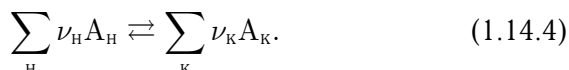
В отличие от обозначений, ранее использованных в п. 1.2, номер фазы будем отмечать верхними римскими цифрами I, II, ..., сохраняя нижние арабские цифры 1, 2, ... для обозначения номера компонента. В применении к конкретным физическим условиям римские цифры будут при необходимости заменяться буквами, явно указывающими на состояние вещества: г — газообразное, п — парообразное, ж — жидкое, т — твердое и к — конденсированное (жидкое или твердое).

Рассмотрим две фазы, содержащие K компонентов и составляющие замкнутую систему, для которой имеем $dU^I = -dU^{II}$, $dV^I = -dV^{II}$, $dn_i^I = -dn_i^{II}$, $i = 1, 2, \dots, K$. На основании фундаментального уравнения Гиббса (1.3.5) записываем дифференциал

в среднем за единицу времени равно числу элементарных актов в обратном направлении.

Условие химического равновесия обеспечивает неизменный состав сосуществующих многокомпонентных фаз при протекании гомогенных или гетерогенных химических реакций. Поскольку условия равновесия фаз являются изобарно-изотермическими, то химические реакции, протекающие в этих фазах, следует описывать с помощью свободной энергии Гиббса $G = \sum_i n_i \mu_i$.

Пусть в системе протекает обратимая химическая реакция, записанная по аналогии с (1.5.14) в обобщенной форме:



Здесь химические компоненты $\text{A}_{\text{н}}$ и $\text{A}_{\text{к}}$, стоящие со своими стехиометрическими коэффициентами $\nu_{\text{н}}$ и $\nu_{\text{к}}$ в левой и правой частях реакции, условно названы начальными (индекс «н») и конечными (индекс «к») участниками данной реакции, несмотря на ее обратимый характер.

Изобарный потенциал химической реакции определим как изменение ΔG свободной энергии системы, происходящее при *одном пробеге реакции слева направо*.

Если в исходном состоянии системы было $n_{\text{н}}$ молей начальных компонентов ($\text{A}_{\text{н}}$) и $n_{\text{к}}$ молей конечных компонентов ($\text{A}_{\text{к}}$), то свободная энергия Гиббса для системы в исходном состоянии:

$$G_0 = \sum_{\text{н}} n_{\text{н}} \mu_{\text{н}} + \sum_{\text{к}} n_{\text{к}} \mu_{\text{к}}.$$

Поскольку для реакции (1.14.4) за один пробег слева направо *исчезает* $\nu_{\text{н}}$ молей компонента $\text{A}_{\text{н}}$ и *возникает* $\nu_{\text{к}}$ молей компонента $\text{A}_{\text{к}}$, то свободная энергия Гиббса в результате такого пробега принимает значение

$$G = \sum_{\text{н}} (n_{\text{н}} - \nu_{\text{н}}) \mu_{\text{н}} + \sum_{\text{к}} (n_{\text{к}} + \nu_{\text{к}}) \mu_{\text{к}}.$$

Отсюда для обобщенной химической реакции (1.14.4) получаем **изобарный потенциал реакции**:

$$\Delta G = G - G_0 = \sum_{\text{к}} \nu_{\text{к}} \mu_{\text{к}} - \sum_{\text{н}} \nu_{\text{н}} \mu_{\text{н}}. \quad (1.14.5)$$

Общее требование изобарно-изотермического равновесия в форме $\Delta G = 0$, примененное к выражению (1.14.5), дает искомое **условие химического равновесия** в следующем виде:

$$\sum_n \nu_n \mu_n = \sum_k \nu_k \mu_k. \quad (1.14.6)$$

Из сравнения формул (1.14.4) и (1.14.6) видно, что для записи условия химического равновесия следует химические символы A_n и A_k участников реакции заменить на соответствующие им химические потенциалы μ_n и μ_k , а на месте стрелок поставить знак равенства.

В практике термодинамических расчетов вместо условия химического равновесия в форме (1.14.6) часто используют так называемый *закон действующих масс* (см. п. 2.12).

Следует отметить, что при рассмотрении кинетики химических реакций нередко вместо изобарного потенциала ΔG используют впервые введенную Де Донде¹⁾ величину с противоположным знаком, $\mathcal{A} = -\Delta G$, называемую **химическим сродством**.

Итак, химическое сродство соответствует изменению свободной энергии системы при *одном пробеге* реакции *справа налево*. По Де Донде принято записывать, используя определение (1.14.5), химическое сродство в форме, содержащей вместо двух сумм единую сумму по i (включая всех участников реакции — начальных с индексом $i = n$ и конечных с индексом $i = k$):

$$\mathcal{A} \equiv -\Delta G = - \sum_i \nu_i \mu_i, \quad \text{где} \quad \nu_i = \begin{cases} -\nu_n < 0, \\ \nu_k > 0. \end{cases} \quad (1.14.7)$$

1.15. Правило фаз Гиббса

Правило фаз устанавливает число *термодинамических степеней свободы*, определяемое как число независимых параметров системы, включая температуру и давление, которые можно произвольно изменять без нарушения фазового и химического равновесия в системе.

Пусть в системе из K компонентов присутствует Φ фаз и протекает X химических реакций типа (1.14.4). Равновесное состояние системы в целом характеризуется температурой T , давлением P и составом каждой из Φ фаз. Состав одной фазы, содержащей K компонентов, однозначно задается их концентрациями, выраженными в мольных долях $x_i = n_i/n$, среди которых лишь $K-1$ независимых величин, так как $\sum_i x_i = 1$, $i = 1, 2, \dots, K$.

¹⁾ Де Донде (De Donder) Теофил (1870–1957) — основатель брюссельской школы неравновесной термодинамики, одним из учеников которого был Нобелевский лауреат Илья Пригожин.

Следовательно, общее число параметров, характеризующих состояние системы, равно $2 + (K - 1)\Phi$. Эти параметры не являются полностью независимыми, поскольку на них наложены условия фазового (1.14.3) и химического (1.14.6) равновесия, куда входят химические потенциалы μ_i , в свою очередь зависящие от температуры, давления и состава фаз. Число фазовых ограничений равняется числу равенств в выражениях (1.14.3): в каждой из K строк имеем $\Phi - 1$ независимых равенств, так что общее их число равно $(\Phi - 1)K$. Число химических ограничений типа (1.14.6) обычно равняется числу X химических реакций ¹⁾.

Таким образом, число термодинамических степеней свободы C равно числу параметров системы, остающихся от их общего числа $2 + (K - 1)\Phi$ за вычетом $(\Phi - 1)K + X$ ограничений, накладываемых условиями фазового и химического равновесия, а именно

$$C = 2 + K - \Phi - X. \quad (1.15.1)$$

Обычно используют число независимых компонентов $K_n = K - X$, вычитая из общего их количества K число X химических ограничений. Тогда **правило фаз Гиббса** (1.15.1) принимает распространенную форму записи:

$$C = 2 + K_n - \Phi. \quad (1.15.2)$$

Число термодинамических степеней свободы возрастает с ростом числа независимых компонентов и уменьшается при увеличении фаз в системе. По числу степеней свободы принято различать *нонвариантные* ($C = 0$), *моновариантные* ($C = 1$), *бивариантные* ($C = 2$) и *поливариантные* ($C > 2$) системы.

Как видно из формулы (1.15.2), число равновесных фаз не может превышать числа независимых компонентов в системе больше, чем на два ($\Phi \leq 2 + K_n$). Максимальное число фаз реализуется в нонвариантной системе, когда $C = 0$. В этом случае все параметры системы строго фиксированы и ни один из них нельзя изменить без нарушения равновесия, что привело бы к исчезновению одной или более фаз.

¹⁾ Строго говоря, величина X может превышать число химических реакций на число так называемых *условий материального баланса*, выражающих сохранение химических элементов в соответствии со стехиометрией реакции и условием ввода исходных реагентов в систему. Так, для газофазной реакции $2\text{HCl} \rightleftharpoons \text{H}_2 + \text{Cl}_2$ ввод в систему хлористого водорода порождает в равных количествах H_2 и Cl_2 . Тогда условие материального баланса в форме $p_{\text{H}_2} = p_{\text{Cl}_2}$ совместно с условием химического равновесия типа (1.14.6) (записываемым обычно в форме закона действующих масс — см. п. 2.12) дает два химических ограничения ($X = 2$). В результате из трех участников реакции независимым остается только один ($K_n = K - X = 1$).

В формулах (1.15.1) и (1.15.2) слагаемое 2 учитывает в качестве управляемых параметров температуру T и давление P . Наиболее часто в технологических процессах полное давление поддерживается равным внешнему давлению (1 атм). В этом частном случае правило фаз (1.15.2) принимает вид

$$C = 1 + K_n - \Phi. \quad (1.15.3)$$

Здесь единица соответствует температуре T как единственному параметру, используемому в дополнение к составу фаз для управления фазовыми и химическими превращениями в системе.

1.16. Условия фазового и электрохимического равновесия в системах с заряженными частицами

До сих пор объектом нашего рассмотрения были макроскопические системы, состоящие из нейтральных частиц (атомов и молекул). Учет заряда частиц должен определенным образом модифицировать полученные ранее соотношения.

Сначала уточним условия фазового равновесия вида (1.14.1)–(1.14.3) в применении к фазам, обменивающимся заряженными частицами, а затем рассмотрим влияние заряда взаимодействующих частиц на условие химического равновесия (1.14.6).

Пусть частица i -го сорта имеет заряд $z_i e$, где $z_i \geq 0$ — зарядовое число, $e = 1,6 \cdot 10^{-19}$ Кл — заряд электрона. В заряженную среду с электрическим потенциалом φ вводим dN_i частиц, несущих элементарный заряд $dq_i = z_i e dN_i = z_i \mathcal{F} dn_i$, где $\mathcal{F} = eN_A = 96485$ Кл/моль — число Фарадея (заряд одного моля).

В результате этого внутренняя энергия системы должна увеличиться на дополнительную величину $\varphi dq_i = z_i \mathcal{F} \varphi dn_i$ по отношению к величине (1.3.5). Тогда суммирование по всем сортам частиц дает дифференциал внутренней энергии:

$$\begin{aligned} dU &= TdS - PdV + \sum_i \mu_i dn_i + \sum_i \varphi dq_i \equiv \\ &\equiv TdS - PdV + \sum_i \tilde{\mu}_i dn_i. \end{aligned} \quad (1.16.1)$$

Здесь вместо химического потенциала μ_i для заряженных частиц введена новая величина в форме

$$\tilde{\mu}_i = \mu_i + z_i \mathcal{F} \varphi, \quad (1.16.2)$$

называемая **электрохимическим потенциалом** частиц i -го сорта с зарядовым числом z_i , которые находятся в среде с электрическим потенциалом φ .

Следовательно, учет заряда частиц приводит к замене химических потенциалов μ_i в формулах (1.10.1)–(1.10.4) их электрохимическими аналогами $\tilde{\mu}_i$ в форме (1.16.2).

Если проследить вывод условий фазового равновесия, сделанный в п. 1.14, то можно убедиться в том, что аналогичной замены требуют лишь условия (1.14.3), а равенства (1.14.1) и (1.14.2) остаются без изменения.

Таким образом, для изобарно-изотермических процессов, протекающих при $T = \text{const}$ и $P = \text{const}$, химические потенциалы μ_i в условии фазового равновесия (1.14.3) должны быть заменены на электрохимические потенциалы $\tilde{\mu}_i$ по формуле (1.16.2):

$$\tilde{\mu}_i^I = \tilde{\mu}_i^{II} = \dots \equiv \tilde{\mu}_i = \text{const}, \quad i = 1, 2, \dots, K. \quad (1.16.3)$$

Итак, фазовое равновесие достигается при постоянстве во всех фазах электрохимического потенциала каждого компонента.

Полезно отметить, что для электронов ($z_e = -1$) в металлах и полупроводниках их электрохимический потенциал $\tilde{\mu}_e$ называется *энергией Ферми* E_F (см. п. 3.5). Равновесное требование постоянства уровня Ферми ($E_F^I = E_F^{II}$) в контактирующих средах I и II, записанное с учетом формулы (1.16.2) в виде $\mu_e^I - \mathcal{F}\varphi^I = \mu_e^{II} - \mathcal{F}\varphi^{II}$, выполняется лишь при наличии на границе раздела сред контактной разности потенциалов:

$$\varphi_{\text{конт}} \equiv \varphi^I - \varphi^{II} = \frac{1}{\mathcal{F}} (\mu_e^I - \mu_e^{II}), \quad (1.16.4)$$

где химический потенциал μ_e определяет положение уровня Ферми в областях с $\varphi = 0$.

Итак, при равновесии сред, обменивающихся заряженными частицами (электронами или ионами), химические потенциалы этих частиц в контактирующих средах не равны между собой, а различие между ними определяет контактную разность электрических потенциалов.

Нетрудно убедиться в том, что для обобщенной химической реакции типа (1.14.4) участие в ней заряженных частиц сохраняет условие химического равновесия в форме (1.14.6), но с заменой химических потенциалов μ_n и μ_k на соответствующие электрохимические потенциалы $\tilde{\mu}_n$ и $\tilde{\mu}_k$. Используя (1.16.2), запишем модифицированную форму (1.14.6) в виде

$$\sum_K \nu_k \mu_k - \sum_N \nu_n \mu_n = \sum_N \nu_n z_n \mathcal{F} \varphi_n - \sum_K \nu_k z_k \mathcal{F} \varphi_k, \quad (1.16.5)$$

где φ_n и φ_k — электрические потенциалы фаз, в которых находятся начальные (н) и конечные (к) участники химической

реакции (1.14.4). В данном случае условие химического (точнее, электрохимического) равновесия (1.16.5) необходимо дополнить требованием электронейтральности:

$$\sum_{\text{н}} \nu_{\text{н}} z_{\text{н}} = \sum_{\text{к}} \nu_{\text{к}} z_{\text{к}}. \quad (1.16.6)$$

Из равенства (1.16.6) следует, что для химической реакции, все участники которой (как без заряда, так и с зарядом) находятся в одной и той же фазе с однородным распределением электрического потенциала ($\varphi_{\text{н}} = \varphi_{\text{к}} \equiv \varphi = \text{const}$), правая часть выражения (1.16.5) обращается в нуль, и оно принимает привычную форму (1.14.6).

Следовательно, любая *гомогенная реакция* с участием как нейтральных, так и заряженных частиц управляется полученным ранее условием химического равновесия (1.14.6), включающим обычные химические (а не электрохимические) потенциалы. Это свойство гомогенных реакций будет использовано в п. 2.17 при анализе ионного равновесия в растворах жидких электролитов.

Учет электрохимических потенциалов необходим только для *гетерогенных реакций*, в частности при рассмотрении электродных процессов в электрохимических системах (см. п. 2.19).

В этом случае изобарный потенциал реакции ΔG , определенный в форме (1.14.5), вместо химических потенциалов $\mu_{\text{н}}$ и $\mu_{\text{к}}$ содержит электрохимические потенциалы $\tilde{\mu}_{\text{н}}$ и $\tilde{\mu}_{\text{к}}$. Используя для них выражение (1.16.2), представляем изобарный потенциал в виде суммы двух вкладов:

$$\Delta G = \Delta G_{\text{хим}} + \Delta G_{\text{эл}}. \quad (1.16.7)$$

Здесь химический вклад в форме (1.14.5) определяется химическими потенциалами частиц,

$$\Delta G_{\text{хим}} = \sum_{\text{к}} \nu_{\text{к}} \mu_{\text{к}} - \sum_{\text{н}} \nu_{\text{н}} \mu_{\text{н}}, \quad (1.16.8)$$

а электрический вклад равен

$$\Delta G_{\text{эл}} = \sum_{\text{к}} \nu_{\text{к}} z_{\text{к}} \mathcal{F} \varphi_{\text{к}} - \sum_{\text{н}} \nu_{\text{н}} z_{\text{н}} \mathcal{F} \varphi_{\text{н}}. \quad (1.16.9)$$

Следовательно, **электрохимическое равновесие**, имеющее место при $\Delta G = 0$, осуществляется в результате взаимной компенсации химического и электрического вкладов:

$$\Delta G_{\text{хим}} = -\Delta G_{\text{эл}}, \quad (1.16.10)$$

что при использовании выражений (1.16.8) и (1.16.9) дает ранее полученное равенство (1.16.5).

1.17. Флуктуационная устойчивость равновесного состояния

Как следует из рассмотрения, выполненного в п. 1.14, фазовое и химическое равновесия в любой системе устанавливаются:

а) при однородном распределении температуры и давления ($T = \text{const}$ и $P = \text{const}$) (см. уравнения (1.14.1) и (1.14.2));

б) при однородном распределении химических потенциалов всех компонентов ($\mu_i = \text{const}$, $i = 1, 2, \dots, K$) с вполне определенными соотношениями между ними как следствие фазового и химического равновесия (см. уравнения (1.14.3) и (1.14.6)).

Флуктуации вышеуказанных интенсивных величин (δT , δP и $\delta \mu_i$) заставляют флуктуировать также и экстенсивные величины, например, энтропию и объем. Поскольку, согласно функциональной зависимости (1.11.7), химический потенциал μ_i является функцией T , P и n_i , то его флуктуации $\delta \mu_i$ порождаются не только δT и δP , но и флуктуациями числа молей δn_i .

На этом основании записываем вариацию $\delta \mu_i$ в следующей форме:

$$\begin{aligned} \delta \mu_i &= \left(\frac{\partial \mu_i}{\partial T} \right)_{P, n_i} \delta T + \left(\frac{\partial \mu_i}{\partial P} \right)_{T, n_i} \delta P + \sum_j \left(\frac{\partial \mu_i}{\partial n_j} \right)_{T, P, n_i \neq n_j} \delta n_j = \\ &= -\bar{S}_i \delta T + \bar{V}_i \delta P + \sum_j \mu_{ij} \delta n_j, \end{aligned} \quad (1.17.1)$$

где $\bar{S}_i$ и $\bar{V}_i$ — парциальные мольные величины, определенные равенствами (1.11.11) и (1.11.12). В выражении (1.17.1) использованы формулы (1.11.16)–(1.11.18) и введено следующее обозначение:

$$\mu_{ij} \equiv \left(\frac{\partial \mu_i}{\partial n_j} \right)_{T, P, n_i \neq n_j} = \left(\frac{\partial \mu_j}{\partial n_i} \right)_{T, P, n_j \neq n_i} \equiv \mu_{ji}. \quad (1.17.2)$$

Для экстенсивных величин S и V , рассматриваемых также как функции T , P и n_i , флуктуации принимают вид

$$\delta S = (\delta S)_{n_i} + \sum_i \bar{S}_i \delta n_i, \quad (1.17.3)$$

$$\delta V = (\delta V)_{n_i} + \sum_i \bar{V}_i \delta n_i, \quad (1.17.4)$$

где

$$(\delta S)_{n_i} = \left(\frac{\partial S}{\partial T} \right)_{P, n_i} \delta T + \left(\frac{\partial S}{\partial P} \right)_{T, n_i} \delta P, \quad (1.17.5)$$

$$(\delta V)_{n_i} = \left(\frac{\partial V}{\partial T} \right)_{P, n_i} \delta T + \left(\frac{\partial V}{\partial P} \right)_{T, n_i} \delta P. \quad (1.17.6)$$

В отличие от (1.17.3) и (1.17.4), для давления как интенсивной величины, зависящей от T и V , записываем флуктуацию в виде

$$\delta P = \left(\frac{\partial P}{\partial T} \right)_{V, n_i} \delta T + \left(\frac{\partial P}{\partial V} \right)_{T, n_i} (\delta V)_{n_i}. \quad (1.17.7)$$

В п. 1.9 было показано, что при равновесии (термодинамическом или текущем) любой системы (замкнутой, закрытой, открытой) ее флуктуационная устойчивость в линейной области обеспечивается отрицательной кривизной поверхности энтропии ($\delta^2 S \equiv 2S_{\text{ex}} < 0$) как функции параметров, флуктуирующих около равновесного значения.

Возникает вопрос: какие физические свойства системы определяют эту кривизну, и что априори гарантирует ее отрицательный знак в равновесных условиях?

Для ответа на этот вопрос надо вычислить вторую вариацию энтропии $S(U, V, n_1, n_2, \dots, n_K)$ как функции внутренней энергии U , объема V и числа молей n_i , вызванную их флуктуациями δU , δV и δn_i . Воспользуемся выражениями (Д.1) и (Д.2) для первого и второго дифференциалов функции n переменных, $f(x_1, \dots, x_n)$, полученных в математическом дополнении к данному параграфу (см. с. 147).

Для записи первой вариации энтропии используем уравнение Гиббса (1.3.5), разрешая его относительно дифференциала dS и делая замену знака дифференциала на знак вариации ($d \rightarrow \delta$):

$$\delta S = \frac{1}{T} \delta U + \frac{P}{T} \delta V - \sum_i \frac{\mu_i}{T} \delta n_i. \quad (1.17.8)$$

Сравнивая (1.17.8) с формулой (Д.1), записываем вторую вариацию энтропии, по аналогии с (Д.2), в следующем виде:

$$\delta^2 S = \delta \left(\frac{1}{T} \right) \delta U + \delta \left(\frac{P}{T} \right) \delta V - \sum_i \delta \left(\frac{\mu_i}{T} \right) \delta n_i. \quad (1.17.9)$$

Вычисляя вариации $\delta(1/T)$, $\delta(P/T)$ и $\delta(\mu_i/T)$, приводим выражение (1.17.9) к виду

$$\begin{aligned} \delta^2 S = & -\frac{\delta T}{T^2} \left(\delta U + P \delta V - \sum_i \mu_i \delta n_i \right) + \\ & + \frac{1}{T} \left(\delta P \delta V - \sum_i \delta \mu_i \delta n_i \right). \end{aligned} \quad (1.17.10)$$

Согласно (1.17.8), первая скобка в выражении (1.17.10) равна $T\delta S$, тогда оно принимает искомую форму:

$$\delta^2 S = -\frac{1}{T} \left(\delta T \delta S - \delta P \delta V + \sum_i \delta \mu_i \delta n_i \right). \quad (1.17.11)$$

Эта формула является центральной формулой для флуктуационной теории термодинамических систем. Ниже она будет обобщена на неравновесные стационарные ситуации путем введения локального квазиравновесия (см. пп. 1.21 и 1.24).

Как следует из формулы (1.17.11), вторая вариация энтропии (или иначе *избыточная энтропия* $S_{\text{ex}} \equiv \delta^2 S/2$) является величиной второго порядка малости относительно флуктуаций термодинамических переменных.

Для того, чтобы выяснить знак $\delta^2 S$, необходимо преобразовать слагаемые, входящие в (1.17.11).

Подставляя вариацию химического потенциала (1.17.1) в выражение (1.17.11), с учетом (1.17.3)–(1.17.6) получаем

$$T\delta^2 S = -(\delta S)_{n_i} \delta T + (\delta V)_{n_i} \delta P - \sum_i \sum_j \mu_{ij} \delta n_i \delta n_j. \quad (1.17.12)$$

Первое слагаемое в правой части (1.17.12) преобразуется с применением формул (1.5.5), (1.10.10) и (1.17.5):

$$\begin{aligned} (\delta S)_{n_i} \delta T &= \left[\frac{C_P}{T} \delta T - \left(\frac{\partial V}{\partial T} \right)_{P, n_i} \delta P \right] \delta T = \\ &= \frac{C_V}{T} (\delta T)^2 - \left(\frac{\partial P}{\partial T} \right)_{V, n_i} (\delta V)_{n_i} \delta T. \end{aligned} \quad (1.17.13)$$

Здесь в последнем равенстве использованы выражения (1.17.7) для δP и (1.5.7) для разности теплоемкостей ($C_P - C_V$).

Использование выражения (1.17.7) для δP во втором слагаемом в правой части формулы (1.17.12) приводит его к следующему виду:

$$(\delta V)_{n_i} \delta P = -\frac{1}{\chi V} (\delta V)_{n_i}^2 + \left(\frac{\partial P}{\partial T} \right)_{V, n_i} (\delta V)_{n_i} \delta T, \quad (1.17.14)$$

где введена изотермическая сжимаемость, определенная как

$$\chi = -\frac{1}{V} \left(\frac{\partial V}{\partial P} \right)_{T, n_i}. \quad (1.17.15)$$

Подстановка равенств (1.17.13) и (1.17.14) в формулу (1.17.12) дает искомое выражение в окончательной форме:

$$\delta^2 S = -\frac{1}{T} \left[\frac{C_V}{T} (\delta T)^2 + \frac{(\delta V)_{n_i}^2}{\chi V} + \sum_i \sum_j \mu_{ij} \delta n_i \delta n_j \right]. \quad (1.17.16)$$

Отсюда непосредственно следует вывод о флуктуационной устойчивости равновесного состояния, реализуемой при $\delta^2 S < 0$. Для этого каждое из трех слагаемых в квадратной скобке должно быть положительным.

Именно это требование реализуется на практике, обеспечивая три типа устойчивости системы по отношению к флуктуациям температуры (δT), объема (δV) и числа частиц (δn_i):

$$C_V > 0 \text{ — тепловая устойчивость,} \quad (1.17.17)$$

$$\chi > 0 \text{ — механическая устойчивость,} \quad (1.17.18)$$

$$\sum_i \sum_j \mu_{ij} \delta n_i \delta n_j > 0 \text{ — материальная устойчивость.} \quad (1.17.19)$$

Три типа устойчивости, выражаемые неравенствами (1.17.17), (1.17.18) и (1.17.19), соответствуют, по крайней мере, названием трем типам контакта системы с внешним окружением, которые были введены ранее в п. 1.3. Устойчивость, выражаемую условием (1.17.19), называют также *диффузионной устойчивостью* (по отношению к процессу диффузии) или *химической устойчивостью* (по отношению к химическим реакциям).

Положительные значения теплоемкости C_V и изотермической сжимаемости χ , выражаемые неравенствами (1.17.17) и (1.17.18), вполне очевидны и всегда реализуются. Однако в случае материального неравенства (1.17.19) ситуация менее очевидная: положительное значение квадратичной формы может возникать лишь при определенных физических условиях. Это доказывается с помощью нулевых соотношений (1.11.19), наложенных на производные μ_{ij} химических потенциалов. С использованием обозначений (1.17.2) эти соотношения можно переписать в виде

$$\sum_{i=1}^K n_i \mu_{ij} = 0 \quad \text{при } j = 1, 2, \dots, K. \quad (1.17.20)$$

Известно, что квадратичная форма типа (1.17.19) является положительно определенной в том и только в том случае, когда положительны все главные миноры определителя, составленного из симметричных коэффициентов ($\mu_{ij} = \mu_{ji}$), определенных по формуле (1.17.2).

Следовательно, для любой пары компонентов должны выполняться, по крайней мере, следующие неравенства:

$$\mu_{ii} > 0, \quad \mu_{jj} > 0, \quad \mu_{ii}\mu_{jj} > \mu_{ij}^2 \quad \text{и т. д.} \quad (1.17.21)$$

В частности, для бинарной системы соотношения (1.17.20) дают

$$\mu_{11} > 0, \quad \mu_{22} > 0, \quad \mu_{12} < 0, \quad (1.17.22)$$

что удовлетворяет условиям (1.17.21) материальной устойчивости. Когда они выполняются, система восстанавливает свой исходный химический состав и пространственную однородность. В тех случаях, когда знак неравенства в формуле (1.17.19) становится отрицательным (т. е. нарушаются неравенства (1.17.22) для бинарной системы), возникают неустойчивые и метастабильные состояния, приводящие, например, к расслоению фаз.

МАТЕМАТИЧЕСКОЕ ДОПОЛНЕНИЕ. Первый и второй дифференциалы функции нескольких переменных.

Запишем первый и второй дифференциалы функции n переменных $f(x_1, \dots, x_k, \dots, x_n) \equiv f(x_k)$:

$$df = \sum_{k=1}^n \frac{\partial f}{\partial x_k} dx_k \equiv f_k^{(1)} dx_k,$$

$$d^2f \equiv d(df) = \sum_{k=1}^n \sum_{l=1}^n \frac{\partial^2 f}{\partial x_k \partial x_l} dx_k dx_l \equiv df_k^{(1)} dx_k.$$

Здесь после знака тождественного равенства подразумевается суммирование по повторяющимся индексам, с учетом введенных ниже обозначений:

$$f_k^{(1)}(x_1, \dots, x_n) = \frac{\partial f(x_1, \dots, x_n)}{\partial x_k},$$

$$df_k^{(1)}(x_1, \dots, x_n) = \frac{\partial f_k^{(1)}(x_1, \dots, x_n)}{\partial x_l} dx_l.$$

Итак, окончательно получаем искомые результаты дифференцирования:

$$df = f_1^{(1)} dx_1 + \dots + f_k^{(1)} dx_k + \dots + f_n^{(1)} dx_n, \quad (Д.1)$$

$$d^2f = df_1^{(1)} dx_1 + \dots + df_k^{(1)} dx_k + \dots + df_n^{(1)} dx_n. \quad (Д.2)$$

Таким образом, в сравнении с первым дифференциалом (Д.1) во втором дифференциале (Д.2) вместо первых производных $f_k^{(1)}$ исходной функции $f(x_k)$ стоят их дифференциалы $df_k^{(1)}$.

1.18. Нарушение условий фазового и химического равновесия

1. Нарушение фазового равновесия. Как известно (см. п. 1.14 и 1.16), любая гетерогенная многокомпонентная система с химически пассивными компонентами находится в условиях фазового равновесия, если все фазы имеют одинаковые температуру и давление ($T = \text{const}$ и $P = \text{const}$), а для каждого i -го компонента его химический потенциал (или электрохимический потенциал для заряженных частиц) одинаков во всех равновесных фазах (μ_i ($\bar{\mu}_i$) = const, $i = 1, 2, \dots, K$ — номер компонента).

Следовательно, нарушение термодинамического равновесия между фазами возможно лишь в условиях неоднородного распределения температуры $T(\mathbf{r})$, давления $P(\mathbf{r})$ и химических потенциалов $\mu_i(\mathbf{r})$, характеризующихся соответствующими градиентами, а именно $\text{grad } T(\mathbf{r}) \neq 0$, $\text{grad } P(\mathbf{r}) \neq 0$ и $\text{grad } \mu_i(\mathbf{r}) \neq 0$. В нестационарных условиях все эти величины зависят также от времени t , не указываемого здесь для упрощения записи.

Ненулевые градиенты являются *движущей силой* процессов массопереноса (или массоотдачи) и теплопереноса (или теплоотдачи). Проявляются эти неравновесные процессы в виде потоков частиц (или массы) и энергии (или тепла). Направление этих потоков таково, что при выключении внешних воздействий они стремятся вернуть систему в равновесное состояние.

На основании общих соображений запишем **плотность потока частиц** i -го компонента в форме векторной величины,

$$\mathbf{J}_i = \alpha_i^T \text{grad } T + \alpha_i^P \text{grad } P + \sum_{j=1}^K \alpha_{ij}^\mu \text{grad } \mu_j, \quad (1.18.1)$$

где последняя сумма учитывает взаимное влияние компонентов.

Формула (1.18.1) содержит в общем случае аддитивные вклады от каждого из трех градиентов с соответствующими коэффициентами пропорциональности. Такой поток, учитывающий линейные вклады градиентов, называют *диффузионным потоком*, а физические процессы, обеспечивающие эти три вклада, носят название *термодиффузии* ($\alpha_i^T \text{grad } T$), *бародиффузии* ($\alpha_i^P \text{grad } P$) и *концентрационной диффузии* ($\sum_j \alpha_{ij}^\mu \text{grad } \mu_j$). Последнее название связано с тем обстоятельством, что химический потенциал i -го компонента определяется его концентрацией c_i в соответствии с равенством (1.13.4):

$$\mu_i = \mu_i^\circ + k_B T \ln a_i \equiv \mu_i^\circ + k_B T \ln \gamma_i c_i. \quad (1.18.2)$$

Здесь химический потенциал отнесен к одной частице, при этом $a_i = \gamma_i c_i$ — активность i -го компонента и γ_i — коэффициент активности, зависящий в общем случае от концентраций c_j всех компонентов системы, включая i -й ($j = 1, 2, \dots, i, \dots, K$).

На практике вклады термодиффузии и бародиффузии обычно играют малозначительную роль и могут быть опущены в формуле (1.18.1) по сравнению с концентрационной диффузией, которую в этом случае называют просто *диффузией*.

Известно (см. п. 1.11), что при неоднородном распределении химического потенциала $\mu_i(\mathbf{r})$ происходит переход частиц i -го сорта из области с большим значением μ_i в область с меньшим его значением. В то же время градиент химического потенциала всегда направлен в сторону возрастания потенциала, что дает $\alpha_i^\mu = -|\alpha_i^\mu| < 0$. Это позволяет, используя выражение (1.18.2) и равенство $\text{grad } \mu_i = (\partial \mu_i / \partial c_i) \text{grad } c_i$, записать выражение (1.18.1) для *плотности диффузионного потока* i -го компонента (в пренебрежении влиянием других компонентов) в виде

$$\mathbf{J}_i = -D_i \text{grad } c_i \quad \text{при } D_i > 0. \quad (1.18.3)$$

Здесь $D_i = |\alpha_i^\mu| (k_B T / c_i) (1 + \partial \ln \gamma_i / \partial \ln c_i)$ означает коэффициент диффузии i -го компонента, имеющий размерность $\text{м}^2/\text{с}$. В общем случае коэффициент диффузии является функцией температуры и концентраций всех компонентов, поскольку для многокомпонентных систем $\gamma_i = f_i(T, c_1, c_2, \dots, c_K)$. Однако при небольших пространственных неоднородностях в сильно разбавленных растворах пренебрегают зависимостью от концентраций и учитывают лишь температурную зависимость $D_i(T)$.

Известно, что для газов такая зависимость обычно имеет вид $D(T) = D_0 (T/273)^m$, где $m = 1,5 - 2,0$ и по порядку величины $D_0 = (0,1 - 1,5) \text{ см}^2/\text{с}$. Для жидкостей и твердых тел температурная зависимость $D(T)$ определяется, как увидим ниже, законом Аррениуса (1.18.16) и по порядку величины дает значения: $D = (10^{-6} - 10^{-4}) \text{ см}^2/\text{с}$ для жидкостей и $D = (10^{-14} - 10^{-10}) \text{ см}^2/\text{с}$ для твердых тел.

Согласно выражению (1.18.3), диффузионный поток вещества всегда вытекает из области с большей концентрацией соответствующих частиц и стремится выровнять их пространственное распределение.

По аналогии с формулой (1.18.1), можно записать *плотность теплового потока*, переносимого всеми частицами,

$$\mathbf{J}_q = \beta^T \text{grad } T + \beta^P \text{grad } P + \sum_{j=1}^K \beta_j^\mu \text{grad } \mu_j. \quad (1.18.4)$$

Здесь, как и для потока частиц (1.18.1), обычно доминирует один из трех вкладов, а именно температурный. Тогда

$$\mathbf{J}_q = -\lambda \text{grad } T \quad \text{при } \lambda > 0, \quad (1.18.5)$$

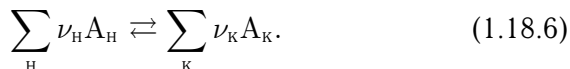
где $\lambda = -\beta^T$ — коэффициент теплопроводности, а минус указывает, что поток тепла направлен от нагретой области к холодной.

Уравнения (1.18.3) и (1.18.5) часто называют, соответственно, **первым законом Фика**¹⁾ и **законом Фурье**²⁾.

Следует подчеркнуть, что выражения (1.18.1) и (1.18.4) для потоков были записаны на основе интуитивных представлений о переносе вещества и энергии в неоднородных средах. При этом в качестве движущих сил были взяты градиенты температуры, давления и химического потенциала.

Более строгие выражения будут получены ниже на основе уравнений баланса числа частиц, энергии и энтропии. При этом будут уточнены выражения для термодинамических сил и установлена связь между ними при помощи коэффициентов Онсагера (см. п. 1.23).

2. Нарушение химического равновесия. В гетерогенной системе обратимая химическая реакция записывается в обобщенной форме:



Здесь $\text{A}_{\text{н}}$ и $\text{A}_{\text{к}}$ — химические символы реагентов, условно названных начальными (стоящие слева с индексом «н») и конечными (стоящие справа с индексом «к»), а $\nu_{\text{н}}$ и $\nu_{\text{к}}$ — соответствующие им стехиометрические коэффициенты, уравнивающие в реакции число атомов каждого элемента (которые могут быть не только целыми, но и дробными числами).

Как известно (см. п. 1.14), в этом случае термодинамическое равновесие обеспечивается, в дополнение к изобарно-изотерми-

¹⁾ Фик (Fick) Адольф (1829–1901) — немецкий физиолог, объяснил законы физиологии на основе математики и механики; выполняя исследования по термодинамике мышц, доказал действенность закона сохранения энергии при мышечном сокращении и впервые (1855) сформулировал законы диффузии, носящие теперь его имя.

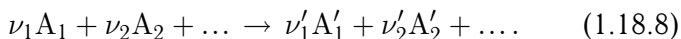
²⁾ Фурье (Fourier) Жан Батист Жозеф (1768–1830) — французский математик и физик, заложил основы теории множеств и теории функций действительного переменного, использовал разработанные математические методы (метод разделения переменных, ряды и интегралы Фурье) в теории распространения тепла, изложенной в классическом трактате «Аналитическая теория тепла», впервые (1822) доказал закон теплопроводности (закон Фурье) и открыл термоэлектрический эффект.

ческим условиям ($P = \text{const}$ и $T = \text{const}$), требованием обращения в нуль изобарного потенциала реакции ($\Delta G = 0$ — см. формулу (1.14.5)). Отсюда следует равенство, накладываемое на химические потенциалы участников реакции (см. уравнение (1.14.6)):

$$\sum_{\text{н}} \nu_{\text{н}} \mu_{\text{н}} = \sum_{\text{к}} \nu_{\text{к}} \mu_{\text{к}}. \quad (1.18.7)$$

При нарушении требования (1.18.7) обратимая химическая реакция (1.18.6) смещается в ту или иную сторону в зависимости от знака изобарного потенциала ΔG , вычисляемого на основании уравнения изотермы Вант-Гоффа (см. ниже формулу (2.13.5)): при $\Delta G < 0$ возможно протекание реакции слева направо, а при $\Delta G > 0$ — справа налево.

Рассмотрим одностороннюю реакцию, протекающую слева направо, которую, в отличие от формы (1.18.6), более удобно для последующего изложения записать в таком виде:



Процесс превращения начальных реагентов A_i в конечные продукты A'_i протекает нестационарно, развиваясь во времени. Кинетическими величинами, характеризующими химический процесс, в данном случае являются мгновенные значения мольно-объемных концентраций для участников реакции.

Скорость изменения этих величин во времени определяет **скорость химической реакции**. Ее можно оценивать либо по уменьшению концентрации $c_{A_i}(t)$ любого из начальных реагентов (нестрихованных), либо по увеличению концентрации $c_{A'_i}(t)$ любого из продуктов реакции (стрихованных), поскольку эти концентрации жестко связаны между собой соотношением стехиометричности:

$$\frac{1}{-\nu_1} \frac{dc_{A_1}(t)}{dt} = \frac{1}{-\nu_2} \frac{dc_{A_2}(t)}{dt} = \dots = \frac{1}{\nu'_1} \frac{dc_{A'_1}(t)}{dt} = \frac{1}{\nu'_2} \frac{dc_{A'_2}(t)}{dt} = \dots \quad (1.18.9)$$

Поскольку здесь стехиометрические коэффициенты для начальных компонентов берутся с минусом ($-\nu_{\text{н}} \equiv -\nu_i < 0$), а для конечных — с плюсом ($\nu_{\text{к}} \equiv \nu'_i > 0$), то удобно воспользоваться нумерацией Де Донде ($\nu_i = \mp \nu_{\text{н(к)}} \leq 0$), связанной с введением химического сродства реакции $\mathcal{A}$ (см. уравнение (1.14.7)).

Следуя Де Донде, возьмем для прямой реакции с номером r (протекающей слева направо) в качестве ее скорости v_r общую для всех участников этой реакции положительную величину,

стоящую в равенствах (1.18.9), и обозначим ее как $d\xi_r/dt > 0$:

$$v_r(t) = \frac{d\xi_r(t)}{dt} \equiv \frac{1}{\nu_{ir}} \frac{dc_i(t)}{dt} > 0 \quad \left(\frac{\text{моль}}{\text{м}^3 \cdot \text{с}} \right). \quad (1.18.10)$$

Отсюда скорость изменения концентрации i -го компонента (реагента при $i = \text{н}$ и $\nu_{ir} = -\nu_{\text{н}} < 0$ или продукта реакции при $i = \text{к}$ и $\nu_{ir} = \nu_{\text{к}} > 0$) в результате протекания r -й реакции равна

$$v_i(t) \equiv \frac{dc_i(t)}{dt} = \nu_{ir} \frac{d\xi_r(t)}{dt} = \nu_{ir} v_r(t). \quad (1.18.11)$$

Новая переменная ξ_r , впервые введенная Де Донде по формуле (1.18.10), имеет разные названия — *химическая переменная* (*химическая координата*), *степень полноты реакции* или **число пробегов реакции**. В нашем случае последнее название кажется более информативным, так как изобарный потенциал химической реакции, $\Delta G = -A$, был определен в п. 1.14 как изменение свободной энергии Гиббса *при одном пробеге* реакции ($\xi_r = 1$) слева направо.

Физический смысл ξ_r , как степени полноты r -й реакции, ясен из результата интегрирования уравнения (1.18.10):

$$c_i(t) = c_i(0) + \nu_{ir} \xi_r(t). \quad (1.18.12)$$

Здесь второе слагаемое показывает, насколько увеличилась (при $\nu_{ir} = \nu_{\text{к}} > 0$ для продукта реакции) или уменьшилась (при $\nu_{ir} = -\nu_{\text{н}} < 0$ для исходного реагента) начальная концентрация $c_i(0)$ за время t протекания прямой реакции с номером r .

Как будет ясно из дальнейшего (см. п. 1.23), скорость химической реакции, определенная как $v_r = d\xi_r/dt$, является величиной, сопряженной с химическим сродством A_r для реакции r ($r = 1, 2, \dots, X$, где X — общее число реакций в системе) таким образом, что их произведение при делении на T дает вклад в производство энтропии $\sigma_s^{\text{хим}}$ от этой реакции (см. (1.23.16)).

Согласно **основному постулату химической кинетики**, скорость прямой реакции с номером r (протекающей слева направо) пропорциональна произведению концентраций исходных реагентов, взятых в степенях, равных их стехиометрическим коэффициентам в реакции (1.18.8):

$$v_r(t) = k_r(T) c_{A_1}^{\nu_1}(t) c_{A_2}^{\nu_2}(t) \dots \equiv k_r(T) \prod_{\text{н}} c_{\text{н}}^{\nu_{\text{н}}}(t). \quad (1.18.13)$$

Коэффициент пропорциональности $k_r(T)$ для каждой реакции r постоянен при неизменной температуре и по этой причине называется **константой скорости** химической реакции. Эта

константа, как увидим ниже, является, в соответствии с законом Аррениуса, функцией только температуры. Она характеризует кинетику химического процесса, под которой понимают закономерность развития процесса во времени. Согласно определению, константа скорости равняется реальной скорости реакции при единичной концентрации каждого из исходных веществ.

Подстановка формулы (1.18.10) в левую часть основного постулата химической кинетики (1.18.13) дает уравнение, называемое *кинетическим уравнением*.

Следует отметить, что присутствие стехиометрических коэффициентов ν_i в качестве показателей степени в формуле (1.18.13) оправдано лишь в двух случаях: для *простых реакций*, протекающих в одну стадию, и для *квазиравновесных реакций*, протекающих в условиях, близких к равновесным.

В общем случае сложная реакция обычно протекает по разветвленному пути, выискивая промежуточные стадии с наибольшей скоростью протекания для того, чтобы обеспечить высокую вероятность достижения конечного состояния. В этом случае уравнение (1.18.13) содержит на месте стехиометрического коэффициента ν_i другой численный коэффициент n_i , называемый *порядком реакции по данному компоненту*.

Каждый из коэффициентов n_i подлежит специальному экспериментальному нахождению. Суммарная величина $n = \sum_i n_i$ определяет *порядок реакции в целом*. По этому признаку различают химические реакции первого порядка (типа $A \rightarrow B + C + \dots$), второго порядка (типа $A + B \rightarrow C + D + \dots$) и т. д., для которых кинетическое уравнение, соответственно, имеет вид (нижний индекс r опущен за ненадобностью):

$$\begin{aligned} v(t) &\equiv -\frac{dc_A(t)}{dt} = k(T) c_A(t), \\ v(t) &\equiv -\frac{dc_A(t)}{dt} = k(T) c_A(t) c_B(t) \quad \text{и т. д.} \end{aligned} \tag{1.18.14}$$

В выражениях (1.18.14) порядок реакции отождествлен с ее *молекулярностью*, определяемой как число $\nu = \sum_i \nu_i$ молекул, одновременно участвующих в акте химического взаимодействия. Отождествление порядка реакции с молекулярностью ($n = \nu$) часто применяют на практике, хотя из вышесказанного ясно, что это справедливо лишь для простых и квазиравновесных реакций.

3. Закон Аррениуса. Рассмотренные неравновесные процессы (диффузионный и кинетический) на элементарном уровне обусловлены переходом частиц из одного состояния в другое

с преодолением потенциального барьера высотой $\mathcal{E}$, изображенного на рис. 1.10. Вероятность w перехода частицы из состояния 1 в состояние 2 зависит только от температуры T и определяется обычной Больцмановской экспонентой:

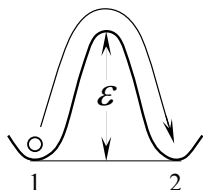


Рис. 1.10. К определению энергии активации термоактивационного процесса

$$w = w_0 e^{-\mathcal{E}/k_B T}, \quad (1.18.15)$$

где w_0 и $\mathcal{E}$ — постоянные величины, характерные для рассматриваемого неравновесного процесса. Высоту потенциального барьера $\mathcal{E}$, который энергетически разделяет соседние положения частицы, принято называть **энергией активации** данного процесса.

Исходя из общих энергетических соображений, можно утверждать, что вероятность осуществления процесса, записанная в форме (1.18.15), характерна как для диффузионного перемещения частиц, так и для химической кинетики. В связи с этим следует ожидать, что величины D и k , характеризующие эти процессы, должны иметь экспоненциальную зависимость от температуры, аналогичную (1.18.15), а именно

$$D(T) = D_0 e^{-\mathcal{E}/k_B T}, \quad (1.18.16)$$

$$k(T) = k_0 e^{-\mathcal{E}/k_B T}. \quad (1.18.17)$$

Отсюда видно, что чем больше энергия активации $\mathcal{E}$ для диффузионного или химического процесса, тем меньше коэффициент диффузии $D(T)$ или константа скорости $k(T)$, и тем сильнее они изменяются с изменением температуры. Более строгое подтверждение соотношений (1.18.16) и (1.18.17) дано в пп. 4.1 и 4.2).

Соотношения (1.18.16) и (1.18.17) определяют так называемый **закон Аррениуса**¹⁾, а процессы, подчиняющиеся этому закону, называют **термоактивационными процессами**. Для таких процессов энергия активации, введенная как высота потенциального барьера, всегда положительна. Это означает, что с ростом температуры коэффициенты $D(T)$ и $k(T)$ всегда возрастают.

¹⁾ Аррениус (Arrhenius) Сванте Август (1859–1927) — шведский физико-химик, один из основателей физической химии, занимался изучением электролитов и исследованием скорости химических реакций, сформулировал теорию электролитической диссоциации, впервые ввел понятие энергии активации и установил характер температурной зависимости константы скорости реакции (уравнение Аррениуса). Нобелевская премия по химии (1903) «как факт признания особого значения его теории электролитической диссоциации для развития химии».

Именно такой смысл вкладывают в термин «термоактивационный процесс», поскольку повышение температуры увеличивает вероятность преодоления частицами энергетического барьера и тем самым активизирует протекание неравновесного процесса.

1.19. Формализм теории Онсагера

Теория Онсагера обосновывает структуру соотношения (1.7.6) для производства энтропии $\mathcal{P}$ и доказывает в рамках линейного приближения для неравновесного процесса α связь между термодинамическими потоками J_α и силами X_α с помощью симметричных коэффициентов (соотношения взаимности Онсагера), построенную на основе принципа микроскопической обратимости (называемого также принципом детального равновесия).

1. Связь между термодинамическими силами и потоками.

Ранее было показано (см. пп. 1.8 и 1.9, а также рис. 1.5 и 1.7), что все системы (замкнутые, закрытые, открытые) эквивалентны в отношении экстремальных свойств внутреннего производства энтропии, определенного соотношением (1.7.4) в виде

$$\mathcal{P}(t) = \frac{d_i S(t)}{dt}. \quad (1.19.1)$$

По этой причине для того, чтобы получить неравновесные соотношения Онсагера, применимые ко всем линейным системам, необходимо рассмотреть адиабатически изолированную систему (лишенную теплового контакта), в которой $d_e S = 0$ и по второму закону термодинамики $d_i S = dS \geq 0$. В этом случае соотношение (1.19.1) принимает уже использованный в (1.8.1) вид:

$$\mathcal{P}(t) = \frac{dS(t)}{dt}. \quad (1.19.2)$$

Пусть состояние неравновесной системы определяется набором n меняющихся во времени параметров $A_1(t), A_2(t), \dots, A_n(t)$ (как, например, температура, давление, концентрации компонентов и другие), от которых зависит неравновесная энтропия $S(A_1(t), A_2(t), \dots, A_n(t))$. В процессе релаксации к равновесию эти параметры монотонно изменяются, стремясь к равновесным значениям $A_1^{(0)}, A_2^{(0)}, \dots, A_n^{(0)}$, обеспечивающим максимальную равновесную энтропию $S_{\text{eq}} = S(A_1^{(0)}, A_2^{(0)}, \dots, A_n^{(0)}) = \max$, при этом из условия максимальности

$$\left(\frac{\partial S}{\partial A_\alpha} \right)_{\text{eq}} = 0 \quad \text{для всех } \alpha = 1, 2, \dots, n. \quad (1.19.3)$$

Следовательно, производство энтропии (1.19.2) принимает вид

$$\mathcal{P} \equiv \frac{dS}{dt} = \sum_{\alpha=1}^n \frac{\partial S}{\partial A_{\alpha}} \frac{dA_{\alpha}}{dt}. \quad (1.19.4)$$

Вводя в выражение (1.19.4) **термодинамические силы**

$$X_{\alpha} = \frac{\partial S}{\partial A_{\alpha}} \quad (1.19.5)$$

и **термодинамические потоки**

$$J_{\alpha} = \frac{dA_{\alpha}}{dt}, \quad (1.19.6)$$

переписываем производство энтропии в требуемой форме (1.7.6):

$$\mathcal{P}(t) = \sum_{\alpha} J_{\alpha}(t) X_{\alpha}(t). \quad (1.19.7)$$

Здесь временная зависимость $\mathcal{P}(t)$ появилась от параметров $A_{\alpha}(t)$ как функций времени.

Потоки и силы, входящие парой в виде произведения в производство энтропии (1.19.7), называются *сопряженными*.

Выражение (1.19.7) для производства энтропии в неравновесных условиях справедливо всегда, включая нелинейные режимы. Однако в линейной термодинамике можно получить определенные упрощения, вводя разностные величины (флуктуации) $x_{\alpha}(t) = A_{\alpha}(t) - A_{\alpha}^{(0)}$ и считая отклонения $x_{\alpha}(t)$ от равновесных значений $A_{\alpha}^{(0)}$ в любой момент времени существенно малыми:

$$|x_{\alpha}(t)| \ll |A_{\alpha}^{(0)}| \quad \text{для всех } \alpha = 1, 2, \dots, n. \quad (1.19.8)$$

Величины $x_{\alpha}(t)$ изменяются таким образом, что в каждый момент времени их набор однозначно характеризует процесс приближения к равновесию в этот момент. Тогда скорость $dx_{\alpha}/dt \equiv J_{\alpha}$ изменения величины x_{α} в каждом неравновесном состоянии является функцией от значений $x_1(t), x_2(t), \dots, x_n(t)$ в этом состоянии, то есть $J_{\alpha} = J_{\alpha}(x_1, x_2, \dots, x_n)$.

Это позволяет в окрестности точки равновесия при разложении потока (1.19.6) в ряд ограничиться линейными слагаемыми:

$$J_{\alpha} \equiv \frac{dx_{\alpha}}{dt} = \sum_{\beta} a_{\alpha\beta} x_{\beta} \quad (1.19.9)$$

с постоянными коэффициентами $a_{\alpha\beta}$. При этом учтено, что в равновесии отсутствуют потоки ($J_{\alpha}^{(0)} = 0$), как и силы, их вызывающие ($X_{\alpha}^{(0)} = 0$). Эти силы в неравновесном состоянии можно

также, согласно (1.19.5), представить в форме линейных сумм вида

$$X_\alpha \equiv \frac{\partial S}{\partial x_\alpha} = \sum_{\beta} b_{\alpha\beta} x_\beta \quad (1.19.10)$$

с постоянными коэффициентами $b_{\alpha\beta}$.

Если в матричном равенстве (1.19.10) величины x_β выразить через силы X_α при помощи обратной матрицы $b_{\alpha\beta}^{-1}$ и подставить их в матричное равенство (1.19.9), то потоки J_α окажутся представленными в виде искоемых линейных комбинаций X_β :

$$J_\alpha = \sum_{\beta} L_{\alpha\beta} X_\beta. \quad (1.19.11)$$

Кинетические коэффициенты $L_{\alpha\beta}$, называемые **коэффициентами Онсагера**, в рамках феноменологической теории не подлежат определению. Это возможно сделать только в рамках молекулярно-кинетической или гидродинамической теории.

Диагональные коэффициенты $L_{\alpha\alpha}$, называемые *собственными коэффициентами*, определяют вклад в поток J_α от сопряженной с ним силы X_α . Коэффициенты с $\alpha \neq \beta$ учитывают *перекрестные эффекты* от необратимых процессов α и β . Важным их свойством является симметрия по индексам α и β :

$$L_{\alpha\beta} = L_{\beta\alpha}. \quad (1.19.12)$$

Такие равенства принято называть *соотношениями взаимности Онсагера*¹⁾. Они были доказаны на основе принципа микроскопической обратимости (или принципа детального равновесия).

2. Соотношения взаимности Онсагера. Чтобы доказать эти соотношения, необходимо знать функцию статистического распределения термодинамических параметров, флуктуирующих около равновесных значений. Знание ее позволит рассчитать корреляцию между флуктуациями x_α и термодинамическими силами X_β , необходимую при доказательстве равенства (1.19.12).

Для нахождения распределения вероятности различных значений флуктуаций, $w(x_1, x_2, \dots, x_n)$, используем формулу Больцмана (1.1.1), связывающую энтропию S с термодинамической вероятностью W . Запишем эту формулу дважды — для неравновесного состояния (S, W) и для равновесного состояния $(S_{\text{eq}}, W_{\text{eq}})$:

$$S = k_B \ln W \quad \text{и} \quad S_{\text{eq}} = k_B \ln W_{\text{eq}}. \quad (1.19.13)$$

¹⁾ Следует обратить внимание, что при наличии магнитного поля $\mathbf{H}$ (не рассматриваемого нами) эти соотношения взаимности принимают такой вид: $L_{\alpha\beta}(\mathbf{H}) = L_{\beta\alpha}(-\mathbf{H})$.

Равновесное состояние более вероятное, т. е. $W_{\text{eq}} > W$. Тогда из выражений (1.19.13) получаем

$$\Delta S \equiv S - S_{\text{eq}} = k_B \ln w \leq 0. \quad (1.19.14)$$

Здесь величина $w = W/W_{\text{eq}} \leq 1$ играет роль плотности вероятности отклонений ΔS от равновесного значения S_{eq} :

$$w(\Delta S) = A e^{\Delta S/k_B}, \quad (1.19.15)$$

при этом нормировочная константа A находится из условия

$$\int_{D(\Delta S)} w(\Delta S) d\Delta S = A \int_{D(\Delta S)} e^{\Delta S/k_B} d\Delta S = 1, \quad (1.19.16)$$

где $D(\Delta S)$ — область определения энтропии ΔS как функции флуктуирующих параметров.

Отклонение ΔS от равновесного или стационарного значения было названо ранее **избыточной энтропией** (см. п. 1.9). С учетом (1.9.24) и (1.17.11), она записывается в следующем виде:

$$\Delta S = S_{\text{ex}} \equiv \frac{\delta^2 S}{2} = -\frac{1}{2T} \left(\delta T \delta S - \delta P \delta V + \sum_i \delta \mu_i \delta n_i \right). \quad (1.19.17)$$

Согласно формуле (1.17.16), избыточная (над равновесным значением) энтропия (1.19.17) всегда отрицательна ($\Delta S < 0$), что обеспечивает флуктуационную устойчивость равновесного состояния благодаря самопроизвольному затуханию флуктуаций (см. п. 1.9 и неравенства (1.17.17)–(1.17.19)).

Флуктуации физических величин (δT , δS и т. д.) выше были обозначены как x_α . Тогда для них отрицательный избыток энтропии, по аналогии с (1.19.17), можно записать в общей форме:

$$\Delta S = -\frac{k_B}{2} \sum_\alpha \sum_\beta g_{\alpha\beta} x_\alpha x_\beta. \quad (1.19.18)$$

В этом случае плотность вероятности (1.19.15) принимает вид функции Гаусса (повторяющиеся индексы подразумевают двойное суммирование по α и β от 1 до n):

$$w(x) = A \exp \left(-\frac{1}{2} g_{\alpha\beta} x_\alpha x_\beta \right), \quad (1.19.19)$$

где под x понимается совокупность всех флуктуирующих параметров $\{x_1, x_2, \dots, x_n\}$.

Несмотря на малость флуктуаций (см. формулу (1.19.8)), область интегрирования по всем x_α в интегралах типа (1.19.16) может быть распространена на значения от $-\infty$ до $+\infty$ ввиду быстрого убывания гауссовой функции с увеличением $|x_\alpha|$.

Используя (1.19.10) и (1.19.14), выражаем термодинамическую силу через функцию распределения вероятности флуктуаций следующим образом:

$$X_\alpha \equiv \frac{\partial \Delta S}{\partial x_\alpha} = k_B \frac{\partial \ln w}{\partial x_\alpha}. \quad (1.19.20)$$

Согласно общепринятому определению корреляции, вычисляем таковую между флуктуацией x_α и термодинамической силой X_β , представленной в форме (1.19.20):

$$\langle x_\alpha X_\beta \rangle = \int_{-\infty}^{\infty} x_\alpha X_\beta w(x) dx \quad \text{при} \quad dx = dx_1 dx_2 \dots dx_n.$$

Подставляем сюда выражение (1.19.20) и интегрируем по частям:

$$\begin{aligned} \langle x_\alpha X_\beta \rangle &= k_B \int_{-\infty}^{\infty} x_\alpha \frac{\partial \ln w(x)}{\partial x_\beta} w(x) dx = k_B \int_{-\infty}^{\infty} x_\alpha \frac{\partial w(x)}{\partial x_\beta} dx = \\ &= k_B x_\alpha w(x) \Big|_{x=-\infty}^{\infty} - k_B \int_{-\infty}^{\infty} \frac{\partial x_\alpha}{\partial x_\beta} w(x) dx. \end{aligned}$$

Здесь первый член обращается в нуль из-за исчезновения гауссовой функции при $|x| \rightarrow \infty$, а второй член дает символ Кронекера $\delta_{\alpha\beta}$, такой что $\delta_{\alpha\alpha} = 1$ и $\delta_{\alpha\beta} = 0$ при $\alpha \neq \beta$. Таким образом получаем искомый коррелятор:

$$\langle x_\alpha X_\beta \rangle = -k_B \delta_{\alpha\beta}. \quad (1.19.21)$$

Следовательно, каждая флуктуация x_α коррелирует только с соответствующей ей силой X_α (с одинаковым индексом α).

Теперь можно доказать соотношение (1.19.12), исходя из принципа микроскопической обратимости, который, согласно Онсагеру, звучит следующим образом: переходы между двумя конфигурациями А и В должны происходить в течение заданного отрезка времени τ одинаково часто как в прямом направлении $A \rightarrow B$, так и в обратном направлении $B \rightarrow A$.

Это утверждение равносильно *принципу детального равновесия*, в соответствии с которым: если две величины x_α и x_β влияют друг на друга в моменты времени, разделенные заданным интервалом τ , то это влияние не зависит от опережения или запаздывания во времени, что записывается в форме следующего корреляционного равенства:

$$\langle x_\alpha(t) x_\beta(t + \tau) \rangle = \langle x_\alpha(t + \tau) x_\beta(t) \rangle.$$

Продифференцировав это равенство по τ и затем положив $\tau = 0$, с учетом (1.19.9) получаем:

$$\left\langle x_\alpha \frac{dx_\beta}{dt} \right\rangle = \left\langle \frac{dx_\alpha}{dt} x_\beta \right\rangle, \quad \text{или} \quad \langle x_\alpha J_\beta \rangle = \langle J_\alpha x_\beta \rangle. \quad (1.19.22)$$

Подставляя в левую и правую части (1.19.22) вместо потоков J_β и J_α их выражения через силы в форме (1.19.11) и используя коррелятор (1.19.21), преобразуем следующим образом

- левую часть равенства (1.19.22):

$$\langle x_\alpha J_\beta \rangle = \sum_{\gamma} L_{\beta\gamma} \langle x_\alpha X_\gamma \rangle = -k_B \sum_{\gamma} L_{\beta\gamma} \delta_{\alpha\gamma} = -k_B L_{\beta\alpha}; \quad (1.19.23)$$

- правую часть равенства (1.19.22):

$$\langle J_\alpha x_\beta \rangle = \sum_{\gamma} L_{\alpha\gamma} \langle x_\beta X_\gamma \rangle = -k_B \sum_{\gamma} L_{\alpha\gamma} \delta_{\beta\gamma} = -k_B L_{\alpha\beta}. \quad (1.19.24)$$

Приравнивая, согласно (1.19.22), правые части формул (1.19.23) и (1.19.24), немедленно приходим к соотношениям взаимности Онсагера в форме (1.19.12).

Подстановка выражения (1.19.11) в формулу (1.19.7) для производства энтропии приводит ее к положительно определенной квадратичной форме:

$$\mathcal{P} = \sum_{\alpha=1}^n \sum_{\beta=1}^n L_{\alpha\beta} X_\alpha X_\beta > 0. \quad (1.19.25)$$

Положительная определенность в (1.19.25) обеспечивается теми же условиями, наложенными на симметричные коэффициенты Онсагера $L_{\alpha\beta}$, которые были сформулированы в п. 1.17 для симметричных коэффициентов μ_{ij} . В частности, для двух неравновесных процессов, одновременно протекающих в системе, при $L_{12} = L_{21}$ имеем (ср. формулу (1.17.21)):

$$L_{11} > 0, \quad L_{22} > 0, \quad L_{11}L_{22} > L_{12}^2. \quad (1.19.26)$$

Среди разнообразных прикладных приложений теории Онсагера важное место занимает ее применение в качестве основы для доказательства теоремы Пригожина.

1.20. Теорема Пригожина о минимуме производства энтропии

1. Доказательство теоремы. Теорема Пригожина формулируется для стационарного неравновесного состояния, реализуемого в любой термодинамической системы с *линейными процессами*, следующим образом:

Если линейная система характеризуется n независимыми силами $(X_1, X_2, \dots, X_n)$, из которых m сил $(X_1, X_2, \dots, X_m)$ зафиксированы **постоянными** при помощи стационарных внешних условий, то оставшиеся потоки $(J_{m+1}, J_{m+2}, \dots, J_n)$, сопряженные со свободными силами $(X_{m+1}, X_{m+2}, \dots, X_n)$, **исчезают**, в то время как потоки $(J_1, J_2, \dots, J_m)$, сопряженные с фиксированными силами, достигают постоянных значений, обеспечивающих **минимальное производство энтропии**.

Для доказательства этой теоремы воспользуемся квадратичной формой (1.19.25) для производства энтропии.

При фиксированных m силах $(X_1, X_2, \dots, X_m = \text{const})$ производство энтропии (1.19.25) как функция остающихся свободными сил $(X_{m+1}, X_{m+2}, \dots, X_n)$ минимально, если

$$\frac{\partial \mathcal{P}}{\partial X_\alpha} = 0 \quad \text{при } \alpha = m+1, m+2, \dots, n, \quad (1.20.1)$$

где при дифференцировании все термодинамические силы, кроме X_α , считаются постоянными. Подставляя в условие (1.20.1) величину $\mathcal{P}$ из уравнения (1.19.25), находим

$$\sum_{\beta=1}^n (L_{\alpha\beta} + L_{\beta\alpha}) X_\beta = 0 \quad \text{при } \alpha = m+1, m+2, \dots, n. \quad (1.20.2)$$

С учетом соотношений взаимности Онсагера (1.19.12), это дает

$$2 \sum_{\beta=1}^n L_{\alpha\beta} X_\beta = 0 \quad \text{при } \alpha = m+1, m+2, \dots, n.$$

Используя выражение (1.19.11) для потоков, получаем искомый результат:

$$J_\alpha = 0 \quad \text{при } \alpha = m+1, m+2, \dots, n. \quad (1.20.3)$$

Итак, теорема Пригожина доказана в самой общей форме, справедливой для стационарного состояния *любых* неравновесных систем (независимо от их физической природы), находящихся вдали от равновесия, но *обязательно* в линейной области протекания процессов. Для такого состояния в формуле (1.19.7) обращаются в нуль те потоки, которые сопряжены со свободными (не фиксированными) силами. Эти силы имеют возможность изменяться для того, чтобы уничтожить свои потоки и не дать вклада в производство энтропии, делая его минимальным в стационарном состоянии. Тогда исчезает зависимость от времени в выражении (1.19.7) для производства энтропии.

Следовательно, сколько бы ни было возможных физических потоков в системе, реально существующих остается лишь m , по числу постоянных сил, фиксированных внешними стационарными условиями. Такое состояние системы принято называть *стационарным состоянием m -го порядка*. В простейшем случае имеют дело с состояниями нулевого, первого и второго порядков. Понятно, что случаю $m = 0$ соответствует состояние термодинамического равновесия. Действительно, при этом на систему не накладывается никаких ограничений, и ни одна из сил не остается постоянной, т. е. системе предоставлена возможность перейти в состояние термодинамического равновесия, при котором как потоки, так и силы исчезают.

2. Применение теоремы Пригожина. Рассмотрим в системе с двумя неравновесными процессами ($n = 2$) переход из состояния второго порядка ($m = 2$) в новое стационарное состояние первого порядка ($m = 1$) путем освобождения одной из сил.

В любом неравновесном состоянии потоки и силы связаны общими соотношениями (1.19.11),

$$J_1(t) = L_{11}X_1(t) + L_{12}X_2(t), \quad (1.20.4)$$

$$J_2(t) = L_{21}X_1(t) + L_{22}X_2(t), \quad (1.20.5)$$

при $L_{21} = L_{12}$ в силу симметрии перекрестных коэффициентов.

Пусть в исходном стационарном состоянии (отмечаемом верхним индексом 0) фиксированы обе термодинамические силы:

$$X_1^0 = \text{const}, \quad X_2^0 = \text{const}. \quad (1.20.6)$$

Тогда, согласно соотношениям (1.20.4)–(1.20.5), в системе существуют исходные *стационарные потоки*:

$$J_1^0 = L_{11}X_1^0 + L_{12}X_2^0, \quad (1.20.7)$$

$$J_2^0 = L_{12}X_1^0 + L_{22}X_2^0. \quad (1.20.8)$$

Начиная с момента времени $t = 0$, выбранного за начало отсчета, создаем в системе неравновесную ситуацию, освобождая от фиксации силу X_2 и сохраняя фиксированной силу X_1 , так что

$$X_1(t) = X_1^0 = \text{const}, \quad X_2(t) = \text{var}. \quad (1.20.9)$$

В результате протекания неравновесного процесса, в соответствии с теоремой Пригожина, должно возникнуть новое стационарное состояние (с верхним индексом «ст») первого порядка

(так как фиксирована одна сила $X_1^{\text{ст}} = X_1^0$) как результат асимптотического перехода к новым стационарным значениям:

$$J_2(t) \rightarrow J_2^{\text{ст}} = 0 \quad \text{при } t \rightarrow \infty, \quad (1.20.10)$$

$$X_2(t) \rightarrow X_2^{\text{ст}} = -\frac{L_{12}}{L_{22}} X_1^0 \quad \text{при } t \rightarrow \infty. \quad (1.20.11)$$

Интересно отметить, что значение (1.20.11), которого достигает сила X_2 в новом стационарном состоянии первого порядка, полностью определяется величиной фиксированной силы X_1^0 и не зависит от начального значения X_2^0 , которое может быть произвольным, в том числе равным нулю.

Полагаем, что сила X_2 , начиная со своего исходного значения (1.20.6), асимптотически приближается к стационарной величине (1.20.11), подчиняясь уравнению релаксации:

$$\begin{aligned} X_2(t) &= -\frac{L_{12}}{L_{22}} X_1^0 + \left(X_2^0 + \frac{L_{12}}{L_{22}} X_1^0 \right) e^{-t/\tau} \equiv \\ &\equiv X_2^{\text{ст}} + \frac{1}{L_{22}} J_2^0 e^{-t/\tau}. \end{aligned} \quad (1.20.12)$$

Тогда общие соотношения (1.20.4) и (1.20.5) позволяют записать *нестационарные потоки* в следующей форме:

$$J_1(t) = L_{11} X_1^0 + L_{12} X_2(t) = J_1^{\text{ст}} + \frac{L_{12}}{L_{22}} J_2^0 e^{-t/\tau}, \quad (1.20.13)$$

$$J_2(t) = L_{12} X_1^0 + L_{22} X_2(t) = J_2^0 e^{-t/\tau}, \quad (1.20.14)$$

где, как и должно быть, при $X_1^{\text{ст}} = X_1^0$ стационарный поток равен

$$J_1^{\text{ст}} = \frac{L_{11}L_{22} - L_{12}^2}{L_{22}} X_1^0 \equiv L_{11} X_1^0 + L_{12} X_2^{\text{ст}}. \quad (1.20.15)$$

Если подставить выражения (1.20.9) и (1.20.12)–(1.20.14) в формулу (1.19.7), то получим зависимость производства энтропии от времени:

$$\begin{aligned} \mathcal{P}(t) &\equiv J_1(t)X_1(t) + J_2(t)X_2(t) = \\ &= J_1^{\text{ст}}X_1^0 + \frac{1}{L_{22}} (J_2^0)^2 e^{-2t/\tau}. \end{aligned} \quad (1.20.16)$$

Следовательно, производство энтропии уменьшается от начального значения (при $t = 0$),

$$\mathcal{P}(0) = J_1^0 X_1^0 + J_2^0 X_2^0 \equiv J_1^{\text{ст}} X_1^0 + \frac{1}{L_{22}} (J_2^0)^2, \quad (1.20.17)$$

до минимального значения в конечном состоянии (при $t \rightarrow \infty$):

$$\mathcal{P}_{\min} = J_1^{\text{ст}} X_1^0 = \left(L_{11} - \frac{L_{12}^2}{L_{22}} \right) (X_1^0)^2 < \mathcal{P}(0). \quad (1.20.18)$$

При этом второй поток, сопряженный со свободной силой X_2 , исчезает ($J_2 = 0$) в полном соответствии с теоремой Пригожина.

Понятно, что если предоставить возможность одновременно изменяться обеим силам, то оба потока будут стремиться к нулю ($J_1(t) \rightarrow 0$ и $J_2(t) \rightarrow 0$). В итоге система достигнет стационарного состояния нулевого порядка, т. е. термодинамического равновесия, при котором исчезают также и силы ($X_1 = 0$ и $X_2 = 0$).

Применим полученные соотношения к *термодиффузии* (называемой *эффектом Соре*), проявляющейся в форме неоднородного пространственного распределения компонентов в системе под влиянием градиента температуры.

Пусть термодинамические силы X_1 и X_2 суть отрицательные градиенты температуры и концентрации i -го компонента, которые в одномерном случае (вдоль координаты x) имеют вид

$$X_1 = -\nabla_x T \equiv -\frac{\partial T}{\partial x} \quad \text{и} \quad X_2 = -\nabla_x c_i \equiv -\frac{\partial c_i}{\partial x}. \quad (1.20.19)$$

Для бинарной системы (с суммарной концентрацией частиц $c = c_1 + c_2$) одна из общепринятых форм записи потока частиц при $P = \text{const}$ такова ($i = 1, 2$, см. ниже формулу (1.23.51)):

$$J_i = -D_{12} \nabla_x c_i - c D_T \nabla_x T / T, \quad (1.20.20)$$

где D_{12} — коэффициент взаимной диффузии ($D_{12} = D_{21}$), D_T — коэффициент термодиффузии. В соответствии с предыдущими обозначениями $J_i = J_2$, тогда в стационарном состоянии $J_i^{\text{ст}} = 0$. Из формулы (1.20.20) получаем стационарный градиент концентрации, порожденный данным градиентом температуры ($i = 1, 2$):

$$(\nabla_x c_i)^{\text{ст}} = -k_T c \nabla_x T / T, \quad (1.20.21)$$

где $k_T = D_T / D_{12}$ — термодиффузионное отношение, пропорциональное произведению концентраций компонентов с коэффициентом пропорциональности, называемым *коэффициентом Соре* S_T (см. формулы (1.23.60) и (1.23.61)).

Сравнение формул (1.20.11) и (1.20.21), с учетом (1.20.19), показывает, что для термодиффузии $L_{12}/L_{22} = k_T c / T$. Поскольку для бинарной смеси $\nabla_x c_1 = -\nabla_x c_2$, в неоднородном температурном поле возникает пространственное разделение компонентов.

Ниже (см. п. 1.23) приведен подробный анализ эффекта Соре (термодиффузии) и обратного ему эффекта Дюфура (диффузионной теплопроводности).

3. Обобщенный принцип Ле Шателье. Распространим известный принцип Ле Шателье¹⁾ (обычно применяемый только к равновесным ситуациям, см. п. 2.15) на случай неравновесных стационарных процессов, управляемых теоремой Пригожина.

Пусть в системе, подверженной действию n термодинамических сил X_α ($\alpha = 1, 2, \dots, n$) и находящейся в стационарном состоянии m -го порядка, одна из $n - m$ свободных сил, например с номером $\alpha = p$ ($m < p < n$), испытывает возмущение δX_p . Все остальные силы при этом сохраняют свои прежние стационарные значения (отмеченные верхним индексом 0). Тогда имеем

$$\begin{aligned} X_\alpha &= X_\alpha^0 \quad \text{при } \alpha = 1, \dots, m, \dots, p-1, p+1, \dots, n, \\ X_\alpha &= X_p = X_p^0 + \delta X_p \quad \text{при } \alpha = p \quad (m < p < n). \end{aligned} \quad (1.20.22)$$

В соответствии с общим уравнением (1.19.11), можно, с учетом (1.20.22), записать поток, сопряженный с возмущенной силой X_p ,

$$J_p = L_{pp} X_p + \sum_{\beta \neq p}' L_{p\beta} X_\beta = J_p^0 + L_{pp} \delta X_p. \quad (1.20.23)$$

Здесь невозмущенное значение термодинамического потока,

$$J_p^0 = \sum_{\beta=1}^n L_{p\beta} X_\beta^0,$$

согласно теореме Пригожина (формула (1.20.3)), должно быть равным нулю, поскольку сила X_p не фиксирована ($m < p < n$). Поэтому выражение (1.20.23) принимает упрощенную форму

$$J_p = L_{pp} \delta X_p. \quad (1.20.24)$$

Так как производство энтропии $\mathcal{P}$ всегда положительно, то делаем вывод о том, что $L_{pp} > 0$. На этом основании из выражения (1.20.24) заключаем, что

$$J_p \delta X_p = L_{pp} (\delta X_p)^2 > 0, \quad (1.20.25)$$

¹⁾ Ле Шателье (Le Chatelier) Анри Луи (1850–1936) — французский физик, химик и металлург, одним из первых проводил фундаментальные исследования химико-технологических процессов в металлургии, создал металлографический микроскоп, термоэлектрический пирометр и усовершенствовал методику исследования металлов и сплавов, впервые (1884) сформулировал общий принцип смещения химического равновесия.

то есть возмущение термодинамической силы и сопряженный с ней поток всегда имеют одинаковые знаки.

Однако, как следует из ранее рассмотренного, положительный поток J_p приводит к отрицательному изменению силы δX_p одного с ним направления, уменьшая ее до нуля. Это связано с тем, что термодинамическая сила направлена противоположно соответствующему градиенту (см. формулы (1.20.19)).

Применим вышеприведенные рассуждения к термодиффузии. Пусть в бинарной смеси постоянный температурный градиент создает, согласно (1.20.21), градиент концентрации i -го компонента. Предположим, что этот компонент концентрируется в высокотемпературной области. Если добавить вещество i в более холодную область, тогда градиент его концентрации $\nabla_x c_i$ уменьшается, а сила $X_2 = -\nabla_x c_i$ увеличивается, то есть получаем положительное возмущение силы ($\delta X_2 > 0$). Условие (1.20.25), выражающее принцип Ле Шателье, требует возникновения положительного потока ($J_2 > 0$). Этот поток перемещает в положительном направлении введенный избыток i -го вещества из холодной области в более горячую. Таким образом, возмущение δX_2 компенсируется возникшим потоком J_2 , как того требует принцип Ле Шателье.

Запишем выражение для производства энтропии в возмущенном состоянии системы, в котором термодинамические силы описываются формулой (1.20.22):

$$\mathcal{P} = \sum_{\alpha=1}^n \sum_{\beta=1}^n L_{\alpha\beta} X_{\alpha}^0 X_{\beta}^0 + \delta X_p \sum_{\beta=1}^n (L_{p\beta} + L_{\beta p}) X_{\beta}^0 + L_{pp} (\delta X_p)^2.$$

Здесь первый член есть минимум производства энтропии в невозмущенном состоянии ($\mathcal{P}_{\min} \equiv \mathcal{P}^0$), а второй член равен нулю в соответствии с требованием теоремы Пригожина (1.20.2). Таким образом, возмущенное значение производства энтропии равно

$$\mathcal{P} = \mathcal{P}_{\min} + L_{pp} (\delta X_p)^2 \rightarrow \mathcal{P}_{\min}. \quad (1.20.26)$$

Следовательно, любое возмущение системы, находящейся в состоянии с минимальным производством энтропии, индуцирует протекание процессов, противодействующих исходному возмущению. В этом и заключается обобщенный принцип Ле Шателье, выражаемый формулой (1.20.26).

В состоянии термодинамического равновесия, когда все силы X_{α}^0 обращаются в нуль, исчезает $\mathcal{P}_{\min}$, и формула (1.20.26) выражает обычный принцип Ле Шателье (см. п. 2.15), нередко называемый *принципом Ле Шателье–Брауна*.

1.21. Локальное квазиравновесие и локальные термодинамические соотношения

До сих пор рассматривались полные значения термодинамических величин, присущие всей системе. Это применимо для описания равновесного состояния, когда система пространственно однородна. Однако в неравновесных условиях пространственная однородность нарушается, так что предыдущее рассмотрение применимо лишь к описанию интегральных свойств системы. Возникает вполне естественный вопрос о том, что же привносит учет локальных особенностей термодинамических систем, находящихся в неравновесном состоянии.

1. Понятие о локальном квазиравновесии. Как уже отмечалось в п. 1.1, время установления равновесия в системе зависит от ее размеров. Действительно, неравновесная замкнутая система с относительно малым числом частиц (но в то же время достаточно большим, чтобы считать ее макроскопической) перераспределяет энергию равновероятно по всем доступным ей степеням свободы более быстро, чем такая же система большего размера. Это позволяет говорить о *глобальном* (или полном) *равновесии* замкнутой системы и *локальном* (или частичном) *равновесии* ее отдельных макрочастей, характеризующихся разными временами релаксации $\tau_{\text{гл}}$ и $\tau_{\text{лок}}$, при этом $\tau_{\text{лок}} \ll \tau_{\text{гл}}$. За времена Δt , такие что $\tau_{\text{лок}} \ll \Delta t \ll \tau_{\text{гл}}$, отдельные подсистемы ведут себя как *квазизамкнутые*, успевая восстановить внутри себя *локальное квазиравновесие*, в то время как система в целом может находиться в неполном равновесии или даже быть далекой от равновесия.

Более того, отдельные физические процессы, протекающие внутри системы с разными скоростями, обычно характеризуют различными временами релаксации.

Например, из-за того, что в реактивной газовой смеси лишь малая часть столкновений из общего их числа приводит к химическому взаимодействию между молекулами, равновесие по отношению к составу смеси (*химическое равновесие*) устанавливается медленнее, чем по отношению к их движению (*газодинамическое* или *гидродинамическое равновесие*). По этой причине химические процессы считают протекающими в пространственно однородных условиях, если, конечно, вдали от равновесия не возникают пространственные (или пространственно-временные) диссипативные структуры, которые сами являются порождением исходной пространственно однородной ситуации.

Другим примером может служить коллектив легких частиц (например, электронов в газовой или твердотельной плазме), взаимодействующих с тяжелыми рассеивателями (молекулами, атомами, ионами и фононами кристалла). При этом процесс рассеяния является квазиупругим, так как из-за большого различия по массе электрон при каждом столкновении практически не теряет энергию, но эффективно рассеивает свой импульс. Поэтому процессы обмена импульсом и энергией между электронами и тяжелыми рассеивателями характеризуются разными временами релаксации — временем релаксации импульса τ_p и временем релаксации энергии τ_ε , такими что $\tau_p \ll \tau_\varepsilon$.

Кроме того, электроны, как термодинамическая подсистема кристалла, характеризуются временем межэлектронных столкновений τ_{ee} . За времена порядка τ_{ee} электроны, обладая равной массой, эффективно обмениваются между собой одновременно энергией и импульсом, стремясь сформировать внутри своего коллектива равновесное максвелловское распределение по скоростям. Это возможно лишь в тех случаях, когда электронная подсистема является *квазизамкнутой* по отношению к процессу обмена с рассеивателями либо только энергией (при $\tau_p \ll \tau_{ee} \ll \tau_\varepsilon$), либо как энергией, так и импульсом (при $\tau_{ee} \ll \tau_p \ll \tau_\varepsilon$).

В первом случае (при $\tau_p \ll \tau_{ee} \ll \tau_\varepsilon$) энергия электронного газа практически неизменна за времена порядка τ_{ee} (из-за $\tau_\varepsilon \gg \tau_{ee}$), так что электроны в состоянии сформировать *максвелловское распределение* с электронной температурой T :

$$f_M(\mathbf{r}, \mathbf{v}, t) = c \left(\frac{m}{2\pi k_B T} \right)^{3/2} \exp \left(-\frac{mv^2}{2k_B T} \right), \quad (1.21.1)$$

где m и c — эффективная масса и концентрация электронов.

Во втором случае (при $\tau_{ee} \ll \tau_p \ll \tau_\varepsilon$) время межэлектронных столкновений становится настолько малым (при увеличении концентрации электронов), что за времена порядка τ_{ee} остаются практически неизменными внутри электронной подсистемы не только энергия (из-за $\tau_\varepsilon \gg \tau_{ee}$), но и импульс (из-за $\tau_p \gg \tau_{ee}$). В этом случае электроны в состоянии перераспределять внутри своего коллектива (без участия рассеивателей) как энергию (формируя электронную температуру T), так и импульс (формируя общую гидродинамическую скорость $\mathbf{v}$), которые определяют так называемое *смещенное максвелловское распределение*:

$$f_M^{cm}(\mathbf{r}, \mathbf{v}, t) = c \left(\frac{m}{2\pi k_B T} \right)^{3/2} \exp \left(-\frac{m(\mathbf{v} - \mathbf{v})^2}{2k_B T} \right). \quad (1.21.2)$$

Строго говоря, появляющееся в первом приближении максвелловское распределение (1.21.1) не есть полная функция распределения электронов по скоростям $\mathbf{v}$, а лишь ее симметричная часть в пространстве скоростей, в отличие от смещенной максвелловской функции (1.21.2). Обе они, как записано выше, зависят не только от $\mathbf{v}$, но и от координат $\mathbf{r}$ и времени t , что отражено в пространственно-временной зависимости макроскопических параметров распределения — $c(\mathbf{r}, t)$, $T(\mathbf{r}, t)$ и $\mathbf{v}(\mathbf{r}, t)$.

Описание физических систем в рамках максвелловских распределений (1.21.1) и (1.21.2) принято называть, соответственно, *квазигидродинамическим приближением* и *гидродинамическим приближением*. В обоих случаях принципиальную роль в формировании этих распределений играют межэлектронные столкновения, большая частота которых обеспечивает соответствующую *квазизамкнутость* электронной подсистемы (только по энергии для (1.21.1) или по энергии и импульсу для (1.21.2)).

В условиях квазигидродинамики (энергетическая квазизамкнутость при $\tau_p \ll \tau_{ee} \ll \tau_\varepsilon$) имеет место так называемый *энергетический контроль* со стороны электронов за перераспределением полученной извне энергии. В условиях гидродинамики (полная квазизамкнутость при $\tau_{ee} \ll \tau_p \ll \tau_\varepsilon$) к энергетическому контролю добавляется *импульсный контроль*, формирующий гидродинамическое движение электронов со средней скоростью $\mathbf{v}$.

Следует подчеркнуть, что в неравновесной ситуации (при наличии потоков) обе максвелловские функции (1.21.1) и (1.21.2) являются *локально квазиравновесными*, так как содержат зависящие от координат и времени величины — скорость $\mathbf{v}(\mathbf{r}, t)$, концентрацию $c(\mathbf{r}, t)$ и температуру $T(\mathbf{r}, t)$ (в общем случае отличную от температуры T_0 окружающей среды). Эти величины отражают вклад внешних рассеивателей в функцию распределения, максвелловская форма которой формируется без их участия самими электронами. И наоборот, пространственно-временные масштабы и свойства макроскопических (гидродинамических) величин формируются без участия электронов внешними для них объектами, включая рассеивающие центры. Последнее обстоятельство говорит в пользу того, что термин «локальное квазиравновесие», используемый нами, правильнее отражает физическую ситуацию в сравнении с термином «локальное равновесие», часто встречающимся в литературе.

Следует иметь в виду, что температура как термодинамическая характеристика произвольной (не только электронной) системы может быть введена в рассмотрение лишь при наличии максвелловского распределения по скоростям внутри коллектива

любых частиц, входящих в состав термодинамической системы. Это основная черта, присущая всем без исключения локально квазиравновесным физическим ситуациям.

2. Локальные величины и термодинамические соотношения. В основе локально квазиравновесного описания неравновесной термодинамической системы лежит разбиение ее объема V на отдельные малые части ΔV , называемые *физически бесконечно малыми частицами (ФБМ-частицами)*. Такая ФБМ-частица, локализованная в точке с координатами $\mathbf{r}$, имеет предельно малые размеры по сравнению с характерным *пространственным масштабом* макроскопического процесса. Она содержит большое число реальных частиц, взаимодействия между которыми практически мгновенно (за времена, малые по сравнению с *временным масштабом* макропроцесса) формируют в момент времени t максвелловское распределение типа (1.21.2) с локальными параметрами — $c(\mathbf{r}, t)$, $\mathbf{v}(\mathbf{r}, t)$ и $T(\mathbf{r}, t)$. Изменение этих параметров в пространстве и во времени описывается *гидродинамическими уравнениями* (называемыми транспортными уравнениями или уравнениями переноса), которые выражают в общем случае баланс числа частиц (или массы), импульса и энергии.

Следовательно, термодинамические величины (как интенсивные, так и экстенсивные), подобно гидродинамическим переменным, характеризующим неравновесную систему, должны стать функциями координат и времени. *Интенсивные величины* при этом сохраняют свой «интенсивный смысл», но рассматриваются в квазиравновесных условиях как величины, зависящие от $\mathbf{r}$ и t :

$$T = T(\mathbf{r}, t), \quad P = P(\mathbf{r}, t), \quad \mu_i = \mu_i(\mathbf{r}, t). \quad (1.21.3)$$

Экстенсивные величины (такие как энтропия S , внутренняя энергия U , энтальпия $H = U + PV$, свободная энергия Гельмгольца $F = U - TS$ и свободная энергия Гиббса $G = H - TS$) меняют свой «экстенсивный смысл» и превращаются в интенсивные величины путем их замены либо *удельной величиной* (на единицу массы), либо *объемной плотностью* (на единицу объема), часто называемой просто плотностью величины.

Несмотря на то, что более распространенным в литературе оказывается первый вариант замены, для дальнейшего рассмотрения воспользуемся плотностями вышеуказанных экстенсивных величин, обозначая их аналогичными малыми буквами, а именно

- *плотности* энтропии, внутренней энергии и энтальпии:

$$s = \frac{S}{V}, \quad u = \frac{U}{V}, \quad h = \frac{H}{V} = u + P; \quad (1.21.4)$$

- *плотности* свободных энергий Гельмгольца и Гиббса:

$$f = \frac{F}{V} = u - Ts, \quad g = \frac{G}{V} = h - Ts. \quad (1.21.5)$$

Кроме того, из общего выражения (1.11.3) для свободной энергии Гиббса получаем

$$g = \frac{1}{V} \sum_i n_i \mu_i = \sum_i c_i \mu_i, \quad (1.21.6)$$

где для i -го компонента ($i = 1, 2, \dots, K$) использована мольно-объемная концентрация (иначе молярная плотность) $c_i = n_i/V$.

Из выражений (1.21.4)–(1.21.6) следует важное для дальнейших преобразований соотношение

$$Ts = u + P - \sum_i c_i \mu_i. \quad (1.21.7)$$

Выражения (1.21.4)–(1.21.6) применимы лишь для гомогенных систем, так как они получены из соответствующих полных величин простым делением на объем V всей системы. Ниже они будут распространены на неоднородные квазиравновесные системы, для которых все величины в формулах (1.21.4)–(1.21.7) становятся локальными функциями $\mathbf{r}$ и t .

Выясним, какой вид принимает фундаментальное уравнение Гиббса (1.3.5), имеющее для полных величин общую форму

$$dU = TdS - PdV + \sum_i \mu_i dn_i, \quad (1.21.8)$$

если его записать для объемных плотностей (1.21.4)–(1.21.6). При приведении к единице объема будем полагать, что объем V системы может изменяться.

Для этого обе части уравнения (1.21.8) необходимо разделить на объем V и использовать для экстенсивных величин U , S и n_i известное правило дифференцирования, записанное в качестве примера для внутренней энергии U :

$$du \equiv d \frac{U}{V} = \frac{dU}{V} - U \frac{dV}{V^2}. \quad (1.21.9)$$

В результате этого из уравнения Гиббса (1.21.8) получаем

$$du = Tds + \sum_i \mu_i d\left(\frac{n_i}{V}\right) - \left[U - TS + PV - \sum_i n_i \mu_i \right] \frac{dV}{V^2}. \quad (1.21.10)$$

Здесь квадратная скобка обращается в нуль, так как изобарный потенциал $G \equiv U - TS + PV$, согласно (1.11.3), равен $\sum_i n_i \mu_i$.

Таким образом, фундаментальное уравнение Гиббса (1.21.8), переписанное для *плотностей экстенсивных величин*, принимает следующий вид:

$$du = Tds + \sum_i \mu_i dc_i. \quad (1.21.11)$$

Следует обратить внимание на тот факт, что полученное уравнение Гиббса (1.21.11) для плотностей экстенсивных величин, в сравнении с уравнением (1.21.8) для полных величин, не содержит объемного вклада, несмотря на отсутствие требования о постоянстве объема¹⁾.

Уравнение (1.21.11) позволяет записать аналогичные дифференциальные выражения для объемной плотности других величин, определенных формулами (1.21.4)–(1.21.5):

$$df = -sdT + \sum_i \mu_i dc_i, \quad (1.21.12)$$

$$dh = Tds + dP + \sum_i \mu_i dc_i, \quad (1.21.13)$$

$$dg = -sdT + dP + \sum_i \mu_i dc_i. \quad (1.21.14)$$

Следует отметить, что в формулах (1.21.13) и (1.21.14) независимое приращение dP имеет множителем единицу как следствие того, что все величины отнесены к единице объема. По этой же причине в формулах (1.21.11) и (1.21.12) полностью отсутствует вклад от приращения dV , несмотря на то, что объем V не считается постоянным.

Из полных дифференциалов объемных плотностей (1.21.11)–(1.21.14) вытекают следующие частные производные (ср. формулы (1.10.5), (1.10.6) и (1.10.8)):

$$\left(\frac{\partial u}{\partial s}\right)_{c_i} = \left(\frac{\partial h}{\partial s}\right)_{P, c_i} = T, \quad (1.21.15)$$

$$\left(\frac{\partial f}{\partial T}\right)_{c_i} = \left(\frac{\partial g}{\partial T}\right)_{P, c_i} = -s, \quad (1.21.16)$$

$$\left(\frac{\partial h}{\partial P}\right)_{s, c_i} = \left(\frac{\partial g}{\partial P}\right)_{T, c_i} = 1, \quad (1.21.17)$$

¹⁾ В отличие от этого, аналогичное уравнение Гиббса для удельных величин имеет вид $d\bar{u} = Td\bar{s} - Pdv + \sum_i \mu_i d\bar{c}_i$, где все величины отнесены к единице массы (что отмечено чертой), а $v = 1/\rho$ означает удельный объем (см. п. 1.24).

а также следующие выражения для химического потенциала i -го компонента (ср. формулу (1.11.1)):

$$\begin{aligned}\mu_i &= \left(\frac{\partial u}{\partial c_i} \right)_{s, c_j \neq c_i} = \left(\frac{\partial h}{\partial c_i} \right)_{s, P, c_j \neq c_i} = \\ &= \left(\frac{\partial f}{\partial c_i} \right)_{T, c_j \neq c_i} = \left(\frac{\partial g}{\partial c_i} \right)_{T, P, c_j \neq c_i}.\end{aligned}\quad (1.21.18)$$

Равенство единице в формуле (1.21.17) отражает тот факт, что плотности величин соответствуют единичному объему, как ясно видно из сравнения с формулой (1.10.8).

По аналогии с химическим потенциалом, являющимся, как следует из (1.21.6), парциальной молярной энергией Гиббса ($g_i \equiv \mu_i$), введем в рассмотрение другие парциальные молярные величины u_i , h_i , f_i , s_i (ср. уравнения (1.11.8)–(1.11.11)):

$$u = \sum_i c_i u_i, \quad \text{где} \quad u_i = \left(\frac{\partial u}{\partial c_i} \right)_{T, P, c_j \neq c_i}; \quad (1.21.19)$$

$$h = \sum_i c_i h_i, \quad \text{где} \quad h_i = \left(\frac{\partial h}{\partial c_i} \right)_{T, P, c_j \neq c_i}; \quad (1.21.20)$$

$$f = \sum_i c_i f_i, \quad \text{где} \quad f_i = \left(\frac{\partial f}{\partial c_i} \right)_{T, P, c_j \neq c_i}; \quad (1.21.21)$$

$$s = \sum_i c_i s_i, \quad \text{где} \quad s_i = \left(\frac{\partial s}{\partial c_i} \right)_{T, P, c_j \neq c_i}. \quad (1.21.22)$$

С учетом формул (1.21.4) и (1.21.5), из выражений (1.21.19)–(1.21.22) следуют равенства $u_i = h_i$ и $f_i = g_i$. Следовательно, для химического потенциала справедливы соотношения:

$$\mu_i \equiv g_i = h_i - T s_i = u_i - T s_i = f_i. \quad (1.21.23)$$

Нетрудно увидеть, что при $T = \text{const}$ и $P = \text{const}$ уравнения (1.21.7) и (1.21.11) дают аналог уравнения Гиббса–Дюгема (1.11.6), записанный для мольно-объемных концентраций:

$$\sum_i c_i (d\mu_i)_{T, P} = 0. \quad (1.21.24)$$

Входящий сюда дифференциал химического потенциала учитывает только вклад мольно-объемных концентраций c_i , в то

время как полный дифференциал при $T = \text{var}$ и $P = \text{var}$ равен

$$d\mu_i = (d\mu_i)_{T,P} + \left(\frac{\partial\mu_i}{\partial T}\right)_{P,c_i} dT + \left(\frac{\partial\mu_i}{\partial P}\right)_{T,c_i} dP, \quad (1.21.25)$$

где частные производные вычисляются с учетом выражений (1.21.16)–(1.21.18) и (1.21.22):

$$\left(\frac{\partial\mu_i}{\partial T}\right)_{P,c_i} = \left(\frac{\partial^2 g}{\partial c_i \partial T}\right)_{P,c_j \neq c_i} = -\left(\frac{\partial s}{\partial c_i}\right)_{T,P,c_j \neq c_i} \equiv -s_i, \quad (1.21.26)$$

$$\left(\frac{\partial\mu_i}{\partial P}\right)_{T,c_i} = \frac{\partial}{\partial P} \left(\frac{\partial g}{\partial c_i}\right)_{T,c_j \neq c_i} = \frac{\partial}{\partial c_i} \left(\frac{\partial g}{\partial P}\right)_{T,c_j \neq c_i} \equiv 0. \quad (1.21.27)$$

Согласно равенству (1.21.27), химические потенциалы μ_i^c (выраженные через мольно-объемные концентрации $c_i = n_i/V$), в отличие от аналогичных величин μ_i (зависящих от числа молей n_i), не чувствуют давления P (ср. формулу (1.11.7)):

$$\mu_i^c = \mu_i(T, c_1, c_2, \dots, c_K). \quad (1.21.28)$$

С учетом формул (1.21.23) и (1.21.24), полный дифференциал (1.21.25) химического потенциала принимает вид

$$d\mu_i^c = (d\mu_i^c)_T + \left(\frac{\partial\mu_i^c}{\partial T}\right)_{c_i} dT = (d\mu_i^c)_T - s_i dT, \quad (1.21.29)$$

где верхний индекс c отражает зависимость химического потенциала от молярных плотностей c_i в формуле (1.21.28).

Основной концепцией **локального квазиравновесия** является утверждение о том, что термодинамические соотношения равновесной термодинамики остаются применимыми к локальным переменным внутри физически бесконечно малого (ФБМ) объема. Теперь они являются функциями $\mathbf{r}$ и t , медленно меняющимися в сравнении с характерными масштабами микроскопических процессов, устанавливающих квазиравновесие внутри ФБМ-объема.

В неравновесных локально неоднородных системах вместо выражений (1.21.4) и (1.21.5) для объемных плотностей (введенных путем деления полных величин на объем системы) используются соответствующие функции координат и времени, например $s(\mathbf{r}, t)$, $u(\mathbf{r}, t)$ и др. В этом случае полные величины (энтропия, внутренняя энергия и др.) выражаются через локальные «полевые» значения в интегральной форме, например

$$S(t) = \int s(\mathbf{r}, t) dV, \quad U(t) = \int u(\mathbf{r}, t) dV \quad \text{и др.} \quad (1.21.30)$$

Следовательно, все приведенные выше термодинамические соотношения для объемных плотностей в условиях локального квазиравновесия остаются применимыми к неравновесным системам. Для их анализа необходимо, в дополнение к термодинамическим уравнениям, иметь транспортные (гидродинамические) уравнения, выражающие баланс физических величин (числа молей, энергии и энтропии).

1.22. Локальные уравнения баланса числа частиц, внутренней энергии и энтропии

Как уже было отмечено, в основе локального описания неравновесной системы лежит разбиение ее объема V на *физически бесконечно малые* подсистемы объемом $\Delta V \ll V$, локально сохраняющие термодинамическую специфику исходной системы. Такие элементарные объемы являются *квазизамкнутыми* в том понимании, что диссипирующие взаимодействия между реальными частицами (нередко называемые *мелкомасштабными флуктуациями*) за относительно короткие (по сравнению с характерным масштабом рассматриваемого макропроцесса) времена $\tau_{\text{лок}}$ устанавливают локальное (внутреннее) равновесие, формируя максвелловское распределение частиц по скоростям (см. с. 80).

Однако такое внутреннее равновесие является лишь *квазиравновесием* из-за взаимодействия каждой подсистемы со своим внешним окружением (резервуаром), в пределе обладающим бесконечным числом степеней свободы. Этот резервуар навязывает (через граничные условия) данной подсистеме локальные интенсивные параметры (1.21.3), под которые подстраиваются ее экстенсивные параметры (1.21.4)–(1.21.5) и др. Локальные значения этих параметров отличаются от их средних по всей системе значений (формируемых за времена $\tau_{\text{гл}}$ установления глобального равновесия), что нередко трактуется как *крупномасштабные флуктуации*, связанные с появлением потоков в системе.

1. Гидродинамический аспект локального квазиравновесия. Пространственно-временные изменения крупномасштабных флуктуаций описываются на основе гидродинамической теории *конвективного* и *кондуктивного* (диффузионного) *переноса*, что выводит рассматриваемую среду за рамки сугубо термодинамического описания. Принципиальная особенность такой теории определяется необходимостью учета как макроскопического движения всей среды в целом (характеризуемого локальной средней скоростью $\mathbf{v}(\mathbf{r}, t)$), так и аналогичного движения ее отдельных

компонентов (характеризуемого локальными средними скоростями $\mathbf{v}_i(\mathbf{r}, t)$, $i = 1, 2, \dots, K$) с диссипацией энергии и производством энтропии.

Если гидродинамические скорости $\mathbf{v}_i$ компонентов известны, то *общая скорость движения среды* определена как

$$\mathbf{v} = \frac{\sum_i c_i \mathbf{v}_i}{\sum_i c_i} \equiv \sum_i x_i \mathbf{v}_i, \quad (1.22.1)$$

где для i -го компонента введена его мольная доля $x_i = c_i/c$.

Следовательно, определенная равенством (1.22.1) скорость $\mathbf{v}$ есть средняя мольная (молярная) скорость движения многокомпонентной среды, создающая общую плотность *конвективного потока* молей (в единицах моль/(м²·с)):

$$\mathbf{J} \equiv c\mathbf{v} = \sum_i c_i \mathbf{v}_i. \quad (1.22.2)$$

Пространственная неоднородность общей гидродинамической скорости движения среды, $\mathbf{v}(\mathbf{r})$, характеризуется *градиентом вектора* $\mathbf{v}$, который в декартовых прямоугольных координатах x_1, x_2, x_3 с единичными ортами $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ имеет следующий вид:

$$\nabla \mathbf{v} = \begin{pmatrix} \frac{\partial v_1}{\partial x_1} & \frac{\partial v_2}{\partial x_1} & \frac{\partial v_3}{\partial x_1} \\ \frac{\partial v_1}{\partial x_2} & \frac{\partial v_2}{\partial x_2} & \frac{\partial v_3}{\partial x_2} \\ \frac{\partial v_1}{\partial x_3} & \frac{\partial v_2}{\partial x_3} & \frac{\partial v_3}{\partial x_3} \end{pmatrix} \equiv \frac{\partial v_l}{\partial x_k} \quad (k, l = 1, 2, 3). \quad (1.22.3)$$

Таким образом, тензор $\nabla \mathbf{v}$ в форме (1.22.3) представляет собой так называемую *диаду*, построенную из векторов набла $\nabla = \mathbf{e}_1(\partial/\partial x_1) + \mathbf{e}_2(\partial/\partial x_2) + \mathbf{e}_3(\partial/\partial x_3)$ и $\mathbf{v} = \mathbf{e}_1 v_1 + \mathbf{e}_2 v_2 + \mathbf{e}_3 v_3$. Последнее тождественное равенство в формуле (1.22.3) с немymi индексами k и l (по которым проводится свертка, то есть суммирование в случае их повторения в разных сомножителях) означает совокупность элементов $(\nabla \mathbf{v})_{kl} = \partial v_l / \partial x_k$ диады $\nabla \mathbf{v}$. След матрицы (1.22.3), равный сумме диагональных элементов, определяет дивергенцию вектора скорости:

$$\operatorname{div} \mathbf{v} = \bar{\mathbf{I}} : \nabla \mathbf{v} \equiv \delta_{kl} \frac{\partial v_l}{\partial x_k} = \frac{\partial v_k}{\partial x_k} \equiv \frac{\partial v_1}{\partial x_1} + \frac{\partial v_2}{\partial x_2} + \frac{\partial v_3}{\partial x_3}. \quad (1.22.4)$$

Здесь введено двойное скалярное произведение (обозначенное $:$) единичного тензора $\bar{\mathbf{I}} \equiv \delta_{kl}$ (символ Кронекера, равный единице при $k = l$ и нулю при $k \neq l$) и диады $\nabla \mathbf{v}$.

Пространственная неоднородность скорости, представленная в форме диады (1.22.3), играет принципиальную роль в формировании *вязкостного движения* среды. Это порождает, как увидим ниже, диссипацию энергии и соответствующий (гидродинамический) вклад в производство энтропии, в дополнение к термодинамическим вкладам за счет теплопроводности, диффузии и химических реакций.

Ближайшая задача состоит в феноменологическом обосновании уравнений баланса числа молей, энергии и энтропии квазиравновесной системы, в которой имеет место гидродинамическое движение среды со средней молярной скоростью (1.22.1).

2. Лагранжево и эйлерово описания сплошной среды. Выделим ФБМ-частицу среды, находящуюся в момент времени t в точке с радиусом-вектором $\mathbf{r}(t)$, как показано на рис. 1.11, и проследим за ее движением вдоль траектории со скоростью

$$\mathbf{v}(t) = \frac{d\mathbf{r}(t)}{dt}. \quad (1.22.5)$$

Подобное определение скорости соответствует **описанию Лагранжа**, в котором основной динамической переменной является радиус-вектор $\mathbf{r}(t)$ как функция времени. Зависимость $\mathbf{r}(t)$ определяет все динамические характеристики процесса в «траекторном» представлении по Лагранжу: не только скорость (1.22.5), но и ускорение ФБМ-частицы, $\mathbf{a}(t) \equiv d\mathbf{v}(t)/dt = d^2\mathbf{r}(t)/dt^2$.

При переходе к **описанию Эйлера** радиус-вектор $\mathbf{r}$ теряет свой динамический смысл и принимает неизменное во времени значение координаты для любой фиксированной точки пространства. В этом случае координата $\mathbf{r}$ и время t работают как независимые *эйлеровы переменные*, от которых зависят все динамические характеристики процесса в «полево» представлении по Эйлеру. Основной динамической величиной в эйлеровом описании является поле скоростей $\mathbf{v}(\mathbf{r}, t)$.

Несмотря на то, что координата $\mathbf{r}$ и время t не зависят друг от друга в эйлеровых переменных, при вычислении полной производной по времени необходимо учесть движение ФБМ-частицы вдоль траектории с помощью функциональной зависимости $\mathbf{r}(t)$. Она теперь связывает между собой $\mathbf{r}$ и t для выбранной частицы.

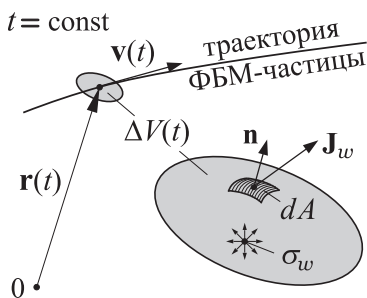


Рис. 1.11. К лагранжеву описанию движения ФБМ-частицы вдоль траектории

Пользуясь этим представлением, вычислим полную производную от скорости (1.22.5) как «полевой» величины $\mathbf{v}(\mathbf{r}, t)$:

$$\frac{d\mathbf{v}(\mathbf{r}, t)}{dt} \equiv \frac{d\mathbf{v}(\mathbf{r}(t), t)}{dt} = \frac{\partial \mathbf{v}(\mathbf{r}, t)}{\partial t} + \frac{\partial \mathbf{v}(\mathbf{r}, t)}{\partial \mathbf{r}} \cdot \frac{d\mathbf{r}(t)}{dt}. \quad (1.22.6)$$

Отсюда, вводя градиент скорости (1.22.3) и учитывая формулу (1.22.5), окончательно получаем искомую связь между полной и частной производными по времени от вектора скорости в виде

$$\frac{d\mathbf{v}}{dt} = \frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v}. \quad (1.22.7)$$

Понятно, что вычисленная таким способом полная производная от скорости $\mathbf{v}(\mathbf{r}, t)$ в эйлеровых переменных полностью совпадает с «траекторным» ускорением частицы в его кинематическом определении по Лагранжу, а именно: $\mathbf{a} \equiv d\mathbf{v}/dt = d^2\mathbf{r}/dt^2$.

Таким образом, левая часть формулы (1.22.7) содержит в рамках *лагранжева описания* производную по времени для данной ФБМ-частицы, движущейся вдоль траектории, которая называется индивидуальной или **субстанциональной (материальной) производной**. Правая часть этой формулы дает правило вычисления такой субстанциональной производной в эйлеровом описании, когда зависящие только от времени «траекторные» величины ($\mathbf{v}(\mathbf{r}(t))$ и другие, ей подобные) представлены в форме «полевых» величин ($\mathbf{v}(\mathbf{r}, t)$ и т. п.), зависящих от эйлеровых переменных $\mathbf{r}$ и t . Частная производная по времени в правой части (1.22.7) характеризует изменение скорости, происходящее за единицу времени в фиксированной точке пространства (но не для конкретной ФБМ-частицы!). Поэтому она называется местной или **локальной производной**, в то время как добавка к ней (типа $\mathbf{v} \cdot \nabla \dots$), учитывающая «траекторный» вклад в эйлерово описание, носит название **конвективной производной**.

3. Общая форма уравнений баланса. Применим полученные соотношения для анализа скорости изменения *объемной плотности* $w(t)$ произвольного экстенсивного свойства $\Delta W(t)$, запасаенного внутри выбранного ФБМ-объема $\Delta V(t)$ в момент времени t (см. рис. 1.11). Плотность величины W определена как

$$w(t) = \frac{\Delta W(t)}{\Delta V(t)}. \quad (1.22.8)$$

Согласно определению ФБМ-частицы, она сохраняет внутри себя все изначально присущие ей реальные частицы, поэтому ее объем $\Delta V(t)$ изменяется во времени вдоль траектории частицы.

Экстенсивное свойство $\Delta W(t)$, присущее выбранной ФБМ-частице, также изменяется во времени по двум причинам.

Во-первых, возможен отток/приток данного свойства через поверхность ΔA , ограничивающую ФБМ-объем ΔV . Это характеризуется **плотностью потока** $\mathbf{J}_w$ в направлении нормали $\mathbf{n}$ к поверхности ΔA ($\mathbf{n} \cdot \mathbf{J}_w$ — количество физического свойства W , *вытекающего* через единичную площадку за единицу времени, см. рис. 1.11 с единичной *внешней нормалью* $\mathbf{n}$).

Во-вторых, процессы, протекающие внутри ФБМ-объема ΔV , дополнительно порождают физическое свойство W . Это характеризуется **плотностью производства** σ_w (количество свойства W , возникающего/исчезающего, соответственно, при $\sigma_w \geq 0$ в единице объема за единицу времени, см. рис. 1.11).

Дифференцируем выражение (1.22.8) в лагранжевом представлении, то есть вдоль траектории движения ФБМ-частицы:

$$\frac{dw(t)}{dt} = \Delta W_c \frac{d}{dt} \left(\frac{1}{\Delta V(t)} \right) + \frac{1}{\Delta V_c} \frac{d}{dt} \Delta W(t), \quad (1.22.9)$$

где нижний индекс c отражает постоянство соответствующей величины при дифференцировании по t .

Первое слагаемое в правой части (1.22.9) учитывает вклад в изменение объемной плотности $w(t)$ за счет перемещения границы $\Delta A(t)$ ФБМ-объема $\Delta V(t)$, движущейся со скоростью $v_n(t) = \mathbf{n} \cdot \mathbf{v}(t)$ в направлении внешней нормали $\mathbf{n}$. Тогда приращение dV элемента объема за конечное время Δt для элементарной площадки dA равно $dV = v_n(t) dA \Delta t$. Для всего объема $\Delta V(t)$ за то же время имеем приращение

$$\frac{d\Delta V(t)}{dt} \Delta t = \int_{\Delta V(t)} dV = \oint_{\Delta A(t)} v_n(t) dA \Delta t.$$

Применяя теорему Остроградского–Гаусса к последнему интегралу в предыдущем выражении, получаем

$$\frac{d\Delta V}{dt} = \oint_{\Delta A} (\mathbf{n} \cdot \mathbf{v}) dA = \int_{\Delta V} (\nabla \cdot \mathbf{v}) dV = \Delta V \operatorname{div} \mathbf{v},$$

где в последнем равенстве учтена малость объема ΔV , в пределах которого $\operatorname{div} \mathbf{v}$ можно считать практически неизменной. Следовательно,

$$\frac{1}{\Delta V} \frac{d\Delta V}{dt} = \operatorname{div} \mathbf{v}. \quad (1.22.10)$$

С учетом (1.22.8) и (1.22.10), первое слагаемое в правой части (1.22.9) принимает вид

$$\Delta W_c \frac{d}{dt} \left(\frac{1}{\Delta V(t)} \right) = - \frac{\Delta W_c}{\Delta V} \left(\frac{1}{\Delta V} \frac{d\Delta V}{dt} \right) = -w \operatorname{div} \mathbf{v}. \quad (1.22.11)$$

Второе слагаемое в правой части (1.22.9) вызвано вышеуказанными процессами внутри ФБМ-объема ($\Delta V(t) = \Delta V_c = \text{const}$), приводящими к изменению $\Delta W(t)$ за счет наличия потока $\mathbf{J}_w$ через границу ($\Delta A(t) = \Delta A_c = \text{const}$) и производства σ_w внутри ΔV . На этом основании записываем второе слагаемое в (1.22.9) (с учетом «физической малости» объема ΔV_c) в виде

$$\begin{aligned} \frac{1}{\Delta V_c} \frac{d}{dt} \Delta W(t) &= - \frac{1}{\Delta V_c} \oint_{\Delta A_c} (\mathbf{n} \cdot \mathbf{J}_w) dA + \frac{1}{\Delta V_c} \int_{\Delta V_c} \sigma_w dV = \\ &= \frac{1}{\Delta V_c} \int_{\Delta V_c} (-\operatorname{div} \mathbf{J}_w + \sigma_w) dV = -\operatorname{div} \mathbf{J}_w + \sigma_w. \end{aligned} \quad (1.22.12)$$

Подстановка (1.22.11) и (1.22.12) в формулу (1.22.9) приводит к первой форме записи *уравнения баланса* для плотности w :

$$\frac{dw}{dt} + w \operatorname{div} \mathbf{v} + \operatorname{div} \mathbf{J}_w = \sigma_w. \quad (1.22.13)$$

Уравнение баланса в форме (1.22.13) содержит полную производную по времени от «траекторной» величины $w(t)$ в лагранжевом представлении. Рассматривая эту величину как поле $w(\mathbf{r}, t)$, можно, по аналогии с формулой (1.22.7), записать полную производную в эйлеровых переменных в виде

$$\frac{dw}{dt} = \frac{\partial w}{\partial t} + \mathbf{v} \cdot \nabla w. \quad (1.22.14)$$

Подставляя (1.22.14) в уравнение (1.22.13), получаем другую форму записи *уравнения баланса* для произвольной «полевой» величины $w(\mathbf{r}, t)$:

$$\frac{\partial w}{\partial t} + \operatorname{div} (w\mathbf{v} + \mathbf{J}_w) = \sigma_w. \quad (1.22.15)$$

Здесь под знаком дивергенции стоят два слагаемых, имеющие смысл потока на единицу площади. Первое из них трактуется как **конвективный поток** величины w , переносимой *упорядоченным движением* частиц среды с гидродинамической скоростью $\mathbf{v}$ ($\mathbf{J}_w^{\text{кон}} \equiv w\mathbf{v}$). Второе слагаемое $\mathbf{J}_w$ учитывает перенос величины w за счет *хаотического движения* частиц по отношению к системе отсчета, движущейся со скоростью $\mathbf{v}$. Это слагаемое часто называют **диффузионным** (или **кондуктивным**) **потоком**

($\mathbf{J}_w^{\text{диф}} \equiv \mathbf{J}_w$). Тогда суммарная плотность потока величины w , стоящая под знаком дивергенции в уравнении (1.22.15), равна

$$\mathbf{J}_w^{\Sigma} = \mathbf{J}_w^{\text{кон}} + \mathbf{J}_w^{\text{диф}} \equiv w\mathbf{v} + \mathbf{J}_w. \quad (1.22.16)$$

В правой части уравнений баланса (1.22.13) и (1.22.15) стоит **локальное производство** величины w , которое учитывает источники (при $\sigma_w > 0$) или стоки (при $\sigma_w < 0$) этой величины внутри рассматриваемой системы. В случае $\sigma_w = 0$ уравнение баланса (1.22.15) превращается в уравнение сохранения величины w .

4. Уравнение баланса числа частиц i -го компонента. Применим уравнение баланса в форме (1.22.13) к молярной плотности c_i компонента i , делая замены $w \rightarrow c_i$, $\mathbf{J}_w \rightarrow \mathbf{J}_i$ и $\sigma_w \rightarrow \sigma_i$:

$$\frac{dc_i}{dt} + c_i \operatorname{div} \mathbf{v} + \operatorname{div} \mathbf{J}_i = \sigma_i. \quad (1.22.17)$$

Здесь $\mathbf{J}_i$ — плотность диффузионного потока (моль/(м²·с)) для i -го компонента ($\mathbf{J}_i^{\text{диф}} \equiv \mathbf{J}_i$).

По аналогии с (1.22.16), запишем полную плотность потока i -го компонента:

$$\mathbf{J}_i^{\Sigma} = \mathbf{J}_i^{\text{кон}} + \mathbf{J}_i^{\text{диф}} \equiv c_i \mathbf{v} + \mathbf{J}_i. \quad (1.22.18)$$

Сумма потоков всех компонентов дает общий конвективный поток, определяемый равенством (1.22.2),

$$\mathbf{J} \equiv c\mathbf{v} = \sum_i c_i \mathbf{v}_i \equiv \sum_i \mathbf{J}_i^{\Sigma}. \quad (1.22.19)$$

Подстановкой (1.22.18) в правую часть равенства (1.22.19) получаем

$$\sum_{i=1}^K \mathbf{J}_i \equiv \sum_{i=1}^K c_i (\mathbf{v}_i - \mathbf{v}) = 0. \quad (1.22.20)$$

Следовательно, по отношению к общей гидродинамической (молярной) скорости $\mathbf{v}$, сумма диффузионных (кондуктивных) потоков равна нулю.

Локальное производство σ_i (моль/(м³·с)) i -го компонента определяется всеми химическими реакциями с участием этого компонента. Согласно (1.18.11), скорость изменения концентрации c_i в реакции с номером r равна $v_i \equiv dc_i/dt = \nu_{ir} v_r$, где $v_r \equiv d\xi_r/dt$ — скорость r -й реакции (моль/(м³·с)).

Суммируя по всем реакциям числом X ($r = 1, 2, \dots, X$), получаем искомую плотность производства молей i -го сорта:

$$\sigma_i = \sum_{r=1}^X \nu_{ir} v_r, \quad i = 1, 2, \dots, K. \quad (1.22.21)$$

Из выражений (1.22.17) и (1.22.21) получаем результирующую скорость изменения мольно-объемной концентрации c_i для выбранной ФБМ-частицы:

$$\frac{dc_i}{dt} = -c_i \operatorname{div} \mathbf{v} - \operatorname{div} \mathbf{J}_i + \sum_r \nu_{ir} v_r. \quad (1.22.22)$$

5. Уравнение баланса внутренней энергии. Применим уравнение баланса в форме (1.22.13) к объемной плотности u внутренней энергии, делая замены $w \rightarrow u$, $\mathbf{J}_w \rightarrow \mathbf{J}_u$ и $\sigma_w \rightarrow \sigma_u$:

$$\frac{du}{dt} + u \operatorname{div} \mathbf{v} + \operatorname{div} \mathbf{J}_u = \sigma_u. \quad (1.22.23)$$

Здесь $\mathbf{J}_u$ — плотность потока внутренней энергии (Вт/м^2), переносимой в системе отсчета, движущейся с общей гидродинамической скоростью $\mathbf{v}$. Ниже $\mathbf{J}_u$ будет связана с плотностью теплового потока $\mathbf{J}_q$, вызванного хаотическим движением частиц (см. формулу (1.23.14)).

Как известно, энергия в объеме ФБМ-частицы выделяется за счет внутренних напряжений (описываемых тензором напряжений $\overline{\mathbf{T}}$) только при неоднородности скорости $\mathbf{v}$, характеризующей ее градиентом (1.22.3). В этом случае локальное производство внутренней энергии (Вт/м^3) определяется выражением

$$\sigma_u = \overline{\mathbf{T}} : \nabla \mathbf{v} \equiv -\overline{\mathbf{P}} : \nabla \mathbf{v}, \quad (1.22.24)$$

где введен обычно используемый тензор давления $\overline{\mathbf{P}} \equiv -\overline{\mathbf{T}}$; двойное скалярное произведение означает $\overline{\mathbf{P}} : \nabla \mathbf{v} = P_{kl} (\partial v_l / \partial x_k)$.

Из выражений (1.22.23) и (1.22.24) получаем результирующую скорость изменения плотности энергии u для выбранной ФБМ-частицы:

$$\frac{du}{dt} = -u \operatorname{div} \mathbf{v} - \operatorname{div} \mathbf{J}_u - \overline{\mathbf{P}} : \nabla \mathbf{v}. \quad (1.22.25)$$

Уравнения (1.22.22) и (1.22.25) будут использованы при выводе выражений для плотностей потока энтропии и производства энтропии ($\mathbf{J}_s$ и σ_s), входящих в уравнение баланса энтропии.

6. Уравнение баланса энтропии. Применим общее уравнение баланса в форме (1.22.13) к объемной плотности s энтропии, делая замены $w \rightarrow s$, $\mathbf{J}_w \rightarrow \mathbf{J}_s$ и $\sigma_w \rightarrow \sigma_s$:

$$\frac{ds}{dt} + s \operatorname{div} \mathbf{v} + \operatorname{div} \mathbf{J}_s = \sigma_s, \quad (1.22.26)$$

где $\mathbf{J}_s$ и σ_s — плотность потока энтропии и локальное производство энтропии, которые в дальнейшем необходимо найти.

Покажем, что локальное уравнение баланса энтропии в форме (1.22.26) приводит к фундаментальному соотношению (1.7.1)

для дифференциала dS полной энтропии системы как суммы вкладов от внешней среды ($d_e S$) и от внутренних неравновесных процессов в системе ($d_i S$).

Для этого вычисляем полную (лагранжеву) производную по времени от энтропии в элементарном объеме ΔV , движущемся вдоль траектории (см. рис. 1.11). Считаем, что от времени зависит как плотность энтропии $s(t)$, так и объем $\Delta V(t)$:

$$\begin{aligned} \frac{dS(t)}{dt} &= \frac{d}{dt} \int_{\Delta V(t)} s(t) dV = \frac{d}{dt} \left[\int_{\Delta V_c} s(t) dV + \int_{\Delta V(t)} s_c dV \right] = \\ &= \int_{\Delta V_c} \frac{ds(t)}{dt} dV + s(t) \frac{d\Delta V(t)}{dt} = \int_{\Delta V_c} \frac{ds}{dt} dV + (s \operatorname{div} \mathbf{v}) \Delta V = \\ &= \int_{\Delta V} \frac{ds}{dt} dV + \int_{\Delta V} s \operatorname{div} \mathbf{v} dV = \int_{\Delta V} \left(\frac{ds}{dt} + s \operatorname{div} \mathbf{v} \right) dV. \end{aligned} \quad (1.22.27)$$

Здесь, как и ранее в равенстве (1.22.9), нижний индекс c отражает постоянство соответствующей величины при дифференцировании по t . При этом использованы соотношения (1.22.10), связывающее $\operatorname{div} \mathbf{v}$ с относительным изменением объема ΔV , и условие «физической малости» этого объема.

В последнем интеграле (1.22.27) стоят первые два члена левой части уравнения (1.22.26). Поэтому интегрирование этого уравнению по объему ΔV с учетом теоремы Остроградского–Гаусса дает

$$\frac{dS}{dt} = - \oint_{\Delta A} (\mathbf{n} \cdot \mathbf{J}_s) dA + \int_{\Delta V} \sigma_s dV. \quad (1.22.28)$$

Поскольку $-\mathbf{n}$ есть *внутренняя* нормаль к поверхности ΔA объема ΔV , то первый интеграл в (1.22.28) интерпретируем как поток энтропии, *втекающей извне* внутрь объема ΔV , а второй — как производство энтропии *внутри* этого объема. На этом основании можем записать (ср. уравнение (1.7.1))

$$\frac{dS}{dt} = \frac{d_e S}{dt} + \frac{d_i S}{dt}, \quad (1.22.29)$$

где для элементарного объема ΔV имеем внешний вклад,

$$\frac{d_e S}{dt} = - \oint_{\Delta A} (\mathbf{n} \cdot \mathbf{J}_s) dA, \quad (1.22.30)$$

и внутреннее производство энтропии, ранее определенное как

$$\mathcal{P} \equiv \frac{d_i S}{dt} = \int_{\Delta V} \sigma_s dV. \quad (1.22.31)$$

Теперь необходимо получить выражения для плотности потока энтропии ($\mathbf{J}_s$) и локального производства энтропии (σ_s), используя уравнение баланса энтропии (1.22.26) и фундаментальное уравнение Гиббса в форме (1.21.11).

1.23. Плотность потока энтропии и локальное производство энтропии

1. Вывод выражений для плотности потока энтропии и локального производства энтропии. Перепишем уравнение Гиббса (1.21.11) для плотностей величин в следующем виде:

$$T \frac{ds}{dt} = \frac{du}{dt} - \sum_i \mu_i \frac{dc_i}{dt}. \quad (1.23.1)$$

Подстановкой уравнений (1.22.22) и (1.22.25) в формулу (1.23.1) надо привести ее к требуемому балансом энтропии виду в форме (1.22.15) с заменами $w \rightarrow s$, $\mathbf{J}_w \rightarrow \mathbf{J}_s$ и $\sigma_w \rightarrow \sigma_s$:

$$\frac{\partial s}{\partial t} + \operatorname{div}(s\mathbf{v} + \mathbf{J}_s) = \sigma_s, \quad (1.23.2)$$

где $\mathbf{J}_s$ и σ_s — плотности потока энтропии и производства энтропии, подлежащие определению.

Указанная подстановка сначала приводит уравнение (1.23.1) к следующему виду:

$$\begin{aligned} \frac{ds}{dt} = & -\frac{1}{T} \left(u + P - \sum_i c_i \mu_i \right) \operatorname{div} \mathbf{v} - \frac{1}{T} \operatorname{div} \mathbf{J}_u - \frac{1}{T} \bar{\pi} : \nabla \mathbf{v} + \\ & + \frac{1}{T} \sum_i \mu_i \operatorname{div} \mathbf{J}_i + \frac{1}{T} \sum_r \mathcal{A}_r v_r. \end{aligned} \quad (1.23.3)$$

Здесь, во-первых, в соответствии с формулой (1.14.7) введено *химическое сродство* для реакции типа (1.14.4) с номером r :

$$\mathcal{A}_r = - \sum_{i=1}^K \nu_{ir} \mu_i, \quad \text{где } \nu_{ir} = \begin{cases} -\nu_{ir} < 0, \\ \nu_{ir} > 0, \end{cases} \quad (1.23.4)$$

при этом стехиометрические коэффициенты для начальных реагентов (стоящих в реакции слева) берутся со знаком минус, а для конечных продуктов (стоящих справа) — со знаком плюс.

Во-вторых, в уравнении (1.23.3) тензор давления $\bar{\mathbf{P}}$, появившийся из (1.22.25), представлен в виде суммы изотропного гидростатического давления P и тензора вязкого давления $\bar{\pi}$,

$$\bar{\mathbf{P}} = P\bar{\mathbf{I}} + \bar{\pi} \equiv P\delta_{kl} + \pi_{kl}, \quad (1.23.5)$$

так что, с учетом (1.22.4), двойное скалярное произведение равно

$$\bar{\mathbf{P}} : \nabla \mathbf{v} = P \operatorname{div} \mathbf{v} + \bar{\pi} : \nabla \mathbf{v}. \quad (1.23.6)$$

В соответствии с равенством (1.21.7), скобка в первом слагаемом уравнения (1.23.3) дает Ts . Тогда, используя тождество $\varphi \operatorname{div} \mathbf{A} = \operatorname{div}(\varphi \mathbf{A}) - \mathbf{A} \cdot \nabla \varphi$ и записывая, по аналогии с (1.22.7), полную производную от плотности энтропии по времени в виде

$$\frac{ds}{dt} = \frac{\partial s}{\partial t} + \mathbf{v} \cdot \nabla s, \quad (1.23.7)$$

приводим уравнение (1.23.3) к искомой форме (1.23.2):

$$\begin{aligned} \frac{\partial s}{\partial t} + \operatorname{div} \left[s\mathbf{v} + \frac{1}{T} \left(\mathbf{J}_u - \sum_i \mu_i \mathbf{J}_i \right) \right] = \\ = \mathbf{J}_u \cdot \nabla \frac{1}{T} - \sum_i \mathbf{J}_i \cdot \nabla \frac{\mu_i}{T} + \frac{1}{T} \sum_r \mathcal{A}_r v_r - \frac{1}{T} \bar{\pi} : \nabla \mathbf{v}. \end{aligned} \quad (1.23.8)$$

Из сравнения уравнения (1.23.8) с требуемой формой (1.23.2) для уравнения баланса энтропии немедленно получаем:

- плотность потока энтропии (Вт/(м²·К))

$$\mathbf{J}_s = \frac{1}{T} \left(\mathbf{J}_u - \sum_i \mu_i \mathbf{J}_i \right); \quad (1.23.9)$$

- локальное производство энтропии (Вт/(м³·К))

$$\begin{aligned} \sigma_s = \mathbf{J}_u \cdot \nabla \frac{1}{T} - \sum_i \mathbf{J}_i \cdot \nabla \frac{\mu_i}{T} + \frac{1}{T} \sum_r \mathcal{A}_r v_r - \frac{1}{T} \bar{\pi} : \nabla \mathbf{v} \equiv \\ \equiv \left(\mathbf{J}_u - \sum_i \mu_i \mathbf{J}_i \right) \cdot \nabla \frac{1}{T} - \frac{1}{T} \sum_i \mathbf{J}_i \cdot \nabla \mu_i + \\ + \frac{1}{T} \sum_r \mathcal{A}_r v_r - \frac{1}{T} \bar{\pi} : \nabla \mathbf{v}. \end{aligned} \quad (1.23.10)$$

Последнее выражение для σ_s не разделяет полностью вклады от градиента температуры и от градиентов концентрации компонентов, поскольку $\nabla \mu_i$ является функцией T и c_i , что подтверждается выражением (1.21.29) для полного дифференциала $d\mu_i^c$

химического потенциала i -го компонента. Так как $d\mu_i = d\mathbf{r} \cdot \nabla \mu_i$ для произвольного $d\mathbf{r}$, то на основании (1.21.29) можно записать

$$\nabla \mu_i = (\nabla \mu_i^c)_T - s_i dT. \quad (1.23.11)$$

Здесь верхний индекс c отражает зависимость химического потенциала от молярных плотностей c_i , а $s_i = (\partial s / \partial c_i)_{T,P}$ есть парциальная молярная энтропия.

Подстановка (1.23.11) в последнее выражение в равенстве (1.23.10) с учетом того, что согласно (1.21.23) $u_i = \mu_i + T s_i$, дает **локальное производство энтропии** в искомой форме записи:

$$\begin{aligned} \sigma_s = & \mathbf{J}_q \cdot \nabla \frac{1}{T} - \frac{1}{T} \sum_i \mathbf{J}_i \cdot (\nabla \mu_i^c)_T + \\ & + \frac{1}{T} \sum_r \mathcal{A}_r v_r - \frac{1}{T} \bar{\pi} : \nabla \mathbf{v}, \end{aligned} \quad (1.23.12)$$

где введена *плотность теплового потока*

$$\mathbf{J}_q = \mathbf{J}_u - \sum_i u_i \mathbf{J}_i. \quad (1.23.13)$$

Подстановка в формулу (1.23.9) выражения для *плотности потока внутренней энергии* в форме

$$\mathbf{J}_u = \mathbf{J}_q + \sum_i u_i \mathbf{J}_i, \quad (1.23.14)$$

приводит **плотность потока энтропии** к следующему виду:

$$\mathbf{J}_s = \frac{1}{T} \mathbf{J}_q + \sum_i s_i \mathbf{J}_i. \quad (1.23.15)$$

Таким образом, в перенос энергии (формула (1.23.14)) и энтропии (формула (1.23.15)), кроме теплового потока $\mathbf{J}_q$ вносят вклад также потоки частиц $\mathbf{J}_i$, которые переносят парциальную молярную энергию $u_i = (\partial u / \partial c_i)_{T,P}$ и парциальную молярную энтропию $s_i = (\partial s / \partial c_i)_{T,P}$; при этом $u_i \equiv h_i = \mu_i + T s_i$, что следует из формулы (1.21.23).

Следует отметить, что при использовании удельных величин (на единицу массы системы) для i -го компонента вместо молярной плотности $c_i = n_i/V$ работает массовая плотность $\rho_i = c_i M_i$ (где M_i — молярная масса, иногда называемая *молекулярным весом*, г/моль). В этом случае полученные выражения для плотности потока энтропии $\mathbf{J}_s$ и плотности производства энтропии σ_s

сохраняют форму записи при выполнении следующих замен:

$$\mathbf{J}_i \left(\frac{\text{моль}}{\text{м}^2 \cdot \text{с}} \right) \rightarrow \mathbf{J}_i M_i \left(\frac{\text{кг}}{\text{м}^2 \cdot \text{с}} \right), \quad \mu_i \left(\frac{\text{Дж}}{\text{моль}} \right) \rightarrow \frac{\mu_i}{M_i} \left(\frac{\text{Дж}}{\text{кг}} \right),$$

$$v_r \left(\frac{\text{моль}}{\text{м}^3 \cdot \text{с}} \right) \rightarrow v_r M_i \left(\frac{\text{кг}}{\text{м}^3 \cdot \text{с}} \right), \quad \mathcal{A}_r \left(\frac{\text{Дж}}{\text{моль}} \right) \rightarrow \frac{\mathcal{A}_r}{M_i} \left(\frac{\text{Дж}}{\text{кг}} \right).$$

В обоих случаях (для плотностей и для удельных величин) соответствующие произведения имеют одинаковую размерность:

$$\mu_i \mathbf{J}_i \left(\frac{\text{Вт}}{\text{м}^2} \right) \quad \text{и} \quad \mathcal{A}_r v_r \left(\frac{\text{Вт}}{\text{м}^3} \right).$$

2. Принцип симметрии Кюри. Плотность производства энтропии (создаваемой в единице объема за единицу времени) в форме выражения (1.23.12), называемая также *локальным производством энтропии*, содержит четыре вклада разного тензорного характера (нулевого, первого и второго рангов):

- «скалярный» вклад, порожденный *химическими реакциями*,

$$\sigma_s^{\text{хим}} = \sum_r J_r X_r, \quad \text{где } J_r = v_r, \quad X_r = \mathcal{A}_r / T; \quad (1.23.16)$$

- «векторный» вклад, порожденный *процессами диффузии*,

$$\sigma_s^{\text{диф}} = \sum_i \mathbf{J}_i \cdot \mathbf{X}_i, \quad \text{где } \mathbf{J}_i \equiv \mathbf{J}_i, \quad \mathbf{X}_i = -(\nabla \mu_i^c)_T / T; \quad (1.23.17)$$

- «векторный» вклад, порожденный *теплопроводностью*,

$$\sigma_s^{\text{теп}} = \mathbf{J} \cdot \mathbf{X}, \quad \text{где } \mathbf{J} = \mathbf{J}_q, \quad \mathbf{X} = \nabla(1/T); \quad (1.23.18)$$

- «тензорный» вклад, порожденный *вязкостью среды*,

$$\sigma_s^{\text{вяз}} = \bar{\mathbf{J}} : \bar{\mathbf{X}}, \quad \text{где } \bar{\mathbf{J}} = \bar{\boldsymbol{\pi}}, \quad \bar{\mathbf{X}} = -\nabla \mathbf{v} / T. \quad (1.23.19)$$

Произведения компонентов векторов (скалярное) и компонентов тензоров (двойное скалярное) в равенствах (1.23.17)–(1.23.19) позволяют единым образом записать все четыре вклада в скалярной форме в виде суммы произведений соответствующих компонентов термодинамических потоков J_α и сил X_α :

$$\sigma_s = \sum_\alpha J_\alpha X_\alpha. \quad (1.23.20)$$

Запись локального производства энтропии σ_s в форме (1.23.20) совпадает с ранее использованным выражением (1.7.6) или (1.9.19) для полного производства энтропии $\mathcal{P}$. Для согласования размерностей следует считать, что эти выражения для $\mathcal{P}$ ранее

были записаны в применении к единичному объему, поскольку в локально квазиравновесных условиях полное производство энтропии определено в виде (см. уравнение (1.22.31))

$$\mathcal{P} \equiv \frac{d_i S}{dt} = \int_V \sigma_s dV = \sum_{\alpha} \int_V J_{\alpha} X_{\alpha} dV. \quad (1.23.21)$$

Согласно теории Онсагера, поток J_{α} линейно выражается через термодинамические силы X_{β} в форме соотношения (1.19.11):

$$J_{\alpha} = \sum_{\beta} L_{\alpha\beta} X_{\beta}, \quad (1.23.22)$$

с кинетическими коэффициентами, подчиняющимися требованию симметрии: $L_{\alpha\beta} = L_{\beta\alpha}$.

Поскольку коэффициенты Онсагера являются скалярными величинами, то они могут связывать между собой необратимые процессы только одинакового тензорного характера (ранга): скалярные со скалярными (нулевого ранга), векторные с векторными (первого ранга) и тензорные с тензорными (второго ранга). Именно это и составляет содержание так называемого *принципа симметрии Кюри* в формулировке Пригожина в применении к изотропным термодинамическим системам.

Следовательно, «скалярный» вклад от химических реакций (числом X), производящих энтропию в форме (1.23.16),

$$\sigma_s^{\text{хим}} = \sum_{r=1}^X v_r \frac{A_r}{T}, \quad (1.23.23)$$

не может вызывать «векторные» вклады, порожденные теплопроводностью и диффузией компонентов (числом K), производящими суммарную энтропию в форме (1.23.17) и (1.23.18),

$$\sigma_s^{\text{теп}} + \sigma_s^{\text{диф}} = \mathbf{J}_q \cdot \text{grad} \frac{1}{T} - \frac{1}{T} \sum_{i=1}^{K-1} \mathbf{J}_i \cdot \text{grad}_T (\mu_i^c - \mu_K^c). \quad (1.23.24)$$

В последней сумме использовано равенство (1.22.20), согласно которому $\mathbf{J}_K = -\sum_{i=1}^{K-1} \mathbf{J}_i$, а градиент разности химических потенциалов вычисляется при постоянстве температуры, что отмечено нижним индексом T .

Таким образом, принцип симметрии Кюри накладывает ограничения на взаимосвязь необратимых процессов и производство энтропии, каждый вклад которых (скалярный, векторный, тензорный) должен быть положительным в соответствии со вторым

началом термодинамики (1.7.5). Взаимное влияние (перекрестный эффект) могут оказывать друг на друга (исключая вязкость):

- а) одновременно протекающие химические реакции (скалярный эффект);
- б) изотермическая диффузия нескольких компонентов (векторный эффект);
- в) диффузия и теплопроводность (векторный эффект).

На основании выражений (1.23.16)–(1.23.18), (1.23.23) и (1.23.24) можно записать для трех вышеуказанных случаев феноменологические уравнения типа (1.23.22), связывающие между собой термодинамические потоки и силы через коэффициенты Онсагера (опуская из рассмотрения вязкостные эффекты):

- для химических реакций ($r = 1, 2, \dots, X$; $L_{rr'} = L_{r'r}$)

$$v_r = \sum_{r'=1}^X L_{rr'} \frac{A_{r'}}{T} \equiv - \sum_{r'=1}^X L_{rr'} \sum_{i=1}^K v_{ir'} \frac{\mu_i}{T}; \quad (1.23.25)$$

- для процессов диффузии ($i = 1, 2, \dots, K$; $L_{ij} = L_{ji}$)

$$\mathbf{J}_i = L_{iq} \text{grad} \frac{1}{T} - \frac{1}{T} \sum_{j=1}^{K-1} L_{ij} \text{grad}_T (\mu_j^c - \mu_K^c); \quad (1.23.26)$$

- для процессов теплопроводности ($L_{qi} = L_{iq}$)

$$\mathbf{J}_q = L_{qq} \text{grad} \frac{1}{T} - \frac{1}{T} \sum_{i=1}^{K-1} L_{qi} \text{grad}_T (\mu_i^c - \mu_K^c). \quad (1.23.27)$$

3. Применение коэффициентов Онсагера к конкретным ситуациям. Установим связь коэффициентов Онсагера, входящих в равенства (1.23.25)–(1.23.27), с экспериментально измеряемыми величинами, такими как скорость химической реакции, коэффициенты диффузии и теплопроводности.

Химические реакции. Скорость химической реакции с номером r была определена формулой (1.18.10) как положительная величина для прямой реакции (протекающей слева направо), отмечаемой теперь верхним индексом $^+$ ($v_r^+(t) > 0$). Для той же реакции, протекающей в обратном направлении (справа налево), скорость принимается отрицательной ($v_r^-(t) < 0$). Поэтому результирующую скорость двусторонней обратимой реакции типа (1.18.6) записываем в виде

$$v_r(t) = v_r^+(t) + v_r^-(t) = v_r^+(t) - |v_r^-(t)|. \quad (1.23.28)$$

Используя основной постулат химической кинетики, запишем скорости прямой и обратной r -й реакции в форме (1.18.13) со своими кинетическими константами $k_r^+(T)$ и $k_r^-(T)$. Тогда результирующая скорость (1.23.28) принимает следующий вид:

$$\begin{aligned} v_r(t) &= v_r^+(t) \left(1 - \frac{k_r^-(T)}{k_r^+(T)} \frac{\prod_{\text{к}} c_{\text{к}}^{\nu_{\text{к}}}(t)}{\prod_{\text{н}} c_{\text{н}}^{\nu_{\text{н}}}(t)} \right) = \\ &= v_r^+(t) \left(1 - K_{cr}^{-1}(T) \prod_i c_i^{\nu_{ir}}(t) \right). \end{aligned} \quad (1.23.29)$$

В последнем равенстве (1.23.29), во-первых, отношение констант скорости заменено единой константой (ср. ф-лу (4.1.5)),

$$K_{cr}(T) = \frac{k_r^+(T)}{k_r^-(T)}, \quad (1.23.30)$$

называемой **константой равновесия** (см. пп. 2.12 и 4.1), и, во-вторых, для начальных и конечных компонентов использована единая нумерация Де Донде (см. формулы (1.14.7) и (1.23.4)).

Появившееся в уравнении (1.23.29) единое произведение по i необходимо связать с химическим сродством $\mathcal{A}_r$ для r -й реакции, введенным по формуле (1.23.4). Для этого используем выражение (1.13.2) для химического потенциала i -го компонента, подстановка которого в (1.23.4) дает

$$\begin{aligned} \mathcal{A}_r &\equiv - \sum_i \nu_{ir} \mu_i = - \sum_i \nu_{ir} \mu_i^\circ(T) - RT \sum_i \nu_{ir} \ln c_i = \\ &= RT \ln \left(K_{cr}(T) \prod_i c_i^{-\nu_{ir}} \right). \end{aligned} \quad (1.23.31)$$

Здесь константа равновесия (1.23.30) введена с помощью уравнения (ср. формулы (2.13.2) и (2.13.3))

$$RT \ln K_{cr}(T) = - \sum_i \nu_{ir} \mu_i^\circ(T). \quad (1.23.32)$$

Из сравнения формул (1.23.29) и (1.23.31) видно, что для химического сродства всегда справедливо соотношение

$$\mathcal{A}_r = RT \ln \frac{v_r^+}{|v_r^-|}, \quad (1.23.33)$$

с помощью которого результирующую скорость реакции (1.23.29) записываем в виде

$$v_r = v_r^+ \left(1 - e^{-\mathcal{A}_r/RT} \right). \quad (1.23.34)$$

Таким образом, в общем случае соотношение между скоростью реакции и ее химическим сродством является нелинейным.

Применение соотношения (1.23.34) к теории Онсагера возможно лишь вблизи равновесия, когда $|\mathcal{A}_r| \ll RT$, так что в линейном приближении

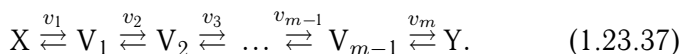
$$v_r = v_r^0 \frac{\mathcal{A}_r}{RT}, \quad (1.23.35)$$

где v_r^0 — «равновесная» скорость, общая как для прямой, так и для обратной реакции на каждой элементарной стадии ($v_r^+ = |v_r^-|$ при $\mathcal{A}_r = 0$), как следствие принципа детального равновесия.

Из сравнения выражения (1.23.35) с феноменологическим соотношением (1.23.25), записанным для одиночной реакции как $v_r = L_{rr}(\mathcal{A}_r/T)$, получаем коэффициент Онсагера:

$$L_{rr} = \frac{v_r^0}{R}, \quad \text{так что} \quad \sigma_s^{\text{хим}} = \frac{v_r^0}{R} \left(\frac{\mathcal{A}_r}{T} \right)^2 > 0. \quad (1.23.36)$$

Рассмотрим теперь реакцию типа $X \rightleftharpoons Y$, которая протекает в несколько последовательных стадий с промежуточными продуктами $V_1, V_2, \dots, V_{m-1}$:



Изменение концентраций всех участвующих реагентов описывается кинетическими уравнениями:

$$\frac{dc_X}{dt} = -v_1, \quad \frac{dc_{V_1}}{dt} = v_1 - v_2, \quad \frac{dc_{V_2}}{dt} = v_2 - v_3, \quad \dots \quad \frac{dc_Y}{dt} = v_m.$$

Пусть из m последовательных стадий процесса последняя реакция $V_{m-1} \rightleftharpoons Y$ является наиболее медленной или, как говорят, *лимитирующей*, поскольку именно она определяет результирующую скорость процесса v . Следовательно, промежуточные продукты, возникая за очень короткие времена, имеют практически постоянные концентрации ($dc_{V_r}/dt = 0$) и

$$v_1 = \dots = v_r = \dots = v_m \equiv v. \quad (1.23.38)$$

Тогда локальное производство энтропии для последовательности из m химических реакций принимает вид

$$\sigma_s^{\text{хим}} = \frac{1}{T} (v_1 \mathcal{A}_1 + \dots + v_r \mathcal{A}_r + \dots + v_m \mathcal{A}_m) = \frac{1}{T} v \mathcal{A}, \quad (1.23.39)$$

где $\mathcal{A} = \sum_{r=1}^m \mathcal{A}_r$ — химическое сродство результирующего процесса $X \rightleftharpoons Y$.

Если для каждой из m реакций $|\mathcal{A}_r|/RT \ll 1$, т. е. выполняется условие применимости линейных феноменологических законов, то по аналогии с (1.23.35) записываем:

$$v_1 = v_1^0 \frac{\mathcal{A}_1}{RT}, \quad \dots \quad v_r = v_r^0 \frac{\mathcal{A}_r}{RT}, \quad \dots \quad v_m = v_m^0 \frac{\mathcal{A}_m}{RT}, \quad (1.23.40)$$

где v_r^0 — «равновесная» скорость прямой r -й реакции в последовательности стадий (1.23.37).

Используя формулы (1.23.38) и (1.23.40), получаем полное химическое сродство результирующего процесса $X \rightleftharpoons Y$:

$$\mathcal{A} = \sum_{r=1}^m \mathcal{A}_r = \sum_{r=1}^m v_r \frac{RT}{v_r^0} = v \sum_{r=1}^m \frac{RT}{v_r^0}.$$

Отсюда следует искомое выражение для результирующей скорости многостадийного процесса:

$$v = v_{\text{эф}}^0 \frac{\mathcal{A}}{RT}, \quad (1.23.41)$$

где эффективная скорость $v_{\text{эф}}^0$ определена равенством

$$\frac{1}{v_{\text{эф}}^0} = \frac{1}{v_1^0} + \dots + \frac{1}{v_r^0} + \dots + \frac{1}{v_m^0}. \quad (1.23.42)$$

Таким образом, линейное соотношение (1.23.41) между скоростью v и суммарным сродством $\mathcal{A} = \sum_{r=1}^m \mathcal{A}_r$ многостадийной реакции (1.23.37) выполняется даже при $\mathcal{A} \gg RT$ только в том случае, если для каждой стадии $\mathcal{A}_r \ll RT$ и концентрации промежуточных продуктов не изменяются (что реализуется, когда последняя реакция $V_{m-1} \rightleftharpoons Y$ является лимитирующей).

Изотермическая диффузия. В многокомпонентной системе при диффузии компоненты оказывают взаимное влияние друг на друга. Поэтому плотность потока i -го компонента, согласно (1.23.26), в изотермических условиях феноменологически выражается через градиенты химического потенциала всех веществ, входящих в систему (включая самое себя), при помощи коэффициентов Онсагера L_{ij} :

$$\mathbf{J}_i = - \sum_{j=1}^K \frac{L_{ij}}{T} \text{grad}_T \mu_j^c = - \sum_{j=1}^{K-1} \frac{L_{ij}}{T} \text{grad}_T (\mu_j^c - \mu_K^c), \quad (1.23.43)$$

где последняя сумма записана с учетом (1.22.20).

С другой стороны, аналогичные выражения принято записывать в форме обобщенного закона Фика, содержащего бинарные

коэффициенты диффузии D_{ij} , измеряемые экспериментально:

$$\mathbf{J}_i = - \sum_{j=1}^K D_{ij} \text{grad } c_j. \quad (1.23.44)$$

Для установления связи между коэффициентами Онсагера и коэффициентами диффузии рассмотрим простейший случай двухкомпонентной системы ($K = 2$). Тогда, согласно соотношению Гиббса–Дюгема (1.21.24), при $T = \text{const}$ и $P = \text{const}$ имеем

$$c_1 d\mu_1^c + c_2 d\mu_2^c = 0, \quad \text{или} \quad c_1 (\nabla \mu_1^c)_T + c_2 (\nabla \mu_2^c)_T = 0. \quad (1.23.45)$$

Здесь второе соотношение между градиентами химических потенциалов следует из первого ввиду того, что $d\mu_i = d\mathbf{r} \cdot \nabla \mu_i$ для произвольного приращения $d\mathbf{r}$.

Полагая $K = 2$ и $i = 1$ в последнем равенстве (1.23.43) и используя (1.23.45), записываем (где $c = c_1 + c_2$):

$$\begin{aligned} \mathbf{J}_1 &= - \frac{L_{11}}{T} \left(1 + \frac{c_1}{c_2} \right) \text{grad}_T \mu_1^c = \\ &= - \frac{L_{11}}{T} \frac{c}{c_2} \frac{\partial \mu_1^c}{\partial c_1} \text{grad } c_1 \equiv -D_{11} \text{grad } c_1. \end{aligned} \quad (1.23.46)$$

Отсюда следует искомое соотношение между феноменологическим коэффициентом L_{11} и коэффициентом диффузии $D_1 \equiv D_{11}$:

$$L_{11} = D_1 T \frac{c_2}{c} \left(\frac{\partial \mu_1^c}{\partial c_1} \right)^{-1}. \quad (1.23.47)$$

Используем для химического потенциала μ_1^c растворенного вещества 1 в растворителе 2 выражение (1.13.2), а именно $\mu_1^c = \mu_1^0(T) + RT \ln c_1$, и переписываем (1.23.47) в следующем виде:

$$L_{11} = \frac{D_1}{R} \frac{c_1 c_2}{c_1 + c_2} \approx D_1 \frac{c_1}{R}, \quad (1.23.48)$$

где приближенное равенство справедливо для сильно разбавленного раствора, когда $c_1 \ll c_2$.

Неизотермическая диффузия (эффекты Соре и Дюфура).

Рассмотрим взаимное влияние (перекрестный эффект) процессов диффузии и теплопроводности, описываемых формулами (1.23.26) и (1.23.27). Для двухкомпонентной системы эти формулы, с учетом связи между градиентами химических потенциалов, даваемой равенством (1.23.45) при $K = 2$ и $i = 1$, имеют вид

$$\mathbf{J}_1 = L_{1q} \text{grad } \frac{1}{T} - L_{11} \frac{1}{T} \left(1 + \frac{c_1}{c_2} \right) \text{grad}_T \mu_1^c,$$

$$\mathbf{J}_q = L_{qq} \operatorname{grad} \frac{1}{T} - L_{q1} \frac{1}{T} \left(1 + \frac{c_1}{c_2} \right) \operatorname{grad}_T \mu_1^c.$$

Введем градиенты температуры T и концентрации c_1 компонента 1, тогда предыдущие формулы принимают вид:

$$\mathbf{J}_1 = -\frac{L_{1q}}{T^2} \operatorname{grad} T - \frac{L_{11}}{T} \frac{c}{c_2} \frac{\partial \mu_1^c}{\partial c_1} \operatorname{grad} c_1, \quad (1.23.49)$$

$$\mathbf{J}_q = -\frac{L_{qq}}{T^2} \operatorname{grad} T - \frac{L_{q1}}{T} \frac{c}{c_2} \frac{\partial \mu_1^c}{\partial c_1} \operatorname{grad} c_1. \quad (1.23.50)$$

Окончательные выражения для плотностей потока частиц и потока тепла получаем из формул (1.23.49) и (1.23.50) в следующем виде:

$$\mathbf{J}_1 = -D_1 \operatorname{grad} c_1 - c(D_T/T) \operatorname{grad} T, \quad (1.23.51)$$

$$\mathbf{J}_q = -\lambda \operatorname{grad} T - D_D \operatorname{grad} c_1. \quad (1.23.52)$$

Здесь введены экспериментально измеряемые коэффициенты:

- коэффициент диффузии ($\text{м}^2/\text{с}$)

$$D_1 = \frac{L_{11}}{T} \frac{c}{c_2} \frac{\partial \mu_1^c}{\partial c_1} = L_{11} \frac{Rc}{c_1 c_2} \left(1 + \frac{\partial \ln \gamma_1}{\partial \ln c_1} \right); \quad (1.23.53)$$

- коэффициент термодиффузии ($\text{м}^2/\text{с}$)

$$D_T = L_{1q} \frac{1}{cT} \equiv k_T D_1 \equiv S_T T \frac{c_1 c_2}{c^2} D_1; \quad (1.23.54)$$

- коэффициент теплопроводности ($\text{Вт}/(\text{м} \cdot \text{К})$)

$$\lambda = \frac{L_{qq}}{T^2}; \quad (1.23.55)$$

- коэффициент Дюфура ($\text{Вт} \cdot \text{м}^2$)

$$\begin{aligned} D_D &= \frac{L_{q1}}{T} \frac{c}{c_2} \frac{\partial \mu_1^c}{\partial c_1} = L_{q1} \frac{Rc}{c_1 c_2} \left(1 + \frac{\partial \ln \gamma_1}{\partial \ln c_1} \right) = \\ &= RT \frac{c^2}{c_1 c_2} \left(1 + \frac{\partial \ln \gamma_1}{\partial \ln c_1} \right) D_T \equiv Q D_1, \end{aligned} \quad (1.23.56)$$

где, с учетом соотношения взаимности Онсагера (1.19.12) в форме $L_{1q} = L_{q1}$, введена энергетическая величина (Дж)

$$Q = \frac{L_{1q}}{L_{11}} = \frac{L_{q1}}{L_{11}}, \quad (1.23.57)$$

называемая *теплотой переноса* (физический смысл дан ниже).

Вторые выражения в формулах (1.23.53) и (1.23.56) получены для химического потенциала в форме (1.13.4), а именно

$$\mu_1^c = \mu_1^\circ(T) + RT \ln \gamma_1 c_1, \quad (1.23.58)$$

где γ_1 — коэффициент активности примеси 1 в растворителе 2.

Коэффициент диффузии (1.23.53) отличается от аналогичной величины D_1 в формуле (1.23.48), полученной для изотермической диффузии (при $\nabla T = 0$), только наличием вклада от коэффициента активности в форме множителя $(1 + \partial \ln \gamma_1 / \partial \ln c_1)$.

При наличии градиента температуры, во-первых, появляется вклад в тепловой поток $\mathbf{J}_q$ с коэффициентом теплопроводности λ (формула (1.23.55)) и, во-вторых, возникают два перекрестных эффекта — *термодиффузия* (называемая *эффектом Соре*) и *диффузионная теплопроводность* (называемая диффузионным термоэффектом, или *эффектом Дюфура*), которые описываются формулами (1.23.54) и (1.23.56), соответственно.

Эффект термодиффузии заключается в возникновении градиента концентрации под влиянием градиента температуры. В этом случае уравнение (1.23.51) при условии $\mathbf{J}_1 = 0$ приводит к следующему соотношению:

$$\nabla c_1 = -k_T \frac{c}{T} \nabla T = -S_T \frac{c_1 c_2}{c} \nabla T, \quad (1.23.59)$$

где введены измеряемые величины (см. формулу (1.23.54)):

- *термодиффузионное отношение*

$$k_T \equiv \frac{D_T}{D_1} = \frac{c_1 c_2}{c^2} \frac{Q}{RT} \left(1 + \frac{\partial \ln \gamma_1}{\partial \ln c_1} \right)^{-1}, \quad (1.23.60)$$

- *коэффициент Соре*

$$S_T \equiv \frac{k_T}{T} \frac{c^2}{c_1 c_2} = \frac{Q}{RT^2} \left(1 + \frac{\partial \ln \gamma_1}{\partial \ln c_1} \right)^{-1}. \quad (1.23.61)$$

Термодиффузионное отношение k_T является безразмерной величиной, а коэффициент Соре S_T оказывается величиной порядка $10^{-2} - 10^{-5} \text{ K}^{-1}$ для газовых и жидких смесей, но в полимерных растворах может принимать большие значения. Явление термодиффузии применяется для разделения смесей газов и жидкостей, в первую очередь изотопов.

Диффузионный термоэффект, являясь обратным по отношению к эффекту термодиффузии, заключается в возникновении градиента температуры при взаимной диффузии веществ с одинаковой исходной температурой. В этом случае уравнение (1.23.52)

при условии $\mathbf{J}_q = 0$ приводит к следующему соотношению

$$\nabla T = -\frac{D_D}{\lambda} \nabla c_1 = -Q \frac{D_1}{\lambda} \nabla c_1. \quad (1.23.62)$$

Физический смысл теплоты переноса Q в форме (1.23.57), входящей в формулы (1.23.60)–(1.23.62), можно понять из выражения (1.23.52), переписанного в виде

$$\mathbf{J}_q = -\lambda \operatorname{grad} T - Q D_1 \operatorname{grad} c_1. \quad (1.23.63)$$

Действительно, в однородном температурном поле ($\operatorname{grad} T = 0$), когда $\mathbf{J}_1 = -D_1 \operatorname{grad} c_1$, из выражения (1.23.63) получаем

$$\mathbf{J}_q = Q \mathbf{J}_1.$$

Следовательно, *теплота переноса* ($Q = |\mathbf{J}_q|$) представляет собой тепловой поток, который создается единичным диффузионным потоком ($|\mathbf{J}_1| = 1$) в изотермических условиях.

Эффект Дюфура (диффузионный термоэффект, или диффузионная теплопроводность) экспериментально наблюдался в газах и приводил к разности температур в несколько градусов. Для жидкостей, как показывают численные оценки, величина D_D примерно в 10^5 раз меньше, а λ в 10^2 раз больше, чем в газах, что дает в тысячи раз более слабый эффект.

На этом ограничим рассмотрение связи между коэффициентами Онсагера и экспериментально измеряемыми величинами. Выше были рассмотрены три простейших случая, а именно:

а) химические реакции, работающие в линейном (квазиравновесном) режиме, т. е. при малых отклонениях от равновесия, когда применимы соотношения Онсагера;

б) взаимное влияние компонентов при изотермической диффузии в двухкомпонентной системе;

в) перекрестное влияние диффузии и теплопроводности (неизотермическая диффузия) в двухкомпонентной системе.

За рамками нашего рассмотрения остались многочисленные перекрестные кинетические явления, анализируемые в рамках линейной теории Онсагера. Например, сюда относятся:

- связанные химические реакции,
- диффузия в многокомпонентных системах,
- термоэлектрические явления (эффекты Зеебека и Пельтье),
- разнообразные гальваноманнитные и термомагнитные явления,
- теплопроводность и термодиффузия в системах с химическими реакциями.

Обсуждение и анализ вышеперечисленных перекрестных кинетических эффектов можно найти в литературе, включая книги Пригожина с соавторами.

В заключение осталось проанализировать локальные условия устойчивости и эволюции неравновесных термодинамических систем, работающих в нелинейном режиме, т. е. в условиях далеких от равновесия, когда неприменима теория Онсагера.

1.24. Локальная устойчивость неравновесного состояния и универсальный принцип эволюции

Флуктуационная устойчивость равновесного состояния, реализуемая при $\delta^2 S < 0$ в форме выражения (1.17.16) для полных термодинамических величин, была доказана в п. 1.17. Как уже отмечалось, переход от равновесных однородно распределенных величин к неравновесным пространственно неоднородным состояниям осуществляется в рамках гипотезы о локальном квазиравновесии (см. п. 1.21).

1. Обоснование флуктуационной устойчивости неравновесных состояний. До сих пор в качестве локальных величин применялись объемные плотности (на единицу объема системы).

Однако обобщение выражения (1.17.16) на локально квазиравновесные ситуации оказывается возможным лишь в применении к *удельным величинам* (отнесенным к единице массы).

Дело в том, что фундаментальное уравнение Гиббса (1.21.8) для полных величин при переходе к плотностям (делением на объем V системы) принимает форму (1.21.11), в которой полностью исчезает объемный вклад даже при $V \neq \text{const}$. В отличие от этого, введение удельных величин (делением на массу m системы) сохраняет структуру исходного уравнения Гиббса, а именно (ср. уравнения (1.21.8) и (1.21.11))

$$d\bar{u} = Td\bar{s} - Pdv + \sum_i \mu_i d\bar{c}_i, \quad (1.24.1)$$

где $\bar{u} = U/m$, $\bar{s} = S/m$, $\bar{c}_i = n_i/m$, $v = V/m$ — соответствующие удельные величины (на единицу массы). При этом химический потенциал μ_i сохраняет прежний смысл парциальной свободной энергии Гиббса на один моль i -го вещества.

Легко заметить, что между удельными величинами (отмеченными чертой) и использованными ранее объемными плотностями (без черты) существуют следующие соотношения:

$$u = \rho\bar{u}, \quad s = \rho\bar{s}, \quad c_i = \rho\bar{c}_i, \quad v = 1/\rho, \quad (1.24.2)$$

где введена суммарная (по всем компонентам) плотность массы,

$$\rho = \frac{m}{V} = \frac{1}{V} \sum_i n_i M_i = \sum_i c_i M_i = \sum_i \rho_i, \quad (1.24.3)$$

составленная из парциальных плотностей, $\rho_i = c_i M_i$, веществ с молярной массой M_i .

Переход от исходного уравнения Гиббса в форме (1.21.8) к его новой форме (1.24.1) для удельных величин осуществляется по той же схеме, которая была выполнена в п. 1.21 для объемных плотностей, только теперь в формулах типа (1.21.9) следует сделать замену $V \rightarrow m$.

Выражение (1.24.1) позволяет записать первую вариацию удельной энтропии (делая замену $d \rightarrow \delta$),

$$\delta \bar{s} = \frac{1}{T} \delta \bar{u} + \frac{P}{T} \delta v - \sum_i \frac{\mu_i}{T} \delta \bar{c}_i. \quad (1.24.4)$$

Отсюда, согласно общим правилам (Д.1) и (Д.2) на с. 147, вытекает следующее выражение для второй вариации:

$$\delta^2 \bar{s} = \delta \left(\frac{1}{T} \right) \delta \bar{u} + \delta \left(\frac{P}{T} \right) \delta v - \sum_i \delta \left(\frac{\mu_i}{T} \right) \delta \bar{c}_i. \quad (1.24.5)$$

Ввиду того, что величина (1.24.5) по структуре полностью идентична формуле (1.17.9) для полных величин, результаты их последующих преобразований будут также сходными. Это позволяет воспользоваться окончательным выражением (1.17.16) для того, чтобы записать следующее выражение:

$$\delta^2 \bar{s} = -\frac{1}{T} \left[\frac{c_v}{T} (\delta T)^2 + \frac{\rho}{\chi} (\delta v)_{\bar{c}_i}^2 + \sum_i \sum_j \mu_{ij} \delta \bar{c}_i \delta \bar{c}_j \right], \quad (1.24.6)$$

при этом в формуле (1.17.16) сделаны замены: $C_V \rightarrow c_v = C_V/m$, $\delta V \rightarrow \delta v$, $V \rightarrow v = 1/\rho$, $\delta n_{i(j)} \rightarrow \delta \bar{c}_{i(j)}$.

Таким образом, как и для термодинамического равновесия, выполнение неравенств (1.17.17)–(1.17.19) обеспечивает условие $\delta^2 \bar{s} \leq 0$. Это гарантирует устойчивость (термическую, механическую, материальную) неравновесного стационарного состояния, характеризуемого минимумом производства энтропии, как показано в точке 1 на рис. 1.7 б. Такая ситуация может реализоваться лишь при небольшом отклонении от равновесия, когда в окрестности минимума $\mathcal{P}_{\min}$ возникают только положительные флуктуации ($\delta \mathcal{P}^+ > 0$), которые самопроизвольно затухают в силу стремления системы к минимуму производства энтропии.

Как уже отмечалось в п. 1.9, совершенно иная ситуация возникает при больших отклонениях от равновесия, когда начинают

проявляться нелинейные свойства системы. Для таких нелинейных систем не существует функций состояния, которые имели бы экстремум в стационарных условиях, в частности, производство энтропии отнюдь не является минимальным (см. поведение $\mathcal{P}$ в окрестности точки 2 на рис. 1.7 б). В подобных ситуациях приходится обращаться к иному критерию, который был сформулирован Пригожиным как общий принцип эволюции нелинейных диссипативных структур.

2. Универсальный критерий эволюции нелинейных систем. Для дальнейшего анализа удобно в скорости изменения производства энтропии $(d\mathcal{P}/dt)$, представленного в общей форме (1.23.21), разделить вклады от термодинамических сил (dX_α/dt) и от термодинамических потоков (dJ_α/dt) , а именно

$$\begin{aligned} \frac{d\mathcal{P}}{dt} &= \sum_{\alpha} \int_{\dot{V}} J_{\alpha} \frac{dX_{\alpha}}{dt} dV + \sum_{\alpha} \int_{\dot{V}} X_{\alpha} \frac{dJ_{\alpha}}{dt} dV \equiv \\ &\equiv \frac{d_X \mathcal{P}}{dt} + \frac{d_J \mathcal{P}}{dt}. \end{aligned} \quad (1.24.7)$$

В области применимости линейных законов и соотношений Онсагера оба слагаемых в формуле (1.24.7) становятся равными. Действительно, используя уравнения (1.19.11) и (1.19.12), получаем (при $L_{\alpha\beta} = L_{\beta\alpha} = \text{const}$)

$$\begin{aligned} \frac{d_X \mathcal{P}}{dt} &\equiv \sum_{\alpha} \int_{\dot{V}} J_{\alpha} \frac{dX_{\alpha}}{dt} dV = \sum_{\alpha} \sum_{\beta} \int_{\dot{V}} L_{\alpha\beta} X_{\beta} \frac{dX_{\alpha}}{dt} dV = \\ &= \sum_{\beta} \int_{\dot{V}} X_{\beta} \frac{dJ_{\beta}}{dt} dV \equiv \frac{d_J \mathcal{P}}{dt} = \frac{1}{2} \frac{d\mathcal{P}}{dt} \leq 0, \end{aligned} \quad (1.24.8)$$

где последнее неравенство является следствием теоремы о минимальном производстве энтропии для линейных процессов.

Вне линейной области протекания неравновесных процессов вклады от сил $(d_X \mathcal{P}/dt)$ и от потоков $(d_J \mathcal{P}/dt)$ в полную производную $d\mathcal{P}/dt$ становятся произвольными, так что эта производная, как и $d_J \mathcal{P}/dt$, не имеет какого-либо общего свойства и не может играть роль критерия.

Однако изменение производства энтропии, обусловленное изменением силы $(d_X \mathcal{P}/dt)$, может принимать как положительный, так и отрицательный знак. Именно отрицательный знак и указывает направление эволюции нелинейной системы к стационарному состоянию. Это, однако, не происходит автоматически

(как было в линейном режиме), а требует вполне определенного выбора параметров системы.

Для доказательства вышесказанного записываем $d_X \mathcal{P}/dt$ на основе выражения (1.23.10) для плотности производства энтропии (следуя общепринятому, опускаем вязкость, $\overline{\pi} = 0$):

$$\begin{aligned} \frac{d_X \mathcal{P}}{dt} \equiv \int_V \frac{d_X \sigma_s}{dt} dV = \int_V \left[\mathbf{J}_u \cdot \frac{d}{dt} \left(\nabla \frac{1}{T} \right) - \right. \\ \left. - \sum_i \mathbf{J}_i \cdot \frac{d}{dt} \left(\nabla \frac{\mu_i}{T} \right) + \sum_r v_r \frac{d}{dt} \left(\frac{A_r}{T} \right) \right] dV. \end{aligned}$$

Применяя теорему Остроградского–Гаусса и выполняя интегрирование по частям, после кропотливых вычислений (детали которых можно найти в книгах Пригожина с соавторами) приходим к окончательному результату:

$$\frac{d_X \mathcal{P}}{dt} = - \int_V \frac{\rho}{T} \left[\frac{c_v}{T} \left(\frac{d}{dt} \frac{1}{T} \right)^2 + \frac{\rho}{\chi} \left(\frac{dv}{dt} \right)_{\bar{c}_i}^2 + \sum_i \sum_j \mu_{ij} \frac{d\bar{c}_i}{dt} \frac{d\bar{c}_j}{dt} \right] dV. \quad (1.24.9)$$

Здесь также использовано предположение о том, что граничные условия не зависят от времени, для того чтобы обеспечить возможность существования стационарных состояний.

Сравнение формул (1.24.6) и (1.24.9) показывает, что условия, обеспечивающие $\delta^2 \bar{s} \leq 0$ (а именно, положительность c_v , χ и двойной суммы по компонентам i и j , см. формулы (1.17.17)–(1.17.19)), одновременно гарантируют и выполнение требования

$$\frac{d_X \mathcal{P}}{dt} \leq 0. \quad (1.24.10)$$

Оно является **универсальным критерием эволюции** неравновесных систем к стационарному состоянию, в котором $d_X \mathcal{P}/dt = 0$.

Универсальность критерия (1.24.10) определяется тем, что он сохраняет свою применимость не только к нелинейным и линейным неравновесным процессам, но и к термодинамически равновесным процессам. Действительно, уравнение (1.24.8) показывает, что требование минимального (*ненулевого*) производства энтропии в стационарном состоянии линейных неравновесных систем (теорема Пригожина) является частным случаем общего критерия (1.24.10). Равновесные процессы также управляются теоремой Пригожина только с требованием *нулевого* производства энтропии в равновесном состоянии системы.

Поскольку величина $d_X \mathcal{P}/dt$, входящая в критерий (1.24.10), порождена суммарным вкладом всех производных dX_α/dt , то в основе *универсального принципа эволюции* термодинамических систем (нелинейных, линейных, равновесных) лежит их стремление *уменьшить термодинамические силы* ($dX_\alpha/dt \leq 0$), вызывающие тот или иной (не только неравновесный, но и равновесный) процесс. Именно уменьшение термодинамических сил снижает производство энтропии как в линейных неравновесных, так и в равновесных процессах окружающего нас мира.

Это заключение действительно является универсальным, так как подтверждено многочисленными опытными данными.

Кроме однородных стационарных состояний нелинейных систем, возможны также их неоднородные состояния в форме колебательных (осциллирующих во времени) структур и пространственно-периодических диссипативных структур. Примером первых может служить известная колебательная химическая реакция Белоусова–Жаботинского, а вторых — стационарные пространственно-периодические структуры Бенара и Тьюринга. Математический анализ таких структур необходимо строить, выходя за рамки «чистой» термодинамики, с привлечением общей теории бифуркаций и применением критериев устойчивости Ляпунова.

Эти вопросы выходят за рамки нашего изложения и могут быть найдены в многочисленных публикациях, включая работы Пригожина с соавторами, приведенные в списке рекомендуемой литературы в конце книги.

1.25. Контрольные вопросы

1. Как принято характеризовать микросостояние и макросостояние физических систем при классическом и квантовом способах их описания?
2. Сформулируйте статистическое определение энтропии по Больцману и укажите ее связь с информационной энтропией по Шеннону.
3. В чем заключается статистический смысл термодинамического определения температуры?
4. Дайте физическую характеристику основных свойств энтропии в применении к равновесной термодинамике.
5. В рамках каких условий справедлив закон возрастания энтропии в традиционной формулировке и какова область его применимости для реальных физических систем?

6. Сформулируйте три начала термодинамики с позиции статистической физики.
7. Какие контакты (каналы взаимодействия) термодинамической системы с окружающей средой вносят вклад во внутреннюю энергию системы?
8. В чем состоит принципиальное различие между замкнутыми, закрытыми и открытыми физическими системами?
9. Запишите фундаментальное уравнение Гиббса и дайте его физическую интерпретацию.
10. Как определены термодинамические функции состояния и функции процесса? Приведите примеры таких функций.
11. Дайте определение адиабатических/изотермических и изохорных/изобарных процессов и приведите примеры их технологической реализации.
12. Какие термодинамические функции системы описывают тепловой эффект процессов, протекающих в изохорных и изобарных условиях?
13. Каковы физические условия действия теплового закона Гесса? Приведите примеры практического применения закона.
14. Какая термодинамическая функция управляет равновесием и направлением самопроизвольного протекания процессов в замкнутой системе?
15. Какие термодинамические функции управляют равновесием и направлением самопроизвольного протекания процессов при изохорно-изотермических и изобарно-изотермических условиях в закрытой системе?
16. Существует ли различие в поведении замкнутой, закрытой и открытой систем, выведенных из состояния термодинамического равновесия?
17. Дайте общую формулировку второго начала термодинамики для открытых систем.
18. Что называют производством энтропии и в чем заключается принцип минимума производства энтропии?
19. Как определены избыточная энтропия и избыточное производство энтропии?
20. Как формируется стационарное состояние открытых систем и какова роль негэнтропии?

21. Какова роль флуктуаций в равновесных структурах, а также в структурной самоорганизации неравновесных систем?
22. Что понимается под флуктуационной устойчивостью равновесных и неравновесных систем и каковы условия ее реализации?
23. С какой целью в дополнение к внутренней энергии вводят в рассмотрение термодинамические потенциалы в форме энтальпии и свободных энергий Гельмгольца и Гиббса?
24. Объясните физический смысл и основные свойства химического потенциала компонента, входящего в состав многокомпонентной гетерогенной системы.
25. Какая физическая причина обеспечивает универсальность газового состояния вещества, описываемого уравнением Клапейрона–Менделеева?
26. Обоснуйте физическую причину появления энтропии смешения в процессе смешивания газов и образования жидких и твердых растворов.
27. В чем общность и различие между моделями идеального газа и идеального раствора?
28. Какова необходимость введения стандартного состояния для материальных веществ и как оно отображается в таблицах стандартных термодинамических величин?
29. Имеется ли различие между стандартными состояниями одной и той же многокомпонентной системы при записи химического потенциала в формах (1.13.2) и (1.13.3)?
30. Сформулируйте общие условия фазового равновесия для многокомпонентных гетерогенных систем без химически активных компонентов.
31. Сформулируйте условия химического равновесия для многокомпонентных гетерогенных систем с химически активными компонентами.
32. Дайте общее определение изобарного потенциала и сродства химической реакции.
33. Как изменяются условия фазового и химического равновесия при учете электрического заряда частиц, входящих в состав многокомпонентной гетерогенной системы?
34. Чем определяется флуктуационная устойчивость термодинамического и текущего равновесия системы?

35. В чем физически проявляется нарушение условий фазового и химического равновесия?
36. Дайте определение термоактивационного процесса и сформулируйте закон Аррениуса.
37. Объясните основное содержание феноменологической теории Онсагера. Какова область ее применимости?
38. В чем состоит принцип микроскопической обратимости (или детального равновесия) по Онсагеру?
39. Сформулируйте теорему Пригожина о минимуме производства энтропии.
40. Какова связь теоремы Пригожина с обобщенным принципом Ле Шателье?
41. Объясните физический смысл локального квазиравновесия в неравновесных системах.
42. В чем состоят особенности лагранжева и эйлера подхода к описанию движения сплошной среды?
43. Какова общая структура уравнения баланса любой физической величины?
44. Запишите уравнения баланса числа частиц, внутренней энергии и энтропии; дайте их физическую интерпретацию.
45. Дайте определение плотности потока энтропии и локального производства энтропии.
46. Какие неравновесные процессы вносят вклад в локальное производство энтропии?
47. В чем состоит принцип симметрии Кюри в формулировке Пригожина для изотропных термодинамических систем?
48. Продемонстрируйте связь коэффициентов Онсагера со скоростью химической реакции. Всегда ли она возможна?
49. Продемонстрируйте связь коэффициентов Онсагера с коэффициентами диффузии и термодиффузии в изотермических и неізотермических условиях (эффекты Соре и Дюфура).
50. Сформулируйте универсальный принцип эволюции термодинамических систем.

Глава 2

УПРАВЛЕНИЕ ФАЗОВЫМИ И ХИМИЧЕСКИМИ РАВНОВЕСИЯМИ В ТЕХНОЛОГИЧЕСКИХ ЗАДАЧАХ ЭЛЕКТРОНИКИ

В основе любого технологического процесса лежит определенное физическое или химическое воздействие на материал с целью управляемого изменения его состояния, структуры или состава. Несмотря на многообразие технологических операций, методов и процессов, применяемых в современных технологиях микро- и нанoeлектроники, они могут быть объединены общностью физико-химических явлений, лежащих в основе их действия. Базовые положения термодинамики позволяют с единых позиций рассмотреть разнообразные процессы технологии.

Например, технологические операции, проводимые в вакууме и газовых средах (такие как вакуумное обезгаживание и очистка материалов, термовакуумное испарение и конденсация, газофазная эпитаксия, легирование полупроводников из газовой фазы и др.), управляются по законам испарения или сублимации конденсированных сред, растворения газообразных компонентов в этих средах с образованием растворов и химических соединений.

Закономерности взаимодействия между жидкой и твердой фазами, описываемые с помощью так называемых диаграмм плавкости, лежат в основе таких технологических операций, как пайка припоями, формирование омических контактов к полупроводникам, жидкофазная эпитаксия, кристаллизационные методы выращивания и очистки веществ и др.

Закономерности химического и электрохимического взаимодействия управляют разнообразными процессами технологии. Среди них, например, очистка, травление и окисление поверхности, химический транспорт веществ при газофазной эпитаксии, формирование люминофорного и термоэмиссионного покрытий в вакуумных приборах и многое другое.

Задача данной главы — рассмотрение с единых позиций физических закономерностей, описывающих равновесие между различными фазами и управляющих изменениями фазового состояния веществ, а также основных положений химической термодинамики для систем с реактивными компонентами, которые описывают равновесие химических реакций и управляют смещением этого равновесия в ту или иную сторону.

2.1. Фазовые превращения в однокомпонентных системах

Рассмотрим однокомпонентную систему без химических взаимодействий, для которой $K = 1$ и $X = 0$. Согласно правилу фаз (1.15.1), такая система имеет число термодинамических степеней свободы $C = 3 - \Phi$. Следовательно, в невариантных условиях (при $C = 0$) однокомпонентная система имеет максимально возможное число фаз, соответствующих *трем* агрегатным состояниям вещества.

Существование *одной фазы* является бивариантным ($C = 2$), т. е. температуру T и давление P можно одновременно изменять независимо друг от друга.

Две фазы не могут находиться в равновесии при произвольных значениях температуры и давления, поскольку система является моновариантной ($C = 1$), т. е. задание одной из величин, T или P , однозначно определяет другую. Требование (1.14.3), наложенное на химические потенциалы фаз, с учетом (1.13.3) для чистого вещества, записывается в следующей форме (нижний индекс i и верхний индекс $^\circ$ за ненадобностью опущены):

$$\mu^{\text{I}}(T, P) = \mu^{\text{II}}(T, P). \quad (2.1.1)$$

Здесь римские цифры нумеруют сосуществующие равновесные фазы I и II. Равенство (2.1.1) представляет собой уравнение для нахождения функциональной зависимости $P = f(T)$.

Состояние однокомпонентной системы принято изображать в виде **фазовой диаграммы** на плоскости P – T , показанной на рис. 2.1. Кривые 1, 2 и 3, найденные из решения уравнения (2.1.1), дают линии равновесия, вдоль которых сосуществуют две фазы: 1 — *кривая сублимации* или *возгонки* (равновесие твердой и газовой фаз), 2 — *кривая испарения* (равновесие жидкой и газовой фаз), 3 — *кривая плавления* (равновесие твердой и жидкой фаз). В отличие от кривых сублимации и плавления, кривая испарения оканчивается в так называемой критической

точке K , где плотности пара и жидкости сравниваются, так что различие между этими фазами исчезает.

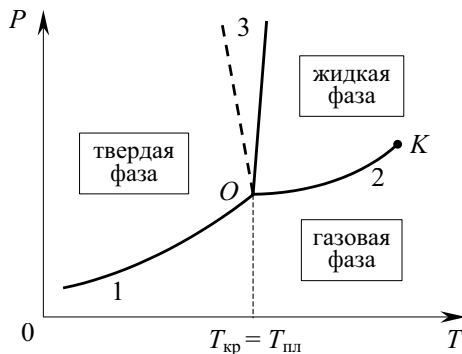


Рис. 2.1. Фазовая диаграмма для чистого вещества: 1 — кривая сублимации (возгонки), 2 — кривая испарения, 3 — кривая плавления; O — тройная точка, K — критическая точка

Каждая кривая равновесия разделяет две смежные однофазные области, в которых система бивариантна. Изменение ее состояния, происходящее вдоль линии, которая пересекает кривую равновесия, соответствует фазовому превращению, называемому *фазовым переходом первого рода*.

Точка O пересечения трех кривых, называемая *тройной точкой*, дает невариантное равновесие твердой, жидкой и газовой фаз. Такая физическая ситуация возникает при кристаллизации твердой фазы из жидкости, находящейся в равновесии с собственным паром, или при обратном процессе плавления твердого тела. Поэтому температуру тройной точки можно принять за температуру кристаллизации $T_{кр}$ жидкости или температуру плавления $T_{пл}$ твердого тела.

Покажем, что фазовые переходы, соответствующие агрегатным превращениям вещества, всегда сопровождаются тепловым эффектом и изменением мольного объема. Для этого качественно изобразим на рис. 2.2 температурный ход химических потенциалов μ^I и μ^{II} для одного и того же вещества, находящегося в двух фазах I и II при $P = \text{const}$. Отрицательный наклон этих кривых соответствует формуле (1.11.21), согласно которой для чистого вещества $\partial\mu/\partial T = -S < 0$ (индексы i и 0 опущены). Аналогично могут быть построены зависимости $\mu^{I,II} = f(P)$ при $T = \text{const}$, но с положительным наклоном, поскольку $\partial\mu/\partial P = V > 0$. Здесь S и V суть мольные величины.

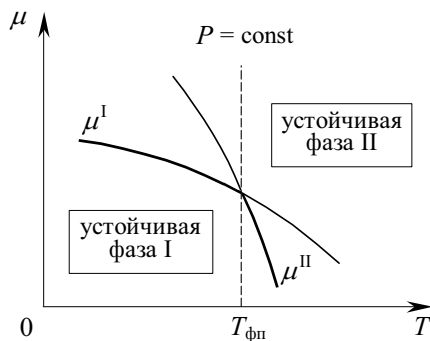


Рис. 2.2. Температурные зависимости химического потенциала чистого вещества, находящегося в двух фазах, для определения устойчивости и равновесия фаз при температуре фазового перехода $T_{\text{фп}}$

Точка пересечения химических потенциалов дает температуру фазового перехода $T_{\text{фп}}$, при которой две фазы, в соответствии с равенством (2.1.1), находятся в равновесии при давлении $P = \text{const}$. С изменением давления температура $T_{\text{фп}}$ изменяется, давая одну из кривых равновесия 1, 2 или 3, показанных на рис. 2.1. При температурах, отличных от $T_{\text{фп}}$, устойчива та фаза, которая имеет меньший химический потенциал (что в однокомпонентной системе при $n = \text{const}$ обеспечивает минимум свободной энергии Гиббса $G = n\mu$). Фаза I на рис. 2.2 является низкотемпературной, поскольку существует при $T < T_{\text{фп}}$, а фаза II, существующая при $T > T_{\text{фп}}$, высокотемпературная. Результирующий температурный ход химического потенциала описывается жирной кривой, которая имеет излом в точке фазового перехода $T_{\text{фп}}$, такой что

$$\left(\frac{\partial \mu^I}{\partial T}\right)_P > \left(\frac{\partial \mu^{II}}{\partial T}\right)_P, \quad \text{откуда} \quad S^I < S^{II}.$$

Следовательно, фазовый переход $I \rightarrow II$, возникающий при повышении температуры, всегда сопровождается скачкообразным увеличением молярной энтропии на величину $\Delta S^{I \rightarrow II} = S^{II} - S^I > 0$. Поскольку переход происходит равновесным образом ($\Delta G = 0$), то изменение энтальпии,

$$\Delta H^{I \rightarrow II} = T_{\text{фп}} \Delta S^{I \rightarrow II} = T_{\text{фп}} (S^{II} - S^I) > 0,$$

проявляется в виде поглощаемой теплоты, называемой *скрытой теплотой фазового перехода*. Этот термин означает, что поглощение тепла не приводит к изменению температуры системы, так как оно расходуется на перестройку структуры вещества

при $T_{\text{фп}} = \text{const}$. Очевидно, что для обратного перехода $\text{II} \rightarrow \text{I}$, происходящего при понижении температуры, имеем $\Delta S^{\text{II} \rightarrow \text{I}} < 0$ и $\Delta H^{\text{II} \rightarrow \text{I}} < 0$, т.е. теплота выделяется. Именно эти скачки энтальпии и энтропии при температуре плавления–кристаллизации $T_{\text{пл}} \equiv T_{\text{кр}}$ и при температуре кипения $T_{\text{кип}}$ качественно показаны на рис. 2.3.

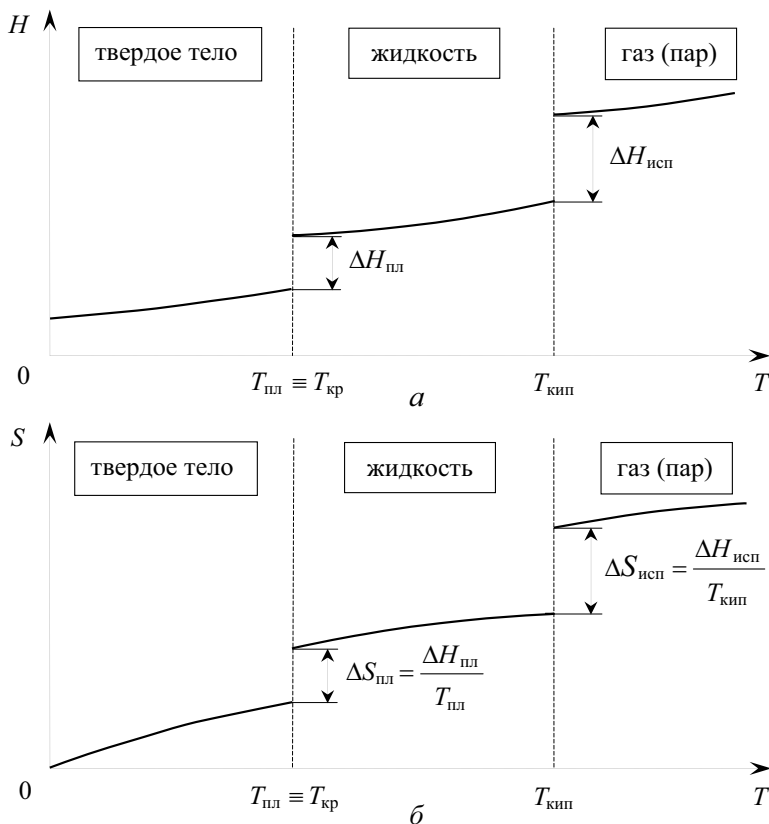


Рис. 2.3. Качественный ход температурных зависимостей энтальпии $H(T)$ и энтропии $S(T)$ с учетом фазовых превращений вещества при температуре плавления–кристаллизации $T_{\text{пл}} \equiv T_{\text{кр}}$ и при температуре кипения $T_{\text{кип}}$

Аналогично рассмотренной выше ситуации, должен наблюдаться излом также и на зависимости $\mu = f(P)$ при $T = \text{const}$, для которого в точке фазового перехода имеем

$$\left(\frac{\partial \mu^{\text{I}}}{\partial P} \right)_T \neq \left(\frac{\partial \mu^{\text{II}}}{\partial P} \right)_T, \quad \text{откуда} \quad V^{\text{I}} \neq V^{\text{II}}.$$

Следовательно, фазовый переход $I \rightarrow II$ всегда сопровождается скачкообразным изменением мольного объема на величину

$$\Delta V^{I \rightarrow II} = V^{II} - V^I \neq 0.$$

Для процессов испарения и сублимации, когда газообразное состояние является конечным, имеем $\Delta V_{\text{исп(суб)}} = V^{\text{г}} - V^{\text{ж}} > 0$, так как мольный объем пара много больше объема, занимаемого тем же молекул в конденсированном (жидком или твердом) состоянии. Например, один моль воды имеет объем, равный 18 см^3 , в то время как при $T = 273 \text{ К}$ и $P = 1 \text{ атм}$ мольный объем идеального газа равен $22,4 \text{ литра}$, то есть $V^{\text{г}} \gg V^{\text{ж}}$.

Плавление большинства твердых тел также происходит с увеличением объема, то есть $\Delta V_{\text{пл}} = V^{\text{ж}} - V^{\text{т}} > 0$. Однако имеются вещества с менее плотной упаковкой атомов в твердом состоянии, чем в жидком. Сюда относятся, например, лед, германий и галлий, объем которых при плавлении уменьшается ($\Delta V_{\text{пл}} < 0$). Этим двум случаям соответствует разный наклон кривой плавления 3 на рис. 2.1: сплошная линия — при $\Delta V_{\text{пл}} > 0$ и штриховая — при $\Delta V_{\text{пл}} < 0$ (см. уравнение Клапейрона–Клаузиуса (2.7.3)).

Скачок первых производных химического потенциала по температуре и давлению, наблюдающийся при агрегатных превращениях вещества, дал им название *фазовых переходов первого рода*. Наряду с этим, в природе наблюдаются физические ситуации, когда в точке фазового перехода химические потенциалы на рис. 2.2 пересекаются с одновременным касанием, обеспечивающим непрерывность их первых производных. Если при этом скачком изменяются вторые частные производные химических потенциалов, то имеют место *фазовые переходы второго рода*. Для этих переходов отсутствуют как тепловой ($\Delta H = 0$), так и объемный ($\Delta V = 0$) эффекты. Однако претерпевают резкие изменения такие физические величины, как изобарный коэффициент теплового расширения α_P , изотермическая сжимаемость χ_T и изобарная теплоемкость C_P , пропорциональные вторым производным химического потенциала:

$$\alpha_P \equiv \frac{1}{V} \left(\frac{\partial V}{\partial T} \right)_P = \frac{1}{V} \frac{\partial^2 \mu}{\partial T \partial P},$$

$$\chi_T \equiv -\frac{1}{V} \left(\frac{\partial V}{\partial P} \right)_T = -\frac{1}{V} \frac{\partial^2 \mu}{\partial P^2},$$

$$C_P \equiv \left(\frac{\partial H}{\partial T} \right)_P = T \left(\frac{\partial S}{\partial T} \right)_P = -T \frac{\partial^2 \mu}{\partial T^2}.$$

Фазовые переходы второго рода наблюдаются, как правило, в твердой фазе и вызваны изменением симметрии твердого тела (переходы ферромагнетик–парамагнетик, сегнетоэлектрик–параэлектрик и т. п.). Такие переходы не представляют существенно-го интереса для технологических применений и в дальнейшем не будут рассматриваться.

2.2. Давление насыщенного пара чистого вещества

Насыщенным паром называется пар, находящийся в равновесии со своей конденсированной (жидкой или твердой) фазой. Кривые сублимации 1 и испарения 2, показанные на рис. 2.1, по своей сути выражают температурную зависимость давления насыщенного пара над чистой конденсированной фазой.

Выведем дифференциальное уравнение, описывающее эту зависимость, а также выясним влияние полного давления P в газовой фазе на давление насыщенного пара, $p_i^0 = f(T, P)$, для данного i -го вещества в *чистом виде* (что отмечено верхним нулевым индексом). При этом не следует путать рассматриваемое здесь давление p_i^0 с ранее принятым для газовой фазы обозначением стандартного значения $p_i^\circ = 1$ атм (см. п. 1.13), которое, будучи фиксированным, было опущено и в дальнейшем не встречается.

Пусть газовая фаза над чистым i -м веществом, помимо его собственного насыщенного пара с давлением p_i^0 , содержит также инертный газ с парциальным давлением $p_{\text{ин}}$, не растворяющийся в конденсированной фазе и не взаимодействующий с ее паром. Последующая задача — найти зависимость p_i^0 от температуры T и общего давления в газовой фазе, $P = p_i^0 + p_{\text{ин}}$.

Исходя из равенства (2.1.1), записываем условие равновесия пара со своей конденсированной фазой (при соответствующей замене верхних индексов I и II на «п» и «к»):

$$\mu_i^{\text{п}}(T, P) = \mu_i^{\text{к}}(T, P). \quad (2.2.1)$$

Согласно выражению (1.13.1), химический потенциал пара равен

$$\mu_i^{\text{п}}(T, P) = \mu_i^{\text{оп}}(T) + RT \ln p_i^0(T, P), \quad (2.2.2)$$

где давление p_i^0 рассматривается как искомая функция T и P .

Химический потенциал конденсированной фазы обычно выражают в форме (1.13.3) через мольную долю x_i , которая в случае чистого вещества равна единице. Поэтому при $x_i = 1$

$$\mu_i^{\text{к}}(T, P) \equiv \mu_i^{\text{ок}}(T, P) + RT \ln x_i = \mu_i^{\text{ок}}(T, P). \quad (2.2.3)$$

Для нахождения зависимости давления $p_i^0(T)$ насыщенного пара чистого вещества от температуры дифференцируем равенство (2.2.1) по T вдоль кривой испарения (или сублимации) при $P = \text{const}$, т. е. учитываем в формуле (2.2.2) неявную зависимость $p_i^0 = f(T, P = \text{const})$:

$$\frac{\partial \mu_i^\pi}{\partial T} + \frac{\partial \mu_i^\pi}{\partial p_i^0} \frac{\partial p_i^0}{\partial T} = \frac{\partial \mu_i^k}{\partial T}, \quad (2.2.4)$$

где, в соответствии с выражением (2.2.2), имеем

$$\frac{\partial \mu_i^\pi}{\partial p_i^0} = \frac{RT}{p_i^0}. \quad (2.2.5)$$

Согласно формуле (1.11.21), первое слагаемое в левой части и правая часть равенства (2.2.4) представляют собой взятые с обратным знаком молярные энтропии i -го вещества в паровой среде ($-S_i^{0\pi}$) и в конденсированном состоянии ($-S_i^{0k}$). С учетом этого, приводим равенство (2.2.4) к форме искомого дифференциального уравнения:

$$\frac{\partial \ln p_i^0}{\partial T} = \frac{\Delta H_i^{0k \rightarrow \pi}}{RT^2}, \quad (2.2.6)$$

где $\Delta H_i^{0k \rightarrow \pi} = T(S_i^{0\pi} - S_i^{0k}) > 0$ — молярная теплота фазового перехода конденсат $\rightarrow$ пар, т. е. теплота испарения (или сублимации) одного моля i -го сорта. Напоминаем, что верхний нулевой индекс в данном случае соответствует чистому веществу.

Для выяснения влияния общего давления P на давление p_i^0 насыщенного пара, дифференцируем равенство (2.2.1) по P при $T = \text{const}$ с учетом неявной зависимости $p_i^0 = f(P, T = \text{const})$ в равенстве (2.2.2):

$$\frac{\partial \mu_i^\pi}{\partial p_i^0} \frac{\partial p_i^0}{\partial P} = \frac{\partial \mu_i^k}{\partial P}. \quad (2.2.7)$$

Согласно формуле (1.11.21), правая часть этого выражения равна молярному объему V_i^{0k} чистого i -го вещества в конденсированном состоянии. С учетом производной (2.2.5), из формулы (2.2.7) получаем искомого дифференциальное уравнение:

$$\frac{\partial \ln p_i^0}{\partial P} = \frac{V_i^{0k}}{RT}. \quad (2.2.8)$$

Согласно уравнению Клапейрона–Менделеева (1.12.1), для одного моля ($n = 1$) идеального газа имеем $p_i^0 V_i^{0\pi} = RT$, тогда

уравнение (2.2.8) принимает следующий вид:

$$\frac{\partial p_i^0}{\partial P} = \frac{V_i^{0\kappa}}{V_i^{0\pi}} \ll 1. \quad (2.2.9)$$

Из-за малости мольного объема конденсата $V_i^{0\kappa}$ в сравнении с объемом $V_i^{0\pi}$, занимаемым одним молекул в газовой фазе ($V_i^{0\kappa} \ll V_i^{0\pi}$ — см. п. 2.1), уравнение (2.2.9) позволяет считать, что $\partial p_i^0 / \partial P \approx 0$, т. е. давление насыщенного пара p_i^0 практически не зависит от полного давления P . Такая зависимость начинает заметно проявляться только при давлениях в десятки и сотни атмосфер. Следовательно, давление паров конструкционных материалов внутри электровакуумного прибора практически не изменяется в результате его откачки и зависит лишь от рабочей температуры деталей прибора.

Нетрудно убедиться в том, что пренебрежение зависимостью p_i^0 от P связано с малостью правой части в уравнении (2.2.7), что позволяет положить там $\partial \mu_i^{\kappa} / \partial P \approx 0$. Это означает, что стандартный химический потенциал (2.2.3) для чистой конденсированной фазы практически не зависит от давления, а, как и для газовой фазы, зависит только температуры:

$$\mu_i^{\text{ок}}(T, P) \approx \mu_i^{\text{ок}}(T). \quad (2.2.10)$$

Температурную зависимость $p_i^0(T)$ описывают уравнением (2.2.6), в котором всегда $\Delta H_i^{0\kappa \rightarrow \pi} > 0$ (см. рис. 2.3), так как процессы испарения и сублимации, как и плавления, требуют подвода тепла. Теплота фазового перехода, строго говоря, сама зависит от температуры. Например, для жидкости теплота испарения уменьшается с ростом T и обращается в нуль в критической точке K на рис. 2.1, где исчезает различие между плотностями жидкости и ее пара. В определенном интервале температур вдали от критической точки можно считать $\Delta H_i^{0\kappa \rightarrow \pi}(T) \approx \text{const}$. Тогда интегрирование уравнения (2.2.6) дает

$$p_i^0(T) = A \exp\left(-\frac{\Delta H_i^{0\kappa \rightarrow \pi}}{RT}\right). \quad (2.2.11)$$

Постоянная интегрирования A для большинства жидкостей примерно одинакова и ее находят из эксперимента при температуре кипения $T_{i, \text{кип}}^0$ (см. формулу (2.2.12)).

Качественный ход зависимости (2.2.11) показан на рис. 2.4. Поскольку для всех веществ $\Delta H_i^{0\kappa \rightarrow \pi} > 0$, то давление насыщенного пара с ростом температуры всегда возрастает. Однако при высоких температурах, когда экспонента приближается

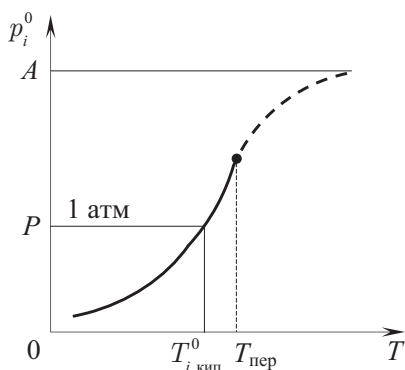


Рис. 2.4. Температурная зависимость давления насыщенного пара для чистого i -го вещества (практический интерес представляет участок экспоненциального нарастания, показанный сплошной линией до точки перегиба при $T_{пер}$)

к единице, зависимость $p_i^0(T)$ выходит на насыщение, где $p_i^0 = A$. Точке перегиба, в которой меняется кривизна кривой и начинается переход от участка экспоненциального нарастания (изображенного сплошной линией) к участку выхода на насыщение (изображенному пунктиром), соответствует температура $T_{пер} = \Delta H_i^{0\text{к}\rightarrow\text{п}}/2R$. Поскольку теплоты испарения и сублимации обычно составляют десятки ккал/моль, то $T_{пер}$ превышает 2500 К, что выше температуры плавления для большинства твердых тел и критической температуры для жидкостей (так, для воды $T_{крит} = 674 \text{ К}$). Следовательно, участок, лежащий выше точки перегиба, не представляет практического интереса. По этой причине на кривых зависимости давления насыщенного пара от температуры обычно изображают лишь начальный участок экспоненциального нарастания при $T < T_{пер}$ (см. рис. 2.6 и 2.8, а также температурные зависимости растворимости и констант химического равновесия, приведенные на рис. 2.9 б, 2.10 б и 2.23).

Внешнее давление, как было установлено выше, практически не влияет на давление насыщенного пара. Однако соотношение между ними видоизменяет характер испарения жидкости. До тех пор, пока p_i^0 меньше полного давления (пусть $P = 1 \text{ атм}$), жидкость свободно испаряется со своей внешней поверхности. При $p_i^0 \approx P$ характер парообразования резко изменяется: начинается внутреннее испарение с образованием пузырьков пара во всем объеме жидкости и на внутренних стенках сосуда, называемое *кипением*. Следовательно, температура кипения $T_{i,кип}^0$, показанная на рис. 2.4, является рядовой точкой на кривой испарения и всегда уменьшается с понижением внешнего давления.

Полагая в формуле (2.2.11) $p_i^0 = 1$ атм и $T = T_{i,\text{кип}}^0$, получаем следующее выражение для постоянной интегрирования:

$$A = \exp\left(\frac{\Delta H_{i,\text{исп}}^0}{RT_{i,\text{кип}}^0}\right) = \exp\left(\frac{\Delta S_{i,\text{исп}}^0}{R}\right), \quad (2.2.12)$$

где $\Delta S_{i,\text{исп}}^0 = \Delta H_{i,\text{исп}}^0 / T_{i,\text{кип}}^0$ — молярная энтропия испарения чистой i -й жидкости при температуре кипения $T_{i,\text{кип}}^0$ (см. рис. 2.3). Для большинства слабо ассоциированных жидкостей справедливо эмпирическое *правило Трутона*, согласно которому энтропия испарения при температуре кипения приблизительно одинакова для всех веществ: $\Delta S_{\text{исп}}^0 \approx 22$ кал/(моль · К).

Соотношение, аналогичное (2.2.12), может быть записано для процесса сублимации. При этом надо брать p_i^0 и $T_{i,\text{пл}}^0$ в тройной точке O на рис. 2.1, соответствующей пару, равновесному одновременно с твердым веществом и его расплавом. Для большинства металлов с плотной упаковкой атомов энтропия плавления примерно одинакова и равна $\Delta S_{\text{пл}}^0 \approx 2,2$ кал/(г-атом · К). Упорядоченные сплавы и химические соединения этому эмпирическому правилу не подчиняются.

Зависимость давления насыщенного пара $p_i^0(T)$, качественно изображенная на рис. 2.4 сплошной кривой, для процесса сублимации твердого вещества оканчивается в тройной точке O , а для процесса испарения жидкости — в критической точке K , как показано на рис. 2.1.

Прежде, чем распространить полученные выше выражения для чистых веществ на многокомпонентные системы, рассмотрим с общих позиций основные соотношения термодинамики твердых и жидких растворов.

2.3. Термодинамика образования твердых и жидких растворов

В различных задачах технологии часто приходится иметь дело с процессами, в которых участвуют как твердые, так и жидкие растворы. Общим признаком всех растворов является хаотическое распределение разнородных частиц в объеме раствора без нарушения его однородности. Для жидких растворов характерен лишь *ближний порядок* в расположении частиц, а в твердых растворах проявляется также и *дальний порядок*, присущий кристаллической решетке либо исходных близких по структуре веществ (GaAs–GaP, AlAs–GaAs, Ge–Si), либо основного вещества (растворителя) при введении в него примесей.

Твердые растворы имеют атомную структуру, поскольку их кристаллическая решетка построена из атомов, связанных между собой силами химического взаимодействия ковалентной, ионной или ионно-ковалентной природы. При взаимном растворении веществ с близкими по структуре и размерам атомами возникают **твердые растворы замещения** путем взаимной замены сходных атомов в узлах кристаллической решетки. Вхождение же в решетку растворителя атомов примеси относительно малых размеров, отличающихся по своим химическим свойствам от основных атомов и по этой причине размещающихся в междоузлиях решетки, дает **твердые растворы внедрения**. Если в междоузлиях размещаются собственные атомы химического соединения (например, атомы В бинарного соединения АВ), то нарушается его стехиометрический состав. Такой твердый раствор внедрения, приводящий к возникновению нестехиометричности соединения, называют *соединением переменного состава* и обозначают как $AB_{1+\delta}$. Стехиометрия соединения может нарушаться также в результате возникновения вакансий из-за отсутствия атомов в узлах решетки, например атомов В. Такое соединение вида $AB_{1-\delta}$ называют построенным по типу **твердого раствора вычитания**. Коэффициент δ (обычно $\delta \ll 1$), используемый в обозначениях $AB_{1\pm\delta}$ (или в более общем виде $A_mB_{n(1\pm\delta)}$), отражает степень нестехиометричности соединения АВ (или A_mB_n), вызванной собственными дефектами — междоузельными атомами и вакансиями (см. п. 3.5).

В микроэлектронике и оптоэлектронике широкое применение нашли твердые растворы полупроводников, построенные на основе двойных (А–В) или тройных (А–В–С) систем. Твердые растворы внедрения из-за различия компонентов чаще являются *ограниченными*, т. е. существуют лишь до определенной концентрации исходных компонентов, выше которой они распадаются. *Неограниченные* твердые растворы, существующие во всем интервале концентраций исходных компонентов (элементов А и В в двойных системах или стехиометрических соединений типа АВ и АС в тройных системах), могут быть построены по типу раствора замещения для компонентов с одинаковым типом кристаллической решетки и с близкими по размерам и электронному строению атомами. Такие соединения принято называть **упорядоченными твердыми растворами**. Их состав выражается в виде $A_{1-x}B_x$ — для раствора элементов А и В, в виде $A_{1-x}B_xC$ — для раствора соединений АС и ВС или в виде $AB_{1-x}C_x$ — для раствора соединений АВ и АС, где x — атомная доля ($0 \leq x \leq 1$).

Жидкие растворы по физическому строению отличаются от только что описанной картины для твердых растворов. В первую очередь, между частицами жидкости должно отсутствовать химическое взаимодействие. В противном случае возникает химически активная система со своими специфическими свойствами, отличными от свойств раствора. В жидких растворах частицы могут находиться в виде как молекул, так и ионов, возникающих в результате электролитической диссоциации. Растворы, содержащие ионы и называемые *растворами электролитов*, исключаются из рассмотрения в ближайших параграфах. Их термодинамические особенности, как и свойства химически активных систем, будут изложены ниже (см. п. 2.17).

Межмолекулярные взаимодействия в жидких растворах обусловлены силами Ван-дер-Ваальса (ориентационного, индукционного и дисперсионного типов), а также водородной и донорно-акцепторной (координационной) связями. Взаимодействие между однородными молекулами растворителя называют *ассоциацией*, а между разнородными молекулами (растворителя и растворенного вещества) — *сольватацией* (или *гидратацией* для водных растворов).

Процессы ассоциации и сольватации формируют ближнюю упорядоченность в растворах путем создания молекулярных комплексов в форме малоустойчивых молекулярных образований переменного состава, называемых, соответственно, *ассоциатами* и *сольватами*. Образование ассоциатов в чистых жидкостях сильнее связывает молекулы между собой (чему отвечает минимум энергии на рис. 2.5), в результате этого избыточная энергия выделяется в виде тепла. Аналогичным образом проявляется и процесс сольватации. Ассоциация и сольватация являются конкурирующими процессами: образование сольватов, сопровождающееся выделением тепла ($\Delta H < 0$), уменьшает степень ассоциации растворителя, что вызывает, наоборот, поглощение теплоты ($\Delta H > 0$). Результирующий тепловой эффект определяет общую теплоту процесса растворения при смешивании жидкостей.

В случае растворения газов и твердых тел в жидкостях тепловой эффект, вызванный процессами ассоциации и сольватации, дает, согласно формуле (1.5.13), вклад $\Delta H_{вз}$ в полную теплоту растворения,

$$\Delta H = \Delta H_{фп} + \Delta H_{вз}, \quad (2.3.1)$$

в дополнение к теплоте фазового перехода $\Delta H_{фп}$, связанной с изменением агрегатного состояния растворяемого вещества.

Наряду с тепловым эффектом $\Delta H_{вз}$, межмолекулярные взаимодействия в растворе приводят при смешении компонентов

к изменению энтропии $\Delta S_{\text{вз}}$ и суммарного объема $\Delta V_{\text{вз}}$. Если растворение сопровождается изменением агрегатного состояния одного или нескольких компонентов, то по аналогии с (2.3.1) к величинам $\Delta S_{\text{вз}}$ и $\Delta V_{\text{вз}}$ должны быть добавлены $\Delta S_{\text{фп}}$ и $\Delta V_{\text{фп}}$.

На этом основании полное изменение объема при растворении равно

$$\Delta V = \Delta V_{\text{фп}} + \Delta V_{\text{вз}}. \quad (2.3.2)$$

В отличие от этого, для получения полного изменения энтропии ΔS необходимо, кроме двух вкладов $\Delta S_{\text{вз}}$ и $\Delta S_{\text{фп}}$, учесть дополнительно также энтропию смешения $\Delta S_{\text{см}}$:

$$\Delta S = \Delta S_{\text{фп}} + \Delta S_{\text{вз}} + \Delta S_{\text{см}}. \quad (2.3.3)$$

Энтропия смешения $\Delta S_{\text{см}}$ раствора аналогична величине (1.12.21), полученной для газов, и учитывает хаотическое перераспределение каждого компонента по всему объему раствора в результате смешивания. Этот эффект увеличивает разупорядочение в системе и, следовательно, по аналогии с (1.12.21) всегда имеем

$$\Delta S_{\text{см}} > 0. \quad (2.3.4)$$

Как ясно из вышесказанного, знаки $\Delta H_{\text{вз}}$, $\Delta V_{\text{вз}}$ и $\Delta S_{\text{вз}}$ определяются конкуренцией между процессами ассоциации и сольватации молекул. Для слабо ассоциированных растворителей (с малополярными молекулами) преобладание сольватации повышает ближнюю упорядоченность в растворах с одновременным уменьшением объема и выделением тепла, так что

$$\Delta H_{\text{вз}} < 0, \quad \Delta V_{\text{вз}} < 0, \quad \Delta S_{\text{вз}} < 0. \quad (2.3.5)$$

Эти же неравенства могут иметь место и для сильно ассоциированных жидкостей с полярными молекулами (какой является вода), если они взаимодействуют с молекулами растворяемых веществ за счет более сильных донорно-акцепторных или водородных связей. В этом случае сольватация также преобладает над ассоциацией.

Для сильно ассоциированных жидкостей возможна и противоположная ситуация в случае, если взаимодействие между молекулами растворителя сильнее их взаимодействия с малополярными молекулами растворяемых веществ. Тогда ассоциация преобладает над сольватацией, т. е. имеют место неравенства, противоположные аналогичным в формуле (2.3.5):

$$\Delta H_{\text{вз}} > 0, \quad \Delta V_{\text{вз}} > 0, \quad \Delta S_{\text{вз}} > 0. \quad (2.3.6)$$

Эти неравенства выражают поглощение тепла, увеличение объема и энтропии, порожденные сильным взаимодействием по-

лярных молекул, приводящим к их ассоциации за счет диполь-дипольных, водородных или донорно-акцепторных связей.

Каждое из выражений (2.3.1)–(2.3.3) содержит дополнительное слагаемое с индексом «фп», вызванное переходом растворяемых частиц из одной фазы в другую. Величины $\Delta H_{\text{фп}}$, $\Delta \tilde{V}_{\text{фп}}$ и $\Delta S_{\text{фп}}$ определяются типом фазового перехода при взаимном растворении компонентов с различным исходным агрегатным состоянием.

Фазовые переходы первого рода, такие как плавление, испарение, сублимация и обратные им, были рассмотрены в п. 2.1. Там показано, что фазовые превращения, происходящие при увеличении температуры, всегда сопровождаются положительными скачками энтальпии и энтропии, как показано на рис. 2.3. Точки плавления $T_{\text{пл}}$ и кипения $T_{\text{кип}}$ разделяют три агрегатные состояния вещества. Именно в этих точках наблюдаются скачкообразные изменения H и S , равные теплотам и энтропиям плавления ($\Delta H_{\text{пл}} > 0$ и $\Delta S_{\text{пл}} = \Delta H_{\text{пл}}/T_{\text{пл}} > 0$) и испарения ($\Delta H_{\text{исп}} > 0$ и $\Delta S_{\text{исп}} = \Delta H_{\text{исп}}/T_{\text{кип}} > 0$). В промежутке между точками фазового перехода температурный ход кривых $H(T)$ и $S(T)$ определяется формулами (1.5.11).

Оценим соотношение между энтальпийным и энтропийным факторами для трех различных случаев растворения с изменением агрегатного состояния растворяемого вещества.

Теплоты плавления ($\Delta H_{\text{пл}}$) и испарения ($\Delta H_{\text{исп}}$) по порядку величины составляют, соответственно, единицы и десятки ккал/моль, что, как правило, превышает $\Delta H_{\text{вз}}$. Следовательно, тепловой эффект (2.3.1) при растворении веществ с изменением их агрегатного состояния определяется, в основном, теплотой фазового перехода $\Delta H_{\text{фп}}$. Что касается энтропии растворения, то из трех слагаемых, входящих в выражение (2.3.3), величина $\Delta S_{\text{вз}}$ также является наименьшей. Поэтому результирующее значение энтропии растворения ΔS определяется соотношением между величинами $\Delta S_{\text{фп}}$ и $\Delta S_{\text{см}}$, при этом, согласно (2.3.4), всегда $\Delta S_{\text{см}} > 0$.

Растворение твердых тел в жидкостях сопровождается разрушением кристаллической решетки. Согласно рис. 2.3, для фазового перехода твердое тело $\rightarrow$ жидкость имеем $\Delta H^{\text{т} \rightarrow \text{ж}} = \Delta H_{\text{пл}} > 0$ и $\Delta S^{\text{т} \rightarrow \text{ж}} = \Delta S_{\text{пл}} > 0$. Следовательно, растворение твердых тел в жидкостях является *эндотермическим* процессом ($\Delta H \approx \Delta H_{\text{пл}} > 0$) и сопровождается ростом энтропии ($\Delta S \approx \Delta S_{\text{пл}} + \Delta S_{\text{см}} > 0$). Отсюда следует вывод о том, что при низких температурах, когда преобладает энтальпийный фактор

($\Delta H > T\Delta S$), растворение практически отсутствует, а с ростом температуры возрастает роль энтропийного фактора, что обеспечивает растворение ($\Delta G = \Delta H - T\Delta S < 0$).

Растворение газов в жидкостях, если оно происходит в молекулярной форме (без диссоциации молекул на атомы и ионы), сопровождается конденсацией газовых молекул. Для фазового перехода газ $\rightarrow$ жидкость всегда имеем $\Delta H^{\text{г} \rightarrow \text{ж}} = -\Delta H_{\text{исп}} < 0$ и $\Delta S^{\text{г} \rightarrow \text{ж}} = -\Delta S_{\text{исп}} < 0$ (см. рис. 2.3). Следовательно, молекулярное растворение газов в жидкостях является *экзотермическим* процессом ($\Delta H \approx -\Delta H_{\text{исп}} < 0$). Возможность его осуществления зависит от энтропии растворения, которая может иметь разный знак ($\Delta S \approx -\Delta S_{\text{исп}} + \Delta S_{\text{см}} >$ или < 0 , при этом всегда $\Delta S_{\text{см}} > 0$). В случае $\Delta S_{\text{см}} > \Delta S_{\text{исп}}$, получаем $\Delta S > 0$, так что газ растворяется в жидкостях при любых температурах, поскольку всегда $\Delta G = -|\Delta H| - T\Delta S < 0$. В тех случаях, если $\Delta S < 0$, растворение газа может прекратиться при высоких температурах, когда $\Delta G = -|\Delta H| + T|\Delta S| > 0$.

Растворение газов в твердых телах может происходить как в молекулярной форме (без диссоциации молекул), так и в атомарной форме (с диссоциацией молекул на атомы).

Молекулярное растворение газов характерно для диэлектрических силикатных материалов типа стекол и керамики, когда газовые молекулы застревают в пустотах силикатной сетки по механизму «молекулярного сита». Этому механизму присущи рассмотренные выше особенности молекулярного растворения газов в жидкостях, в частности выделение тепла ($\Delta H < 0$).

Атомарное растворение газов характерно для металлов. Так как диссоциация молекул на атомы всегда требует подвода энергии, то теплота процесса (2.3.1), в дополнение к двум вкладам, должна содержать третье слагаемое — теплоту диссоциации $\Delta H_{\text{дис}} > 0$:

$$\Delta H = \Delta H_{\text{фп}} + \Delta H_{\text{вз}} + \Delta H_{\text{дис}}. \quad (2.3.7)$$

В отличие от жидких растворов, вклад $\Delta H_{\text{вз}}$, учитывающий химическое взаимодействие растворенных атомов с решеткой, может быть сравним с двумя другими слагаемыми в выражении (2.3.7). При этом обязательно $\Delta H_{\text{вз}} < 0$, поскольку силы притяжения, обеспечивающие это взаимодействие, уменьшают потенциальную энергию системы, как показано ниже на рис. 2.5. Согласно рис. 2.3, фазовый переход газ $\rightarrow$ твердое тело всегда является *экзотермическим* процессом, тогда $\Delta H_{\text{фп}} = \Delta H^{\text{г} \rightarrow \text{т}} \approx -(\Delta H_{\text{исп}} + \Delta H_{\text{пл}}) < 0$. Результирующий тепловой эффект определяется соотношением между двумя величинами

разных знаков: $\Delta H_{\text{дис}} > 0$ и $(\Delta H_{\text{фп}} + \Delta H_{\text{вз}}) < 0$. Принципиально возможны следующие две физические ситуации.

1. Растворение газов в обычных *конструкционных материалах* (таких, как Fe, Cu, Ni, Mo, W и др., в том числе их сплавах) происходит с образованием *твердых растворов* внедрения или замещения. В этом случае энергия связи растворенных атомов с решеткой относительно невелика, что обеспечивает $|\Delta H_{\text{фп}} + \Delta H_{\text{вз}}| < \Delta H_{\text{дис}}$, то есть растворение носит *эндотермический характер* ($\Delta H > 0$ — поглощение тепла). Для осуществления процесса растворения необходимо иметь $\Delta S > 0$, так как в противном случае $\Delta G = \Delta H + T|\Delta S| > 0$. Следовательно, при достаточно низких температурах, когда вклад энтропийного фактора $T\Delta S$ уменьшается, растворение газа в конструкционных металлах может прекратиться.

2. Растворение газов в активных *газопоглощающих материалах* (таких, как Ti, Ta, Zr, Nb и др.) происходит с образованием в твердой фазе *химических соединений* (типа гидридов, нитридов, оксидов). Энергия связи атомов газа в химических соединениях достаточно велика для того, чтобы обеспечить $|\Delta H_{\text{фп}} + \Delta H_{\text{вз}}| > \Delta H_{\text{дис}}$, то есть растворение носит *экзотермический характер* ($\Delta H < 0$ — выделение тепла). С точки зрения возможности осуществления процесса здесь ситуация аналогична той, что имеет место при растворении газов в жидкостях.

Знание полных величин ΔH и ΔS позволяет в изобарных условиях выяснить возможность образования раствора из чистых компонентов по знаку изобарного потенциала $\Delta G = \Delta H - T\Delta S$. При $\Delta G < 0$ процесс растворения идет самопроизвольно и раствор остается ненасыщенным. Условие равновесия $\Delta G = 0$ соответствует *насыщенному раствору*, в котором присутствует избыток растворяемых веществ в виде осадка. Концентрация растворенного компонента в насыщенном (равновесном) растворе называется *растворимостью* данного вещества. Она зависит от температуры и давления (а также от присутствия других компонентов в растворе) и выражается в виде функциональной зависимости $c_i = f(T, P)$ или $x_i = f(T, P)$.

Парогазовая среда над раствором, находящаяся с ним в равновесии, называется *насыщенной*, а парциальное давление каждого компонента в газе называют *давлением насыщенного пара* этого компонента. Давление насыщенного пара в общем случае может зависеть как от температуры, так и от давления, $p_i = f(T, P)$, хотя обычно влияние последнего пренебрежимо мало (см. пп. 2.2 и 2.5).

Нахождение зависимостей $c_i = f(T, P)$ и $p_i = f(T, P)$ составляет основную задачу термодинамического анализа растворов. Эти зависимости следуют из условий фазового и химического равновесия. Анализ физических закономерностей, вытекающих из этих условий, и выяснение принципиальных возможностей для управления ими в технологических процессах будут рассмотрены ниже, а также в гл. 3.

2.4. Модели и термодинамические свойства растворов

В общем случае процесс образования растворов из чистых компонентов сопровождается тепловым эффектом ΔH , изменениями объема ΔV , энтропии ΔS и энергии Гиббса ΔG .

Ниже будем рассматривать изобарно-изотермическое смешение чистых компонентов без изменения их агрегатного состояния, при котором справедливы следующие равенства (ср. формулы (2.3.1)–(2.3.3) при $\Delta H_{\text{фп}} = 0$):

$$\begin{aligned}\Delta H &= \Delta H_{\text{вз}}, & \Delta S &= \Delta S_{\text{вз}} + \Delta S_{\text{см}}, \\ \Delta V &= \Delta V_{\text{вз}}, & \Delta G &= \Delta G_{\text{вз}} + \Delta G_{\text{см}}.\end{aligned}\tag{2.4.1}$$

Здесь величины с индексом «вз» отражают взаимодействие между частицами в растворе, а величины $\Delta S_{\text{см}}$ и $\Delta G_{\text{см}}$ суть энтропия смешения и изобарный потенциал смешения компонентов.

Запишем приращения (2.4.1), используя парциальные величины, введенные формулами (1.11.3), (1.11.9), (1.11.11) и (1.11.12):

$$\Delta H = H - H_0 = \sum_i n_i (\bar{H}_i - H_i^0),\tag{2.4.2}$$

$$\Delta V = V - V_0 = \sum_i n_i (\bar{V}_i - V_i^0),\tag{2.4.3}$$

$$\Delta S = S - S_0 = \sum_i n_i (\bar{S}_i - S_i^0),\tag{2.4.4}$$

$$\Delta G = G - G_0 = \sum_i n_i (\mu_i - \mu_i^0).\tag{2.4.5}$$

Здесь, по аналогии с формулами (1.12.15)–(1.12.17), нижние нулевые индексы отвечают исходному состоянию чистых веществ до растворения, а верхние нули обозначают соответствующие им молярные величины (1.12.14).

Многообразие и сложность взаимодействий между частицами в реальных растворах делают невозможным их строгое описание. Поэтому в термодинамике растворов обычно используют

формальный прием путем введения специальной величины a_i , называемой **активностью** ¹⁾ i -го компонента. Это позволяет сохранить для химического потенциала форму записи (1.13.3), полученную в рамках модели идеального газа, заменив в ней мольную долю x_i компонента на его активность $a_i = \gamma_i x_i$:

$$\mu_i = \mu_i^\circ(T, P) + RT \ln a_i = \mu_i^{\text{ид}} + RT \ln \gamma_i. \quad (2.4.6)$$

Здесь γ_i — коэффициент активности, отражающий взаимодействие между частицами в растворе и зависящий в общем случае от его состава и температуры. Этот коэффициент определяет избыточный химический потенциал i -го компонента,

$$\Delta\mu_i^{\text{изб}} \equiv \mu_i - \mu_i^{\text{ид}} = RT \ln \gamma_i, \quad (2.4.7)$$

по отношению к химическому потенциалу идеальной системы,

$$\mu_i^{\text{ид}} = \mu_i^\circ(T, P) + RT \ln x_i, \quad (2.4.8)$$

а $\mu_i^\circ \equiv \mu_i^0 = H_i^0 - TS_i^0$ означает химический потенциал чистого i -го компонента.

По знаку избыточного химического потенциала принято различать:

- *положительные растворы*, для которых $\Delta\mu_i^{\text{изб}} > 0$ и $\gamma_i > 1$, т. е. активность компонента больше его мольной доли ($a_i > x_i$);
- *отрицательные растворы*, для которых $\Delta\mu_i^{\text{изб}} < 0$ и $\gamma_i < 1$, т. е. активность компонента меньше его мольной доли ($a_i < x_i$).

Пользуясь формулой (1.11.15), можно выразить избыточный химический потенциал $\Delta\mu_i^{\text{изб}}$ через избыточные значения для парциальных величин ²⁾:

$$\Delta\mu_i^{\text{изб}} = (\bar{H}_i - \bar{H}_i^{\text{ид}}) - T(\bar{S}_i - \bar{S}_i^{\text{ид}}) \equiv \Delta\bar{H}_i^{\text{изб}} - T\Delta\bar{S}_i^{\text{изб}}. \quad (2.4.9)$$

Из формул (2.4.7) и (2.4.9) получаем общее выражение для коэффициента активности:

$$\ln \gamma_i = \frac{\Delta\bar{H}_i^{\text{изб}}}{RT} - \frac{\Delta\bar{S}_i^{\text{изб}}}{R}. \quad (2.4.10)$$

¹⁾ По аналогии с активностью $a_i = \gamma_i x_i$, для реальных газов вводят так называемую *фугитивность* (летучесть) $f_i = \gamma_i p_i$, сохраняя при этом химический потенциал в форме (1.13.1) с заменой p_i на f_i . Однако в дальнейшем будем считать газовую фазу идеальной ($\gamma_i = 1$), что, как правило, достаточно хорошо соответствует действительности и позволяет вместо фугитивностей работать с парциальными давлениями газообразных компонентов.

²⁾ Здесь избыток энтропии $\Delta\bar{S}_i^{\text{изб}}$ следует отличать от избыточной энтропии $S_{\text{ex}} = \delta^2 S/2$, ранее введенной формулой (1.9.24).

Таким образом, отклонение реального раствора от идеальных систем определяется двумя факторами — *энтальпийным*, $\Delta \bar{H}_i^{\text{изб}} = \bar{H}_i - \bar{H}_i^{\text{ид}}$, и *энтропийным*, $\Delta \bar{S}_i^{\text{изб}} = \bar{S}_i - \bar{S}_i^{\text{ид}}$. Путем последовательного учета этих факторов строят модели растворов, для которых справедливы следующие соотношения:

- *модель идеального раствора*

$$\Delta \bar{H}_i^{\text{изб}} = 0, \quad \Delta \bar{S}_i^{\text{изб}} = 0, \quad \ln \gamma_i = 0; \quad (2.4.11)$$

- *модель атермального раствора*

$$\Delta \bar{H}_i^{\text{изб}} = 0, \quad \Delta \bar{S}_i^{\text{изб}} \neq 0, \quad \ln \gamma_i = -\Delta \bar{S}_i^{\text{изб}}/R; \quad (2.4.12)$$

- *модель регулярного раствора*

$$\Delta \bar{H}_i^{\text{изб}} \neq 0, \quad \Delta \bar{S}_i^{\text{изб}} = 0, \quad \ln \gamma_i = \Delta \bar{H}_i^{\text{изб}}/RT. \quad (2.4.13)$$

Рассмотрим термодинамические свойства, присущие растворам трех вышеуказанных типов.

1. Идеальные растворы. Модель идеального раствора является простейшим приближением к реальности, сильно упрощающим термодинамическое описание растворов. Согласно условиям (2.4.11), к идеальным растворам применимы все термодинамические соотношения, полученные для идеальных газов, тогда в соответствии с формулами (1.12.15), (1.12.17), (1.12.20) и (1.12.21) имеем:

$$\Delta H = H - H_0 = \sum_i n_i (\bar{H}_i^{\text{ид}} - H_i^0) = 0, \quad (2.4.14)$$

$$\Delta V = V - V_0 = \sum_i n_i (\bar{V}_i^{\text{ид}} - V_i^0) = 0, \quad (2.4.15)$$

$$\begin{aligned} \Delta S = S - S_0 &= \sum_i n_i (\bar{S}_i^{\text{ид}} - S_i^0) = \\ &= -nR \sum_i x_i \ln x_i \equiv \Delta S_{\text{см}} > 0, \end{aligned} \quad (2.4.16)$$

$$\begin{aligned} \Delta G = G - G_0 &= \sum_i n_i (\mu_i^{\text{ид}} - \mu_i^0) = \\ &= nRT \sum_i x_i \ln x_i \equiv \Delta G_{\text{см}} < 0. \end{aligned} \quad (2.4.17)$$

Таким образом, образование идеального раствора происходит *самопроизвольно* ($\Delta G < 0$) путем *необратимого* ($\Delta S > 0$) перераспределения растворяемых компонентов по всему объему

раствора, что *не сопровождается* тепловым эффектом ($\Delta H = 0$) и объемным эффектом ($\Delta V = 0$).

Для идеальных растворов, в отличие от идеальных газов, нельзя пренебречь потенциальной энергией взаимодействия между молекулами по сравнению с кинетической энергией их теплового движения. Приближение идеальности в данном случае означает пренебрежимо малое различие в межмолекулярных взаимодействиях между однородными и разнородными молекулами в растворе. Для этого разнородные молекулы (растворителя и растворенного вещества) должны иметь примерно одинаковые размеры и близкое химическое строение. Подобными свойствами обладают, например, жидкие растворы бензола и толуола, метилового и этилового спиртов, твердые растворы изотопов различных веществ, смеси изомерных углеводородов и т. п.

2. Атермальные растворы. Модель атермального раствора является более общей по сравнению с моделью идеального раствора. Согласно (2.4.12), для атермальных растворов, как и для идеальных, отсутствует теплота образования ($\Delta H = 0$).

Отличие между ними заключается, во-первых, в существовании объемного эффекта при атермальном растворении:

$$\Delta V = V - V_0 = \sum_i n_i (\bar{V}_i - V_i^0) > 0. \quad (2.4.18)$$

Во-вторых, ненулевая избыточная энтропия, $\Delta \bar{S}_i^{\text{изб}} = \bar{S}_i - \bar{S}_i^{\text{ид}}$, атермальных растворов обеспечивает дополнительный к энтропии смешения вклад, равный $\Delta \bar{S}_{\text{вз}} = \sum_i n_i \Delta \bar{S}_i^{\text{изб}}$, так что

$$\begin{aligned} \Delta S &= S - S_0 = \sum_i n_i (\bar{S}_i - S_i^0) = \\ &= \sum_i n_i (\bar{S}_i^{\text{ид}} - S_i^0) + \sum_i n_i \Delta \bar{S}_i^{\text{изб}} = \\ &= -nR \sum_i x_i \ln x_i + \sum_i n_i \Delta \bar{S}_i^{\text{изб}} \equiv \Delta S_{\text{см}} + \Delta S_{\text{вз}}. \end{aligned} \quad (2.4.19)$$

Атермальные растворы возникают при смешивании неполярных жидкостей с сильно различающимися по размерам молекулами. Сюда относятся, в частности, растворы высокомолекулярных веществ в обычных растворителях. Практически нулевая теплота образования этих растворов связана с тем, что обычно молекулы растворителя близки по размерам и молекулярной природе к звеньям полимерных цепей. В результате этого энергия взаимодействия между однородными и разнородными молекулами одного порядка по величине. Разветвленность и гибкость

длинных полимерных цепей приводит к высокой степени неупорядоченности в их пространственном размещении. В результате этого $\Delta S_{\text{см}} > 0$ и $\Delta \bar{S}_i^{\text{изб}} > 0$. Большие положительные значения энтропии растворения ($\Delta S = \Delta S_{\text{см}} + \Delta S_{\text{вз}} > 0$) обеспечивают самопроизвольное протекание этого процесса в силу того, что $\Delta G = -T\Delta S < 0$.

Атермальные растворы относятся к отрицательным растворам ($\gamma_i < 1$), так как при $\Delta \bar{S}_i^{\text{изб}} > 0$ из выражения (2.4.12) следует, что $\ln \gamma_i = -\Delta \bar{S}_i^{\text{изб}}/R < 0$.

3. Регулярные растворы. Модель регулярного раствора соответствует противоположному предельному случаю по отношению к модели атермального раствора. Согласно (2.4.13), регулярные растворы близки по свойствам к идеальным растворам из-за отсутствия избыточной энтропии ($\Delta \bar{S}_i^{\text{изб}} \equiv \bar{S}_i - \bar{S}_i^{\text{ид}} = 0$), т. е. энтропия их образования определяется только энтропией смешения:

$$\begin{aligned} \Delta S &= S - S_0 = \sum_i n_i (\bar{S}_i - S_i^0) = \sum_i n_i (\bar{S}_i^{\text{ид}} - S_i^0) = \\ &= -nR \sum_i x_i \ln x_i \equiv \Delta S_{\text{см}} > 0. \end{aligned} \quad (2.4.20)$$

Как и для идеальных растворов, в модели регулярного раствора пренебрегают изменением молярного объема при растворении каждого компонента ($\bar{V}_i = V_i^0$), тогда

$$\Delta V = V - V_0 = \sum_i n_i (\bar{V}_i - V_i^0) = 0. \quad (2.4.21)$$

Результат взаимодействия разнородных частиц в регулярном растворе проявляется в виде избыточной энтальпии $\Delta \bar{H}_i^{\text{изб}} = \bar{H}_i - \bar{H}_i^{\text{ид}}$, дающей основной вклад в тепловой эффект:

$$\begin{aligned} \Delta H &= H - H_0 = \sum_i n_i (\bar{H}_i - H_i^0) = \sum_i n_i (\bar{H}_i^{\text{ид}} - H_i^0) + \\ &+ \sum_i n_i \Delta \bar{H}_i^{\text{изб}} = \sum_i n_i \Delta \bar{H}_i^{\text{изб}} \equiv \Delta H_{\text{вз}}. \end{aligned} \quad (2.4.22)$$

Здесь учтено, что для идеального раствора теплота образования равна нулю, так как $\bar{H}_i^{\text{ид}} = H_i^0$. Результирующий знак теплоты растворения (2.4.22) определяется соотношением избыточных энтальпий $\Delta \bar{H}_i^{\text{изб}}$ для компонентов раствора, которые, как увидим ниже, могут иметь разный знак.

Теория регулярных растворов строится на основе учета взаимодействия только между ближайшими соседями. Зависимость потенциальной энергии u_{ij} взаимодействия двух частиц (с номерами i и j) от расстояния r_{ij} между ними качественно показана на рис. 2.5. Устойчивому расположению частиц на расстоянии r_{ij}^0 соответствует потенциальный минимум с энергией $u_{ij}^0 < 0$. Величина u_{ij}^0 определяет избыточную энтальпию $\Delta \bar{H}_i^{\text{изб}}$ и коэффициент активности γ_i или знак раствора, определяемый как знак $\ln \gamma_i$. Покажем это на примере бинарного раствора, состоящего из компонентов А и В.

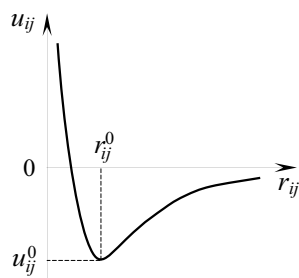


Рис. 2.5. Зависимость потенциальной энергии u_{ij} взаимодействия двух частиц с номерами i и j от расстояния r_{ij} между ними

Если для идеальных растворов полностью пренебрегают различием в размерах, форме и характере взаимодействия частиц, то в модели регулярного раствора следует учитывать эти факторы. Для обеспечения соотношений (2.4.20) и (2.4.21), при- сущих как идеальному, так и регулярному растворам, необходимо потребовать, чтобы указанные различия были невелики. В этом случае замена однородных пар взаимодействующих частиц (А–А и В–В) на смешанные пары (А–В) не нарушает их хаотического пространственного распределения, что важно для нахождения энтропии смешения.

Образование смешанной связи А–В с энергией $u_{AB}^0 < 0$ происходит в результате разрыва двух однородных связей А–А и В–В, обладающих энергиями $u_{AA}^0 < 0$ и $u_{BB}^0 < 0$. Энергия взаимодействия, отнесенная к одной связи А–В, в этом случае равняется $u_{AB}^0 - (u_{AA}^0 + u_{BB}^0)/2$.

В пересчете на один моль раствора получаем энергию взаимообмена U_{AB} (называемую иногда параметром взаимодействия и обозначаемую α_{AB}), которая пропорциональна изменению энергии частицы при замене всех ее соседей частицами другого сорта, а именно:

$$\begin{aligned} U_{AB} &= z\mathcal{N}_A \left(u_{AB}^0 - \frac{u_{AA}^0 + u_{BB}^0}{2} \right) = \\ &= z\mathcal{N}_A \left(\frac{|u_{AA}^0| + |u_{BB}^0|}{2} - |u_{AB}^0| \right). \end{aligned} \quad (2.4.23)$$

Здесь $\mathcal{N}_d$ — постоянная Авогадро, z — координационное число, равное числу ближайших соседей (максимальное значение $z = 12$ достигается при плотной упаковке частиц одинакового размера).

Следовательно, тепловой эффект при образовании одного моля бинарного раствора, имеющего состав $x_A = n_A/(n_A + n_B) = 1 - x_B$, равен $x_A x_B U_{AB}$.

Отсюда получаем энтальпию образования раствора с общим числом молей $n = n_A + n_B$:

$$\Delta H = n x_A x_B U_{AB} = \frac{n_A n_B}{n_A + n_B} U_{AB}. \quad (2.4.24)$$

Из равенства (2.4.22) следует, что

$$\Delta \bar{H}_i^{\text{изб}} = \frac{\partial \Delta H}{\partial n_i}. \quad (2.4.25)$$

Применяя (2.4.25) к выражению (2.4.24), из формулы (2.4.13) получаем следующие соотношения:

$$\ln \gamma_A = \frac{\Delta \bar{H}_A^{\text{изб}}}{RT} = x_B^2 \frac{U_{AB}}{RT} = (1 - x_A)^2 \frac{U_{AB}}{RT}, \quad (2.4.26)$$

$$\ln \gamma_B = \frac{\Delta \bar{H}_B^{\text{изб}}}{RT} = x_A^2 \frac{U_{AB}}{RT} = (1 - x_B)^2 \frac{U_{AB}}{RT}. \quad (2.4.27)$$

Как видно из выражений (2.4.23) и (2.4.24), знак теплоты образования раствора зависит от соотношения между энергиями взаимодействия однородных и разнородных частиц, входящих в состав раствора:

а) если взаимодействие однородных частиц в растворе сильнее, чем разнородных, т. е. $|u_{AA}^0 + u_{BB}^0| > 2|u_{AB}^0|$, тогда $U_{AB} > 0$ и $\Delta H > 0$, что дает *поглощение тепла*;

б) если преобладает взаимодействие между разнородными частицами, т. е. $|u_{AA}^0 + u_{BB}^0| < 2|u_{AB}^0|$, тогда $U_{AB} < 0$ и $\Delta H < 0$, что дает *выделение тепла*.

Из формул (2.4.26) и (2.4.27) следует, что в первом случае регулярный раствор является *положительным* по обоим компонентам ($\gamma_A > 1$ и $\gamma_B > 1$), а во втором случае — *отрицательным* ($\gamma_A < 1$ и $\gamma_B < 1$). Следовательно, знак регулярного раствора всегда совпадает со знаком теплового эффекта и, как увидим ниже, со знаком отклонения от закона Рауля (см. п. 2.5).

Для идеальных растворов энергии взаимодействия однородных и разнородных частиц равны, так что $U_{AB} = 0$, $\Delta H = 0$ и $\gamma_A = \gamma_B = 1$.

В технологии полупроводниковых приборов и интегральных микросхем в настоящее время применяются материалы высокой степени очистки, в которых содержание примесей на 5–10 порядков меньше количества атомов основного вещества. Такие твердые растворы являются *предельно разбавленными* и близки по свойствам к модели идеального раствора.

Полагаем, что компонент А в бинарном растворе является растворителем, для которого $x_A \approx 1$, а компонент В — малой примесью ($x_B \ll 1$). Тогда из выражений (2.4.26) и (2.4.27) получаем $\gamma_A = 1$ и $RT \ln \gamma_B = \Delta \bar{H}_B^{\text{изб}} = U_{AB} = \text{const}$, где константа не зависит ни от температуры, ни от состава раствора. Полученный результат можно обобщить на многокомпонентные предельно разбавленные растворы, приписывая растворителю индекс $i = 1$, а примесям — индекс $i \neq 1$. В этом случае

$$\gamma_1 = 1 \quad \text{и} \quad RT \ln \gamma_i = \Delta \bar{H}_i^{\text{изб}} = U_{1i} = \text{const}. \quad (2.4.28)$$

Таким образом, в предельно разбавленных растворах растворитель подчиняется модели идеального раствора, в частности

$$\mu_1 = \mu_1^\circ + RT \ln x_1. \quad (2.4.29)$$

Для примесей (отмечаемых индексом $i \neq 1$) из выражений (2.4.6), (2.4.8) и (2.4.28) получаем следующее выражение для химического потенциала,

$$\begin{aligned} \mu_i &= \mu_i^{\text{ид}} + RT \ln \gamma_i = \\ &= (\mu_i^\circ + U_{1i}) + RT \ln x_i \equiv \mu_i^* + RT \ln x_i, \end{aligned} \quad (2.4.30)$$

которое по виду совпадает с (2.4.29). Однако стандартное значение химического потенциала в выражении (2.4.30) отличается от μ_i° , так как в стандартном состоянии ($x_i = 1$) имеем $\mu_i = \mu_i^*$, при этом $\mu_i^* = \mu_i^\circ(T, P) + U_{1i}(u_{11}^0, u_{1i}^0, u_{ii}^0)$.

Следовательно, в рамках модели регулярного раствора химический потенциал i -й примеси в стандартном состоянии вещества определяется как свойствами чистого вещества (учитываемыми величиной μ_i°), так и взаимодействием примеси с растворителем (учитываемым величиной U_{1i}).

Тот факт, что величина U_{1i} не зависит ни от температуры T , ни от концентраций x_i , приближает описание примесей в предельно разбавленном растворе с помощью химического потенциала в форме (2.4.30) к описанию идеального раствора. И это несмотря на то, что для них коэффициент активности $\gamma_i \neq 1$

и, как следует из выражения (2.4.28), экспоненциально зависит от температуры:

$$\gamma_i(T) = \exp\left(\frac{U_{1i}}{RT}\right) = \exp\left(\frac{\Delta\bar{H}_i^{\text{изб}}}{RT}\right). \quad (2.4.31)$$

Величина γ_i при данной температуре является константой, не зависящей от присутствия других примесей в растворе. Следовательно, активность любой примеси в предельно разбавленном растворе пропорциональна ее мольной доле: $a_i = \gamma_i(T)x_i$.

Как видно из выражения (2.4.31), температурный ход коэффициента активности $\gamma_i(T)$ определяется знаком теплоты $\Delta\bar{H}_i^{\text{изб}} \propto \Delta H$: для положительных растворов (когда $\Delta H > 0$) величина γ_i с ростом температуры монотонно уменьшается от бесконечно больших значений до единицы, а для отрицательных растворов (когда $\Delta H < 0$) — увеличивается от нуля до единицы. Так как в обоих случаях $\gamma_i \rightarrow 1$ при $T \rightarrow \infty$, то предельно разбавленные растворы при возрастании температуры приближаются к идеальному раствору, для которого $\Delta G < 0$, согласно (2.4.17). Этот результат означает, что в природных условиях все вещества самопроизвольно загрязняются естественными примесями с образованием твердых растворов, близких по свойствам к модели предельно разбавленного или идеального раствора.

Теперь, вооружившись общими представлениями из термодинамики растворов, перейдем к анализу процессов формирования газовой фазы над растворами, растворимости газов в конденсированных средах и равновесных закономерностей кристаллизации/плавления в бинарных системах.

2.5. Давление насыщенных паров над растворами и смесями веществ в конденсированном состоянии

Если при определенных условиях взаимная растворимость веществ в конденсированном (твердом или жидком) состоянии пренебрежимо мала, то их одновременное сосуществование образует смесь этих веществ. Она представляет собой многофазную (по числу веществ) систему, например:

а) *поликристалл* как смесь в твердом состоянии кристаллов чистого вещества А (или твердого раствора α) и чистого вещества В (или твердого раствора β);

б) *расслаивающаяся жидкость* как смесь взаимно нерастворимых жидкостей;

в) *осадок* нерастворимого твердого вещества в жидкости.

В последних двух случаях при высокой степени дисперсности (измельчения веществ) они образуют соответственно *эмульсии* и *суспензии* (см. подробнее на с. 613).

В отсутствие взаимной растворимости термодинамические свойства компонентов в смеси остаются такими же, как и для чистых веществ. Так, парциальное давление насыщенного пара $p_i^0(T)$ для i -го компонента над смесью описывается формулой (2.2.11). С смесь взаимно нерастворимых жидкостей (эмульсия) закипает при той температуре, при которой суммарное давление паров, формируемое испарением всех ее компонентов, равняется внешнему давлению P (в частности $P = 1$ атм). Иными словами, температура кипения смеси $T_{\text{кип}}^{\text{см}}$ находится из равенства

$$\sum_i p_i^0(T_{\text{кип}}^{\text{см}}) = P = 1 \text{ атм.} \quad (2.5.1)$$

Графическое решение уравнения (2.5.1) для бинарной системы из взаимно нерастворимых жидкостей А и В качественно показано на рис. 2.6. Очевидно, что для смеси температура $T_{\text{кип}}^{\text{см}}$ ниже температур кипения $T_{\text{А,кип}}^0$ и $T_{\text{В,кип}}^0$ чистых компонентов. Это свойство используется на практике для перегонки слабо летучей жидкости В путем ее совместного кипения с более летучей жидкостью А.

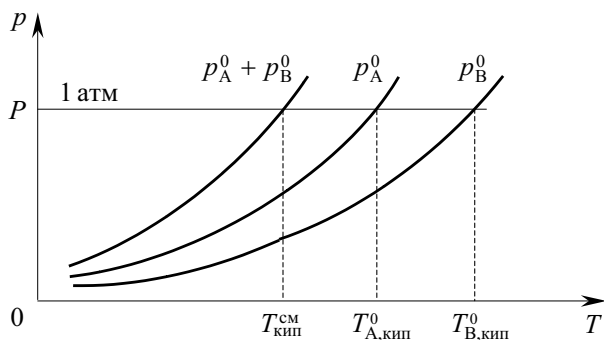


Рис. 2.6. Графическое нахождение температуры кипения $T_{\text{кип}}^{\text{см}}$ для бинарной смеси взаимно нерастворимых жидкостей А и В

Рассмотрим давление насыщенных паров над твердыми и жидкими растворами. Первоначально будем считать конденсированную фазу идеальным раствором известного состава x_i , где $i = 1, 2, \dots, K$. Задача заключается в нахождении зависимости парциального давления p_i насыщенного пара для каждого i -го компонента от температуры, давления и состава раствора.

Условие фазового равновесия имеет тот же вид (2.2.1), что и для чистого вещества. Однако в данном случае имеем $p_i \neq p_i^0$ и $x_i \neq 1$, так что вместо равенств (2.2.2) и (2.2.3) работают общие выражения (1.13.1) и (1.13.3):

$$\mu_i^{\text{п}}(T, P) = \mu_i^{\text{оп}}(T) + RT \ln p_i(T, P), \quad (2.5.2)$$

$$\mu_i^{\text{к}}(T, x_i) = \mu_i^{\text{ок}}(T) + RT \ln x_i. \quad (2.5.3)$$

Парциальное давление $p_i(T, P)$ в формуле (2.5.2) рассматриваем как искомую функцию температуры T и давления P . В силу равенства (2.2.10) в выражении (2.5.3) опущена зависимость стандартного химического потенциала $\mu_i^{\text{ок}}(T, P)$ от давления P (именно это приведет ниже к отсутствию зависимости p_i от P).

Подстановка (2.5.2) и (2.5.3) в условие фазового равновесия (2.2.1) дает выражение

$$p_i = x_i f(T), \quad \text{где} \quad f(T) = \exp\left(\frac{\mu_i^{\text{ок}}(T) - \mu_i^{\text{оп}}(T)}{RT}\right). \quad (2.5.4)$$

Поскольку при $x_i = 1$ имеем дело с чистым веществом, для которого $p_i \equiv p_i^0$, то получаем $f(T) = p_i^0(T)$. Тогда равенство (2.5.4) принимает форму **закона Рауля**¹⁾,

$$p_i(T, x_i) = x_i p_i^0(T), \quad (2.5.5)$$

который гласит: *парциальное давление p_i равновесного пара i -го компонента над идеальным раствором равно давлению p_i^0 насыщенного пара этого вещества в чистом виде, умноженному на его мольную долю x_i в растворе.*

Искомая температурная зависимость $p_i(T)$ для раствора определяется выражением (2.2.11), дающим зависимость $p_i^0(T)$ для чистого компонента, в то время как зависимость $p_i(P)$ практически исчезла из-за пренебрежения зависимостью от давления для чистого вещества (см. формулу (2.2.9)).

Влияние состава бинарного раствора на давление насыщенного пара по закону Рауля (2.5.5) показано на рис. 2.7, где левый край соответствует чистому компоненту А ($x_A = 1$), а правый край — чистому компоненту В ($x_B = 1$). Суммарное давление $p_A + p_B$ над идеальным бинарным раствором при любом его

¹⁾ Рауль (Raoult) Франсуа-Мари (1830–1901) — французский химик и физик, опубликовал многочисленные исследования по переходу химической энергии в энергию тока, но известность получил (1882) за изучение закономерности (носящей ныне его имя) понижения упругости пара и повышения температуры кипения растворителя при введении в него растворенного вещества.

составе лежит, как видно из рис. 2.7, между давлениями насыщенного пара p_A^0 и p_B^0 для чистых компонентов А и В.

Однако закону Рауля с достаточной точностью подчиняются лишь немногие растворы близких по структуре веществ (бензол и толуол, метиловый и этиловый спирты и т. п.). В реальных условиях наблюдаются отклонения от прямой пропорциональности, показанной на рис. 2.7, в сторону как увеличения, так и уменьшения давлений (см. ниже).

Как уже было установлено в п. 2.4, растворитель в предельно разбавленном растворе подчиняется закономерностям идеального раствора. Для него справедлив закон Рауля (2.5.5), переформулированный следующим образом: введение i -й примеси в чистый растворитель (отмечаемый индексом $i = 1$) уменьшает давление его насыщенного пара на величину $\Delta p_1 = p_1^0 - p_1$, пропорциональную мольной доле x_i примеси в растворе, а именно:

$$\Delta p_1 = x_i p_1^0 \quad \text{при} \quad x_i \ll 1 \quad (i \neq 1). \quad (2.5.6)$$

Так как в предельно разбавленном растворе каждая примесь взаимодействует с растворителем независимо от других, то формула (2.5.6) справедлива для всех примесей, которые аддитивно понижают парциальное давление паров растворителя. Тогда

$$\Delta p_1 = \sum_{i \neq 1} x_i p_1^0 \quad \text{при} \quad \sum_{i \neq 1} x_i \ll 1. \quad (2.5.7)$$

На рис. 2.8 показаны температурные зависимости $p_1^0(T)$ и $p_1(T)$, соответственно, для растворителя в чистом виде и в растворе с малым количеством растворенных примесей, обеспечивающих, согласно (2.5.7), понижение давления паров растворителя на величину Δp_1 . Отсюда видно, что при введении примесей в раствор его температура кипения повышается на величину $\Delta T_{1,\text{кип}} = T_{1,\text{кип}} - T_{1,\text{кип}}^0$, пропорциональную сумме мольных долей примесей в растворе:

$$\Delta T_{1,\text{кип}} = \frac{\Delta p_1 / p_1^0}{\partial \ln p_1^0 / \partial T} = \frac{R (T_{1,\text{кип}}^0)^2}{\Delta H_{1,\text{исп}}^0} \sum_{i \neq 1} x_i, \quad (2.5.8)$$

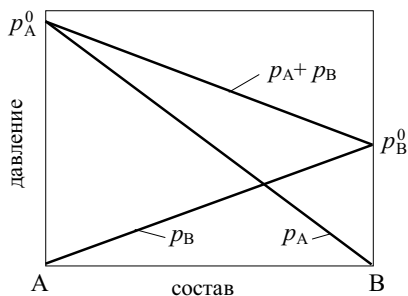


Рис. 2.7. Влияние состава бинарной системы на давление насыщенных паров над идеальным раствором в соответствии с законом Рауля

где в процессе преобразований были использованы формулы (2.2.6) и (2.5.7).

Закон Рауля для растворителя в форме (2.5.6) будет использован ниже (см. п. 2.7) для построения диаграмм плавкости различных типов.

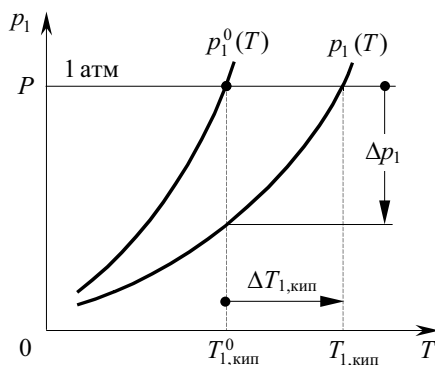


Рис. 2.8. Понижение давления пара растворителя Δp_1 и повышение температуры кипения $\Delta T_{1,\text{кип}}$ для предельно разбавленного раствора по отношению к чистому растворителю с давлением насыщенного пара $p_1^0(T)$

В отличие от растворителя, примесь в предельно разбавленном растворе в общем случае не подчиняется модели идеального раствора. Как уже отмечалось, химический потенциал (2.4.30) для примесей ($i \neq 1$) равен величине μ_i^* , отличной от стандартного химического потенциала μ_i° чистого компонента на величину потенциальной энергии U_{1i} взаимодействия этого компонента с растворителем. По этой причине для примесей в предельно разбавленном растворе вместо закона Рауля (2.5.5) имеет место **закон Рауля–Генри**:

$$p_i = G_i x_i \quad \text{при} \quad i \neq 1. \quad (2.5.9)$$

Равенство (2.5.9) выражает линейную связь между парциальным давлением пара примеси и ее мольной долей в растворе, аналогичную (2.5.5), но с коэффициентом пропорциональности G_i , который отличается от давления p_i^0 насыщенного пара чистого вещества. Соотношение между коэффициентом Генри G_i и давлением p_i^0 будет получено ниже.

Для реальных растворов область сильного разбавления ограничена достаточно малыми концентрациями растворенных веществ. В широком диапазоне концентраций наблюдаются отклонения от линейного закона Рауля. Это отклонение называют положительным или отрицательным в зависимости от того,

больше или меньше давление насыщенных паров над реальным раствором по сравнению с аналогичными давлениями для идеального раствора, показанными на рис. 2.7 в виде прямых линий. В качестве примера для растворов с положительным отклонением от закона Рауля могут служить растворы: вода–метиловый спирт, бензол–ацетон и этиловый спирт–ацетон, а для растворов с отрицательным отклонением — растворы: ацетон–хлороформ и бензол–хлороформ.

Покажем, что знак отклонения от закона Рауля совпадает со знаком раствора, введенным ранее (см. п. 2.4), в соответствии со значением коэффициента активности: значения $\gamma_i > 1$ и $\gamma_i < 1$ характеризуют положительный и отрицательный раствор по i -му компоненту. Для этого из закона Рауля (2.5.5) выразим величину $x_i = p_i/p_i^0$ и подставим ее в (2.5.3), тогда

$$\mu_i^{\text{к}} = \mu_i^{\text{ок}} + RT \ln \frac{p_i}{p_i^0}. \quad (2.5.10)$$

Полученное выражение обобщается на реальные растворы путем введения активности $a_i = \gamma_i x_i$ с помощью равенства (2.4.6), так что формулу (2.5.10) переписываем в виде

$$\mu_i^{\text{к}} = \mu_i^{\text{ок}} + RT \ln a_i, \quad \text{где} \quad a_i = \gamma_i x_i = p_i/p_i^0. \quad (2.5.11)$$

Введение активности для компонентов реального раствора позволяет на основании выражения (2.5.11) записать более общее, чем закон Рауля–Генри, соотношение,

$$p_i = a_i p_i^0 = \gamma_i x_i p_i^0 \equiv \gamma_i p_i^{\text{ид}}, \quad (2.5.12)$$

справедливое для любых растворов. Здесь $p_i^{\text{ид}} = x_i p_i^0$ означает давление пара i -го компонента над идеальным раствором, соответствующее закону Рауля (2.5.5).

Из сравнения выражений (2.5.9) и (2.5.12) получаем коэффициент Генри для примесей в предельно разбавленном растворе,

$$\Gamma_i = \gamma_i p_i^0, \quad (2.5.13)$$

зависящий в данном случае только от температуры в силу температурных зависимостей $\gamma_i(T)$ и $p_i^0(T)$, выражаемых формулами (2.4.31) и (2.2.11).

Из формулы (2.5.12) видно, что для положительного ($\ln \gamma_i > 0$) или отрицательного ($\ln \gamma_i < 0$) раствора имеем положительное ($p_i > p_i^{\text{ид}}$) или отрицательное ($p_i < p_i^{\text{ид}}$) отклонение от закона Рауля. Это означает, что соответствующие этим растворам зависимости проходят выше или ниже прямых линий на рис. 2.7.

Следовательно, атермальные растворы, для которых $\gamma_i < 1$, имеют только отрицательные отклонения от закона Рауля.

Как известно (см. п. 2.4), знак раствора ($\ln \gamma_i > 0$ или < 0) для регулярных растворов совпадает со знаком теплоты растворения ($\Delta H > 0$ или $\Delta H < 0$). Следовательно, знак отклонения от закона Рауля также совпадает со знаком теплоты растворения.

Последнее физически объясняется тем, что, например, при $\Delta H < 0$ (когда теплота выделяется) разнородные молекулы (растворителя и растворенного вещества) взаимодействуют между собой сильнее, чем однородные. Это затрудняет переход молекул в газовую фазу и снижает давление насыщенного пара по сравнению с идеальным раствором (в котором разнородные и однородные молекулы взаимодействуют одинаково), что и обеспечивает неравенство $p_i < p_i^{\text{ид}}$.

Для нахождения температурной зависимости $p_i(T)$ давления пара i -го компонента над реальным раствором логарифмируем равенство (2.5.12) с последующим дифференцированием по T при $x_i = \text{const}$:

$$\frac{\partial \ln p_i}{\partial T} = \frac{\partial \ln p_i^0}{\partial T} + \frac{\partial \ln \gamma_i}{\partial T}. \quad (2.5.14)$$

Первое слагаемое в правой части (2.5.14) описывается уравнением (2.2.6). Второе слагаемое, учитывающее температурную зависимость коэффициента активности $\gamma_i(T)$, наиболее просто записывается для предельно разбавленных растворов, для которых, согласно выражению (2.4.28), имеем $RT \ln \gamma_i = \Delta \bar{H}_i^{\text{изб}} = \text{const}$. В этом случае формула (2.5.14) дает для реального раствора следующее уравнение:

$$\frac{\partial \ln p_i}{\partial T} = \frac{\Delta \bar{H}_{i,\text{исп}}}{RT^2}, \quad (2.5.15)$$

которое совпадает по форме с аналогичным уравнением (2.2.6), полученным для чистых веществ.

Входящая в уравнение (2.5.15) парциальная молярная теплота испарения $\Delta \bar{H}_{i,\text{исп}} \equiv T(\bar{S}_i^{\text{п}} - \bar{S}_i^{\text{к}}) = \Delta H_i^{0_{\text{к} \rightarrow \text{п}}} - \Delta \bar{H}_i^{\text{изб}}$ содержит: а) теплоту фазового перехода конденсат $\rightarrow$ пар, $\Delta H_i^{0_{\text{к} \rightarrow \text{п}}} = T(S_i^{0_{\text{п}}} - S_i^{0_{\text{к}}})$, для чистого i -го компонента; б) избыточную энтальпию его растворения $\Delta \bar{H}_i^{\text{изб}}$. Последняя величина отражает неидеальность раствора и входит в $\Delta \bar{H}_{i,\text{исп}}$ с обратным знаком, поскольку испарение является процессом, обратным растворению газа в конденсированной фазе. Аналогичная картина имеет место и для процесса сублимации с заменой $\Delta \bar{H}_{i,\text{исп}}$ на $\Delta \bar{H}_{i,\text{суб}}$.

Решение уравнения (2.5.15) при $\Delta\bar{H}_{i,\text{исп(суб)}}(T) = \text{const}$ дает температурную зависимость $p_i(T)$ для насыщенного пара компонента i над раствором, качественно показанную на рис. 2.8 и по виду аналогичную формуле (2.2.11), но определяемую парциальной молярной теплотой испарения (сублимации).

2.6. Молекулярная и атомарная растворимость газов в конденсированных средах

Как и в предыдущем параграфе, будем рассматривать равновесие газа с конденсированной средой. Однако в данном случае задача прямо противоположная — найти равновесную мольно-объемную концентрацию $c_i = n_i/V$ газовых частиц i -го сорта (молекул или атомов), растворенных в конденсированной фазе, зная парциальное давление p_i этого компонента в газовой фазе. Образующийся в равновесных условиях раствор (жидкий или твердый) является *насыщенным*, а концентрация в нем газовых частиц называется *растворимостью* газа.

Как отмечалось в п. 2.3, растворение многоатомных газов может происходить либо в молекулярной форме (без диссоциации газовых молекул на атомы), либо в атомарной форме (в виде атомов, возникающих в результате диссоциации молекул). Возможно также ионное растворение, реализуемое в растворах жидких электролитов (см. п. 2.17) и в твердых растворах, образуемых примесями в кристаллах полупроводников (см. п. 3.8). Специфика механизма отражается в закономерностях, управляющих процессом растворения. В данном параграфе ограничиваем рассмотрение анализом молекулярного и атомарного механизмов растворения газов.

1. Молекулярное растворение газов управляется законами фазового равновесия. Для i -го компонента по-прежнему выполняется условие равновесия типа (2.2.1),

$$\mu_i^{\Gamma}(T, p_i) = \mu_i^{\text{K}}(T, p_i), \quad (2.6.1)$$

в котором химические потенциалы имеют вид (1.13.1) и (1.13.2):

$$\mu_i^{\Gamma}(T, p_i) = \mu_i^{\text{ог}}(T) + RT \ln p_i, \quad (2.6.2)$$

$$\mu_i^{\text{K}}(T, p_i) = \mu_i^{\text{ок}}(T) + RT \ln c_i(T, p_i). \quad (2.6.3)$$

Здесь, согласно (2.2.10), оба стандартных химических потенциала $\mu_i^{\text{ог}}$ и $\mu_i^{\text{ок}}$ зависят лишь от температуры T , а выражение (2.6.3) содержит искомую зависимость $c_i(T, p_i)$.

Подставляя (2.6.2) и (2.6.3) в равенство (2.6.1) и разрешая относительно c_i , получаем закон молекулярной растворимости,

$$c_i(T, p_i) = s_i(T) p_i, \quad (2.6.4)$$

называемый **законом Генри**. Согласно этому закону, растворимость газа в молекулярной форме пропорциональна его парциальному давлению в газовой фазе над раствором, что качественно показано на рис. 2.9 а.

Закон молекулярной растворимости (2.6.4) по своей сути является обращением закона Рауля (2.5.5) для идеального раствора или закона Рауля–Генри (2.5.9) для примесей в предельно разбавленном растворе. Обычно этому закону подчиняется растворение газов в слабо полярных жидкостях, а также в твердых силикатных материалах типа стекол и керамики, в которых газовые молекулы размещаются в пустотах силикатной структуры по механизму «молекулярного сита» (см. п. 4.6).

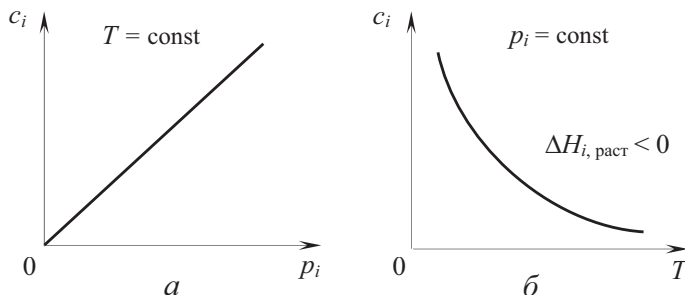


Рис. 2.9. Зависимость молекулярной растворимости газа в конденсированной фазе по закону Генри от парциального давления p_i (а) и от температуры T (б)

Температурная зависимость в законе Генри выражается при помощи *коэффициента растворимости*: $s_i(T) \equiv \exp[(\mu_i^{\text{ок}}(T) - \mu_i^{\text{ог}}(T))/RT]$. Для нахождения его температурной зависимости дифференцируем условие фазового равновесия (2.6.1) по T при $p_i = \text{const}$ с учетом искомой зависимости $c_i = f(T, p_i = \text{const})$:

$$\frac{\partial \mu_i^{\text{г}}}{\partial T} = \frac{\partial \mu_i^{\text{к}}}{\partial T} + \frac{\partial \mu_i^{\text{к}}}{\partial c_i} \frac{\partial c_i}{\partial T}, \quad (2.6.5)$$

где равенство $\partial \mu_i^{\text{к}} / \partial c_i = RT / c_i$ следует из выражения (2.6.3).

Согласно (1.11.16), вводим в рассмотрение парциальные энтропии для i -го компонента в газовой и конденсированной фазах:

$$\frac{\partial \mu_i^{\text{г}}}{\partial T} = -\bar{S}_i^{\text{г}} \quad \text{и} \quad \frac{\partial \mu_i^{\text{к}}}{\partial T} = -\bar{S}_i^{\text{к}}. \quad (2.6.6)$$

Так как растворение газа происходит равновесным образом ($\Delta \bar{G}_i = 0$), то можно записать парциальную теплоту молекулярного растворения газового моля как

$$\Delta \bar{H}_{i,\text{раст}} \equiv \Delta \bar{H}_i^{\text{г} \rightarrow \text{к}} = T(\bar{S}_i^{\text{к}} - \bar{S}_i^{\text{г}}). \quad (2.6.7)$$

Эта величина противоположна по знаку парциальной теплоте испарения (сублимации) $\Delta \bar{H}_{i,\text{исп(суб)}} = T(\bar{S}_i^{\text{п}} - \bar{S}_i^{\text{к}})$, определяющей температурную зависимость давления насыщенного пара $p_i(T)$ в уравнении (2.5.15).

Подстановка выражения (2.6.6) в уравнение (2.6.5) с учетом (2.6.7) дает искомое дифференциальное уравнение,

$$\frac{\partial \ln c_i}{\partial T} \equiv \frac{d \ln s_i}{dT} = \frac{\Delta \bar{H}_{i,\text{раст}}}{RT^2}, \quad (2.6.8)$$

которое по внешнему виду аналогично уравнению (2.5.15). При $\Delta \bar{H}_{i,\text{раст}}(T) = \text{const}$ решение уравнения (2.6.8) имеет вид

$$s_i(T) = s_{i0} \exp\left(-\frac{\Delta \bar{H}_{i,\text{раст}}}{RT}\right). \quad (2.6.9)$$

Как показано в п. 2.3, молекулярное растворение газов является экзотермическим процессом ($\Delta \bar{H}_{i,\text{раст}} < 0$), т. е. сопровождается выделением тепла, так как межмолекулярные взаимодействия в конденсированных средах сильнее, чем в газах. Следовательно, с увеличением температуры молекулярная растворимость газов, как правило, уменьшается по экспоненциальному закону (2.6.9), как показано на рис. 2.9 б.

2. Атомарное растворение газов, в отличие от молекулярного, управляется законами не фазового, а химического равновесия. Это связано с тем, что процесс диссоциации газовой молекулы на атомы с последующим их растворением в конденсированной фазе можно представить в виде химической реакции. Например, растворение газовой молекулы водорода $\text{H}_2(\text{г})$ записывается в виде химической реакции



где справа стоят два атома водорода $\text{H}(\text{к})$, растворенные в конденсированной фазе. Для реакции (2.6.10) условие химического равновесия (1.14.6) принимает вид

$$\mu_{\text{H}_2}^{\text{г}} = 2\mu_{\text{H}}^{\text{к}}, \quad \text{или} \quad \mu_i^{\text{г}}(T, p_i) = 2\mu_i^{\text{к}}(T, p_i). \quad (2.6.11)$$

Здесь в последнем равенстве «водородные» индексы H_2 и H заменены на единый индекс i , чтобы получаемые ниже выражения были аналогичны ранее записанным для закона Генри.

Записывая $\mu_i^r(T, p_i)$ в форме (2.6.2), а $\mu_i^k(T, p_i)$ в форме (2.6.3) и подставляя их в равенство (2.6.11), получаем закон атомарной растворимости двухатомных газов,

$$c_i(T, p_i) = s_i(T) p_i^{1/2}. \quad (2.6.12)$$

называемый **законом Сивертса**. Согласно этому закону, атомарная растворимость двухатомного газа пропорциональна квадратному корню из парциального давления этого газа над раствором. Качественный вид закона Сивертса показан на рис. 2.10 а.

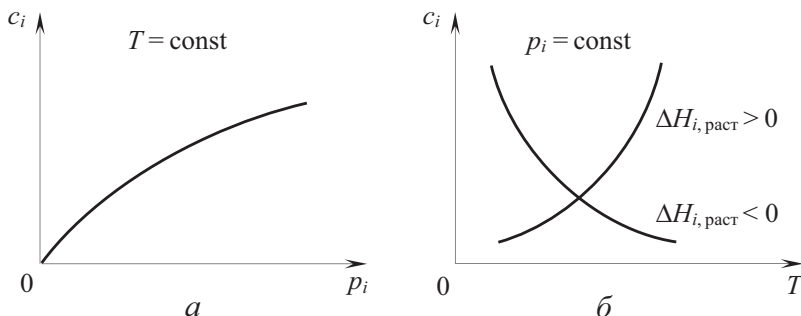


Рис. 2.10. Зависимость атомарной растворимости газа в конденсированной фазе по закону Сивертса от парциального давления p_i (а) и от температуры T (б)

Температурная зависимость в законе Сивертса (2.6.12), как и в законе Генри (2.6.4), выражается *коэффициентом растворимости* $s_i(T)$. Для нахождения его температурной зависимости дифференцируем условие химического равновесия (2.6.11) (с единым индексом i) по T при $p_i = \text{const}$ с учетом искомой функциональной зависимости $c_i = f(T, p_i = \text{const})$:

$$\frac{\partial \mu_i^r}{\partial T} = 2 \frac{\partial \mu_i^k}{\partial T} + 2 \frac{\partial \mu_i^k}{\partial c_i} \frac{\partial c_i}{\partial T}, \quad (2.6.13)$$

где равенство $\partial \mu_i^k / \partial c_i = RT / c_i$ следует из выражения (2.6.3).

Подставляя выражение (2.6.6) в формулу (2.6.13) и записывая для двухатомного газа, в отличие от (2.6.7), парциальную теплоту атомарного растворения газового моля в виде

$$\Delta \bar{H}_{i, \text{раст}} \equiv \Delta \bar{H}_i^{r \rightarrow k} = T(2\bar{S}_i^k - \bar{S}_i^r), \quad (2.6.14)$$

получаем искомое дифференциальное уравнение:

$$\frac{\partial \ln c_i}{\partial T} \equiv \frac{d \ln s_i}{dT} = \frac{\Delta \bar{H}_{i, \text{раст}}}{2RT^2}. \quad (2.6.15)$$

Множитель 2 в формулах (2.6.14) и (2.6.15) учитывает тот факт, что при атомарном растворении число молей в конденсированной фазе удваивается по сравнению с двухатомным газом. При $\Delta\bar{H}_{i,\text{раст}}(T) = \text{const}$, решение уравнения (2.6.15) принимает следующий вид:

$$s_i(T) = s_{i0} \exp\left(-\frac{\Delta\bar{H}_{i,\text{раст}}}{2RT}\right). \quad (2.6.16)$$

Известно (см. п. 2.3), что атомарное растворение газов характерно для металлов, при этом *в конструкционных материалах* (типа Fe, Ni, Cu и др.) процесс растворения эндотермический ($\Delta\bar{H}_{i,\text{раст}} > 0$), а *в активных газопоглотителях* (типа Ti, Zr, Ta и др.) — экзотермический ($\Delta\bar{H}_{i,\text{раст}} < 0$). В соответствии с этим, формула (2.6.16) дает два противоположных случая температурного хода атомарной растворимости, показанные на рис. 2.10 б.

В конструкционных материалах газ растворяется с поглощением тепла и с ростом температуры растворимость возрастает. В противоположность этому, растворение газа в активных газопоглотителях идет с выделением тепла, так что с ростом температуры растворимость уменьшается.

Качественное объяснение различия в механизмах растворения газов для конструкционных и газопоглощающих металлов с участием процессов адсорбции, растворения и диффузии будет дано в п. 5.12 (см. рис. 5.45).

Закономерности испарения и сублимации конденсированных веществ и растворения газов, изученные в пп. 2.2–2.6, лежат в основе управления технологическими процессами, проводимыми в вакууме и газовых средах (такими как вакуумное обезгаживание и очистка материалов, термовакуумное испарение и конденсация, молекулярная и газофазная эпитаксия, легирование полупроводников из газовой фазы и т. п.), а также их используют для оценки остаточного давления в электровакуумных приборах и в вакуумных камерах технологических установок.

Наряду с этим, широкое распространение в технологии электронных приборов получили процессы, основанные на фазовых превращениях жидкость $\rightleftharpoons$ твердое тело, такие как пайка припоями, формирование вплавленных омических контактов к полупроводникам, кристаллизационные методы очистки веществ, жидкофазная эпитаксия и др. Нижеследующие параграфы посвящены рассмотрению фазового равновесия между жидкой и твердой фазами, которое лежит в основе построения так называемых *диаграмм плавкости*, необходимых для изучения основных закономерностей кристаллизации и плавления веществ.

2.7. Равновесие жидкой и твердой фаз в однокомпонентных и бинарных системах

В однокомпонентных системах равновесие между жидкой и твердой фазами имеет место вдоль кривой плавления 3, исходящей из тройной точки O на рис. 2.1. Условие равновесия для чистого i -го вещества записывается в виде равенства химических потенциалов (ср. уравнение (2.1.1)),

$$\mu_i^T(T, P) = \mu_i^J(T, P). \quad (2.7.1)$$

Это уравнение позволяет найти равновесную температуру (плавления $T_{\text{пл}}$ при переходе $t \rightarrow ж$ или кристаллизации $T_{\text{кр}}$ при обратном переходе $ж \rightarrow t$) как функцию давления P , т. е. зависимость $T = f(P)$. Для нахождения этой зависимости продифференцируем равенство (2.7.1) по P вдоль кривой плавления 3 на рис. 2.1:

$$\left(\frac{\partial \mu_i^T}{\partial P}\right)_T + \left(\frac{\partial \mu_i^T}{\partial T}\right)_P \frac{\partial T}{\partial P} = \left(\frac{\partial \mu_i^J}{\partial P}\right)_T + \left(\frac{\partial \mu_i^J}{\partial T}\right)_P \frac{\partial T}{\partial P}. \quad (2.7.2)$$

Используя формулу (1.11.21), преобразуем равенство (2.7.2) к виду

$$\frac{\partial T}{\partial P} = \frac{V_i^{0ж} - V_i^{0т}}{S_i^{0ж} - S_i^{0т}} \equiv \frac{T \Delta V_{i, \text{пл}}^0}{\Delta H_{i, \text{пл}}^0}. \quad (2.7.3)$$

Это уравнение называют **уравнением Клапейрона–Клаузиуса**.

Уравнение такого типа применимо не только к кривой плавления 3, но и для кривых сублимации 1 и испарения 2 на рис. 2.1. В частности, из него может быть получено уравнение (2.2.6) для давления насыщенного пара чистого вещества при использовании уравнения Клапейрона–Менделеева (1.12.1) и соответствующей замене величин в формуле (2.7.3).

Поскольку плавление твердых веществ всегда требует подвода теплоты, т. е. $\Delta H_{i, \text{пл}}^0 \equiv T(S_i^{0ж} - S_i^{0т}) > 0$, то знак производной в уравнении (2.7.3), дающей наклон кривых 3 на рис. 2.1, определяется только знаком объемного эффекта при плавлении, $\Delta V_{i, \text{пл}}^0 \equiv V_i^{0ж} - V_i^{0т}$.

Как было отмечено в п. 2.1, встречаются вещества как с $\Delta V_{i, \text{пл}}^0 > 0$, так и с $\Delta V_{i, \text{пл}}^0 < 0$. Первым веществам (которых в природе явно больше) соответствует сплошная линия 3 с положительным наклоном, а вторым — штриховая линия 3 с отрицательным наклоном. Однако в обоих случаях величина $|\Delta V_{i, \text{пл}}^0|$ невелика, так как молярные объемы веществ в жидком и твердом состояниях примерно одинаковы. Поэтому кривые плавления 3

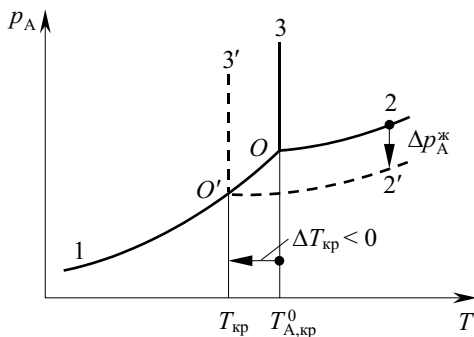


Рис. 2.11. Кривые равновесия фаз для чистого вещества А (сплошные линии) и для раствора примеси В в растворителе А (штриховые линии) при наличии растворимости только в жидком состоянии

на рис. 2.1 выходят из тройной точки O почти вертикально, то есть температура плавления–кристаллизации практически не зависит от давления. Например, для воды, имеющей $\Delta V_{пл}^0 < 0$, понижение давления от атмосферного (760 Торр) до $P = 4,6$ Торр в тройной точке увеличивает температуру ее замерзания с 0°C всего до $0,01^\circ\text{C}$.

На основании вышесказанного, за температуру кристаллизации–плавления $T_{кр} \equiv T_{пл}$ обычно принимают температуру, соответствующую пересечению кривых сублимации и испарения в тройной точке для всех веществ, в том числе для жидких растворителей с примесями. Физически это вполне объяснимо, так как возникновение твердой фазы при кристаллизации раствора требует равенства давления паров растворителя над твердой и жидкой фазами, ибо газовая фаза для них общая.

Выясним влияние растворенных веществ (примесей) на температуру кристаллизации–плавления растворителя. Для этой цели воспользуемся законом Рауля в форме (2.5.6), дающей понижение давления пара растворителя при введении в него примеси.

Будем рассматривать бинарную систему А–В, в которой оба компонента неограниченно взаимно растворимы в жидком состоянии. В зависимости от существования или отсутствия взаимной растворимости компонентов в твердой фазе поведение температуры кристаллизации $T_{кр}$ при изменении состава жидкой фазы будет различным.

Для определенности в качестве растворителя возьмем компонент А, для которого зависимость давления насыщенного пара от температуры показана на рис. 2.11 и 2.12. Сплошными линиями 1, 2 и 3 изображены кривые равновесия соответствующих

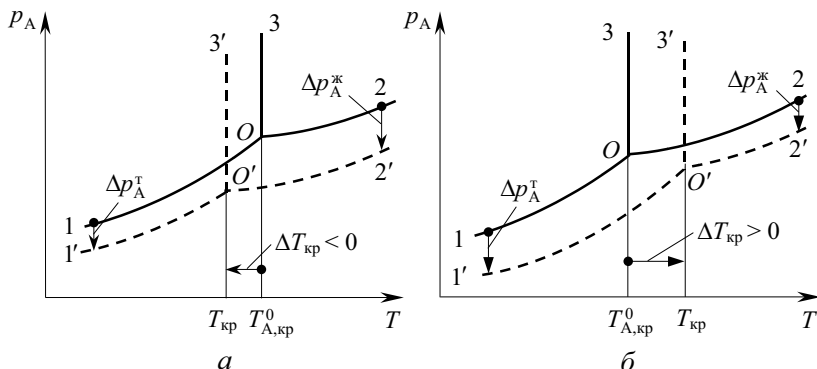


Рис. 2.12. Кривые равновесия фаз для чистого вещества А (сплошные линии) и для раствора примеси В в растворителе А (штриховые линии) при наличии растворимости как в жидком, так и в твердом состоянии, соответствующие случаям понижения (а) и повышения (б) температуры кристаллизации раствора $T_{кр}$ по отношению к аналогичной температуре $T_{А,кр}^0$ для чистого растворителя

фаз для чистого вещества А (аналогичные кривым на рис. 2.1), а штриховыми линиями 1', 2' и 3' — аналогичные кривые при введении в растворитель А примеси В. В соответствии с законом Рауля (2.5.6) штриховые кривые сублимации 1' и испарения 2' смещены вниз по отношению к аналогичным сплошным кривым 1 и 2, соответственно, на $\Delta p_A^T = x_B^T p_A^{0T}$ и $\Delta p_A^ж = x_B^ж p_A^{0ж}$.

Рисунок 2.11 соответствует случаю отсутствия растворимости компонента В в твердом растворителе А ($x_B^T \equiv 0$), так что закон Рауля проявляется лишь для жидкой фазы: $\Delta p_A^T \equiv 0$ и $\Delta p_A^ж \neq 0$. В этом случае температура кристаллизации жидкого растворителя А обязательно понижается в результате растворения в нем примеси В, т. е. $\Delta T_{кр} = T_{кр} - T_{А,кр}^0 < 0$.

Рисунок 2.12 отвечает существованию растворимости компонента В как в твердом, так и в жидком растворителе А, так что $\Delta p_A^T \neq 0$ и $\Delta p_A^ж \neq 0$. В этом случае принципиально возможны две ситуации, соответствующие как понижению ($\Delta T_{кр} < 0$ на рис. 2.12 а), так и повышению ($\Delta T_{кр} > 0$ на рис. 2.12 б) температуры кристаллизации $T_{кр}$ жидкого раствора по отношению к аналогичной температуре $T_{А,кр}^0$ чистого растворителя.

Для получения количественных соотношений будем полагать содержание компонента В в растворителе А достаточно малым, чтобы рассматривать жидкий и твердый растворы как предельно разбавленные. Тогда к растворителю А применимы соотношения для идеальных растворов, в частности следующие выражения

для химических потенциалов (см. формулу (2.4.29)):

$$\mu_A^T = \mu_A^{0T} + RT \ln x_A^T \quad \text{и} \quad \mu_A^{\text{ж}} = \mu_A^{0\text{ж}} + RT \ln x_A^{\text{ж}}. \quad (2.7.4)$$

Из равенства химических потенциалов ($\mu_A^T = \mu_A^{\text{ж}}$) следует, что составы твердого и жидкого растворов различны, т. е. $x_A^T \neq x_A^{\text{ж}}$, так как разным агрегатным состояниям одного и того же чистого вещества А соответствуют разные химические потенциалы ($\mu_A^{0T} \neq \mu_A^{0\text{ж}}$). С изменением температуры равновесный состав растворов изменяется, давая две зависимости $x_A^T = f_1(T)$ и $x_A^{\text{ж}} = f_2(T)$. Кривую $x_A^T = f_1(T)$, выражающую зависимость состава твердого раствора от температуры, принято называть **линией солидуса** и обозначать на диаграммах плавкости буквой *s*. Аналогичная кривая $x_A^{\text{ж}} = f_2(T)$ для жидкого раствора, равновесного с твердым, называется **линией ликвидуса** и обозначается буквой *l*. На диаграммах плавкости, рассматриваемых ниже, эти линии изображают в координатах T – x (температура–состав).

Растворы в равновесном состоянии принято характеризовать *коэффициентами распределения* (или *сегрегации*) компонентов между твердой и жидкой фазами в виде

$$K_A(T) = \frac{x_A^T(T)}{x_A^{\text{ж}}(T)} \quad \text{и} \quad K_B(T) = \frac{x_B^T(T)}{x_B^{\text{ж}}(T)}, \quad (2.7.5)$$

которые являются функциями только температуры.

Приравниваем друг другу химические потенциалы (2.7.4) и дифференцируем полученное равенство по T при $P = \text{const}$:

$$\left(\frac{\partial \mu_A^T}{\partial T} \right)_P + \left(\frac{\partial \mu_A^T}{\partial x_A^T} \right)_T \frac{\partial x_A^T}{\partial T} = \left(\frac{\partial \mu_A^{\text{ж}}}{\partial T} \right)_P + \left(\frac{\partial \mu_A^{\text{ж}}}{\partial x_A^{\text{ж}}} \right)_T \frac{\partial x_A^{\text{ж}}}{\partial T}. \quad (2.7.6)$$

Входящие сюда производные химических потенциалов следуют из выражений (2.7.4) в виде $\partial \mu_A^T / \partial x_A^T = RT / x_A^T$ и $\partial \mu_A^{\text{ж}} / \partial x_A^{\text{ж}} = RT / x_A^{\text{ж}}$. На основании (1.11.16) записываем

$$\left(\frac{\partial \mu_A^T}{\partial T} \right)_P = -\bar{S}_A^T \quad \text{и} \quad \left(\frac{\partial \mu_A^{\text{ж}}}{\partial T} \right)_P = -\bar{S}_A^{\text{ж}}. \quad (2.7.7)$$

Подстановка (2.7.7) в равенство (2.7.6) дает уравнение для коэффициента распределения вещества А:

$$\frac{d \ln K_A}{dT} = \frac{\Delta \bar{H}_{A, \text{кр}}}{RT^2}, \quad (2.7.8)$$

где $\Delta \bar{H}_{A, \text{кр}} \equiv \Delta \bar{H}_A^{\text{ж} \rightarrow \text{т}} = T(\bar{S}_A^T - \bar{S}_A^{\text{ж}})$ — парциальная теплота кристаллизации растворителя А, при этом $\Delta \bar{H}_{A, \text{кр}} = -\Delta \bar{H}_{A, \text{пл}} < 0$.

Считая $\Delta \bar{H}_{A,пл} \approx \Delta H_{A,пл}^0 = \text{const}$, проинтегрируем уравнение (2.7.8) от температуры $T_{A,кр}^0$ для чистого вещества А (при которой $K_A(T_{A,кр}^0) = 1$) до некоторого значения $T_{кр}$ (при котором имеем $K_A(T_{кр})$). В результате интегрирования получаем **уравнение Шредера–Ван-Лаара**:

$$\ln K_A = \frac{\Delta H_{A,пл}^0}{R} \left(\frac{1}{T_{кр}} - \frac{1}{T_{A,кр}^0} \right). \quad (2.7.9)$$

Отсюда можно найти изменение температуры кристаллизации, $\Delta T_{кр} = T_{кр} - T_{A,кр}^0$, при введении как в жидкий, так и в твердый растворитель А примеси В с концентрациями $x_B^* = 1 - x_A^*$ и $x_B^T = 1 - x_A^T$, соответственно. Используя соотношение $\ln x_A = \ln(1 - x_B) \approx -x_B$ при $x_B \ll 1$, нетрудно из (2.7.9) при $\Delta T_{кр} \ll T_{A,кр}^0$ получить равенство (ср. с формулой (2.5.8))

$$\Delta T_{кр} = (K_B - 1) \frac{R(T_{A,кр}^0)^2}{\Delta H_{A,пл}^0} x_B^*. \quad (2.7.10)$$

Это равенство показывает прямую пропорциональность между смещением температуры кристаллизации $\Delta T_{кр}$ и концентрацией примеси x_B^* в жидкой фазе с коэффициентом пропорциональности, зависящим от коэффициента распределения K_B для этой примеси.

Формула (2.7.10) является количественным выражением закономерностей, качественно представленных на рис. 2.11 и 2.12. Действительно, в отсутствие растворимости в твердом состоянии имеем $K_B = 0$, так что из формулы (2.7.10) получаем $\Delta T_{кр} < 0$ (см. рис. 2.11). Итак, с увеличением мольной доли примеси в жидком растворе температура его кристаллизации $T_{кр}$ линейно (при $x_B^* \ll 1$) понижается от значения $T_{A,кр}^0$ для чистого растворителя, как показано на рис. 2.13 а (где нижний индекс «кр» для простоты опущен).

Если примесь В перераспределяется между твердым и жидким растворителями с коэффициентом распределения $K_B < 1$, то формула (2.7.10) дает всегда понижение температуры кристаллизации раствора ($\Delta T_{кр} < 0$), как показано на рис. 2.12 а. В случае $K_B > 1$ из формулы (2.7.10) получаем $\Delta T_{кр} > 0$, что соответствует рис. 2.12 б.

При одновременном существовании жидкого и твердого растворов построение фазовой диаграммы в координатах T – x требует (в отличие от случая $K_B = 0$ на рис. 2.13 а) изображения двух линий — *ликвидуса* l и *солидуса* s . Эти линии характеризуют

составы жидкой (x_B^*) и твердой (x_B^T) фаз, находящихся в равновесии при температуре T , и показаны на рис. 2.13 б и в для случаев $K_B < 1$ и $K_B > 1$. Важно помнить, что на диаграммах T - x (температура–состав) линии солидуса и ликвидуса следует читать в обратных координатах как $x_B^T = f_1(T)$ и $x_B^* = f_2(T)$.

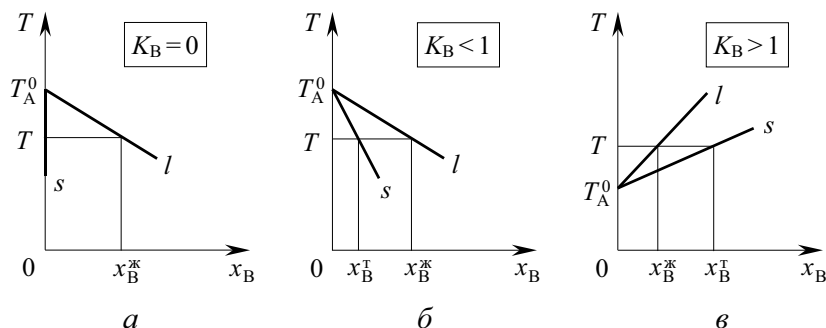


Рис. 2.13. Расположение линий ликвидуса l и солидуса s на диаграмме температура–состав при различных коэффициентах распределения примеси В между твердым и жидким растворителем А: а — $K_B = 0$, б — $K_B < 1$, в — $K_B > 1$

Отсутствие растворимости в твердой фазе (рис. 2.13 а) можно рассматривать как частный случай общей ситуации, отвечающей случаю $K_B < 1$ на рис. 2.13 б. Наклонная линия на рис. 2.13 а является линией ликвидуса l (выражающей состав жидкого раствора), в то время как линия, соответствующая чистому кристаллу А, идущая по оси ординат вертикально вниз из точки T_A^0 , играет роль линии солидуса s . Действительно, полагая $K_B \rightarrow 0$ на рис. 2.13 б (т. е. $x_B^T \rightarrow 0$ при $x_B^* = \text{const}$), приходим к картине, изображенной на рис. 2.13 а с вертикальным расположением линии солидуса, которую назовем *вырожденным солидусом*.

Диаграммы фазового состояния бинарной системы, построенные в координатах T - x (температура–состав) при $P = \text{const}$, называются **диаграммами плавкости**. По оси состава слева размещается чистый компонент А ($x_A = 1, x_B = 0$), а справа — чистый компонент В ($x_B = 1, x_A = 0$). При этом в качестве компонентов А и В могут выступать не только элементарные вещества (Si-Ge, Ga-As и др.), но и химические соединения (NaCl-H₂O, AlAs-GaAs и др.).

Приведенные на рис. 2.13 линии ликвидуса l и солидуса s в виде прямых отрезков дают начальный участок диаграммы плавкости в области малых концентраций примеси (при $x_B \ll 1$ и $x_A \ll 1$ для левого и правого углов диаграммы, соответственно). При удалении от этих углов линии ликвидуса и солидуса

искривляются в соответствии с типом диаграммы. Эти линии, смыкаясь между собой или с другими линиями равновесия твердых фаз, разбивают всю плоскость T – x на ряд областей, в пределах которых система может быть однофазной, двухфазной или трехфазной.

Для бинарной системы равновесие трех фаз возможно лишь в одномерной области, вырожденной в горизонтальную линию $T = \text{const}$, называемую *эвтектической* или *перитектической горизонталью* (см. ниже пп. 2.8 и 2.10). Максимальное число фаз в бинарной системе при $P = \text{const}$, в соответствии с правилом фаз (1.15.3), равно трем ($\Phi = 3$), при этом система становится инвариантной ($C = 0$).

Нахождение системы в той или иной области диаграммы плавкости и последовательность смены этих областей при изменении температуры определяются исходным состоянием жидкого раствора $(T_0, x_0^{\text{ж}})$, из которого начинается его принудительное охлаждение с целью последующей кристаллизации. Охлаждение производится путем отвода тепла от системы с достаточно малой скоростью, при которой практически сохраняется термодинамическое равновесие в системе. Такой процесс кристаллизации, протекающий в соответствии с диаграммой плавкости, принято называть *равновесной кристаллизацией*.

В ходе равновесной кристаллизации точка начального состояния $(T_0, x_0^{\text{ж}})$, называемая *фигуративной точкой*, при охлаждении смещается вниз по вертикальной линии $x_0^{\text{ж}} = \text{const}$. При этом она пересекает линии ликвидуса, солидуса и эвтектические или перитектические горизонталы, тем самым переводя систему из одной области (одно-, двух- или трехфазной) в другую. Для конкретной температуры $T = \text{const}$ в двухфазных областях проводят горизонтальные отрезки линий, называемые *коннодами*, оканчивающиеся на границах данной области. Концы коннод, упирающиеся в линии ликвидуса и солидуса (или в эквивалентные им линии, в том числе вырожденные солидусы для чистых элементов и соединений, идущие из их точек плавления вертикально вниз), показывают по оси x равновесный состав фаз, сосуществующих в этой области. Конкретная область диаграммы при данной температуре $T < T_0$ определяется положением в ней фигуративной точки, движущейся при снижении температуры вниз по вертикали $x_0^{\text{ж}} = \text{const}$. Фигуративная точка делит определенную конноду двухфазной области в пропорции, равной отношению числа молей в сосуществующих фазах, в соответствии с так называемым *правилом рычага* (см. п. 2.9).

Ниже, основываясь на полученных результатах, качественно обсудим построение различных типов диаграммы плавкости и характерные для них особенности равновесной кристаллизации.

2.8. Диаграмма плавкости бинарных систем без твердых растворов

Если в бинарной системе компоненты А и В неограниченно растворимы друг в друге в жидком состоянии и полностью нерастворимы в твердом, то нет расслоения жидкой фазы, и при кристаллизации из нее выпадают раздельно чистые кристаллы A^T и B^T . Следовательно, твердое состояние представляет собой поликристалл — двухфазную систему в виде смеси мелких кристаллов A^T и B^T .

Диаграмма плавкости показана на рис. 2.14 а. Вблизи каждого чистого компонента А или В ход диаграммы соответствует изображенному на рис. 2.13 а. По мере увеличения концентрации растворенного вещества линии ликвидуса l_A и l_B , опускаясь вниз из точек T_A^0 и T_B^0 , искривляются и пересекаются в точке э, называемой *эвтектической точкой* (от греческого слова «легко плавящийся»). В результате диаграмма плавкости принимает единственно возможный вид, показанный на рис. 2.14 а, часто называемый диаграммой типа *простой эвтектики*.

Выше линий ликвидуса l_A и l_B существует лишь жидкий раствор, т.е. система однофазная ($\Phi = 1$). Линии ликвидуса l_A и l_B характеризуют состав жидкого раствора, находящегося в равновесии, соответственно, с кристаллами A^T и B^T , при этом система двухфазная ($\Phi = 2$). Эти равновесные фазы $A^T + ж$ и $ж + B^T$ указаны внутри «треугольных» областей $T_A^0 \varepsilon T_\varepsilon$ и $T_B^0 \varepsilon T_\varepsilon$, лежащих выше эвтектической горизонтали $T_\varepsilon = \text{const}$. Ниже температуры эвтектики жидкость исчезает и система становится твердой двухфазной ($A^T + B^T$). На эвтектической горизонтали T_ε система является трехфазной: эвтектический раствор-расплав состава x_ε находится в равновесии с кристаллами A^T и B^T .

Рассмотрим процесс равновесной кристаллизации жидких растворов разного исходного состава $x_0^ж$, отмеченных на рис. 2.14 а фигуративными точками и стрелками с номерами от 1 до 5. Фигуративным точкам 1 и 5 соответствуют чистые вещества А и В, а точке 3 — раствор эвтектического состава x_ε . Соответствующие кривые охлаждения (в виде зависимости температуры T от времени) показаны на рис. 2.14 б под теми же номерами. На разных участках кривых охлаждения буквами указаны кристаллизующиеся вещества (A^T , B^T и $A^T + B^T$).

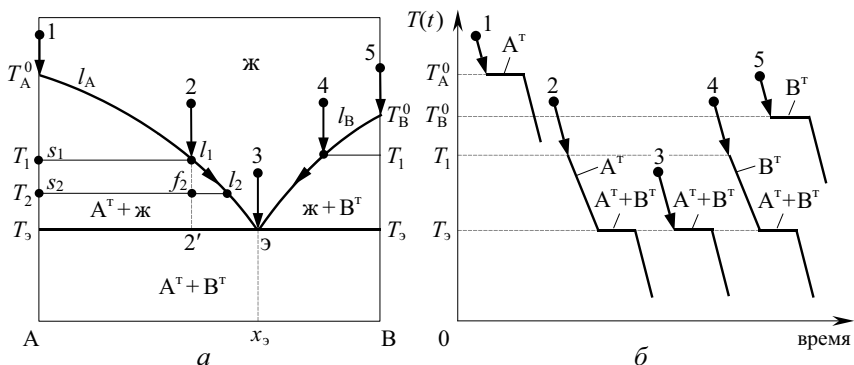


Рис. 2.14. Диаграмма плавкости типа простой эвтектики для бинарных систем без твердых растворов (а) и кривые охлаждения жидкого раствора (б) для пяти различных составов, обозначенных точками 1–5 на рисунке а. Коннода s_2l_2 с фигуративной точкой f_2 приведена для демонстрации правила рычага. На разных участках кривых охлаждения буквами указаны кристаллизующиеся вещества (A^T , B^T , $A^T + B^T$)

Для чистого вещества А или В число степеней свободы системы при $P = \text{const}$, согласно правилу фаз (1.15.3), равно $C = 2 - \Phi$. До начала кристаллизации жидкий расплав $A^ж$ или $B^ж$ является моновариантной системой ($C = 1$), т. е. температура может изменяться в ходе принудительного охлаждения. При температуре кристаллизации ($T = T_A^0$ или $T = T_B^0$) появляются две равновесные фазы ($A^ж + A^T$ или $B^ж + B^T$), так что система становится невариантной ($C = 0$). Это означает, что температура не может изменяться в процессе кристаллизации до тех пор, пока весь расплав не затвердеет и система снова не станет однофазной и моновариантной ($C = 1$). Следовательно, невариантному состоянию системы в процессе кристаллизации соответствует неизменная температура на кривых охлаждения 1 и 5 в форме *площадки невариантности*. Постоянная температура поддерживается путем отвода теплоты кристаллизации.

Для бинарной системы при $P = \text{const}$ число степеней свободы, $C = 3 - \Phi$, изменяется по мере охлаждения в соответствии с изменением числа фаз. Выше линий ликвидуса l_A и l_B однофазная система является бивариантной ($C = 2$), т. е. могут независимо изменяться температура и состав жидкого раствора. В ходе принудительного охлаждения исходного раствора, состав которого соответствует фигуративной точке 2 или 4 на рис. 2.14 а, система остается бивариантной вплоть до температуры T_1 , при которой вертикальная линия 2 или 4 достигает линии ликвидуса l_A или l_B . В этот момент выпадают кристаллы A^T (для

состава 2) или B^T (для состава 4) и система становится двухфазной и моновариантной ($C = 1$). Это означает, что при сохранении двух фаз ($A^T + ж$ или $ж + B^T$) допустимо изменять температуру, каждому значению которой отвечает строго определенный состав раствора в соответствии с линиями ликвидуса l_A или l_B . Выделяющаяся теплота кристаллизации дополнительно подогревает систему, так что ниже температуры T_1 кривая охлаждения 2 или 4 на рис. 2.14 б идет менее круто по сравнению с кривой принудительного охлаждения при $T > T_1$.

Для исходного состава 2 по мере выделения кристаллов A^T жидкий раствор обогащается компонентом В, т. е. при охлаждении ниже температуры T_1 состав раствора изменяется вдоль линии ликвидуса l_A в направлении стрелки к эвтектической точке э. При температуре T_3 , наряду с кристаллами A^T (для исходного состава 2), начинают выпадать кристаллы B^T и система становится трехфазной и невариантной ($C = 0$). При этом на кривой охлаждения 2 (рис. 2.14 б) появляется площадка невариантности, которая сохраняется до тех пор, пока не исчезнет жидкая фаза и система вновь не станет двухфазной ($A^T + B^T$) и моновариантной ($C = 1$). Аналогично протекает процесс кристаллизации для исходного состава 4, т. е. по мере снижения температуры, система движется вдоль линии ликвидуса l_B в направлении стрелки к эвтектической точке э. Следует отметить, что начальная температура T_1 выбрана одинаковой для процессов кристаллизации 2 и 4 только ради удобства графического изображения на рис. 2.14.

Если исходный раствор-расплав имеет эвтектический состав x_3 (точка 3 на рис. 2.14 а), то при его охлаждении кристаллизация начинается при температуре эвтектики T_3 . Из раствора одновременно кристаллизуются чистые вещества A^T и B^T , т. е. при $T = T_3$ система сразу из бивариантной ($C = 2$) превращается в невариантную ($C = 0$). При этом на кривой охлаждения 3 (рис. 2.14 б) появляется площадка невариантности, исчезающая после кристаллизации всего раствора-расплава. В процессе кристаллизации эвтектического раствора возникает поликристаллическая смесь $A^T + B^T$.

Особенностью систем без твердых растворов является первоначальная кристаллизация из жидкой фазы только одного из веществ в чистом виде. Другие вещества (примеси) будут концентрироваться в последних порциях раствора-расплава, которые кристаллизуются при температуре эвтектики. Это свойство лежит в основе метода зонной очистки веществ.

Диаграмма плавкости типа простой эвтектики (рис. 2.14 а) характерна для ряда металлов, входящих в состав электродных сплавов, предназначенных для формирования выпрямляющих и невыпрямляющих (омических) контактов к кремнию и германию, например, Si–Au (Ag, Ga, Sn, Sb), Ge–Au (In, Sn, Pb) и др. Так, для системы Si–Au диаграмма на рис. 2.14 а имеет следующие параметры: $T_{\text{Si}}^0 = 1417^\circ\text{C}$, $T_{\text{Au}}^0 = 1063^\circ\text{C}$, $T_3 = 370^\circ\text{C}$, состав эвтектики $x_3 = 69\%$ (ат.) Si.

Эвтектическую диаграмму имеют многие водные растворы солей, таких как NaCl, AgNO₃, CaCl₂, CuSO₄ и др. Затвердевшие эвтектические смеси соли и льда называются *криогидратами*, а их эвтектические температуры — криогидратными температурами. Например, для водного раствора поваренной соли NaCl при ее эвтектическом содержании 22,4% (вес.) криогидратная температура равна $T_3 = -21,2^\circ\text{C}$. Эти же соли могут образовывать с водой химические соединения типа NaCl · nH₂O ($n = 1, 2$), CuSO₄ · nH₂O ($n = 1, 3, 5$), CaCl₂ · nH₂O ($n = 1, 2, 4, 6$), называемые *кристаллогидратами*. В отличие от криогидратов, являющихся смесью кристаллов льда и соли, в кристаллогидратах молекулы воды химически связаны с ионной решеткой соли. При высоких температурах кристаллизационная вода полностью или частично удаляется и возникает безводная соль или низший кристаллогидрат. При низких температурах кристаллизуются высшие кристаллогидраты (с большим n), которые в криогидратной точке вместе с кристаллами льда образуют криогидраты. Полная диаграмма состояния соль–вода с учетом образования кристаллогидратов, строго говоря, относится к диаграммам с инконгруэнтно плавящимися соединениями, рассмотренным в п. 2.11.

2.9. Диаграммы плавкости бинарных систем с неограниченным твердым раствором

Пусть вещества А и В, кроме полной взаимной растворимости в жидком состоянии, также неограниченно растворимы друг в друге в твердой фазе с образованием непрерывного твердого раствора α. В этом случае принципиально возможны три типа диаграмм плавкости, изображенные на рис. 2.15 и 2.16, которые часто называются диаграммами типа «сигара» (рис. 2.15), «сигара с максимумом» (рис. 2.16 а) и «сигара с минимумом» (рис. 2.16 б). Угловые участки диаграмм вблизи чистых веществ А и В представляют собой три возможные комбинации двух вариантов понижения и повышения температуры кристаллизации, показанные на рис. 2.13 б и в.

Выше линии ликвидуса существует жидкий раствор ж , ниже линии солидуса — твердый раствор α непрерывного состава, а между этими линиями — область разделения системы на жидкий и твердый растворы, находящиеся в равновесии.

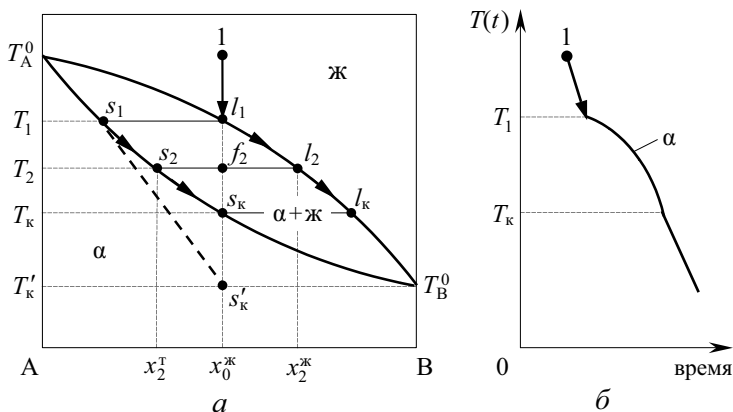


Рис. 2.15. Диаграмма плавкости типа «сигара» для бинарных систем с неограниченным твердым раствором (а) и кривая охлаждения жидкого раствора (б) для состава, обозначенного точкой 1 на рисунке а. Коннода s_2l_2 с фигуративной точкой f_2 приведена для демонстрации правила рычага. Штриховой кривой на рисунке а показан ход неравновесной кристаллизации

Рассмотрим процесс кристаллизации жидкого раствора, который соответствует фигуративной точке 1 с начальным составом $x_0^{\text{ж}}$ на рис. 2.15 а. При охлаждении системы этот состав сохраняется неизменным вплоть до температуры T_1 в точке l_1 на линии ликвидуса. В этот момент начинают выпадать первые порции твердого раствора α состава, отвечающего точке s_1 на линии солидуса. Кристаллы s_1 беднее компонентом В по сравнению с жидким раствором l_1 ($K_B < 1$), так что при их осаждении последний обогащается компонентом В. Следовательно, по мере охлаждения жидкого раствора его состав должен изменяться вдоль линии ликвидуса в направлении стрелки от точки l_1 к l_2 и далее к конечной точке l_k . Соответственно этому, состав твердого раствора α при равновесной кристаллизации меняется вдоль линии солидуса s_1 к s_2 и далее к конечной точке s_k . Кристаллизация полностью заканчивается при температуре T_k .

В процессе кристаллизации система остается двухфазной и моновариантной ($C = 1$), так что на кривой охлаждения не возникает площадки невариантности (рис. 2.15 б). Это отличает системы с неограниченным твердым раствором (рис. 2.15 и 2.16)

от систем с эвтектикой (рис. 2.14). Выделение теплоты кристаллизации, как и прежде, замедляет темп охлаждения. Строго говоря, наклон кривой охлаждения ниже температуры T_k может отличаться от наклона на начальном участке (при $T > T_1$), так как теплоемкости жидкого и твердого растворов в общем случае различны.

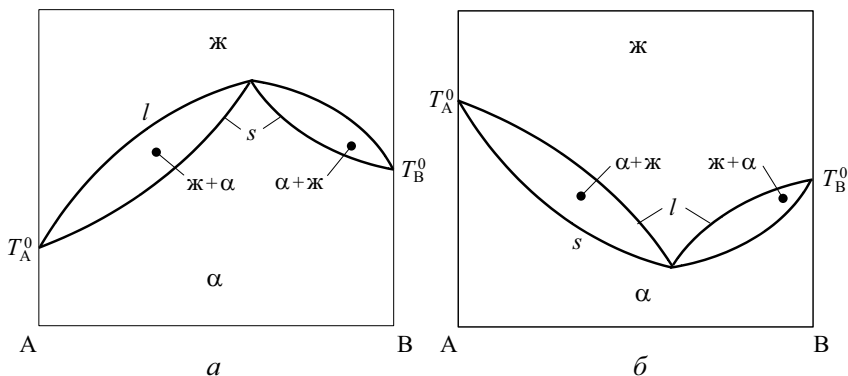


Рис. 2.16. Диаграммы плавкости типа «сигара с максимумом» (а) и «сигара с минимумом» (б) для бинарных систем с неограниченным твердым раствором

Описанный процесс равновесной кристаллизации следует осуществлять настолько медленно, чтобы ранее образовавшиеся кристаллы s_1 , s_2 и т. д. успевали за счет процесса диффузии изменять свой состав в соответствии с линией солидуса. В результате весь объем раствора после затвердевания при температуре T_k будет иметь однородный состав с концентрацией, равной концентрации x_0^* исходной жидкости.

При промежуточной температуре T_2 положение фигуративной точки f_2 на конноде s_2l_2 позволяет найти относительную долю закристаллизовавшегося раствора. Это делается с помощью так называемого *правила рычага*, доказываемого ниже.

Пусть в бинарном жидком растворе состава x_0^* (вычисленного по любому компоненту А или В) исходное количество молей равно $n_0^* = n_{A0}^* + n_{B0}^*$. Как показано на рис. 2.15 а, при некоторой температуре T_2 существует равновесие твердого и жидкого растворов с концентрациями x_2^* и x_2^* (по тому же компоненту). Надо найти соотношение между количеством молей в твердой фазе ($n_2^* = n_{A2}^* + n_{B2}^*$) и в жидкой фазе ($n_2^* = n_{A2}^* + n_{B2}^*$). Для этого воспользуемся условиями сохранения общего числа молей ($n_0^* = n_2^* + n_2^*$) и числа молей выбранного компонента ($x_0^* n_0^* = x_2^* n_2^* + x_2^* n_2^*$). Отсюда получаем искомое *правило рычага*,

записанное в форме следующего соотношения:

$$\frac{n_2^{\text{ж}}}{n_2^{\text{т}}} = \frac{x_0^{\text{ж}} - x_2^{\text{т}}}{x_2^{\text{ж}} - x_0^{\text{ж}}} \equiv \frac{s_2 f_2}{f_2 l_2}. \quad (2.9.1)$$

Правило рычага гласит: *число молей в сосуществующих фазах обратно пропорционально отрезкам, на которые делит конноду $s_2 l_2$ фигуративная точка f_2* (соответствующая температуре T_2 на рис. 2.15 а).

Правило (2.9.1) применимо также к кристаллизации чистого вещества, например, для компонента А на рис. 2.14 а, где при температуре T_2 показана коннода $s_2 l_2$, для которой $x_2^{\text{т}} = x_{\text{А}}^{\text{т}} \equiv 1$. Аналогичным образом строится правило рычага и на диаграммах более сложного вида (см., например, конноду $s_2 l_2$ с фигуративной точкой f_2 на рис. 2.19 а).

В отличие от равновесной кристаллизации, при *неравновесном* (быстром) *охлаждении*, диффузия в твердой фазе, будучи медленным процессом, не успевает выравнивать состав растущих кристаллов. В этом случае процесс кристаллизации, показанный на рис. 2.15 а, начинается при температуре T_1 и идет вдоль некоторой штриховой линии, проходящей ниже линии солидуса. Это вызвано тем, что кристаллы, выпадающие при температуре $T_2 < T_1$, оказываются более обедненными компонентом В по сравнению с равновесным составом вдоль линии солидуса. При достижении температуры $T_{\text{к}}$, соответствующей концу равновесной кристаллизации, суммарное число молей компонентов А и В в твердой фазе не равно их исходному числу $n_0^{\text{ж}}$ в растворе-расплаве, который в данных условиях более обогащен компонентом В. Последующая неравновесная кристаллизация (при температурах ниже $T_{\text{к}}$) еще более обогащает расплав компонентом В. Как результат этого, процесс затвердевания полностью заканчивается при температуре, равной температуре кристаллизации $T_{\text{В}}^0$ чистого вещества В. Таким образом, при неравновесной кристаллизации возникает неоднородный твердый раствор, зерна которого имеют в центре состав s_1 , а на периферии — чистое вещество В. Результирующее изменение состава твердой фазы при неравновесной кристаллизации показано на рис. 2.15 а штриховой линией, идущей из точки s_1 при температуре T_1 в точку $s'_{\text{к}}$ при температуре $T'_{\text{к}} = T_{\text{В}}^0$.

Непрерывные твердые растворы характерны для многих двух-компонентных систем, состоящих из близких материалов, таких как Si–Ge, Au–Ag, Au–Ni, Ni–Cu, Ni–Fe, Mo–W и др., а также для сложных трехкомпонентных систем типа NaCl–NaBr,

$\text{Al}_2\text{O}_3\text{--Cr}_2\text{O}_3$ или для полупроводниковых соединений типа AlAs--GaAs , GaP--GaAs , InP--InAs , InSb--InAs , ZnTe--ZnSe и др.

2.10. Диаграммы плавкости бинарных систем с ограниченными твердыми растворами

Часто компоненты, полностью взаимно растворимые в жидкой фазе, имеют ограниченную взаимную растворимость в твердом состоянии. В этом случае при малых концентрациях растворенного вещества краевые участки диаграмм плавкости (вблизи чистых компонентов А и В) могут иметь только два вида, показанные на рис. 2.13 б и в. Эти краевые участки по форме должны быть частью диаграмм, характерных для систем с непрерывными твердыми растворами, которые приведены на рис. 2.15 и 2.16.

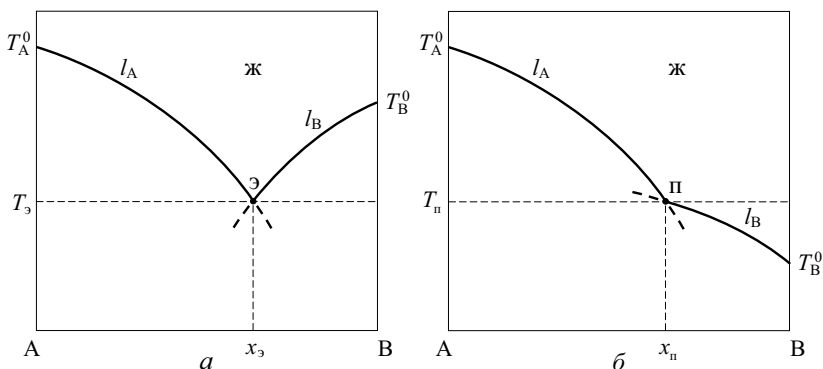


Рис. 2.17. Кривые ликвидуса в бинарных системах с ограниченными твердыми растворами: а — диаграмма с эвтектикой, б — диаграмма с перитектикой

В зависимости от характера взаимного расположения и пересечения двух линий ликвидуса l_A и l_B (являющихся геометрическим местом температур начала кристаллизации двух твердых растворов α (В в A^T) и β (А в B^T)) возможны два случая, которые приведены на рис. 2.17:

- точка пересечения линий ликвидуса (инвариантная *эвтектическая точка* э) соответствует температуре $T_э$, лежащей ниже обеих температур T_A^0 и T_B^0 кристаллизации чистых компонентов;
- точка пересечения линий ликвидуса (инвариантная *перитектическая точка* п) соответствует температуре $T_п$, лежащей между температурами T_A^0 и T_B^0 кристаллизации чистых компонентов.

Возникающие диаграммы плавкости по характеру невариантного равновесия в первом случае называют диаграммой *эвтектического типа* (или типа сложной эвтектики), а во втором случае — *перитектического типа* (или типа перитектики).

Диаграммы плавкости эвтектического и перитектического типов показаны соответственно на рис. 2.18 а и 2.19 а. Эти диаграммы, в отличие от приведенных на рис. 2.15 и 2.16, содержат две линии солидуса s_A и s_B , соответствующие двум твердым растворам — α (В в A^T) и β (А в B^T). Каждый из этих растворов имеет предельную растворимость x_α^T и x_β^T . Это означает, что при достаточно больших концентрациях растворенного вещества (В в A^T и А в B^T), превосходящих предельные растворимости x_α^T и x_β^T , система в твердом состоянии распадается на две равновесные фазы — твердые растворы α и β , существующие совместно ($\alpha + \beta$). Если, наряду с твердыми фазами, в системе имеется также жидкий раствор, то система невариантна ($C \equiv 3 - \Phi = 0$).

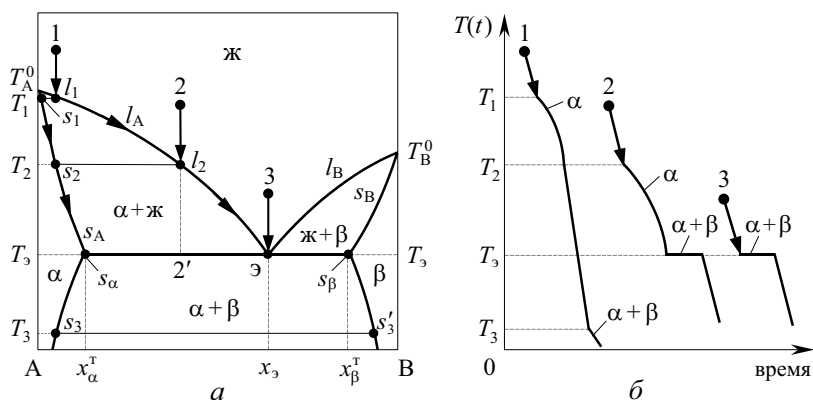


Рис. 2.18. Диаграмма плавкости типа сложной эвтектики для бинарных систем с ограниченными твердыми растворами (а) и кривые охлаждения жидкого раствора (б) для трех различных составов, обозначенных точками 1, 2, 3 на рисунке а. На разных участках кривых охлаждения буквами указаны кристаллизующиеся вещества (α , $\alpha + \beta$)

На обеих диаграммах невариантное (трехфазное) равновесие может быть реализовано лишь на горизонтальных отрезках $s_\alpha э s_\beta$ (рис. 2.18 а) и $s_\alpha s_\beta п$ (рис. 2.19 а), называемых соответственно *эвтектической* и *перитектической горизонталью*. Точкам s_α и s_β соответствуют предельные растворимости x_α^T и x_β^T , в то время как эвтектическая (э) и перитектическая (п) точки образуются пересечением линий ликвидуса l_A и l_B .

Области моновариантного (двухфазного) равновесия, обозначенные как $\alpha + \text{ж}$, $\beta + \text{ж}$ и $\alpha + \beta$, соответствуют равновесию твердых растворов α и β с жидкостью и между собой. В областях, обозначенных буквами α , β и ж , независимо существуют отдельные фазы бивариантной системы.

Переход из одной области в другую происходит при движении фигуративной точки в процессе охлаждения системы из начального положения (обозначенного цифрами 1, 2, 3) вниз по вертикали. В частности, когда фигуративная точка попадает на эвтектическую или перитектическую горизонталь (положения 1' и 2'), возникает разделение невариантной системы на три фазы ($\alpha + \beta + \text{ж}$). За время пребывания системы на площадке невариантности (см. кривые охлаждения, обозначенные цифрами 2, 3 на рис. 2.18 б и 1, 2 на рис. 2.19 б) одна из этих фаз полностью исчезает, в результате чего система снова превращается в моновариантную ($C = 1$).

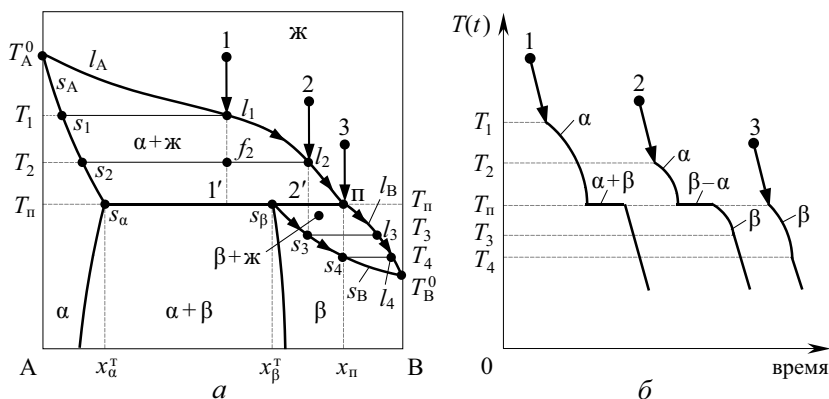


Рис. 2.19. Диаграмма плавкости типа перитектики для бинарных систем с ограниченными твердыми растворами (а) и кривые охлаждения жидкого раствора (б) для трех различных составов, обозначенных точками 1, 2, 3 на рисунке а. Коннода s_2l_2 с фигуративной точкой f_2 приведена для демонстрации правила рычага. На разных участках кривых охлаждения буквами указаны кристаллизующиеся вещества (α , β , $\alpha + \beta$, $\beta - \alpha$)

В эвтектической точке э жидкий раствор состава $x_э$ находится в равновесии с двумя твердыми фазами, из которых одна (s_α) обогащена компонентом А, а другая (s_β) — компонентом В. В отличие от этого, в перитектической точке п жидкий раствор состава x_p находится в равновесии с твердыми растворами s_α и s_β , каждый из которых богаче одним из компонентов (в данном случае компонентом А) по сравнению с составом x_p .

Из сравнения рис. 2.18 *а* с рис. 2.14 *а* и 2.16 *б* видно, что диаграмма типа сложной эвтектики (рис. 2.18 *а*) является комбинацией диаграммы типа простой эвтектики (рис. 2.14 *а*) с диаграммой типа «сигара с минимумом» (рис. 2.16 *б*). Действительно, диаграмма на рис. 2.18 *а* превращается либо в простую эвтектику при исчезновении твердых растворов (при $x_\alpha^T \rightarrow A$ и $x_\beta^T \rightarrow B$), либо в «сигару с минимумом» при смыкании концов s_α и s_β эвтектической горизонтали.

Общим признаком для всех трех диаграмм является понижение температуры кристаллизации каждого компонента при добавлении в его расплав другого вещества. В отличие от этого, для диаграммы с перитектикой (рис. 2.19 *б*) введение компонента В в расплав А понижает, а компонента А в расплав В повышает температуру кристаллизации. В этом смысле перитектическая диаграмма приближается к диаграмме типа «сигара» (рис. 2.15 *а*) и переходит в нее при смыкании точек s_α и s_β .

Рассмотрим протекание процесса равновесной кристаллизации в бинарных системах со сложной эвтектикой и перитектикой для трех исходных положений 1, 2 и 3 фигуративной точки на рис. 2.18 и 2.19.

В эвтектической системе (рис. 2.18) кристаллизация в случае 1 протекает аналогично ходу кристаллизации на рис. 2.15. Она начинается при температуре T_1 с осаждения кристаллов α состава s_1 и заканчивается при температуре T_2 осаждением кристаллов состава s_2 , равного исходному x_0^* в точке 1. Совпадение точки s_2 конца кристаллизации в случае 1 с точкой начала кристаллизации в случае 2 выбрано только ради удобства изображения на рис. 2.18 *а*.

Отличие между случаями 1, показанными на рис. 2.15 *б* и рис. 2.18 *б*, состоит в том, что для последнего при температуре $T \leq T_3$ система вновь становится двухфазной, поскольку твердый раствор α состава s_3 частично распадается с образованием твердого раствора β состава, отвечающего точке s'_3 . В процессе дальнейшего медленного охлаждения происходит перекристаллизация твердых растворов в соответствии с ходом линий $s_\alpha s_3$ и $s_\beta s'_3$. Твердая фаза в этом случае представляет собой поликристаллическую смесь твердых растворов α и β .

Кристаллизация в системах со сложной эвтектикой для случаев 2 и 3 (рис. 2.18) протекает в точности так же, как для аналогичных случаев в системе с простой эвтектикой (рис. 2.14). Разница состоит лишь в том, что вместо чистых кристаллов А^т и В^т выпадают кристаллы твердых растворов α и β , указанные

на соответствующих участках кривых охлаждения. В любой эвтектической системе при нахождении ее на площадке невариантности, когда фигуративная точка $2'$ или ε лежит на эвтектической горизонтали, из трех равновесных фаз ($\alpha + \beta + \text{ж}$) всегда исчезает жидкий раствор, т.е. заканчивается кристаллизация, а дальше возможна лишь перекристаллизация типа $\alpha \rightleftharpoons \beta$.

В отличие от этого, в перитектических системах (рис. 2.19) на площадке невариантности может исчезать либо жидкая фаза (для случая 1), либо один из твердых растворов (α для случая 2) в зависимости от расположения фигуративных точек $1'$ и $2'$ на перитектической горизонтали $s_\alpha s_\beta \text{п}$.

Начальный этап кристаллизации для случая 1 (показанный на рис. 2.19 а стрелками, идущими вдоль линий ликвидуса l_A и солидуса s_A от точек l_1 и s_1 к точкам п и s_α) при температурах выше температуры перитектики $T_{\text{п}}$ ничем не отличается от аналогичного случая 2 на рис. 2.18. Когда жидкий раствор окажется в точке п, а твердый раствор α в точке s_α , дальнейший отбор теплоты от системы приводит к появлению кристаллов твердого раствора β . Так как состав s_β этих кристаллов лежит между s_α и п, то на их образование расходуются обе фазы (α и ж). В ходе кристаллизации жидкая фаза полностью исчезает, и остаются часть неиспользованного раствора α и вновь возникшая фаза β (см. $\alpha + \beta$ на площадке невариантности для кривой 1 на рис. 2.19 б). При дальнейшем охлаждении возможна перекристаллизация типа $\alpha \rightleftharpoons \beta$.

При кристаллизации в случае 2 начальный этап до температуры перитектики $T_{\text{п}}$ ничем не отличается от предыдущего. Положение фигуративной точки $2'$ на перитектической горизонтали правее точек s_α и s_β свидетельствует о том, что в данном случае система обеднена компонентом А по сравнению как с ранее возникшими кристаллами α (состава x_α^T в точке s_α), так и с вновь образующимися кристаллами β (состава x_β^T в точке s_β). Поэтому образование последних идет за счет полного исчезновения первых совместно с частичным потреблением вещества из жидкой фазы. Обозначение $\beta - \alpha$ на площадке невариантности для кривой охлаждения 2 на рис. 2.19 б отражает процесс появления фазы β и исчезновения фазы α . После ее полного исчезновения система вновь становится моновариантной, содержащей фазу β состава x_β^T и жидкость состава $x_{\text{п}}$. Дальнейшее охлаждение сопровождается выделением кристаллов β вплоть до температуры T_3 , при которой вертикаль, идущая вниз от исходной точки 2, пересекает линию солидуса s_β в точке s_3 .

В процессе кристаллизации составы жидкого и твердого растворов изменяются по линиям ликвидуса l_B и солидуса s_B в направлении стрелок, идущих от точек p и s_β к точкам l_3 и s_3 . При температуре T_3 процесс кристаллизации полностью заканчивается и в системе существуют лишь кристаллы твердого раствора β , состав которых равен исходному составу $x_0^{\text{ж}}$ жидкого раствора в точке 2.

Кристаллизация в случае 3 на рис. 2.19 (или по любому другому пути правее точки 3) аналогична кристаллизации одного твердого раствора (в данном случае β) из раствора-расплава, например, на рис. 2.15. В рассматриваемом случае из жидкого раствора перитектического состава x_p (или более богатого компонентом В) начинают осаждаться только кристаллы β состава $x_\beta^{\text{т}}$, так как в этом случае кристаллы α на перитектической горизонтали неустойчивы. При охлаждении ниже температуры T_p составы жидкого и твердого растворов изменяются вдоль линий ликвидуса l_B и солидуса s_B в направлении стрелок, идущих от точек p и s_β к точкам l_4 и s_4 . В этих точках при температуре T_4 заканчивается кристаллизация жидкого раствора перитектического состава.

Принципиальная разница между процессами кристаллизации в бинарных системах с эвтектикой и перитектикой состоит в следующем. В системах с эвтектикой из жидкой фазы, обогащенной по сравнению с эвтектическим составом компонентом А (или В), всегда выделяются кристаллы твердого раствора α (или β), более богатые тем же компонентом А (или В), а в эвтектической точке одновременно выпадают две твердые фазы. В системах же с перитектикой из жидкой фазы любого состава выделяются кристаллы твердого раствора α или β , всегда обогащенные одним и тем же компонентом (А на рис. 2.19) по сравнению с исходным жидким раствором. В перитектической точке одна твердая фаза выпадает, а другая растворяется.

Бинарные системы с ограниченной растворимостью компонентов в твердой фазе довольно многочисленны. Например, системе $\text{Cu}-\text{Ag}$ соответствует диаграмма плавкости типа сложной эвтектики (рис. 2.18 а) с характерными температурами: $T_{\text{Cu}}^0 = 1083^\circ\text{C}$, $T_{\text{Ag}}^0 = 960^\circ\text{C}$, $T_3 = 780^\circ\text{C}$, имеющая следующий состав на эвтектической горизонтали: $x_\alpha^{\text{т}} = 8\%$ (вес.), $x_3 = 72\%$ (вес.), $x_\beta^{\text{т}} = 91,2\%$ (вес.) для серебра. Эвтектический сплав этой системы широко применяется в качестве припоя, известного под маркой ПСр72, с температурой плавления $T_3 = 780^\circ\text{C}$.

Преимущество эвтектического припоя проявляется при многоступенчатой пайке узлов из медных деталей. В процессе пайки при температуре $T_0 > T_э$ происходит растворение меди в расплаве, что соответствует на диаграмме плавкости (рис. 2.18 а) смещению из точки э вверх влево по линии ликвидуса l_A . Быстрое охлаждение деталей после выдержки при температуре T_0 формирует паяный шов с температурой плавления, превышающей $T_э$. Это допускает повторные нагревы паяного шва и последующие пайки с другими деталями из меди при последовательном снижении температуры в диапазоне от T_0 до $T_э$.

Диаграмма плавкости перитектического типа (рис. 2.19 а) присуща, например, системе Pt–Ag с характерными температурами: $T_{Pt}^0 = 1770^\circ\text{C}$, $T_{Ag}^0 = 960^\circ\text{C}$, $T_n = 1185^\circ\text{C}$, имеющей следующий состав на перитектической горизонтали: $x_\alpha^T = 20\%$ (вес.), $x_\beta^T = 61\%$ (вес.), $x_n = 80\%$ (вес.) для серебра.

Кроме вышеуказанных систем, ограниченную взаимную растворимость имеют также Si–Al (Pb), Ge–Ag (Sb), Pb–Cu (Zn, Sn, Al), NaNO_3 – KNO_3 (AgNO_3) и др.

2.11. Диаграммы плавкости бинарных систем с химическими соединениями в твердой фазе

Между компонентами А и В в твердой фазе возможно взаимодействие с образованием химических соединений постоянного (стехиометрического) или переменного (нестехиометрического) состава с кристаллической решеткой, отличной от решеток A^T и B^T . Возможные химические соединения могут быть отнесены к одному из нижеследующих типов.

1. Химические соединения постоянного состава A_mB_n , построенные по правилам валентности из элементов разных групп (I–VII) периодической таблицы, например, $A^I B^{VII}$, $A^{II} B^{VI}$, $A^{III} B^V$, $A^{IV} B^{IV}$, $A^I B_3^V$, $A_2^I B_3^{VI}$, $A^{II} B_2^V$, $A_3^{II} B_2^{VII}$, $A_2^{III} B^{VI}$, $A_2^{III} B_2^{VII}$. Каждое из этих соединений имеет так называемую *электронную конфигурацию* e_1 , равную 4. Она определяется как среднее число валентных электронов, приходящихся на один атом соединения:

$$e_1 = \frac{me_A + ne_B}{m + n}.$$

Здесь e_A — число валентных электронов в атоме А, равное I, II, III, IV; e_B — число валентных электронов в атоме В, равное IV, V, VI, VII.

Наибольшее применение нашли химические соединения: а) типа $A^I B^{VII}$ (LiF, NaCl, CuCl, AgI и др.), называемые *галогенидами*; б) типа $A^{II} B^{VI}$ (CdS, CdSe, CdTe, ZnS, HgTe и др.), называемые *халькогенидами*; в) типа $A^{III} B^V$ (GaAs, GaP, GaN, InP, AlAs и др.); г) типа $A^{IV} B^{IV}$ (карбид кремния SiC различных политипных модификаций).

2. Интерметаллические соединения металлов $A_m B_n$, построенные с нарушением правила валентности, для которых электронная конфигурация принимает значения: а) $e_1 = 3/2$ (CuZn, Cu₃Al, Cu₅Sn, FeAl и др.); б) $e_1 = 21/13$ (Cu₅Zn₈, Cu₉Al₄, Cu₃₁Sn₈, Fe₅Zn₂₁ и др.); в) $e_1 = 7/4$ (CuZn₃, Cu₃Sn, Cu₅Al₃, CuCd₃, FeZn₇ и др.).

3. Химические соединения переменного состава $A_m B_{n(1 \pm \delta)}$, построенные на основе стехиометрического соединения $A_m B_n$ по типу твердых растворов замещения, внедрения или вычитания (см. п. 2.3). Сюда, в частности, относятся: а) *карбиды* (WC, W₂C, TiC, TaC и др.); б) *нитриды* (TiN, TaN, ZrN, Fe₂N, Fe₄N и др.); в) *оксиды* (TiO, TiO₂, FeO, Fe₂O₃, BaTiO₃ и др.). Интервал изменения состава в окрестности соединения $A_m B_n$ (выражаемый параметром δ), в пределах которого существует соединение переменного состава, принято называть *областью гомогенности*. Химические соединения переменного состава, которые могут быть выражены конкретной формулой $A_m B_{n(1 \pm \delta)}$ и обладают достаточно высокой степенью упорядоченности, получили название *дальтони́дов*.

4. Фазы переменного состава, возникшие по типу твердых растворов, в области существования которых (также называемой *областью гомогенности*) нет, однако, ни одного стехиометрического соединения типа $A_m B_n$. Такие промежуточные фазы, не выражаемые химической формулой и имеющие малую степень упорядоченности, называют *бертоллидами*.

Приведенное разделение на четыре типа химических соединений в твердой фазе является до некоторой степени условным. Практически не существует строго стехиометрических соединений с нулевой областью гомогенности. В соединениях первого типа область гомогенности, хоть и значительно более узкая по сравнению с дальтонидами, но обычно присутствует. В частности, это касается халькогенидов типа $A^{II} B^{VI}$ (CdS, CdSe, CdTe и др.), в которых оба элемента достаточно летучие, что приводит к нарушению стехиометричности за счет дефектов решетки. Интерметаллические соединения также обычно имеют достаточно

широкую область существования и могут быть отнесены к соединениям переменного состава.

Диаграммы плавкости бинарных систем с химическими соединениями первых трех типов получаются комбинацией ранее рассмотренных диаграмм эвтектического и перитектического типа, которые приведены на рис. 2.14 а, 2.18 а и 2.19 а. Совокупность всех химических или интерметаллических соединений A_mB_n с разными m и n изображается на диаграмме в виде вертикальных линий, идущих вниз из точек плавления соединений. Эти линии разбивают всю плоскость T – x на ряд областей, каждая из которых содержит одну из диаграмм эвтектического или перитектического типа. Две возможные комбинации трех вышеуказанных диаграмм показаны на рис. 2.20 а и б для системы с одним химическим или интерметаллическим соединением. На первом из них соединение A_mB_n строго стехиометрического состава, на втором — переменного состава с областью гомогенности в виде двух твердых растворов γ_1 и γ_2 (один из которых может отсутствовать, тогда область гомогенности смещается в одну сторону от стехиометрического состава).

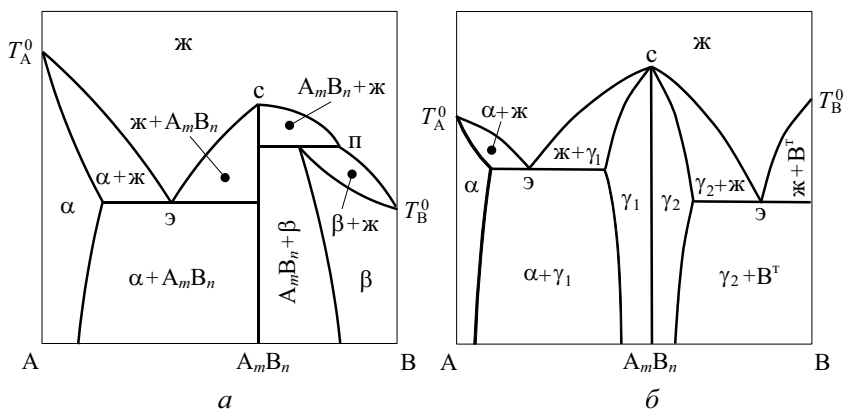


Рис. 2.20. Дистектические диаграммы плавкости бинарных систем с конгруэнтно плавящимся химическим соединением A_mB_n постоянного состава (а) и переменного состава (б)

Фазы, существующие в различных областях диаграммы, указаны теми же буквами, что и ранее, а буквы э, п и с отмечают, соответственно, точку эвтектики, точку перитектики и точку кристаллизации химического соединения из раствора того же состава (называемую *дистектической точкой*). Процесс кристаллизации в каждой из областей протекает аналогично

описанному в пп. 2.8 и 2.10, только вместо чистых веществ А и В или твердых растворов α и β в соответствующих областях кристаллизуется соединение A_mB_n или твердые растворы γ_1 и γ_2 на его основе. Так как исходный состав жидкого раствора в дистектической точке строго стехиометричен, то система при охлаждении ведет себя как однокомпонентная: при температуре кристаллизации соединения выпадают кристаллы A_mB_n , а на кривой охлаждения появляется площадка невариантности, как в случае чистых веществ А и В (см. кривые 1 и 5 на рис. 2.14 б).

Диаграммы состояния, аналогичные показанным на рис. 2.20 (т. е. содержащие соединения постоянного или переменного состава с химической формулой типа A_mB_n , кристаллизующиеся из жидкого раствора стехиометрического состава), называют *дистектическими диаграммами*. В отличие от них, для систем с бертоллидными фазами в области гомогенности отсутствует вертикальная линия, соответствующая химическому соединению. Так, на рис. 2.20 б в случае бертоллида исчезает вертикальная прямая, разделяющая твердые растворы γ_1 и γ_2 , на месте которых размещается непрерывный твердый раствор γ , образующий единую бертоллидную фазу.

Все вышесказанное в отношении дистектических диаграмм справедливо для химических соединений с так называемым *конгруэнтным плавлением*. В этом случае соединение A_mB_n устойчиво вплоть до температуры своего плавления, выше которой образуется раствор–расплав строго стехиометрического состава ($mA^ж + nB^ж$). Наряду с этим, в природе существуют и соединения, плавящиеся инконгруэнтно («(ин)конгруэнтный» от лат. *(не)совпадающий*). *Инконгруэнтное плавление* происходит в тех случаях, когда соединение неустойчиво при высоких температурах. Не доходя до своей истинной температуры плавления, оно разлагается по схеме $A_mB_n \rightleftharpoons ж + т$ (или п) с образованием жидкости ж и новой твердой фазы т (или пара п). При этом жидкий раствор ($m'A^ж + n'B^ж$) отличается по стехиометрии от исходной твердой фазы A_mB_n , из которой эта жидкость образовалась.

Типичный вид фазовой диаграммы с инконгруэнтно плавящимся соединением приведен на рис. 2.21. Здесь $T_{пл}$ является реальной температурой инконгруэнтного плавления соединения A_mB_n с образованием чистого твердого компонента $A^т$ и жидкости состава, отвечающего точке п. Горизонтальная линия $T_{пл,сп}$, вдоль которой выполняется невариантное равновесие трех фаз ($A^т + A_mB_n + ж$), есть не что иное, как перитектическая горизонталь, а точка п — перитектическая точка, которые рассмотрены

в предыдущем параграфе. Истинная температура плавления соединения соответствует максимуму штриховой кривой и лежит в области неустойчивости соединения A_mB_n . Кристаллизация жидкого раствора, имеющего состав правее точки п на рис. 2.21, ничем не отличается от аналогичных процессов, рассмотренных для систем с эвтектикой (см. рис. 2.14 и 2.18).

Растворы с составами левее точки п кристаллизуются по перитектическому типу, аналогично показанному на рис. 2.19. Для исходной фигуративной точки 1 (рис. 2.21) сначала кристаллизуется чистое вещество A^T при охлаждении вплоть до температуры $T_{пл}$, когда жидкость окажется в точке п, а система в целом — в точке 1'. Здесь начинают выпадать кристаллы A_mB_n до тех пор, пока не будет израсходован весь раствор-расплав. В результате возникает смесь кристаллов $A + A_mB_n$.

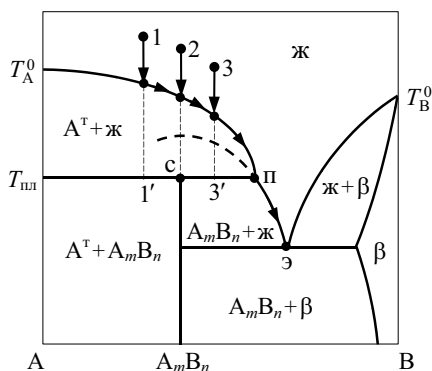


Рис. 2.21. Диаграмма плавокости бинарных систем с инконгруэнтно плавящимся химическим соединением A_mB_n

При охлаждении жидкого раствора из точки 2 (рис. 2.21) начальный этап кристаллизации протекает аналогично рассмотренному с образованием A^T . На этапе перитектической кристаллизации при температуре $T_{пл}$ (где имеет место невариантное равновесие трех фаз — A^T , A_mB_n и ж) образование кристаллов A_mB_n неизбежно происходит за счет растворения ранее появившихся кристаллов A^T . Они исчезнут одновременно с полным израсходованием жидкого раствора. В результате этого останутся лишь кристаллы химического соединения A_mB_n .

При охлаждении раствора из точки 3 (рис. 2.21) выделившиеся кристаллы A^T будут полностью израсходованы на образование кристаллов A_mB_n еще до того, как исчезнет вся жидкость. Поэтому на данном этапе кристаллизация не заканчивается, так

что дальнейшее охлаждение сопровождается изменением состава жидкого раствора вдоль линии ликвидуса от точки P к точке $Э$ с образованием кристаллов A_mB_n . При этом возникает новое инвариантное эвтектическое равновесие трех фаз — кристаллов A_mB_n , жидкого раствора $ж$ эвтектического состава и твердого раствора β . Здесь полностью заканчивается процесс кристаллизации с образованием эвтектической смеси кристаллов $A_mB_n + \beta$.

Многие соединения на основе слабо летучего компонента A (элементы I, II или III группы) и легко летучего неметаллического компонента B (элементы VII группы Cl, Br, I (галогены), элементы VI группы S, Se, Te (халькогены) или элементы V группы P, As, Sb) при нагревании, еще находясь в твердом состоянии, начинают разлагаться с выделением легко летучего компонента. Далее они инконгруэнтно плавятся и образуют не только жидкость, но и паровую фазу, состоящую из одного или обоих компонентов. Так ведут себя многие полупроводниковые соединения типа $A^{III}B^V$ ($GaAs, GaP, InP, InSb$ и др.) и $A^{II}B^{VI}$ ($ZnS, CdS, CdSe, CdTe$ и др.). В первом случае газовая фаза образована летучим компонентом B^V , а во втором — обоими компонентами A^{II} и B^{VI} .

Для описания бинарной системы в такой ситуации необходимо учитывать три независимые переменные: температуру T , состав раствора x и давление паров летучего компонента p_B (газовые молекулы которого могут иметь разные молекулярные формы B, B_2, B_4 и т. п.). Для примера на рис. 2.22 *a* приведена диаграмма состояния системы $Ga-As$, имеющей одно химическое соединение — арсенид галлия $GaAs$, который находит широкое применение в полупроводниковых приборах и интегральных микросхемах сверхвысокочастотного и оптического диапазона.

Показанная на рис. 2.22 *a* кривая ликвидуса системы $Ga-As$ может быть получена только при обеспечении в газовой фазе равновесного давления паров мышьяка в форме молекул As_2 и As_4 . Это давление меняется вдоль кривой ликвидуса, так что: а) $p_{As_2} > p_{As_4}$ при $T < 930^\circ C$; б) $p_{As_2} = p_{As_4} \approx 2 \cdot 10^{-5}$ атм при $T = 930^\circ C$; в) $p_{As_2} = 0,074$ атм и $p_{As_4} = 0,902$ атм при температуре плавления $GaAs$, равной $1238^\circ C$.

Для чистого мышьяка его плавление при $817^\circ C$ может быть наблюдаемо лишь при обеспечении избыточного давления паров мышьяка вплоть до 36 атм (в обычных условиях мышьяк сублимируется без плавления). Наоборот, галлий с низкой температурой плавления ($T_{Ga} = 29,8^\circ C$) имеет очень малое давление насыщенного пара ($p_{Ga} \approx 10^{-4}$ атм при $T_{пл} = 1238^\circ C$).

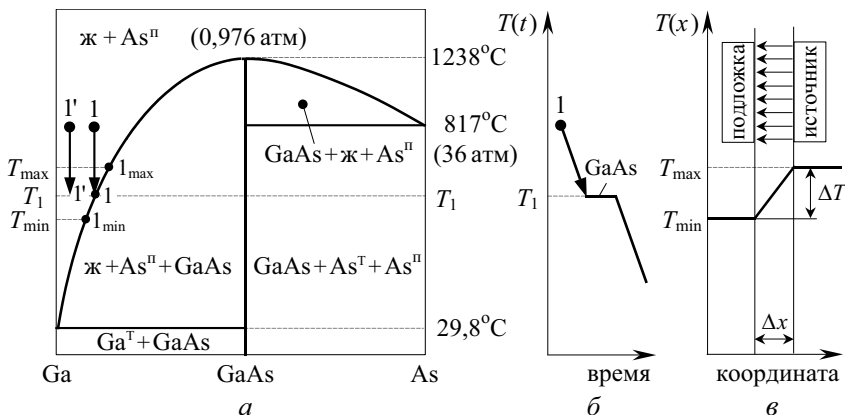


Рис. 2.22. Диаграмма фазового состояния системы Ga–As (а), кривая принудительного охлаждения жидкого раствора (б) и пространственное распределение температуры при кристаллизации с температурным градиентом (в). На рисунке а в скобках указано равновесное давление паров мышьяка As^{II} , возрастающее при движении вдоль линии ликвидуса от Ga к As

Диссоциация арсенида галлия начинается с температур около 600°C , при этом $p_{\text{As}_2} = p_{\text{Ga}} \approx 10^{-8}$ атм. Жидкая фаза состоит из расплава галлия, насыщенного в соответствии с кривой ликвидуса мышьяком, который находится в равновесии с собственными парами, обозначенными на диаграмме как As^{II} . Так как в правой части диаграммы (при $x_{\text{As}} > 0,5$) давление паров мышьяка велико (более 1 атм), то для практических целей при выращивании GaAs из жидкой фазы используют расплав галлия с малым содержанием мышьяка: при $x_{\text{As}}^{\text{ж}} = 0,1$ температура начала кристаллизации примерно 850°C .

Процессы кристаллизации с принудительным охлаждением раствора-расплава (см. рис. 2.22 б), а также с использованием температурного градиента (см. рис. 2.22 в) будут рассмотрены ниже в п. 4.13.

Во всех рассмотренных диаграммах плавкости жидкая фаза предполагалась однородной при любом составе системы. В ряде случаев однородность жидкой фазы может нарушаться путем ее расслоения на два несмешивающихся жидких раствора ($\text{B}^{\text{ж}}$ в $\text{A}^{\text{ж}}$ и $\text{A}^{\text{ж}}$ в $\text{B}^{\text{ж}}$). Эти два раствора существуют на плоскости $T-x$ в определенной области жидкого состояния, ограниченной линиями расслаивания, которые характеризуют составы равновесных жидких фаз. Подобная картина может наблюдаться и в твердом состоянии, когда при некоторых условиях однородный твердый

раствор распадается на два раствора, одновременно существующие в определенной области диаграммы состояния.

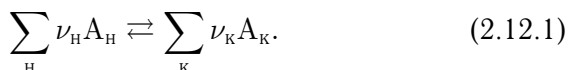
Описанная неустойчивость жидкой или твердой фазы, приводящая к появлению новых областей двухфазного равновесия, может существенно усложнить вид диаграмм плавкости.

До сих пор рассматривались основные закономерности равновесия между фазами и управления фазовыми превращениями в системах с химически пассивными компонентами. Учет химических соединений, выполненный в п. 2.11, проводился только с позиции их влияния на диаграммы плавкости бинарных систем.

В нижеследующих параграфах переходим к последовательно-му рассмотрению основных положений химической термодинамики, то есть термодинамики систем с химически реактивными компонентами. Общую основу такого рассмотрения составляют закон действующих масс, характерный для равновесных химических реакций, и уравнения изотермы, изобары и изохоры Вант-Гоффа¹⁾. Ионное равновесие в водных растворах электролитов и термодинамика электрохимических систем, включая электродные процессы, завершают материал данной главы.

2.12. Закон действующих масс и константы химического равновесия

Пусть в системе протекает обратимая химическая реакция, записанная в обобщенной форме



Здесь индексы «н» и «к» относятся к химическим реагентам A_n и A_k со стехиометрическими коэффициентами ν_n и ν_k , которые условно названы *начальными* (с индексом «н» в левой части уравнения) и *конечными* (с индексом «к» в правой части уравнения) реагентами. Так как общие условия равновесия фаз, выражаемые равенствами (1.14.1) и (1.14.2), являются изобарно-

¹⁾ Вант-Гофф (Van't Hoff) Якоб Хендрик (1852–1911) — голландский химик, один из основателей физической химии и стереохимии, исследовал химическое сродство и кинетику реакций, вывел основные уравнения химической термодинамики (уравнения изотермы, изобары и изохоры Вант-Гоффа), заложил основы теории разбавленных растворов, вывел закон осмотического давления (закон Вант-Гоффа). Нобелевская премия по химии (1901) «в знак признания огромной важности открытия законов химической динамики и осмотического давления в растворах».

изотермическими, то для описания химических реакций следует использовать свободную энергию Гиббса G .

Как показано в п. 1.14, общее изобарно-изотермическое условие равновесия $\Delta G = 0$ для химической реакции (2.12.1) принимает вид, даваемый уравнением (1.14.6):

$$\sum_{\text{н}} \nu_{\text{н}} \mu_{\text{н}} = \sum_{\text{к}} \nu_{\text{к}} \mu_{\text{к}}, \quad (2.12.2)$$

где $\mu_{\text{н}}$ и $\mu_{\text{к}}$ — химические потенциалы начальных и конечных участников реакции.

Установим соотношение между равновесными концентрациями веществ (или парциальными давлениями газов), участвующих в химической реакции, которое принято называть *законом действующих масс*.

1. Гомогенные химические реакции представляют простейший случай, рассмотрение которого начнем с газофазных реакций. Пусть в системе протекает реакция (2.12.1), все участники которой являются газообразными компонентами с химическими потенциалами μ_i в форме (1.13.1). Подставим это выражение в равенство (2.12.2) и преобразуем его к виду

$$\begin{aligned} \sum_{\text{к}} \nu_{\text{к}} \ln p_{\text{к}} - \sum_{\text{н}} \nu_{\text{н}} \ln p_{\text{н}} = \\ = -\frac{1}{RT} \left[\sum_{\text{к}} \nu_{\text{к}} \mu_{\text{к}}^{\circ}(T) - \sum_{\text{н}} \nu_{\text{н}} \mu_{\text{н}}^{\circ}(T) \right] \equiv f(T). \end{aligned} \quad (2.12.3)$$

Здесь в левой части под логарифмами стоят безразмерные парциальные давления $p_{\text{к(н)}}$ (равные числу атмосфер), а правая часть зависит только от температуры.

Вводя новую функцию температуры $K_p(T) = \exp f(T)$ и заменяя в левой части (2.12.3) сумму логарифмов логарифмом произведения, получаем в окончательной форме искомый **закон действующих масс**:

$$\frac{\prod_{\text{к}} p_{\text{к}}^{\nu_{\text{к}}}}{\prod_{\text{н}} p_{\text{н}}^{\nu_{\text{н}}}} = K_p(T). \quad (2.12.4)$$

Величину $K_p(T)$, зависящую только от температуры и нечувствительную к суммарному давлению P и исходному составу системы, принято называть **константой химического равновесия**. В левой части формулы (2.12.4) стоит произведение равновесных парциальных давлений для конечных участников реакции, взятых в степенях, равных их стехиометрическим коэффициентам, деленное на аналогичное произведение для начальных реагентов.

В качестве примера записи закона действующих масс рассмотрим газообразные компоненты — кислород, окись и двуокись углерода, взаимодействующие по обратимой реакции



Для этой реакции общее условие химического равновесия (2.12.2) и закон действующих масс (2.12.4) принимают следующий вид:

$$2\mu_{\text{CO}} + \mu_{\text{O}_2} = 2\mu_{\text{CO}_2} \quad \text{и} \quad \frac{p_{\text{CO}_2}^2}{p_{\text{CO}}^2 p_{\text{O}_2}} = K_p(T).$$

В выражении (2.12.4) парциальные давления можно записать как $p_i = c_i RT$ (при $i = \text{н, к}$) и выразить закон действующих масс через мольно-объемные концентрации:

$$\frac{\prod_{\text{к}} c_{\text{к}}^{\nu_{\text{к}}}}{\prod_{\text{н}} c_{\text{н}}^{\nu_{\text{н}}}} = K_c(T). \quad (2.12.6)$$

Здесь введена новая константа равновесия $K_c(T)$, связанная с константой $K_p(T)$ соотношением

$$K_c(T) = K_p(T)(RT)^{-\Delta\nu^r}, \quad (2.12.7)$$

при этом

$$\Delta\nu^r = \sum_{\text{к}} \nu_{\text{к}}^r - \sum_{\text{н}} \nu_{\text{н}}^r \quad (2.12.8)$$

— изменение числа молей в газовой фазе, происходящее при одном пробеге реакции слева направо (для реакции (2.12.5) $\Delta\nu^r = -1$).

Парциальные давления можно представить в форме $p_i = x_i P$ (при $i = \text{н, к}$), тогда закон действующих масс (2.12.4) принимает вид

$$\frac{\prod_{\text{к}} x_{\text{к}}^{\nu_{\text{к}}}}{\prod_{\text{н}} x_{\text{н}}^{\nu_{\text{н}}}} = K_x(T, P), \quad (2.12.9)$$

где введена другая константа равновесия $K_x(T, P)$, связанная с $K_p(T)$ следующим соотношением:

$$K_x(T, P) = K_p(T) P^{-\Delta\nu^r}. \quad (2.12.10)$$

Если химическая реакция, оставаясь гомогенной, протекает не в газовой, а в конденсированной фазе, то химический потенциал μ_i каждого компонента ($i = \text{н, к}$) выражается через его активность $a_i = \gamma_i x_i$ в виде (2.5.11). В этом случае прямым

аналогом «газовой» формы (2.12.9) является запись закона действующих масс в виде

$$\frac{\prod_{\text{к}} a_{\text{к}}^{\nu_{\text{к}}}}{\prod_{\text{н}} a_{\text{н}}^{\nu_{\text{н}}}} = K_a(T, P). \quad (2.12.11)$$

Константа равновесия $K_a(T, P)$, входящая в (2.12.11), сформирована, по аналогии с выражениями (2.12.3) и (2.12.4), из стандартных химических потенциалов $\mu_i^{\text{ок}}(T, P)$ для чистых конденсированных фаз (при $x_i = 1$). В общем случае они зависят как от температуры T , так и общего давления P в системе. Однако последней зависимостью обычно возможно пренебречь ($\mu_i^{\text{ок}}(T, P) \approx \mu_i^{\text{ок}}(T)$) в силу равенства (2.2.10), то есть можно считать $K_a(T, P) \approx K_a(T)$.

Для идеального раствора каждый компонент имеет $\gamma_i = 1$ и $a_i = x_i$. Тогда закон действующих масс (2.12.11) принимает форму (2.12.9) с выражением состава через мольные доли x_i , полученную для идеального газа. Поскольку в рамках модели идеального раствора отсутствуют объемные эффекты ($\Delta V = 0$, см. п. 2.4), то к идеальным растворам применима также и другая «газовая» форма (2.12.6), построенная на мольно-объемных концентрациях $c_i = n_i/V$.

В частном случае предельно разбавленного раствора левая часть закона действующих масс (2.12.11) не содержит вклада от растворителя (с индексом $i = 1$), так как для него, согласно (2.4.28), $\gamma_1 = 1$ и $a_1 = x_1 \approx 1$. Остающийся вклад присутствующих в малых количествах примесей придает выражению (2.12.11) форму (2.12.9). Действительно, согласно (2.4.31), коэффициенты активности примесей γ_i ($i \neq 1$) зависят только от температуры, тогда их вклад может быть объединен с константой $K_a(T)$ в виде

$$K_x(T) = K_a(T) \frac{\prod_{\text{н}} \gamma_{\text{н}}^{\nu_{\text{н}}}(T)}{\prod_{\text{к}} \gamma_{\text{к}}^{\nu_{\text{к}}}(T)}. \quad (2.12.12)$$

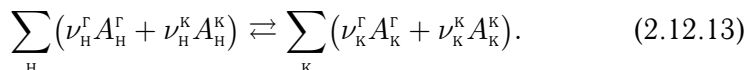
К предельно разбавленным растворам также применим закон действующих масс в форме (2.12.6), содержащей мольно-объемные концентрации $c_i = x_i c$. При этом общая концентрация раствора $c = n/V$ определяется в основном растворителем (с индексом $i = 1$) и может считаться постоянной в силу сравнительно малого содержания примесей, т. е. $c \approx c_1 = n_1/V = \text{const}$. Тогда подстановка равенства $x_i = c_i/c$ в выражение (2.12.9) приводит его к виду (2.12.6), а константой химического равновесия стано-

вится величина $K_c(T) = K_x(T) c^{\Delta\nu^k}$, где $\Delta\nu^k = \sum_k \nu_k^k - \sum_n \nu_n^k$ — изменение числа молей в конденсированной фазе, происходящее за один пробег реакции слева направо.

Следовательно, для идеальных и предельно разбавленных растворов всегда применим закон действующих масс, записанный в одной из форм (2.12.6) или (2.12.9), которые были получены для идеального газа. «Концентрационная» форма (2.12.6) закона действующих масс используется также при квазихимическом описании дефектов в полупроводниках, ведущих себя подобно примесям в предельно разбавленных растворах (см. гл. 3).

До сих пор рассматривались лишь гомогенные химические реакции, протекающие в одной фазе (газовой, жидкой или твердой). В этом случае равновесный состав многокомпонентной фазы характеризовался парциальными давлениями p_i , мольно-объемными концентрациями c_i , мольными долями x_i или активностями a_i компонентов.

2. Гетерогенные химические реакции определенным образом видоизменяют форму закона действующих масс. Пусть имеем дело с реакцией между конденсированной смешанной фазой, содержащей твердые и жидкие растворы (характеризуемые активностями компонентов $a_i = \gamma_i x_i$, $i = n, k$), и газовой фазой. Последнюю по-прежнему считаем идеальной, что не требует замены парциальных давлений p_i их фугитивностями $f_i = \gamma_i p_i$ (см. сноску на с. 223). По аналогии с формулой (2.12.1) запишем эту реакцию в обобщенной форме:



Здесь по-прежнему нижние индексы «n» и «k» соответствуют начальным и конечным реагентам, а верхние индексы «г» и «к» отражают газообразное и конденсированное состояния веществ.

На основании вышеизложенного закон действующих масс для реакции (2.12.13) записывается в виде

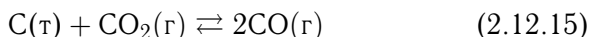
$$\frac{\prod_k p_k^{\nu_k^g} a_k^{\nu_k^k}}{\prod_n p_n^{\nu_n^g} a_n^{\nu_n^k}} = K_{pa}(T). \quad (2.12.14)$$

Уравнение (2.12.14) есть наиболее общая форма записи закона действующих масс. Если конденсированная фаза подчиняется модели идеального раствора, то активности соответствующих веществ заменяют на их мольные доли. Для чистых конденсированных фаз (как и для растворителя в предельно разбавленном

растворе) имеем $\gamma_i = 1$ и $a_i = x_i \approx 1$. В результате этого они исключаются из равенства (2.12.14), которое принимает форму (2.12.4). В то же время для примесей в предельно разбавленном растворе коэффициенты активности $\gamma_i(T)$, будучи функциями только температуры, оказываются включенными в константу равновесия $K_{px}(T)$, связанную с $K_{pa}(T)$ тем же дробным множителем, который входит в соотношение (2.12.12) между $K_x(T)$ и $K_a(T)$. При этом в равенстве (2.12.14) активности a_i примесей заменены на их мольные доли x_i .

Таким образом, если в реакции участвуют только газы и *чистые* конденсированные вещества (без жидких и твердых растворов), то закон действующих масс сохраняет форму (2.12.4), полученную для гомогенной газофазной реакции. В этом случае левая часть (2.12.4) содержит парциальные давления только газообразных компонентов, несмотря на участие в реакции и конденсированных веществ. Последнее может быть интерпретировано как результат включения парциальных давлений $p_i(T)$ насыщенного пара чистых веществ (зависящих, согласно (2.2.11), только от T) в температурную зависимость константы равновесия $K_{pa}(T)$, которая входит в (2.12.14) и заново переобозначена в выражении (2.12.4) как $K_p(T)$.

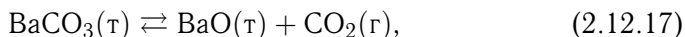
Например, для гетерогенной химической реакции



закон действующих масс записываем в виде

$$K_p(T) = \frac{p_{\text{CO}}^2}{p_{\text{CO}_2}}. \quad (2.12.16)$$

Если в реакции участвует всего один газообразный продукт, возникающий как результат разложения исходного твердого соединения, то равновесное давление этого газа называется **давлением диссоциации** химического соединения. Например, для реакции разложения карбоната бария BaCO_3 (играющей важную роль в процессе формирования эмиссии оксидного катода при его термообработке на откачном посту, см. приложение А.1),



закон действующих масс, включающий давление диссоциации p_{CO_2} , имеет вид

$$K_p(T) = p_{\text{CO}_2}. \quad (2.12.18)$$

Следовательно, давление диссоциации твердого химического соединения BaCO_3 (созданное молекулами газа CO_2), как и давление его насыщенного пара (созданное молекулами BaCO_3

в газовой фазе), зависит только от температуры T и практически не зависит от полного давления P в системе. Однако последнее оказывает влияние на характер разложения соединения. Так как процесс разложения вещества всегда требует подвода тепла ($\Delta H > 0$), то, как увидим ниже (см. п. 2.14), с увеличением температуры константа равновесия $K_p(T)$, а значит и давление $p_{\text{CO}_2}(T)$, всегда возрастают.

Температуру, при которой давление диссоциации сравнивается с внешним давлением, называют *температурой разложения* соединения ($T_{\text{разл}}$). При $T < T_{\text{разл}}$ процесс разложения происходит медленно и спокойно, а при $T \geq T_{\text{разл}}$ начинается интенсивное разложение с бурным газовыделением. В этом смысле температура разложения твердого вещества аналогична температуре кипения жидкости.

Закон действующих масс, записанный в одной из вышеуказанных форм, связывает между собой равновесные парциальные давления газов, концентрации или активности всех участников реакции таким образом, что при фиксированной температуре ничто не может быть изменено без нарушения равновесия. При анализе химических процессов константы равновесия для любой формы записи закона действующих масс считаются известными функциями температуры, найденными экспериментально или рассчитанными в стандартных условиях по известной методике (описанной, например, в приложении А).

Ниже будет установлен математический критерий в форме уравнения изотермы Вант-Гоффа, который показывает направление возможного протекания химической реакции при нарушении исходных равновесных условий.

2.13. Направление протекания химической реакции. Уравнение изотермы Вант-Гоффа

В п. 1.14 для химической реакции типа (2.12.1) был определен изобарный потенциал ΔG в виде (1.14.5) как изменение $G_{\text{к}} - G_{\text{н}}$, происходящее в системе при химическом превращении веществ за один пробег реакции слева направо, т. е. в расчете на один акт взаимодействия начальных веществ, записанных слева в реакции (2.12.1).

Из этого определения вытекают следующие заключения:

- а) при $\Delta G = 0$ любая реакция находится в равновесии;
- б) при $\Delta G = G_{\text{к}} - G_{\text{н}} < 0$ — реакция смещена вправо, т. е. самопроизвольно протекает слева направо, так как $G_{\text{н}} > G_{\text{к}}$;

в) при $\Delta G = G_{\text{к}} - G_{\text{н}} > 0$ — реакция смещена влево, т. е. самопроизвольно протекает справа налево, так как $G_{\text{к}} > G_{\text{н}}$.

Два последних заключения являются следствием того, что при $P = \text{const}$ система стремится к минимуму свободной энергии Гиббса (см. рис. 1.6).

Сначала рассмотрим гетерогенные химические реакции с участием газовой фазы и чистых конденсированных веществ в отсутствие растворов. Полученные ниже результаты нетрудно будет обобщить на реакции с участием растворов, используя результаты предыдущего параграфа.

Как известно, газообразные реагенты, в соответствии с выражением (1.13.1), имеют химический потенциал

$$\mu_i^{\text{г}}(T, p_i) = \mu_i^{\text{ог}}(T) + RT \ln p_i.$$

Для чистых конденсированных веществ их химический потенциал практически зависит только от температуры T , поскольку, согласно (2.2.10), имеем $\mu_i^{\text{ок}}(T, P) \approx \mu_i^{\text{ок}}(T)$. Тогда *изобарный потенциал* (1.14.5) химической реакции равен

$$\begin{aligned} \Delta G &\equiv G_{\text{к}} - G_{\text{н}} = \sum_{\text{к}} \nu_{\text{к}} \mu_{\text{к}} - \sum_{\text{н}} \nu_{\text{н}} \mu_{\text{н}} = \\ &= \Delta G_T^{\circ} + RT \left(\sum_{\text{к}} \nu_{\text{к}} \ln p_{\text{к}} - \sum_{\text{н}} \nu_{\text{н}} \ln p_{\text{н}} \right). \end{aligned} \quad (2.13.1)$$

Здесь введена функция температуры

$$\Delta G_T^{\circ} = \sum_{\text{к}} \nu_{\text{к}} \mu_{\text{к}}^{\circ}(T) - \sum_{\text{н}} \nu_{\text{н}} \mu_{\text{н}}^{\circ}(T), \quad (2.13.2)$$

называемая *стандартным изобарным потенциалом* реакции. В стандартных условиях, когда $p_{\text{н}} = p_{\text{н}}^{\circ} = 1$ атм и $p_{\text{к}} = p_{\text{к}}^{\circ} = 1$ атм, из равенства (2.13.1) действительно получаем $\Delta G = \Delta G_T^{\circ}$.

Следует обратить внимание на то, что в выражении (2.13.2) при суммировании учитываются все участники реакции (т. е. $\nu_{\text{н,к}} = \nu_{\text{н,к}}^{\text{г}}$ и $\nu_{\text{н,к}}^{\text{к}}$), а в последнем равенстве (2.13.1) — только газообразные компоненты (т. е. $\nu_{\text{н,к}} = \nu_{\text{н,к}}^{\text{г}}$). При этом парциальные давления $p_{\text{н}}$ и $p_{\text{к}}$ в выражении (2.13.1) соответствуют реальной *неравновесной* ситуации.

Чтобы отличить реальные давления компонентов от парциальных давлений в условиях химического равновесия, обозначим последние в виде $\bar{p}_{\text{н}}$ и $\bar{p}_{\text{к}}$ (в отличие от предыдущего параграфа, где все равновесные давления, входящие в закон действующих масс (2.12.4), имели обозначения $p_{\text{н}}$ и $p_{\text{к}}$).

Полагая при равновесии $\Delta G = 0$, из выражения (2.13.1) получаем

$$\Delta G_T^\circ = -RT \left(\sum_{\kappa} \nu_{\kappa} \ln \bar{p}_{\kappa} - \sum_{\text{н}} \nu_{\text{н}} \ln \bar{p}_{\text{н}} \right) \equiv -RT \ln K_p(T), \quad (2.13.3)$$

где на основании закона действующих масс (2.12.4) (с переобозначением равновесных давлений в виде $\bar{p}_{\text{н}}$ и $\bar{p}_{\kappa}$) имеем

$$K_p(T) = \frac{\prod_{\kappa} \bar{p}_{\kappa}^{\nu_{\kappa}}}{\prod_{\text{н}} \bar{p}_{\text{н}}^{\nu_{\text{н}}}}. \quad (2.13.4)$$

Выражение (2.13.3) дает важное для термодинамических расчетов соотношение между стандартным изобарным потенциалом ΔG_T° и константой химического равновесия $K_p(T)$.

Подстановка равенства (2.13.3) в формулу (2.13.1) приводит изобарный потенциал химической реакции в неравновесных условиях, когда $p_i \neq \bar{p}_i$ (при $i = \text{н}, \kappa$), к форме

$$\Delta G = RT \left[\left(\sum_{\kappa} \nu_{\kappa} \ln p_{\kappa} - \sum_{\text{н}} \nu_{\text{н}} \ln p_{\text{н}} \right) - \ln K_p(T) \right], \quad (2.13.5)$$

называемой **уравнением изотермы Вант-Гоффа**.

В выражении (2.13.5) первое слагаемое, заключенное в круглые скобки, по структуре аналогично $\ln K_p(T)$, записанному с учетом закона действующих масс (2.13.4). Однако оно содержит парциальные давления газовых компонентов ($p_{\text{н}}$ и p_{κ} , без черточек), взятые в реальных условиях, отличных от равновесных. Введем обозначение в виде новой функции,

$$\Pi(p_i) \equiv \frac{\prod_{\kappa} p_{\kappa}^{\nu_{\kappa}}}{\prod_{\text{н}} p_{\text{н}}^{\nu_{\text{н}}}}, \quad (2.13.6)$$

зависящей от реальных давлений $p_i = \{p_{\text{н}}, p_{\kappa}\}$ в системе. По структуре выражение (2.13.6) в точности совпадает с константой равновесия $K_p(T)$ в законе действующих масс (2.13.4). Последняя, в отличие от $\Pi(p_i)$, выражается через равновесные давления $\bar{p}_i$ и зависит лишь от температуры T .

С учетом обозначения (2.13.6), уравнение изотермы Вант-Гоффа (2.13.5) принимает следующий вид:

$$\Delta G = \Delta G_T^\circ + RT \ln \Pi(p_i) \equiv RT \left[\ln \Pi(p_i) - \ln K_p(T) \right]. \quad (2.13.7)$$

Уравнение (2.13.7) показывает, что соотношение между величинами $\Pi(p_i)$ и $K_p(T)$ определяет *химическое равновесие*

($\Delta G = 0$ при $\Pi = K_p$) и самопроизвольное протекание обратной реакции (2.12.1) в обоих направлениях: слева направо при $\Pi < K_p$ ($\Delta G < 0$) и справа налево при $\Pi > K_p$ ($\Delta G > 0$).

Для гетерогенных химических реакций типа (2.12.13) с участием газовой и смешанных конденсированных фаз (включая чистые вещества и растворы) закон действующих масс имеет наиболее общую форму (2.12.14). Эта форма отражает состав газа и растворов в виде определенной комбинации, связывающей парциальные давления $\bar{p}_i$ и активности $\bar{a}_i$ химических реагентов в равновесных условиях.

Напомним, что в выражении (2.12.14) черта сверху, отражающая равновесность ситуации, была опущена (для сравнения см. ниже формулу (2.13.10)).

Легко понять, что в неравновесных условиях аналогичная комбинация, включающая реальные давления p_i и активности a_i , должна дать новую функцию $\Pi(p_i, a_i)$, тождественную с (2.13.6) по смыслу и совпадающую с (2.12.14) по структуре:

$$\Pi(p_i, a_i) \equiv \frac{\prod_{\text{к}} p_{\text{к}}^{\nu_{\text{к}}^{\text{г}}} a_{\text{к}}^{\nu_{\text{к}}^{\text{к}}}}{\prod_{\text{н}} p_{\text{н}}^{\nu_{\text{н}}^{\text{г}}} a_{\text{н}}^{\nu_{\text{н}}^{\text{к}}}}. \quad (2.13.8)$$

В этом случае уравнение изотермы Вант-Гоффа (2.13.7) и закон действующих масс в форме (2.12.14) (с переобозначением равновесных величин черточкой сверху в виде $\bar{p}_i$ и $\bar{a}_i$) переписываются в следующей форме:

$$\Delta G = RT \left[\ln \Pi(p_i, a_i) - \ln K_{pa}(T) \right], \quad (2.13.9)$$

$$K_{pa}(T) = \frac{\prod_{\text{к}} \bar{p}_{\text{к}}^{\nu_{\text{к}}^{\text{г}}} \bar{a}_{\text{к}}^{\nu_{\text{к}}^{\text{к}}}}{\prod_{\text{н}} \bar{p}_{\text{н}}^{\nu_{\text{н}}^{\text{г}}} \bar{a}_{\text{н}}^{\nu_{\text{н}}^{\text{к}}}}. \quad (2.13.10)$$

Понятно, что для предельно разбавленного раствора функция $\Pi(p_i, a_i)$ вида (2.13.8) не содержит вклада от растворителя (с индексом $i = 1$), так как для него $\gamma_1 = 1$ и $a_1 = x_1 \approx 1$. Вклад же примесей учитывается в виде соответствующей комбинации их мольных долей x_i (с индексом $i \neq 1$), поскольку коэффициенты активности $\gamma_i(T)$ для них, как функции только температуры, включены в константу равновесия $K_{pa}(T)$.

До сих пор речь шла об управлении химическими реакциями, протекающими в условиях постоянного давления ($P = \text{const}$). Если же реакция осуществляется при постоянном объеме ($V = \text{const}$), то ее протеканием управляет свободная энергия Гельм-

гольца $F \equiv G - PV = \sum_i n_i \mu_i - PV$. Изменение F , равное $\Delta F \equiv \Delta G - \Delta PV$, происходящее при одном пробеге реакции слева направо и называемое *изохорным потенциалом* реакции, с учетом (2.13.1) записываем в виде

$$\begin{aligned} \Delta F &\equiv F_{\text{к}} - F_{\text{н}} = \sum_{\text{к}} \nu_{\text{к}} \mu_{\text{к}} - \sum_{\text{н}} \nu_{\text{н}} \mu_{\text{н}} - \Delta PV = \\ &= \Delta F_T^\circ + RT \left(\sum_{\text{к}} \nu_{\text{к}} \ln c_{\text{к}} - \sum_{\text{н}} \nu_{\text{н}} \ln c_{\text{н}} \right). \end{aligned} \quad (2.13.11)$$

При выводе (2.13.11) было использовано выражение (1.13.2) для химического потенциала, $\mu_i = \mu_i^\circ(T) + RT \ln c_i$, а изменение общего давления ΔP в объеме $V = \text{const}$, вызванное изменением числа газовых молей $\Delta \nu^{\text{г}}$ в виде (2.12.8), записано как $\Delta P = \Delta \nu^{\text{г}} RT/V$. В равновесных условиях, когда $\Delta F = 0$ и $c_i = \bar{c}_i$, выражение (2.13.11) дает равенство,

$$\Delta F_T^\circ = -RT \ln K_c(T), \quad (2.13.12)$$

определяющее связь между *стандартным изохорным потенциалом* ΔF_T° и константой химического равновесия $K_c(T)$. Равенство (2.13.12) аналогично (2.13.3), при этом константы равновесия $K_c(T)$ и $K_p(T)$ связаны между собой соотношением (2.12.7).

Подстановка (2.13.12) в (2.13.11) дает окончательное выражение для изохорного потенциала химической реакции в неравновесных условиях, когда $c_i \neq \bar{c}_i$ (при $i = \text{н, к}$),

$$\Delta F = RT \left[\left(\sum_{\text{к}} \nu_{\text{к}} \ln c_{\text{к}} - \sum_{\text{н}} \nu_{\text{н}} \ln c_{\text{н}} \right) - \ln K_c(T) \right], \quad (2.13.13)$$

по форме совпадающее с (2.13.5) и также называемое **уравнением изотермы Вант-Гоффа**.

Для идеальных газов, взаимодействующих с идеальными и предельно разбавленными растворами и чистыми конденсированными фазами, использование равенств $p_i = c_i RT$ и $\bar{p}_i = \bar{c}_i RT$ позволяет убедиться в том, что $\Delta F = \Delta G$. Действительно, различие между K_p и K_c , выражаемое формулой (2.12.7), компенсируется в равенствах (2.13.5) и (2.13.13) различием между p_i и c_i . В более общих случаях с участием реальных газов и растворов возможно $\Delta F \neq \Delta G$.

Таким образом, изобарный (ΔG) и изохорный (ΔF) потенциалы в реальных неравновесных условиях являются движущей силой процесса химического (а также и фазового) превращения, характеризуя так называемое *пересыщение*, возникающее в системе при условиях, отличных от равновесных. Это пересыщение

можно рассчитать по уравнению изотермы Вант-Гоффа (2.13.5) или (2.13.13), причем в случае идеальных и практически чистых сред величины ΔG и ΔF равны между собой. Однако их стандартные значения,

$$\Delta G_T^\circ = -RT \ln K_p(T) \quad \text{и} \quad \Delta F_T^\circ = -RT \ln K_c(T),$$

численно различны и используются для расчета температурных зависимостей $K_p(T)$ и $K_c(T)$. В практике термодинамических расчетов обычно вычисляют стандартный изобарный потенциал $\Delta G_T^\circ = \Delta H_T^\circ - T\Delta S_T^\circ$, пользуясь таблицами стандартных термодинамических величин (см. п. 1.13). Методика термодинамического расчета изложена в приложении А.

2.14. Температурная зависимость констант химического равновесия. Уравнения изобары и изохоры Вант-Гоффа

Для вывода уравнений, описывающих температурную зависимость констант равновесия $K_p(T)$ и $K_c(T)$, будем использовать уравнения Гиббса–Гельмгольца, полученные в п. 1.10. Поскольку при выводе этих уравнений не использовались какие-либо ограничения, то они применимы всегда, включая стандартные условия. На этом основании перепишем уравнения (1.10.15) и (1.10.16) для стандартного состояния системы, заменяя приращения термодинамических функций соответствующими стандартными приращениями (отмеченными верхним кружком):

$$\Delta U_T^\circ = \Delta F_T^\circ - T \frac{\partial \Delta F_T^\circ}{\partial T} \equiv -T^2 \frac{\partial}{\partial T} \frac{\Delta F_T^\circ}{T}, \quad (2.14.1)$$

$$\Delta H_T^\circ = \Delta G_T^\circ - T \frac{\partial \Delta G_T^\circ}{\partial T} \equiv -T^2 \frac{\partial}{\partial T} \frac{\Delta G_T^\circ}{T}. \quad (2.14.2)$$

Для стандартных свободных энергий ΔG_T° и ΔF_T° , входящих в уравнения (2.14.2) и (2.14.1), в предыдущем параграфе были получены выражения (2.13.3) и (2.13.12), связывающие их с константами химического равновесия $K_p(T)$ и $K_c(T)$.

Для реакций, протекающих в условиях $P = \text{const}$, подстановка выражения $\Delta G_T^\circ = -RT \ln K_p(T)$ в уравнение (2.14.2) дает **уравнение изобары Вант-Гоффа**:

$$\frac{d \ln K_p}{dT} = \frac{\Delta H_T^\circ}{RT^2}. \quad (2.14.3)$$

Величина ΔH_T° , называемая *стандартной изобарной теплотой* химической реакции, связана с ΔG_T° соотношением типа (1.8.14):

$$\Delta G_T^\circ = \Delta H_T^\circ - T \Delta S_T^\circ, \quad (2.14.4)$$

где ΔS_T° — *стандартная энтропия* химической реакции.

Стандартные величины ΔH_T° и ΔS_T° рассчитывают по таблицам стандартных термодинамических величин (см. приложение А). Так, для обобщенной химической реакции (2.12.1) эти величины вычисляют как изменение энтальпии и энтропии системы за один пробег реакции слева направо в стандартных условиях (см. формулу (1.5.15)):

$$\Delta H_T^\circ \equiv H_K^\circ(T) - H_H^\circ(T) = \sum_K \nu_K \Delta H_T^\circ(A_K) - \sum_H \nu_H \Delta H_T^\circ(A_H), \quad (2.14.5)$$

$$\Delta S_T^\circ \equiv S_K^\circ(T) - S_H^\circ(T) = \sum_K \nu_K S_T^\circ(A_K) - \sum_H \nu_H S_T^\circ(A_H). \quad (2.14.6)$$

Здесь $\Delta H_T^\circ(A_{K,H})$ и $S_T^\circ(A_{K,H})$ — стандартные мольные теплоты образования и стандартные энтропии, взятые из таблиц при температуре T (в частности, ΔH_{298}° и S_{298}° при 298 К, пересчитанные с помощью (1.5.10) и (1.5.11) на температуру T) для конечных (A_K) и начальных (A_H) участников реакции типа (2.12.1).

Вычисленные значения (2.14.5) и (2.14.6), отвечающие стандартному состоянию, в общем случае отличаются от аналогичных величин ΔH и ΔS при равновесии, когда $\Delta G \equiv \Delta H - T \Delta S = 0$. Последние вычисляются как изменения энтальпии ($\Delta H = H_K - H_H$) и энтропии ($\Delta S = S_K - S_H$) системы, происходящие в равновесных условиях при одном пробеге реакции слева направо. Именно эти равновесные величины $\Delta H(T)$ и $\Delta S(T)$ принято называть *изобарной теплотой* и *энтропией* химической реакции. Их отличие от стандартных величин ΔH_T° и ΔS_T° проявляется уже в том, что при равновесии имеем $\Delta H = T \Delta S$, в то время как $\Delta H_T^\circ \neq T \Delta S_T^\circ$ в силу соотношения (2.14.4), где в общем случае $\Delta G_T^\circ = -RT \ln K_p(T) \neq 0$.

Покажем, что для реакций с участием газообразных компонентов и чистых конденсированных фаз можно положить $\Delta H(T) = \Delta H_T^\circ$, в то время как $\Delta S(T) \neq \Delta S_T^\circ$.

Поскольку стандартное состояние для конденсированных фаз (в том числе и для растворов) соответствует совокупности компонентов, взятых в чистом виде при $x_i = 1$ (см. п. 1.13), то в данном случае различие между стандартным и равновесным состояниями касается только газовой фазы. Для газообразных

компонентов в стандартном состоянии $p_i = p_i^\circ \equiv 1$ атм, а при равновесии $p_i = \bar{p}_i$.

Согласно (1.12.7), энтропия одного газового моля для i -го компонента в равновесных условиях равна

$$S_i = S_i^\circ - R \ln \bar{p}_i, \quad i = \text{н, к.} \quad (2.14.7)$$

Учтем тот факт, что стандартное и равновесное состояния для чистых конденсированных фаз одинаковы, а термодинамические свойства (в частности, химический потенциал $\mu_i^{\text{ок}}(T)$, даваемый формулой (2.2.10)) практически не зависят от общего давления P . Тогда вычисление энтропии химической реакции,

$$\Delta S(T) = \sum_{\text{к}} \nu_{\text{к}} S_{\text{к}}(T) - \sum_{\text{н}} \nu_{\text{н}} S_{\text{н}}(T), \quad (2.14.8)$$

с учетом формул (2.13.4), (2.14.6) и (2.14.7), дает соотношение

$$\Delta S(T) = \Delta S_T^\circ - R \ln K_p(T) \neq \Delta S_T^\circ, \quad (2.14.9)$$

где ΔS_T° вычисляется по формуле (2.14.6).

Отсюда изобарная теплота химической реакции, с учетом формул (2.13.3), (2.14.4) и (2.14.9), принимает вид

$$\Delta H(T) \equiv T \Delta S(T) = T \Delta S_T^\circ + \Delta G_T^\circ \equiv \Delta H_T^\circ. \quad (2.14.10)$$

На этом основании обычно не делают различия между величинами $\Delta H(T)$ и ΔH_T° . Это позволяет вычислять изобарный тепловой эффект реакции по формуле (2.14.5), используя стандартные теплоты образования компонентов из простых веществ, что и было ранее учтено при записи выражения (1.5.15). Сделанное заключение в отношении взаимодействия идеальной газовой фазы с чистыми конденсированными фазами распространяется также на идеальные и предельно разбавленные растворы, так как в них каждый компонент ведет себя независимо от других, как если бы он был взят в чистом виде (в частности, коэффициент активности (2.4.31) не зависит от состава раствора).

Для химических реакций, протекающих в условиях $V = \text{const}$, подстановка выражения $\Delta F_T^\circ = -RT \ln K_c(T)$ в уравнение (2.14.1) дает **уравнение изохоры Вант-Гоффа**:

$$\frac{d \ln K_c}{dT} = \frac{\Delta U_T^\circ}{RT^2}. \quad (2.14.11)$$

Величина ΔU_T° , называемая *стандартной изохорной теплотой* химической реакции, связана с ΔF_T° соотношением типа (1.8.14):

$$\Delta F_T^\circ = \Delta U_T^\circ - T \Delta S_T^\circ. \quad (2.14.12)$$

По аналогии с формулой (2.14.10), обычно не делают различия между стандартной и реальной теплотами химической реакции, т. е. полагают $\Delta U(T) = \Delta U_T^\circ$.

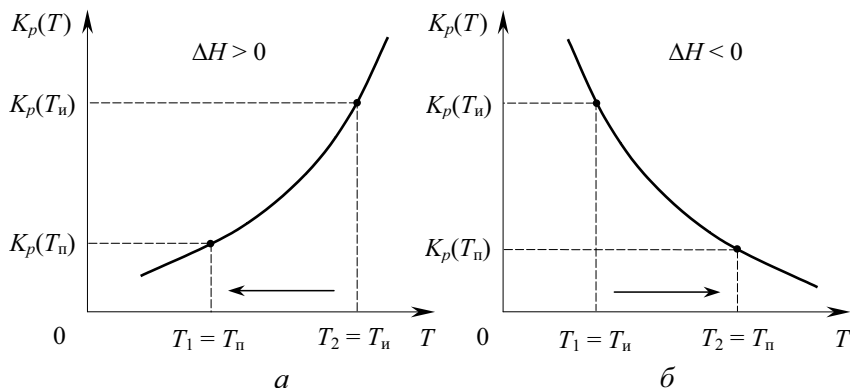


Рис. 2.23. Качественный ход температурной зависимости константы равновесия $K_p(T)$ для эндотермических реакций (а) и для экзотермических реакций (б). Стрелками указано направление транспорта твердого вещества с помощью газотранспортной химической реакции из зоны источника с температурой T_n в зону подложки с температурой T_n (см. п. 2.16)

Таким образом, температурный ход констант химического равновесия определяется, как следует из уравнений (2.14.3) и (2.14.11), тепловым эффектом реакции: с ростом температуры константа $K_p(T)$ или $K_c(T)$ возрастает для *эндотермических реакций* (при $\Delta H > 0$ или $\Delta U > 0$) и убывает для *экзотермических реакций* (при $\Delta H < 0$ или $\Delta U < 0$).

Характер температурной зависимости констант равновесия, получаемой из решения уравнений (2.14.3) и (2.14.11), если в них положить $\Delta H_T^\circ = \Delta H(T) = \text{const}$ и $\Delta U_T^\circ = \Delta U(T) = \text{const}$, является экспоненциальным:

$$K_p(T) = K_{p0} \exp\left(-\frac{\Delta H}{RT}\right), \quad (2.14.13)$$

$$K_c(T) = K_{c0} \exp\left(-\frac{\Delta U}{RT}\right).$$

Температурная зависимость константы равновесия $K_p(T)$ для эндотермических и экзотермических реакций качественно показана на рис. 2.23.

Как увидим в дальнейшем (см. п. 2.16), разный характер температурного хода константы равновесия для эндотермических и экзотермических реакций будет давать различное направление

переноса твердого вещества через газовую фазу из одной температурной зоны в другую при помощи газотранспортных химических реакций.

2.15. Управление химическими реакциями. Принцип Ле Шателье

Управление химическими реакциями в технологических процессах выполняется для решения двух взаимосвязанных задач:

1) создание оптимальных физических условий для протекания химической реакции в направлении достижения желаемого технологического эффекта, например, для максимального выхода полезного продукта реакции;

2) подавление нежелательных химических реакций, конкурирующих с основной реакцией и тем самым препятствующих решению первой задачи.

Для этой цели условия проведения процесса, дающие наилучший технологический эффект, должны быть выбраны такими, при которых изобарный потенциал ΔG конкурирующих реакций обеспечивает их самопроизвольное протекание в направлении, не препятствующем выходу полезного продукта за счет основной химической реакции.

Анализ влияния внешних условий на протекание обобщенной химической реакции (2.12.1) с участием газовой и чистых конденсированных фаз будем проводить с использованием уравнения изотермы Вант-Гоффа в форме (2.13.7).

В качестве исходного состояния берем равновесие ($\Delta G = 0$), при котором $p_i = \bar{p}_i$ и $\Pi(\bar{p}_i) = K_p(T)$, согласно (2.13.7). Функция $\Pi(p_i)$, определенная в виде (2.13.6), для конкретной реакции меняется соответствующим образом при изменении парциальных давлений p_i в реальных технологических условиях. При этом константа равновесия $K_p(T)$ как функция температуры считается известной из литературных источников или полученной из термодинамического расчета.

Для управления химической реакцией будем изменять внешние условия путем варьирования ряда физических параметров системы, среди которых:

- температура T ,
- состав газовой фазы (т. е. парциальные давления p_i),
- общее давление P в системе,
- разбавление газовой смеси инертным газом.

1. Температура влияет на смещение химического равновесия через температурную зависимость константы $K_p(T)$, входящей

в уравнение (2.13.7) для изотермы химической реакции. Согласно выражению (2.14.13), возрастание температуры увеличивает $K_p(T)$ для эндотермических реакций ($\Delta H > 0$) и уменьшает для экзотермических реакций ($\Delta H < 0$) (см. рис. 2.23). На этом основании из уравнения (2.13.7) следует, что рост температуры смещает эндотермическую реакцию вправо ($\Delta G < 0$), а экзотермическую — влево ($\Delta G > 0$).

Следовательно, обратимая химическая реакция протекает в эндотермическом направлении (с поглощением тепла: $\Delta H > 0$) при повышении температуры (за счет подвода тепла к системе) и в противоположном экзотермическом направлении (с выделением тепла: $\Delta H < 0$) при понижении температуры (за счет отбора тепла от системы).

2. Состав газовой фазы изменяется путем введения (или удаления) дополнительного количества химически активного компонента, что изменяет его парциальное давление p_i и отражается на величине $\Pi(p_i)$ в уравнении (2.13.7). Согласно (2.13.6), числитель функции $\Pi(p_i)$ содержит парциальные давления конечных химических реагентов, стоящих в реакции справа. Поэтому увеличение этих давлений при $T = \text{const}$ обеспечивает неравенство $\Pi(p_i) > K_p(T)$, что смещает реакцию (2.12.1) влево (в силу $\Delta G > 0$). Наоборот, дополнительное введение начальных химических реагентов, стоящих в реакции слева, смещает ее вправо (в силу $\Delta G < 0$).

Следовательно, введение в равновесную систему дополнительного количества какого-либо химического компонента всегда сдвигает реакцию в том направлении, которое обеспечивает уменьшение парциального давления (или концентрации) этого компонента.

3. Общее давление в системе влияет на величину $\Pi(p_i)$, входящую в уравнение (2.13.7), через соотношение $p_i = x_i P$, что приводит функцию (2.13.6) к форме с явной зависимостью от P :

$$\Pi(p_i) = \Pi(x_i) P^{\Delta \nu^\Gamma}, \quad \text{где} \quad \Pi(x_i) \equiv \frac{\prod_{\text{к}} x_{\text{к}}^{\nu_{\text{к}}}}{\prod_{\text{н}} x_{\text{н}}^{\nu_{\text{н}}}}. \quad (2.15.1)$$

При этом изменение числа газовых молей $\Delta \nu^\Gamma$, происходящее в ходе реакции слева направо, описывается формулой (2.12.8). Для химических реакций, протекающих в изобарных условиях, изменение числа молей на величину $\Delta \nu^\Gamma$ изменяет объем газовой фазы на $\Delta V = \Delta \nu^\Gamma RT/P$.

Как видно из формулы (2.15.1), при постоянном составе системы ($x_i = \text{const}$) давление P по-разному влияет на величину $\Pi(p_i)$ (а через нее и на знак ΔG) в зависимости от величины и знака $\Delta \nu^r$:

а) при $\Delta \nu^r = 0$ изменение P не влияет на величину $\Pi(p_i)$ и не нарушает равновесия реакции;

б) при $\Delta \nu^r > 0$ (или $\Delta \nu^r < 0$) возрастание P увеличивает (или уменьшает) величину $\Pi(p_i)$, при этом соответственно $\Delta G > 0$ (или $\Delta G < 0$), т. е. реакция смещается влево (или вправо).

в) при $\Delta \nu^r > 0$ (или $\Delta \nu^r < 0$) понижение P уменьшает (или увеличивает) величину $\Pi(p_i)$, при этом соответственно $\Delta G < 0$ (или $\Delta G > 0$), т. е. реакция смещается вправо (или влево).

Следовательно, сжатие (или расширение) газовой системы, происходящее при повышении (или понижении) в ней давления, смещает химическое равновесие (при $\Delta \nu^r \neq 0$) всегда в том направлении, которое обеспечивает уменьшение (или увеличение) числа молей в газовой фазе.

Иными словами, повышение (или понижение) давления благоприятствует образованию веществ, занимающих при данных условиях меньший (или больший) объем, т. е. усиливает то направление реакции, которое уменьшает (или увеличивает) число газовых молей в системе.

4. Разбавление газовой смеси путем введения в систему инертного газа (при этом $P = \text{const}$) всегда уменьшает мольную долю x_i (или парциальное давление $p_i = x_i P$) каждого i -го химического реагента в газовой фазе.

Действительно, при добавлении Δn молей инертного газа к n молям химически активных газов мольная доля x_i снижается от исходного значения n_i/n до $n_i/(n + \Delta n) = n_i/n(1 + \vartheta)$, где $\vartheta = \Delta n/n$ — степень разбавления газовой смеси. Иными словами, введение инертного газа при $P = \text{const}$ уменьшает мольную долю x_i в $(1 + \vartheta)$ раз. С точки зрения влияния на парциальное давление ($p_i = x_i P$) это эквивалентно уменьшению общего давления P в $(1 + \vartheta)$ раз при $x_i = \text{const}$.

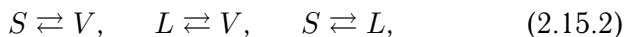
Это позволяет обратиться к формуле (2.15.1), откуда видно, что химические реакции с $\Delta \nu^r = 0$ нечувствительны к введению инертного газа. Для реакций, сопровождающихся увеличением объема газа ($\Delta V > 0$ при $\Delta \nu^r > 0$), введение инертного газа уменьшает $\Pi(p_i)$, что дает $\Delta G < 0$, т. е. реакция сдвигается вправо. Наоборот, реакции с $\Delta \nu^r < 0$ смещаются влево при разбавлении газовой смеси инертным газом.

Смещение равновесия гетерогенной химической системы, осуществляемое путем изменения одного или нескольких ее параметров, необходимо осуществлять в соответствии с правилом фаз Гиббса, а именно (см. п. 1.14): число варьируемых параметров не должно превышать числа термодинамических степеней свободы. В противном случае, по крайней мере, одна из взаимодействующих фаз будет исчезать и состояние системы изменится.

Легко убедиться в том, что все рассмотренные случаи смещения химического равновесия подчиняются так называемому **принципу Ле Шателье** (см. обобщенный принцип на с. 165): *внешнее воздействие, выводящее систему из состояния термодинамического равновесия, стимулирует в ней протекание процессов, противодействующих этому воздействию.*

Этот принцип управляет не только химическими, но и фазовыми превращениями веществ. Более того, любое фазовое превращение может быть описано методами химической термодинамики с использованием закона действующих масс (для константы равновесия K_p) и уравнения изотермы Вант-Гоффа (для изобарного потенциала ΔG).

Например, фазовые переходы первого рода, связанные с обратимым изменением агрегатного состояния вещества, такие как сублимация $\rightleftharpoons$ кристаллизация, испарение $\rightleftharpoons$ конденсация, плавление $\rightleftharpoons$ кристаллизация, можно записать в форме следующих квазихимических реакций:



где S , L и V обозначают твердое (*Solid*), жидкое (*Liquid*) и парообразное (*Vapor*) состояния вещества. Применением закона действующих масс к этим квазихимическим реакциям можно получить все соотношения, ранее полученные на основании условий фазового равновесия.

Как уже отмечалось, управление химическими превращениями производится с целью реализации таких технологических условий, при которых обеспечивается наибольший выход полезного продукта. Наряду с основной химической реакцией, обеспечивающей желаемый эффект, могут протекать конкурирующие реакции, которые должны быть по возможности исключены.

Поэтому основной задачей термодинамического расчета является нахождение оптимальных условий осуществления того или иного технологического процесса, при которых протекание конкурирующих реакций подавлено, а основная реакция эффективно осуществляется с наибольшей степенью превращения исходных веществ в полезный продукт.

Термодинамическая возможность или невозможность протекания какой-либо реакции в нужном направлении оценивается по знаку изобарного потенциала ($\Delta G < 0$ или $\Delta G > 0$), который вычисляется из уравнения изотермы Вант-Гоффа, записанного в одной из форм (2.13.5), (2.13.7) или (2.13.9). Для расчета ΔG по этим формулам необходимо знать, во-первых, реальные парциальные давления p_i для всех компонентов газовой фазы и, во-вторых, температурный ход константы химического равновесия $K_p(T)$ или величины $\Delta G_T^\circ = -RT \ln K_p(T)$, рассчитываемой в стандартных условиях на основании табличных данных (см. приложение А).

Парциальные давления реальных газов входят в изобарный потенциал ΔG в форме величины $\Pi(p_i)$, прибавляемой к ΔG_T° (см. формулу (2.13.7)). Логарифмические вклады парциальных давлений ($\ln p_i$) являются существенными лишь в условиях, близких к равновесным, когда $RT|\ln \Pi(p_i)| \approx |\Delta G_T^\circ|$. При достаточно больших значениях $|\Delta G_T^\circ|$ (> 10 ккал/моль) эти логарифмические вклады в суммарную величину $\Delta G = \Delta G_T^\circ + RT \ln \Pi(p_i)$ не могут изменить ее результирующий знак по сравнению со знаком ΔG_T° .

В таких условиях стандартный изобарный потенциал ΔG_T° свидетельствует о принципиальной возможности протекания реакции либо слева направо (при $\Delta G_T^\circ < -10$ ккал/моль), либо справа налево (при $\Delta G_T^\circ > 10$ ккал/моль) не только в стандартных условиях (при $p_i = 1$ атм), но и для любых реальных ситуаций. При $|\Delta G_T^\circ| \leq 10$ ккал/моль для выяснения принципиальной возможности или невозможности осуществления желаемого технологического процесса следует использовать полное выражение $\Delta G = \Delta G_T^\circ + RT \ln \Pi(p_i)$, учитывающее реальные давления в системе. Методика подобного термодинамического расчета изложена в приложениях А.1 и А.2.

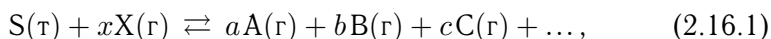
Следует иметь в виду, что практическая осуществимость того или иного процесса (в отличие от принципиальной термодинамической возможности) и скорость его протекания определяются не только термодинамическими, но и кинетическими факторами, которые будут рассмотрены в гл. 4.

2.16. Газотранспортные химические реакции

Химическое взаимодействие между твердым веществом и реактивной газовой фазой с образованием газообразных продуктов реакции играет важную роль в различных технологических процессах электроники.

Газотранспортной химической реакцией (ГХР) называется обратимая гетерогенная реакция, протекающая в двух температурных зонах (травления и осаждения), с помощью которой в *зоне травления* труднолетучее твердое вещество переводится из твердой фазы в газовую в виде продуктов реакции, после транспорта которых в *зону осаждения* происходит их обратное превращение в твердый продукт. Механизм транспорта и кинетика химического переноса веществ будут изучены ниже в п. 4.14 и 4.15. Далее рассмотрим основные термодинамические закономерности реализации ГХР.

Запишем обратимую газотранспортную химическую реакцию в обобщенной форме,



приведенной к одному молю твердого вещества S ($x, a, b, c, \dots$ — соответствующие стехиометрические коэффициенты для газов $X, A, B, C, \dots$).

Компонент X , осуществляющий газовое травление вещества S и перевод его в газовую фазу в виде одной или нескольких молекулярных форм $A, B, C, \dots$ (которые содержат химические элементы, входящие в состав S), называется *транспортным агентом*. Именно эти газообразные компоненты осуществляют перенос через газовую фазу исходного твердого вещества S в «разобранном виде».

В проточных системах в газовую фазу вводят *газ-носитель*, парциальное давление которого существенно превышает давления других компонентов газовой фазы (см. ниже рис. 4.23). Основная функция газа-носителя состоит в создании общего газового потока, осуществляющего доставку транспортного агента в зону травления (где размещается исходное вещество S , называемое *источником*) и дальнейший перенос химически активных газовых молекул из зоны травления в зону осаждения (где размещается *подложка*, на которой осаждается продукт S).

В качестве газа-носителя может выступать как инертный газ (например, аргон), так и химически активный участник реакции из числа $A, B, C, \dots$ (например, H_2, N_2 и др.). Присутствие газа-носителя необходимо в технологических процессах, проводимых в проточных системах, в которых транспорт осуществляется за счет конвективного массопереноса (см. п. 4.14). Массоперенос может осуществляться также в результате молекулярной диффузии в газовой фазе при использовании так называемой сэндвич-системы (см. п. 4.15).

Покажем, что химический перенос веществ с помощью ГХР возможен не только из горячей зоны в холодную (как для процесса испарения–конденсации), но и в обратном направлении. При этом направление транспорта определяется знаком теплоты химической реакции. В соответствии с (2.14.13), на рис. 2.23 качественно показан температурный ход константы равновесия $K_p(T)$ для эндотермической ($\Delta H > 0$) и экзотермической ($\Delta H < 0$) реакции. Температуры зоны источника (зоны травления) и зоны подложки (зоны осаждения) обозначены, соответственно, как $T_{\text{и}}$ и $T_{\text{п}}$, в то время как T_1 и T_2 всегда таковы, что $T_1 < T_2$. Распределение температуры в проточной и сэндвич-системах показано ниже на рис. 4.23 и 4.24.

Для реакции (2.16.1) закон действующих масс (2.13.4) имеет следующий вид (вещество S — чистое):

$$K_p(T) = \frac{\bar{p}_A^a \bar{p}_B^b \bar{p}_C^c \cdots}{\bar{p}_X^x}. \quad (2.16.2)$$

Отсюда видно, что бóльшие значения $K_p(T)$ соответствуют бóльшему выходу газообразных продуктов реакции A, B, C, ..., в то время как уменьшение $K_p(T)$ означает преимущественный сдвиг реакции (2.16.1) влево с образованием твердого продукта S. На этом основании температурная зона с бóльшим значением $K_p(T_{\text{и}})$ должна быть зоной источника (травления), а с меньшим значением $K_p(T_{\text{п}})$ — зоной подложки (осаждения).

Газовая смесь, образующаяся в зоне источника при температуре $T_{\text{и}}$ и транспортируемая тем или иным способом (за счет газового потока или диффузии) в зону подложки, является для нее неравновесной. Парциальные давления $\bar{p}_i(T_{\text{и}})$, равновесные при температуре $T_{\text{и}}$, играют роль реальных неравновесных давлений p_i для зоны подложки с температурой $T_{\text{п}}$. Тогда для зоны подложки функция (2.13.6) принимает вид

$$\Pi(p_i) \Big|_{T_{\text{п}}} \equiv \frac{p_A^a p_B^b p_C^c \cdots}{p_X^x} \Big|_{T_{\text{п}}} = \frac{\bar{p}_A^a \bar{p}_B^b \bar{p}_C^c \cdots}{\bar{p}_X^x} \Big|_{T_{\text{и}}} \equiv K_p(T_{\text{и}}).$$

Следовательно, согласно уравнению изотермы Вант-Гоффа в форме (2.13.7), в зоне подложки возникает пересыщение,

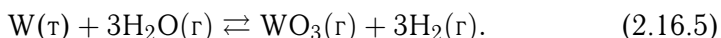
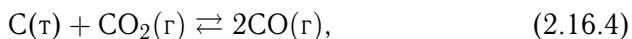
$$\begin{aligned} \Delta G(T_{\text{п}}) &= RT_{\text{п}} \left[\ln \Pi(p_i) \Big|_{T_{\text{п}}} - \ln K_p(T_{\text{п}}) \right] = \\ &= RT_{\text{п}} \left[\ln K_p(T_{\text{и}}) - \ln K_p(T_{\text{п}}) \right] > 0, \end{aligned} \quad (2.16.3)$$

знак которого соответствует сдвигу реакции (2.16.1) в сторону осаждения твердого продукта S. Выделяющийся в этой зоне

транспортный агент X диффундирует в обратном направлении (если этому не препятствует газовый поток). Это результат совместно с отводом газов А, В, С, ... из зоны источника создает в ней пересыщение, противоположное по знаку (2.16.3), что обеспечивает травление твердого источника S.

Как следует из неравенства в формуле (2.16.3), должно быть $K_p(T_{\text{и}}) > K_p(T_{\text{п}})$, то есть химический перенос твердого вещества в газотранспортной системе всегда осуществляется из температурной зоны с большим значением константы равновесия $K_p(T_{\text{и}})$ в температурную зону с ее меньшим значением $K_p(T_{\text{п}})$, что показано стрелками на рис. 2.23 а и б.

Для эндотермических реакций (с поглощением тепла) транспорт вещества всегда происходит из высокотемпературной зоны в низкотемпературную ($T_2 \rightarrow T_1$ на рис. 2.23 а). Примером таких реакций являются следующие:



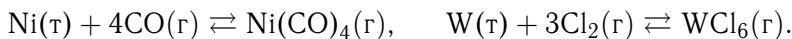
Реакция (2.16.4) играет важную роль в процессе формирования термоэмиссионных свойств оксидного катода в электровакуумных приборах. Оксидная масса наносится на kern катода в виде суспензии карбонатов бария, стронция и кальция (BaCO_3 , SrCO_3 и CaCO_3), приготовленной на органическом связующем веществе (биндере). Разложение биндера в процессе термообработки создает большое количество углерода, в результате чего оксидный слой чернеет и увеличивается его теплоизлучающая способность. В таком случае возникала бы необходимость повышения мощности подогревателя для обеспечения рабочей температуры катода.

Реакция (2.16.4) способствует удалению углерода и осветлению оксидного покрытия, при этом транспортный агент CO_2 образуется в результате разложения карбонатов по реакции (2.12.17). Следовательно, в данном случае газотранспортная химическая реакция (2.16.4) играет положительную роль.

В противоположность этому, химическая реакция (2.16.5) является вредной, так как обеспечивает перенос парами воды тугоплавкого вольфрама в виде летучего окисла WO_3 с горячей нити накаливания (в осветительных лампах) или с подогревателя катода (в электровакуумных приборах) в более холодные области прибора (экраны, стенки колбы и т. п.). В результате этого вольфрамовые нити и подогреватели утоньшаются и со временем

перегорают. Следовательно, наличие в электронных приборах остаточных паров воды снижает их срок службы.

Для экзотермических реакций (с выделением тепла) транспорт вещества всегда осуществляется из низкотемпературной зоны в высокотемпературную ($T_1 \rightarrow T_2$ на рис. 3.1 ,б). Такими реакциями, например, являются следующие:



Эти реакции лежат в основе карбонильного и хлоридного методов извлечения тугоплавких металлов из руд.

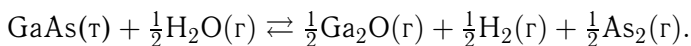
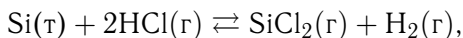
Химический транспорт веществ из холодной зоны в горячую с помощью экзотермических газотранспортных реакций особенно опасен для катодов электровакуумных приборов. Перенос на горячий катод ряда веществ может снижать его термоэмиссионную способность в результате следующих процессов:

а) химического отравления — при переносе химически активных веществ (кислорода, хлора, серы и т. п.), взаимодействующих с атомами бария в оксидном покрытии;

б) спекания оксидного слоя — при переносе инертных легкоплавких веществ;

в) почернения поверхности оксидного слоя и снижения его рабочей температуры — при переносе инертных тугоплавких веществ (типа углерода и карбидов).

Газотранспортные химические реакции широко применяются в процессах эпитаксиального роста монокристаллических полупроводниковых слоев. В частности, химический транспорт кремния и арсенида галлия с помощью, соответственно, хлористого водорода и паров воды осуществляется с применением реакций



В результате этих реакций образуются газообразные продукты $\text{SiCl}_2(\text{г})$ и $\text{Ga}_2\text{O(г)}$, транспортирующие, соответственно, кремний и галлий из высокотемпературной зоны в низкотемпературную, поскольку обе реакции являются эндотермическими. Методика термодинамического расчета химического транспорта арсенида галлия парами воды изложена в приложении А.2.

Другие примеры использования газотранспортных реакций в технологических процессах эпитаксиального роста пленок газофазными методами будут рассмотрены ниже в п. 4.13.

2.17. Ионное равновесие и термодинамические свойства растворов электролитов

До сих пор при рассмотрении процессов с участием жидких растворов мы пренебрегали существованием в них ионов, возникающих в результате электролитической диссоциации.

Электролитическая диссоциация, проявляющаяся в виде распада молекул растворяемого вещества на ионы, возможна лишь в растворителях с сильно полярными молекулами, имеющих высокую диэлектрическую проницаемость (каковой является, например, вода с $\epsilon = 81$). Вещества, которые в растворе состоят полностью или частично из ионов, называются соответственно *сильными* и *слабыми электролитами*.

Растворы электролитов проводят электрический ток в результате движения ионов, что сопровождается процессами массопереноса и химическими превращениями веществ на электродах, объединенных общим названием *электролиз* (см. п. 2.18 и 2.19). Ионы с положительным зарядом, движущиеся при электролизе к катоду, принято называть *катионами*, а отрицательные ионы, идущие к аноду, — *анионами*.

Процесс электролитической диссоциации протекает в форме четырех последовательных этапов (см. п. 2.3):

а) ориентационное окружение молекул растворяемого вещества полярными молекулами растворителя (воды), притягивающимися друг к другу за счет взаимодействия диполь-дипольного типа (для полярных молекул типа HCl) или ионно-дипольного типа (для ионных кристаллов типа NaCl);

б) ионизация полярных молекул (типа HCl) силовым полем окружающих их диполей растворителя, возникающая как следствие сильной поляризации ковалентной связи внутри молекулы растворяемого вещества с превращением ее в ионную связь;

в) диссоциация молекул вещества с ионной связью (естественной для кристаллов типа NaCl или наведенной для молекулы HCl) на ионы, возникающая как следствие ослабления связи между ионами в молекуле растворяемого вещества за счет ионно-дипольных разрывающих усилий со стороны полярных молекул растворителя;

г) сольватация (гидратация) путем окружения возникших ионов растворяемого вещества полярными молекулами растворителя (воды) с образованием сольватных (гидратных) оболочек, которые экранируют заряды ионов и препятствуют обратному соединению ионов в молекулы.

Будем рассматривать лишь водные растворы электролитов. Как показано в п. 1.16, гомогенные химические реакции, связанные с ионными взаимодействиями в однофазной системе, управляются теми же закономерностями, что и реакции между нейтральными частицами, с использованием обычных химических (а не электрохимических) потенциалов в форме закона действующих масс и уравнения изотермы Вант-Гоффа, полученных в п. 2.12 и 2.13. Водные растворы электролитов будем считать сильно разбавленными, что позволяет для их описания использовать концентрации взаимодействующих ионов без введения активностей. Чистая вода диссоциирует по реакции



Водородный ион H^+ (протон) неустойчив в воде и соединяется с молекулой воды H_2O , образуя устойчивый ион гидроксония H_3O^+ (см. п. 3.1). Однако в дальнейшем будем для простоты говорить о водородных ионах H^+ (понимая под ними ионы гидроксония H_3O^+), так как это не отразится на результатах в силу тождественного равенства их концентраций ($c_{\text{H}^+} \equiv c_{\text{H}_3\text{O}^+}$).

Равновесие между молекулами воды H_2O и образовавшимися ионами H^+ и OH^- управляется законом действующих масс (2.12.6), записанным для реакции (2.17.1) в форме

$$K_c(T) = \frac{c_{\text{H}^+} c_{\text{OH}^-}}{c_{\text{H}_2\text{O}}} \quad (2.17.2)$$

Поскольку степень диссоциации воды очень мала, то равновесная концентрация молекул практически равна общей мольной концентрации воды, $c_{\text{H}_2\text{O}} = 55,56 \text{ моль/л} = \text{const}$, объединяя которую с K_c , можно переписать равенство (2.17.2) в виде

$$K_w(T) = c_{\text{H}^+} c_{\text{OH}^-}, \quad (2.17.3)$$

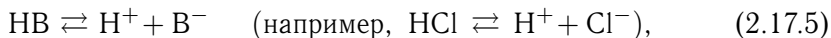
где константа $K_w(T)$ называется *ионным произведением воды*. Для чистой воды из равенства (2.17.3) и условия электронейтральности получаем

$$c_{\text{H}^+} = c_{\text{OH}^-} = \sqrt{K_w(T)}. \quad (2.17.4)$$

При $T = 298 \text{ К}$ имеем $K_w = 10^{-14}$ и $c_{\text{H}^+} = c_{\text{OH}^-} = 10^{-7} \text{ г-ион/л}$, что дает очень малую степень диссоциации воды, $\xi \equiv c_{\text{H}^+}/c_{\text{H}_2\text{O}} = 1,8 \cdot 10^{-9}$. Электрохимическое объяснение малой величины ионного произведения воды будет дано в п. 2.20.

Концентрацию ионов водорода H^+ (точнее, ионов гидроксония H_3O^+) и гидроксила OH^- можно изменять добавлением кислоты

НВ (например, HCl) или основания АОН (например, щелочи NaOH), диссоциирующих в воде по реакциям



Водородные и гидроксильные ионы, возникающие по трем различным механизмам, описываемым реакциями (2.17.1), (2.17.5) и (2.17.6), неразличимы в растворах электролитов. Поэтому в водном растворе, содержащем произвольное количество кислот, щелочей и солей, концентрации c_{H^+} и $c_{\text{ОН}^-}$ не могут изменяться произвольно. Их произведение, согласно (2.17.3), сохраняется неизменным при $T = \text{const}$. Таким образом, закон действующих масс (2.17.3) управляет *ионным равновесием* в воде: любое повышение концентрации водородных ионов, характерное для кислых сред, где $c_{\text{H}^+} > c_{\text{ОН}^-}$, вызывает соответствующее уменьшение концентрации гидроксильных ионов, и наоборот для щелочных сред, где $c_{\text{ОН}^-} > c_{\text{H}^+}$.

Характер раствора — кислый (кислотный) или основной (щелочной) — принято оценивать с помощью так называемого **водородного показателя**:

$$\text{pH} = -\lg c_{\text{H}^+} \equiv -\lg c_{\text{H}_3\text{O}^+}, \quad (2.17.7)$$

где c_{H^+} — безразмерная величина, равная числу грамм-ионов в литре раствора. Согласно сказанному, при $T = 298 \text{ K}$ в нейтральной среде $\text{pH} = 7$, в кислых средах $\text{pH} < 7$ (тем меньше, чем выше кислотность), в щелочных средах $\text{pH} > 7$ (тем больше, чем выше щелочность).

Знание величины pH и поддержание ее постоянной важно при использовании растворов электролитов в технологических процессах, например, для химического травления полупроводников, так как от значения pH зависят термодинамические и кинетические закономерности процесса травления. На практике величину pH оценивают с помощью окрашенных индикаторов или измеряют специальными pH -метрами.

Для создания растворов с практически неизменным значением pH , называемых *буферными растворами*, используется кислота НВ, слабо диссоциирующая по реакции (2.17.5), и ее соль АВ, почти полностью диссоциирующая по реакции $\text{AB} \rightleftharpoons \text{A}^+ + \text{В}^-$. Исходные концентрации НВ и АВ в водном растворе обозначим в виде $c_{\text{кисл}}$ и $c_{\text{соль}}$.

На основании закона действующих масс для реакции (2.17.5) можно записать

$$c_{\text{H}^+} = K_{\text{НВ}} \frac{c_{\text{НВ}}}{c_{\text{В}^-}}, \quad (2.17.8)$$

где $K_{\text{НВ}}$ — константа химического равновесия реакции (2.17.5), называемая *константой диссоциации* кислоты НВ.

Так как степень диссоциации для соли $\xi_{\text{АВ}} \approx 1$, а для кислоты $\xi_{\text{НВ}} \ll 1$, то число ионов В^- , образующихся из молекул АВ, много больше числа аналогичных ионов, образующихся из кислоты НВ. Это позволяет для равновесных концентраций НВ и В^- в растворе записать $c_{\text{НВ}} \approx c_{\text{кисл}}$ и $c_{\text{В}^-} \approx c_{\text{соль}}$. В этом случае равенство (2.17.8) принимает следующий вид:

$$c_{\text{H}^+} = K_{\text{НВ}} \frac{c_{\text{кисл}}}{c_{\text{соль}}} \quad \text{или} \quad \text{pH} = -\lg K_{\text{НВ}} - \lg \frac{c_{\text{кисл}}}{c_{\text{соль}}}. \quad (2.17.9)$$

Основное свойство буферных растворов — малая чувствительность водородного показателя pH к разбавлению. Действительно, добавление воды, уменьшая $c_{\text{кисл}}$ и $c_{\text{соль}}$, оставляет их отношение практически без изменения.

Кроме того, буферные растворы оказываются малочувствительными к добавлению умеренного количества другой сильно диссоциирующей кислоты или щелочи. Действительно, если концентрация добавленной кислоты невелика ($\Delta c \ll c_{\text{кисл}}$ и $c_{\text{соль}}$), то при полной ее диссоциации в одном литре дополнительно возникает Δc ионов H^+ . В соответствии с рассмотренным в п. 2.15, это сдвигает реакцию (2.17.5) влево с образованием слабой кислоты НВ. При этом концентрация молекул НВ увеличивается, а ионов В^- уменьшается на одну и ту же величину Δc , т. е. $c_{\text{НВ}} = c_{\text{кисл}} + \Delta c$ и $c_{\text{В}^-} = c_{\text{соль}} - \Delta c$. Тогда из равенства (2.17.8) получаем

$$c_{\text{H}^+} = K_{\text{НВ}} \frac{c_{\text{кисл}}(1 + \Delta c/c_{\text{кисл}})}{c_{\text{соль}}(1 - \Delta c/c_{\text{соль}})} \approx K_{\text{НВ}} \frac{c_{\text{кисл}}}{c_{\text{соль}}}, \quad (2.17.10)$$

т. е. pH раствора практически не изменяется при $\Delta c \ll c_{\text{кисл}}$ и $c_{\text{соль}}$. При добавлении небольшого количества сильной щелочи, полностью диссоциирующей по реакции (2.17.6), концентрация ионов гидроксила OH^- увеличивается, а водородных ионов H^+ уменьшается. Если, как и ранее, это уменьшение Δc мало, то величины $c_{\text{НВ}} = c_{\text{кисл}} - \Delta c$ и $c_{\text{В}^-} = c_{\text{соль}} + \Delta c$ обеспечивают приближенное равенство в формуле (2.17.10), т. е. постоянство водородного показателя pH.

Буферные растворы важны для поддержания стабильными условий проведения химических реакций не только в технологи-

ческих процессах, но и в биологических системах, в том числе для существования живых организмов (в человеческой крови $\text{pH} \approx 7,4$).

Рассмотрим обратимый процесс растворения в воде твердого сильного электролита АВ по уравнению



выражающему равновесие ионов A^+ и B^- в растворе с твердым осадком АВ(т). В случае малорастворимой соли, дающей сильно разбавленный раствор, это равновесие управляется законом действующих масс, записанным через концентрации в виде

$$K_{\text{AB}}(T) = \frac{c_{\text{A}^+} c_{\text{B}^-}}{c_{\text{AB}}}. \quad (2.17.12)$$

В силу постоянства c_{AB} для твердого осадка из (2.17.12) получаем закон действующих масс

$$L(T) = c_{\text{A}^+} c_{\text{B}^-}, \quad (2.17.13)$$

где величина $L(T) \equiv K_{\text{AB}}(T) c_{\text{AB}}$ названа *произведением растворимости*.

Если в растворе отсутствуют другие электролиты, образующие ионы A^+ и B^- , то $c_{\text{A}^+} = c_{\text{B}^-} = L^{1/2}$ из-за стехиометрии соединения АВ. При добавлении к насыщенному раствору АВ другого электролита, содержащего общий с ним ион (например, ион A^+), увеличивается концентрация c_{A^+} . Равенство (2.17.13) требует уменьшения концентрации ионов B^- в растворе, что может быть обеспечено лишь при выпадении в осадок части ранее растворенного вещества АВ. Следовательно, растворимость твердой соли понижается при введении в раствор другого электролита, содержащего общие с ней ионы.

Рассмотрим ту же обратимую реакцию (2.17.11), но для процесса диссоциации жидкого электролита АВ(ж) в водном растворе. Если исходная концентрация электролита c_0 , а степень его диссоциации ξ , то равновесные концентрации в растворе равны:

$$c_{\text{A}^+} = c_{\text{B}^-} = \xi c_0 \quad \text{и} \quad c_{\text{AB}} = (1 - \xi) c_0.$$

Подставив эти значения в закон действия масс (2.17.12), получим

$$K_{\text{AB}}(T) = \frac{\xi^2}{1 - \xi} c_0.$$

Отсюда при слабой диссоциации ($\xi \ll 1$) получаем следующее соотношение:

$$\xi(T) \approx \sqrt{K_{\text{AB}}(T)/c_0}, \quad (2.17.14)$$

известное как **закон разбавления Оствальда**¹⁾. Согласно этому закону, разбавление раствора увеличивает степень диссоциации слабого жидкого электролита обратно пропорционально квадратному корню из его молярной концентрации.

Следует иметь в виду, что все приведенные выше соотношения справедливы для сильно разбавленных растворов. В противном случае концентрации компонентов c_i должны быть заменены их активностями a_i .

Ограничимся этим кратким рассмотрением ионных процессов, что составляет лишь малую часть общей теории жидких электролитов, детально изложенной в литературе. Более подробное рассмотрение ионных процессов будет продолжено в следующей главе в применении к заряженным примесям в полупроводниках как твердых электролитах.

2.18. Термодинамика электрохимических процессов

Электрохимической системой называют совокупность проводников первого рода (с электронным механизмом электропроводности) и второго рода (с ионным механизмом электропроводности). Химические взаимодействия в таких системах происходят на поверхности металлических электродов по *гетерогенному механизму* путем обмена электронами между ионами в растворе и металлом, выполняющим роль резервуара с неограниченным запасом электронов. Аналогичные гомогенные реакции, идущие в жидких электролитах с передачей электрона в актах непосредственного столкновения реагирующих частиц, называют *окислительно-восстановительными реакциями*. При этом окислению соответствует потеря электрона частицей, а восстановлению — его приобретение (см. п. 2.19).

В электрохимических реакциях процессы окисления и восстановления протекают на разных электродах, т. е. пространственно разделены в отличие от химических реакций. Именно это позволяет превратить большую часть химической энергии в электрическую. Электронные явления на поверхности электродов сопровождаются процессами массопереноса в объеме раствора с возможным растворением или осаждением материала электродов.

¹⁾ Оствальд (Ostwald) Вильгельм Фридрих (1853–1932) — немецкий химик и философ, известен работами по теории электролитической диссоциации, установил закон разбавления электролитов и разработал фундаментальные вопросы химической кинетики и катализа. Нобелевская премия по химии (1909) «за изучение природы катализа и основополагающие исследования скоростей химических реакций».

Протекание обратимой окислительно-восстановительной реакции на металлических электродах порождает возникновение *электродвижущей силы* (ЭДС), вызывающей ток во внешней электрической цепи (*гальванические процессы*). И наоборот, протекание через раствор электролита тока от внешнего источника сопровождается химическими превращениями веществ на электродах (*электролитические процессы*). Соответственно этому, электрохимические системы применяют на практике в следующих двух вариантах:

а) в виде *гальванического элемента* для создания химических источников тока,

б) в виде *электролитической ячейки* для осаждения или удаления вещества в процессе электролиза.

Положительный и отрицательный электроды в электролитической ячейке называются *анодом* и *катодом*, а в гальваническом элементе — *положительным* и *отрицательным полюсом*.

В электролитической ячейке положительный потенциал анода, созданный внешним источником, «оттягивает» электроны от поверхности внутрь металла, «оголяя» его ионную решетку. Тем самым обеспечивается переход ионов металла в раствор электролита. Следовательно, на аноде происходит процесс окисления (путем отбора электрона от аниона в растворе) одновременно с возможным растворением материала электрода — *анодное окисление и растворение*. На катоде такие процессы идут в обратном направлении, создавая избыток электронов на поверхности металла, и тем самым обеспечивают восстановление (путем передачи электрона катиону в растворе) с одновременным осаждением вещества на электроде — *катодное восстановление и осаждение*.

В гальваническом элементе отрицательный полюс, посылая электроны во внешнюю цепь, ведет себя как анод в электролитической ячейке, обеспечивая протекание процессов окисления и растворения. Напротив, положительный полюс, подобно катоду электролитической ячейки, обеспечивает восстановление и осаждение материала на своей поверхности.

Характерной особенностью любой электрохимической системы является самопроизвольное возникновение обратимой ЭДС между металлическими электродами как результат протекания обратимой химической реакции. Термин «обратимая электродвижущая сила» означает, что один и тот же электрохимический элемент может быть обратимо использован двояко: как в гальваническом режиме, так и в электролитическом режиме. Для этого

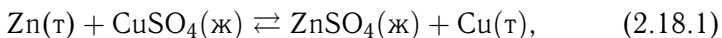
надо выполнить вполне определенное соотношение между двумя противодействующими друг другу электродвижущими силами — между внутренней ЭДС элемента (ΔE) и внешней ЭДС, приложенной к элементу ($E_{\text{вн}}$).

Для работы в гальваническом режиме необходимо внутреннюю ЭДС частично скомпенсировать внешней ЭДС, так чтобы незначительное превышение (при $\Delta E > E_{\text{вн}}$) обеспечивало бы протекание бесконечно малого тока во внешней цепи. Обращение гальванического элемента в электролитический достигается при обратном неравенстве ($E_{\text{вн}} > \Delta E$), обеспечивающем протекание бесконечно малого тока в противоположном направлении. Требование бесконечно малого тока обеспечивает, как увидим ниже, *электрохимическое равновесие* в системе (см. п. 1.16).

Протекание тока конечной величины вызывает необратимость процесса. В этом случае превышение ЭДС $|E_{\text{вн}} - \Delta E|$, называемое *перенапряжением*, складывается из омического падения напряжения в электролите $U_{\text{ом}}$ и так называемой *электродной поляризации* $E_{\text{пол}}$. При наличии перенапряжения внешняя ЭДС равна: а) $E_{\text{вн}} = \Delta E - (U_{\text{ом}} + E_{\text{пол}}) < \Delta E$ — для гальванического элемента, б) $E_{\text{вн}} = \Delta E + (U_{\text{ом}} + E_{\text{пол}}) > \Delta E$ — для электролитической ячейки. Омические потери напряжения $U_{\text{ом}} = IR$ связаны с необратимым выделением тепла Джоуля–Ленца в виде $Q = I^2 R$, за счет протекания тока I через раствор электролита с электрическим сопротивлением R .

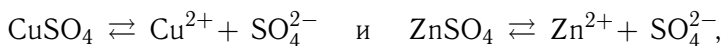
Возникновение электродной поляризации $E_{\text{пол}}$ вызвано неравновесностью электродных процессов при протекании тока конечной величины, изменяющего концентрацию ионов в электролите по сравнению с равновесной. Это, как увидим ниже, влияет на потенциалы электродов, т. е. вызывает так называемую *концентрационную поляризацию*. В применении к электролизу нередко говорят также о *химической поляризации*, понимая под этим возникновение ЭДС ΔE между электродами электролитической ячейки. Действительно, выделение продуктов электролиза на электродах приводит к образованию гальванической пары, порождающей ΔE . Химическая и концентрационная поляризации противодействуют электролизу. Как результат этого, ток $I = [E_{\text{вн}} - (\Delta E + E_{\text{пол}})]/R$ возникает под действием превышения внешней ЭДС $E_{\text{вн}}$ над суммарной (химической и концентрационной) поляризацией ($\Delta E + E_{\text{пол}}$).

В качестве примера рассмотрим обратимую химическую реакцию



работающую в элементе Даниэля–Якоби, который схематически изображен на рис. 2.24. Цинку (Zn), стоящему в реакции слева, соответствует левый электрод, а меди (Cu), стоящей в реакции справа, — правый электрод. Электроды погружены в растворы своих солей ZnSO_4 и CuSO_4 , разделенные пористой перегородкой, показанной двойным пунктиром. Эта перегородка предотвращает взаимное перемешивание электролитов, но способна проводить ионы SO_4^{2-} . Фигурная стрелка с надписью «анионы» указывает направление движения общего для обоих электролитов аниона SO_4^{2-} (движение катионов Zn^{2+} и Cu^{2+} противоположно). Знаки + и –, поставленные на электродах, соответствуют направлению электрического поля от правого электрода к левому, что показано жирной стрелкой на рис. 2.24, при этом знаки внутренней ЭДС, как увидим ниже, различны ($\Delta E \geq 0$).

Исходная реакция (2.18.1) с учетом диссоциации солей,



путем удаления общего иона SO_4^{2-} может быть записана в ионной форме:



Гальванический режим работы, когда $\Delta E \geq E_{\text{вн}}$, показан на рис. 2.24 а. В этом случае направление тока во внешней цепи (обозначенное сплошными стрелками, показывающими движение положительных зарядов) таково, что внутри растворов общие для них отрицательные ионы (анионы SO_4^{2-}) движутся от плюса к минусу, то есть против действия внутренней ЭДС (ΔE). Такое «противодвижение» поддерживается силами химической природы, которые порождены самопроизвольным протеканием реакции (2.18.2) в прямом направлении (слева направо).

Движение анионов SO_4^{2-} из раствора CuSO_4 в раствор ZnSO_4 вызвано тем, что в первом растворе их заряд преобладает над зарядом катионов Cu^{2+} , а во втором, наоборот, доминирует заряд катионов Zn^{2+} . Действительно, на положительном полюсе в результате процесса восстановления катионы Cu^{2+} превращаются в атомы Cu (осаждающиеся на медном электроде). На отрицательном полюсе идет процесс окисления с превращением атомов Zn в ионы Zn^{2+} , переходящие как катионы в раствор электролита (т. е. цинковый электрод растворяется). При таком направлении движения ионов SO_4^{2-} поддерживается электронейтральность растворов и обеспечивается непрерывность электронного и ионного токов.

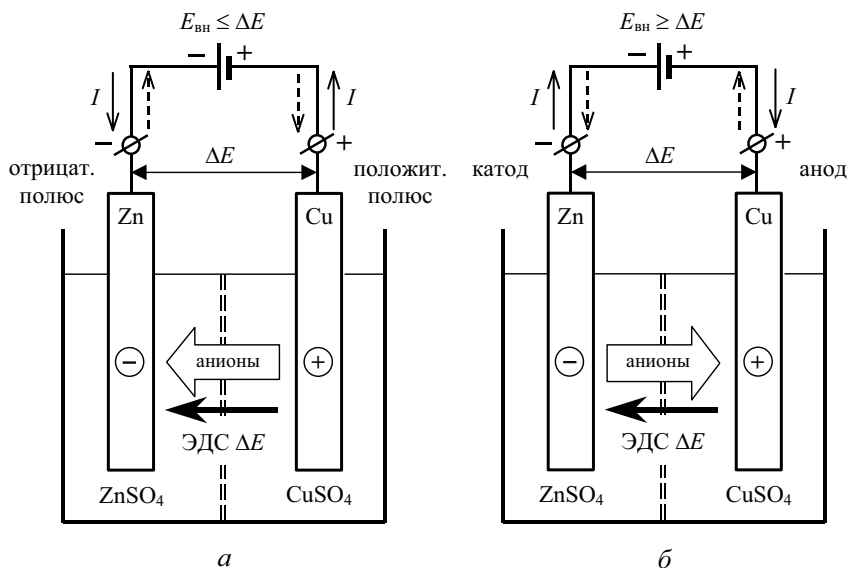


Рис. 2.24. Схематическое изображение элемента Даниэля-Якоби при работе в гальваническом режиме (а) и в электролитическом режиме (б). Сплошными стрелками показано направление тока в электрической цепи, замыкающее движение катионов внутри элемента. Пунктирные стрелки указывают направление движения электронов между электродами во внешней электрической цепи

Электролитический режим работы, когда $\Delta E \leq E_{вн}$, показан на рис. 2.24 б. Направление тока внешней цепи (обозначенное сплошными стрелками) меняется на противоположное по сравнению с рис. 2.24 а. В этом случае анионы SO_4^{2-} движутся под действием внешней ЭДС ($E_{вн}$) от минуса к плюсу. Это обеспечивает подвод электрической энергии от внешнего источника к химической системе и делает возможным протекание реакции (2.18.2) в обратном направлении (справа налево, когда $\Delta G > 0$). На катоде катионы Zn^{2+} восстанавливаются до атомов Zn, а на аноде идет процесс окисления с переходом в раствор атомов Cu в форме катиона Cu^{2+} , то есть при электролизе медный электрод растворяется, а цинк выделяется из раствора.

Пунктирными стрелками на рис. 2.24 показано направление движения электронов во внешней электрической цепи. Из сравнения рис. 2.24 а и б видно, что электронные процессы на отрицательном полюсе гальванического элемента и на аноде (положительном электроде) электролитической ячейки одинаковы: и тот, и другой поставляют электроны во внешнюю цепь, отбирая их от нейтральных атомов металла (путем перевода ионов металла

из кристаллической решетки в раствор электролита в форме катиона). Противоположным образом ведут себя положительный полюс гальванического элемента и катод (отрицательный электрод) электролитической ячейки: они оба принимают электроны от внешней цепи, передавая их катионам раствора с переводом последних в кристаллическую решетку металла. В терминах окислительно-восстановительных реакций первый случай соответствует процессу окисления, а второй — процессу восстановления (см. п. 2.19).

Покажем, что обратимая электродвижущая сила ΔE электрохимического элемента однозначно определяется изобарным потенциалом ΔG химической реакции, лежащей в основе работы этого элемента.

Пусть ЭДС ΔE перемещает по электрической цепи заряд, равный заряду одного грамм-иона $z\mathcal{F}$, где $\mathcal{F} = e\mathcal{N}_A = 96485$ Кл — постоянная Фарадея, а z — число электронов, участвующих в зарядовом обмене на электродах и определяющих заряд ионов в растворе (не следует путать $z > 0$ с зарядовым числом $z_i \geq 0$ частиц i -го сорта, входящим в электрохимический потенциал (1.16.2), в частности для электронов $z_e = -1$). При перемещении заряда $z\mathcal{F}$ совершается электрическая работа $W_{\text{эл}} = z\mathcal{F}\Delta E$.

Для гальванического элемента, как источника тока, электрическая работа всегда положительна (что соответствует термодинамическому правилу знаков, см. п. 1.3), поскольку элемент производит работу над внешней цепью, в которой $E_{\text{вн}} \leq \Delta E$. Электрическая работа ($W_{\text{эл}}$) и механическая работа ($W_{\text{мех}} \equiv P\Delta V$) совместно с тепловым эффектом обратимого процесса ($Q_{\text{обр}} \equiv T\Delta S$) дают вклад во внутреннюю энергию системы, равный $\Delta U = Q_{\text{обр}} - (W_{\text{мех}} + W_{\text{эл}})$ (см. п. 1.6).

С учетом вышесказанного, закон сохранения энергии (1.3.1), записанный для конечных приращений, принимает вид

$$\Delta U = T\Delta S - P\Delta V - z\mathcal{F}\Delta E. \quad (2.18.3)$$

В этом случае свободная энергия Гиббса, $G \equiv U + PV - TS$, получает приращение

$$\Delta G = -S\Delta T + V\Delta P - z\mathcal{F}\Delta E. \quad (2.18.4)$$

Отсюда следует, что электрическая работа в изобарно-изотермических условиях (при $\Delta T = \Delta P = 0$) равна

$$W_{\text{эл}} \equiv z\mathcal{F}\Delta E = -\Delta G. \quad (2.18.5)$$

Это выражение получено для *гальванического режима*, когда химическая реакция *самопроизвольно* протекает слева направо

в силу $\Delta G < 0$, что обеспечивает положительную внутреннюю ЭДС ($\Delta E > 0$). В этом случае (при $\Delta E \geq E_{\text{вн}}$) электрохимический элемент под действием сил химической природы перемещает ионы в электролите против внутреннего электрического поля (анионы — от плюса к минусу, а катионы — от минуса к плюсу), совершая положительную работу над внешней цепью ($W_{\text{эл}} > 0$), т. е. вырабатывая электрическую энергию. Именно такая картина изображена на рис. 2.24 а, чему соответствует реакция (2.18.1), самопроизвольно протекающая слева направо.

Выражение (2.18.5) справедливо и для противоположного направления *той же реакции* (справа налево при $\Delta G > 0$). Однако такая ситуация реализуема лишь в *электролитическом режиме*, когда внешняя ЭДС такова, что $E_{\text{вн}} \geq \Delta E$. Это соотношение заставляет химическую реакцию протекать в *несамопроизвольном* для нее направлении. При этом внешняя ЭДС совершает электрическую работу ($W_{\text{эл}} < 0$) по перемещению ионов электролита в направлении внутреннего электрического поля (анионы — от минуса к плюсу, а катионы — от плюса к минусу), что несвойственно самой реакции. Именно такая картина изображена на рис. 2.24 б, где анионы SO_4^{2-} движутся от отрицательного электрода к положительному, а катионы — в обратном направлении.

Следует отметить, что для одного и того же электрохимического элемента внутреннее электрическое поле всегда направлено одинаково как в гальваническом, так и в электролитическом режимах, а движение ионов в растворе по отношению к внутреннему полю для этих режимов противоположно. Это ясно видно из сравнения рис. 2.24 а и б. Несмотря на то, что жирная стрелка, обозначающая направление поля, в обоих случаях одного направления, знак обратимой внутренней ЭДС различен.

Электродвижущая сила считается *положительной* ($\Delta E > 0$), если анионы в растворе движутся от правого электрода к левому (одинаковое направление фигурной и жирной стрелок на рис. 2.24 а), и *отрицательной* ($\Delta E < 0$) — при обратном движении анионов (встречное направление фигурной и жирной стрелок на рис. 2.24 б).

Изобарный потенциал химической реакции задается уравнением изотермы Вант-Гоффа (2.13.9), переписанным через активности a_i химических компонентов с учетом (2.13.8), в следующей форме (пренебрегая влиянием газовой фазы):

$$\Delta G = RT \left[\left(\sum_{\text{к}} \nu_{\text{к}} \ln a_{\text{к}} - \sum_{\text{н}} \nu_{\text{н}} \ln a_{\text{н}} \right) - \ln K_a(T) \right] \equiv$$

$$\equiv \Delta G_T^\circ + RT \ln \frac{\prod_{\kappa} a_{\kappa}^{\nu_{\kappa}}}{\prod_{\text{H}} a_{\text{H}}^{\nu_{\text{H}}}}. \quad (2.18.6)$$

Константа химического равновесия $K_a(T)$, входящая в закон действующих масс (2.13.10) (где $\bar{a}_i$ — равновесные активности),

$$K_a(T) = \frac{\prod_{\kappa} \bar{a}_{\kappa}^{\nu_{\kappa}}}{\prod_{\text{H}} \bar{a}_{\text{H}}^{\nu_{\text{H}}}}, \quad (2.18.7)$$

связана со стандартным изобарным потенциалом ΔG_T° соотношением

$$\Delta G_T^\circ = -RT \ln K_a(T). \quad (2.18.8)$$

Из выражений (2.18.5) и (2.18.6) получаем обратимую ЭДС электрохимического элемента в реальных условиях,

$$\Delta E = -\frac{\Delta G}{z\mathcal{F}} = \Delta E_T^\circ - \frac{RT}{z\mathcal{F}} \ln \frac{\prod_{\kappa} a_{\kappa}^{\nu_{\kappa}}}{\prod_{\text{H}} a_{\text{H}}^{\nu_{\text{H}}}}. \quad (2.18.9)$$

Введенная здесь стандартная ЭДС (при $a_i = 1$) равна

$$\Delta E^\circ = -\frac{\Delta G_T^\circ}{z\mathcal{F}} = \frac{RT}{z\mathcal{F}} \ln K_a \equiv \frac{k_{\text{B}}T}{ze} \ln K_a. \quad (2.18.10)$$

Формула (2.18.9) является прямым аналогом уравнения изотермы Вант-Гоффа и называется **уравнением Нернста для ЭДС гальванического элемента**.

Это уравнение вместе с уравнением (2.18.10) для стандартной ЭДС позволяет, рассчитав для любой реакции термодинамические величины K_a и ΔG , найти реальную и стандартную электродвижущую силу элемента. Можно, наоборот, по измеренным электрическим величинам ΔE и ΔE° вычислить термодинамические характеристики химической реакции. Как следует из уравнения Гиббса–Гельмгольца (1.10.16), чтобы найти ΔH и ΔS для химической реакции, протекающей в электрохимическом элементе, необходимо знать температурную зависимость $\Delta E(T)$. Тогда

$$\Delta H = \Delta G - T \left(\frac{\partial \Delta G}{\partial T} \right)_P = -z\mathcal{F} \left(\Delta E - T \frac{\partial \Delta E}{\partial T} \right),$$

$$\Delta S \equiv \frac{\Delta H - \Delta G}{T} = - \left(\frac{\partial \Delta G}{\partial T} \right)_P = z\mathcal{F} \frac{\partial \Delta E}{\partial T}.$$

Легко убедиться в том, что уравнение Нернста (2.18.9) соответствует *электрохимическому равновесию* в системе. Для этого надо отождествить рассмотренные здесь величины ΔG и $W_{\text{эл}}$ с химическим ($\Delta G_{\text{хим}}$) и электрическим ($\Delta G_{\text{эл}}$) вкладом в изобарный потенциал реакции, введенными формулами (1.16.8) и (1.16.9), т. е. положить $\Delta G \equiv \Delta G_{\text{хим}}$ и $W_{\text{эл}} \equiv \Delta G_{\text{эл}}$. Подстановка этих тождеств в формулу (2.18.5) приводит ее к условию электрохимического равновесия (1.16.10).

Таким образом, модуль и знак обратимой ЭДС ΔE электрохимического элемента определяются модулем и знаком изобарного потенциала ΔG химической реакции, протекающей в этом элементе. Если условия таковы, что реакция находится в химическом равновесии ($\Delta G = 0$), то ЭДС отсутствует ($\Delta E = 0$). Она возникает лишь в том случае, когда химическая реакция протекает *самопроизвольно* ($\Delta G < 0$), будучи способной совершать электрическую работу $W_{\text{эл}} > 0$ *лишь в гальваническом режиме*.

Как известно (см. п. 1.14), изобарным потенциалом ΔG химической реакции, записанной в обобщенной форме (2.12.1), принято называть изменение свободной энергии $G_{\text{к}} - G_{\text{н}}$, происходящее в системе при одном пробеге реакции *слева* (где записаны начальные химические реагенты, имеющие $G_{\text{н}}$) *направо* (где стоят конечные продукты реакции, имеющие $G_{\text{к}}$). Такая условность, принятая в отношении $\Delta G = G_{\text{к}} - G_{\text{н}}$, делает вполне определенным знак обратимой ЭДС ΔE . С точки зрения расположения электродов в электрохимическом элементе условились величину ΔE вычислять как разность вводимых ниже *электродных потенциалов* для правого ($E_{\text{п}}$) и левого ($E_{\text{л}}$) электродов:

$$\Delta E = E_{\text{п}} - E_{\text{л}}. \quad (2.18.11)$$

В соответствии с ранее установленным правилом, *в гальваническом режиме*, когда реакция самопроизвольно протекает слева направо (при $\Delta G < 0$), имеем $\Delta E > 0$, т. е. правый электрод положительнее левого ($E_{\text{п}} > E_{\text{л}}$), при этом обязательно ток в растворе (движение положительных ионов) должен протекать также слева направо. Если же ток оказывается противоположного направления (движение катионов от правого электрода к левому) *в гальваническом режиме*, то это означает $\Delta E < 0$, т. е. левый электрод положительнее правого ($E_{\text{л}} > E_{\text{п}}$), что возможно лишь при протекании обратимой реакции справа налево (при $\Delta G > 0$).

Следовательно, выбор знаков таков, что всегда *в гальваническом режиме* направление возрастания электродных потенциалов совпадает с направлением самопроизвольного протекания реакции и с направлением тока (с движением катионов в растворе):

при $\Delta G < 0$ и $\Delta E > 0$ все три направления идут слева направо, а при $\Delta G > 0$ и $\Delta E < 0$ — справа налево (см. рис. 2.24 *a* для реакции (2.18.1), соответствующей первому случаю).

Итак, в гальваническом элементе обратимая ЭДС ($\Delta E > 0$ или $\Delta E < 0$) возникает лишь при самопроизвольном протекании реакции (слева направо при $\Delta G < 0$ или справа налево при $\Delta G > 0$). Та же реакция, протекающая в обратном направлении, не только не может служить источником электрической энергии, но, наоборот, требует затраты энергии извне путем приложения внешней противоЭДС, такой что $|E_{\text{вн}}| \geq |\Delta E|$, т. е. может осуществляться лишь в режиме электролиза.

В качестве примера вычислим, пользуясь таблицами стандартных термодинамических величин, значения изобарного потенциала при образовании соединений из простых веществ для реакции (2.18.1) при 298 К: $\Delta G^\circ_{\text{CuSO}_4(\text{ж})} = -161,8$ ккал/моль, $\Delta G^\circ_{\text{ZnSO}_4(\text{ж})} = -212,5$ ккал/моль, при этом для простых веществ $\Delta G^\circ_{\text{Cu}(\text{т})} = 0$ и $\Delta G^\circ_{\text{Zn}(\text{т})} = 0$. Отсюда стандартный изобарный потенциал реакции равен $\Delta G^\circ_{298} = \Delta G^\circ_{\text{ZnSO}_4(\text{ж})} - \Delta G^\circ_{\text{CuSO}_4(\text{ж})} = -50,7$ ккал/моль. Как уже было отмечено (см. п. 2.15), логарифмический вклад в величину ΔG , определяемую формулой (2.18.6), может быть опущен при $|\Delta G^\circ_{298}| > 10$ ккал/моль. Поэтому в данном случае $\Delta G \approx \Delta G^\circ_{298} = -50,7$ ккал/моль, т. е. реакция (2.18.1) самопроизвольно протекает слева направо. Это означает, что цинк растворяется с образованием ZnSO_4 и вытесняет из раствора медь, осаждающуюся на медном электроде. Соответствующая этому процессу стандартная ЭДС элемента Даниэля–Якоби равна $\Delta E^\circ = -\Delta G^\circ_{298}/2\mathcal{F} = +1,1$ В. Отсюда, согласно установленному правилу, при $\Delta E^\circ > 0$ правый (медный) электрод заряжается положительно по отношению к левому (цинковому) электроду, что и показано на рис. 3.2 *a*. Ниже мы подтвердим полученный результат численного расчета рассмотрением окислительно-восстановительных реакций на электродах.

2.19. Электродные процессы в электрохимических системах

Основное различие между химическими и электрохимическими реакциями состоит в том, что в акте электрохимического взаимодействия непосредственный контакт между реагирующими частицами (имеющий место в химической реакции) заменен на контакт участников реакции с электродами. В результате этого

происходит обмен ионами между электродом и раствором электролита. Электроды, обеспечивающие обмен катионами (положительными ионами), называют *обратимыми относительно катиона*, а электроды, обеспечивающие обмен анионами (отрицательными ионами) — *обратимыми относительно аниона*.

1. Окислительно-восстановительные реакции на электродах. Существуют электроды, которые не обмениваются ионами с раствором, а обеспечивают только подвод или отвод электронов, необходимых для осуществления окислительно-восстановительной реакции на своей поверхности. Эти электроды принято называть *окислительно-восстановительными* или *редокс-электродами* (сокращение от англ. *reduction* — восстановление, *oxidation* — окисление). Для таких электродов, в отличие от обратимых, отсутствует массоперенос их материала через раствор.

Следовательно, в электрохимических системах электронные и ионные процессы на электродах играют ключевую роль.

Выясним физическую природу возникновения обратимой ЭДС в электрохимическом элементе путем изучения термодинамики электродных процессов на металле, погруженном в водный раствор электролита. Для этого сначала покажем, что при соприкосновении металла M с раствором, содержащим ионы M^{z+} того же металла, на границе электрод–раствор возникает двойной заряженный слой.

Электрохимический потенциал, введенный формулой (1.16.2), для иона M^{z+} в кристаллической решетке металла ($\tilde{\mu}_M^T$) можно считать постоянным при данной температуре, а в растворе электролита ($\tilde{\mu}_M^{\text{ж}}$) — зависящим от его концентрации (активности). При $\tilde{\mu}_M^T > \tilde{\mu}_M^{\text{ж}}$ начинается переход ионов M^{z+} из металла в раствор. При этом металл заряжается отрицательно, а раствор по отношению к нему положительно, т. е. в районе границы раздела возникает двойной заряженный слой с электрическим полем, способствующим переходу ионов в обратном направлении. По мере прямого перехода ионов в электролит их концентрация в растворе возрастает, что увеличивает электрохимический потенциал $\tilde{\mu}_M^{\text{ж}}$. Так продолжается до тех пор, пока скорости прямых и обратных переходов металл $\rightleftharpoons$ раствор не уравниваются, т. е. пока не наступит фазовое равновесие, при котором, согласно (1.16.3),

$$\tilde{\mu}_M^T = \tilde{\mu}_M^{\text{ж}}, \quad \text{или} \quad (\mu_M^T - \mu_M^{\text{ж}}) + z\mathcal{F}(\varphi^T - \varphi^{\text{ж}}) = 0. \quad (2.19.1)$$

Здесь φ^T и $\varphi^{\text{ж}}$ — электрические потенциалы металла и раствора электролита, при этом в рассматриваемом случае имеем $\varphi^T < \varphi^{\text{ж}}$ и $\mu_M^T > \mu_M^{\text{ж}}$. Следовательно, разность электрических по-

тенциалов при равновесии скомпенсирована разностью химических потенциалов.

Если исходная концентрация электролита такова, что $\tilde{\mu}_M^T < \tilde{\mu}_M^J$, то при погружении металла в раствор часть ионов M^{z+} из раствора переходит в металл, заряжая его положительно по отношению к раствору, так что при равновесии $\varphi^T > \varphi^J$.

Таким образом, в обоих случаях (при $\tilde{\mu}_M^T > \tilde{\mu}_M^J$ и $\tilde{\mu}_M^T < \tilde{\mu}_M^J$) на границе раздела электрод–раствор возникает двойной электрический слой, одной обкладкой которого является заряд на металле, локализованный практически на его поверхности. Равный и противоположный по знаку заряд в растворе создается ионами: катионами при $\varphi^T < \varphi^J$ и анионами при $\varphi^T > \varphi^J$. Так как ионы в растворе совершают тепловое движение, то ионная обкладка двойного слоя частично размыта в так называемый *диффузный слой* (или *слой Гуи*). Оставшаяся меньшая часть ионного заряда, локализованная между диффузным слоем и поверхностью металла, называется *плотным слоем* (или *слоем Гельмгольца*).

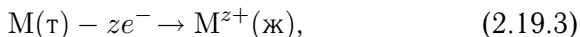
Результирующее падение напряжения $\varphi = \varphi^T - \varphi^J$ на диффузном и плотном слоях называют *электрическим потенциалом электрода* по отношению к раствору (не путать с вводимым ниже *электродным потенциалом E*) и находят из равенства (2.19.1) в виде

$$\varphi \equiv \varphi^T - \varphi^J = \frac{1}{zF} (\mu^J - \mu^T). \quad (2.19.2)$$

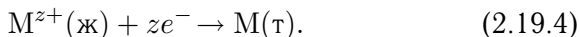
Из сравнения выражения (2.19.2) с ранее записанной формулой (1.16.4) видно, что электрический потенциал электрода φ совпадает с контактной разностью потенциалов $\varphi_{\text{конт}}$, обычно вводимой в электронную теорию металлов и полупроводников.

Описанные процессы обмена ионами M^{z+} между кристаллической решеткой металла и жидким раствором электролита могут быть представлены как две химические реакции:

а) *реакция ионизации* нейтральных атомов металла $M(t)$, которые отдают электроны во внешнюю цепь с образованием ионов $M^{z+}(ж)$ в растворе:



б) *реакция разряда* ионов $M^{z+}(ж)$ на электроде, получающем электроны от внешней цепи, с образованием нейтральных атомов $M(t)$ металла:



Как уже ранее отмечалось, процесс, протекающий по реакции (2.19.3) и сопровождаемый увеличением заряда частицы (в данном случае атома) за счет потери электронов, принято называть

окислением. Обратный процесс, описываемый реакцией (2.19.4) и связанный с уменьшением заряда частицы (иона) в результате приобретения электронов, называется *восстановлением*.

На основании вышесказанного, процесс окисления (2.19.3) протекает всегда на отрицательном полюсе гальванического элемента или на аноде электролитической ячейки (*анодное окисление* при электролизе). Окисление обязательно сопровождается переходом ионов металла в раствор (*анодное растворение* при электролизе).

В противоположность этому, процесс восстановления (2.19.4) идет всегда на положительном полюсе гальванического элемента или на катоде электролитической ячейки (*катодное восстановление* при электролизе). Восстановление обязательно сопровождается переходом ионов из раствора на поверхность металла (*катодное осаждение* при электролизе).

Для рассмотренного в п. 2.18 элемента Даниэля–Якоби, где левый электрод цинковый, а правый — медный (рис. 2.24), обратимая химическая реакция (2.18.1), лежащая в основе работы этого электрохимического элемента, была записана в ионной форме (2.18.2), которая в свою очередь может быть представлена в виде двух электродных реакций:

а) в *гальваническом режиме* для прямой реакции (протекающей слева направо)

$\text{Zn} - 2e^- \rightarrow \text{Zn}^{2+}$ (окисление и растворение Zn на отрицательном полюсе),

$\text{Cu}^{2+} + 2e^- \rightarrow \text{Cu}$ (восстановление и осаждение Cu на положительном полюсе);

б) в *электролитическом режиме* для обратной реакции (протекающей справа налево)

$\text{Cu} - 2e^- \rightarrow \text{Cu}^{2+}$ (окисление и растворение Cu на аноде),

$\text{Zn}^{2+} + 2e^- \rightarrow \text{Zn}$ (восстановление и осаждение Zn на катоде).

Таким образом, на электродах электрохимического элемента всегда протекает обратимая *окислительно-восстановительная реакция*, записанная в обобщенной форме как



где Ox — окисленная форма (или *окислитель* от англ. *Oxidant*), Red — восстановленная форма (или *восстановитель* от англ. *Reductant*). Одна из этих двух форм может быть нейтральной.

В уравнении (2.19.5) прямая реакция приобретения электронов окислителем ($\text{Ox} + ze^- \rightarrow \text{Red}$), описывает его восстано-

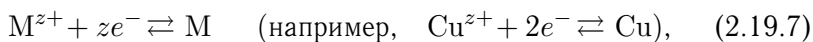
ние, а обратная реакция, связанная с потерей электронов восстановителем ($\text{Red} - ze^- \rightarrow \text{Ox}$), соответствует его окислению.

Поскольку заряд электрона отрицательный, то всегда окислитель, принимающий электроны, изначально имеет больший заряд, чем его восстановленная форма, на величину $ze > 0$, где e — абсолютная величина заряда электрона. Поэтому из квазихимического уравнения (2.19.5) следует условие электронейтральности:

$$\text{заряд}\{\text{Ox}\} - \text{заряд}\{\text{Red}\} = ze. \quad (2.19.6)$$

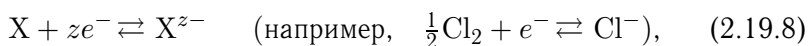
Окислительно-восстановительная форма уравнения (2.19.5) описывает обратимый процесс, называемый *редокс-процессом* (от англ. сокращения *redox*).

К редокс-процессам типа (2.19.5) относятся реакции (2.19.3) и (2.19.4) для металла М, которые имеют общую обратимую форму записи,



называемую *обратимой относительно катиона*. На электроде, обратимом относительно катиона, в процессе восстановления (при протекании реакции слева направо) выделяется металл.

К редокс-процессам типа (2.19.5) относятся также реакции, аналогичные (2.19.3) и (2.19.4), но с участием неметалла Х, записанные в такой форме,



которая называется *обратимой относительно аниона*. На электроде, обратимом относительно аниона, в процессе окисления (при протекании реакции справа налево) выделяется неметалл (типа S, Se и Te) или химически активный газ (типа F_2 , Cl_2 , O_2 и H_2 — газовые электроды).

2. Электродные потенциалы. Электрический потенциал (φ) электрода по отношению к раствору, введенный равенством (2.19.2), невозможно измерить экспериментально. Для этого обязательно потребовался бы еще один измерительный электрод, контактирующий с раствором без скачка потенциалов, называемый *нулевым электродом*. Такого универсального электрода, пригодного для электролитов разного состава, в природе не существует. Практический интерес представляет только разность электрических потенциалов. Поэтому условились отсчитывать потенциал любого электрода от потенциала так называемого *нормального водородного электрода*, принятого в качестве электрода сравнения, чей *стандартный* электродный потенциал

положен нулевым. Иными словами, водородный электрод выполняет функцию своеобразного «электрохимического заземления».

Потенциал электрода, определенный по водородной шкале (относительно нормального водородного электрода), принято называть **электродным потенциалом** и обозначать его буквой E . Для определенности условились в *стандартном* гальваническом элементе водородный электрод всегда брать *левым* с электрическим потенциалом $\varphi_{\text{л}} \equiv \varphi_{\text{Н}}$. Для другого электрода (правого) с электрическим потенциалом $\varphi_{\text{п}} \equiv \varphi$ его электродный потенциал равен $E \equiv \varphi_{\text{п}} - \varphi_{\text{л}} = \varphi - \varphi_{\text{Н}}$. При этом соответствующая химическая реакция записана в виде (2.19.5), где окисленная форма стоит слева, а восстановленная форма — справа.

Электроды, обратимые относительно катиона, ведут себя таким образом, что их электродный потенциал E : а) *положителен*, если реакция (2.19.5) на данном электроде протекает в направлении восстановления (слева направо при $\Delta G < 0$) с переходом катионов из раствора на электрод (осаждение вещества); б) *отрицателен* при обратном (окислительном) направлении реакции (справа налево при $\Delta G > 0$), сопровождаемом переходом катионов с электрода в раствор (растворение вещества).

Отсюда ясно видно, что в реальном гальваническом элементе металл с более низким электродным потенциалом *самопроизвольно* растворяется и вытесняет из раствора электролита ионы другого металла, имеющего больший электродный потенциал. Переход ионов металла в обратном направлении требует приложения внешней ЭДС, то есть он возможен лишь в режиме электролиза.

Электроды, обратимые относительно аниона, ведут себя прямо противоположно — электрод приобретает: а) *положительный* потенциал E при переходе анионов с электрода в раствор (растворение вещества); б) *отрицательный* потенциал E при обратном переходе анионов из раствора на электрод (осаждение вещества). В этом случае неметалл вытесняет из раствора другие неметаллы с меньшим электродным потенциалом.

Более строгое термодинамическое обоснование правила вытеснения металлов и неметаллов из раствора электролитов будет дано в конце параграфа.

В реальном электрохимическом элементе с электрическими потенциалами правого и левого электродов ($\varphi_{\text{п}}$ и $\varphi_{\text{л}}$) их электродные потенциалы (вводимые как $E_{\text{п}}$ и $E_{\text{л}}$, отсчитанные от электрического потенциала водородного электрода $\varphi_{\text{Н}}$), соответственно, равны $E_{\text{п}} = \varphi_{\text{п}} - \varphi_{\text{Н}}$ и $E_{\text{л}} = \varphi_{\text{л}} - \varphi_{\text{Н}}$. В этом случае из

определения (2.18.11) получаем обратимую ЭДС элемента:

$$\Delta E \equiv E_{\text{п}} - E_{\text{л}} = (\varphi_{\text{п}} - \varphi_{\text{Н}}) - (\varphi_{\text{л}} - \varphi_{\text{Н}}) = \varphi_{\text{п}} - \varphi_{\text{л}} \equiv \Delta \varphi.$$

Следовательно, ЭДС элемента, вычисляемая как разность электродных потенциалов правого и левого электродов, равна разности электрических потенциалов этих электродов по отношению к раствору. В величину ΔE вносит вклад также контактная разность потенциалов $\varphi_{\text{конт}}$ между металлами с разными работами выхода электрона, даваемая формулой (1.16.4).

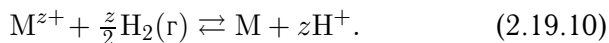
Водородный электрод обычно выполняется в виде платиновой пластины с губчатым поверхностным слоем, опущенной в раствор серной кислоты и омываемой потоком газообразного водорода. На поверхности платины, насыщенной растворенным в ней водородом, протекает реакция



аналогичная реакции (2.19.7), идущей на поверхности металлических электродов, обратимых относительно катиона. Платина Pt, как и другие благородные металлы типа Au, Pd, Ir, играет роль инертного материала, в котором растворенный водород имеет устойчивый электродный потенциал.

Для нахождения электродного потенциала металла М, на котором протекает реакция (2.19.7), обратимая относительно катиона, составляется стандартный гальванический элемент, состоящий из водородного электрода (левого) и металлического электрода (правого).

При записи химической реакции, протекающей в этом элементе, в уравнении (2.19.7) слева и справа добавляется слагаемое $z\text{H}^+$. Тогда с учетом реакции (2.19.9) получаем



ЭДС такого элемента, вычисленная по формуле (2.18.9) для реакции (2.19.10), равна

$$\Delta E = \Delta E^\circ - \frac{RT}{z\mathcal{F}} \ln \frac{a_{\text{М}} a_{\text{H}^+}^z}{a_{\text{М}^{z+}} p_{\text{H}_2}^{z/2}}. \quad (2.19.11)$$

Согласно определению ЭДС в виде (2.18.11), имеем $\Delta E \equiv E_{\text{п}} - E_{\text{л}} = E_{\text{М}} - E_{\text{Н}}$ и $\Delta E^\circ = E_{\text{М}^{z+}}^\circ - E_{\text{H}^+}^\circ$.

Тогда из выражения (2.19.11) можно выделить два электродных потенциала:

а) для металлического электрода М ($a_M = 1$ — чистый металл)

$$E_M = E_{M^{z+}}^\circ - \frac{RT}{z\mathcal{F}} \ln \frac{a_M}{a_{M^{z+}}} \equiv E_{M^{z+}}^\circ + \frac{RT}{z\mathcal{F}} \ln a_{M^{z+}}; \quad (2.19.12)$$

б) для водородного электрода (по реакции (2.19.9))

$$E_H = E_{H^+}^\circ - \frac{RT}{F} \ln \frac{p_{H_2}^{1/2}}{a_{H^+}}. \quad (2.19.13)$$

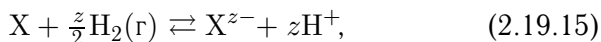
Действительно, разность выражений (2.19.12) и (2.19.13) дает формулу (2.19.11).

Водородная шкала электродных потенциалов построена таким образом, что стандартному состоянию соответствуют $p_{H_2}^\circ = 1$ атм и $a_{H^+}^\circ = 1$ (или $c_{H^+}^\circ = 1$ г-ион/л); при этом стандартный потенциал водородного электрода при всех температурах принят равным нулю ($E_{H^+}^\circ = 0$). Тогда из (2.19.13) получаем потенциал водородного электрода при $p_{H_2} = 1$ атм:

$$E_H = \frac{RT}{\mathcal{F}} \ln a_{H^+} = -2,3 \frac{RT}{\mathcal{F}} \text{ рН}. \quad (2.19.14)$$

Здесь $\text{рН} = -\lg a_{H^+}$ — водородный показатель, записанный по аналогии с выражением (2.17.7) через активность a_{H^+} водородных ионов (точнее, ионов гидроксония H_3O^+).

Аналогично реакция (2.19.8) для неметаллического электрода Х, обратимая относительно аниона, вместе с водородной реакцией (2.19.9) дает суммарную реакцию



для которой формула (2.18.9) дает обратимую ЭДС:

$$\Delta E = \Delta E^\circ - \frac{RT}{z\mathcal{F}} \ln \frac{a_{X^{z-}} a_{H^+}^z}{a_X p_{H_2}^{z/2}}. \quad (2.19.16)$$

Используя потенциал водородного электрода (2.19.14), из формулы (2.19.16) получаем электродный потенциал неметаллического электрода Х (для чистого вещества $a_X = 1$):

$$E_X = E_{X^{z-}}^\circ - \frac{RT}{z\mathcal{F}} \ln \frac{a_{X^{z-}}}{a_X} \equiv E_{X^{z-}}^\circ - \frac{RT}{z\mathcal{F}} \ln a_{X^{z-}}. \quad (2.19.17)$$

Формулы типа (2.19.12), (2.19.13) и (2.19.17) часто называют **уравнениями Нернста для электродного потенциала**.

Сопоставление каждой из этих формул с соответствующей химической реакцией (2.19.7), (2.19.9) и (2.19.8) позволяет легко установить следующее правило их написания: логарифмическое

слагаемое, добавляемое с множителем $(-RT/z\mathcal{F})$ к стандартному потенциалу, содержит вклады восстановленной формы (X^{z-}) в числителе и окисленной формы (M^{z+}) в знаменателе, а вклад электронов исключен.

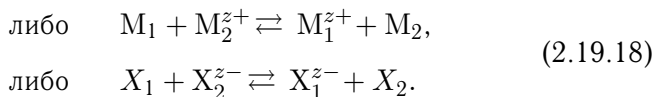
В реальных условиях расчет электродного потенциала по уравнению Нернста возможен, если для обратимых ионов в растворе известны как их активности $a_{M^{z+}}$ и $a_{X^{z-}}$ (или концентрации $c_{M^{z+}}$ и $c_{X^{z-}}$ при сильном разбавлении), так и стандартные электродные потенциалы $E_{M^{z+}}^{\circ}$ и $E_{X^{z-}}^{\circ}$. Электродные потенциалы обычно измеряют в стандартных условиях при $T = 298$ К и $a_{M^{z+}}^{\circ} = a_{X^{z-}}^{\circ} = 1$ (или $c_{M^{z+}}^{\circ} = c_{X^{z-}}^{\circ} = 1$ г-ион/л). Для различных веществ в справочной литературе приведены стандартные электродные потенциалы в виде так называемого **ряда напряжений**, построенного в порядке возрастания значений E° , начиная с лития (сильного восстановителя, имеющего $E_{Li^{+}}^{\circ} = -3,045$ В) вверх ряда и кончая фтором (сильным окислителем, имеющим $E_{F^{-}}^{\circ} = +2,65$ В) вниз ряда.

В отличие от стандартного изобарного потенциала реакции ΔG° , электродный потенциал вещества E° не изменяется при пропорциональном изменении стехиометрических коэффициентов в реакциях типа (2.19.5). Дело в том, что величина E° приведена к одному электрону, о чем свидетельствует наличие числа z в знаменателе формул (2.19.12) и (2.19.17). Например, для хлора имеем $E_{Cl^{-}}^{\circ} = +1,36$ В независимо от записи реакции в форме $\frac{1}{2}Cl_2 + e^{-} \rightleftharpoons Cl^{-}$ или в форме $Cl_2 + 2e^{-} \rightleftharpoons 2Cl^{-}$. Однако величина E° для одного и того же вещества зависит от заряда его иона в растворе. Так, для однозарядных и двухзарядных ионов меди имеем $E_{Cu^{+}}^{\circ} = +0,521$ В и $E_{Cu^{2+}}^{\circ} = +0,337$ В.

Знание электродных потенциалов позволяет рассчитать ЭДС гальванического элемента по формуле (2.18.11). В случае элемента Даниэля–Якоби (рис. 2.24) для цинкового (левого) и медного (правого) электродов из таблиц находим $E_{Zn^{2+}}^{\circ} = -0,763$ В и $E_{Cu^{2+}}^{\circ} = +0,337$ В. Отсюда получаем стандартную ЭДС как $\Delta E^{\circ} = E_{Cu^{2+}}^{\circ} - E_{Zn^{2+}}^{\circ} = +1,1$ В. Это значение оказалось в точности равным величине, полученной в предыдущем параграфе для реакции (2.18.1) путем расчета ее стандартного изобарного потенциала ΔG_{298}° .

Электродные потенциалы могут применяться для анализа не только электрохимических реакций, но и обычных реакций окислительно-восстановительного типа, протекающих по химическому пути без участия металлических электродов, поскольку $\Delta E = -\Delta G/z\mathcal{F}$ (см. формулы (2.18.6) и (2.18.9)).

Рассмотрим применение такого подхода для анализа взаимодействия между солями M_1X_1 и M_2X_2 в водном растворе. Если рассматриваемые соли имеют общий анион $X_1^{z_1-} = X_2^{z_2-} \equiv X^{z-}$ или катион $M_1^{z_1+} = M_2^{z_2+} \equiv M^{z+}$, то этот ион может быть исключен из уравнения реакции при записи ее в ионной форме (см. формулы (2.18.1) и (2.18.2)):



Уравнение (2.18.9) для обратимой ЭДС принимает соответствующий вид (для чистых веществ $a_{M_1} = a_{M_2} = a_{X_1} = a_{X_2} = 1$):

$$\begin{aligned} \text{либо} \quad \Delta E &= \Delta E_M^\circ - \frac{RT}{z\mathcal{F}} \ln \frac{a_{M_1^{z+}}}{a_{M_2^{z+}}} \equiv E_{M_2} - E_{M_1}, \\ \text{либо} \quad \Delta E &= \Delta E_X^\circ - \frac{RT}{z\mathcal{F}} \ln \frac{a_{X_1^{z-}}}{a_{X_2^{z-}}} \equiv E_{X_1} - E_{X_2}, \end{aligned} \quad (2.19.19)$$

где использованы следующие обозначения для обратимой ЭДС в стандартных условиях, а именно

$$\Delta E_M^\circ = E_{M_2^{z+}}^\circ - E_{M_1^{z+}}^\circ \quad \text{и} \quad \Delta E_X^\circ = E_{X_1^{z-}}^\circ - E_{X_2^{z-}}^\circ.$$

При записи тождественных равенств в (2.19.19) были использованы выражения (2.19.12) и (2.19.17) для электродных потенциалов металлов $E_{M_{1,2}}$ и неметаллов $E_{X_{1,2}}$, соответственно.

Пусть для определенности $\Delta G < 0$, т. е. каждая из реакций (2.19.18) протекает слева направо с превращением вещества с индексом 1 в ионы раствора и с выделением из раствора вещества с индексом 2. При этом $\Delta E > 0$ и выражения (2.19.19) дают следующие соотношения между электродными потенциалами: $E_{M_1} < E_{M_2}$ и $E_{X_1} > E_{X_2}$.

Следовательно, металл M_1 с меньшим электродным потенциалом растворяется в виде катиона и вытесняет из раствора катионы другого металла M_2 с большим потенциалом, восстанавливая их до чистого металла. Для неметаллов правило вытеснения прямо противоположное: любой неметалл X_1 вытесняет из раствора другие неметаллы X_2 , имеющие меньший электродный потенциал.

Все вышесказанное касается в общем случае электродных потенциалов, учитывающих, в соответствии с уравнениями Нернста (2.19.12) и (2.19.17), логарифмический вклад активности (или концентрации при сильном разбавлении) для обратимого

иона в реальном растворе. На практике для этих целей обычно используют стандартные значения электродных потенциалов, представленные, как уже говорилось, в виде *ряда напряжений*.

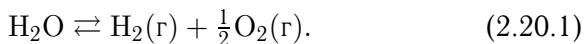
Вверху ряда напряжений расположены: *активные металлы* (литий с $E_{\text{Li}^+}^\circ = -3,045$ В, калий с $E_{\text{K}^+}^\circ = -2,925$ В, натрий с $E_{\text{Na}^+}^\circ = -2,714$ В), а также *пассивные неметаллы* (теллур с $E_{\text{Te}^{2-}}^\circ = -1,14$ В, селен с $E_{\text{Se}^{2-}}^\circ = -0,77$ В, сера с $E_{\text{S}^{2-}}^\circ = -0,51$ В) — *сильные восстановители* (отдающие электроны) с наименьшими (отрицательными) значениями E° .

Внизу ряда напряжений расположены: *пассивные металлы* (серебро с $E_{\text{Ag}^+}^\circ = +0,8$ В, платина с $E_{\text{Pt}^{2+}}^\circ = +1,2$ В, золото с $E_{\text{Au}^{3+}}^\circ = +1,5$ В и $E_{\text{Au}^+}^\circ = +1,69$ В), а также *активные неметаллы* (бром с $E_{\text{Br}^-}^\circ = +1,065$ В, хлор с $E_{\text{Cl}^-}^\circ = +1,36$ В, фтор с $E_{\text{F}^-}^\circ = +2,65$ В) — *сильные окислители* (принимающие электроны) с наибольшими (положительными) значениями E° (см. на рис. 2.25 левую и правую вертикальные оси с направлением, обратным ряду напряжений).

Следовательно, при контактировании через водную среду двух металлов первым начинает растворяться более активный металл (с меньшим значением $E_{\text{M}^{z+}}^\circ$) и только после его полного исчезновения может начаться растворение благородного металла (с большим значением $E_{\text{M}^{z+}}^\circ$). По этой причине цинк с $E_{\text{Zn}^{2+}}^\circ = -0,763$ В защищает от коррозии железо (оцинкованное железо) с $E_{\text{Fe}^{2+}}^\circ = -0,44$ В и $E_{\text{Fe}^{3+}}^\circ = -0,036$ В, а золото, имеющее наибольший стандартный электродный потенциал, устойчиво в большинстве химически активных сред.

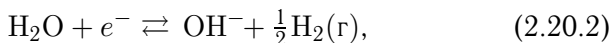
2.20. Управление ионными процессами в водных растворах электролитов

До сих пор вода рассматривалась как химически пассивный растворитель. Однако она может принимать активное участие в химических и электрохимических реакциях с выделением газообразного водорода или кислорода. В основе водных реакций лежит процесс электролитического разложения воды, описываемый реакцией

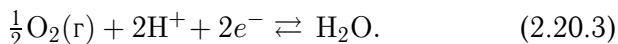


При протекании реакции (2.20.1) по электрохимическому пути она распадается на два электродных процесса:

а) разложение воды путем *восстановления* с выделением водорода (при протекании реакции слева направо),



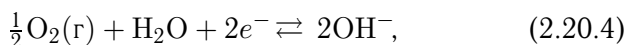
б) разложение воды путем *окисления* с выделением кислорода (при протекании реакции справа налево),



Легко убедиться в том, что суммирование электродных реакций (2.20.2) и (2.20.3) дает общую реакцию разложения воды в виде (2.20.1).

Реакция (2.20.2) является другой формой записи реакции восстановления водорода (2.19.9), видоизмененной за счет прибавления к обеим частям гидроксильного иона OH^- . Поэтому к реакции восстановления воды (2.20.2) применимы формулы (2.19.13) и (2.19.14) для электродного потенциала **водородного электрода**.

Реакция (2.20.3) также прибавлением 2OH^- к обеим частям может быть представлена в новой форме,



отвечающей так называемому **кислородному электроду**, для которого стандартный электродный потенциал $E_{\text{OH}^-}^\circ = +0,401 \text{ В}$.

В условиях, отличных от стандартных, электродный потенциал кислородного электрода записывается по аналогии с водородным электродом (см. формулы (2.19.9) и (2.19.13)) в виде уравнения Нернста для реакции (2.20.4):

$$E_{\text{O}} = E_{\text{OH}^-}^\circ - \frac{RT}{F} \ln \frac{a_{\text{OH}^-}}{p_{\text{O}_2}^{1/4}}. \quad (2.20.5)$$

Здесь активность воды $a_{\text{H}_2\text{O}}$ включена, как постоянная, в стандартный потенциал $E_{\text{OH}^-}^\circ$. Отсюда при $p_{\text{O}_2} = 1 \text{ атм}$ получаем потенциал кислородного электрода в форме

$$\begin{aligned} E_{\text{O}} &= E_{\text{OH}^-}^\circ - \frac{RT}{F} \ln a_{\text{OH}^-} \equiv \\ &\equiv E_{\text{OH}^-}^\circ - 2,3 \frac{RT}{F} [\lg K_w(T) + \text{pH}], \end{aligned} \quad (2.20.6)$$

где использовано ионное произведение воды $K_w(T)$, для которого из закона действующих масс (2.17.3) получаем

$$\lg K_w(T) = \lg a_{\text{H}^+} + \lg a_{\text{OH}^-} \equiv -\text{pH} + \lg a_{\text{OH}^-}, \quad (2.20.7)$$

а $\lg K_w(298) = -14$. Как видно из формулы (2.20.7), стандартному состоянию кислородного электрода ($a_{\text{OH}^-}^\circ = 1$) соответствует предельно щелочная среда ($\text{pH} = 14$), что отличает его от водородного электрода, для которого стандартной является предельно кислая среда ($\text{pH} = 0$ при $a_{\text{H}^+}^\circ = 1$).

Таким образом, электродные процессы восстановления воды по уравнению (2.20.2) и окисления воды по уравнению (2.20.3) управляются, соответственно, реакцией (2.19.9), протекающей на водородном электроде с потенциалом E_{H} в форме (2.19.14), и реакцией (2.20.4), протекающей на кислородном электроде с потенциалом E_{O} в форме (2.20.6).

Уравнения (2.19.14) и (2.20.6) при $T = 298 \text{ K}$, когда $2,3RT/\mathcal{F} \equiv 2,3k_{\text{B}}T/e = 0,059 \text{ В}$, принимают форму,

$$E_{\text{H}} = -0,059 \text{ рН} \quad \text{и} \quad E_{\text{O}} = 0,401 + 0,059 (14 - \text{рН}), \quad (2.20.8)$$

выражающую в явном виде зависимость электродных потенциалов от водородного показателя рН раствора.

Зависимости (2.20.8) изображены на рис. 2.25 сплошными прямыми линиями, где возрастание потенциала E снизу вверх противоположно расположению электродных потенциалов сверху вниз, принятому в ряду напряжений (см. с. 317). Слева и справа от графика приведены значения электродных потенциалов для металлов ($E_{\text{M}^{z+}}^\circ$) и неметаллов ($E_{\text{X}^{z-}}^\circ$) с указанием зарядового состояния их ионов.

Следует иметь в виду, что выражения (2.20.8) получены из уравнений (2.19.14) и (2.20.6), справедливых лишь при $p_{\text{H}_2} = p_{\text{O}_2} = 1 \text{ атм}$. Более общие уравнения (2.19.13) и (2.20.5), учитывающие реальные давления водорода и кислорода в системе, дают:

$$E_{\text{H}} = -0,059 \left(\text{рН} + \lg p_{\text{H}_2}^{1/2} \right),$$

$$E_{\text{O}} = 0,401 + 0,059 \left(14 - \text{рН} + \lg p_{\text{O}_2}^{1/4} \right).$$

Как известно, в атмосферном воздухе парциальные давления водорода и кислорода равны $p_{\text{H}_2} \approx 5 \cdot 10^{-7} \text{ атм}$ и $p_{\text{O}_2} \approx 0,21 \text{ атм}$. Это приводит к тому, что на рис. 2.25 линия E_{H} для водородного потенциала параллельно смещается вверх на 0,186 В, а линия E_{O} для кислородного потенциала — вниз на 0,01 В, что отображено штриховыми линиями. Этим различием обычно пренебрегают и на практике пользуются формулами (2.20.8) и соответствующими им сплошными прямыми на рис. 2.25. В этом случае при всех значениях рН реакция (2.20.2) преимущественно сдвинута влево ($\Delta E \equiv E_{\text{H}} < 0$ и $\Delta G > 0$), а реакция (2.20.3) сдвинута

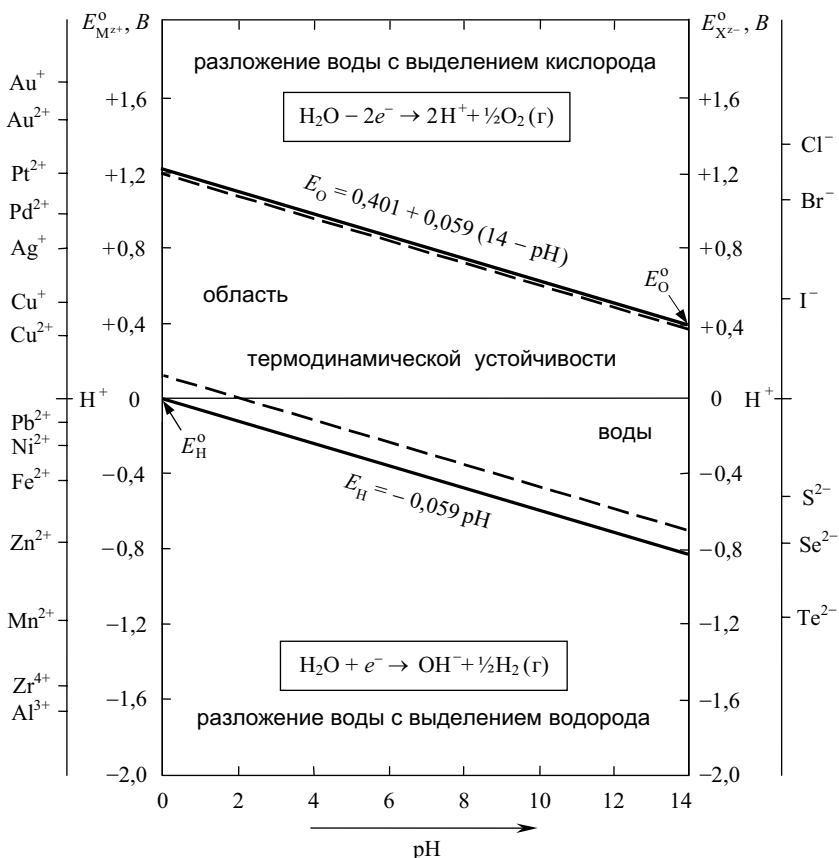


Рис. 2.25. Зависимость электродных потенциалов E_O и E_H для кислородного и водородного электродов от pH раствора с указанием областей электрохимического поведения воды. Вертикальные оси слева и справа содержат электродные потенциалы некоторых металлов ($E_{M^{z+}}^0$) и неметаллов ($E_{X^{z-}}^0$) с указанием зарядового состояния их ионов

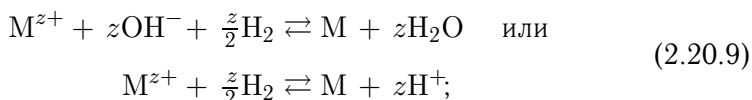
вправо ($\Delta E \equiv E_O > 0$ и $\Delta G < 0$), т. е. обе реакции протекают в направлении образования молекул H_2O . Именно этот факт объясняет малое ионное произведение воды ($K_w = a_{H^+} a_{OH^-} = 10^{-14}$ при 298 K). Однако присутствие в водном растворе соответствующих ионов может обращать протекание этих реакций, т. е. приводить к разложению воды либо с выделением H_2 по реакции (2.20.2), либо с выделением O_2 по реакции (2.20.3).

Для анализа взаимодействия металлов и неметаллов с водой будем действовать по обычному правилу, используя электродные пары, составленные из электрода сравнения и металла М или

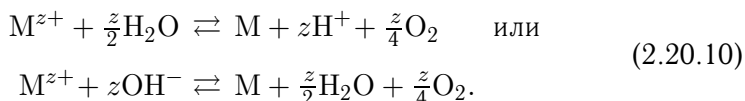
неметалла X. В качестве электрода сравнения (левого) будем брать либо водородный электрод, либо кислородный электрод с электродными потенциалами, даваемыми формулами (2.20.8) для реакций разложения воды (2.20.2) и (2.20.3).

Для **металлов** M результирующие реакции взаимодействия с водой получаются из уравнения (2.19.7): а) вычитанием уравнений (2.20.2) и (2.19.9), умноженных соответственно на z и $z/2$, — для водородного электрода; б) вычитанием уравнений (2.20.3) и (2.20.4), умноженных на $z/2$, — для кислородного электрода. В результате этого получаем:

а) для водородного электрода (комбинация реакции (2.19.7) с реакциями (2.20.2) или (2.19.9))



б) для кислородного электрода (комбинация реакции (2.19.7) с реакциями (2.20.3) или (2.20.4))



Для реакций (2.20.9) и (2.20.10) обратимая ЭДС, вычисленная по уравнению Нернста (2.18.9), равна:

а) для водородного электрода

$$\Delta E = \Delta E_{M-H}^\circ - \frac{RT}{z\mathcal{F}} \ln \frac{a_{H^+}^z}{a_{M^{z+}} p_{H_2}^{z/2}} \equiv E_M - E_H, \quad (2.20.11)$$

где $\Delta E_{M-H}^\circ = E_{M^{z+}}^\circ - E_{H^+}^\circ$;

б) для кислородного электрода

$$\Delta E = \Delta E_{M-O}^\circ - \frac{RT}{z\mathcal{F}} \ln \frac{p_{O_2}^{z/4}}{a_{M^{z+}} a_{OH^-}^z} \equiv E_M - E_O, \quad (2.20.12)$$

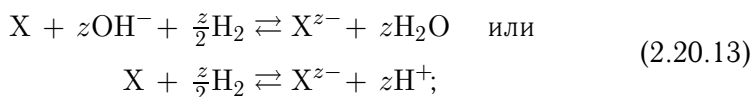
где $\Delta E_{M-O}^\circ = E_{M^{z+}}^\circ - E_{OH^-}^\circ$.

При этом для электродных потенциалов металлического (E_M), водородного (E_H) и кислородного (E_O) электродов использованы, соответственно, формулы Нернста (2.19.12), (2.19.13) и (2.20.5)

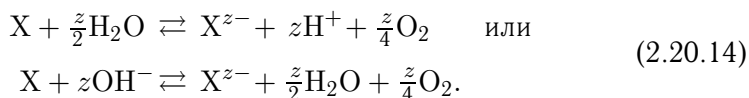
Для **неметаллов** X результирующие реакции взаимодействия с водой получаются из уравнения (2.19.8): а) вычитанием уравнений (2.20.2) и (2.19.9), умноженных соответственно на z и $z/2$, — для водородного электрода; б) вычитанием уравнений (2.20.3)

и (2.20.4), умноженных на $z/2$, — для кислородного электрода. В результате этого получаем:

а) для водородного электрода (комбинация реакции (2.19.8) с реакциями (2.20.2) или (2.19.9))



б) для кислородного электрода (комбинация реакции (2.19.8) с реакциями (2.20.3) или (2.20.4))



Для реакций (2.20.13) и (2.20.14) обратимая ЭДС, вычисленная по уравнению Нернста (2.18.9), равна

а) для водородного электрода

$$\Delta E = \Delta E_{\text{X-H}}^\circ - \frac{RT}{z\mathcal{F}} \ln \frac{a_{\text{X}^{z-}} a_{\text{H}^+}^z}{p_{\text{H}_2}^{z/2}} \equiv E_{\text{X}} - E_{\text{H}}, \quad (2.20.15)$$

$$\text{где} \quad \Delta E_{\text{X-H}}^\circ = E_{\text{X}^{z-}}^\circ - E_{\text{H}^+}^\circ;$$

б) для кислородного электрода

$$\Delta E = \Delta E_{\text{X-O}}^\circ - \frac{RT}{z\mathcal{F}} \ln \frac{a_{\text{X}^{z-}} p_{\text{O}_2}^{z/4}}{a_{\text{OH}^-}^z} \equiv E_{\text{X}} - E_{\text{O}}, \quad (2.20.16)$$

$$\text{где} \quad \Delta E_{\text{X-O}}^\circ = E_{\text{X}^{z-}}^\circ - E_{\text{OH}^-}^\circ.$$

При этом для электродных потенциалов неметаллического (E_{X}), водородного (E_{H}) и кислородного (E_{O}) электродов использованы соответственно формулы Нернста (2.19.17), (2.19.13) и (2.20.5).

2.21. Анализ процессов в гальваническом и электролитическом режимах работы элемента

В гальваническом режиме работы элемента (а также в применении к обычным реакциям, протекающим по *химическому пути*) анализ полученных результатов будем проводить на основе стандартных электродных потенциалов для металлов и неметаллов, некоторые из которых указаны на рис. 2.25. Такой анализ позволяет сделать нижеследующие выводы в отношении самопроизвольности протекания рассмотренных выше реакций.

1. Металлы и неметаллы с E_M и $E_X < E_H$ (чьи электродные потенциалы лежат *ниже* линии E_H на рис. 2.25), согласно уравнениям (2.20.11) и (2.20.15), дают $\Delta E < 0$ и $\Delta G > 0$ для «водородных» реакций (2.20.9) и (2.20.13), что соответствует их протеканию *справа налево*. Следовательно, обе реакции обеспечивают *разложение воды с выделением водорода*, при этом активные металлы (Li, K, Na и др.) термодинамически неустойчивы и растворяются в воде с образованием катионов M^{z+} , а неметаллы (Te, Se и S при $pH < 8$) устойчивы в силу превращения анионов X^{z-} в атомы.

2. Металлы и неметаллы с E_M и $E_X > E_H$ (чьи электродные потенциалы лежат *выше* линии E_H на рис. 2.25), согласно уравнениям (2.20.11) и (2.20.15), дают $\Delta E > 0$ и $\Delta G < 0$ для «водородных» реакций (2.20.9) и (2.20.13), что соответствует их протеканию *слева направо*. Следовательно, такие вещества *не разлагают воду*, а наоборот, взаимодействуя с OH^- , превращают в воду газообразный водород, при этом металлы термодинамически устойчивы, а неметаллы растворяются в воде с образованием анионов X^{z-} .

3. Металлы и неметаллы с E_M и $E_X < E_O$ (чьи электродные потенциалы лежат *ниже* линии E_O на рис. 2.25), согласно уравнениям (2.20.12) и (2.20.16), дают $\Delta E < 0$ и $\Delta G > 0$ для «кислородных» реакций (2.20.10) и (2.20.14), что соответствует их протеканию *справа налево*. Следовательно, такие вещества *не разлагают воду*, а наоборот, взаимодействуя с H^+ , они превращают газообразный кислород в воду, при этом неметаллы термодинамически устойчивы, а металлы растворяются в воде с образованием катионов M^{z+} .

4. Металлы и неметаллы с E_M и $E_X > E_O$ (чьи электродные потенциалы лежат *выше* линии E_O на рис. 2.25), согласно уравнениям (2.20.12) и (2.20.16), дают $\Delta E > 0$ и $\Delta G < 0$ для «кислородных» реакций (2.20.10) и (2.20.14), что соответствует их протеканию *слева направо*. Следовательно, обе реакции обеспечивают *разложение воды с выделением кислорода*, при этом активные неметаллы (F_2 , Cl_2 и Br_2 при $pH > 3$) термодинамически неустойчивы и растворяются в воде с образованием анионов X^{z-} , а металлы (Au, Pt и Pd при $pH > 4$) устойчивы в силу превращения катионов M^{z+} в атомы.

Таким образом, *вода термодинамически устойчива* при контакте с веществами (металлами и неметаллами), электродный потенциал которых лежит между двумя линиями E_H и E_O , соответствующими водородному и кислородному электродам. Выше

верхней линии вода разлагается с выделением кислорода, а ниже нижней линии — с выделением водорода.

В водных растворах, контактирующих с кислородом в газовой фазе, *термодинамически устойчивы металлы*, электродный потенциал E_M которых положительнее потенциала E_O кислородного электрода (лежит выше верхней линии на рис. 2.25), а также *неметаллы*, электродный потенциал E_X которых отрицательнее E_O (лежит ниже верхней линии на рис. 2.25). К первой группе относятся благородные металлы (Ag, Pd, Pt и Au) по отношению к нейтральным и щелочным средам ($7 < \text{pH} < 14$). Однако повышение кислотности среды ($0 < \text{pH} < 7$) приводит к растворению даже благородных металлов, за исключением только золота.

В водных растворах, контактирующих с водородом в газовой фазе, *термодинамически устойчивы металлы*, электродный потенциал E_M которых положительнее потенциала E_H водородного электрода (лежит выше нижней линии на рис. 2.25), а также *неметаллы*, электродный потенциал E_X которых отрицательнее E_H (лежит ниже нижней линии на рис. 2.25). Сюда относятся Te, Se и S, находящиеся в нейтральных и кислых средах ($\text{pH} \leq 8$).

В электролитическом режиме работы электрохимического элемента рассмотренные процессы управляют конкуренцией за участие в электродных процессах между металлами (неметаллами) и молекулами воды.

Как уже отмечалось, при электролизе *анодный процесс* всегда *окислительный* и при участии в нем металла анода сопровождается его растворением. Физически это объясняется тем, что положительный заряд на аноде делает его способным принимать электроны (то есть быть окислителем) и одновременно оттягивает вовнутрь собственные электроны металла, оголяя его ионную решетку на границе с раствором электролита. Последнее способствует взаимодействию ионов решетки с полярными молекулами воды и переходу катионов металла в раствор (*анодное растворение*).

Напротив, *катодный процесс* всегда *восстановительный* и сопровождается осаждением металла. Действительно, сообщение отрицательного заряда катоду делает его способным отдавать электроны (то есть быть восстановителем). При этом избыточный электронный заряд на поверхности металла экранирует его ионную решетку от взаимодействия с молекулами воды, тем самым подавляя растворение металла и способствуя выделению катионов из раствора (*катодное осаждение*).

При электролизе водных растворов солей катодное осаждение (по восстановительной реакции (2.19.4): $M^{z+} + ze^- \rightarrow M$) и анодное растворение (по окислительной реакции (2.19.3): $M - ze^- \rightarrow M^{z+}$) испытывают конкуренцию со стороны процесса электролитического разложения воды, протекающего:

а) на катоде по реакции восстановления воды (2.20.2) с выделением водорода за счет разряда водородных ионов ($H^+ + e^- \rightarrow H \rightarrow \frac{1}{2}H_2$), что характеризуется линией E_H на рис. 2.25;

б) на аноде по реакции окисления воды (2.20.3) с выделением кислорода за счет разряда гидроксильных ионов ($OH^- - e^- \rightarrow OH \rightarrow \frac{1}{2}H_2O + \frac{1}{4}O_2$), что характеризуется линией E_O на рис. 2.25.

На катоде в конкурентной борьбе за восстановление участвуют катионы M^{z+} и H^+ . Как известно (см. п. 2.19), металлы с отрицательным электродным потенциалом ($E_{M^{z+}}^\circ < 0$), измеренным по водородной шкале, склонны к окислению с переходом катионов из ионной решетки металла в раствор электролита, что не может быть обеспечено на катоде. Поэтому при электролизе водных растворов солей таких металлов, как щелочные и щелочноземельные металлы, Mg, Al, Zr и др. (электродный потенциал которых лежит ниже линии E_H на рис. 2.25), на катоде выделяется водород без осаждения этих металлов. Следовательно, для катодного осаждения указанных легких металлов надо использовать обезвоженные расплавы их солей.

Металлы, имеющие положительный электродный потенциал ($E_{M^{z+}}^\circ > 0$), проявляют тенденцию к восстановлению с переходом катионов из раствора на электрод. В первую очередь на катоде разряжаются катионы благородных металлов с большими значениями $E_{M^{z+}}^\circ$, так как для них реакция (2.19.7) имеет $\Delta G^\circ < 0$ с наибольшим значением $|\Delta G^\circ|$. По этой причине для таких металлов, как Au, Pt, Pd и Ag, энергетически выгоднее осаждение из водных растворов их солей без выделения водорода.

По мере уменьшения $E_{M^{z+}}^\circ$ (при движении сверху вниз от благородных металлов к химически активным на рис. 2.25) с приближением к линии E_H , соответствующей катодному разложению воды с выделением водорода, возрастает конкуренция со стороны этого процесса. Для металлов (от меди до марганца) с электродным потенциалом (положительным или отрицательным), лежащим недалеко от линии E_H , при электролизе водных растворов их солей одновременно происходит восстановление катионов металла и водорода. Чем больше значение $E_{M^{z+}}^\circ$ и выше концентрация электролита, тем вероятнее первый процесс

и больше выход осаждаемого металла. Выделение водорода при этом приводит к тому, что раствор электролита у катода приобретает более щелочной характер.

На аноде в конкурентной борьбе за окисление принимают участие атомы металла M , гидроксильные ионы OH^- и анионы солей или кислот. Растворение материала анода при электролизе возможно лишь при использовании в качестве анода достаточно активных металлов, таких как Zn , Fe , Ni и Cu , для которых величина $E_{M^{z+}}^\circ$ находится в середине ряда напряжений. Щелочные и щелочноземельные металлы всегда термодинамически неустойчивы, т. е. растворяются в растворах с любыми pH , и по этой причине неприменимы для использования в качестве анода.

Аноды из благородных металлов типа Au , Pt и Pd являются инертными, т. е. практически не растворяются при любых токах электролиза. На инертном аноде окисляются анионы электролита, как правило, с выделением соответствующего газа. В первую очередь разряжаются анионы X^{z-} с наименьшими значениями $E_{X^{z-}}^\circ$ (Te^{2-} , Se^{2-} , S^{2-}), имеющие наибольшую склонность к окислению с выделением вещества X , ибо для них реакция (2.19.8) имеет $\Delta G^\circ > 0$ с наибольшим значением ΔG° .

По мере увеличения $E_{X^{z-}}^\circ$ (при движении снизу вверх к химически активным неметаллам на рис. 2.25) с приближением к линии E_O , соответствующей анодному разложению воды с выделением кислорода, возрастает конкуренция со стороны этого процесса. Поэтому на аноде может оказаться термодинамически вероятнее процесс разряда ионов гидроксила OH^- , которому отвечает линия E_O для кислородного электрода. Так, при электролизе щелочей (NaOH , KOH), кислородосодержащих кислот (H_2SO_4 , HNO_3 , H_3PO_4) и соответствующих солей щелочных и щелочноземельных металлов единственно возможным процессом является разложение воды с выделением водорода на катоде и кислорода на аноде. Это объясняется тем, что кислородосодержащие ионы (SO_4^{2-} , NO_3^- , PO_4^{3-} и др.) имеют достаточно высокие значения $E_{X^{z-}}^\circ > 0$ (лежащие выше линии E_O на рис. 2.25), так что их разряд на аноде энергетически невыгоден по сравнению с разрядом гидроксильных ионов OH^- .

Иная картина наблюдается при электролизе соляной кислоты HCl и водных растворов ее солей (NaCl , NiCl_2 , CuCl_2 и др.): на катоде выделяется водород (для HCl и NaCl) или соответствующий металл (для NiCl_2 и CuCl_2), а на аноде — обычно газообразный хлор Cl_2 (при инертном аноде из Pt). Последнее противоречит данным рис. 2.25, так как стандартный электрод-

ный потенциал разряда иона хлора ($E_{\text{Cl}^-}^\circ = +1,36 \text{ В}$) превышает потенциал разряда иона OH^- , лежащий в интервале от $+1,23 \text{ В}$ при $\text{pH} = 0$ до $+0,401 \text{ В}$ при $\text{pH} = 14$ (см. линию E_{O} для кислородного электрода). Следовательно, по термодинамическим соображениям выделение хлора на аноде должно быть подавлено разложением воды с выделением кислорода. Однако на практике это не наблюдается.

Дело в том, что в дополнение к термодинамическим (равновесным) факторам существенное значение реально приобретают кинетические (неравновесные) факторы. Как отмечалось в п. 2.18, кинетические эффекты проявляются в форме возникновения поляризации электродов при протекании тока конечной величины. Такой ток вызывает отклонение концентрации ионов в растворе вблизи поверхности электродов и их электродных потенциалов от соответствующих равновесных значений. При этом катодная поляризация всегда отрицательная (потенциал электрода с током ниже равновесного), а анодная поляризация — положительная (потенциал электрода с током выше равновесного). Следовательно, в реальных условиях потенциал разряда ионов на электроде отличается от равновесного электродного потенциала на некоторую величину, называемую *электродным перенапряжением*, которая зависит от вида иона, материала электрода, состава раствора и силы тока.

Перенапряжения, влияющие на процессы выделения водорода и кислорода, называют соответственно *перенапряжением водорода* и *кислорода*. Перенапряжение водорода (вызванное отрицательной поляризацией катода) и перенапряжение кислорода (вызванное положительной поляризацией анода) приводят к тому, что в неравновесных условиях потенциалы разряда ионов гидроксония H_3O^+ и гидроксила OH^- лежат, соответственно, ниже равновесной линии E_{H} для водородного электрода и выше равновесной линии E_{O} для кислородного электрода, приведенных на рис. 2.25. Перенапряжения водорода и кислорода в зависимости от материала электрода, состава раствора, температуры и силы тока могут достигать десятых долей вольта (вплоть до 1 В).

Возвращаясь к электролизу с участием хлора, следует учесть перенапряжение кислорода на аноде, повышающее потенциал выделения кислорода с $E_{\text{O}} = +0,814 \text{ В}$ при $\text{pH} = 7$ до значения $1,5 \text{ В}$. Это превышает потенциал разряда ионов хлора ($E_{\text{Cl}^-}^\circ = +1,36 \text{ В}$) и способствует преимущественному выделению хлора перед кислородом. Однако в разбавленных растворах

снижение концентрации (активности) ионов хлора увеличивает, согласно (2.19.17), величину $E_{\text{Cl}} = E_{\text{Cl}}^{\circ} - (RT/\mathcal{F}) \ln a_{\text{Cl}^-} = +1,36 - 0,059 \lg a_{\text{Cl}^-}$ (где численные значения соответствуют $T = 298 \text{ K}$). Значение ее может превысить 1,5 В, так что выделение хлора сменится выделением кислорода.

Перенапряжение водорода имеет принципиальное значение для катодного осаждения таких металлов, как Pb, Sn, Ni, Fe, Zn, Сг и др., которые стоят выше водорода в ряду напряжений. Их осаждение может быть объяснено только различием в величине перенапряжения для металлов и водорода. Для этих металлов значения $E_{\text{M}^{z+}}^{\circ}$ лежат при определенных рН ниже равновесной линии E_{H} на рис. 2.25. Следовательно, по термодинамическому критерию разряд ионов H^+ с выделением водорода должен преобладать над разрядом ионов M^{z+} и подавлять осаждение металла на катоде. Однако в реальных условиях электролиза перенапряжение металла величиною в сотые доли вольта значительно меньше перенапряжения водорода, что опускает вниз линию E_{H} на рис. 2.25 в неравновесных условиях. Тогда между потенциалами разряда H^+ и M^{z+} возникает обратное соотношение, соответствующее преимущественному осаждению металла по сравнению с выделением водорода.

В конкурентной борьбе за участие в электродных процессах при электролизе преимущество остается за теми реакциями, которые при данном токе обеспечивают минимальное значение $\Delta E + E_{\text{пол}}$, где ΔE — ЭДС гальванической пары, возникающей при выделении на электродах продуктов электролиза, $E_{\text{пол}}$ — суммарная поляризация катода и анода (см. п. 2.18). В случае, когда материал анода совпадает с осадком на катоде (например, медный анод в растворе CuSO_4 или CuCl_2), гальванической пары не образуется, т. е. $\Delta E = 0$. В результате этого преимущественно протекает процесс растворения анода с выделением материала его на катоде (так называемый электролиз с растворимым анодом, применяемый на практике для рафинирования металлов).

Помимо рассмотренных первичных реакций при электролизе возможно протекание вторичных (совмещенных) химических реакций с участием продуктов электролиза. Совмещенные реакции существенно усложняют физическую картину электродных процессов. При электролизе водных растворов электролитов химически активные газы (водород и кислород), выделяющиеся в результате разложения воды, могут приводить к образованию гидридов на катоде (где выделяется H_2) и оксидов на аноде (где выделяется O_2). Так как обычно гидриды легколетучие, то вторичная реакция их образования может быть использована для

катодного травления полупроводников, например, для травления Ge или Si в кислых растворах с катодным выделением газообразных продуктов — моногермана GeH_4 или моносилана SiH_4 .

Наоборот, окислы, возникающие как вторичный продукт в анодном процессе, обычно являются достаточно химически стойкими и надежно защищают (пассивируют) поверхность металла, делая его малочувствительным к внешним воздействиям. Примером является алюминий, который всегда защищен от коррозии тонким слоем окисла Al_2O_3 .

2.22. Контрольные вопросы

1. В чем заключается специфика фазовых превращений первого и второго рода в однокомпонентных системах?
2. Какая термодинамическая функция определяет температурный ход кривых сублимации и испарения на фазовой диаграмме, качественно показанной на рис. 2.1?
3. Что такое насыщенный пар чистой конденсированной фазы и как управлять давлением этого пара?
4. Дайте термодинамическую оценку вклада энтальпийного и энтропийного факторов при образовании жидких и твердых растворов.
5. В чем состоит специфика растворения газов в конструкционных и газопоглощающих материалах?
6. Какую роль играет активность компонентов в термодинамической теории неидеальных растворов?
7. Каковы основные термодинамические свойства растворов в рамках моделей идеального, атермального и регулярного раствора?
8. Сформулируйте закон Рауля для идеальных растворов и обоснуйте отклонения от него в применении к реальным растворам.
9. Как изменяется температура кипения жидкости при введении в нее растворимой или нерастворимой жидкой добавки?
10. Сформулируйте законы Генри и Сивертса для молекулярного и атомарного механизмов растворения. В каких материалах преимущественно реализуются эти механизмы?

11. Какая термодинамическая функция определяет температурную зависимость молекулярной и атомарной растворимости газов в конденсированных средах?
12. Чем определяется наклон кривой плавления на фазовой диаграмме, качественно показанной на рис. 2.1?
13. В чем состоит принципиальное различие между температурными зависимостями для термоактивационных процессов и для равновесных процессов испарения и растворения?
14. Поясните роль закона Рауля в обосновании типа диаграммы плавкости для бинарных систем с неограниченным растворением компонентов в жидкой фазе.
15. Каково назначение линий ликвидуса и солидуса на бинарных диаграммах плавкости?
16. Что такое фигуративная точка, коннода и правило рычага на диаграммах плавкости?
17. Как используется правило фаз Гиббса для обоснования хода кривой охлаждения при равновесной кристаллизации из бинарного жидкого раствора?
18. Какой тип диаграммы плавкости характерен для бинарных систем в отсутствие взаимной растворимости компонентов в твердой фазе?
19. Какие типы диаграммы плавкости характерны для бинарных систем с неограниченным твердым раствором?
20. Как видоизменяет диаграмму плавкости ограниченная растворимость компонентов в твердой фазе?
21. Какие диаграммы плавкости принято называть диаграммами эвтектического типа и перитектического типа?
22. Чем отличаются процессы равновесной кристаллизации в бинарных системах с диаграммами плавкости эвтектического и перитектического типов?
23. В чем состоит принципиальное видоизменение диаграммы плавкости при наличии химического соединения в твердой фазе?
24. Что такое химическое соединение переменного состава и как определена его область гомогенности на диаграмме плавкости?
25. Какие диаграммы плавкости принято называть дистектическими диаграммами?

26. Чем отличаются процессы кристаллизации химических соединений с конгруэнтным и инконгруэнтным плавлением?
27. Поясните физическое содержание закона действующих масс для обратимых химических реакций.
28. Запишите условия химического равновесия для реакций, приведенных в п. 2.16 и ниже в п. 4.13: а) в общей форме (через химические потенциалы) и б) в форме закона действующих масс. Почему в последнем случае исчезает вклад чистых конденсированных фаз?
29. Дайте общее определение изобарного потенциала химической реакции для реальных и стандартных условий.
30. Какая связь существует между стандартным изобарным потенциалом ΔG_T° реакции и константой химического равновесия $K_p(T)$?
31. Какова методика расчета температурной зависимости константы химического равновесия $K_p(T)$ на основе таблиц стандартных термодинамических величин?
32. Поясните физическое содержание уравнения изотермы Вант-Гоффа.
33. Каким образом обеспечивается термодинамическое подавление нежелательных конкурирующих реакций?
34. Каков физический смысл уравнения изобары и уравнения изохоры Вант-Гоффа?
35. Какие термодинамические функции определяют температурную зависимость констант химического равновесия $K_p(T)$ и $K_c(T)$?
36. Чем отличаются эндотермические и экзотермические химические реакции?
37. Каким образом изменение температуры и давления смещает химическое равновесие?
38. Как влияет изменение состава системы на протекание химической реакции?
39. Может ли введение инертного газа повлиять на газофазную химическую реакцию?
40. Сформулируйте в общей форме принцип Ле Шателье и приведите конкретные примеры его проявления.

41. Что такое газотранспортные химические реакции и какова область их практического применения?
42. Чем определяется направление химического транспорта твердого вещества из одной температурной зоны в другую?
43. Как определены ионное произведение воды и водородный показатель рН водного раствора электролитов?
44. Каким образом управлять смещением ионного равновесия в водных растворах электролитов?
45. В чем состоит принципиальное различие между химическими и электрохимическими реакциями?
46. В каком направлении протекает окислительно-восстановительная реакция в электрохимическом элементе при работе в гальваническом и электролитическом режимах?
47. Обоснуйте соотношение между обратимой электродвижущей силой электрохимического элемента и изобарным потенциалом химической реакции, лежащей в основе работы этого элемента.
48. Как связано уравнение Нернста для обратимой ЭДС электрохимического элемента с уравнением изотермы Вант-Гоффа для изобарного потенциала химической реакции?
49. Поясните с общих позиций, каким образом протекают процессы окисления/восстановления и растворения/осаждения материалов на электродах электрохимического элемента.
50. Какова роль электрохимического разложения воды с выделением водорода и кислорода в электродных процессах?

Глава 3

УПРАВЛЕНИЕ СТРУКТУРНЫМИ РАВНОВЕСИЯМИ И ДЕФЕКТООБРАЗОВАНИЕМ В КРИСТАЛЛАХ

Любые нарушения периодичности кристаллической решетки принято называть *дефектами структуры*. Некоторые физические свойства твердых тел, такие как плотность, упругость, решеточная теплоемкость и диэлектрическая проницаемость, малочувствительны к несовершенству кристаллической решетки, то есть являются *дефекто-нечувствительными*. Они определяются химическим составом вещества и симметрией сил связи между частицами, что адекватно описывается в рамках модели *идеального кристалла*. Поэтому для анализа таких свойств важно знать кристаллическую структуру твердого тела, исходя из кристаллохимических и кристаллографических особенностей пространственной решетки идеального кристалла.

Однако в подавляющем большинстве случаев физические свойства твердых тел (электрические, магнитные, оптические, механические, тепловые) существенным образом зависят от степени несовершенства кристалла, то есть являются *дефекто-чувствительными*. Реальные твердые тела обычно неидеальны, поскольку всегда имеются нарушения периодической структуры: а) изменения расстояний до ближайших соседей (за счет тепловых колебаний решетки); б) отсутствия атомов/ионов в узле решетки или замены их инородной частицей; в) смещения атома/иона из узла в междоузлие или размещения там инородной частицы; г) местного нарушения структуры в результате облучения кристалла частицами высоких энергий и т. п.

Анализ дефектообразования и влияния дефектов на физические свойства материалов является одной из ключевых проблем физики твердого тела. Особенно актуальна эта проблема

в применении к физике полупроводников, поскольку свойства полупроводниковых кристаллов в наибольшей степени, по сравнению с другими материалами, изменяются в зависимости от концентрации и типа дефектов.

Задача данной главы — изучение особенностей кристаллической структуры и физических закономерностей, определяющих поведение точечных дефектов в кристаллах, и управление ими для целенаправленного изменения электрофизических свойств полупроводниковых материалов путем их легирования, то есть искусственного создания точечных дефектов в кристаллах.

3.1. Химические связи в твердых телах

В неизменных внешних условиях твердые тела (в отличие от жидкостей) сохраняют не только свой объем, но и форму. Это связано с наличием как *ближнего порядка* в расположении частиц (подобно жидкостям), так и *дальнего порядка*, вызванного пространственной симметрией в силовом периодическом взаимодействии между частицами, формирующем кристаллическую структуру. Твердые вещества существуют в трех разновидностях:

- *монокристаллы* с распространением дальнего порядка на весь объем вещества (о квазикристаллах см. п. 5.2);
- *поликристаллы* с сохранением дальнего порядка только в ограниченных объемах (называемых кристаллитами с размерами в десятки и сотни тысяч периодов решетки), представляющие собой совокупность беспорядочно ориентированных кристаллитов;
- *аморфные вещества* (неупорядоченные структуры) с сохранением только ближнего порядка в расположении частиц.

Дальняя неупорядоченность аморфных веществ вызвана, во-первых, большой протяженностью молекул (полимеры, пластмассы, смолы, резины, органические стекла и т. п.) и, во-вторых, достаточно быстрым охлаждением жидкости (плавленый кварц, аморфный углерод, неорганические стекла и т. п.). В последнем случае возникает лишь ближний порядок в расположении частиц, но не успевает образоваться кристаллическая решетка. Поэтому аморфное состояние не является термодинамически устойчивым. Со временем (быстрее или медленнее, в зависимости от температуры) происходит самопроизвольный процесс кристаллизации с выделением энергии по причине стремления системы к минимуму свободной энергии.

Отличительной особенностью кристаллических тел является, во-первых, резко выраженный переход в жидкое состояние при

температуре плавления и, во-вторых, наличие анизотропии в виде зависимости физических свойств от направления.

Первая особенность характерна для монокристаллов и поликристаллов. Это отличает их от аморфных тел, не имеющих температуры плавления и характеризующихся интервалом размягчения, в пределах которого с возрастанием температуры увеличивается текучесть за счет постепенного уменьшения вязкости.

Анизотропия присуща лишь монокристаллам, что отличает их от поликристаллов и аморфных тел, свойства которых изотропны, т. е. одинаковы по всем направлениям. Следует отметить, что существуют органические вещества, называемые *жидкими кристаллами*, которые обладают одновременно свойствами текучести (как у жидкостей) и анизотропии (как у кристаллов, но с дальним порядком в одном или двух направлениях).

Огромное разнообразие кристаллических веществ определяется их кристаллографической структурой, химическим составом, размерами атомов, ионов или молекул и природой сил, действующих между ними. Начнем изложение с краткого рассмотрения кристаллохимических и кристаллографических элементов теории идеальных (бездефектных) кристаллов.

Кристаллические тела принято классифицировать двояко:

а) по химическому признаку, то есть по типу химической связи между частицами;

б) по геометрическому признаку, то есть по характеру симметрии кристаллической решетки.

По характеру химической связи между частицами (молекулами, атомами, ионами), образующими кристаллическую структуру, принято различать четыре типа кристаллов: молекулярные, металлические, ионные и ковалентные (в том числе полярные) кристаллы.

При сближении частиц между ними возникает силовое взаимодействие, характер которого зависит от атомной структуры и строения электронных оболочек. Во взаимодействии участвуют в первую очередь валентные электроны внешних электронных оболочек, в то время как внутренние оболочки атомов практически не оказывают влияния. Характер межатомного взаимодействия в значительной степени определяется способностью атомов отдавать или принимать электроны. Эту способность принято характеризовать, соответственно, потенциалом ионизации (I) и сродством к электрону (E).

Потенциалом ионизации характеризуют энергию, затрачиваемую на отрыв наименее прочно удерживаемого валентного

электрона от нейтрального газообразного атома, находящегося в основном состоянии, т. е. энергию процесса ионизации (2.19.3): $M - e^- \rightarrow M^+$. Наименьший потенциал ионизации ($I \approx 4-5$ эВ) имеют щелочные металлы Cs, Rb, K, Na, Li, у которых всего один валентный электрон, а наибольший ($I \approx 12-25$ эВ) — инертные газы Xe, Kr, Ar, Ne, He. У них внешняя электронная оболочка имеет полностью застроенную форму «октета» ns^2np^6 ¹⁾; исключение составляет атом гелия, для которого $n = 1$ и $l = n - 1 = 0$, так что его электронная конфигурация имеет вид $1s^2$.

Сродством к электрону характеризуют энергию, выделяемую при захвате электрона нейтральным газообразным атомом, находящимся в состоянии с наименьшей энергией, т. е. энергию процесса типа (2.19.8): $X + e^- \rightarrow X^-$. Наименьшее сродство к электрону ($E < 1$ эВ) имеют щелочные металлы Cs, Rb, K, Na, Li, а наибольшее ($E \approx 3-4$ эВ) — галогены At, I, Br, Cl, F.

Обе величины (как I , так и E) указывают на большую или меньшую способность атома переходить в ионное состояние. Для качественного суждения о распределении электронной плотности между взаимодействующими атомами используют так называемую *электроотрицательность* x , определенную в простейшем виде как $x = (I + E)/2$. В литературе встречаются и другие, более сложные, определения x , но во всех случаях наименьшую электроотрицательность имеют щелочные металлы (обладающие малыми как I , так и E), а наибольшую — галогены (обладающие большими как I , так и E); в частности, по Полингу (в эВ):

$$x_{\text{Cs}} = 1,75, \quad x_{\text{Rb}} = 2,0, \quad x_{\text{K}} = 2,0, \quad x_{\text{Na}} = 2,25, \quad x_{\text{Li}} = 2,5; \\ x_{\text{At}} = 5,5, \quad x_{\text{I}} = 6,25, \quad x_{\text{Br}} = 7,0, \quad x_{\text{Cl}} = 7,5, \quad x_{\text{F}} = 10,0.$$

Следовательно, с ослаблением металлических свойств элементов их электроотрицательность всегда возрастает. Парной комбинацией элементов с различным соотношением значений x можно качественно объяснить появление ионной, ковалентной, полярной ковалентной и металлической связи между ними.

Ионная связь в чистой форме (или в преимущественно ионной форме) возникает между элементами, стоящими на противоположных краях ряда электроотрицательностей (с одного края

¹⁾ Напоминаем, что в общепринятых обозначениях n означает главное квантовое число, нумерующее электронные оболочки $K, L, M, \dots$ (He, Ne, Ar, Kr, Xe соответствуют значения $n = 1, 2, 3, 4, 5$). Буквами s и p обозначены подоболочки с орбитальным (азимутальным) квантовым числом $l = 0$ и $l = 1$ (при $l_{\text{max}} = n - 1$). Максимальное заполнение подоболочки электронами с учетом спинового вырождения равно $2(2l + 1)$, т. е. 2 и 6 при $l = 0$ и $l = 1$, что указано верхними цифрами (s^2 и p^6).

щелочные и щелочноземельные металлы, а с другого — галогены и халькогены). При этом максимум электронной плотности между атомами предельно смещен в сторону более электроотрицательного элемента (с большим значением x) с превращением атомов в противоположно заряженные ионы, удерживаемые кулоновским притяжением. Силы притяжения ионных зарядов уравниваются силами отталкивания полностью заполненных внешних электронных оболочек ионов.

Для щелочно-галогидных кристаллов (типа NaCl и CsCl) такие оболочки имеют электронную конфигурацию инертных газов в форме заряженного «ионного октета» ns^2np^6 (см. сноску на с. 336). Сферическая симметрия таких конфигураций делает ионную связь ненаправленной и склонной к достаточно плотной упаковке, имеющей место, например, в кристаллах кубической сингонии (см. ниже). В этом случае вводят в рассмотрение *ионные радиусы* катиона и аниона (так, по Полингу $r_{\text{Na}^+} = 0,98 \text{ \AA}$, $r_{\text{Cs}^+} = 1,65 \text{ \AA}$ и $r_{\text{Cl}^-} = 1,81 \text{ \AA}$), сумма которых равна межъядерному расстоянию, а отношение — отношению эффективных зарядов ядер (с учетом их экранировки внутренними электронами).

В структуры ионного типа кристаллизуются галогениды щелочных металлов, халькогениды щелочноземельных металлов и большая часть простых (бинарных) оксидов. Ионные кристаллы представляют собой ярко выраженные диэлектрики с энергией связи, достигающей 150–200 ккал/моль (напоминаем, что $1 \text{ эВ} \approx 23 \text{ ккал/моль}$). Поэтому они имеют достаточно высокие температуры плавления и теплоты плавления (сублимации).

С уменьшением разности электроотрицательностей взаимодействующих атомов (т. е. по мере их приближения с двух краев ряда электроотрицательности к его середине), начиная уже с перехода от галогенидов $\text{A}^{\text{I}}\text{B}^{\text{VII}}$ к халькогенидам $\text{A}^{\text{II}}\text{B}^{\text{VI}}$, начинает проявляться частичный вклад пространственно направленного характера ковалентной связи.

Ковалентная связь в чистой форме (или в преимущественно ковалентной форме) возникает между атомами, имеющими равные (или близкие) электроотрицательности. Физическая природа этой связи проявляется в сугубо квантовой форме так называемого *обменного взаимодействия* между атомами с одинаковыми или примерно одинаковыми электронными конфигурациями внешних валентных оболочек.

На языке классического описания обменное взаимодействие представляют как попарное обобществление валентных электронов с противоположными спинами, изначально принадлежащих

различным атомам (простейший пример — молекула водорода H_2). При этом обобществленная пара электронов как бы принадлежит каждому из связанных атомов, что позволяет ему достигнуть электронной конфигурации инертного газа в форме незапряженного «ковалентного октета» ns^2np^6 (см. сноску на с. 336). Следовательно, ковалентная связь между атомами организована по правилу « $8 - N$ », где N — число валентных электронов атома, равное номеру группы периодической таблицы элементов.

Чисто ковалентная связь существует в двухатомных молекулах, состоящих из одинаковых атомов, для которых $4 \leq N \leq 7$ в соответствии с правилом « $8 - N$ ». Действительно, атому галогена (VII группа) недостает одного электрона до полного октета из 8 электронов. Обобществление одной электронной пары дает полностью насыщенную ковалентную связь, когда отсутствуют электроны, которые могли бы взаимодействовать с другими атомами. Поэтому во всех агрегатных состояниях галогены состоят из двухатомных молекул (например, F_2 , Cl_2 и т. п.), которые в конденсированном состоянии связываются между собой за счет межмолекулярных сил (см. ниже). Аналогичным образом ведут себя атомы VI и V групп, обобществляющие в двухатомных молекулах две и три электронные пары, соответственно.

Элементам IV группы периодической таблицы (C, Si, Ge, Sn) недостает четырех электронов до полного октета. Для образования одинарной ковалентной связи атом должен иметь состояние с одним неспаренным электроном, в то время как для установления четырех связей в решетке алмаза (см. ниже) атому углерода необходимы четыре электрона. Однако для атома C электронная конфигурация $1s^2 2s^2 2p^2$ с двумя неспаренными электронами (см. сноску на с. 336) позволяет сформировать только две ковалентные связи. Вступая во взаимодействие, атом углерода возбуждается с переводом одного электрона из состояния $2s$ в менее устойчивое состояние $2p$, создавая электронную конфигурацию $1s^2 2s^2 2p^3$, имеющую теперь четыре необходимых валентных электрона. Для каждого атома C (или Si и Ge) обобществление этих электронов в ковалентной связи с четырьмя соседними атомами, расположенными по углам правильного тетраэдра, создает кристаллическую структуру типа алмаза (см. рис. 3.9 а).

Ковалентная связь, в отличие от ионной, характеризуется валентной насыщенностью и строгой пространственной направленностью (от центрального атома к угловым атомам тетраэдра). Так как распределение электронной плотности (создаваемое попарно обобществленными электронами) симметрично относительно се-

редины межатомного (точнее, межъядерного) расстояния, то атомам приписывают *ковалентные радиусы*, равные половине такого расстояния ($r_C = 0,77 \text{ \AA}$, $r_{Si} = 1,17 \text{ \AA}$, $r_{Ge} = 1,22 \text{ \AA}$).

Разновидностью ковалентной связи является так называемая *донорно-акцепторная* (иначе *координационная*) *связь*. В этом случае общая электронная пара создается не обобществлением двух электронов, взятых от разных атомов, а предоставляется одним атомом (донором) другому атому (акцептору), обладающему незанятыми электронными орбитами. Разные механизмы образования ковалентной и донорно-акцепторной связей приводят к одинаковому результату и часто реализуются одновременно в химических соединениях типа $A^{II}B^{VI}$ и $A^{III}B^V$. Например, в кристалле сульфида цинка имеются две ковалентные связи и две донорно-акцепторные, при этом Zn — акцептор, а S — донор.

Полярная ковалентная связь возникает между различными атомами с неравными, но близкими электроотрицательностями. Тогда центр электронного облака смещен от середины межатомного расстояния к более электроотрицательному атому. Такое смещение называют степенью полярности (ионности) связи, которая пропорциональна разности электроотрицательностей. Так, для атомов Si и C (IV группа) электроотрицательности равны $x_{Si} = 4,5 \text{ эВ}$ и $x_C = 6,25 \text{ эВ}$. Это свидетельствует о полярной связи между ними со смещением электронного заряда к атому углерода в тетраэдрической структуре карбида кремния SiC. Длину связи оценивают как сумму ковалентных радиусов: для SiC имеем $r_{Si} + r_C = 1,94 \text{ \AA}$, что совпадает с результатами рентгеноструктурных измерений. Это подтверждает широко используемую модель взаимно касающихся атомных (ионных) шаров.

Описанная модель характерна для структур с достаточно плотной и плотнейшей упаковками (например, в гексагональной решетке и в гранецентрированной или объемноцентрированной кубических решетках). Тетраэдрические структуры ковалентных кристаллов (типа алмаза) образуются комбинацией двух решеток с плотнейшей упаковкой (см. ниже).

Ковалентные кристаллы обычно проявляют полупроводниковые свойства и имеют достаточно большую энергию связи, достигающую 100 ккал/моль. В большинстве своем, они имеют высокие температуры плавления и теплоты плавления (сублимации) и отличаются большой твердостью, особенно алмаз.

Как ясно из вышесказанного, ионная связь возникает только между разными атомами, поэтому она относится к *гетерополярным связям*. В отличие от этого, ковалентная связь может

быть как между различными элементами (типа химических соединений $A^{III}B^V$ и $A^{II}B^{VI}$), так и между одинаковыми атомами (типа C, Si, Ge), образуя кристаллы с *гомеополярной связью*, к которой относятся также и металлические кристаллы.

Металлическая связь образована электрическим взаимодействием между положительно заряженными ионами и подвижными электронами, возникшими в результате потери атомами своих валентных электронов. К таким атомам относятся те, у которых недостаточно валентных электронов для связи либо электронными парами (по правилу « $8 - N$ »), либо отдельными электронами, т. е. атомы первых групп периодической таблицы ($N < 4$). В этом случае связь обеспечивается валентными электронами, принадлежащими всей совокупности атомов. Они обобществлены в единый коллектив, образующий квазисвободный электронный газ, удерживающий внутри себя решетку положительных ионов.

Более сложная ситуация возникает в металлических сплавах (построенных на основе твердых растворов — см. п. 2.3) и интерметаллических соединениях (построенных с нарушением правил валентности — см. п. 2.11). Но во всех случаях металлическая связь очень прочная (с энергией до 150 ккал/моль), что обеспечивает высокие механические свойства металлов.

В силу сферической симметрии ионов в узлах решетки металлическая связь не является направленной. Поэтому подавляющее большинство металлов кристаллизуется в наиболее плотные упаковки гексагональной и кубической сингонии (см. ниже).

Молекулярная связь образована диполь–дипольными силами Ван-дер-Ваальса и силами водородной связи между валентно насыщенными электронными конфигурациями молекул и атомов инертных газов. Насыщенность внешних электронных оболочек приводит к тому, что валентные силы химической природы не играют никакой роли в образовании таких кристаллов. В этом случае принципиально важными становятся эффективные дипольные моменты, вызванные пространственной асимметрией в распределении зарядов внутри полярной молекулы или наведенные внутри неполярной молекулы (с симметричным распределением зарядов) электрическим полем другого диполя.

Диполь–дипольные взаимодействия принято делить на три группы: а) *ориентационное* взаимодействие между полярными молекулами с постоянными дипольными моментами; б) *индукционное* взаимодействие между полярной и неполярной молекулами с постоянным и индуцированным дипольными моментами; в) *дисперсионное* взаимодействие между неполярными молеку-

лами, вызванное квантовой синхронизацией мгновенных диполей взаимодействующих частиц (см. п. 5.9).

Дисперсионная связь в чистом виде реализуется только при кристаллизации инертных газов, атомы которых, имея валентно насыщенную и сферически симметричную оболочку, обладают нулевым сродством к электрону и нулевым дипольным моментом. При кристаллизации двухатомных и многоатомных молекул (типа N_2 , O_2 , CH_4 и т. п.) к дисперсионным силам добавляются ориентационные и индукционные силы Ван-дер-Ваальса.

Водородная связь встречается в молекулярных кристаллах, имеющих атом водорода в составе молекулы (типа HF , HCl , H_2O и т. п.), связанный с сильно электроотрицательным атомом (типа F , Cl , O и т. п.). В этом случае атом H , отдавая свой единственный электрон, превращается в протон H^+ с размерами, в тысячи раз меньшими размеров других атомов. Это позволяет ему приближаться к таким атомам, взаимодействуя с их электронами, и даже внедряться в их электронную оболочку. Именно так ведет себя H^+ в воде, образуя с молекулой H_2O ион гидроксония H_3O^+ (см. п. 2.17). Водородная связь играет также существенную роль в структуре льда, обеспечивая взаимодействие между атомами кислорода через водородный атом.

Таким образом, водородная связь образуется как бы второй валентностью атома водорода, которую он проявляет к сильно электроотрицательным атомам, если основная валентность связывает его с таким же электроотрицательным атомом в молекуле. Водородная связь свойственна любым агрегатным состояниям вещества. Она может существовать между как одинаковыми, так и различными молекулами, в том числе между разными частями одной и той же молекулы. Наиболее распространенной является водородная связь между молекулами, содержащими гидроксильные группы OH (см. п. 5.9).

Атомы инертных газов, обладая сферической формой и сферически симметричными (ненаправленными) связями, при кристаллизации образуют гранецентрированную кубическую решетку, обладающую плотнейшей упаковкой (см. ниже). Для молекулярных кристаллов с двухатомными и многоатомными молекулами реализуются более сложные решетки.

В молекулярных кристаллах энергия межмолекулярной связи мала (порядка 1–5 ккал/моль). Поэтому они обладают малой твердостью, низкими температурами плавления и теплотами сублимации, большой сжимаемостью и значительной летучестью. Молекулярные кристаллы являются диэлектриками и прозрачны для электромагнитных волн вплоть до дальнего ультрафиолета.

3.2. Классификация пространственных решеток кристаллов

Характерным признаком всех без исключения кристаллов является так называемая *трансляционная симметрия*. Это означает следующее: при определенном выборе в произвольной точке трех некопланарных отрезков можно путем их параллельного переноса (*трансляции*) плотно заполнить все бесконечное пространство периодической совокупностью одинаковых параллелепипедов. Точки, связанные трансляциями, образуют трехмерную систему *эквивалентных узлов*, называемую пространственной решеткой, базирующейся на *элементарной ячейке* в форме исходного параллелепипеда.

Таким образом, *пространственная решетка* есть результат *геометрического построения*, позволяющего нам представить периодичность в пространственном расположении материальных частиц (или групп частиц) и помогающего выявить законы симметрии кристаллической структуры в виде набора ее симметричных преобразований. В отличие от решетки, *кристаллическая структура* является *физической реальностью* в форме расположения в пространстве материальных частиц (атомов, ионов, молекул), входящих в состав конкретного кристалла.

Принципиальное различие между структурой кристалла и его кристаллической решеткой состоит в том, что узел решетки не должен обязательно совпадать с центром материальной частицы (хотя именно так чаще и делается на практике), а число частиц может превышать число узлов в элементарной ячейке при структуре кристалла, составленной из нескольких решеток.

Классы симметрии, пространственные группы, сингонии и решетки Браве. Выбор элементарной ячейки, построенной на основной тройке трехмерных трансляций, в принципе неоднозначен. Поэтому принято выбирать ячейку таким образом, чтобы она а) обладала наименьшим объемом (кратчайшие трансляции), б) имела прямые углы и равные стороны (если это возможно), в) отражала симметрию кристалла максимальным образом.

Симметрия кристалла подразумевает инвариантность его формы по отношению к той или иной совокупности геометрических преобразований, называемой *точечной группой симметрии*. Результатом таких преобразований является совмещение кристалла как геометрической фигуры с самим собой, при этом остается неподвижной по крайней мере одна точка фигуры. Простейшими элементами точечной группы симметрии являются:

- *плоскость симметрии* (совмещение фигуры путем ее зеркального отражения в плоскости);
- *ось симметрии* n -го порядка (совмещение фигуры путем ее поворота вокруг оси на угол $2\pi/n$, $n = 1, 2, 3, 4, 6$);
- *центр симметрии* (совмещение фигуры путем ее инверсии относительно некоторой точки).

Не вдаваясь в детали теории симметрии, отметим лишь, что полная симметрия любого геометрического объекта (кристаллического многогранника) может быть получена комбинацией плоскостей, осей и центров симметрии. Групповой анализ показывает, что число таких комбинаций ограничено 32 точечными группами симметрии, которые называют *классами симметрии*. Внешняя симметрия кристалла, как геометрического объекта, находится в прямой зависимости от его внутренней структуры и, следовательно, определяется дискретным расположением материальных частиц в соответствии с типом пространственной решетки.

Включение в рассмотрение пространственной решетки требует добавления к трем основным элементам точечной группы (плоскость симметрии, ось симметрии, центр симметрии) четвертого элемента симметрии — *пространственной трансляции*, которая действует не на одну ячейку, а на всю решетку в целом. Комбинация пространственной трансляции с элементами точечной симметрии дает два новых элемента симметрии — *винтовую ось* (поворот вокруг оси с параллельным переносом вдоль нее) и *плоскость скольжения* (зеркальное отражение в плоскости с параллельным переносом вдоль нее).

Анализ возможных случаев симметрии решеток показывает, что из шести вышеупомянутых операций симметрии (отражение, поворот, инверсия, трансляция, винтовое вращение, скользящее отражение) можно образовать 230 комбинаций, называемых ***пространственными группами симметрии***, которые определенным образом распределены по 32 классам точечной симметрии.

Указанный выше рецепт выбора элементарной ячейки позволяет построить так называемую *примитивную ячейку*, содержащую один узел решетки. Ячейки с большим числом узлов считают *сложными*. При подсчете узлов (u) в ячейке следует руководствоваться правилом: каждый узел вносит свою часть, равную

- $1/8$ — для углового узла, принадлежащего восьми ячейкам;
- $1/4$ — для узла на ребре, принадлежащем четырем ячейкам;
- $1/2$ — для узла в центре грани, принадлежащей двум ячейкам;
- 1 — для узла в центре данной ячейки.

Самой общей формой примитивной ячейки является параллелепипед с восемью узлами, расположенными в его вершинах, так что $y = 1/8 \times 8 = 1$. Среди сложных ячеек принято различать:

- а) **объемноцентрированные ячейки** — с одним дополнительным узлом в центре ячейки, так что $y = 1/8 \times 8 + 1 = 2$;
 б) **базоцентрированные ячейки** — с дополнительными узлами в центре нижнего и верхнего оснований, так что $y = 1 + 1/2 \times 2 = 2$;
 в) **гранецентрированные ячейки** — с дополнительными узлами в центре всех шести граней, так что $y = 1/8 \times 8 + 1/2 \times 6 = 4$;

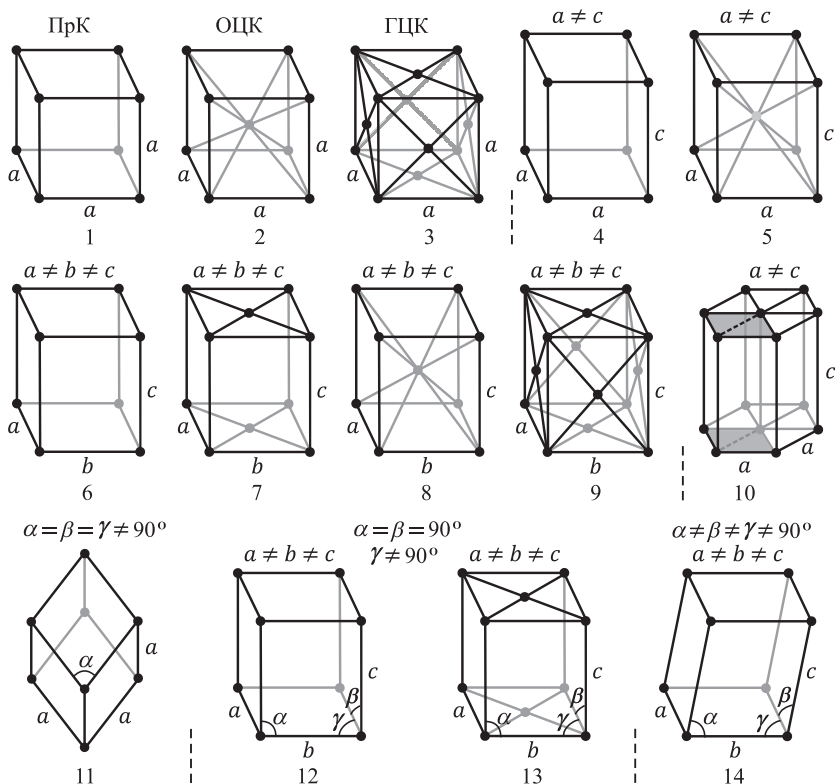


Рис. 3.1. Распределение решеток Браве по сингониям: 1–3 — решетки *кубической сингонии* (ПрК — примитивная, ОЦК — объемноцентрированная, ГЦК — гранецентрированная); 4–5 — решетки *тетрагональной сингонии* (примитивная, базоцентрированная, объемноцентрированная, гранецентрированная); 6–9 — решетки *ромбической сингонии* (примитивная, базоцентрированная, объемноцентрированная, гранецентрированная); 10 — решетка *гексагональной сингонии*; 11 — решетка *ромбоэдрической сингонии*; 12–13 — решетки *моноклинной сингонии* (примитивная, базоцентрированная); 14 — решетка *триклинной сингонии*. Сингонии отделены друг от друга вертикальными штриховыми линиями, проведенными между номерами

Ячейки, которые имеют те же элементы точечной симметрии, что и решетка в целом, называют **ячейками Браве**. Таких ячеек оказалось всего 14, из которых примитивными являются семь.

По виду примитивной ячейки Браве классифицируют семь типов так называемых кристаллических систем, или **сингоний** (от греч. «сходноугольность»). В одну сингонию объединяются примитивная ячейка Браве (всегда) и сложные ячейки (если есть таковые), у которых одинаковы как элементы точечной симметрии, так и кристаллографическая система координат.

Геометрическая форма 14 ячеек Браве (примитивных и сложных) и распределение их по 7 сингониям изображены на рис. 3.1. Свойства примитивных ячеек (углы между ребрами и соотношение их длин) с указанием формы ячеек Браве, их числа и количества точечных групп (классов симметрии) и пространственных групп в каждой сингонии приведены в табл. 3.1.

Кристаллографические оси координат, будучи обычно направленными вдоль ребер примитивной ячейки, однозначно предопределены типом сингонии. Примитивная ячейка в форме параллелепипеда характеризуется шестью параметрами: тремя *длинами ребер* — a , b (в основании), c (высота); тремя *углами* между ними — α (между b и c), β (между c и a), γ (между a и b). Отрезки a , b , c суть периоды трансляции или *периоды идентичности*.

Координатные оси X , Y , Z соответствуют *декартовой системе* в случае кубической ячейки ($a = b = c$) и *аффинной системе* (с разными масштабами вдоль осей), если длина хотя бы одного ребра отлична от других ($a = b \neq c$ или $a \neq b \neq c$). В обоих случаях координатная система может быть либо *прямоугольной* ($\alpha = \beta = \gamma = 90^\circ$), либо *косугольной*, если хотя бы один из углов отличен от 90° ($\alpha = \beta = 90^\circ \neq \gamma$ или $\alpha \neq \beta \neq \gamma \neq 90^\circ$). Как правило, координатные оси X и Y размещают в плоскости основания вдоль ребер a и b , а ось Z направляют вдоль ребра c . Однако для ромбоэдрической ячейки (см. фигуру 11 на рис. 3.1), представляющей собой куб, растянутый (или сжатый) вдоль одной из его пространственных диагоналей, именно она принимается за ось Z косугольной декартовой системы координат. Кристаллографические оси для остальных пяти сингоний (тетрагональной, ромбической, гексагональной, моноклинной, триклинной) соответствуют аффинной системе координат.

Особое место занимает гексагональная ячейка в форме призмы с центрированным шестиугольником в основании (имеющим равные стороны a) и перпендикулярной ему оси c вдоль высоты призмы (см. фигуру 10 на рис. 3.1). В этом случае примитивной ячейкой является прямая призма с основаниями в форме ромба

Таблица 3.1. Сингонии и свойства примитивных ячеек Браве с указанием распределения по сингониям решеток Браве (общим числом 14), точечных групп (общим числом 32) и пространственных групп (общим числом 230)

Сингония: углы и ребра ячейки Браве	Форма ячейки Браве	Число решеток Браве (всего 14)	Число точечных групп (всего 32)	Число простран. групп (всего 230)
кубическая: $\alpha = \beta = \gamma = 90^\circ$ $a = b = c$	куб	3	5	36
тетрагональная: $\alpha = \beta = \gamma = 90^\circ$ $a = b \neq c$	призма с квадратным основанием	2	7	68
(орто)ромбическая: $\alpha = \beta = \gamma = 90^\circ$ $a \neq b \neq c$	прямоуголь- ный паралле- лелепид	4	3	59
гексагональная: $\alpha = \beta = 90^\circ, \gamma = 120^\circ$ $a = b \neq c$	правильная шестиуголь- ная призма	1	7	27
ромбоэдрическая: $\alpha = \beta = \gamma \neq 90^\circ$ $a = b = c$	ромбоэдр (растянутый куб)	1	5	25
моноклинная: $\alpha = \beta = 90^\circ, \gamma \neq 90^\circ$ $a \neq b \neq c$	прямой параллелепи- пед	2	3	13
триклинная: $\alpha \neq \beta \neq \gamma \neq 90^\circ$ $a \neq b \neq c$	косоугольный параллелепи- пед	1	2	2

(выделенного серым цветом) с углом 120° в центре шестиугольника, образованного тремя такими призмами. Примитивная ячейка (затемненная *ортромбическая призма*) составлена из двух *тригональных призм*, разделенных пунктирной линией на фигуре 10 рис. 3.1, с основанием в форме правильного треугольника.

По числу параметров (с максимальным числом 6), характеризующих сингонии, принято разбивать их на три категории:

- а) высшая категория, характеризующаяся одним параметром;
- б) средняя категория, характеризующаяся двумя параметрами;
- в) низшая категория, характеризующаяся числом параметров, большим двух (3, 4, 6).

Высшая категория содержит одну единственную сингонию — *кубическую*, состоящую из примитивной (ПрК), объемноцентрированной (ОЦК) и гранецентрированной (ГЦК) кубических решеток, параметром которых является период решетки a .

Средняя категория содержит три сингонии: *тетрагональную* (с примитивной и объемноцентрированной решетками), *гексагональную* и *ромбоэдрическую*. Первые две характеризуются периодами решетки a и c , а последняя — периодом решетки a и углом α (см. фигуры 4–5, 10 и 11 на рис. 3.1 и табл. 3.1).

Низшая категория содержит три оставшиеся сингонии:

а) *(орто)ромбическую сингонию* (с примитивной, базоцентрированной, объемноцентрированной и гранецентрированной решетками, см. фигуры 6–9 на рис. 3.1 и табл. 3.1), имеющую 3 параметра — периоды решетки a , b , c ;

б) *моноклинную сингонию* (с примитивной и базоцентрированной решетками, см. фигуры 12–13 на рис. 3.1 и табл. 3.1), имеющую 4 параметра — периоды решетки a , b , c и угол наклона γ ;

в) *триклинную сингонию* (см. фигуру 14 на рис. 3.1 и табл. 3.1), имеющую все 6 параметров — периоды решетки a , b , c и углы наклона α , β , γ .

Индексы Миллера обозначают пространственную ориентацию кристаллографических плоскостей и направлений в решетке.

Кристаллографическую плоскость (в том числе грани кристалла) обозначают с помощью отрезков ta , nb , pc , отсекаемых этой плоскостью на трех осях X , Y , Z координатной системы, принятой для данной сингонии. Отношение трех чисел $m : n : p$ (в общем случае являющихся рациональными) характеризует наклон этой плоскости (а значит и всего семейства параллельных плоскостей) к осям координат.

Индексы Миллера h , k , l вводят как обратные величины по отношению к исходным числам m , n , p , приведенные к взаимно простым (целым) числам (с фактором приведения A):

$$h : k : l = \frac{A}{m} : \frac{A}{n} : \frac{A}{p}.$$

Так, при $m : n : p = 2 : 3 : 4$ имеем $A = 12$ и $h : k : l = 6 : 4 : 3$. Целые числа h , k , l называют *индексами плоскости*, а заключенные в круглые скобки (hkl) — *символом плоскости* (и всех параллельных плоскостей); в приведенном примере это — (643) .

Индексы Миллера означают, что на координатных осях внутри ячейки Браве отсекаются отрезки a/h , b/k , c/l . Если символ записан в виде $(\bar{h}kl)$, $(h\bar{k}l)$, $(hk\bar{l})$, то соответствующий индекс

(отмеченный чертой) надо брать со знаком минус, т. е. плоскость отсекает отрезок в отрицательном направлении вдоль этой оси.

Если плоскость параллельна оси, то отсекаемый ею отрезок на этой оси становится бесконечным, а соответствующий индекс равным 0. Тогда символы трех координатных плоскостей, параллельных двум осям, записывают как (100) , (010) , (001) . Для кубической решетки эти плоскости и три другие с указанием их символов приведены на рис. 3.2. Совокупность *эквивалентных* (в силу симметрии) *плоскостей* принято записывать в фигурных скобках. Например, на рис. 3.2 для трех плоскостей в верхнем ряду имеем совокупную запись в форме $\{100\}$, а для совокупности плоскостей, симметрично эквивалентных каждой из трех в нижнем ряду, соответственно, имеем $\{110\}$, $\{111\}$, $\{112\}$.

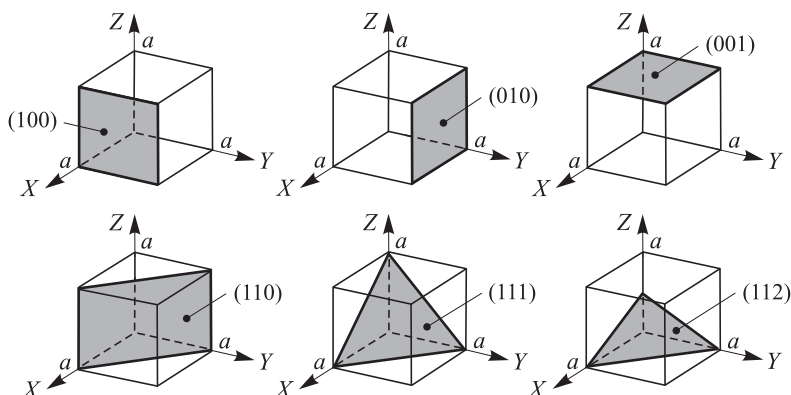


Рис. 3.2. Положение и символы некоторых плоскостей в кубической ячейке

Для описания гексагональных кристаллов (см. фигуру 10 на рис. 3.1) удобно вместо символа (hkl) использовать четырехосный символ $(hkil)$ путем введения третьей оси U , характеризующей индексом i . Координатные оси X, Y, U в плоскости шестиугольного основания расположены под углом 120° друг к другу. Поскольку геометрия кристалла исчерпывающе описывается тремя индексами Миллера, то дополнительно введенный индекс i зависит от первых двух, а именно $h + k + i = 0$. В четырехосной записи плоскость основания имеет символ (0001) , а для шести боковых граней призмы символы таковы: $(1\bar{1}00)$, $(0\bar{1}10)$, $(\bar{1}010)$, $(\bar{1}100)$, $(01\bar{1}0)$, $(10\bar{1}0)$, что совокупно записывается как $\{1\bar{1}00\}$.

Кристаллографическое направление на любой узел решетки обозначают с помощью его координат u, v, w , принятых для данной сингонии. При движении вдоль этого направления соотноше-

ние между координатами ($u : v : w$) сохраняется и может быть использовано для индексации направления. В отличие от символа плоскости (hkl), для символа направления применяют квадратные скобки $[uvw]$, а совокупность эквивалентных направлений обозначают угловыми скобками $\langle uvw \rangle$. Следовательно, направления координатных осей (и параллельные им) получают символы $[100]$ (вдоль a), $[010]$ (вдоль b), $[001]$ (вдоль c), а противоположные им обозначают с чертой над единицей. В гексагональной системе вертикальной оси c соответствует символ $[0001]$.

В кубической сингонии направление, перпендикулярное плоскости (hkl), имеет те же индексы Миллера, только указанные в квадратных скобках $[hkl]$ (для других систем это не справедливо). Например, направление пространственной диагонали $[111]$ в кубической ячейке перпендикулярно плоскости (111) , показанной на рис. 3.2.

3.3. Пространственная упаковка и координация частиц в кристаллических структурах

Точечное расположение узлов в решетке является абстракцией, необходимой для рассмотрения свойств симметрии кристаллов и вводимой в пренебрежении конечными размерами частиц. Однако физические свойства кристаллических структур зависят от химической природы вещества, размеров атомов/ионов и сил взаимодействия между ними.

Поэтому вводят в рассмотрение *эффективные радиусы* атомов и ионов, понимая под этим радиус сферы их действия, причем атом/ион считается несжимаемым шаром. Атомные и ионные радиусы элементов, экспериментально найденные из рентгеноструктурных измерений и согласованные с квантово-механическими расчетами, приведены в литературе в форме таблиц.

В структуре кристалла атомы/ионы рассматриваются как соприкасающиеся жесткие шары, так что расстояние между их центрами равно сумме эффективных радиусов. Изменение характера химической связи между атомами сказывается на межатомных расстояниях: один и тот же элемент имеет разный эффективный радиус в кристаллах разного типа. Действительно, модель несжимаемых шаров не учитывает деформацию электронной оболочки атомов в электрическом поле, которая приводит к их *поляризуемости*, зависящей от типа связи атомов. Установлены следующие закономерности для химических элементов.

1. В периодической таблице внутри одной группы элементов радиусы атомов (ионов с одинаковым зарядом) увеличиваются

с ростом атомного номера (сверху вниз), поскольку возрастает число внутренних электронных оболочек.

2. Ионный радиус элемента уменьшается с увеличением положительного заряда ядра (сильнее притягивающего электроны) и возрастает с увеличением отрицательного заряда электронов; поэтому радиус аниона, получившего избыток электронов, всегда больше радиуса катиона, потерявшего валентные электроны.

3. Эффекты поляризуемости тем меньше, чем ближе электронная оболочка атома к конфигурации благородного газа.

4. Поляризуемость аниона (отрицательного иона) возрастает с увеличением его радиуса, поскольку поле ядра, будучи экранированным внутренними электронными оболочками, слабее удерживает от смещения электроны на внешних валентных орбитах (собственные в атоме вместе с захваченными в анионе).

5. Поляризующее действие на анион поля катиона (положительного иона) возрастает с уменьшением радиуса и с увеличением положительного заряда катиона; при этом сами катионы взаимно почти не поляризуются, так как их меньший электронный заряд прочно удерживается большим зарядом ядра.

В кристаллической структуре на один центральный поляризуемый ион (анион) действуют несколько симметрично расположенных поляризующих ионов (катионов). В этом случае диполи могут не возникать, а эффект поляризации проявляется в уменьшении межатомных расстояний и эффективных радиусов анионов. Симметричному расположению ионов одного знака вокруг центрального иона противоположного знака способствуют силы отталкивания между одноименными ионами. Нарушение симметрии в пространственной ориентации окружающих частиц порождается сближением их электроотрицательностей. Это меняет взаимную координацию частиц как результат появления ковалентного вклада и перехода от чисто ионной связи к смешанной (ионно-ковалентной) и в пределе к ковалентной связи.

1. Пространственная упаковка. Если представлять атомы или ионы твердыми несжимаемыми шарами, между которыми действуют силы взаимного притяжения и отталкивания, то особенности кристаллического строения большинства структур можно объяснить, рассматривая их как упаковку таких шаров.

Для качественного суждения о типе пространственной упаковки атомов/ионов важную роль, кроме эффективных радиусов, играет так называемое *координационное число* z данного атома (иона), равное числу его ближайших соседей в структуре кристалла. Соединив мысленно прямыми линиями центры соседних

частиц друг с другом, получим *координационный многогранник*; при этом атом (ион), для которого строится такой многогранник, находится в его центре.

Координационное число и отношение эффективных радиусов частиц определенным образом связаны друг с другом: чем меньше различие в размерах частиц, тем больше z . Обоснование этого утверждения выполним на примере ионной связи.

Ионная связь, будучи ненаправленной (в пренебрежении поляризацией ионов), определяется, главным образом, геометрическим фактором, т. е. отношением радиусов катиона (меньшего) и аниона (большого), $r_k/r_a < 1$. Объяснение на качественном уровне следует из рис. 3.3, где показана последовательность перехода от исходного устойчивого состояния в форме катиона, окруженного четырьмя анионами (поз. I), к новому устойчивому окружению катиона тремя анионами (поз. IV). Это происходит в результате постепенного уменьшения размера катиона.

В устойчивой конфигурации ионов взаимное отталкивание анионов уравновешено их притяжением со стороны центрального катиона; при этом анионы соприкасаются только с катионом, но не между собой (поз. I). С уменьшением радиуса катиона такая конфигурация сохраняется устойчивой вплоть до момента взаимного касания анионов (поз. II). При дальнейшем уменьшении размера катиона угловая симметрия в расположении анионов нарушается и начинает преобладать кулоновское расталкивание между ними (поз. III). Это приводит к отщеплению одного из анионов и к формированию новой устойчивой конфигурации треугольной формы (поз. IV).

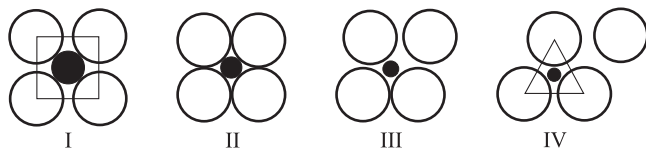


Рис. 3.3. Влияние размера катиона (черного шара) на конфигурацию окружения его анионами (белыми шарами): трансформация квадратного окружения (поз. I) в треугольное окружение (поз. IV) с уменьшением радиуса катиона

Таким образом, структурой, которая разделяет две ближайшие устойчивые конфигурации с разным координационным числом, является предельная структура с одновременным касанием всех шаров малого и большого радиуса (обозначенных как r_k и r_a для катионов и анионов, соответственно) — см. поз. II на рис. 3.3. В такой ситуации отношение r_k/r_a определяет нижний предел устойчивости ионной конфигурации.

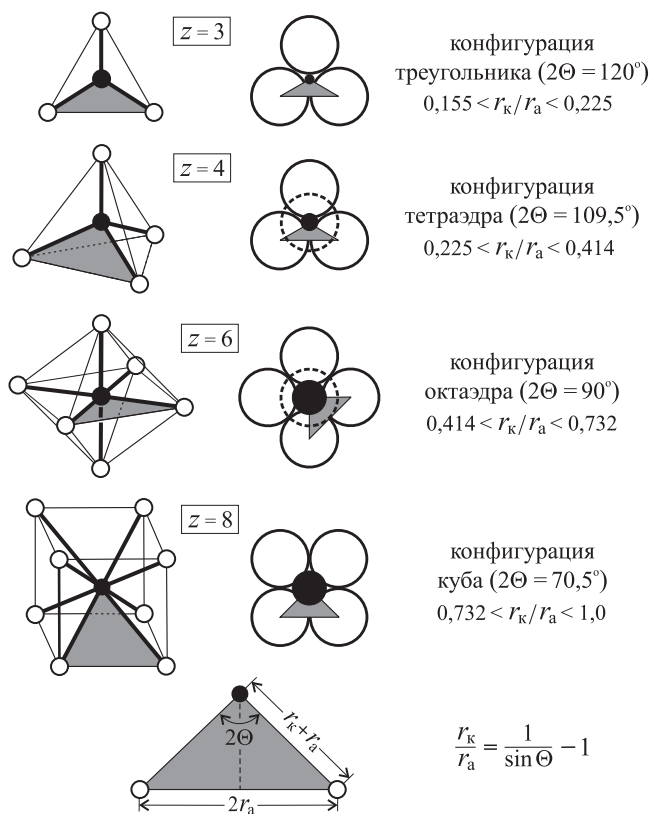


Рис. 3.4. Конфигурация координационного многогранника и нижнее предельное отношение радиусов катиона (черные шары) и аниона (белые шары) для четырех координационных чисел $z = 3, 4, 6, 8$. Внизу изображен (в плоскости рисунка) определяющий треугольник, из которого следует формула для r_k/r_a

Подобные предельные ситуации изображены на рис. 3.4 для четырех значений $z = 3, 4, 6, 8$. Конфигурации координационного многогранника даны в двух видах — в аксонометрическом изображении (левый столбец с ионами в форме малых кружков белого и черного цвета) и в проекции на плоскость рисунка при виде сверху (правый столбец с соприкасающимися анионами). Серым цветом выделены определяющие треугольники с углом 2Θ при их вершине, примыкающей к центральному катиону. Из геометрии треугольника следует формула для r_k/r_a , приведенная внизу рисунка. Расчет по этой формуле дает нижнее предельное значение, указанное в соответствующих неравенствах, выше которого существует та или иная ионная конфигурация.

Таким образом, с уменьшением различия в размерах катиона и аниона происходит смена ионных конфигураций в следующей последовательности: *треугольная* ($z = 3$ при $0,155 < r_k/r_a < 0,225$) $\rightarrow$ *тетраэдрическая* ($z = 4$ при $0,225 < r_k/r_a < 0,414$) $\rightarrow$ *октаэдрическая* ($z = 6$ при $0,414 < r_k/r_a < 0,732$) $\rightarrow$ *кубическая* ($z = 8$ при $0,732 < r_k/r_a < 1,0$). С ростом координационного числа растет плотность упаковки частиц в кристаллической структуре. Однако значение $z = 8$ отнюдь не обеспечивает максимальную упаковку, которая возможна лишь для шаров одинакового размера при $z = 12$ и называется **плотнейшей упаковкой**.

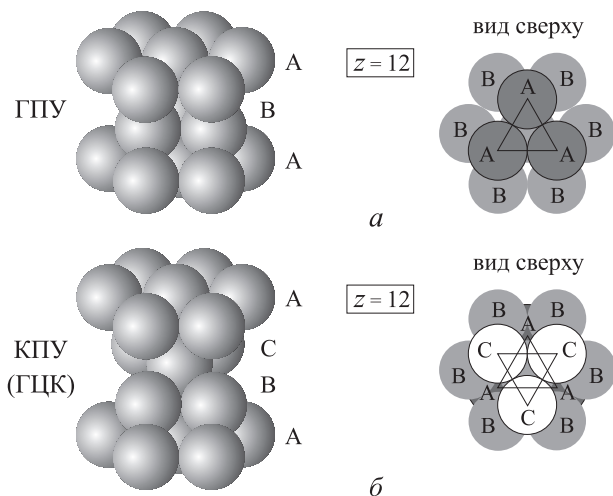


Рис. 3.5. Последовательность слоев при упаковке шаров одинакового размера в структурах двух типов: *а* — АВАВ... при гексагональной плотнейшей упаковке (ГПУ); *б* — АВСАВС... при кубической плотнейшей упаковке (КПУ), дающей гранецентрированную кубическую (ГЦК) решетку

Два разных типа структуры с плотнейшей упаковкой шаров одинакового радиуса показаны на рис. 3.5 *а* и *б* в двух видах — в аксонометрическом изображении (левый столбец) и в проекции на плоскость рисунка при виде сверху (правый столбец). За основу построения в плоскости рисунка взяты шары слоя В светло-серого цвета, которые образуют правильный шестиугольник вокруг центрального шара.

Шары слоя А лежат в лунках первого слоя А, расположенного ниже плоскости рисунка. Поэтому они явно не просматриваются на рис. 3.5 *а*, поскольку закрыты такими же шарами верхнего слоя А темно-серого цвета, образующими вокруг центрального шара В *базисный треугольник* ААА (с вершиной, направленной

вверх). На рис. 3.5 б шары нижнего слоя А просматриваются как темно-серые пустоты в слое В, которые отмечены таким же, как на рис. 3.5 а, базисным треугольником ААА (см. ниже рис. 5.4).

Размещение шаров в первых двух слоях А (темно-серые) и В (светло-серые) на обоих рисунках одинаково. Различие начинается лишь в третьем слое: а) либо шары повторяют размещение первого слоя А (см. на рис. 3.5 а *базисный треугольник* ААА темно-серых шаров); б) либо шары С белого цвета располагаются в других лунках слоя В с поворотом на 60° (см. на рис. 3.5 б *базисный треугольник* ССС белых шаров, направленный вершиной вниз, т. е. антипараллельный треугольнику ААА). Во втором случае выше слоя С снова размещается слой А, отмеченный на рис. 3.5 б треугольником ААА (без показа темно-серых шаров).

Таким образом, чередование слоев на рис. 3.5 а и б можно записать, соответственно, как АВ АВ... (двухслойная структура) и АВ С АВ С... (трехслойная структура). Первая структура называется **гексагональной плотнейшей упаковкой** (ГПУ), а вторая — **кубической плотнейшей упаковкой** (КПУ).

В отличие от простой гексагональной ячейки, имеющей произвольную высоту c (фигура 10 на рис. 3.1), ячейка ГПУ из-за наличия дополнительного слоя В между основаниями А имеет вполне определенное соотношение осей $c/a = 2\sqrt{2}/3 = 1,633$. На рис. 3.5 а вид сверху изображен в плоскости шестиугольного основания гексагональной ячейки, перпендикулярной оси c .

Ячейка КПУ, как нетрудно заметить, полностью совпадает с ячейкой ГЦК (гранецентрированной кубической) решетки, изображенной в виде фигура 3 на рис. 3.1. При этом плотнейшие слои лежат в плоскостях $\{111\}$, перпендикулярных пространственным диагоналям куба $\langle 111 \rangle$. Следовательно, на рис. 3.5 б вид сверху изображен в кристаллографической плоскости (111).

Для плотнейших упаковок координационное число $z = 12$, что очевидно для центрального шара в слое В. Его ближайшее окружение составляют 6 шаров слоя В и по 3 шара ниже и выше его (в слоях А и А на рис. 3.5 а или в слоях А и С на рис. 3.5 б).

Кроме координационного числа важно знать форму и размер пустот между шарами, в которых могут размещаться частицы меньших размеров. В плоскости своего слоя пустота образована тремя шарами и ее форма зависит от того, каким образом она «закрыта» сверху или снизу — шаром или лункой соседнего слоя. В ГПУ- и ГЦК-структурах существуют пустоты двух типов:¹⁾

¹⁾ Использованы греческие слова: грань — *эдра*, угол — *гония* и цифры: 4 — *тетра*, 5 — *пента*, 6 — *гекса*, 8 — *окта*, 10 — *дека*, 12 — *додека*, 20 — *икоса*.

- **тетраэдрические пустоты**, «закрытые» (сверху или снизу) шаром соседнего слоя, так что пустота окружена четырьмя шарами;
- **октаэдрические пустоты**, «закрытые» (сверху или снизу) лункой соседнего слоя, так что пустота окружена шестью шарами.

Тетраэдрическая и октаэдрическая пустоты возникают из конфигураций тетраэдра и октаэдра, изображенных на рис. 3.4, при удалении центрального катиона (черного шара). Из приведенных там неравенств следует, что в ГПУ- и ГЦК-решетках при радиусе шаров r в пустоту помещается малый шар с радиусом, равным

$$r_{\text{т.п.}} = 0,225 r \text{ — для тетраэдрической пустоты,}$$

$$r_{\text{о.п.}} = 0,414 r \text{ — для октаэдрической пустоты.}$$

На рис. 3.5 *а* для ГПУ тетраэдрическая пустота в центре треугольника ААА образована тремя шарами А и снизу закрыта центральным шаром В светло-серого цвета. Там же видны маленькие белые площадки за пределами тройки шаров А, которые образуют между слоями А и В октаэдрические пустоты. Они являются сквозными в том смысле, что продолжаютсЯ из слоя в слой как сплошные каналы, по которым может происходить диффузия примесей в кристаллах. Каждый шар окружен шестью октаэдрическими и восемью тетраэдрическими пустотами, а соответствующие пустоты образованы шестью и четырьмя шарами. Отсюда следует вывод, что при ГПУ на каждый шар приходится одна октаэдрическая пустота и две тетраэдрические пустоты.

При КПУ сложнее выглядит картина размещения пустот относительно кристаллографических осей ГЦК-решетки, так как нормаль к плоскости рис. 3.5 *б* (вид сверху) совпадает с направлением $[111]$ пространственной диагонали куба. Положение пустот в ГЦК-структуре пояснено на рис. 3.6, где четыре атома, окружающие тетраэдрическую пустоту, выделены серым цветом (с засветкой), а пять атомов, не примыкающих к пустотам, показаны белыми шарами меньшего диаметра.

Тетраэдрические пустоты (т.п.) существуют в центре тетраэдров, построенных при каждой из 8 вершин куба. Исходящие из вершины три ребра заканчиваются в центре трех прилегающих граней куба. Соединение этих трех центров дает грань тетраэдра (затемненную

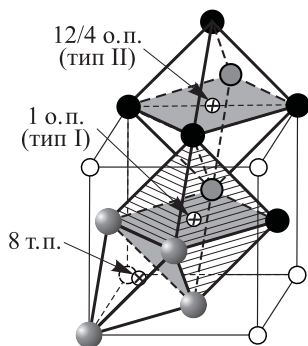


Рис. 3.6. Расположение тетраэдрических пустот (т.п.) и октаэдрических пустот (о.п.) в гранецентрированной кубической структуре

для ясности), которая перпендикулярна пространственной диагонали куба. Центр тетраэдрической пустоты (показанный на рисунке крестиком) располагается от вершины на расстоянии $d/4$, где $d = \sqrt{3}a$ — длина пространственной диагонали куба, a — период решетки. Поскольку ГЦК-ячейка содержит четыре атома, то на каждый из них приходится две тетрагональные пустоты, как и в ГПУ-структуре.

Октаэдрические пустоты (о.п.) в структуре ГЦК бывают двух типов (I и II). Тип I полностью принадлежит одной ячейке и имеет форму октаэдра, образованного шестью атомами в центре граней куба. Центр этой пустоты (показанный на рисунке крестиком) расположен в середине базисного квадрата (затемненного для ясности) и совпадает с центром куба. Октаэдр типа I соприкасается с угловыми тетраэдрами всеми гранями, одна из которых показана затемненной на рисунке. Октаэдры типа II выходят за пределы данной ячейки и центрированы на середине ребер куба, каковых в ячейке 12 (один из них изображен на рисунке выше центрального октаэдра и содержит два атома соседней ячейки). Поскольку каждое ребро принадлежит четырем ячейкам, то общее число октаэдрических пустот ($1 + 12/4 = 4$) равно числу атомов в ГЦК-ячейке, как и в случае ГПУ структуры.

Рассмотренные выше принципы пространственной упаковки атомов/ионов служат основанием для качественного суждения о типе кристаллической структуры твердых веществ.

2. Основные типы кристаллических структур. При кристаллизации вещества система стремится обеспечить минимум свободной энергии. Одним из факторов, снижающих потенциальную энергию взаимодействия между частицами, является их максимальное сближение и установление взаимной связи с возможно большим числом частиц, т. е. стремление к более плотной упаковке частиц с наибольшим координационным числом.

Тенденция к реализации плотнейшей упаковки свойственна всем типам структур, но сильнее всего она выражена в металлических, ионных и молекулярных (преимущественно диэлектрических) кристаллах. В них связи ненаправленные или слабо направленные, так что для атомов, ионов и молекул вполне приемлемой является модель твердых несжимаемых шаров. Однако в кристаллах, построенных с преобладанием гетерополярности, имеет место связь между координационным числом (z) и отношением ионных радиусов (r_k/r_a), представленная на рис. 3.4.

Полупроводниковые кристаллы используют направленную ковалентную (неполярную или полярную) связь, которая обычно

реализуется путем комбинации по крайней мере двух решеток, по отдельности достаточно плотно упакованных, но в итоге обеспечивающих малые координационные числа (вплоть до $z = 4$).

Трансляционными решетками Браве (общим число 14), приведенными на рис. 3.1 и в табл. 3.1, отнюдь не исчерпываются все возможные варианты, возникающие при построении кристаллических структур, в первую очередь для химических соединений. Дело в том, что периодическое повторение ячейки Браве дает трансляционную решетку, состоящую только из частиц (молекул, атомов, ионов) одного сорта. Поэтому структуру сложного соединения можно построить комбинацией обычно двух решеток Браве, определенным образом вставленных одна в другую.

Кристаллические структуры ряда веществ сходны по размерам частиц, расстоянию между ними и пространственному расположению атомов/ионов (или атомных/ионных групп) и отличаются только параметрами (но не типом) кристаллической решетки. Поэтому структуры подобных веществ объединены в один *структурный тип*, названный по имени его характерного представителя. Для описания конкретной структуры надо указать параметры решетки и структурный тип, к которому относится данный кристалл.

Число структурных типов ограничено и наиболее часто встречаются следующие типы структур, характерные для металлических, диэлектрических и полупроводниковых кристаллов:

- *в металлических кристаллах*
 - 1) структура вольфрама (ОЦК-решетка),
 - 2) структура меди (ГЦК-решетка),
 - 3) структура магния (ГПУ-решетка);
- *в диэлектрических кристаллах*
 - 1) структура хлористого натрия (сдвоенная ГЦК-решетка),
 - 2) структура хлористого цезия (сдвоенная ПрК-решетка);
- *в полупроводниковых кристаллах*
 - 1) структура алмаза (сдвоенная ГЦК-решетка),
 - 2) структура сфалерита (сдвоенная ГЦК-решетка),
 - 3) структура вюрцита (сдвоенная ГПУ-решетка).

Следовательно, наиболее распространенными являются кристаллические структуры, построенные на решетках Браве кубической и гексагональной сингонии (см. рис. 3.1 и табл. 3.1).

Кроме координационного числа z различные структуры характеризуются также *плотностью упаковки*, вводимой как отношение объема $V_{\text{ат}}$, занимаемого атомами, к объему $V_{\text{яч}}$ всей

ячейки Браве. Атомы представляются твердыми шарами радиуса r , поэтому $V_{\text{ат}} = n(4\pi/3)r^3$, где n — число атомов в ячейке. Объем кубической ячейки $V_{\text{яч}} = a^3$, где a — период решетки. Для ячейки ГПУ с площадью шестиугольного основания $A = 3\sqrt{3}a^2/2$ и высотой $c = 2\sqrt{2/3}a$ получаем $V_{\text{яч}} = 3\sqrt{2}a^3$.

Соответствующие параметры кристаллических структур — примитивной кубической (ПрК), объемноцентрированной кубической (ОЦК), гранецентрированной кубической (ГЦК), гексагональной плотноупакованной (ГПУ) — вместе с рассчитанной на их основе плотностью упаковки приведены в табл. 3.2. Радиусы атомов записаны с учетом того, что они соприкасаются вдоль ребер куба в ПрК ($2r = a$), вдоль пространственных диагоналей ($4r = \sqrt{3}a$) в ОЦК и вдоль диагоналей граней ($4r = \sqrt{2}a$) в ГЦК.

Таблица 3.2. Параметры кубических и гексагональных кристаллов

Параметры кристалла	ПрК	ОЦК	ГЦК	ГПУ
Координационное число z	6	8	12	12
Число атомов в ячейке, n	1	2	4	6
Радиус атома, r	$a/2$	$a\sqrt{3}/4$	$a\sqrt{2}/4$	$a/2$
Объем одного атома, $V_{\text{ат}}/n$	$\frac{\pi}{6}a^3$	$\frac{\pi\sqrt{3}}{16}a^3$	$\frac{\pi\sqrt{2}}{24}a^3$	$\frac{\pi}{6}a^3$
Плотность упаковки, $V_{\text{ат}}/V_{\text{яч}}$	$\frac{\pi}{6} = 0,52$	$\frac{\pi\sqrt{3}}{8} = 0,68$	$\frac{\pi\sqrt{2}}{6} = 0,74$	$\frac{\pi\sqrt{2}}{6} = 0,74$

Таким образом, в структурах с плотнейшей упаковкой (ГЦК и ГПУ), имеющих $z = 12$, объем ячейки на 74% занят атомами. При уменьшении координационного числа до 8 и 6 плотность упаковки снижается соответственно до 68% (ОЦК) и 52% (ПрК).

Рассмотрим кратко особенности и реализуемость перечисленных выше структур и соответствующие им решетки Браве.

Структура вольфрама (объемноцентрированная кубическая решетка), показанная на рис. 3.7 а, не является структурой с полнейшей упаковкой. Тем не менее к ней, помимо щелочных и щелочно-земельных металлов, относится также большинство тугоплавких металлов кроме вольфрама W, такие как хром Cr, железо Fe, молибден Mo, цирконий Zr, тантал Ta, ниобий Nb и др. Последнее находит следующее объяснение. В ячейке ОЦК для

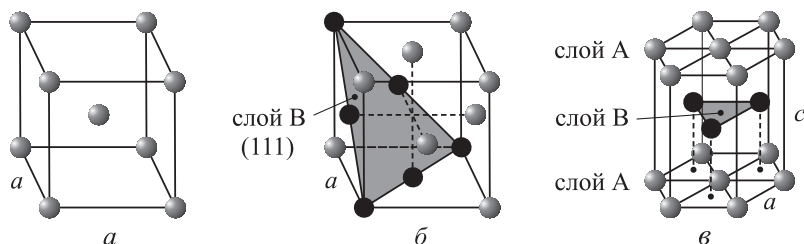


Рис. 3.7. Ячейки Браве: *а* — объемноцентрированная кубическая (ОЦК, структура вольфрама), *б* — гранецентрированная кубическая (ГЦК, структура меди), *в* — гексагональная плотноупакованная (ГПУ, структура магния)

центрального атома ближайшими соседями являются атомы в вершинах куба ($z = 8$). Они отстоят друг от друга на расстоянии $d = a\sqrt{3}/2 = 0,866a$ и обеспечивают плотность упаковки 68%. Кроме них на расстоянии $d = a$ от данного атома располагаются шесть центральных атомов в соседних ячейках (вторая координационная сфера), что практически увеличивает координационное число до $z \approx 14$. Это дает суммарный выигрыш энергии, компенсирующий отрицательный вклад от небольшого увеличения средних расстояний между атомами по сравнению с ГЦК, где атомы находятся на расстоянии $d = a\sqrt{2}/2 = 0,707a$. В результате повышается прочность кристаллов, проявляющаяся в их высокой температуре плавления, достигающей у вольфрама 3422°C .

Структура меди (гранецентрированная кубическая решетка), показанная на рис. 3.7 б, кроме меди Cu характерна для многих металлов, таких как золото Au, серебро Ag, платина Pt, никель Ni, алюминий Al, свинец Pb, палладий Pd, торий Th и др. Многие из них образуют непрерывные твердые растворы и интерметаллические соединения (см. п. 2.11). Эти металлы сравнительно мягкие и пластичные. Последнее объясняется тем, что в структурах типа меди тетраэдрические и октаэдрические пустоты в ГЦК-решетке, показанные на рис. 3.6, не заполнены другими частицами. Это допускает в силу ненаправленности связей между атомами их смещение по так называемым *плоскостям скольжения*. В решетке ГЦК таковыми являются плоскости наибольшей упаковки $\{111\}$, одна из которых изображена заштрихованной на рис. 3.7 б. Атомы, принадлежащие этой плоскости, для наглядности выделены черным цветом.

Структура магния (гексагональная плотноупакованная решетка), показанная на рис. 3.7 в, характерна не только для магния Mg, но и для кадмия Cd, цинка Zn, титана Ti, таллия Tl, бериллия Be и др., а также для большинства редкоземельных

элементов и ряда интерметаллических соединений. В отличие от простой решетки, изображенной на рис. 3.1 (фигура 10), ГПУ-решетка на рис. 3.7 *в* имеет слой В (для наглядности затемненный) расположенный посередине между базисными слоями А на фиксированном расстоянии $c/2 = a\sqrt{2}/3$ (с наблюдаемым отклонением вплоть до 10% для некоторых металлов). Атомы в слоях В (выделенные черным цветом), размещаются над центрами треугольников в базисной плоскости (0001) с плотнейшей упаковкой.

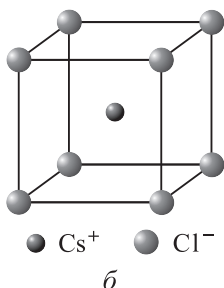
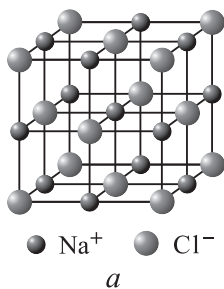


Рис. 3.8. Структуры каменной соли (*a*) и хлористого цезия (*б*)

Структура каменной соли (сдвоенная гранецентрированная кубическая решетка), показанная на рис. 3.8 *а*, может быть описана как две ГЦК-решетки (типа структуры меди), сдвинутые на полпериода решетки ($a/2$) вдоль любого из ребер $\langle 100 \rangle$. В структуры этого типа кристаллизуются практически все галогениды щелочных металлов, большая часть оксидов и халькогенидов (сульфидов, селенидов, теллуридов) щелочноземельных металлов и т. п. Крупные анионы хлора Cl^- занимают узлы ГЦК-ячейки и образуют кубическую плотнейшую упаковку, в которой катионы натрия Na^+ , имея меньший размер, заполняют только октаэдрические пустоты, оставляя тетраэдрические свободными (см. рис. 3.6). Действительно, для NaCl отношение эффективных ионных радиусов таково: $r_{\text{Na}^+}/r_{\text{Cl}^-} = 0,95/1,81 \approx 0,52$. Тогда неравенства $0,414 < r_{\text{к}}/r_{\text{а}} < 0,732$ из рис. 3.4 обеспечивают октаэдрическое окружение с

координационным числом $z = 6$ для каждого иона. Именно это имеет место в структуре NaCl: ион натрия окружен шестью ионами хлора и наоборот.

Структура хлористого цезия (сдвоенная примитивная кубическая решетка), показанная на рис. 3.8 *б*, на первый взгляд кажется объемноцентрированной, подобно решетке вольфрама на рис. 3.7 *а*. Но это не так, поскольку в вершинах и в центре куба стоят разные ионы, что не обеспечивает трансляционную симметрию ОЦК-решетки Браве (нет трансляции из вершины ячейки в ее центр). Поэтому структура CsCl состоит из двух примитивных кубических решеток, сдвинутых на половину объемной диагонали. Дело в том, что ионы цезия больше ионов натрия и

не могут поместиться в октаэдрических (и тем более в тетраэдрических) пустотах решетки хлора, если бы она была ГЦК, как в структуре NaCl. Для CsCl отношение эффективных ионных радиусов таково: $r_{\text{Cs}^+}/r_{\text{Cl}^-} = 1,69/1,81 \approx 0,93$. В этом случае неравенства $0,732 < r_{\text{к}}/r_{\text{а}} < 1,0$ на рис. 3.4 обеспечивают кубическое окружение иона с координационным числом $z = 8$. Именно это имеет место в структуре CsCl: каждый ион цезия окружен восемью ионами хлора и наоборот. В структуры этого типа кристаллизуются и другие галогениды, например Cs(Br, I), Rb(Br, I), Tl(Br, Cl), и многие сплавы редкоземельных элементов.

Структура алмаза (сдвоенная гранецентрированная кубическая решетка), показанная на рис. 3.9 а, представляет собой две ГЦК-решетки со сдвигом на четверть длины пространственной диагонали, показанной штриховой линией на рис. 3.9 а. Помимо алмаза, в структуру этого типа кристаллизуются также элементарные полупроводники — кремний Si и германий Ge.

Если через центр исходной ГЦК-ячейки мысленно провести три взаимно перпендикулярных плоскости, то получаем восемь кубиков, один из которых показан тонкими линиями на рис. 3.9 а. Середина кубика соответствует центру тетраэдрической пустоты ГЦК-решетки (см. рис. 3.6). Именно сюда помещен атом другой решетки, так что заполненными оказываются четыре кубика (через один) из восьми. В результате этого исчезает исходная полнейшая упаковка ГЦК и возникает тетраэдрическая связь (со стереоуглом $109^\circ 28'$, с плотностью упаковки $\pi\sqrt{3}/16 = 34\%$ при $z = 4$), показанная толстыми линиями на рис. 3.9 а.

Здесь для зрительного восприятия сделано двойное выделение идентичных атомов углерода — цветом и размерами. Четыре атома в вершинах и шесть в центрах граней изображены, соответственно, черными и светло-серыми (с засветкой) и уменьшающимися при пространственном удалении, атомы же в центрах тетраэдров — темно-серыми. Четыре атома в вершинах куба, не участвующие в тетраэдрической связи внутри данной ячейки, показаны малыми черными точками. Результирующая сдвоенная ячейка содержит восемь атомов (4 в центрах тетраэдров, $6 \cdot 1/2$ в центрах граней, $8 \cdot 1/8$ в вершинах куба).

Следовательно, при периоде решетки a (длина ребра куба) число атомов N , содержащихся в единице объема кристалла, равно $8/a^3$, а именно (с указанием межатомного расстояния d):

алмаз: $a = 3,56 \text{ \AA}$, $d_{\text{C-C}} = 1,54 \text{ \AA}$, $N_{\text{C}} = 1,77 \cdot 10^{23} \text{ см}^{-3}$;
 кремний: $a = 5,43 \text{ \AA}$, $d_{\text{Si-Si}} = 2,35 \text{ \AA}$, $N_{\text{Si}} = 5,0 \cdot 10^{22} \text{ см}^{-3}$;
 германий: $a = 5,66 \text{ \AA}$, $d_{\text{Ge-Ge}} = 2,45 \text{ \AA}$, $N_{\text{Ge}} = 4,4 \cdot 10^{22} \text{ см}^{-3}$.

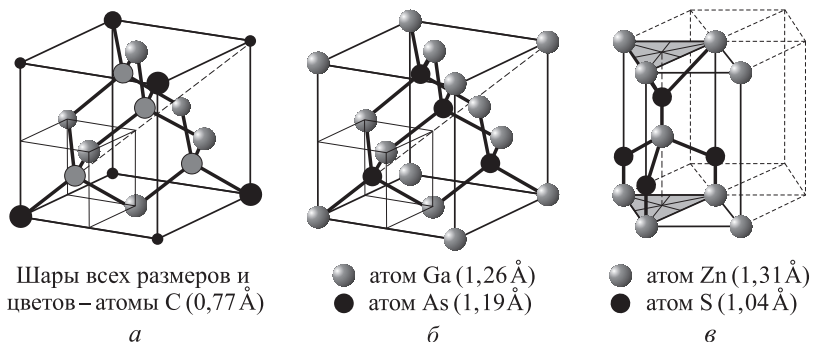


Рис. 3.9. Структуры алмаза (а), сфалерита (арсенид галлия GaAs) (б) и вюрцита (сульфид цинка ZnS) (в); в скобках — тетраэдрические радиусы атомов

В кристаллах типа алмаза вдоль направлений $\langle 110 \rangle$ имеются сквозные шестигранные каналы, по которым особенно легко происходит диффузия примесей, что типично для Si и Ge.

Структура сфалерита (сдвоенная гранецентрированная кубическая решетка), показанная на рис. 3.9 б, характерна для кубической модификации сульфида цинка ZnS, называемой *цинковой обманкой*. Такой структурой обладают также многие полупроводниковые соединения типа $A^{III}B^V$ (арсенид галлия GaAs, фосфид галлия GaP, фосфид индия InP, антимонид индия InSb и др.) и типа $A^{II}B^{VI}$ (селенид цинка ZnSe, теллурид цинка ZnTe, сульфид кадмия CdS, селенид кадмия CdSe и др.).

На рис. 3.9 б приведена ячейка GaAs с указанием ковалентных радиусов галлия и мышьяка. Структура сфалерита идентична структуре алмаза с тетраэдрическим окружением атомов (рис. 3.9 а), только одна ГЦК-подрешетка занята атомами Ga, а другая — атомами As, которые на рисунке различаются размером и цветом шаров. В ячейке GaAs отсутствует центр симметрии, т. е. структура полярна по четырем направлениям $\langle 111 \rangle$. Наблюдается различие между плотноупакованными плоскостями (111) и $(\bar{1}\bar{1}\bar{1})$: если одна из них содержит атомы Ga, то другая — атомы As. Это обуславливает анизотропию свойств поверхности (микротвердость, адсорбция, химическое травление и т. п.).

Ниже приведены периоды решетки a и длина тетрагональной связи d (межатомное расстояние) для ряда полупроводников:

GaP:	$a = 5,45 \text{ Å}$,	$d_{\text{Ga-P}} = 2,36 \text{ Å}$;
GaAs:	$a = 5,65 \text{ Å}$,	$d_{\text{Ga-As}} = 2,45 \text{ Å}$;
InP:	$a = 5,87 \text{ Å}$,	$d_{\text{In-P}} = 2,54 \text{ Å}$;
InSb:	$a = 6,48 \text{ Å}$,	$d_{\text{In-Sb}} = 2,80 \text{ Å}$.

Нетрудно видеть, что в идеальных структурах с тетраэдрическими межатомными связями (структуры алмаза и сфалерита) межатомное расстояние и период решетки связаны соотношением $d/a = \sqrt{3}/4 = 0,433$. Приведенные выше экспериментальные значения a и d (в том числе для C, Si, Ge) дают величины d/a , очень близкие к 0,433.

Структура вюрцита (сдвоенная гексагональная плотноупакованная решетка), изображенная на рис. 3.9 в, характерна для гексагональной модификации сульфида цинка. Такой структурой обладают близкие к ZnS полупроводники, например, сульфид кадмия CdS и селенид кадмия CdSe. Для большинства соединений $A^{II}B^{VI}$ характерен фазовый переход сфалерит $\rightleftharpoons$ вюрцит. Структура вюрцита реализуется, если атом металла имеет малые размеры и большую электроотрицательность.

На рис. 3.9 в приведена примитивная ячейка вюрцита для ZnS в форме прямой призмы с ромбом в основании и углом 120° в центре шестиугольника, образованного тремя такими призмами (две из которых показаны на рисунке пунктиром). Эта элементарная ячейка совпадает с той, что ранее была выделена затемнением на фигуре 10 рис. 3.1.

Орторомбическая призма составлена из двух тригональных призм, треугольные основания одной из которых отмечены серым цветом на рис. 3.9 в. В эту тригональную призму дополнительно введены пять атомов: а) один атом Zn (светлый шар с засветкой) и один атом S (черный шар) на вертикальной линии, проходящей через центры треугольных оснований; б) три атома S на ребрах призмы. Толстыми линиями показаны тетраэдрические связи между атомами цинка и серы.

Таким образом, в гексагональной призме (составленной из шести тригональных призм) атомы Zn располагаются в ее вершинах, в центрах оснований и внутри трех тригональных призм (через одну по кругу), а атомы S — в тех же трех тригональных призмах и на всех вертикальных ребрах гексагональной призмы. Из сравнения рис. 3.9 в с рис. 3.7 в следует, что в структуре вюрцита катионы (атомы Zn) образуют ГПУ-решетку с плотноупакованными слоями ($z = 12$), параллельными плоскости базиса (0001). Анионы (атомы S) также образуют ГПУ-решетку, заполняя половину тетраэдрических пустот в ГПУ-подрешетке Zn. Результирующая сдвоенная ГПУ-решетка имеет по катионам и анионам координационное число $z = 4$. Направление [0001] (подобно направлениям $\{111\}$ в сфалерите) является полярным, тогда плоскости (0001) и (000 $\bar{1}$) заполнены разными атомами

(катионами и анионами), что проявляется в анизотропии физических и химических свойств на различных гранях кристаллов.

3. Политипизм, полиморфизм и физические свойства. До сих пор, рассматривая плотноупакованные структуры, имели дело только с двухслойной упаковкой АВ АВ АВ ... (ГПУ) и с трехслойной упаковкой АВС АВС АВС ... (КПУ или ГЦК). Эти структуры являются плотнейшими с максимальной плотностью упаковки 74% (см. табл. 3.2). Возможно существование n -слойных упаковок ($n > 3$) с меньшей, но достаточно высокой плотностью. С ростом числа слоев увеличивается количество вариантов реализации n -слойной упаковки: $n = 4$ и $n = 5$ — по одной, $n = 6$ — две, $n = 7$ — три, $n = 8$ — шесть, $n = 10$ — 16, $n = 12$ — 43 и т. д. Символы слоев не могут иметь рядом одинаковые буквы, что означало бы расположение шаров друг над другом, а не в лунках:

АВАС АВАС АВАС ... — четырехслойная упаковка;

АВСАС АВСАС АВСАС ... — пятислойная упаковка;

АВСАСВ АВСАСВ АВСАСВ ... — шестислойная упаковка;

АВАВАС АВАВАС АВАВАС ... — шестислойная упаковка.

Свойство кристаллов образовывать структуры из одних и тех же слоев, но с разной последовательностью их чередования называют *политипизмом*. Параметры решетки в плоскости слоев политипов одинаковы, а по нормали к слоям различны, но всегда кратны расстоянию между ближайшими слоями.

Наибольшим числом политипов (вплоть до $n = 1200$) обладает *карбид кремния* (химическое соединение SiC высокой твердости с температурой плавления 2830°C). Политипы SiC записываются в форме nH (от англ. *Hexagonal*) или nR (от англ. *Rhombohedral*) для n -слойных структур с гексагональной и ромбоэдрической ячейками ($a = 3,08 \text{ \AA}$ и $c = 2,52 \text{ \AA}$) с периодом повторения nc вдоль оси c (достигающим 150 нм у политипа 594R). Для карбида кремния найдено более 50 политипов, среди которых наиболее распространенными являются 4H, 6H и 15R.

Политипизм представляет собой частный случай более общего явления, известного как *полиморфизм* (по-гречески *морфо* означает «форма», «образ», «вид»). Под этим понимают свойство некоторых веществ существовать в нескольких кристаллических фазах, отличающихся симметрией и типом структуры. Каждая из таких фаз, называемая *полиморфной модификацией*, устойчива в определенной области температур и давлений. Полиморфные превращения относятся к фазовым переходам I рода (см. п. 2.1) со скачкообразными изменениями энтальпии (тепловой эффект), мольного объема (плотность массы) и структуры кристалла, что

обычно сопровождается изменением других физических свойств (например, удельной теплоемкости, теплопроводности, электропроводности и т. п.).

Вышеуказанные политипы карбида кремния присущи его гексагональной модификации, обозначаемой как α -SiC. Кроме того, карбид кремния встречается в кубической модификации сфалерита β -SiC, записываемой в форме $3C$ (от англ. *Cubic*, $n = 3$ — чередование трех слоев ABC...). Аналогичная ситуация имеет место, что уже было отмечено, для сульфида цинка ZnS, который кристаллизуется как в виде цинковой обманки (кубического сфалерита $3C$) или гексагонального вюрцита ($2H$), так и в виде политипов $4H$, $6H$, $8H$, $10H$, $9R$, $12R$, $15R$ и др.

Карбид кремния относится к широкозонным полупроводникам. Различие политипных модификаций проявляется в ширине запрещенной зоны: 2,2 эВ для $3C$; 2,9 эВ для $6H$; 3,3 эВ для $4H$. Это приводит к изменению физических свойств (в первую очередь оптических), например, двойного лучепреломления света.

Среди полиморфных модификаций других веществ наиболее известны две *аллотропные модификации*¹⁾ углерода в форме кубической структуры алмаза (со sdвоенной ГЦК-решеткой на рис. 3.9 а) и гексагональной структуры графита (см. рис. 3.10).

В структуре графита атомы располагаются послойно в вершинах правильных шестиугольников и удерживаются ковалентными связями на расстоянии 1,42 Å. Эта связь, формирующая плоскую гексагональную сетку с координационным числом $z = 3$, обеспечивается тремя из четырех валентных электронов углерода. Оставшийся у каждого атома электрон участвует в межслойном взаимодействии за счет более слабых сил (типа вандер-ваальсовых), при этом расстояние между слоями возрастает до 3,35 Å. Каждый слой смещен по отношению к двум соседним (период 6,7 Å), точно повторяющим друг друга,

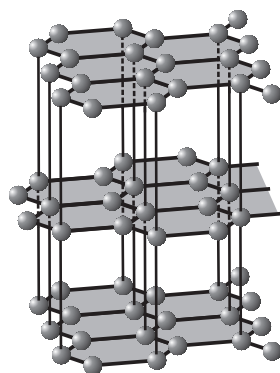


Рис. 3.10. Структура графита

¹⁾ Термин «аллотропия» по смыслу близок к полиморфизму, но применим только к *химическим элементам* и выражает их способность существовать в форме различных простых веществ, отличающихся по строению и свойствам, включая разные агрегатные состояния вещества. В отличие от этого, полиморфизм подразумевает существование лишь разных *кристаллических форм*.

на половину диагонали гексагона. В результате этого, три его атома (через один) совпадают с атомами соседних слоев, а три оставшихся — с центрами соседних гексагонов (см. рис. 3.10).

Возможность взаимного скольжения слабо связанных слоев определяет известные свойства графита — мягкость, пластичность и малую твердость. В противоположность этому, сильные межатомные связи, тетраэдрически координированные в пространстве под жестким углом $109^{\circ}28'$, обеспечивают алмазу самую высокую твердость среди известных природных материалов.

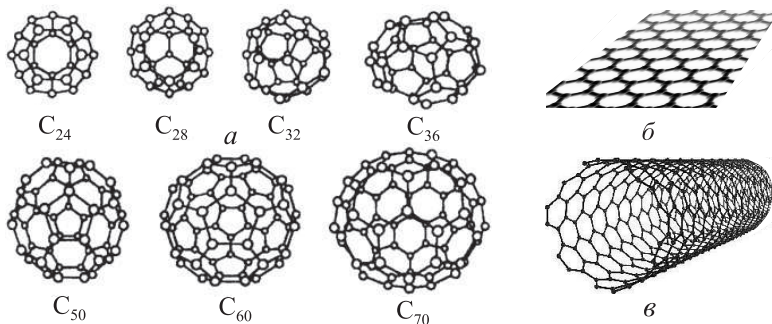


Рис. 3.11. Новые аллотропные модификации углерода: а — фуллерены, б — графен, в — углеродная нанотрубка (однослойная)

В последние два-три десятилетия были открыты и синтезированы новые аллотропные (кристаллические) формы углерода:

- *фуллерены* — сфероидные структуры, представляющие собой многоатомные углеродные молекулы C_n ($n = 24, 28, 32, \dots, 50, 60, 70, \dots$) в форме замкнутых многогранников (рис. 3.11 а);
- *графены* — планарные структуры, представляющие собой двумерный кристалл, который состоит из одиночного слоя атомов углерода, собранных в гексагональные ячейки (рис. 3.11 б);
- *углеродные нанотрубки* — цилиндрические структуры в форме гексагональных графитовых плоскостей (графенов), определенным образом свернутых в трубку (рис. 3.11 в).

Эти новые физические объекты, перспективные для нанoeлектронных применений, были подробно описаны с физических и физико-технологических позиций во введении (см. с. 44–55).

Полиморфные превращения могут происходить не только за счет фазовых переходов I рода, но и II рода. В этом случае отсутствуют изменения энтальпии и мольного объема, а кристаллическая структура меняется не скачком, а постепенно, вслед за изменением симметрии одной из фаз.

Примером может служить титанат бария BaTiO_3 , имеющий при температуре выше 120°C (точка Кюри) кубическую структуру *перовскита*, изображенную на рис. 3.12. Анионы кислорода центрируют грани куба с катионами бария в его вершинах и образуют кислородный октаэдр (базисный квадрат выделен серым цветом), центр которого занят катионом титана малого размера. Следовательно, кубическая ячейка бария Ba^{2+} является одновременно гранецентрированной (ионами O^{2-}) и объемноцентрированной (ионами Ti^{4+}). Подсчет числа ионов на одну ячейку дает необходимое стехиометрическое соотношение $\text{Ba}(8 \cdot 1/8) : \text{Ti}(1) : \text{O}(6 \cdot 1/2) =$

$= 1 : 1 : 3$, при котором сохраняется зарядовая нейтральность ячейки. Подобная ячейка имеет центр симметрии и не является полярной, что соответствует так называемой *параэлектрической фазе*. Ниже температуры Кюри катион титана смещается из центра кислородного октаэдра на $0,06 \text{ \AA}$. Несмотря на очень малое смещение, этого достаточно для того, чтобы деформировать ионную кубическую ячейку и превратить ее в тетрагональную с отношением осей $c/a = 1,01$. В результате сдвига ионов возникает асимметрия их расположения, что создает спонтанную электрическую поляризацию, характерную для *сегнетоэлектрической фазы*. Таким образом происходит фазовый переход параэлектрик $\rightarrow$ сегнетоэлектрик.

Кроме титаната бария к группе сегнетоэлектриков относятся и другие оксидные соединения (например, титанат стронция SrTiO_3) с кубической структурой перовскита в высокотемпературной (параэлектрической) фазе, которые превращаются в тетрагональную сегнетоэлектрическую фазу при температуре Кюри.

Вещества, близкие по химическому и геометрическому строению, т. е. состоящие из частиц одинаковой валентности и типа связи с примерно равными радиусами (отклонение менее 5–7%) принято называть *изоморфными*. Изоморфные вещества могут замещать друг друга в кристаллической решетке, образуя непрерывный ряд кристаллов переменного состава в форме *упорядоченных* твердых растворов замещения (см. п. 2.3). Физическими свойствами изоморфных твердых растворов можно управлять, изменяя содержание компонентов, и тем самым синтезировать материалы с нужными свойствами.

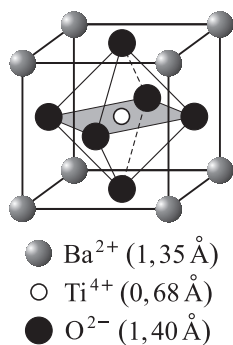


Рис. 3.12. Структура перовскита BaTiO_3 ; в скобках — ионные радиусы

В применении к сегнетоэлектрикам многообещающими для технических применений являются изоморфные твердые растворы $\text{Ba}_x\text{Sr}_{1-x}\text{TiO}_3$, сокращенно называемые BSTO. Сегнетоэлектрические пленки на их основе из-за сильной диэлектрической нелинейности и сравнительно низких потерь оказались перспективным материалом для создания различного рода устройств СВЧ-диапазона, работающих как при криогенных, так и при комнатных температурах.

В заключение отметим тесную связь анизотропии физических свойств с геометрической анизотропией структуры кристаллов. Последняя определяется в конечном счете различием межатомных связей в разных направлениях. Ярким подтверждением этого является слоистая структура графита с плотной упаковкой атомов в каждом слое, слабо связанном с соседними слоями, что позволяет им смещаться друг по отношению к другу.

Как известно, физические свойства характеризуются скалярными, векторными и тензорными величинами, которые в определенной степени отражают симметрию кристаллической структуры. Для кристаллов одной и той же сингонии разным физическим свойствам могут соответствовать тензоры разного ранга. Например, кубические кристаллы по механическим (упругим) свойствам анизотропны, а по оптическим — изотропны. В оптике все среды, описываемые симметричными тензорами второго ранга, принято разделять на изотропные, одноосные и двуосные.

Такое деление полностью совпадает с разбиением кристаллов по категориям (см. с. 347): высшая категория (кубическая сингония) — *изотропные среды*; средняя категория (тетрагональная, гексагональная и ромбоэдрическая сингонии) — *одноосные кристаллы*; низшая категория (ромбическая, моноклиновая и триклинная сингонии) — *двуосные кристаллы*. Таким образом, понижение симметрии кристаллов при переходе от высшей категории к низшей изменяет физические свойства от изотропных до двуосно-анизотропных.

При переходе от монокристаллов к поликристаллам анизотропия, характерная для кристалла, сохраняется внутри отдельных зерен, но исчезает для образца в целом. Изотропность поликристаллов вызвана угловым усреднением свойств по всем направлениям, в случае если зерна не имеют преимущественной ориентации. В противном случае, когда навязывается доминирующая ориентация (например, в процессе кристаллизации, отжига или механической прокатки), в образце вновь проявляется анизотропия в форме так называемой *текстуры* (от лат. «ткань», «строе-

ние»). Такие структуры, в отличие от кристаллических, принято называть текстурированными структурами.

На этом закончим рассмотрение модели идеального кристалла и перейдем к реальным кристаллам, дефекты которых оказывают принципиальное влияние на физические свойства.

3.4. Классификация и свойства атомных дефектов кристаллической структуры

Классификацию дефектов обычно проводят по размерному признаку, а именно по числу измерений макроскопической протяженности дефекта. Принято различать:

1) *точечные дефекты* (нульмерные), вызванные нарушениями структуры, локализованными в ее отдельных точках, с размерами микроскопического масштаба, не превышающими нескольких периодов решетки по каждому из трех измерений;

2) *линейные дефекты* (одномерные), имеющие макроскопическую протяженность лишь в одном измерении на расстояния, много бóльшие периода решетки, с поперечными размерами микроскопического масштаба (порядка нескольких межатомных расстояний) — краевые и винтовые дислокации;

3) *поверхностные дефекты* (двумерные), имеющие макроскопическую протяженность, соизмеримую с размерами кристалла, одновременно в двух измерениях — границы зерен (блоков), двойников, межфазные границы, дефекты упаковки и внешние поверхности кристалла;

4) *объемные дефекты* макроскопического масштаба (трехмерные) — трещины, полости и включения инородной фазы.

По энергетическому состоянию принято различать равновесные и неравновесные дефекты. Точечные дефекты образуются, в частности, как результат тепловых колебаний решетки и поэтому существуют даже в равновесных условиях, находясь в термодинамическом равновесии с кристаллической структурой. При облучении кристаллов фотонами и частицами высоких энергий появляются неравновесные точечные дефекты, называемые *радиационными дефектами*. Линейные, поверхностные и объемные дефекты также относятся к неравновесным, поскольку обычно они возникают в неравновесных условиях, наряду с точечными дефектами, например при кристаллизации и термической или механической обработке материалов.

1. Точечные дефекты представляют собой нарушение периодичности структуры в отдельных точках кристаллической

решетки, достаточно удаленных друг от друга. Область силового воздействия дефекта на окружающие его частицы распространяется на расстояния порядка нескольких периодов решетки.

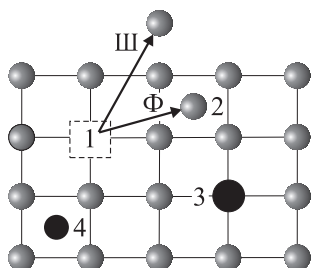


Рис. 3.13. Основные типы точечных дефектов: 1 — вакансия, 2 — междоузельный атом, 3 — примесь замещения, 4 — примесь внедрения

Простейшими точечными дефектами являются *вакансии* (незанятые узлы решетки) и *междоузельные атомы*, располагающиеся в тетраэдрических или октаэдрических пустотах решетки. При этом междоузелье может занимать как собственный атом (ион), так и чужеродный, называемый *примесью внедрения*. Если чужеродный атом (ион) внедряется на место вакансии, то его называют *примесью замещения*, так как он заменяет основную частицу в узле решетки. Схематическое изображение вышеперечисленных дефектов приведено на рис. 3.13.

Понятно, что в достаточно плотноупакованных структурах (ГПУ, ГЦК и ОЦК) внедрение атомов в междоузлия (пустоты структуры) наиболее вероятно для примесных частиц малых размеров. Именно так ведут себя в металлах внедренные малые атомы H, C, Si, O, N, образуя химические соединения — гидриды, карбиды, силициды, оксиды, нитриды.

В полупроводниковых кристаллах со сравнительно рыхлой структурой типа алмаза или сфалерита в тетраэдрических и октаэдрических междоузлиях структуры довольно легко размещаются не только примеси, но и собственные атомы кристалла. В этом случае переход атома из узла в междоузелье создает *парный дефект*, называемый **дефектом Френкеля**.

Механизм генерации собственных дефектов по Френкелю показан на рис. 3.13 стрелкой Ф. Стрелка Ш соответствует **дефекту Шоттки**, который возникает по механизму дефектообразования, предложенному Шоттки для кристаллов с плотной упаковкой. По Шоттки вакансия возникает как *одиночный дефект* в результате выхода атома на поверхность (с возможным последующим испарением его в газовую среду). Возникшая в приповерхностном слое вакансия диффундирует в объем кристалла путем встречного движения атомов по вакансиям. При генерации дефектов по механизму Френкеля плотность кристалла остается неизменной, а по механизму Шоттки — незначительно снижается из-за пренебрежимо малого увеличения объема кристалла.

Собственные дефекты структуры, возникшие в кристаллах по механизмам Френкеля¹⁾ и Шоттки²⁾, часто называют *разупорядочением по Френкелю и по Шоттки*.

Для образования вакансии атому необходимо разорвать связи со своим окружением, на что требуется энергия порядка единиц электронвольт. Например, для Ge она составляет 2 эВ (по 0,5 эВ на каждую тетраэдрическую связь), а для Si — 2,3 эВ. При этом увеличивается внутренняя энергия кристалла (на величину ΔU), но одновременно и возрастает энтропия (на величину ΔS) из-за увеличения беспорядка в расположении частиц. Противоборство энергетического и энтропийного факторов обеспечивает равновесную концентрацию дефектов из условия минимума свободной энергии, когда $\Delta F \equiv \Delta U - T\Delta S = 0$ (анализ концентрации дефектов по Френкелю и Шоттки см. в п. 3.10).

В ионных или в ковалентных полярных кристаллах равновесная концентрация собственных дефектов (вакансий и междоузельных атомов) управляется не только требованием термодинамического равновесия, но и *условием электронейтральности* кристалла. Это условие обеспечивается образованием равного количества положительно и отрицательно заряженных дефектов. При этом междоузельный катион (анион) имеет положительный (отрицательный) заряд, а катионная (анионная) вакансия (как отсутствие соответствующего заряда в узле) несет эффективный отрицательный (положительный) заряд.

Для обеспечения электронейтральности дефекты по Шоттки в ионных кристаллах возникают как *парный дефект*: вакансии одновременно появляются в катионной и анионной подрешетках с равными концентрациями (если нет причин для нарушения стехиометрии химического соединения). Образование дефектов по

¹⁾ Френкель Яков Ильич (1894–1952) — советский физик-теоретик, специалист в области физики твердого тела, ядерной физики, кинетической теории, сформулировал основные квантово-механические концепции электропроводности металлов, движения атомов и ионов в кристаллах (дефекты по Френкелю) и поглощения света в диэлектриках (экситон Френкеля), впервые объяснил природу ферромагнетизма обменным взаимодействием и создал теорию доменного строения ферромагнетиков. Автор первого отечественного учебника по теоретической физике.

²⁾ Шоттки (Schottky) Вальтер (1886–1976) — немецкий физик, специалист в области термодинамики, физики твердого тела, вакуумной и полупроводниковой электроники, открыл зависимость тока эмиссии от электрического поля (эффект Шоттки), изучал дефекты в полупроводниках (дефекты по Шоттки) и протекание тока через контакт «металл–полупроводник» (барьер Шоттки), широко применяемый в современных приборах (диод Шоттки и полевой транзистор Шоттки).

Френкелю возможно лишь в одной (например, катионной) подрешетке, так как положительный заряд катиона в междоузлии компенсируется эффективным отрицательным зарядом его вакансии. Из-за большого размера аниона дефекты по Френкелю в анионной подрешетке менее вероятны, чем в катионной.

Заряженные собственные дефекты, в первую очередь вакансии, играют первостепенную роль в ионных кристаллах диэлектриков щелочно-галогидного типа со структурой каменной соли NaCl (см. рис. 3.8 а). Именно они определяют основные электрические и оптические свойства диэлектриков. В основе поведения таких дефектов лежит их взаимодействие с подвижными носителями заряда (электронами и дырками), возникающими для обеспечения электронейтральности кристалла. Известным примером являются так называемые *центры окраски* в виде точечных структурных дефектов, обладающих собственной частотой оптического поглощения в видимом диапазоне и изменяющих в результате этого окраску кристалла.

Простейшим из центров окраски является *F-центр* (от нем. *Farbenzentrum*), схематически изображенный на рис. 3.14. Он представляет собой анионную (галогидную) вакансию, которая, будучи положительно заряженной, для нейтрализации своего эффективного заряда захватывает электрон (показанный на нижнем рисунке в октаэдрическом окружении катионами). Ионизация F-центра путем отрыва электрона при освещении видимым светом приводит к избирательному поглощению на той частоте, которая соответствует энергии связи электрона с вакансией. Как результат этого, бесцветный щелочно-галогидный кристалл становится окрашенным. Так, кристаллы NaCl и KCl имеют максимум поглощения на длинах волн $\lambda = 465$ нм (синяя область спектра) и $\lambda = 563$ нм (зеленая область спектра), что и придает этим кристаллам, соответственно, желто-коричневую и фиолетовую окраску (в дополнительных цветах к цвету линии поглощения).

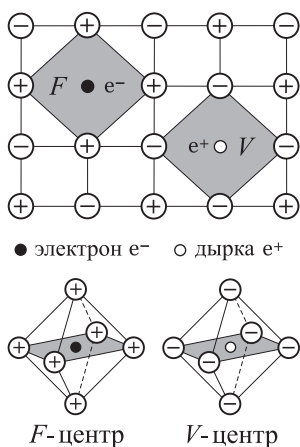


Рис. 3.14. Схема центров окраски в щелочно-галогидных кристаллах типа NaCl

Появление F-центров вызвано отжигом щелочно-галогидных кристаллов в парах щелочного металла (термодинамику образования дефектов при отжиге см. в п. 3.11). Важно отметить, что

указанные спектральные полосы не зависят от того, какой щелочной металл (Na, K или другой) использован для создания паров. Это подтверждает вакансионную природу F -центра как собственного точечного дефекта в ионном кристалле.

Кроме F -центра (электронного центра окраски) возможен и дырочный центр окраски, называемый V -центром. Он изображен на рис. 3.14 в виде катионной вакансии, захватившей дырку, которая находится в октаэдрическом окружении анионами. Такой дефект создается путем отжига щелочно-галоидного кристалла в атмосфере галогена. Кроме простых (одиначных) центров наблюдаются и сложные центры окраски, составленные из простых и обозначаемые буквами F' , M , N , R .

Влияние точечных дефектов (собственных и примесных) на физические свойства кристаллов рассмотрено ниже в п. 3.5.

2. Линейные дефекты представляют собой нарушение структуры, простирающееся вдоль некоторой линии на расстояние, сравнимое с размерами образца. Поперечный размер линейного дефекта не превышает нескольких межатомных расстояний.

Линейные дефекты могут создаваться в форме цепочки междоузельных атомов, называемой *кроддионом* (от англ. *crowd*), которые, однако, являются нестабильными образованиями.

К устойчивым линейным дефектам относятся *дислокации*, созданные обрывом или сдвигом атомных слоев. Это чаще всего является результатом механического воздействия на материал, приводящего к пластическому сдвигу в форме некоторой линии дислокации. В зависимости от ориентации этой линии по отношению к вектору сдвига (иначе называемому вектором Бюргерса) различают *краевые* и *винтовые дислокации*, схематически изображенные на рис. 3.15.

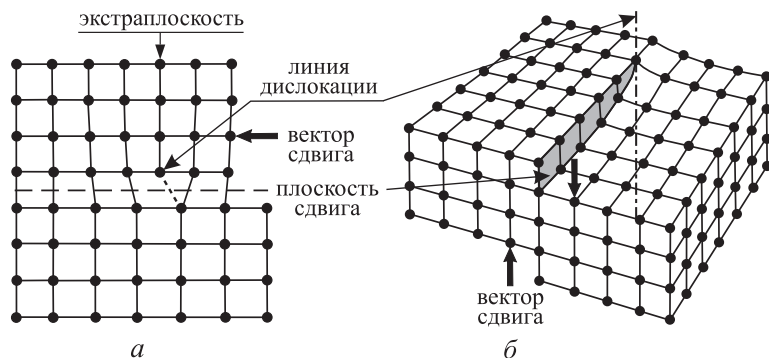


Рис. 3.15. Линейные дефекты: краевая дислокация (а), винтовая дислокация (б)

Краевая дислокация (рис. 3.15 а) образована экстраплоскостью — лишней атомной полуплоскостью, вдвинутой между плоскостями идеального кристалла. Край этой плоскости и образует *линию краевой дислокации* в направлении, нормальном вектору сдвига, вызывающего смещение одной части кристалла (верхней на рисунке) от другой по плоскости сдвига, перпендикулярной плоскости рисунка. Под действием сдвигающего усилия происходит перемещение экстраплоскости на межатомное расстояние путем разрыва связи по плоскости сдвига, что показано пунктирной линией на рис. 3.15 а. Вокруг дислокационной линии число ближайших атомов может отличаться от координационного числа данной структуры. Однако уже на расстоянии нескольких периодов от центра структурные искажения практически исчезают. Поэтому искажения сосредоточены вблизи дислокационной линии, где и запасается энергия упругих деформаций.

Винтовая дислокация (рис. 3.15 б) также возникает при сдвиге одной части кристалла относительно другой. Но при этом, в отличие от краевой дислокации, линия винтовой дислокации параллельна вектору сдвигу. Здесь отсутствует экстраплоскость, а процесс образования винтовой дислокации можно представить в виде разреза межатомных связей и сдвига вдоль плоскости разреза. Особенность винтовой дислокации заключается в том, что она состоит не из параллельных атомных плоскостей, а как бы из одной плоскости, закрученной по винту. Ось винта и есть *линия винтовой дислокации*. Направление закрутки может быть как лево-, так и правовинтовым, а шаг винта может составлять от одного до нескольких межатомных расстояний. Выход дислокации на поверхность заканчивается незарастающей ступенью, которая играет важную роль в механизме спирального роста кристаллов (см. п. 5.15). Как и в случае краевой дислокации, искажения сосредоточены вблизи дислокационной линии и зависят от расстояния до центра вращения (оси дислокации).

Дислокации (краевые и винтовые) могут возникать только под действием внешних напряжений. Они не появляются в идеальном кристалле при однородном нагреве, поскольку энергия тепловых колебаний решетки существенно меньше энергии их генерации. Эта величина на 1 см длины линии дислокации по оценкам составляет около $5 \cdot 10^{-11}$ Дж, что в пересчете на одно межатомное расстояние дает примерно $2,5 \cdot 10^{-18}$ Дж (≈ 15 эВ!).

Характеристикой качества кристаллов является *плотность дислокаций*, определяемая как число дислокационных линий, пересекающих единичную площадку в кристалле. Обычными мето-

дами кристаллизации из расплава получают плотность дислокаций 10^4 – 10^8 см $^{-2}$. Путем отжига она снижается до 10^3 – 10^4 см $^{-2}$ и увеличивается на несколько порядков в результате пластической деформации металлов (при плотности дислокаций 10^8 см $^{-2}$ общая протяженность дислокационных линий в 1 мм ^{3} кристалла составляет около 1 км!). Наиболее совершенные полупроводниковые кристаллы имеют плотность дислокаций 10 – 10^2 см $^{-2}$.

3. Поверхностные дефекты образованы любой границей толщиной в несколько межатомных расстояний, которая разделяет области с различной кристаллографической ориентацией. К поверхностным дефектам относят границы кристаллитов (зерен, фрагментов, блоков), двойники и дефекты упаковки.

Большинство веществ в естественном состоянии представляют собой поликристаллы, состоящие из *зерен* с размерами в доли миллиметра, ориентированных относительно друг друга произвольным образом с углом разориентации вплоть до 20° . Границы между такими зернами, называемые *большеугловыми*, запасают избыточную энергию упругих деформаций до 10^{-4} Дж/см ^{2} , что сравнимо с теплотой плавления, отнесенной к одному атому.

Зерна могут быть разделены на *фрагменты* размерами в десятки раз меньшими зерен и с разориентацией в единицы градусов, а фрагменты состоят из отдельных *блоков*, создающих *мозаичную структуру*. Такая структура с размерами блоков до 10^{-5} – 10^{-2} см и с углами разориентации в минуты или секунды присуща также и монокристаллам. В пределах каждого блока кристаллическая решетка почти идеальна (если не учитывать точечные дефекты), а границы между ними, называемые *малоугловыми*, построены из краевых дислокаций, как показано на рис. 3.16 а. Энергия малоугловой границы примерно равна сумме энергий дислокаций, составляющих эту границу.

В процессе роста кристаллов и при механических воздействиях возможно образование *двойниковых структур* типа изображенной на рис. 3.16 б. Составляющие двойник кристаллы могут быть получены один из другого путем зеркального отражения в плоскости двойникового или при повороте на 180° вокруг некой двойниковой оси (не показанной на рисунке). Каждая атомная плоскость, параллельная плоскости двойникового, смещается под действием силы **F** вдоль самой себя на расстояние, пропорциональное ее удалению от плоскости зеркального отражения. При этом каждая атомная плоскость в области двойникового отстоит от предыдущей на одну и ту же часть периода решетки. Плоскостями двойникового обычно служат кристаллографические

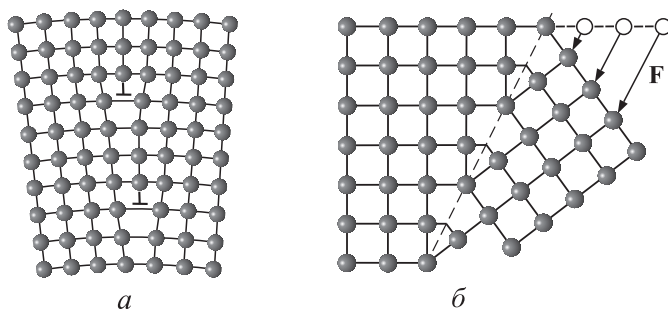


Рис. 3.16. Поверхностные дефекты: *a* — малоугловая граница зерен, построенная из краевых дислокаций, линии которых отмечены знаком $\perp$; *б* — двойниковая структура, правая часть которой, полученная зеркальным отражением левой части в плоскости двойникового (штриховая линия), возникла как результат смещения атомов (светлые кружки) под действием силы $\mathbf{F}$

плоскости с малыми индексами Миллера, например, (112) для ГЦК-кристаллов и $(10\bar{1}2)$ для ГПУ-кристаллов.

К поверхностным дефектам относятся и так называемые *дефекты упаковки*, представляющие собой нарушения в порядке чередования атомных слоев в плотноупакованных структурах. Например, в гранецентрированной кубической решетке исходная последовательность заполнения слоев $ABC\ ABC\ \dots$ может быть нарушена в виде последовательности $ABC\ AB\ AB\ ABC\ \dots$. Чередование $AB\ AB$ характерно для гексагональной плотноупакованной структуры. Здесь дефект упаковки представляет собой тонкую прослойку ГПУ, вставленную в ГЦК-структуру. Изменение взаимного расположения слоев сохраняет координационное число. Поэтому энергетические различия структур, содержащих дефекты упаковки, невелики. Значит энергия таких дефектов тоже мала, например, по сравнению с теплотой плавления: обычно она на полтора-два порядка меньше энергии межзеренных границ.

Поверхностные дефекты, в частности межкристаллитные границы, оказывают влияние на физические свойства материалов — механические, электрические, оптические и т. д. Наличие границ приводит к тому, что диффузия примесей в поликристаллах заметно эффективнее, чем в монокристаллах.

Границы между зернами характеризуются большей реакционной способностью, и зачастую отличаются по химическому составу от самих зерен и блоков. Нередко здесь располагаются прослойки различных примесей, которые могут находиться не только в кристаллическом, но и в стеклообразном состоянии, что характерно, например, для керамических материалов.

3.5. Точечные дефекты и физические свойства кристаллов

Точечными дефектами, как уже отмечалось, являются несовершенства периодической структуры, область силового влияния которых простирается по всем направлениям на расстояния в несколько периодов решетки. К точечным дефектам относятся:

1) **собственные дефекты** кристаллической структуры:

- *вакансии* в виде незанятых атомами или ионами узлов кристаллической решетки;
- *междоузельные атомы* в виде собственных атомов кристалла, находящихся в междоузельных пустотах решетки;
- *антиструктурные дефекты* кристаллических соединений (например, бинарных химических соединений MX) в виде собственных атомов кристалла, занимающих несвойственные им узлы чужой подрешетки (M_X — атом M на месте атома X или X_M — атом X на месте атома M);

2) **примесные дефекты** кристаллической структуры:

- *примеси замещения* в виде чужеродных атомов, заменяющих атомы в узлах решетки или заполняющих ее вакансии;
- *примеси внедрения* в виде чужеродных атомов, занимающих междоузлия решетки или заменяющих собственные междоузельные атомы.

В результате хаотического распределения дефектов с увеличением их концентрации может возникать взаимодействие между ними, что приводит к появлению комплексов, образованных несколькими дефектами и называемых *ассоциатами*. По своим электронным свойствам они, как правило, отличаются от исходных простых дефектов, но по размерному признаку также относятся к разряду точечных. Примером ассоциатов могут служить *бивакансии*, образующиеся в результате спаривания электронов двух соседних вакансий.

Устойчивое местоположение дефекта в решетке определяется двумя факторами — энергетическим (как результат стремления системы к минимуму свободной энергии) и геометрическим (как необходимость выполнения определенных соотношений между эффективными размерами дефекта и соответствующего для него места в кристаллической решетке).

Энергетический фактор зависит от соотношения сил притяжения и отталкивания, обусловленного близостью электронных конфигураций дефекта и атомов решетки. По этой причине примеси замещения в соединении MX стремятся занять узлы более

близких к ним по химической природе атомов решетки: атомы металлов занимают узлы М, а атомы неметаллов — узлы Х. *Геометрический фактор* может стать определяющим для атомов внедрения (примесных или собственных) в плотноупакованных решетках (гранецентрированной кубической и гексагональной), где объем междоузельных пустот относительно невелик.

Электронная структура дефекта определяет его электрическое поведение в кристалле. Если в дополнение к необходимому для закрепления в решетке числу валентных электронов дефект (примесный или собственный) имеет один или несколько избыточных электронов (локализованных около этого дефекта и слабо связанных с ним), тогда в результате тепловых колебаний решетки эти электроны могут быть оторваны от своего центра и обобществлены кристаллом в виде подвижных *электронов проводимости*. Такие дефекты называют *донорами*, а создаваемую ими проводимость — *электронной проводимостью* или проводимостью *n*-типа.

Если для закрепления в решетке дефекту (примесному или собственному) недостает собственных валентных электронов, то недостающие электроны он забирает из валентных связей между основными атомами кристалла с образованием одной или нескольких положительно заряженных пустот в этих связях, называемых *дырками*. Такие дефекты называют *акцепторами*, а создаваемую ими проводимость — *дырочной проводимостью* или проводимостью *p*-типа.

Примесь, способную выполнять функции как донора, так и акцептора в зависимости от ее положения в решетке одного и того же кристалла, принято называть *амфотерной примесью*.

Энергия, необходимая: а) для отрыва электрона от донорного атома D с возникновением подвижного электрона проводимости e^- и неподвижного положительного иона D^+ и б) для присоединения электрона к акцепторному атому A с возникновением подвижной дырки e^+ (часто обозначаемой как h^+ от англ. *hole*) и отрицательного иона A^- , называется *энергией ионизации* донорного или акцепторного центра. Эти величины будем обозначать, соответственно, как ΔE_D или ΔE_A .

Можно сформулировать следующие общие правила, позволяющие в большинстве случаев предсказывать электронные свойства точечных дефектов.

Атомы замещения (примесные и собственные в виде антиструктурных дефектов) действуют как *доноры*, если их валентность больше числа валентных электронов замещаемого атома,

и как *акцепторы* — в противном случае. Например, в полупроводниковых кристаллах Si и Ge (элементы IV группы) атомы элементов V группы (P, As, Sb) и III группы (B, Al, Ga) являются, соответственно, донорами и акцепторами. В то же время атомы Si и Ge являются амфотерными примесями для полупроводникового соединения GaAs (типа $A^{III}B^V$), выполняя функции донора при размещении в подрешетке галлия и акцептора при размещении в подрешетке мышьяка. В кристалле GaAs антиструктурный дефект Ga_{As} является двухзарядным акцептором, а дефект As_{Ga} — двухзарядным донором.

Атомы внедрения (примесные и собственные), занимающие междоузельные положения, действуют как *доноры*, если их внешняя электронная оболочка заполнена меньше, чем наполовину (что характерно для металлов), и как *акцепторы* — в противном случае (что характерно для неметаллов). Междоузельных доноров существует сравнительно много (Li в Si, Ge, GaSb, ZnO; Cu в PbS и др.), тогда как междоузельных акцепторов встречается в природе довольно мало.

Вакансии действуют как *доноры*, если число неспаренных электронов вблизи вакансии меньше половины числа валентных электронов, имеющихся в аналогичном месте совершенного кристалла, и как *акцепторы* — в противном случае. Например, в кристалле GaAs насыщенная связь между Ga и As создается с помощью восьми валентных электронов. В результате вакансии галлия, окруженная пятью неспаренными электронами атомов мышьяка, для связывания с ними может принимать электроны, заряжаясь отрицательно и выполняя функции акцептора. Наоборот, вакансии мышьяка, окруженная тремя неспаренными электронами атомов галлия, является донором.

Как уже ранее отмечалось, в большинстве случаев зависимость физических свойств реального кристалла от дефектов оказывается настолько сильной, что в конечном итоге эти свойства определяются не столько исходной структурой твердого тела, сколько концентрацией и типом дефектов. В частности, сюда относятся такие дефекто-чувствительные свойства, как электропроводность, люминесценция и термоэлектронная эмиссия, кратко описываемые ниже.

1. Электрические свойства полупроводников (в частности, электропроводность) определяются в основном поведением *мелких примесей*, энергия ионизации которых (ΔE_D для доноров и ΔE_A для акцепторов) сравнима с тепловой энергией $k_B T$ (равной 0,026 эВ при 300 K). Это обеспечивает практически

полную тепловую ионизацию мелких примесей. Для описания электронных состояний в кристалле используют энергетические зонные диаграммы, изображенные на рис. 3.17.

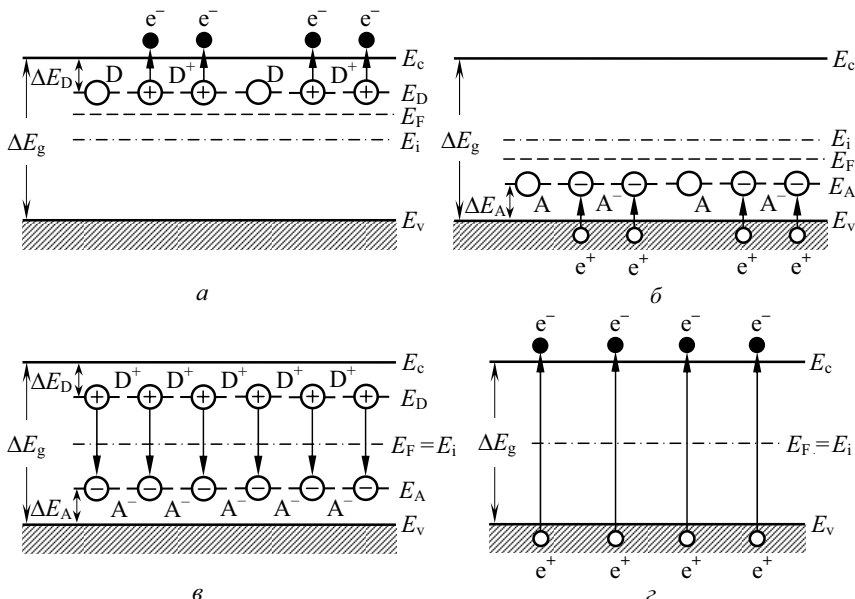


Рис. 3.17. Энергетические зонные диаграммы, построенные для полупроводников n -типа (а) и p -типа (б), для компенсированных полупроводников (в) и для собственных полупроводников (г)

Как известно, точечные дефекты (примесные и собственные) создают *локализованные* электронные состояния — донорные и акцепторные, в отличие от *нелокализованных* состояний, соответствующих подвижным электронам и описываемых функциями Блоха¹⁾. На зонной диаграмме нелокализованным состояниям соответствуют зона *проводимости* для электронов и *валентная зона* для дырок, разделенные областью энергий, запрещенных для нелокализованных состояний, шириною ΔE_g (индекс от англ. *gap*). В *запрещенной зоне* располагаются энергетические

¹⁾ Блох (Bloch) Феликс (1905–1983) — швейцарско-американский физик, основоположник современных методов физики твердого тела, заложил основы зонной теории кристаллов и теории ферромагнетизма, впервые ввел понятие спиновых волн, разработал основы теории и осуществил первые эксперименты по исследованию ядерного магнитного резонанса. Нобелевская премия по физике (совместно с Э. Перселлом, 1952) «за создание новых точных методов ядерных магнитных измерений и связанные с ними открытия».

уровни, отвечающие только локализованным состояниям. При этом одному и тому же дефекту может принадлежать несколько локальных уровней — по числу электронов, которыми дефект в состоянии обменивается с зоной проводимости или с валентной зоной. Это так называемые *многозарядные центры*: например, примесь Au в Ge может находиться в одном из зарядовых состояний $+1e$, $-1e$, $-2e$, $-3e$, каждому из которых в запрещенной зоне соответствует один уровень донорного типа (при заряде $+1e$) или акцепторного типа (при зарядах $-1e$, $-2e$, $-3e$).

Мелкая донорная примесь создает в запрещенной зоне локальный уровень с энергией E_D вблизи дна зоны проводимости E_C (рис. 3.17 а). Аналогично мелкая акцепторная примесь создает локальный уровень с энергией E_A вблизи потолка валентной зоны E_V (рис. 3.17 б). Тепловая ионизация примесных центров изображена на рис. 3.17 а и б направленными вверх стрелками, которые соответствуют электронным переходам:

а) для *полупроводников n-типа* — с уровня донора E_D в зону проводимости с образованием донорного иона D^+ и электрона проводимости e^- (рис. 3.17 а),

б) для *полупроводников p-типа* — из валентной зоны на уровень акцептора E_A с образованием акцепторного иона A^- и подвижной дырки e^+ (рис. 3.17 б).

Процессы, связанные с тепловым перебросом электронов на вышележащие уровни (что изображено стрелками, идущими снизу вверх), в равновесном состоянии уравниваются обратными переходами, что отображается стрелками, направленными сверху вниз (которые не показаны на рис. 3.17 а и б). Такие переходы изображены на рис. 3.17 в для так называемых *компенсированных полупроводников*, в которых донорные и акцепторные центры присутствуют примерно в одинаковых концентрациях. В этом случае электроны, стремясь к минимуму энергии, самопроизвольно переходят от донорных атомов к акцепторным (на энергетической диаграмме переходы сверху вниз, т. е. с донорного уровня на акцепторный), превращая их в ионы. При полной компенсации примесей отсутствуют подвижные носители заряда, порожденные ионизацией доноров и акцепторов. По характеру изменения концентрации электронов и дырок в зависимости от температуры компенсированные полупроводники близки к *собственным* (бездефектным) *полупроводникам*, зонная диаграмма которых приведена на рис. 3.17 г (без указания обязательных в условиях равновесия электронных переходов сверху вниз).

В собственных полупроводниках единственным источником генерации подвижных носителей заряда является *межзонное*

возбуждение. Так называется процесс «вырыва» электрона из валентной связи между основными атомами кристалла. Ему соответствуют стрелки, идущие вверх из валентной зоны в зону проводимости на рис. 3.17 *г*, что характеризует парное возникновение электронов и дырок при межзонном возбуждении. Самопроизвольный процесс, обратный межзонному возбуждению, называют *межзонной рекомбинацией*. Понятно, что тепловое межзонное возбуждение имеет место во всех без исключения полупроводниках. Тогда на рис. 3.17 *а*, *б* и *в* должны быть дополнительно изображены такие же стрелки, как и на рис. 3.17 *г*.

Глубина залегания энергетического уровня в запрещенной зоне отсчитывается от верхней (для доноров) или нижней (для акцепторов) ее границы и равняется энергии ионизации примесного центра: $\Delta E_D = E_c - E_D$ и $\Delta E_A = E_A - E_v$. При этом для мелких примесей ΔE_D и $\Delta E_A \ll \Delta E_g$, где $\Delta E_g = E_c - E_v$ — ширина запрещенной зоны, равная энергии, необходимой для вырыва электрона из валентной связи кристалла в результате процесса межзонного возбуждения. Для таких полупроводников, как германий, кремний и арсенид галлия, при $T = 300$ К имеем: $\Delta E_g = 0,67$ эВ (Ge), $1,12$ эВ (Si) и $1,43$ эВ (GaAs). В то же время глубина залегания уровня для атомов фосфора Р и бора В (наиболее часто применяемых для легирования кремния в качестве донора и акцептора) равна $\Delta E_D = 0,044$ эВ (Р в Si) и $\Delta E_A = 0,046$ эВ (В в Si). В арсениде галлия энергия ионизации практически всех мелких доноров очень мала ($\Delta E_D \approx 0,006$ эВ¹⁾), и они полностью ионизованы при комнатной температуре.

Как известно, физика твердого тела дает следующие выражения для равновесной концентрации электронов и дырок в невырожденных полупроводниках:

$$n = N_c \exp\left(-\frac{E_c - E_F}{k_B T}\right) \quad \text{и} \quad p = N_v \exp\left(-\frac{E_F - E_v}{k_B T}\right), \quad (3.5.1)$$

где k_B — постоянная Больцмана и E_F — уровень Ферми. Величины N_c и N_v называются эффективной плотностью электронных состояний в зоне проводимости и в валентной зоне и определены соответственно как

$$N_{c(v)} = 2 \left(\frac{m_{n(p)} k_B T}{2\pi \hbar^2} \right)^{3/2}, \quad (3.5.2)$$

¹⁾ В физике твердого тела энергию частиц обычно измеряют в электрон-вольтах, при этом $1 \text{ эВ} = 1,6 \cdot 10^{-19} \text{ Дж}$, что в пересчете на один моль (путем умножения на постоянную Авогадро $N_A = 6,022 \cdot 10^{23} \text{ частица/моль}$) дает соотношение: $1 \text{ эВ/частица} = 96,485 \text{ кДж/моль} = 23,06 \text{ ккал/моль}$.

где $m_{n(p)}$ — эффективная масса электрона (дырки), $\hbar$ — постоянная Планка¹⁾, равная $1,05 \cdot 10^{-34}$ Дж·с.

Положение уровня Ферми²⁾ внутри запрещенной зоны находится из условия электронейтральности полупроводника:

$$n + N_{A-} = p + N_{D+}, \quad (3.5.3)$$

где N_{D+} и N_{A-} — концентрация ионизованных донорных и акцепторных центров.

Равенства (3.5.1) позволяют записать следующее выражение для уровня Ферми:

$$E_F = E_c - k_B T \ln \frac{N_c}{n} = E_v + k_B T \ln \frac{N_v}{p}. \quad (3.5.4)$$

Отсюда ясно видно, что в собственном полупроводнике (отмечаемом нижним индексом i от англ. *intrinsic*) или полностью компенсированном полупроводнике, когда $n = p \equiv n_i$, уровень Ферми проходит близко к середине запрещенной зоны, в частности точно посередине при равных эффективных массах электрона и дырки. Действительно, согласно (3.5.2), для частного случая, когда $m_n = m_p$, имеем $N_c = N_v$, тогда из равенства (3.5.4) получаем $E_F = E_i \equiv (E_c + E_v)/2$.

Равенства (3.5.4) показывают, что уровень Ферми смещается от середины запрещенной зоны: а) *вверх*, приближаясь к дну зоны проводимости E_c при введении донорной примеси, когда увеличивается n ; б) *вниз*, приближаясь к потолку валентной зоны E_v при введении акцепторной примеси, когда увеличивается p . Для многих полупроводников, в том числе для Ge, Si и GaAs, мелкие донорные и акцепторные примеси практически полностью ионизованы при комнатной температуре, что обеспечивает соответствующую электропроводность n -типа или p -типа.

¹⁾ Планк (Planck) Макс Карл Эрнст Людвиг (1858–1947) — выдающийся немецкий физик, основатель квантовой теории, давший новое направление развитию физики с начала XX века, ввел квант действия (1900) и, исходя из идеи квантования, получил закон излучения, названный его именем; широко известен также трудами по термодинамике, теории относительности и философии естествознания. Нобелевская премия по физике (1918) «в знак признания его заслуг в деле развития физики благодаря открытию квантов энергии».

²⁾ Ферми (Fermi) Энрико (1901–1954) — виднейший итальянский физик, автор многочисленных фундаментальных работ в области квантовой и статистической теории, физики элементарных частиц; участник и руководитель теоретических и экспериментальных работ по нейтронной физике и ядерной энергетике, участвовал в создании первой атомной бомбы. Нобелевская премия по физике (1938) «за открытие искусственной радиоактивности, вызванной медленными нейтронами».

При сильном легировании полупроводников описанная выше ситуация существенно изменяется. Дело в том, что локальность примесных уровней в запрещенной зоне сохраняется до тех пор, пока примеси хаотически распределены в кристалле на достаточно больших расстояниях друг от друга, т. е. при сравнительно малой их концентрации. В этом случае каждому типу примесей соответствует свой энергетический уровень в запрещенной зоне с N -кратным вырождением, где N — концентрация примесей рассматриваемого типа. С ростом N расстояния между примесными атомами уменьшаются и может возникнуть перекрытие волновых функций электронов (или дырок), локализованных на соседних донорных (или акцепторных) центрах. Это приводит к снятию энергетического вырождения и к расщеплению примесных уровней в так называемую «примесную подзону». Поскольку такая подзона образуется в результате взаимодействия хаотически распределенных атомов примеси, то движению электрона в ней нельзя ставить в соответствие волну Блоха, что отличает примесную подзону от разрешенных зон кристалла. При дальнейшем увеличении числа примесей примесная подзона сливается с основной зоной, что соответствует *сильно легированным полупроводникам*. Это происходит при $N_D \approx 2 \cdot 10^{17} \text{ см}^{-3}$ для Ge, при $N_D \approx 8 \cdot 10^{18} \text{ см}^{-3}$ для Si, при $N_D \approx 1,5 \cdot 10^{16} \text{ см}^{-3}$ для GaAs.

Наряду с дефектами, создающими мелкие уровни, в полупроводниках существуют и многозарядные дефекты, образующие, как правило, *глубокие уровни*. Эти уровни играют роль *центров прилипания* (или *ловушек захвата*) и *центров рекомбинации* (или *рекомбинационных ловушек*). Первые располагаются относительно близко от границ запрещенной зоны и могут захватывать только носители одного знака с последующим их тепловым высвобождением (электронные и дырочные центры прилипания). Вторые, располагаясь вблизи середины запрещенной зоны, попеременно захватывают носители обоих знаков, обеспечивая их рекомбинацию (излучательную или безызлучательную).

2. Электрические свойства диэлектриков с ионной структурой определяются, в первую очередь, не примесями, как в полупроводниках, а собственными дефектами (вакансиями и междоузельными атомами) при условии *нестехиометрического состава* кристаллического соединения. Строго говоря, нестехиометричность кристаллов более естественна в процессе их роста, нежели стехиометричность. Точный стехиометрический состав неорганических соединений в принципе является скорее исключением, чем правилом. В особой мере это относится к оксидным

соединениям металлов, которые при равновесии с окружающей средой чаще всего нестехиометричны.

Нестехиометричностью можно управлять с помощью отжига кристалла в парах собственных элементов, составляющих его кристаллическую структуру (см. п. 3.11). Такой отжиг изменяет соотношение между концентрациями собственных дефектов в катионной и анионной подрешетках. А так как эти дефекты несут эффективный заряд, то электронейтральность нестехиометрического соединения восстанавливается за счет изменения электронно-дырочного равновесия в кристалле. Иными словами, в результате отжига изменяется тип проводимости изначально стехиометрического (или близкого к стехиометрии) непроводящего (или слабо проводящего) диэлектрика.

В зависимости от характера отклонения от стехиометрии могут возникать два типа дефектных структур, характерных в первую очередь для оксидов металлов или близких к ним по структуре халькогенидов, записанных в виде MX_2 или MX (см. п. 3.6).

Кристаллы с недостатком анионов X^{2-} (в частности, кислорода O^{2-}) реализуются в двух вариантах:

а) как результат *удаления* из решетки атомов X с преобладанием положительно заряженных *анионных вакансий*, что принято записывать в виде $\text{MX}_{2-\delta}$ или $\text{MX}_{1-\delta}$;

б) как результат *введения* в решетку атомов M с преобладанием положительно заряженных *междоузельных катионов*, что принято записывать в виде $\text{M}_{1+\delta}\text{X}_2$ или $\text{M}_{1+\delta}\text{X}$.

В обоих вариантах в кристалле появляются свободные электроны, обеспечивая *электронную проводимость n-типа*. Дело в том, что элементы M и X вводятся и удаляются в атомарной форме, а в решетке они присутствуют в форме ионов.

В первом варианте анионная вакансия для своей нейтрализации удерживает два электрона в катионном окружении (подобно F -центру на рис. 3.14). Эти слабо удерживаемые электроны за счет тепловых колебаний решетки отрываются от вакансии и становятся электронами проводимости. Следовательно, анионная вакансия выполняет функцию *донора*.

Во втором варианте введенному в кристалл атому M для размещения в междоузлии геометрически выгодно уменьшить свой размер путем потери валентных электронов и превращения в положительный ион M_1^{2+} (см. обозначения в п. 3.6). Следовательно, междоузельный атом катиона является *донором*.

Примером ионных соединений, которые могут обнаруживать *электронную проводимость* как следствие дефицита аниона

(типа кислорода) или избытка катиона (металла), являются следующие: BaO, BeO, CdO, TiO₂, ZnO, ZrO₂, Nb₂O₅, CdS и др.

Кристаллы с недостатком катионов M²⁺ реализуются в двух возможных вариантах:

а) как результат *удаления* из решетки атомов М с преобладанием отрицательно заряженных *катионных вакансий*, что принято записывать в виде M_{1-δ}X₂ или M_{1-δ}X;

б) как результат *введения* в решетку атомов Х с преобладанием отрицательно заряженных *междоузельных анионов*, что принято записывать в виде MX_{2+δ} или MX_{1+δ}.

В обоих вариантах в кристалле должны появиться свободные дырки, обеспечивающие *дырочную проводимость p-типа*, по аналогичной, как и ранее, причине, вызванной тем фактом, что элементы М и Х удаляются и вводятся в кристалл обязательно в атомарной форме.

В первом варианте катионная вакансия для своей нейтрализации удерживает две дырки в анионном окружении (подобно V-центру на рис. 3.14). Эти слабо удерживаемые дырки за счет тепловых колебаний решетки отрываются от вакансии и обеспечивают проводимости *p-типа*. Следовательно катионная вакансия выполняет функцию *акцептора*.

Второй вариант с междоузельным анионом маловероятен из-за большого размера аниона. Однако в некоторых случаях, например в кристаллах UO_{2+δ}, методом дифракции нейтронов был обнаружен сложный дефект, состоящий из двух междоузельных ионов кислорода, связанных с кислородной вакансией (каждый междоузельный ион кислорода вытесняет ближайший кислородный ион из узла в соседнее междоузлие).

Для оксидов переходных металлов типа Fe, Ni, Co и др., имеющих переменную валентность, более вероятен другой механизм компенсации отрицательного заряда избыточных вакансий двухвалентного катиона M²⁺. Для этого отдельные ионы металла в узлах решетки переходят в состояние с более высоким зарядом. Например, в окисле FeO с преобладанием вакансий двухвалентного железа ионы Fe²⁺ меняют свою валентность: Fe²⁺ → Fe³⁺. Таким образом, каждая вакансия двухвалентного иона железа с эффективным зарядом -2 компенсируется двумя трехвалентными ионами Fe³⁺ в узлах решетки. Избыточный положительный заряд этих ионов притягивает электрон из соседней основной связи двухвалентного железа с кислородом, создавая там дырку.

Хорошо известными ионными соединениями с дефектами структуры типа катионных вакансий, обеспечивающих *дыроч-*

ную проводимость, являются следующие: FeO, NiO, CoO, MnO, Ag₂O, Cu₂O, Cu₂S, FeS и др.

Оба типа рассмотренных дефектов (как при дефиците анионов, так и при дефиците катионов) могут возникать только на поверхности, контактирующей с окружающей газовой средой, а также при взаимодействии с высшим/низшим окислом или металлом, например на границе раздела металл–окисная пленка.

3. Люминесцентные свойства кристаллов определяются дефектами (примесными или собственными) с глубоко залегающими энергетическими уровнями, поскольку их энергия ионизации или возбуждения (E) близка к кванту $\hbar\omega$ «рабочего» оптического излучения. Вынужденные электронные переходы с нижележащих уровней (в том числе из валентной зоны) на вышележащие уровни (включая зону проводимости), сопровождаемые генерацией неравновесных носителей, происходят при поглощении оптического кванта с энергией $E = \hbar\omega$ ¹⁾. Обратные спонтанные переходы возбужденных электронов обеспечивают соответствующее некогерентное оптическое излучение, избыточное над тепловым и называемое люминесценцией.

Люминесценция как спонтанный процесс некогерентного излучения отличается от когерентного индуцированного (лазерного) излучения, вызванного электронными переходами с уровней *инверсной населенности*, искусственно создаваемой с помощью так называемой *накачки*. В отличие от лазерного излучения, индуцированного светом, послесвечение при люминесценции вызывается тепловыми колебаниями решетки. Такое зависимое от температуры длительное во времени послесвечение принято называть *фосфоресценцией*, а соответствующие ей материалы — *фосфорами* (или *кристаллофосфорами*). Кратковременное излучение, прекращающееся практически мгновенно (за времена менее 10^{-9} с) после возбуждения, называют *флуоресценцией*.

Кроме *фотолюминесценции*, вызванной оптическим возбуждением, возможны и другие виды люминесценции: *катодолюминесценция* — при возбуждении быстрыми электронами (катодными лучами), *рентгенолюминесценция* — при возбуждении квантами рентгеновского излучения и *электролюминесценция* — при

¹⁾ Энергия кванта ($E = \hbar\omega$, измеренная в электронвольтах) и длина электромагнитной волны в вакууме ($\lambda = 2\pi c/\omega$, измеренная в микрометрах) связаны соотношением $E\lambda = 2\pi\hbar c = 1,24$ эВ·мкм. Следовательно, диапазон длин волн видимого света от $\lambda = 0,76$ мкм (красный цвет) до $\lambda = 0,38$ мкм (фиолетовый цвет) соответствует диапазону энергий от $E = 1,63$ эВ до $E = 3,26$ эВ (ширина запрещенной зоны полупроводников и диэлектриков).

возбуждении электрическим полем (например, полем p – n -перехода с рекомбинацией инжектированных носителей заряда).

Независимо от способа возбуждения, обеспечивающего неравновесную концентрацию электронов в зоне проводимости или появление электронов на верхних возбужденных уровнях дефекта, называемого *активатором*, структура люминесцентного излучения определяется исключительно обратными излучательными переходами. В ряде случаев для повышения эффективности люминесценции активатора в кристаллофосфóры добавляют другую примесь, называемую *сенсibiliзатором*. Она является центром поглощения возбуждающего света с последующей резонансной передачей части поглощенной энергии активатору как излучающему *центру люминесценции*.

Переход возбужденного электрона из зоны проводимости в валентную зону или на локальный уровень активатора дает люминесценцию, называемую *рекомбинационной*. Типичными представителями кристаллофосфóров с рекомбинационной люминесценцией являются халькогениды цинка и кадмия, например, ZnS , ZnSe , CdS , CdSe и другие соединения типа $A^{II}B^{VI}$, а также соединения типа $A^{III}B^V$, такие как GaAs , GaP , InAs , InP , GaN и др.

Если излучение связано с электронными переходами между основным и возбужденными уровнями, лежащими *внутри* запрещенной зоны кристалла и принадлежащими одному и тому же дефекту, называемому *центром свечения*, то такая люминесценция относится к *внутрицентральной* (или *нерекомбинационной*). Подобные люминофоры имеют значительно большую ширину запрещенной зоны, чем рекомбинационные люминофоры, поскольку внутри запрещенной зоны должны размещаться энергетические уровни центров свечения.

Типичными кристаллофосфóрами с внутрицентральной люминесценцией являются щелочно-галогидные кристаллы NaCl , KCl и другие соединения типа $A^I B^{VII}$. Роль центра свечения в них могут исполнять *центры окраски* типа F -центра или V -центра, изображенные на рис. 3.14. Дефектами с внутрицентральной люминесценцией являются также примеси редкоземельных элементов (Er , Yb и др.) в германии, кремнии и в соединениях типа $A^{III}B^V$.

Как было сказано, основой кристаллофосфóров служат сульфиды, селениды, теллуриды Zn , Cd , оксиды Ca , Mn , щелочно-галогидные и некоторые другие соединения. В качестве активаторов используют ионы металлов (Mn , Cu , Ag , Co , Eu , Tl и т. д.).

Кристаллофосфóры представляют собой твердые растворы активатора в кристаллической структуре люминофора. В большинстве случаев такие растворы являются растворами замещения,

в образовании которых важную роль играет валентность активатора. По этому признаку все активаторы делят на две группы:

- *изовалентные активаторы*, валентность которых совпадает с валентностью заменяемых атомов люминофора: ZnS:Mn с заменой Zn^{2+} на Mn^{2+} ; $\text{Al}_2\text{O}_3:\text{Cr}$ с заменой Al^{3+} на Cr^{3+} и т. п.

- *иновалентные активаторы* с валентностью, отличной от валентности атомов люминофора: ZnS:Cu или ZnS:Ag , где ионы Cu^+ или Ag^+ занимают место Zn^{2+} в узле решетки люминофора.

При изовалентном замещении проблемы не возникают, и концентрация активатора может достигать десятых долей процента.

Иновалентное замещение создает электрически заряженные дефекты. Поэтому образование твердого раствора возможно лишь при наличии так называемых *соактиваторов* — дополнительных вводимых примесей, создающих *компенсирующие дефекты* для обеспечения электронейтральности.

Принцип компенсации заряда/валентности рассмотрим на примере цинк-сульфидного люминофора ZnS:Cu с одновалентными атомами меди в качестве активатора (зеленая люминесценция). В структуре сфалерита ион Cu^+ на месте Zn^{2+} в узле создает нехватку положительного заряда, т. е. эффективный отрицательный заряд -1 . Поскольку при активации примесью кристалл остается электрически нейтральным, то появившийся локальный заряд дефекта Cu_{Zn}^- (см. обозначения в п. 3.6) должен быть скомпенсирован зарядом противоположного знака. Для осуществления этого принципиально возможны три варианта.

1. *Компенсация анионными вакансиями* возможна потому, что отсутствие иона серы (S^{2-}) свидетельствует об эффективном *положительном* заряде $+2$ его вакансии. Поэтому одна анионная (серная) вакансия компенсирует заряды двух примесных дефектов активатора Cu_{Zn}^- . Однако образование вакансий энергетически менее выгодно, чем введение компенсирующих примесей.

2. *Компенсация трехвалентными катионами*, такими как Al^{3+} , Ga^{3+} , In^{3+} , осуществима потому, что замещение ими иона Zn^{2+} приводит к появлению эффективно *положительного* однозарядного дефекта (Al_{Zn}^+ , Ga_{Zn}^+ , In_{Zn}^+) в узле решетки. Следовательно, концентрации примесей активатора (Cu или Ag) и соактиватора (Al, Ga или In) должны быть равными.

3. *Компенсация одновалентными анионами* является наиболее эффективной для цинк-сульфидных люминофоров ZnS:Cu и ZnS:Ag с применением ионов Cl^- в роли компенсирующего соактиватора. Действительно, замена двухзарядного иона S^{2-} однозарядным ионом Cl^- дает эффективный *положительный* заряд

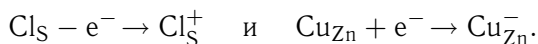
в анионном узле в форме точечного дефекта Cl_S^+ . Следовательно, полная компенсация заряда достигается при равных концентрациях примесей активатора (Cu или Ag) и соактиватора (Cl или других галогенов).

Вышеприведенный анализ компенсации зарядов основывался на рассмотрении полупроводникового соединения типа $\text{A}^{\text{II}}\text{B}^{\text{VI}}$ в пределе чисто ионной связи, хотя в сфалерите ZnS связь в большей степени ковалентная, нежели ионная. В этом случае рассуждения строятся на *принципе компенсации валентности*.

Применим этот принцип к варианту компенсации активатора Cu атомами галогена Cl. Встраивание в узлы ковалентного кристалла ZnS равного числа атомов меди и хлора эквивалентно растворению в нем соединения активатора CuCl. При образовании твердого раствора, несмотря на *гетеровалентное замещение*, не нарушается тетраэдрическое расположение атомов в структуре сфалерита и не возникают ни вакансии, ни междоузельные атомы. Этому способствует как сходный тип связи, так и близость соответствующих ковалентных радиусов в соединениях ZnS и CuCl ($r_{\text{Zn}} = 1,31 \text{ \AA}$, $r_{\text{S}} = 1,04 \text{ \AA}$ и $r_{\text{Cu}} = 1,35 \text{ \AA}$, $r_{\text{Cl}} = 0,99 \text{ \AA}$).

Для возникновения ковалентной связи Zn–S два валентных электрона от атома серы должны быть переданы двухвалентному атому цинка. В результате этого оба атома Zn и S получают по четыре электрона и приобретают способность к образованию направленных связей тетраэдрического типа (см. структуру сфалерита на рис. 3.9 б).

При замене Zn одновалентным атомом Cu возникает дефицит электрона в ковалентной связи Cu–S, т. е. атом Cu_{Zn} (активатор) является *акцептором* для ZnS. Замена шестивалентного атома S на атом Cl (соактиватора), имеющего семь валентных электронов, порождает *донор* Cl_S в ZnS. При равных концентрациях акцептора (активатора) и донора (соактиватора) возникает *компенсированный полупроводник*, зонная диаграмма которого приведена на рис. 3.17 в. В таком полупроводнике не происходит тепловой ионизации донора и акцептора, как в случае мелких примесей. Вместо этого «лишний» электрон от донора переходит к акцептору, что записывают в форме квазихимических реакций:



Процесс возникновения иона активатора Cu_{Zn}^- (играющего роль центра свечения) и иона соактиватора Cl_S^+ (восстанавливающего электронейтральность кристалла) энергетически выгоден, так как происходит с понижением энергии в силу перехода элек-

трона при компенсации сверху вниз (см. рис. 3.17 в): с донорного уровня Cl_S на акцепторный уровень Cu_{Zn} .

Таким образом, конечный результат компенсирующего действия соактиватора получается одинаковым вне зависимости от того, какая связь преобладает — ионная или ковалентная.

На рис. 3.18 приведена схема расположения энергетических уровней для цинк-сульфидного люминофора, активированного медью (ZnS:Cu, Cl — слева) и серебром (ZnS:Ag, Cl — справа). В обоих случаях роль соактиватора выполняет хлор (принято записывать после двоеточия активатор и соактиватор через запятую).

Акцепторный дефект Cu_{Zn}^- или Ag_{Zn}^- , созданный активатором (Cu или Ag) и обладающий электроном в отсутствие возбуждения, является *центром свечения*. Созданный соактиватором (атомом Cl) донорный дефект Cl_S^+ в отсутствие электрона играет роль ловушки или *центра захвата*. Наличие или отсутствие электрона на соответствующем уровне изображено черным или белым кружком на рис. 3.18.

Стрелки показывают *электронные переходы*: 1 — при поглощении возбуждающего света или любом другом возбуждении с переходом электрона в зону проводимости; 2 — при люминесцентном излучении, вызванном обратным переходом электрона; 3 — при захвате электрона ловушкой; 4 — при *тепловом* освобождении электрона из ловушки.

Кроме активаторного поглощения (стрелки 1), возможно и *фундаментальное поглощение* с межзонным электронным переходом (не показанным на рисунке), если квант возбуждения превышает ширину запрещенной зоны ($\Delta E_g = 3,54$ эВ для ZnS). В этом случае результатом поглощения фотона является рождение электронно-дырочных пар. Дырки довольно быстро захватываются активаторным центром Cu_{Zn}^- или Ag_{Zn}^- , нейтрализуя его (точнее, превращая в двухзарядный ион Cu^{2+} , аналогичный основному иону Zn^{2+}). После этого активатор (с захваченной дыркой) приобретает способность к рекомбинационному захвату электрона, ранее переведенного в зону проводимости.

Показанное на рисунке энергетическое положение уровней Cu_{Zn}^- и Ag_{Zn}^- обеспечивает зеленую люминесценцию (с центром

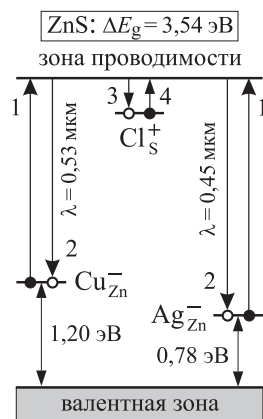


Рис. 3.18. Зонная диаграмма люминесценции цинк-сульфидного кристаллофосфора

полосы при $\lambda = 0,53$ мкм) и синюю люминесценцию (с центром полосы при $\lambda = 0,45$ мкм) для ZnS:Cu и ZnS:Ag кристаллофосфоров, соответственно. Полосы люминесценции не зависят от природы соактиватора, который, будучи центром захвата возбужденных электронов, определяет время жизни электрона на ловушке и тем самым изменяет лишь длительность послесвечения (фосфоресценции).

4. Термоэлектронная эмиссия с поверхности диэлектриков и полупроводников (и в меньшей степени металлов) также определяется дефектностью кристаллической структуры. Как известно, плотность тока термоэлектронной эмиссии дается *формулой Ричардсона–Дэшмана*¹⁾:

$$i = \overline{D} A T^2 \exp\left(-\frac{\Phi}{k_B T}\right). \quad (3.5.5)$$

Здесь $A = e m k_B^2 / 2\pi^2 \hbar^3 \approx 120 \text{ A}/(\text{см}^2 \cdot \text{K}^2)$ — постоянная Ричардсона (e и m — заряд и масса электрона в вакууме); $\overline{D}$ — средний коэффициент прозрачности поверхностного барьера, зависящий от формы барьера и состояния поверхности (для металлов изменяется в широких пределах: $\overline{D} = 0,02-0,5$).

Величина $\Phi = E_0 - E_F$ представляет собой *работу выхода* электрона из поверхности твердого тела, определяемую как разность между уровнем вакуума (E_0) и уровнем Ферми (E_F). *Уровень вакуума* соответствует энергии электрона, покоящегося в вакууме, то есть находящегося вне тела за пределами области действия его приповерхностных полей.

На энергетических диаграммах (см. рис. 3.19) уровень вакуума E_0 расположен выше дна зоны проводимости E_c на расстоянии χ , называемом *электронным сродством* (или сродством к электрону, или *внешней работой выхода*), т. е. $E_0 = E_c + \chi$. Величина $\chi > 0$ является параметром кристалла, не зависящим от наличия дефектов и различным лишь для разных его граней. Следовательно, работа выхода может быть представлена как

$$\Phi \equiv E_0 - E_F = (E_0 - E_c) + (E_c - E_F) = \chi + (E_c - E_F). \quad (3.5.6)$$

¹⁾ Ричардсон (Richardson) Оуэн Уилланс (1879–1959) — английский физик, впервые (1902) установил зависимость между температурой металла и скоростью эмиссии электронов с его поверхности (закон Ричардсона), занимался также эмиссией электронов под действием химических реагентов, фотоэлектрическим эффектом, спектроскопией и взаимодействием рентгеновского излучения с веществом. Нобелевская премия по физике (1928) «за работы по термодинамическим исследованиям и особенно за открытие закона, носящего его имя». С. Дэшман квантово-механически обосновал (1927) закон Ричардсона.

Отсюда видно, что при $\chi = \text{const}$ работа выхода снижается с уменьшением разности $(E_c - E_F)$, то есть согласно (3.5.4), с увеличением концентрации дефектов *донорного типа* (собственных или примесных) в объеме полупроводника/диэлектрика.

Однако формула (3.5.6) не учитывает изгиба зонных границ в приповерхностном слое. Для этого надо различать положение дна зоны проводимости внутри объема полупроводника (E_{cb} , индекс b от англ. *bulk*) и на поверхности (E_{cs} , индекс s от англ. *surface*). Работа выхода определяется положением дна зоны проводимости по отношению к уровню Ферми именно на поверхности. Поэтому формула (3.5.6) должна быть переписана в виде

$$\Phi = \chi + (E_{cs} - E_F), \quad (3.5.7)$$

где электронное средство приведено к поверхности (см. рис. 3.19):

$$\chi = E_0 - E_{cs}. \quad (3.5.8)$$

Уровень Ферми E_F определяет *электрохимический потенциал* электронов ($\tilde{\mu}_e$). При отсчете от дна зоны проводимости вдали от поверхности (E_{cb}), где отсутствует электрическое поле и зоны плоские, он может быть записан как

$$\tilde{\mu}_e = E_{cb} - E_F. \quad (3.5.9)$$

В условиях равновесия уровень Ферми, а значит и электрохимический потенциал остаются неизменными по системе (см. (1.16.3)):

$$\tilde{\mu}_e(x) = \text{const}. \quad (3.5.10)$$

В приповерхностной области, где зоны искривлены ($E_c(x) \neq \text{const}$), по аналогии с (3.5.9) можно ввести *химический потенциал* электронов как разность

$$\mu_e(x) = E_c(x) - E_F \quad (3.5.11)$$

(при любом направлении оси x , не показанной на рис. 3.19 а и б, с началом координат $x = 0$ на правом краю рисунка).

Химический потенциал, в отличие от электрохимического, может изменяться, так как его неоднородность компенсируется

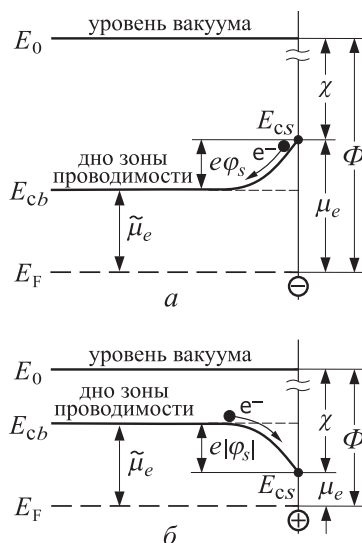


Рис. 3.19. Зонные диаграммы при наличии поверхностных состояний акцепторного типа (а) и донорного типа (б)

неоднородным распределением электрического потенциала, когда $\varphi(x) \neq \text{const}$, обеспечивая постоянство электрохимического потенциала во всех точках системы. Действительно, согласно определению (1.16.2), электрохимический и химический потенциалы электронов связаны друг с другом соотношением (в расчете на один электрон, $z_e = -1$, $\mathcal{F} \rightarrow e$):

$$\tilde{\mu}_e \equiv \mu_e(x) + z_e e \varphi(x) = \mu_e(x) - e \varphi(x) = \text{const}. \quad (3.5.12)$$

В частности, для границы кристалла при $x = 0$, где $E_c(0) = E_{cs}$ и $\varphi(0) = \varphi_s$, как показано на рис. 3.19 а и б, из формул (3.5.11) и (3.5.12) вытекают следующие соотношения:

$$\mu_e = E_{cs} - E_F \quad \text{и} \quad \tilde{\mu}_e = \mu_e - e \varphi_s. \quad (3.5.13)$$

Подстановка соотношений (3.5.13) в уравнение (3.5.7) приводит к искомому выражению для работы выхода (см. рис. 3.19):

$$\Phi = \chi + \mu_e = (\chi + \tilde{\mu}_e) + e \varphi_s. \quad (3.5.14)$$

Отсюда видно, что на работу выхода электронов с поверхности полупроводника оказывают влияние три фактора:

а) кристаллографическая ориентация поверхности, определяющая величину *электронного сродства* $\chi = E_0 - E_{cs}$;

б) соотношение между количеством дефектов донорного и акцепторного типов (собственных и примесных) в объеме кристалла, определяющее положение уровня Ферми в запрещенной зоне, т. е. *электрохимический потенциал* электронов $\tilde{\mu}_e = E_{cb} - E_F$;

в) состояние поверхности полупроводника, определяющее искривление зон, т. е. *поверхностный потенциал* $\varphi_s = (E_{cs} - E_{cb})/e$.

При фиксированной ориентации поверхности ($\chi = \text{const}$) снижения работы выхода, как уже отмечалось, можно добиться за счет объемных свойств материала при преобладающем влиянии доноров над акцепторами с целью уменьшения вклада $\tilde{\mu}_e$ в Φ .

Как видно из сравнения рис. 3.19 а и б, третий вклад ($e \varphi_s$) в величину работы выхода (3.5.14), порожденный поверхностными состояниями (ПС), может изменять знак:

$$e \varphi_s = \begin{cases} E_{cs} - E_{cb} > 0 & \text{— акцепторные ПС (рис. 3.19 а),} \\ E_{cs} - E_{cb} < 0 & \text{— донорные ПС (рис. 3.19 б).} \end{cases} \quad (3.5.15)$$

Акцепторные ПС принимают электроны из объема кристалла и заряжают поверхность отрицательно (знак минус в кружке $\ominus$ на рис. 3.19 а). Это создает в приповерхностной области обедненный электронами слой, так что $\mu_e > \tilde{\mu}_e$, зоны изгибаются вверх, и работа выхода возрастает на величину $e \varphi_s$.

Донорные ПС, наоборот, отдают электроны в объем кристалла и заряжают поверхность положительно (знак плюс в кружке $\oplus$ на рис. 3.19 б). Это создает в приповерхностной области обогащенный электронами слой, так что $\mu_e < \tilde{\mu}_e$, зоны изгибаются вниз, и работа выхода уменьшается на величину $e|\varphi_s|$.

Движение электронов (e^-) под действием приповерхностного электрического поля, приводящее к образованию обедненного и обогащенного слоев, показано на рис. 3.19 а и б стрелками вдоль искривленной границы зоны проводимости.

Следовательно, как и при объемных дефектах, для эффективной термоэмиссии предпочтительнее создавать на поверхности донорные центры, снижающие работу выхода электронов. Изгиб энергетических зон характеризуется поверхностным потенциалом с типичным значением $|\varphi_s| \approx (0,1 - 0,2)$ В. Из сравнения этой величины с возможными изменениями уровня Ферми в полупроводниках/диэлектриках с запрещенной зоной в единицы электронвольт ясно, что определяющее влияние на работу выхода и термоэмиссию оказывают объемные дефекты материала.

В качестве термоэмиттирующего покрытия наиболее широко используются оксидные слои на металлической подложке (никелевом керне), называемые оксидным катодом. Оксидное покрытие представляет собой твердый раствор окислов бария (BaO), стронция (SrO) и кальция (CaO) с преобладанием BaO. Оксид бария является диэлектриком, имеющим $\Delta E_g \approx 3,8$ эВ и $\chi \approx 0,9$ эВ, так что в отсутствие дефектов $\Phi_{\max} = \chi + \Delta E_g/2 \approx 2,8$ эВ. Следовательно, для уменьшения работы выхода надо создавать дефекты донорного типа.

Кристалл BaO имеет структуру типа каменной соли NaCl (см. рис. 3.8 а с заменой однозарядных ионов Na^+ и Cl^- на двухзарядные — катион Ba^{2+} и анион O^{2-}). Оксид бария относится к группе ионных соединений с недостатком анионов, т. е. имеющих электронный тип проводимости (см. с. 386).

В реальных условиях кристалл оксида бария содержит катионные и анионные вакансии, среди которых преобладающими по концентрации должны быть анионные (кислородные) вакансии, проявляющие донорные свойства. Это достигается введением избыточных атомов бария, заполняющих катионные вакансии и отдающих два своих валентных электрона соседней вакансии кислорода. Такая вакансия в подрешетке аниона (кислорода), нейтрализованная двумя электронами, полученными от атома бария, аналогична F -центру в щелочно-галогенидных соединениях (см. рис. 3.14). Различие лишь в числе нейтрализующих электронов — один для NaCl (KCl, RbCl) и два — для BaO (SrO, CaO).

Следовательно, желательными дефектами донорного типа в оксиде бария являются двухзарядные F -центры, имеющие два донорных уровня, как показывает эксперимент, с глубиной залегания $\Delta E_{D1} \approx 1,4$ эВ и $\Delta E_{D2} \approx 2,3$ эВ. Понятно, что любые примеси акцепторного типа и вещества, связывающие свободный барий, отравляют оксидный катод, т. е. снижают его эмиссию. Поэтому в керн катода добавляют активирующие присадки (типа Mg, Al, Si и др.), восстанавливающие потерянный барий, например, по реакции $BaO + Mg \rightarrow MgO + Ba$.

Преобладание анионных вакансий над катионными на практике достигается в результате специальной термообработки катода, называемой *активировкой*. Во время активировки атомы бария, появляющиеся при разложении оксида ($BaO \rightarrow Ba + \frac{1}{2}O_2$), заполняют катионные вакансии, вследствие чего их концентрация уменьшается. В результате процесса активировки уровень Ферми поднимается вверх, обеспечивая работу выхода оксидного катода, равную $\Phi = (1,5-1,8)$ эВ. Это приводит к плотности тока термоэмиссии около 1 А/см^2 при температурах $(700-800)^\circ\text{С}$.

Поверхностные состояния электронов, упоминаемые выше (рис. 3.19), впервые были получены из решения одноэлектронного уравнения Шредингера независимо друг от друга Таммом¹⁾ и Шокли (см. сноску на с. 18) в рамках модели *идеальной поверхности*. Идеализация поверхности подразумевает ограничение кристалла плоскостью с сохранением того расположения атомов, которое было характерно для бесконечного кристалла.

Распределение потенциала $V(x)$ в одномерном кристалле с периодом a , изображено на рис. 3.20 *а* слева от плоскости $x = x_0$, соответствующей обрыву решетки. Такое распределение потенциала допускает аналитическое решение краевой задачи на собственные значения, построенной на основе одноэлектронного уравнения Шредингера.

Не вдаваясь в детали решения квантово-механической задачи, отметим, что в этом случае возникают два типа решений, соответствующих (см. рис. 3.20 *б* и *в*):

¹⁾ Тамм Игорь Евгеньевич (1895–1971) — советский физик-теоретик в области квантовой механики, физики твердого тела, теории излучения и ядерной физики, ввел понятие фононов и произвел квантование акустических волн, построил квантовую теорию фотоэффекта и рассеяния света в кристаллах, доказал существование особых состояний электрона на поверхности кристаллов (уровни Тамма), создал (совместно с И.М. Франком) теорию излучения «сверхсветового» электрона в среде, объясняющую природу излучения Вавилова–Черенкова. Нобелевская премия по физике (совместно с И.М. Франком и П.А. Черенковым, 1958) «за открытие и истолкование эффекта Черенкова».

а) *объемным состояниям* с волновыми функциями $\psi(x)$, имеющими постоянную амплитуду в объеме кристалла (при $x < x_0$) и экспоненциально затухающими вне кристалла (при $x > x_0$);

б) *поверхностным состояниям* с волновыми функциями $\psi(x)$, локализованными в районе поверхности, т. е. экспоненциально затухающими как вне кристалла (при $x > x_0$), так и внутри его (при $x < x_0$ с осцилляциями, уменьшающимися при удалении от границы $x = x_0$).

Распределение электронной плотности в форме зависимости $|\psi(x)|^2$ для объемных и поверхностных состояний качественно показано на рис. 3.20 б и в. Поверхностным состояниям соответствуют энергетические уровни в запрещенной зоне полупроводника, называемые уровнями Тамма или Шокли.

Детальный анализ локализованных энергетических уровней, полученных в свое время Таммом (1932 г.) и Шокли (1939 г.), показывает, что между ними нет принципиального различия: состояния (уровни) Тамма получаются при $x_0 = 0$ (обрыв решетки в плоскости атомов), а состояния (уровни) Шокли — при $x_0 = a/2$ (обрыв решетки по середине между атомами). Исторически было принято различать эти два типа состояний, приписывая им разные физические причины образования. По современным воззрениям, состояния Тамма порождены нарушением периодичности кристаллической решетки в районе поверхности, а состояния Шокли являются следствием изменения химической связи между атомами в поверхностном слое. Однако в обоих случаях поверхность кристалла считается идеальной, т. е. лишенной дефектов, что, естественно, является модельным приближением.

На поверхности реального кристалла всегда присутствуют, во-первых, структурные дефекты (в том числе, вызванные перестройкой атомов на идеальной поверхности в форме *релаксации и реконструкции поверхности*, см. п. 5.3), и, во-вторых, адсорбированные частицы. Эти поверхностные неоднородности, имея ненасыщенные валентные связи, могут обмениваться носителями заряда с объемом кристалла. Им также принято приписывать энергетические уровни, отличные от уровней Тамма и Шокли.

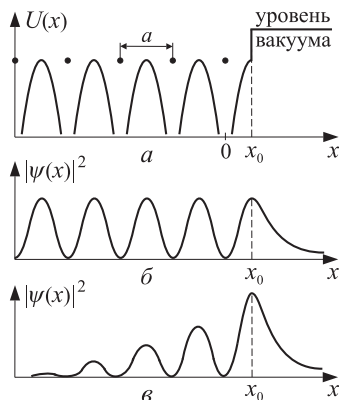


Рис. 3.20. Одномерный потенциал полубесконечной решетки (а) и качественные картины распределения электронной плотности для объемных (б) и поверхностных (в) состояний

Поверхностные состояния распределены по ширине запрещенной зоны с плотностью от 10^{10} до 10^{14} эВ $^{-1}$ ·см $^{-2}$ (на единственный интервал энергии и на единицу площади поверхности). Именно подобного рода донорные и акцепторные ПС определяют физические и физико-химические поверхностные свойства полупроводников, образуя описанные выше приповерхностные слои, обогащенные или обедненные носителями заряда (см. уравнение (3.5.15) и рис. 3.19 а и б).

В настоящее время наиболее исследованной является поверхность кремния (Si) и его граница раздела с собственным окислом (Si–SiO $_2$), широко используемые в МОП-технологии современных процессоров (см. введение).

Как известно, идеально чистая поверхность полупроводника, полученная сколом кристалла в сверхвысоком вакууме (с давлением менее 10^{-9} мм рт. ст.), содержит атомы с незаполненными валентными связями. Эти связи склонны захватывать электроны и, следовательно, ведут себя подобно акцепторам, заряжая поверхность отрицательно с плотностью электронов около 10^{15} см $^{-2}$. Действительно, для идеальной поверхности (100) в структуре алмаза (см. рис. 3.9 а) квадратная грань $a \times a$ содержит два атома (один в центре грани и $4 \cdot 1/4$ в ее вершинах). Отсюда поверхностная плотность атомов равна $2/a^2$, что для Si (с периодом решетки $a = 5,43$ Å) дает $N_{\text{Si}}^{(100)} \approx 6,78 \cdot 10^{14}$ см $^{-2}$.

В реальных условиях поверхность кремния покрыта слоем диоксида SiO $_2$ (естественного или искусственно созданного), так что поверхностные состояния возникают в районе интерфейса Si–SiO $_2$. В этом случае ковалентные связи Si замыкаются на тетраэдры SiO $_2$, создавая поверхностные состояния донорного типа, заряжающие интерфейс положительно и изгибающие зоны вниз (см. рис. 3.19 б). Эти ПС, по поведению близкие к состояниям Тамма–Шокли, принято называть *быстрыми состояниями*, так как для них времена обмена зарядами с объемом полупроводника измеряется микросекундами. Эксперимент показал, что плотность быстрых ПС на границе раздела Si–SiO $_2$ приблизительно равна $9 \cdot 10^{10}$ см $^{-2}$, $2 \cdot 10^{11}$ см $^{-2}$ и $5 \cdot 10^{11}$ см $^{-2}$, соответственно, для плоскостей (100), (110) и (111). Это дает преимущество кристаллографической ориентации [100] по сравнению с другими.

Кроме быстрых ПС, в районе интерфейса Si–SiO $_2$ возникают также *медленные состояния*, локализованные в переходном слое внутри SiO $_2$ толщиной около 100 Å. Здесь имеется избыток ионов кремния (а также подвижных ионов Na $^+$ и т. п.), создающих положительный пространственный заряд с поверхностной

плотностью около 10^{11} ион/см². Медленные ПС имеют весьма большие времена релаксации (вплоть до месяцев и лет), что является причиной долговременного дрейфа параметров приборов.

3.6. Квазихимический метод описания дефектов

В реальных условиях количество точечных дефектов (собственных и примесных) в полупроводниках относительно невелико, т. е. их концентрация много меньше числа основных атомов в единице объема. Так, для кремния имеем $N_{\text{Si}} = 5 \cdot 10^{22}$ см⁻³, в то время как концентрация легирующих примесей обычно на несколько порядков ниже (от 10^{12} – 10^{13} см⁻³ для предельно очищенных материалов с естественным уровнем остаточных примесей до 10^{19} – 10^{20} см⁻³ для сильно легированных полупроводников). На этом основании реальные кристаллы полупроводников можно рассматривать как твердые растворы, близкие по свойствам к предельно разбавленным растворам.

Следовательно, для анализа поведения дефектов в полупроводниках можно использовать термодинамические соотношения из теории растворов (см. п. 2.4 и 2.3) с тем, чтобы процессы, связанные с образованием и исчезновением дефектов, описывать при помощи уравнений, аналогичных по форме с уравнениями для обычных химических реакций (см. п. 2.12–2.14). Такой подход к описанию дефектообразования в кристаллах, базирующийся на основных положениях химической термодинамики, называют *квазихимическим методом*.

Квазихимический метод в применении к полупроводникам требует использования специальных обозначений для структурных элементов кристалла и их концентраций (записанных путем заключения соответствующего элемента в квадратные скобки):

M — элементарный полупроводник (например, Ge и Si);

MX — бинарный полупроводник типа A^{II}B^{VI} и A^{III}B^V (например, ZnS, GaAs и др.);

M_I и X_I — междоузельные атомы M и X (например, Ga_I и As_I);

[M_I] и [X_I] — концентрация междоузельных атомов M и X (например, [Ga_I] и [As_I]);

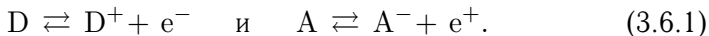
M_I⁺ и X_I⁻ — ионизованные междоузельные атомы M и X (например, Ga_I⁺ и As_I⁻);

[M_I⁺] и [X_I⁻] — концентрация ионизованных междоузельных атомов M и X (например, [Ga_I⁺] и [As_I⁻]);

- V_M и V_X — вакансии в подрешетках М и Х (например, V_{Ga} и V_{As});
 $[V_M]$ и $[V_X]$ — концентрация вакансий в подрешетках М и Х (например, $[V_{Ga}]$ и $[V_{As}]$);
 V_M^- и V_X^+ — заряженные вакансии в подрешетках М и Х (например, V_{Ga}^- и V_{As}^+);
 $[V_M^-]$ и $[V_X^+]$ — концентрация заряженных вакансий в подрешетках М и Х (например, $[V_{Ga}^-]$ и $[V_{As}^+]$);
D и A — донорный и акцепторный атомы (например, P и B в Si);
 $[D]$ и $[A]$ — концентрация донорных и акцепторных атомов (например, [P] и [B] в Si);
 D^+ и A^- — донорный и акцепторный ионы (например, P^+ и B^- в Si);
 $[D^+]$ и $[A^-]$ — концентрация донорных и акцепторных ионов (например, $[P^+]$ и $[B^-]$ в Si);
 $[e^-] \equiv n$ — концентрация электронов;
 $[e^+] \equiv p$ — концентрация дырок.

Междоузельные атомы и вакансии в нейтральном и заряженном состояниях, примесные атомы и ионы (донорные, акцепторные, амфотерные) совместно с электронами и дырками рассматриваются как компоненты квазихимических реакций. Эти реакции, как и обычные химические реакции, в равновесном состоянии подчиняются закону действующих масс и могут быть охарактеризованы константами равновесия, зависящими от температуры (см. п. 2.12).

Процессы ионизации донорного атома D и акцепторного атома A с образованием неподвижных ионов D^+ и A^- и подвижных электрона e^- и дырки e^+ можно записать в форме следующих квазихимических реакций:



Применение закона действующих масс (2.12.6) к уравнениям (3.6.1) дает следующий результат:

$$K_D(T) = \frac{[D^+]}{[D]} n \quad \text{и} \quad K_A(T) = \frac{[A^-]}{[A]} p, \quad (3.6.2)$$

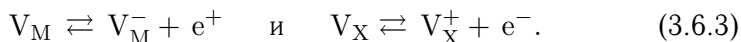
где $K_D(T)$ и $K_A(T)$ — константы равновесия процессов ионизации донора и акцептора.

В полупроводниковых соединениях типа MX собственные дефекты могут находиться в заряженном состоянии и, как уже

говорилось в п. 3.5, выполнять функции донора или акцептора. Это связано с тем, что химическая связь между атомами М и Х в кристаллической решетке МХ не является чисто ковалентной, как в случае гомеополярных кристаллов Ge или Si. Она имеет определенную степень ионности, то есть элементы М и Х несут некоторый эффективный заряд.

В полупроводниках типа МХ обычно элемент М относится к II или III группе периодической таблицы элементов и по свойствам близок к металлам. Поэтому в решетке атом М несет положительный эффективный заряд. Элемент Х относится к VI или V группам и имеет в решетке отрицательный эффективный заряд.

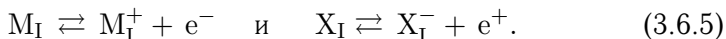
При образовании вакансии V_M или V_X соответствующий элемент М или Х удаляется из узла решетки в виде нейтрального атома. Оставшийся на освободившемся месте эффективный заряд, ранее принадлежащий ушедшему элементу, сохраняется вакансией, обеспечивая ее нейтральность. В результате тепловых колебаний решетки этот заряд может быть высвобожден в виде дырки e^+ или электрона e^- с образованием, соответственно, заряженной вакансии V_M^- или V_X^+ . Иными словами, эффективный заряд катионной вакансии V_M^- создается за счет некомпенсированных отрицательных зарядов анионов X^- , окружающих эту вакансию. Процесс ионизации вакансий можно описать квазихимическими уравнениями:



Закон действующих масс для квазихимических реакций (3.6.3) имеет следующий вид:

$$K_{V_M}(T) = \frac{[V_M^-]p}{[V_M]} \quad \text{и} \quad K_{V_X}(T) = \frac{[V_X^+]n}{[V_X]} . \quad (3.6.4)$$

Собственные атомы кристалла, оказавшиеся в междоузлиях, также могут ионизоваться в соответствии с квазихимическими уравнениями:



Вторая реакция (3.6.5) менее вероятна, поскольку присоединение электрона приводит к увеличению эффективного размера элемента Х, что затрудняет его размещение в междоузлии. Закон действующих масс для квазихимических реакций (3.6.5) записываем по аналогии с формулами (3.6.4) для реакций (3.6.3):

$$K_{M_I}(T) = \frac{[M_I^+]n}{[M_I]} \quad \text{и} \quad K_{X_I}(T) = \frac{[X_I^-]p}{[X_I]} . \quad (3.6.6)$$

3.7. Электронно-дырочное равновесие в полупроводниках

Квазихимическое изучение механизмов возникновения подвижных носителей заряда начинаем с чистого (беспримесного) полупроводника, называемого **собственным полупроводником**. В таком полупроводнике предполагаются полностью отсутствующими как примесные, так и собственные дефекты, что является модельной идеализацией. Однако в природе не существует абсолютно чистых веществ, поскольку они самопроизвольно загрязняются естественными примесями из-за отрицательности изобарного потенциала смещения ($\Delta G_{\text{см}} < 0$ — см. п. 1.12 и 2.4).

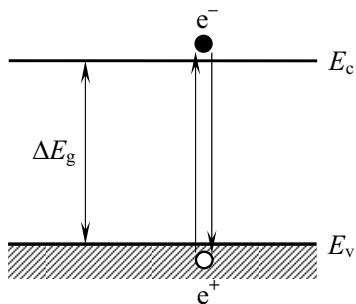


Рис. 3.21. Энергетическая диаграмма для собственного полупроводника, отображающая процесс межзонного возбуждения–рекомбинации

В собственном полупроводнике носители заряда возникают в результате межзонного возбуждения. На рис. 3.21 показана соответствующая энергетическая диаграмма, где, в отличие от рис. 3.17 з, одновременно отображены обратимые процессы межзонного возбуждения и рекомбинации в виде стрелок, направленных вверх и вниз. Вследствие тепловых колебаний решетки электрон может перебрасываться из валентной зоны в зону проводимости, преодолевая энергетический барьер, равный ширине запрещенной зоны ΔE_g . Вероятность

межзонного переброса должна возрастать с ростом температуры. В результате этого процесса в зоне проводимости возникает отрицательно заряженная частица e^- (электрон), а в валентной зоне — положительно заряженная частица e^+ (дырка).

Уравнение квазихимической реакции, описывающее процесс межзонного возбуждения–рекомбинации, имеет форму



которой соответствует закон действующих масс

$$K_i(T) = n_i p_i, \quad (3.7.2)$$

где индекс i происходит от английского слова «intrinsic» («присущий», «свойственный»). В данном случае условие электронейтральности (3.5.3) включает лишь собственные концентрации электронов n_i и дырок p_i , так что для них получаем

$$n_i = p_i. \quad (3.7.3)$$

Подстановка равенства (3.7.3) в уравнение (3.7.2) дает соотношение

$$K_i(T) = n_i^2 \equiv p_i^2, \quad (3.7.4)$$

позволяющее вычислять собственную концентрацию носителей заряда при температуре T , если известна температурная зависимость константы равновесия $K_i(T)$.

Для нахождения $K_i(T)$ используем известные термодинамические соотношения. Тепловую генерацию носителей заряда в полупроводниках обычно рассматривают происходящей в изохорно-изотермических условиях, так как объем кристалла практически неизменен. В этом случае процессы управляются свободной энергией Гельмгольца F , изменение которой в стандартных условиях, записанное в соответствии с (2.13.12) как $\Delta F_T^\circ = -RT \ln K_c(T)$, определяет температурный ход константы равновесия $K_c(T)$.

Для реакции (3.7.1) роль константы равновесия K_c играет K_i , поэтому для нее можем записать:

$$\Delta F_T^\circ = -RT \ln K_i(T). \quad (3.7.5)$$

Как известно, стандартный изохорный потенциал ΔF_T° химической реакции можно выразить через изменения внутренней энергии ΔU_T° и энтропии ΔS_T° в стандартных условиях:

$$\Delta F_T^\circ = \Delta U_T^\circ - T \Delta S_T^\circ. \quad (3.7.6)$$

Из формул (3.7.4)–(3.7.6) следует выражение для собственной константы равновесия:

$$K_i(T) \equiv \exp\left(-\frac{\Delta F_T^\circ}{RT}\right) = K_{i0} \exp\left(-\frac{\Delta U_T^\circ}{RT}\right), \quad (3.7.7)$$

где

$$K_{i0} = \exp\left(\frac{\Delta S_T^\circ}{R}\right).$$

Поскольку прямая реакция (3.7.1) межзонного возбуждения требует затраты тепловой энергии, т. е. является эндотермической ($\Delta U_T^\circ > 0$), то константа равновесия $K_i(T)$ всегда возрастет с увеличением температуры.

Для выяснения физического содержания величин ΔU_T° и ΔS_T° , входящих в (3.7.7), воспользуемся сведениями из физики полупроводников, приведенными в п. 3.5. Собственная концентрация определяется выражениями (3.5.1) при $n_i = p_i$, а именно

$$n_i^2(T) = N_c N_v \exp\left(-\frac{\Delta E_g}{k_B T}\right). \quad (3.7.8)$$

Из сравнения выражений (3.7.7) и (3.7.8), с учетом (3.7.4) и $R = k_B \mathcal{N}_A$, получаем

$$\Delta U_T^\circ = \Delta E_g \mathcal{N}_A, \quad K_{i0} \equiv \exp\left(\frac{\Delta S_T^\circ}{R}\right) = N_c N_v \equiv \frac{1}{2} \left(\frac{k_B T}{\pi \hbar^2}\right)^3 (m_n m_p)^{3/2}, \quad (3.7.9)$$

где $\mathcal{N}_A = 6,022 \cdot 10^{23}$ моль⁻¹ — постоянная Авогадро.

Таким образом, стандартная внутренняя энергия ΔU_T° , как следует из (3.7.9), равна энергии, израсходованной на межзонную генерацию носителей заряда в одном моле собственного полупроводника. Величина K_{i0} , входящая в (3.7.7) и определяемая стандартной энтропией ΔS_T° , равна произведению эффективных плотностей состояний в зонах ($N_c N_v$) и слабо зависит от температуры ($\propto T^3$) по сравнению с экспоненциальной зависимостью для $K_i(T)$ в формуле (3.7.7).

Собственная концентрация носителей заряда является физическим параметром любого полупроводника и зависит только от температуры. Например, для таких полупроводников, как германий, кремний, арсенид галлия при $T = 300$ К: $n_i = 2,4 \cdot 10^{13}$ см⁻³ (Ge), $n_i = 1,5 \cdot 10^{10}$ см⁻³ (Si), $n_i = 1,8 \cdot 10^6$ см⁻³ (GaAs).

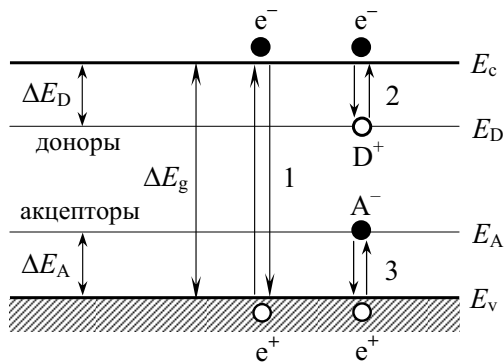


Рис. 3.22. Энергетическая диаграмма для примесного полупроводника, отображающая три процесса генерации–рекомбинации носителей заряда

Рассмотрим **примесный полупроводник**, где в общем случае присутствуют как донорные, так и акцепторные примеси. Возможны три механизма генерации подвижных носителей заряда, показанные направленными вверх стрелками 1, 2 и 3 на рис. 3.22 (аналогичные стрелки, направленные вниз, соответствуют обратным процессам рекомбинации):

1 — тепловое межзонное возбуждение с одновременным возникновением «парных» электрона и дырки;

2 — ионизация донорной примеси с образованием лишь электрона в зоне проводимости (без «парной» дырки) в результате его отрыва от донорного центра;

3 — ионизация акцепторной примеси с образованием только дырки в валентной зоне (без «парного» электрона) в результате захвата электрона акцепторным центром.

Используя выражения (3.6.1), (3.6.2), (3.7.1) и (3.7.2), запишем уравнения квазихимических реакций и соответствующие им законы действующих масс для трех вышеуказанных процессов:

$$0 \rightleftharpoons e^- + e^+, \quad K_i(T) \equiv n_i^2 = np, \quad (3.7.10)$$

$$D \rightleftharpoons D^+ + e^-, \quad K_D(T) = [D^+] n / [D], \quad (3.7.11)$$

$$A \rightleftharpoons A^- + e^+, \quad K_A(T) = [A^-] p / [A]. \quad (3.7.12)$$

Константа равновесия $K_i(T)$ для процесса межзонного возбуждения по-прежнему определяется выражением (3.7.7), которое с учетом (3.7.9) может быть переписано в виде

$$K_i(T) = K_{i0} \exp \left(-\frac{\Delta E_g}{k_B T} \right). \quad (3.7.13)$$

Выражения для констант равновесия $K_D(T)$ и $K_A(T)$, соответствующих реакциям ионизации донорной и акцепторной примеси, можно по аналогии с (3.7.13) записать в следующем виде (см. рис. 3.22):

$$K_D(T) = K_{D0} \exp \left(-\frac{\Delta E_D}{k_B T} \right), \quad (3.7.14)$$

$$K_A(T) = K_{A0} \exp \left(-\frac{\Delta E_A}{k_B T} \right), \quad (3.7.15)$$

где $\Delta E_D = E_c - E_D$ и $\Delta E_A = E_A - E_v$ — глубина залегания донорного и акцепторного уровней.

Поскольку $\Delta E_D > 0$ и $\Delta E_A > 0$, то с возрастанием температуры константы равновесия (3.7.14) и (3.7.15) всегда увеличиваются. Однако скорость нарастания констант $K_D(T)$ и $K_A(T)$ меньше, чем $K_i(T)$, поскольку ΔE_D и ΔE_A меньше, чем ΔE_g .

В дополнение к трем уравнениям (3.7.10)–(3.7.12), в форме закона действующих масс, запишем еще три уравнения:

$$n + [A^-] = p + [D^+], \quad (3.7.16)$$

$$[D] + [D^+] = N_D, \quad (3.7.17)$$

$$[A] + [A^-] = N_A. \quad (3.7.18)$$

Уравнение (3.7.16) представляет условие электронейтральности (3.5.3), переписанное в новых обозначениях, а уравнения (3.7.17) и (3.7.18) задают полное количество доноров N_D и акцепторов N_A , которые введены в единицу объема полупроводника.

Система, состоящая из шести уравнений (3.7.10)–(3.7.12) и (3.7.16)–(3.7.18), позволяет найти неизвестные концентрации: n , p , $[D]$, $[A]$, $[D^+]$ и $[A^-]$, характеризующие равновесное состояние полупроводника при разных температурах. Для этого необходимо знать степень легирования полупроводника (путем задания $N_D = \text{const}$ и $N_A = \text{const}$) и температурную зависимость констант равновесия $K_D(T)$, $K_A(T)$ и $K_i(T)$.

Следует иметь в виду, что при постоянной температуре произведение концентраций электронов и дырок, $np = n_i^2(T) = \text{const}$, не зависит от количества введенных примесей, а константа равновесия $K_i(T) \equiv n_i^2(T)$ определяется, согласно (3.7.13), только шириной ΔE_g запрещенной зоны. Следовательно, введение в полупроводник добавочных доноров увеличивает концентрацию электронов (n) и с неизбежностью уменьшает концентрацию дырок ($p = n_i^2(T)/n$) или наоборот. Именно в этом и заключается физический смысл **электронно-дырочного равновесия**.

В заключение интересно обратить внимание на аналогию между электронно-дырочным равновесием в полупроводниках, выражаемым соотношением (3.7.10), и ионным равновесием в воде, записанным в форме (2.17.3). Это равенство выражает соотношение между равновесными концентрациями водородных и гидроксильных ионов при их возникновении по трем различным механизмам: диссоциация молекул воды по реакции (2.17.1) (аналог процесса межзонного возбуждения (3.7.10)) и диссоциация кислот и щелочей по реакциям (2.17.5) и (2.17.6) (аналог процессов ионизации доноров и акцепторов (3.7.11) и (3.7.12)).

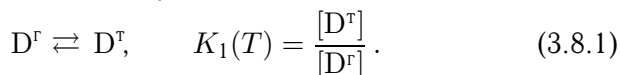
3.8. Растворимость примесей в полупроводниках с учетом ионизации примесных атомов

В п. 2.6 были изучены закономерности молекулярного и атомарного растворения газов в конденсированных средах. Наряду с этим возможен ионный механизм, реализуемый в растворах

жидких электролитов, что уже было изучено в п. 2.17, и в твердых растворах, образуемых примесями в полупроводниках, что будет рассмотрено ниже.

Цель рассмотрения — нахождение закономерностей взаимного влияния примесей при легировании полупроводников из газовой фазы и анализ температурного хода растворимости легирующих примесей при ионном механизме. В отличие от предыдущего параграфа, где полная концентрация примесей считалась заданной и неизменной ($N_D = \text{const}$ и $N_A = \text{const}$), в данном случае предполагаем равновесие между полупроводником и газовой фазой, содержащей легирующую примесь с заданным парциальным давлением; при этом концентрации N_D и N_A подлежат определению. Поскольку концентрацию теперь измеряем *числом частиц* в единице объема, то ниже везде вместо R применяем k_B .

Для определенности будем рассматривать донорную примесь, имеющую парциальное давление p_D в газовой фазе. Пусть это давление создается газом в атомарной форме D^g , который растворяется в твердой фазе с превращением в форму D^t . Такой процесс записываем в виде квазихимической реакции с соответствующим законом действующих масс:

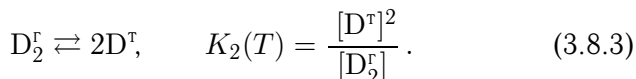


Считая парциальное давление $p_D = [D^g]k_B T$ заданным, найдем концентрацию растворенных атомов в твердой фазе (опуская для простоты верхний индекс «t»):

$$[D] = s_D(T) p_D, \quad (3.8.2)$$

где $s_D(T) = K_1(T)/k_B T$ — коэффициент растворимости. Соотношение (3.8.2) выражает в новых обозначениях *закон Генри* (2.6.4), справедливый при растворении газа в твердой фазе без диссоциации газовых молекул (в данном случае атомов газа).

Пусть газ имеет молекулярную форму D_2^g , тогда атомарное растворение в твердой фазе сопровождается диссоциацией двухатомных молекул на атомы по реакции:



Аналогично сделанному выше, при заданном парциальном давлении $p_D = [D_2^g]k_B T$ в газовой фазе, из равенства (3.8.3) получаем атомарную растворимость газа в твердой фазе:

$$[D] = s_D(T) p_D^{1/2}, \quad (3.8.4)$$

где $s_D(T) = [K_2(T)/k_B T]^{1/2}$ — коэффициент растворимости. Соотношение (3.8.4) выражает в новых обозначениях закон Сиверкса (2.6.12) для атомарного механизма растворения.

Таким образом, равновесная концентрация донорных атомов, растворенных в полупроводнике в нейтральном состоянии, однозначно определяется соответствующим парциальным давлением в газовой фазе, согласно законам Генри или Сиверкса.

Тепловая ионизация донорных атомов D порождает ионы D^+ в соответствии с квазихимической реакцией (3.7.11). Из закона действующих масс для этой реакции получаем равновесную концентрацию донорных ионов:

$$[D^+] = [D] \frac{K_D(T)}{n}. \quad (3.8.5)$$

С учетом выражения (3.8.5) общая растворимость донорной примеси (в виде атомов и ионов) равна

$$N_D \equiv [D] + [D^+] = [D] \left(1 + \frac{K_D(T)}{n} \right). \quad (3.8.6)$$

Аналогично записываем растворимость акцепторной примеси:

$$N_A \equiv [A] + [A^-] = [A] \left(1 + \frac{K_A(T)}{p} \right). \quad (3.8.7)$$

Формулы (3.8.6) и (3.8.7) позволяют выяснить взаимное влияние примесей на их растворимость в полупроводниках. Если в полупроводник, содержащий примеси с концентрациями N_D и N_A , равными (3.8.6) и (3.8.7), вводить другие донорные атомы при $T = \text{const}$, то в силу ионизации последних концентрация электронов n увеличивается, а дырок $p = n_i^2/n$ уменьшается. Тогда растворимость исходной донорной примеси, определяемая формулой (3.8.6), падает при величине $[D]$, поддерживаемой постоянной с помощью внешней газовой фазы, в которой $p_D = \text{const}$. В то же время, согласно (3.8.7), растворимость акцепторной примеси растет при $p_A = \text{const}$ в газовой фазе.

Следовательно, примеси одного типа (или доноры, или акцепторы) при растворении в полупроводнике взаимно уменьшают растворимость друг друга, а растворимость примесей противоположного типа (доноров и акцепторов) взаимно увеличивается.

Исследуем температурную зависимость $N_D(T)$ для растворимости доноров, считая, что полупроводник n -типа находится в равновесии с газовой фазой, содержащей донорную примесь при парциальном давлении $p_D = \text{const}$. Для этого проанализируем формулу (3.8.6) в двух температурных областях.

Область низких температур, где все атомы примеси практически неионизованы ($[D^+] \ll [D]$), обеспечивает, с учетом равенств (3.8.2), (3.8.4) и (3.8.6), существование соотношения

$$N_D \approx [D] = s_D(T) p_D^m. \quad (3.8.8)$$

В этом случае примеси растворяются в неионизованной форме либо по закону Генри ($m = 1$), либо по закону Сивертса ($m = 1/2$). Температурная зависимость коэффициентов растворимости $s_D(T)$ обсуждалась в п. 2.6 (см. формулы (2.6.9) и (2.6.16)).

Область средних и высоких температур, где практически вся примесь ионизована ($[D^+] \gg [D]$), обеспечивает, с учетом равенств (3.8.6) и (3.8.8), существование соотношения

$$N_D \approx [D^+] = [D] \frac{K_D(T)}{n} = s_D(T) K_D(T) \frac{p_D^m}{n}. \quad (3.8.9)$$

Обычно теплота растворения (ΔU_D) для атомов примеси относительно невелика, т. е. $\Delta U_D / \mathcal{N}_A \ll \Delta E_D$. Поэтому вкладом температурной зависимости коэффициента растворимости $s_D(T)$ в величину (3.8.9) можно пренебречь и учитывать только температурный ход константы равновесия $K_D(T)$.

Воспользуемся общим условием электронейтральности в форме (3.7.16), которое для рассматриваемого полупроводника n -типа (при $[A^-] = 0$) принимает следующий вид:

$$n = [D^+] + p \approx N_D + \frac{n_i^2}{n}. \quad (3.8.10)$$

Здесь использовано приближенное равенство в (3.8.9) и условие электронно-дырочного равновесия (3.7.10).

Выражая n из равенства (3.8.9) и подставляя его в (3.8.10), получаем уравнение для нахождения $N_D(T)$, откуда

$$N_D(T) = \frac{[D] K_D(T)}{\sqrt{[D] K_D(T) + n_i^2(T)}}. \quad (3.8.11)$$

Температурную зависимость, описываемую общей формулой (3.8.11), проанализируем для двух частных случаев.

1. *Средние температуры* обеспечивают малый вклад процесса межзонного возбуждения в концентрацию подвижных носителей заряда, так что $n_i^2 \ll [D] K_D$. В этом случае $p \approx 0$ и условие электронейтральности (3.8.10) принимает вид $n \approx [D^+]$. Температурный ход ионной растворимости следует из выражения (3.8.11) в форме

$$N_D(T) \approx \sqrt{[D] K_D(T)} \propto K_D^{1/2}(T) = K_{D0}^{1/2} \exp\left(-\frac{\Delta E_D}{2k_B T}\right). \quad (3.8.12)$$

Здесь использовано выражение (3.7.14) для константы равновесия $K_D(T)$. Как видно из выражения (3.8.12), с возрастанием температуры растворимость $N_D \approx [D^+]$ увеличивается, что изображено в виде начального участка кривой на рис. 3.23.

2. *Высокие температуры* делают вклад процесса межзонного возбуждения в концентрацию подвижных носителей заряда определяющим. В этом случае вся примесь, хоть и будучи ионизованной, но дает малый вклад ($[D^+] \ll p$), так что условие электронейтральности (3.8.10) принимает вид $n \approx p \approx n_i$. Это означает, что $[D]K_D \ll n_i^2$ в формуле (3.8.11), в результате чего

$$N_D(T) \approx \frac{[D]K_D(T)}{n_i(T)} \propto K_D(T) n_i^{-1}(T) = K_{D0} K_{i0}^{-1/2} \exp\left(\frac{\Delta E_g/2 - \Delta E_D}{k_B T}\right). \quad (3.8.13)$$

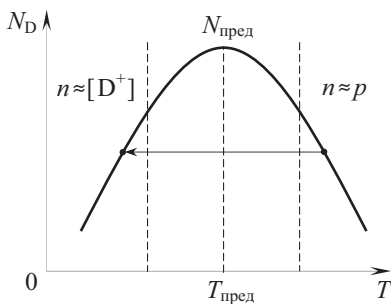


Рис. 3.23. Качественный вид ретроградного хода температурной зависимости ионной растворимости примесей в полупроводниках

Здесь были использованы выражения (3.7.13) и (3.7.14) для констант равновесия $K_D(T)$ и $K_i(T) \equiv n_i^2(T)$. Поскольку $\Delta E_D < \Delta E_g/2$, то величина $N_D(T)$, даваемая формулой (3.8.13), уменьшается с ростом температуры, что изображено в виде конечного участка кривой на рис. 3.23.

Таким образом, температурный ход ионной растворимости примесей отличается от законов Генри и Сивертса, характеризующих монотонной зависимостью от температуры (см. рис. 2.9 б и 2.10 б). В данном случае, как показывает рис. 3.23, температурная зависимость не является монотонной и называется *ретроградным ходом* кривой растворимости. Термин «ретроградный» означает возврат с увеличением температуры к значениям ионной растворимости, уже имевшим место при более низких температурах (показано стрелкой).

Для каждой пары примесь–полупроводник при вполне определенной температуре $T_{пред}$ существует предельная растворимость $N_{пред}$, выше которой невозможно увеличить концентрацию данной примеси в полупроводнике. Так, для кремния (с температурой плавления $T_{пл} = 1417^\circ\text{C}$) донорная примесь фосфора имеет $N_{пред} = 1 \cdot 10^{21} \text{ см}^{-3}$ при $T_{пред} = 1180^\circ\text{C}$, а акцепторная примесь бора имеет $N_{пред} = 6 \cdot 10^{20} \text{ см}^{-3}$ при $T_{пред} = 1250^\circ\text{C}$.

3.9. Взаимодействие заряженных дефектов: компенсация и ассоциация

В предыдущем параграфе с общих позиций было показано, что примесные дефекты, растворенные в ионной форме, взаимно влияют на растворимость друг друга: она увеличивается для ионов разных знаков (доноры/акцепторы) и уменьшается для ионов одного знака (доноры/доноры или акцепторы/акцепторы).

Такое влияние часто трактуют как *непрямое взаимодействие* заряженных дефектов путем сдвига электронно-дырочного равновесия через условие электронейтральности. Увеличение растворимости противоположно заряженных ионов (доноров и акцепторов) вполне объяснимо с позиции *компенсации дефектов* (см. ранее рис. 3.17 в).

Наряду с этим, возможно *прямое взаимодействие* в процессе *ассоциации дефектов*, как бы «выпадающих в осадок» в форме ассоциатов. Для заряженных дефектов разных знаков этому способствует их электростатическое (кулоновское) притяжение.

Ниже на примере примесей в полупроводниках покажем, что всегда ассоциация дефектов, как и их взаимная компенсация, увеличивает концентрацию дефектов за счет их ассоциированной формы, дополняющей ионную растворимость.

Будем искать концентрацию N_D донорной примеси в присутствии акцепторной примеси. Последнюю считаем заранее введенной в кристалл с заданной концентрацией N_A , которую можно изменять как параметр задачи.

Пусть растворение доноров происходит из газовой фазы, содержащей легирующую примесь (в виде атомов или двухтомных молекул) с парциальным давлением p_D , которое определяет атомарную концентрацию доноров в полупроводнике (см. формулы (3.8.2), (3.8.4) и (3.8.8)):

$$[D] = s_D(T) p_D^m, \quad (3.9.1)$$

где $m = 1$ и $1/2$ для законов Генри и Сивертса, соответственно.

Будем считать обе примеси (донорную и акцепторную) мелкими, так что

$$[D] \ll [D^+] \quad \text{и} \quad [A] \ll [A^-]. \quad (3.9.2)$$

В дополнение к формуле (3.9.1), характеризующей равновесие между полупроводником и газовой фазой, запишем остальные необходимые соотношения. Сюда относятся:

- реакция межзонного возбуждения (см. формулу (3.7.10)),

$$0 \rightleftharpoons e^- + e^+, \quad n_i^2(T) = np; \quad (3.9.3)$$

- реакция ионизации донорных атомов D (см. формулу (3.7.11)),

$$D \rightleftharpoons D^+ + e^-, \quad K_D(T) = \frac{[D^+]n}{[D]}; \quad (3.9.4)$$

- реакция ассоциации акцепторных и донорных ионов с образованием нейтрального ассоциата AD,

$$A^- + D^+ \rightleftharpoons AD, \quad K_{AD}(T) = \frac{[AD]}{[A^-][D^+]}; \quad (3.9.5)$$

- условие электронейтральности (см. формулу (3.7.16)),

$$n + [A^-] = p + [D^+]; \quad (3.9.6)$$

- условие сохранения общего числа акцепторных атомов (изначально введенных в кристалл), варьируемого как параметр,

$$N_A = [A^-] + [AD] = \text{const}; \quad (3.9.7)$$

- искомая концентрация донорных атомов, растворенных в ионной и ассоциированной форме,

$$N_D = [D^+] + [AD]. \quad (3.9.8)$$

В выражениях (3.9.7) и (3.9.8) учтено, что, согласно (3.9.2), акцепторы и доноры практически полностью ионизированы, т. е. опущены нейтральные атомы в общем балансе числа частиц.

В предыдущем параграфе было получено выражение (3.8.11) для ионной растворимости *одиной* донорной примеси (в отсутствие других примесей), которое перепишем в виде

$$N_D^0 = \frac{\sqrt{K_D(T)[D]}}{\sqrt{1+\delta}}. \quad (3.9.9)$$

Величину N_D^0 будем называть *собственной растворимостью* донорной примеси (без компенсирующего влияния акцепторов). В формуле (3.9.9) введено обозначение:

$$\delta = \frac{n_i^2(T)}{K_D(T)[D]} \ll 1. \quad (3.9.10)$$

Здесь условие малости δ справедливо для области средних температур, представляющей наибольший интерес. Как показано в п. 3.8, существенный вклад межзонного возбуждения (определяющий ретроградный ход растворимости) проявляется лишь при очень высоких температурах, близких к температуре плавления.

При выводе выражения (3.9.10) были использованы формулы (3.9.3), (3.9.4), (3.9.6) при $[A^-] = 0$ и (3.9.8) при $[AD] = 0$.

Из законов действующих масс (3.9.3) и (3.9.4) следует, что $p \equiv n_i^2/n = \delta[D^+]$. Поэтому, принимая условие (3.9.10), то есть полагая $\delta = 0$, можно опустить p в условии электронейтральности (3.9.6), а собственную растворимость (3.9.9),

$$N_D^0 = \sqrt{K_D(T)[D]} = \sqrt{K_D(T)s_D(T)p_D^m} = \text{const}, \quad (3.9.11)$$

считать известной величиной при заданных значениях T и p_D .

Будем сначала рассматривать влияние только компенсирующей акцепторной примеси, а потом учтем одновременное влияние компенсации и ассоциации.

1. Эффект компенсации без учета ассоциации доноров и акцепторов описываем системой уравнений (3.9.3)–(3.9.8), исключая из нее уравнение (3.9.5) и полагая $[AD] = 0$ в двух последних уравнениях. Тогда условие электронейтральности (3.9.6) с учетом того, что $p = 0$, $[A^-] = N_A$ и $[D^+] = N_{D0}$, принимает вид

$$n + N_A = N_{D0}. \quad (3.9.12)$$

При этом концентрация донорных ионов, $N_{D0} \neq N_D^0$, растворенных в полупроводнике при наличии акцепторов, является искомой величиной, учитывающей эффект компенсации. Здесь нижний нулевой индекс ($N_{D0} = [D^+]$) отражает наличие как донорных, так и акцепторных ионов *без учета ассоциации*, а верхний нулевой индекс (N_D^0 , собственная растворимость донорной примеси) соответствует одиночным донорам (без акцепторов).

Из закона действующих масс (3.9.3), с учетом (3.9.11), получаем

$$n = \frac{K_D[D]}{[D^+]} = \frac{(N_D^0)^2}{N_{D0}}. \quad (3.9.13)$$

Введем в рассмотрение *степень компенсации* α_0 как долю донорных ионов, компенсированных акцепторными ионами с концентрацией $[A^-] = N_A$, по отношению к их полному числу N_D^0 , полученному при ионной растворимости одиночных доноров:

$$\alpha_0 = \frac{[A^-]}{N_D^0} = \frac{N_A}{N_D^0}. \quad (3.9.14)$$

Величину α_0 (с нижним нулевым индексом, отражающим отсутствие ассоциации примесей) будем называть *степенью начальной компенсации* (ср. α в формуле (3.9.33)).

Подстановка (3.9.13) в равенство (3.9.12) приводит к уравнению для нормированной величины $X_0 = N_{D0}/N_D^0$:

$$X_0^2 - \alpha_0 X_0 - 1 = 0. \quad (3.9.15)$$

Таким образом, ионная растворимость доноров с учетом их компенсации акцепторами (без эффекта ассоциации) равна

$$N_{D0} = X_0 N_D^0, \quad \text{где } X_0 = \frac{\alpha_0}{2} + \sqrt{1 + \left(\frac{\alpha_0}{2}\right)^2}. \quad (3.9.16)$$

Как и следовало ожидать, эффект компенсации увеличивает растворимость примеси, поскольку всегда $X_0 > 1$ при $\alpha_0 \neq 0$.

2. Совместный эффект компенсации и ассоциации описываем полной системой уравнений (3.9.3)–(3.9.8), в которой теперь $[AD] \neq 0$, $[A^-] \neq N_A$, $[D^+] \neq N_D$ (см. (3.9.7) и (3.9.8)).

Ассоциацию будем учитывать, вводя *степень ассоциации* β как долю акцепторных ионов, образовавших нейтральные ассоциаты с концентрацией $[AD]$, по отношению к исходной концентрации акцепторов N_A :

$$\beta = \frac{[AD]}{N_A}. \quad (3.9.17)$$

Тогда из равенств (3.9.7) и (3.9.17) получаем концентрацию акцепторных ионов:

$$[A^-] \equiv N_A - [AD] = (1 - \beta)N_A = \alpha_0(1 - \beta)N_D^0, \quad (3.9.18)$$

где в последнем равенстве использована формула (3.9.14).

Для концентрации электронов используем выражение, аналогичное (3.9.13), но содержащее $[D^+] \neq N_D$ (см. формулу (3.9.8)):

$$n = \frac{K_D[D]}{[D^+]} = \frac{(N_D^0)^2}{[D^+]}. \quad (3.9.19)$$

Подстановкой выражений (3.9.18) и (3.9.19) в условие электронейтральности (3.9.6) при $p = 0$ получаем уравнение для нахождения концентрации доноров, растворенных в ионной форме,

$$[D^+]^2 - \alpha_0(1 - \beta)N_D^0[D^+] - (N_D^0)^2 = 0. \quad (3.9.20)$$

С учетом (3.9.18), из закона действующих масс для реакции ассоциации (3.9.5) получаем связь между β и $[D^+]$ в виде

$$\beta = \frac{C[D^+]}{C[D^+] + N_D^0}, \quad \text{откуда } C \frac{[D^+]}{N_D^0} = \frac{\beta}{1 - \beta}. \quad (3.9.21)$$

Здесь введена безразмерная величина (считаемая известной)

$$C(T) = K_{AD}(T)N_D^0 \equiv K_{AD}(T)\sqrt{K_D(T)s_D(T)p_D^m}, \quad (3.9.22)$$

которую будем называть *параметром ассоциации*.

Введением нормированной величины $X = [D^+]/N_D^0$ приводим уравнение (3.9.20) к следующему виду (ср. уравнение (3.9.15)):

$$X^2 - \frac{\alpha_0}{1 + CX} X - 1 = 0, \quad (3.9.23)$$

где использована формула (3.9.21), откуда $1 - \beta = (1 + CX)^{-1}$.

Уравнение (3.9.23) является кубическим относительно искомой величины X . В отсутствие ассоциации (при $C = 0$) оно переходит в ранее полученное квадратное уравнение (3.9.15) для X_0 .

Искомую концентрацию донорных атомов, растворенных одновременно в ионной и ассоциированной форме, с помощью формул (3.9.8), (3.9.14) и (3.9.17) записываем в нормированном виде:

$$\frac{N_D}{N_D^0} = X + \alpha_0 \beta = X \left(1 + \frac{\alpha_0 C}{1 + CX} \right). \quad (3.9.24)$$

Здесь собственная растворимость N_D^0 одиночной донорной примеси, даваемая формулой (3.9.11), считается известной.

Рассмотрим два предельных случая малой ($\beta \ll 1$) и большой ($\beta \approx 1$) степени ассоциации доноров и акцепторов.

Малая степень ассоциации: $\beta \ll 1$ и $CX \approx \beta \ll 1$.

В этом случае кубическое уравнение (3.9.23) превращается в следующее квадратное уравнение,

$$(1 + \alpha_0 C)X^2 - \alpha_0 X - 1 = 0, \quad (3.9.25)$$

решение которого после подстановки в (3.9.24) при $X \ll C^{-1}$ дает

$$\frac{N_D}{N_D^0} \approx (1 + \alpha_0 C)X = \frac{\alpha_0}{2} + \sqrt{1 + \left(\frac{\alpha_0}{2}\right)^2 + \alpha_0 C}. \quad (3.9.26)$$

Большая степень ассоциации: $\beta \approx 1$ и $CX \approx (1 - \beta)^{-1} \gg 1$.

В этом случае кубическое уравнение (3.9.23) превращается в следующее квадратное уравнение,

$$CX^2 - \alpha_0 - C = 0, \quad (3.9.27)$$

решение которого после подстановки в (3.9.24) при $X \gg C^{-1}$ дает

$$\frac{N_D}{N_D^0} \approx X + \alpha_0 = \alpha_0 + \sqrt{1 + \frac{\alpha_0}{C}}. \quad (3.9.28)$$

Как видно из выражения (3.9.26), при «включении» механизма ассоциации растворимость донорной примеси начинает возрастать по отношению к той величине (3.9.16), которая была

сформирована компенсирующими акцепторами в отсутствие ассоциации (при $C = 0$):

$$N_{D0} = X_0 N_D^0 > N_D^0. \quad (3.9.29)$$

По мере возрастания параметра ассоциации C наблюдается переход от выражения (3.9.26) к выражению (3.9.28), так что при $C \gg \alpha_0$ получаем

$$N_D \approx (\alpha_0 + 1) N_D^0. \quad (3.9.30)$$

Это означает, что $X \approx 1$ или $[D^+] \approx N_D^0$. Тогда, согласно формулам (3.9.8) и (3.9.30), концентрация ассоциатов равна

$$[AD] \equiv N_D - [D^+] \approx N_D - N_D^0 = \alpha_0 N_D^0. \quad (3.9.31)$$

Отсюда, в соответствии с равенством (3.9.7), для концентрации акцепторных ионов получаем

$$[A^-] \equiv N_A - [AD] \approx N_A - \alpha_0 N_D^0 \approx 0, \quad (3.9.32)$$

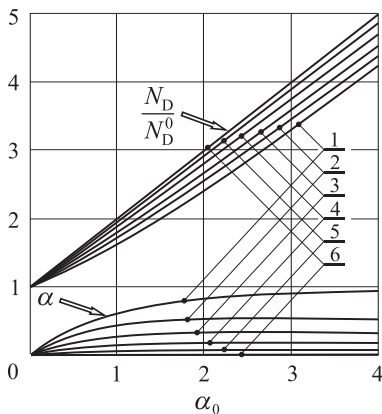


Рис. 3.24. Зависимость растворимости N_D/N_D^0 и степени реальной компенсации α от степени начальной компенсации α_0 для шести значений параметра ассоциации: $C = 0$ (кривые 1), $C = 0,3$ (кривые 2), $C = 1$ (кривые 3), $C = 3$ (кривые 4), $C = 10$ (кривые 5), $C = 100$ (кривые 6)

где приближенное нулевое равенство соответствует определению α_0 в форме (3.9.14). Поскольку, согласно (3.9.18), $[A^-] = (1 - \beta)N_A$, то обращение (3.9.32) в нуль происходит при $\beta \lesssim 1$.

Следовательно, с увеличением ассоциации акцепторная примесь, введенная в полупроводник при фиксированной концентрации ($N_A = \text{const}$), все в большей степени превращается в нейтральный ассоциат AD, и практически полностью исчезает ее ионная форма ($[A^-] \approx 0$) при $\beta \approx 1$.

Результаты численных расчетов по точным формулам (3.9.23) и (3.9.24) приведены на рис. 3.24 для шести зна-

чений параметра ассоциации $C \equiv K_{AD} N_D^0$ (см. формулу (3.9.22)). При этом нижняя кривая 1 из верхнего семейства кривых соответствует начальной растворимости доноров, сформированной компенсирующими акцепторами в отсутствие ассоциации ($C = 0$).

Таким образом, при любой степени начальной компенсации ($\alpha_0 > 0$) и любом параметре ассоциации ($C \geq 0$) наблюдается превышение растворимости донора N_D над собственной растворимостью N_D^0 (в отсутствие акцепторов). Как видно из рис. 3.24, это превышение может достигать нескольких раз (вплоть до 10).

Следует иметь в виду, что величина α_0 (*степень начальной компенсации*), введенная в форме (3.9.14) по отношению к собственной растворимости доноров N_D^0 (в отсутствие акцепторов) и принимающая значения, большие единицы, не отражает действительную степень компенсации в полупроводнике при наличии ассоциации. В этом случае *степень реальной компенсации* α (без нулевого индекса) следует вводить по отношению к той растворимости доноров, $N_{D0} = X_0 N_D^0$, которую сформировали компенсирующие акцепторы (см. формулу (3.9.16)). Итак,

$$\alpha = \frac{[A^-]}{N_{D0}} = \frac{[A^-]}{X_0 N_D^0} \neq \frac{N_A}{N_D^0} = \alpha_0, \quad (3.9.33)$$

Учитывая (3.9.17) и (3.9.33), переписываем равенство (3.9.7) в следующем виде:

$$1 = \frac{[A^-]}{N_A} + \frac{[AD]}{N_A} = \frac{[A^-]}{N_{D0}} \frac{N_{D0}}{N_A} + \frac{[AD]}{N_A} = \frac{\alpha}{\alpha_0} X_0 + \beta.$$

Отсюда следует соотношение, связывающее величины (3.9.14), (3.9.17) и (3.9.33), а именно

$$\alpha = \frac{\alpha_0}{X_0} (1 - \beta) = \frac{\alpha_0}{X_0} \frac{1}{1 + CX}, \quad (3.9.34)$$

где в последнем равенстве использовано выражение (3.9.21).

Формула (3.9.34) позволяет получить зависимость α от α_0 при разных значениях параметра ассоциации C . Эти зависимости приведены в нижней части рис. 3.24. Отсюда видно, что степень реальной компенсации α всегда меньше единицы. В отсутствие ассоциации (при $C = 0$) верхняя кривая 1 монотонно приближается к единице с ростом α_0 , а все другие кривые имеют сильно размытый максимум при $\alpha_0 \approx 2 \div 3$. Этот максимум примерно равен 0,55, 0,34, 0,18, 0,074 и 0,009 при $C = 0,3, 1, 3, 10$ и 100 , соответственно. Как и следовало ожидать, с увеличением параметра ассоциации степень компенсации уменьшается практически до нуля, поскольку исчезают компенсирующие акцепторные ионы, будучи связанными в форме нейтральных ассоциатов AD.

Приведенный выше термодинамический анализ процесса ассоциации доноров и акцепторов применим к любым заряженным дефектам, включая вакансии и междоузельные атомы.

3.10. Внутреннее равновесие собственных и примесных дефектов

Как уже отмечалось, модель совершенного кристалла является физической идеализацией, полезной для описания некоторых свойств кристаллов. Однако даже в пренебрежении влиянием примесей она нереализуема при конечных температурах решетки.

Наряду с тем, что все атомы совершают тепловые колебательные движения (что само по себе рассматривается как один из видов дефектности, приводящей к генерации квазичастиц, называемых *фононами*), некоторые из них нарушают идеальный порядок в заселении узлов кристаллической решетки. Именно такие дефекты в виде вакансий, междоузельных атомов и атомов в антиструктурных положениях принято называть *собственными дефектами* кристалла.

Равновесная концентрация собственных дефектов обусловлена тем обстоятельством, что ее возрастание до определенного уровня приводит к уменьшению *свободной энергии Гельмгольца* F (пренебрегая незначительным изменением объема кристалла при генерации вакансий по механизму Шоттки). Если создание одной вакансии требует затраты энергии ΔE_V , то при их концентрации $N_V \equiv [V]$ увеличение внутренней энергии единицы объема равно $\Delta U = N_V \Delta E_V$.

Наряду с этим возрастает энтропия кристалла на величину $\Delta S = \Delta S_{\text{кон}} + N_V \Delta S_{\text{кол}}$, состоящую из двух вкладов. Первый вклад $\Delta S_{\text{кон}}$, называемый *конфигурационной энтропией*, представляет собой не что иное, как энтропию смешения, возникающую в результате статистического распределения дефектов (вакансий) по всем доступным узлам решетки. Второй вклад, называемый *колебательной энтропией*, вызван понижением частоты колебаний атомов, окружающих одиночную вакансию, по сравнению с аналогичной частотой до образования вакансии. Поэтому всегда для колебательной энтропии $\Delta S_{\text{кол}} > 0$, как и для конфигурационной энтропии $\Delta S_{\text{кон}} > 0$.

Вычислим конфигурационную энтропию для кристалла, имеющего N атомов в единице объема, которые до образования вакансий занимали N узлов решетки.

Пусть возникло N_V вакансий путем перехода соответствующих атомов на поверхность кристалла с образованием N_V новых узлов. Статистический вес W такого состояния определяется как число возможных комбинаций, приводящих к распределению атомов по $N + N_V$ узлам решетки. Он равен общему числу перестановок $(N + N_V)!$, из которого надо исключить перестановки

одинаковых структурных элементов, а именно, $N!$ перестановок атомов и $N_V!$ перестановок вакансий. Тогда статистический вес для кристалла с вакансиями, равный $W = (N + N_V)! / (N! N_V!)$, определяет его конфигурационную энтропию, вычисляемую по общей формуле (1.1.1):

$$\begin{aligned}\Delta S_{\text{кон}} &\equiv k_B \ln W = \\ &= k_B \ln \frac{(N + N_V)!}{N! N_V!} \approx k_B \left(N \ln \frac{N + N_V}{N} + N_V \ln \frac{N + N_V}{N_V} \right),\end{aligned}$$

где приближенное равенство записано для больших значений N и N_V , когда справедлива формула Стирлинга $\ln x! \approx x \ln x - x$.

На основании вышесказанного свободная энергия образования N_V вакансий представляется в виде

$$\begin{aligned}\Delta F &\equiv \Delta U - T \Delta S = \\ &= N_V (\Delta E_V - T \Delta S_{\text{кол}}) - k_B T \left(N \ln \frac{N + N_V}{N} + N_V \ln \frac{N + N_V}{N_V} \right).\end{aligned}$$

Отсюда условие минимума изохорного потенциала ΔF , записанное в следующей форме:

$$\left(\frac{\partial \Delta F}{\partial N_V} \right)_T = \Delta E_V - T \Delta S_{\text{кол}} - k_B T \ln \frac{N + N_V}{N_V} = 0,$$

дает искомую равновесную концентрацию вакансий в кристалле при данной температуре T (считая $N_V \ll N$):

$$N_V \equiv [V] = N \exp \left(\frac{\Delta S_{\text{кол}}}{k_B} \right) \exp \left(- \frac{\Delta E_V}{k_B T} \right). \quad (3.10.1)$$

Выполненный анализ соответствует генерации вакансий по механизму *Шоттки*, описанному в п. 3.4, для которого характерно возникновение вакансии путем выхода атома на поверхность кристалла без образования междоузельного атома.

Наряду с этим, как известно, возможен и другой механизм, называемый *механизмом Френкеля*. Согласно этому механизму, атом из узла решетки переходит в междоузелье с парным образованием вакансии и междоузельного атома, обеспечивая равные их концентрации ($N_V = N_I$).

Аналогично сделанному выше, записываем для механизма Френкеля статистический вес кристалла с вакансиями и междоузельными атомами в следующем виде:

$$W = \frac{N!}{(N - N_V)! N_V!} \frac{N'!}{(N' - N_I)! N_I!}.$$

Это дает равновесную концентрацию дефектов по Френкелю:

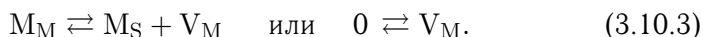
$$N_V = N_I = (NN')^{1/2} \exp\left(\frac{\Delta S_{\text{кол}}}{k_B}\right) \exp\left(-\frac{\Delta E_I}{2k_B T}\right), \quad (3.10.2)$$

где ΔE_I — энергия, необходимая для перевода атома из узла в междоузлие, N' — число междоузлий в единице объема кристалла до образования дефектов по Френкелю.

Следовательно, при любой конечной температуре устойчивому состоянию соответствует дефектный кристалл, содержащий вакансии и междоузельные атомы, концентрация которых, согласно (3.10.1) и (3.10.2), экспоненциально возрастает с увеличением температуры. Это служит подтверждением ранее высказанного утверждения о практической нереализуемости модели идеального (бездефектного) кристалла.

Применим квазихимический подход для описания генерации собственных дефектов по механизмам Шоттки и Френкеля.

1. Дефекты по Шоттки. В элементарном полупроводнике М атом M_M , находящийся в своем узле, из приповерхностного слоя переходит на поверхность кристалла в форме атома M_S , а на его месте возникает вакансия V_M . Созданная таким образом вакансия диффундирует в глубь кристалла как результат встречного движения атомов по вакансиям. Этот процесс записывают в виде квазихимического уравнения

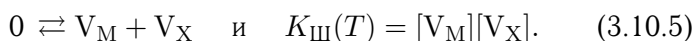


Вторая форма записи в (3.10.3) получена из первой путем сокращения M_M и M_S , являющихся обозначением одного и того же атома. Эта форма носит название *реакции растворения вакуума* и ей соответствует закон действующих масс в следующем виде:

$$K_{\text{Ш}}(T) = [V_M]. \quad (3.10.4)$$

В полупроводниковом соединении МХ с ионными связями вакансии по механизму Шоттки возникают одновременно в двух подрешетках (катионной и анионной). В противном случае, при выходе на поверхность ионов только одного знака она (поверхность) оказалась бы заряженной.

Квазихимическая реакция растворения вакуума для бинарного полупроводника и соответствующий закон действующих масс имеют вид:



Реакция (3.10.5) отражает парное возникновение вакансий в обеих подрешетках, что сохраняет электронейтральность ионного соединения MX и его стехиометрию при $[\text{V}_\text{M}] = [\text{V}_\text{X}]$.

Сравнение формул (3.10.1) и (3.10.4) дает выражение для константы равновесия по Шоттки,

$$K_\text{Ш}(T) = K_\text{Ш0} \exp\left(-\frac{\Delta E_\text{V}}{k_\text{B}T}\right), \quad (3.10.6)$$

экспоненциально зависящей от температуры.

2. Дефекты по Френкелю. В элементарном полупроводнике M возникают одновременно вакансии V_M и междоузельный атом M_I . Квазихимическая реакция и соответствующий закон действующих масс имеют следующий вид:

$$0 \rightleftharpoons \text{M}_\text{I} + \text{V}_\text{M} \quad \text{и} \quad K_\text{ФМ}(T) = [\text{M}_\text{I}][\text{V}_\text{M}]. \quad (3.10.7)$$

В полупроводниковых соединениях MX дефекты по Френкелю возникают независимо в двух подрешетках. Для подрешетки M справедлива та же квазихимическая реакция (3.10.7), что и для элементарного полупроводника M . В подрешетке X механизм Френкеля управляется аналогичными соотношениями:

$$0 \rightleftharpoons \text{X}_\text{I} + \text{V}_\text{X} \quad \text{и} \quad K_\text{ФХ}(T) = [\text{X}_\text{I}][\text{V}_\text{X}]. \quad (3.10.8)$$

В силу парного возникновения вакансий и междоузельных атомов из закона действующих масс (3.10.7) получаем их равные концентрации:

$$[\text{V}_\text{M}] = [\text{M}_\text{I}] = K_\text{ФМ}^{1/2}(T). \quad (3.10.9)$$

Сравнение формул (3.10.2) и (3.10.9) дает выражение для константы равновесия по Френкелю:

$$K_\text{Ф}(T) = K_\text{Ф0} \exp\left(-\frac{\Delta E_\text{I}}{k_\text{B}T}\right), \quad (3.10.10)$$

где ΔE_I — энергия, затрачиваемая на переход атома из узла в междоузельное положение.

Из формул (3.10.6) и (3.10.10) видно, что температурная зависимость констант равновесия определяется энергией образования соответствующего дефекта. Поскольку $\Delta E_\text{V} > 0$ и $\Delta E_\text{I} > 0$, то с увеличением температуры константы равновесия $K_\text{Ш}(T)$ и $K_\text{Ф}(T)$ всегда возрастают.

Механизм генерации дефектов по Шоттки (реакции (3.10.3) и (3.10.5)) наиболее вероятен, поскольку отсутствует геометрический фактор, который может стать определяющим при возникновении дефектов по Френкелю (реакции (3.10.7) и (3.10.8)),

особенно в плотноупакованных решетках. В этом случае размер атома должен быть достаточно малым для возможности размещения его в междоузлии решетки.

Представляет интерес выяснение вопроса о взаимном влиянии собственных и примесных дефектов кристалла в процессе их образования.

Рассмотрим бинарный полупроводник МХ, из собственных дефектов которого будем учитывать только вакансии в подрешетках М и Х, возникающие по механизму Шоттки в соответствии с квазихимической реакцией (3.10.5). В реальной ситуации существенными могут оказаться также и дефекты по Френкелю, антиструктурные дефекты, бивакансии и комплексы с участием собственных и примесных атомов. Это естественно усложняет задачу, но принципиально не изменяет порядок ее рассмотрения.

В дополнение к реакции в форме (3.10.5), описывающей генерацию вакансий по Шоттки, запишем остальные квазихимические реакции с соответствующими законами действующих масс и другие необходимые соотношения. Сюда относятся:

- реакция ионизации вакансий в подрешетке М (см. формулы (3.6.3) и (3.6.4)),

$$V_M \rightleftharpoons V_M^- + e^+, \quad K_{V_M}(T) = \frac{[V_M^-]p}{[V_M]}; \quad (3.10.11)$$

- реакция ионизации вакансий в подрешетке Х (см. формулы (3.6.3) и (3.6.4)),

$$V_X \rightleftharpoons V_X^+ + e^-, \quad K_{V_X}(T) = \frac{[V_X^+]n}{[V_X]}; \quad (3.10.12)$$

- реакция ионизации донорных атомов D (см. формулу (3.7.11)),

$$D \rightleftharpoons D^+ + e^-, \quad K_D(T) = \frac{[D^+]n}{[D]}; \quad (3.10.13)$$

- реакция ионизации акцепторных атомов А (см. ф-лу (3.7.12)),

$$A \rightleftharpoons A^- + e^+, \quad K_A(T) = \frac{[A^-]p}{[A]}; \quad (3.10.14)$$

- реакция межзонного возбуждения (см. формулу (3.7.10)),

$$0 \rightleftharpoons e^- + e^+, \quad n_i^2(T) = np; \quad (3.10.15)$$

- условие электронейтральности (ср. формулу (3.7.16)),

$$n + [A^-] + [V_M^-] = p + [D^+] + [V_X^+]; \quad (3.10.16)$$

- условие сохранения общего числа донорных атомов, введенных в кристалл,

$$N_D = [D] + [D^+] = \text{const}; \quad (3.10.17)$$

- условие сохранения общего числа акцепторных атомов, введенных в кристалл,

$$N_A = [A] + [A^-] = \text{const}; \quad (3.10.18)$$

- условие сохранения стехиометрии полупроводникового соединения,

$$[V_M] + [V_M^-] = [V_X] + [V_X^+]. \quad (3.10.19)$$

Система, состоящая из десяти уравнений (3.10.5) и (3.10.11)–(3.10.19), принципиально разрешима относительно искомых концентраций — n , p , $[V_M]$, $[V_M^-]$, $[V_X]$, $[V_X^+]$, $[D]$, $[D^+]$, $[A]$, $[A^-]$, характеризующих равновесное состояние полупроводника при разных температурах. Для этого нужно знать степень легирования полупроводника (путем задания $N_D = \text{const}$ и $N_A = \text{const}$) и температурную зависимость всех констант равновесия, входящих в законы действующих масс (3.10.5) и (3.10.11)–(3.10.15).

Термодинамические величины, необходимые для расчета констант равновесия реакций образования дефектов, на практике обычно рассчитывают в рамках вполне определенных приближений с использованием экспериментальных данных по межфазовым равновесиям.

Для качественного анализа запишем общую концентрацию вакансий с помощью уравнений (3.10.11) и (3.10.12):

$$N_{V_M} \equiv [V_M] + [V_M^-] = [V_M] \left(1 + \frac{K_{V_M}(T)}{p} \right), \quad (3.10.20)$$

$$N_{V_X} \equiv [V_X] + [V_X^+] = [V_X] \left(1 + \frac{K_{V_X}(T)}{n} \right). \quad (3.10.21)$$

Если в кристалл MX вводится донорная примесь D , то концентрация электронов n возрастает, а дырок $p = n_i^2/n$ снижается. Тогда на основании выражений (3.10.20) и (3.10.21) при $T = \text{const}$ концентрация вакансий в подрешетке M должна увеличиться, а в подрешетке X — уменьшиться. Противоположное заключение можно сделать в отношении акцепторной примеси, вводимой в кристалл с вакансиями.

Обратное влияние вакансий на растворимость примесных атомов при снятии условий (3.10.17) и (3.10.18) следует из формул (3.8.6) и (3.8.7), а именно: увеличение концентрации $[V_M]$ для вакансий акцепторного типа (или уменьшение концентрации

$[V_X]$ для вакансий донорного типа) увеличивает растворимость донорной примеси, $N_D \equiv [D] + [D^+]$, и снижает растворимость акцепторной примеси, $N_A \equiv [A] + [A^-]$.

Полученный результат является обобщением сделанного ранее в п. 3.10 заключения о растворимости доноров и акцепторов. Он позволяет сформулировать общее правило взаимного влияния заряженных собственных и примесных дефектов в кристаллах: *растворимость дефекта (как примесного, так и собственного) в ионной форме возрастает, если присутствуют ионизованные дефекты противоположного знака*, и, в свою очередь, повышает концентрацию этих дефектов; *если же ионизованные дефекты имеют заряды одного знака, то они взаимно снижают растворимость друг друга*.

3.11. Управление собственными дефектами путем отжига кристаллов в парогазовой среде

В предыдущем параграфе было рассмотрено внутреннее равновесие дефектов в полупроводниковом соединении MX в отсутствие его контакта с окружающей средой. В этом случае возникающие собственные дефекты не нарушают стехиометрический состав соединения (см. формулу (3.10.19)). Желаемое отклонение от стехиометрии можно получить путем управляемого обмена атомами между полупроводником и внешней газовой фазой. Это достигается в процессе высокотемпературного отжига кристалла MX в газе, содержащем избыточные компоненты M или X (последний чаще в молекулярной форме X_2).

Обычно компонент X является более летучим, что изначально приводит к нарушению стехиометрии за счет преобладания вакансий V_X . Поэтому длительная выдержка кристалла в атмосфере газа X_2 при высокой температуре (*отжиг* для установления равновесной концентрации дефектов) и последующее быстрое охлаждение до комнатной температуры (*закалка* для сохранения требуемой дефектности кристалла) позволяют управлять нестехиометричностью кристалла путем заполнения вакансий атомами газа. Таким способом можно регулировать величину и тип проводимости полупроводника. Возникшее «замороженное» состояние дефектов является неравновесным, однако достаточно устойчивым. Действительно, при комнатной температуре диффузионные перемещения атомов настолько медленные, что практически исключается миграция дефектов, и сохраняется нестехиометричность кристалла, достигнутая на этапе отжига.

Рассмотрим термодинамическое равновесие в системе, состоящей из бинарного соединения MX (например, оксида или халькогенида типа $\text{A}^{\text{II}}\text{B}^{\text{VI}}$) и двух газообразных компонентов $\text{M}(\text{г})$ и $\text{X}_2(\text{г})$. Сначала пренебрежем влиянием примесей и будем считать, что среди собственных дефектов преобладающую роль играют дефекты по Шоттки $V_{\text{М}}$ и V_{X} , причем $V_{\text{М}}$ действует всегда как акцептор, а V_{X} — как донор.

Искомые равновесные концентрации электронов, дырок и вакансий подчиняются системе уравнений, включающей:

- уравнение генерации вакансий по механизму Шоттки,

$$0 \rightleftharpoons V_{\text{М}} + V_{\text{X}}, \quad K_{\text{Ш}}(T) = [V_{\text{М}}][V_{\text{X}}]; \quad (3.11.1)$$

- уравнение ионизации вакансий в подрешетке М,

$$V_{\text{М}} \rightleftharpoons V_{\text{М}}^- + e^+, \quad K_{V_{\text{М}}}(T) = \frac{[V_{\text{М}}^-]p}{[V_{\text{М}}]}; \quad (3.11.2)$$

- уравнение ионизации вакансий в подрешетке X,

$$V_{\text{X}} \rightleftharpoons V_{\text{X}}^+ + e^-, \quad K_{V_{\text{X}}}(T) = \frac{[V_{\text{X}}^+]n}{[V_{\text{X}}]}; \quad (3.11.3)$$

- уравнение межзонного возбуждения,

$$0 \rightleftharpoons e^- + e^+, \quad n_i^2(T) = np; \quad (3.11.4)$$

- уравнение электронейтральности,

$$n + [V_{\text{М}}^-] = p + [V_{\text{X}}^+]; \quad (3.11.5)$$

- уравнение равновесия кристалла с газом, имеющим избыточные атомы М,

$$\text{M}_{\text{М}} \rightleftharpoons \text{M}(\text{г}) + V_{\text{М}}, \quad K_{\text{М}}(T) = p_{\text{М}}[V_{\text{М}}]; \quad (3.11.6)$$

- уравнение равновесия кристалла с газом, имеющим избыточные атомы X,

$$\text{X}_{\text{X}} \rightleftharpoons \frac{1}{2}\text{X}_2(\text{г}) + V_{\text{X}}, \quad K_{\text{X}}(T) = p_{\text{X}_2}^{1/2}[V_{\text{X}}]. \quad (3.11.7)$$

Комбинация квазихимических реакций (3.11.6) и (3.11.7), с учетом уравнения (3.11.1), дает общую реакцию диссоциации твердого вещества MX на газовые компоненты:

$$\text{MX} \rightleftharpoons \text{M}(\text{г}) + \frac{1}{2}\text{X}_2(\text{г}), \quad K_{\text{MX}}(T) = p_{\text{М}}p_{\text{X}_2}^{1/2}. \quad (3.11.8)$$

Следовательно, при $T = \text{const}$ нельзя независимо регулировать парциальные давления $p_{\text{М}}$ и p_{X_2} , то есть управляемым

параметром является одно из этих давлений. Тогда шесть уравнений (3.11.1)–(3.11.5) и (3.11.6) или (3.11.7) принципиально позволяют рассчитать искомые концентрации n , p , $[V_M]$, $[V_M^-]$, $[V_X]$, $[V_X^+]$, при условии, что соответствующие константы равновесия, как функции температуры, известны.

Качественно исследуем влияние парциального давления p_{X_2} на концентрацию электронов и дырок в полупроводнике при его изотермическом отжиге в среде, содержащей газ X_2 .

Сначала рассмотрим кристалл MX , помещенный в запаянную ампулу, внутри которой в результате диссоциации соединения при $T = \text{const}$ формируется газовая фаза с равновесными давлениями $\bar{p}_M$ и $\bar{p}_{X_2}$, связанными между собой законом действующих масс (3.11.8). В этом случае сохраняется стехиометрия кристалла, ибо парная генерация вакансий в подрешетках обеспечивает их равные концентрации $[V_M] = [V_X] = K_{\text{Ш}}^{1/2}(T)$. Такая ситуация соответствует компенсированному полупроводнику (рис. 3.17 в) с условием электронейтральности (3.11.5) в форме $n = p$, которая аналогична собственному полупроводнику (рис. 3.17 г).

Пусть теперь в газовую фазу вводится избыток компонента X_2 под давлением $p_{X_2} > \bar{p}_{X_2}$. Тогда из закона действующих масс (3.11.7) следует, что концентрация $[V_X]$ вакансий донорного типа уменьшается, а для вакансий акцепторного типа их концентрация $[V_M]$, в соответствии с (3.11.1), увеличивается. В этом случае стехиометрия соединения нарушается, так как $[V_M] > [V_X]$. Это обеспечивает неравенство $p > n$, т. е. кристалл превращается в полупроводник p -типа. При введении в газовую фазу избытка компонента M под давлением $p_M > \bar{p}_M$ из формулы (3.11.8) получаем $p_{X_2} < \bar{p}_{X_2}$. Ситуация становится прямо противоположной: $[V_X] > [V_M]$ и $n > p$, что соответствует полупроводнику n -типа.

Иллюстрация вышеизложенного качественно изображена на рис. 3.25 в виде зависимости $|n - p|$ от парциального давления p_{X_2} . Поскольку эти величины на практике изменяются в широких пределах, то использован логарифмический масштаб, отображенный на рис. 3.25 в виде знака $\lg$ перед переменными вдоль координатных осей.

Рассмотренный случай отсутствия легирующих примесей иллюстрируется средней кривой в форме «птички», центральный участок («носик») которой соответствует высокоомному компенсированному полупроводнику, в котором $[V_M] \approx [V_X]$ и $n \approx p$. По обе стороны от этого участка компенсация нарушается за счет преобладающего действия вакансий в одной из подрешеток: слева $[V_X] > [V_M]$ и $n \approx [V_X^+]$, а справа $[V_M] > [V_X]$ и $p \approx [V_M^-]$.

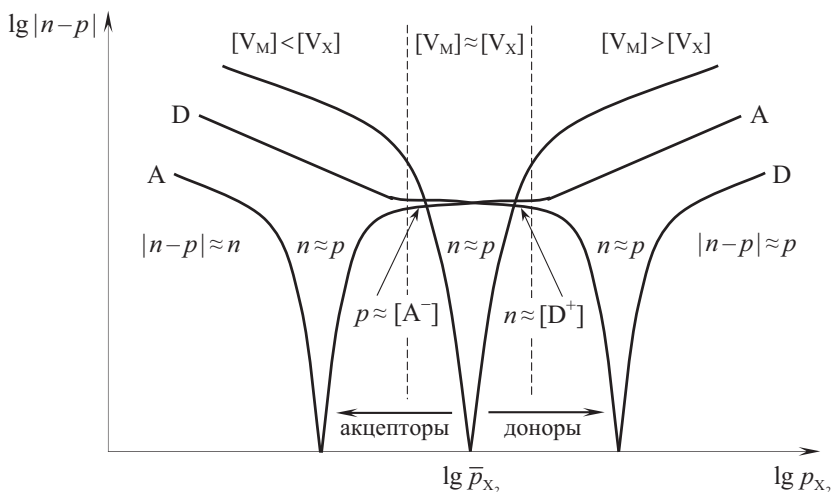


Рис. 3.25. Качественный вид зависимости концентрации носителей заряда в бинарном полупроводнике MX от парциального давления p_{X_2} в газовой фазе: центральная кривая соответствует отсутствию легирующих примесей; горизонтальными стрелками показано направление смещения центральной кривой при наличии легирующих примесей: для доноров кривая смещается вправо, для акцепторов — влево (соответствующие ветви отмечены буквами D и A)

Выясним характер зависимости концентрации дырок p от p_{X_2} в области высоких давлений, где $p \approx [V_M^-]$. Из закона действующих масс (3.11.2) получаем $p^2 \approx K_{VM}[V_M] = K_{VM}K_{III}[V_X]^{-1}$, при этом последнее равенство записано с учетом (3.11.1). Найдя из уравнения (3.11.7) значение $[V_X]$ и подставив его в предыдущее равенство, окончательно получаем искомое выражение:

$$p = \left(\frac{K_{III}(T)K_{VM}(T)}{K_X(T)} \right)^{1/2} p_{X_2}^{1/4}.$$

Следовательно, в области высоких давлений (при $p_{X_2} \gg \bar{p}_{X_2}$) концентрация дырок p нарастает как $p \propto p_{X_2}^{1/4}$ (правое «дырочное крыло птички»).

Аналогично можно показать, что в области низких давлений (при $p_{X_2} \ll \bar{p}_{X_2}$ или $p_M \gg \bar{p}_M$) концентрация электронов падает как $n \propto p_{X_2}^{-1/4}$ (левое «электронное крыло птички»). Особенностью средней кривой является симметрия «крыльев» относительно $\bar{p}_{X_2}$, которая нарушается при наличии примесей в кристалле.

Покажем, что если в полупроводник с дефектами по Шоттки вводятся примеси, то они смещают кривую в направлении

стрелок на рис. 3.25: вправо для донорной примеси и влево для акцепторной примеси.

Действительно, при $p_{X_2} = \bar{p}_{X_2}$, когда наблюдается полная взаимная компенсация вкладов вакансий, полупроводник с мелкими донорами, для которых $[D] \approx [D^+]$, все еще сохраняет электронную проводимость за счет ионизации доноров, так что $n \approx [D^+]$. В этом случае компенсация может быть достигнута в результате преобладающего влияния вакансий V_M акцепторного типа, что реализуется при $p_{X_2} > \bar{p}_{X_2}$. Следовательно, при введении донорной примеси «носик птички» смещается вправо, в результате чего при давлении $\bar{p}_{X_2}$ располагается ее «электронное крыло». Аналогично при наличии акцепторной примеси в кристалле, когда $p \approx [A^-]$ при $p_{X_2} = \bar{p}_{X_2}$, здесь расположено «дырочное крыло птички», а ее «носик» смещен влево.

Именно это показано на рис. 3.25 в виде двух кривых, смещенных относительно средней «птички», соответствующие ветви которых отмечены буквами D и A. В центральной части, ограниченной вертикальными пунктирными линиями, условие электронейтральности $n \approx p$ для компенсированного полупроводника без примесей (средняя кривая) заменяется на условия $n \approx [D^+] = \text{const}$ и $p \approx [A^-] = \text{const}$, соответственно, для доноров и акцепторов (две крайние кривые). По этой причине возникает асимметрия «крыльев» для этих кривых на рис. 3.25: в центральной части соответствующего «крыла» наблюдается близкий к горизонтальному ход, который в дальнейшем переходит в зависимость типа $n \propto p_{X_2}^{-1/4}$ (для левого «электронного крыла донорной птички») или $p \propto p_{X_2}^{1/4}$ (для правого «дырочного крыла акцепторной птички»).

3.12. Распределение амфотерной примеси в кристаллической решетке полупроводников

Известно, что в определенных случаях одна и та же примесь может занимать различные места в решетке полупроводника и соответственно находиться в разных ионизационных состояниях, выполняя функцию либо донора, либо акцептора, т. е. является амфотерной примесью. Например, атом меди, расположенный в междоузлиях германия или кремния, служит донором электронов, а при нахождении в узлах решетки выполняет функцию акцептора с кратностью ионизации, равной трем.

Другим примером амфотерного поведения примеси служит кремний в кристаллической решетке арсенида галлия.

Как известно (см. рис. 3.9 б), кристалл GaAs состоит из двух подрешеток: одна построена из атомов галлия Ga (элемент III группы периодической таблицы), другая — из атомов мышьяка As (элемент V группы). Кремний Si (элемент IV группы) может занимать узлы в обеих подрешетках GaAs, выполняя либо функцию донора на месте галлия, либо функцию акцептора на месте мышьяка.

Термодинамический анализ такой системы требует применения следующей совокупности уравнений:

- уравнение перераспределения атомов Si между подрешетками кристалла GaAs,

$$\text{Si}_{\text{Ga}} \rightleftharpoons \text{Si}_{\text{As}}, \quad K_{\text{Si}}(T) = \frac{[\text{Si}_{\text{As}}]}{[\text{Si}_{\text{Ga}}]}; \quad (3.12.1)$$

- уравнение ионизации атомов кремния,

$$\text{Si}_{\text{Ga}} \rightleftharpoons \text{Si}_{\text{Ga}}^+ + e^-, \quad K_{\text{Si}_{\text{Ga}}}(T) = \frac{[\text{Si}_{\text{Ga}}^+] n}{[\text{Si}_{\text{Ga}}]}; \quad (3.12.2)$$

$$\text{Si}_{\text{As}} \rightleftharpoons \text{Si}_{\text{As}}^- + e^+, \quad K_{\text{Si}_{\text{As}}}(T) = \frac{[\text{Si}_{\text{As}}^-] p}{[\text{Si}_{\text{As}}]}; \quad (3.12.3)$$

- уравнение межзонного возбуждения,

$$0 \rightleftharpoons e^- + e^+, \quad n_i^2(T) = np; \quad (3.12.4)$$

- уравнение электронейтральности кристалла,

$$n + [\text{Si}_{\text{As}}^-] = p + [\text{Si}_{\text{Ga}}^+]; \quad (3.12.5)$$

- уравнение сохранения общего числа атомов кремния, введенных в GaAs,

$$[\text{Si}_{\text{Ga}}] + [\text{Si}_{\text{As}}] + [\text{Si}_{\text{Ga}}^+] + [\text{Si}_{\text{As}}^-] = [\text{Si}]. \quad (3.12.6)$$

Численное решение задачи, описываемой системой уравнений (3.12.1)–(3.12.6), в принципе возможно только при известной температурной зависимости констант равновесия и заданной общей концентрации введенного кремния $[\text{Si}]$.

Используя законы действующих масс (3.12.2) и (3.12.3), запишем полную концентрацию атомов кремния, размещенных в двух подрешетках GaAs:

$$N_{\text{Si}_{\text{Ga}}} \equiv [\text{Si}_{\text{Ga}}] + [\text{Si}_{\text{Ga}}^+] = [\text{Si}_{\text{Ga}}] \left(1 + \frac{K_{\text{Si}_{\text{Ga}}}(T)}{n} \right), \quad (3.12.7)$$

$$N_{\text{Si}_{\text{As}}} \equiv [\text{Si}_{\text{As}}] + [\text{Si}_{\text{As}}^-] = [\text{Si}_{\text{As}}] \left(1 + \frac{K_{\text{Si}_{\text{As}}}(T)}{p} \right). \quad (3.12.8)$$

На основании этих уравнений проведем качественный анализ влияния $[\text{Si}]$ на концентрацию электронов проводимости в GaAs.

Из эксперимента известно, что при малом содержании кремния примерно 98% его атомов располагаются в подрешетке Ga, выполняя роль мелкого донора, т. е. $[\text{Si}_{\text{Ga}}] \ll [\text{Si}_{\text{Ga}}^+]$. Поскольку собственная концентрация для GaAs невелика ($n_i = 1,8 \cdot 10^6 \text{ см}^{-3}$ при $T = 300 \text{ K}$), то $p = n_i^2/n \ll n$. Тогда уравнения (3.12.5) и (3.12.6) позволяют записать $n \approx [\text{Si}_{\text{Ga}}^+] \approx [\text{Si}]$. Следовательно, концентрация электронов проводимости в арсениде галлия растет прямо пропорционально степени легирования его кремнием. Такая картина сохраняется вплоть до значений $[\text{Si}] \approx 10^{18} \text{ см}^{-3}$, как показано на рис. 3.26.

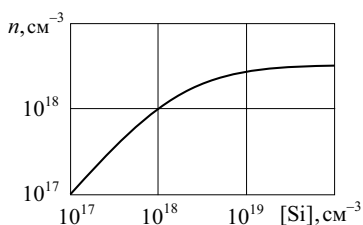


Рис. 3.26. Концентрация электронов проводимости как функция содержания Si в арсениде галлия

По мере дальнейшего увеличения $[\text{Si}]$ возрастание концентрации электронов n вызывает, в соответствии с (3.12.4), настолько существенное снижение концентрации дырок p , что реакция (3.12.3) заметным образом смещается вправо, обеспечивая рост концентрации $[\text{Si}_{\text{As}}^-]$. Из равенства (3.12.8) следует, что с уменьшением концентрации p полная величина $N_{\text{Si}_{\text{As}}}$ возрастает. Это означает,

что при достаточном увеличении общей концентрации вводимого кремния он начинает размещаться в подрешетке мышьяка, где является акцептором. Поэтому на рис. 3.26 увеличение концентрации электронов n замедляется и, начиная со значений $[\text{Si}] \approx 10^{19} \text{ см}^{-3}$, полностью прекращается, обеспечивая предельную растворимость $4 \cdot 10^{18} \text{ см}^{-3}$ для кремния как донора.

Если в арсенид галлия ввести другую донорную примесь (например, Te или Se), то возрастает растворимость кремния в подрешетке мышьяка. При наличии же акцепторной примеси (например, Zn или Cd) увеличивается растворимость кремния в подрешетке галлия. Такое заключение следует из соотношений (3.12.7) и (3.12.8) и полностью согласуется с общим правилом о взаимном влиянии ионизованных дефектов в полупроводниках, сформулированном в конце п. 3.10.

Выше был рассмотрен простейший вариант задачи с амфотерным поведением кремния в арсениде галлия. Реальная ситуация в этом случае (как, впрочем, и во многих других) оказывается значительно сложнее. При высокой концентрации кремния не все

его атомы располагаются в узлах галлиевой и мышьяковой под-решеток с образованием мелких доноров Si_{Ga} и акцепторов Si_{As} . Оставшиеся атомы кремния образуют между собой ассоциаты (например, донорно-акцепторного типа $\text{Si}_{\text{Ga}}-\text{Si}_{\text{As}}$) или комплексы совместно с собственными дефектами решетки и с неконтролируемыми примесями, которые всегда присутствуют в реальном полупроводниковом кристалле.

Модельное описание таких ассоциатов и комплексов обычно трудно осуществить, что усложняет задачу анализа, и чаще она решается с помощью эксперимента. Так, экспериментально установлено, что один из комплексов, образованный с участием Si, дает акцепторный уровень с энергией ионизации $\Delta E_A \approx 0,1$ эВ. Концентрация этих акцепторных комплексов зависит от ряда факторов и, в частности, увеличиваясь с ростом полной концентрации $[\text{Si}]$, она может превзойти концентрацию мелких доноров $[\text{Si}_{\text{Ga}}]$. В результате этого происходит инверсия проводимости арсенида галлия от электронного типа к дырочному.

Описанная ситуация изменяет ход кривой на рис. 3.26 в области больших значений $[\text{Si}]$. Инверсия типа проводимости GaAs за счет комплексообразования с участием амфотерных атомов Si используется на практике при выращивании светодиодных $p-n$ -структур методом жидкофазной эпитаксии. При этом управление инверсией осуществляют выбором температуры эпитаксиального роста, которая зависит не только от количества примеси кремния, введенной в жидкий раствор-расплав Ga-As, но и от других технологических факторов.

3.13. Контрольные вопросы

1. В чем состоит общность и различие между монокристаллическими, поликристаллическими и аморфными фазами?
2. Сформулируйте особенности химической связи в ионных, ковалентных (неполярных и полярных), металлических и молекулярных кристаллах.
3. Дайте определение таким понятиям, как: кристаллическая решетка, кристаллическая структура и кристаллическая система (сингония); трансляционная симметрия и точечная симметрия; точечные и пространственные группы симметрии; примитивные ячейки, сложные ячейки и ячейки Браве.
4. Каков геометрический смысл введения индексов Миллера в кристаллических решетках?

5. Какова связь между координационным числом и пространственными упаковками ионов, изображенными на рис. 3.4?
6. Дайте характеристику плотнейших упаковок шаров одинакового диаметра (ГПУ и КПУ).
7. Дайте характеристику двух типов пустот в плотнейших шаровых упаковках (ГПУ и КПУ).
8. Опишите основные типы кристаллических структур, характерных для металлических, диэлектрических и полупроводниковых кристаллов.
9. Объясните смысл аллотропных, полиморфных и политипных модификаций вещества.
10. В чем выражается политипизм карбида кремния SiC?
11. Дайте характеристику основных аллотропных модификаций углерода.
12. Охарактеризуйте дефекты кристаллической структуры как нульмерные (точечные), одномерные (линейные), двумерные (поверхностные) и трехмерные (объемные).
13. Дайте физическое обоснование классификации точечных дефектов в реальных кристаллах как собственных дефектов и примесных дефектов.
14. Чем определяются электронные свойства примесных и собственных дефектов в кристаллах полупроводников?
15. По какому признаку различают донорные, акцепторные и амфотерные примеси? Приведите примеры для конкретных полупроводников.
16. Дайте определение собственного полупроводника, компенсированного полупроводника и полупроводников с электронным и дырочным типом проводимости.
17. Каким образом точечные дефекты определяют электрические свойства полупроводников и диэлектриков?
18. Каким образом точечные дефекты определяют люминесцентные свойства полупроводников?
19. Каким образом точечные дефекты определяют термоэмиссионные свойства твердых тел?
20. Что такое поверхностные электронные состояния и какова их роль в работе полупроводниковых приборов?

21. В чем заключается основная идея квазихимического подхода к описанию точечных дефектов в полупроводниках?
22. Приведите примеры записи квазихимических реакций для примесных дефектов в полупроводниках с соответствующими законами действующих масс.
23. Сформулируйте физическое содержание электронно-дырочного равновесия в полупроводниках и проведите аналогию с ионным равновесием в водных растворах электролитов.
24. Сравните вывод закона Генри в форме (3.8.2) при квазихимическом подходе с аналогичным выводом в форме (2.6.4), основанным на условии фазового равновесия при молекулярном механизме растворения.
25. Сравните вывод закона Сивертса в форме (3.8.4) при квазихимическом подходе с аналогичным выводом в форме (2.6.12), основанным на условии химического равновесия при атомарном механизме растворения.
26. В чем состоит специфика ионного механизма растворения примесей в полупроводниках по сравнению с молекулярным и атомарным механизмами растворения?
27. Каким образом ассоциация заряженных дефектов влияет на их растворимость в полупроводниках?
28. Каковы механизмы тепловой генерации собственных дефектов по Шоттки и по Френкелю?
29. Приведите примеры записи квазихимических реакций генерации собственных дефектов с соответствующими законами действующих масс.
30. Сформулируйте правило взаимного влияния заряженных точечных дефектов (собственных и примесных) на их растворимость в кристаллической решетке полупроводника.
31. В чем заключается основная идея управления концентрацией собственных дефектов в полупроводниках при изотермическом отжиге кристаллов в парогазовой среде?
32. Качественно объясните ход кривых на рис. 3.25 с обоснованием влияния донорных и акцепторных примесей на положение кривой.
33. На основе уравнений (3.12.1)–(3.12.6) объясните поведение амфотерной примеси кремния в кристаллической решетке арсенида галлия.

Глава 4

УПРАВЛЕНИЕ ДИФФУЗИОННЫМИ И КИНЕТИЧЕСКИМИ ПРОЦЕССАМИ В ТЕХНОЛОГИЧЕСКИХ ЗАДАЧАХ ЭЛЕКТРОНИКИ

В двух предыдущих главах были рассмотрены основные физические закономерности, лежащие в основе равновесных процессов, управляющих фазовыми, химическими и электрохимическими превращениями веществ и процессами дефектообразования в кристаллах. Принципиальная возможность смещения термодинамического равновесия в нужном для рассматриваемого процесса направлении выяснялась на основании знака изобарно-изотермического потенциала ΔG (или изохорно-изотермического потенциала ΔF).

Однако практическая осуществимость процесса определяется не только термодинамическими факторами. Важную роль при этом, в частности при управлении скоростью протекания процессов, играют существенно неравновесные явления, лежащие в основе диффузионной и химической кинетики и широко используемые в разнообразных задачах электронной технологии.

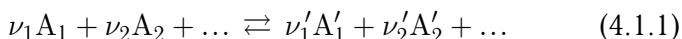
Основополагающие принципы управления как равновесными, так и неравновесными термодинамическими процессами были изложены в гл. 1. В частности, было установлено, что нарушение условий фазового равновесия вызывает появление в системе диффузионного и теплового потоков, выражаемых равенствами (1.18.3) и (1.18.5). При нарушении условия химического равновесия наблюдаются макроскопические изменения состава системы, вызванные протеканием химической реакции, скорость которой определяется в соответствии с основным постулатом химической кинетики в форме (1.18.13).

Задача данной главы — последовательное и более детальное рассмотрение физических явлений, возникающих при нарушении условий фазового и химического равновесия в применении

к технологическим проблемам электроники, включая разнообразные диффузионные и кинетические задачи, принципиально важные не только для твердотельной, но и для современной вакуумной электроники.

4.1. Механизм кинетики химических реакций. Уравнение Аррениуса

Задачей данного параграфа является обоснование закона Аррениуса, полученного из общих соображений для константы скорости химической реакции в форме (1.18.17). С этой целью рассмотрим обратимую химическую реакцию, записанную по аналогии с необратимой реакцией (1.18.8) в виде



В соответствии с основным постулатом химической кинетики (1.18.13) представляем скорости прямой и обратной реакции (4.1.1) в следующем виде:

$$\begin{aligned} v_1(t) &= k_1(T) c_{A_1}^{\nu_1}(t) c_{A_2}^{\nu_2}(t) \dots, \\ v_2(t) &= k_2(T) c_{A'_1}^{\nu'_1}(t) c_{A'_2}^{\nu'_2}(t) \dots, \end{aligned} \quad (4.1.2)$$

где $k_1(T)$ и $k_2(T)$ — константы скорости этих реакций, рассматриваемые как известные функции температуры.

Определим начальные условия, задавая при $t = 0$ ненулевыми только концентрации исходных (нештрихованных) компонентов, входящих в левую часть реакции (4.1.1):

$$c_{A_i}(0) \neq 0 \quad \text{и} \quad c_{A'_i}(0) = 0.$$

Эти начальные условия, примененные к (4.1.2), дают начальные скорости для прямой и обратной реакции:

$$v_1(0) \neq 0 \quad \text{и} \quad v_2(0) = 0.$$

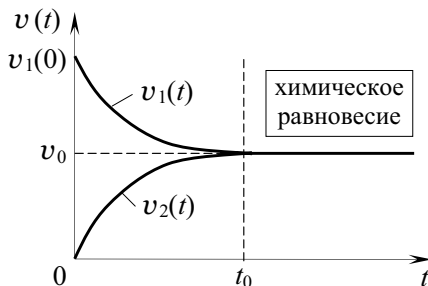


Рис. 4.1. Зависимость скорости прямой (v_1) и обратной (v_2) реакций от времени

По мере расходования начальных реагентов A_i и возникновения конечных продуктов реакции A'_i , скорость $v_1(t)$ будет с течением времени уменьшаться, а скорость $v_2(t)$ возрастать, как показано качественно на рис. 4.1. В момент времени t_0 возникает

химическое равновесие, при котором скорости прямой и обратной реакций становятся равными:

$$v_1(t_0) = v_2(t_0) \equiv v_0. \quad (4.1.3)$$

Равенство (4.1.3), начинающее действовать при $t \geq t_0$, содержит равновесные концентрации в выражениях (4.1.2), которые, как и ранее в п. 2.13, обозначаем чертой сверху, т. е. $\bar{c}_{A_i}$ и $\bar{c}_{A'_i}$. Тогда подстановка выражений (4.1.2) в равенство (4.1.3) дает

$$\frac{k_1(T)}{k_2(T)} = \frac{\bar{c}_{A'_1}^{\nu'_1} \bar{c}_{A'_2}^{\nu'_2} \dots}{\bar{c}_{A_1}^{\nu_1} \bar{c}_{A_2}^{\nu_2} \dots} \equiv \frac{\prod_k \bar{c}_{A_k}^{\nu'_k}}{\prod_n \bar{c}_{A_n}^{\nu_n}} = K_c(T), \quad (4.1.4)$$

где начальным (нештрихованным) и конечным (штрихованным) реагентам присвоены прежние индексы «н» и «к». Появление константы химического равновесия $K_c(T)$ обусловлено законом действующих масс (2.12.6), записанным в виде последнего равенства в формуле (4.1.4).

Таким образом, константа химического равновесия равняется отношению констант скорости для прямой и обратной реакций:

$$K_c(T) = \frac{k_1(T)}{k_2(T)}. \quad (4.1.5)$$

Этот результат является следствием динамического характера равновесия в химических реакциях (см. п. 1.14).

Воспользуемся уравнением изохоры Вант-Гоффа (2.14.11), которому подчиняется константа равновесия:

$$\frac{d \ln K_c(T)}{dT} = \frac{\Delta U}{RT^2}, \quad (4.1.6)$$

где ΔU — изохорный тепловой эффект реакции.

Подставляя соотношение (4.1.5) в уравнение (4.1.6) и полагая $\Delta U = U'_1 - U'_2$, получаем

$$\frac{d \ln k_1}{dT} - \frac{d \ln k_2}{dT} = \frac{U'_1}{RT^2} - \frac{U'_2}{RT^2}. \quad (4.1.7)$$

В уравнении (4.1.7) введены два энергетических параметра U'_1 и U'_2 , которые, как увидим далее, определяют высоту потенциальных барьеров, которую вынуждены преодолевать частицы, участвующие в прямой и обратной реакциях. Уравнение (4.1.7) можно расчленить на два уравнения,

$$\frac{d \ln k_1}{dT} = \frac{U'_1}{RT^2} + B \quad \text{и} \quad \frac{d \ln k_2}{dT} = \frac{U'_2}{RT^2} + B, \quad (4.1.8)$$

относящиеся к прямой и обратной реакциям и содержащие одинаковую величину B . Эта величина, как показывают многочисленные эксперименты, всегда равна нулю. Опуская индексы 1 и 2, записываем оба уравнения (4.1.8) в общем виде

$$\frac{d \ln k(T)}{dT} = \frac{\mathcal{E}}{k_B T^2}. \quad (4.1.9)$$

Здесь $\mathcal{E} = U'/\mathcal{N}_A$ — энергия активации химической реакции (прямой или обратной), отнесенная к одной частице, $\mathcal{N}_A$ — число Авогадро.

Уравнение (4.1.9) получило название **уравнения Аррениуса**. Его интегрирование при $\mathcal{E} = \text{const}$ дает температурный ход константы скорости в форме закона Аррениуса (1.18.17). Таким образом, подтверждается зависимость $k(T)$, которая в п. 1.18 была записана на основании выражения (1.18.15) для вероятности преодоления частицей потенциального барьера высотой $\mathcal{E}$.

Уточним смысл термина «преодоление потенциального барьера» на примере обратимой реакции замещения $A + BC \rightleftharpoons AB + C$. Соответствующая энергетическая диаграмма изображена на рис. 4.2.

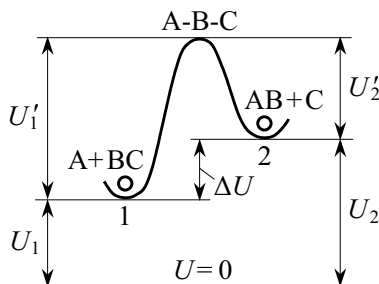


Рис. 4.2. Энергетическая диаграмма для обратимой реакции $A + BC \rightleftharpoons AB + C$

Рассмотрим механизм взаимодействия между частицами для прямой реакции. Исходная система $A + BC$, состоящая из атомов A и молекул BC , переходит из состояния 1 с внутренней энергией U_1 , преодолевая потенциальный барьер высотой U'_1 , в состояние 2 для системы $AB + C$, состоящей из молекул AB и атомов C , с энергией U_2 . Максимуму диаграммы соответствует так называемое *переходное состояние*, которое система выбирает как наиболее вероятное среди возможных промежуточных состояний. Здесь система находится в возбужденном состоянии, при котором возникает неустойчивый *активированный комплекс* в виде трех связанных атомов $A-B-C$.

Появление активированного комплекса, сопровождающееся возрастанием энергии системы, обусловлено двумя факторами.

Во-первых, при сближении атома А и молекулы ВС они испытывают силы взаимного отталкивания как структуры с полностью насыщенными электронными конфигурациями. Во-вторых, по мере их сближения ослабляется связь В–С внутри молекулы из-за влияния полей атома А.

Оба эти фактора увеличивают энергию системы, что и объясняет возрастающий характер потенциального профиля на рис. 4.2 при движении системы из состояния 1 в направлении к переходному состоянию. В этом состоянии парные химические связи внутри системы из трех атомов не имеют избирательного характера, так что возникает некоторая конфигурация А–В–С, названная активированным комплексом, которая обладает максимальной энергией.

Естественно, что переходное состояние не является устойчивым, и активированный комплекс при преобладающем парном взаимодействии между атомами А и В может образовать новую молекулу АВ. В результате этого энергия понижается, что соответствует «скатыванию» системы с максимума энергетической диаграммы в правый потенциальный минимум 2. Величина ΔU была введена в формуле (4.1.7) как разность энергий активации для прямой ($U'_1 \equiv \mathcal{E}_1 \mathcal{N}_A$) и обратной ($U'_2 \equiv \mathcal{E}_2 \mathcal{N}_A$) реакций. Как видно из рис. 4.2, одновременно с $\Delta U = U'_1 - U'_2$ имеем $\Delta U = U_2 - U_1$, что определяет тепловой эффект реакции в изохорных условиях. Следовательно, энергии активации для прямой и обратной реакций отличаются друг от друга на величину теплового эффекта ($\mathcal{E}_1 - \mathcal{E}_2 = \Delta U / \mathcal{N}_A$).

Таким образом, при столкновении частиц в химическое взаимодействие вступают лишь достаточно активные частицы, имеющие избыток энергии (в виде кинетической энергии поступательного движения молекулы или потенциальной энергии колебательного и вращательного движения атомов в молекуле) по сравнению со средней энергией частиц при данной температуре. Эта избыточная энергия и составляет энергию активации, которая идет на преодоление активными частицами потенциального барьера для того, чтобы принимать участие в акте химического взаимодействия.

Для управления скоростью химической реакции (прямой или обратной) в систему нередко вводят вещества, которые напрямую не участвуют в реакции, а предназначены для понижения или повышения активационного барьера реакции. Первые, называемые *катализаторами*, ускоряют реакцию, а вторые, называемые *ин-*

гибиторами (отрицательными катализаторами), замедляют ее. Это происходит по той причине, что атомы катализатора лишь участвуют в образовании активированного комплекса, уменьшая или увеличивая энергию активации, но не входят в состав конечных продуктов. Следовательно, катализатор не влияет на константу равновесия $K_c = k_1/k_2$, так как одновременно изменяет обе константы скорости k_1 и k_2 для прямой и обратной реакций, сохраняя неизменным их отношение.

4.2. Механизмы диффузии атомов в твердом теле

Диффузия в твердом теле представляет собой процесс перемещения инородных атомов растворенного вещества (примесей) в кристаллической решетке растворителя, называемый *гетеродиффузией*. Кроме примесных атомов, диффундировать могут и собственные дефекты решетки при их неоднородном распределении, что принято называть *самодиффузией*. Количественной характеристикой процесса является коэффициент диффузии D , введенный уравнением (1.18.3), выражающим первый закон Фика. Он связывает градиент концентрации диффундирующих атомов с плотностью их потока.

С микроскопической точки зрения процесс диффузии заключается в случайных перемещениях атомов в объеме твердого тела. Перемещения происходят в форме последовательных скачков из одного потенциального минимума в другой, которые разделены потенциальным барьером, называемым *энергией активации* процесса диффузии. Величина энергии активации $\mathcal{E}$ зависит от конкретного механизма диффузии.

Различают следующие механизмы диффузии, энергетические диаграммы которых приведены на рис. 4.3.

1. *Обменный механизм* реализуется при взаимном обмене местами двух атомов, расположенных в соседних узлах решетки (рис. 4.3 а).

2. *Кольцевой механизм* реализуется при согласованном обмене местами группы атомов, расположенных по окружности (рис. 4.3 б). Оба обменных механизма характерны для процесса самодиффузии.

3. *Вакансионный механизм* реализуется при движении атомов по вакансиям (рис. 4.3 в). В районе вакансии энергетический барьер $\mathcal{E}$ понижен по сравнению с обычным потенциальным профилем (пунктирные кривые), так что атом может его достаточно легко преодолеть. Движение атома по вакансиям сопровождается встречным движением вакансии как собственного дефекта

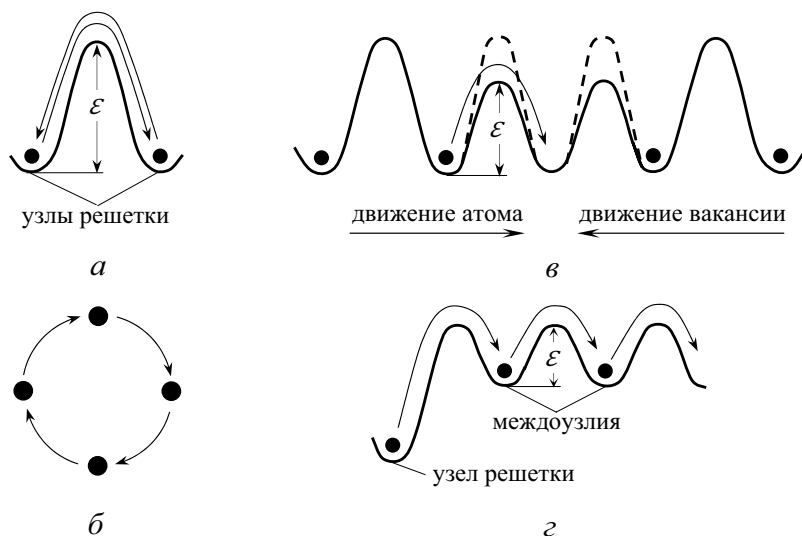


Рис. 4.3. Энергетические диаграммы и перемещения атомов для разных механизмов диффузии: *а* — обменный механизм, *б* — кольцевой механизм, *в* — вакансионный механизм, *г* — междоузельный механизм

решетки. Очевидно, что скорость диффузии пропорциональна концентрации вакансий, которая сильно зависит от температуры. Данный механизм характерен как для процесса самодиффузии, так и для гетеродиффузии примесных атомов замещения, находящихся в узлах кристаллической решетки.

4. *Междоузельный механизм* осуществляется путем движении атомов по междоузлиям (рис. 4.3 *г*). Энергия атома в междоузлии выше, чем для атома в узле, однако потенциальный барьер $\mathcal{E}$, разделяющий соседние междоузлия, ниже. Поэтому междоузельный механизм приводит к большим скоростям диффузии, нежели предыдущие механизмы. Данный механизм характерен для гетеродиффузии примесных атомов внедрения, которые при растворении располагаются в междоузлиях решетки.

5. *Комбинированные механизмы* (диссоциативный и эстафетный) имеют место в том случае, если атомы растворенного вещества занимают как узлы, так и междоузлия твердого растворителя. Так как для примесных атомов коэффициент диффузии по вакансиям значительно меньше, чем по междоузлиям, то при *диссоциативном механизме* атомы примеси покидают узлы и перемещаются лишь по междоузлиям. При *эстафетном механизме* междоузельный атом примеси вытесняет из соседнего узла собственный атом решетки, который в свою очередь замещает

примесный атом в ближайшем соседнем узле, вытесняя его в междоузлие и т. д.

Кроме вышеуказанных механизмов диффузии, существенную роль в дефектных материалах играет диффузия вдоль дислокаций и по границам зерен, а также поверхностная диффузия адсорбированных атомов и молекул.

Дадим на качественном уровне микроскопическое обоснование закона Аррениуса (1.18.16) для коэффициента диффузии атомов в твердых телах, воспользовавшись аналогией с диффузионным движением молекул в газах. В обоих случаях перемешивание неоднородно распределенных веществ происходит в результате их теплового движения, однако в твердых телах из-за больших сил сцепления между частицами эти процессы протекают в миллиарды раз медленнее, чем в газах. Скорость перемешивания веществ в твердом состоянии определяется вероятностью w (или частотой) перехода частицы в единицу времени из одного положения в другое в соответствии с законом Больцмана (1.18.15). Френкель ввел в рассмотрение вместо w обратную величину $\tau = 1/w$, назвав ее *временем оседлой жизни*, которое характеризует среднее время пребывания атома в минимуме потенциального рельефа решетки.

Согласно соотношению (1.18.15), для времени оседлой жизни атома справедливо выражение

$$\tau = \tau_0 e^{\mathcal{E}/k_B T}, \quad (4.2.1)$$

где $\mathcal{E}$ — высота энергетического барьера, разделяющего соседние потенциальные минимумы (энергия активации). Постоянная τ_0 по порядку величины близка к периоду колебаний атома около положения равновесия и определяется упругостью межатомных связей (см. ниже п. 5.9). Поэтому она не сильно изменяется для разных кристаллов, так что обычно считают $\tau_0 \approx 10^{-13}$ с. Как и следовало ожидать, с повышением энергии активации $\mathcal{E}$ и с понижением температуры T возрастает среднее время жизни атома в потенциальном минимуме.

Таким образом, каждый атом в среднем через время τ перескакивает в соседний потенциальный минимум, смещаясь в пространстве на период решетки a . Это позволяет ввести в рассмотрение среднюю скорость перемещения атомов в кристалле, определив ее как $v = a/\tau$. Отсюда следует аналогия с газами, для которых средняя тепловая скорость молекул равна $\bar{v} = \bar{\lambda}/\bar{\tau}$, где $\bar{\lambda}$ и $\bar{\tau}$ — средние длина и время свободного пробега молекулы между двумя столкновениями. Кинетическая теория газов дает

известное выражение для коэффициента диффузии молекул:

$$D = \frac{1}{3} \bar{v}^2 \bar{\tau} = \frac{1}{3} \frac{\bar{\lambda}^2}{\bar{\tau}}.$$

По аналогии с этой формулой, отождествляя $\bar{v}$ с $v = a/\tau$ и $\bar{\tau}$ с $\tau = \tau_0 \exp(\mathcal{E}/k_B T)$, записываем коэффициент диффузии атомов в твердом теле в следующей форме:

$$D = \alpha v^2 \tau = \alpha \frac{a^2}{\tau} \equiv D_0 e^{-\mathcal{E}/k_B T}, \quad (4.2.2)$$

где $D_0 \equiv \alpha a^2/\tau_0$ и $\mathcal{E}$ — энергия активации процесса диффузии. Коэффициент пропорциональности α принимает различные значения для кристаллических решеток разного типа (в частности для простой, гранецентрированной и объемноцентрированной кубических решеток α принимает значения 1/6, 1/12 и 1/24).

Формула (4.2.2) совпадает с выражением (1.18.16), записанным на основании общих энергетических соображений, что подтверждает термоактивационную природу процесса диффузии. Значения коэффициента D_0 и энергии активации $\mathcal{E}$, входящих в выражение (4.2.2), существенным образом зависят от механизма диффузии, имеющего место в реальных кристаллах.

В заключение приведем ориентировочные значения D_0 и $\mathcal{E}$ для самодиффузии атомов кремния Si, а также для гетеродиффузии атомов фосфора P и бора B в кремнии, выполняющих, соответственно, функцию донора и акцептора:

$D_0 = 1,8 \cdot 10^3 \text{ см}^2/\text{с}$ (Si), $D_0 = 10,5 \text{ см}^2/\text{с}$ (P в Si), $D_0 = 5,1 \text{ см}^2/\text{с}$ (B в Si);

$\mathcal{E} = 4,48 \text{ эВ}$ (Si), $\mathcal{E} = 3,69 \text{ эВ}$ (P в Si), $\mathcal{E} = 3,70 \text{ эВ}$ (B в Si).

Эти величины и подобные им для других полупроводников практически вычисляют на основе экспериментально разработанных методик, и они могут быть найдены в справочной литературе.

4.3. Законы Фика. Начальные и граничные условия в задачах диффузии

Рассмотрим произвольный объем V , внутри которого распределены частицы с неоднородной концентрацией $c(\mathbf{r}, t)$, зависящей также от времени t в нестационарном состоянии. Изменение числа частиц в объеме V связано с их вытеканием через поверхность S , ограничивающую этот объем. Если выделить на поверхности элементарную площадку dS с единичной внешней нормалью $\mathbf{n}$, то количество частиц, *вытекающих* через нее за

единицу времени, равно $J_n dS$, где $J_n = \mathbf{n} \cdot \mathbf{J}$ — нормальная компонента плотности потока частиц $\mathbf{J}$.

Скорость убыли частиц внутри объема V дается поверхностным интегралом $\oint_S J_n dS$, поэтому очевидным становится следующее интегральное соотношение:

$$\frac{\partial}{\partial t} \int_V c(\mathbf{r}, t) dV = - \oint_S J_n dS. \quad (4.3.1)$$

С учетом теоремы Остроградского–Гаусса,

$$\int_V \operatorname{div} \mathbf{J} dV = \oint_S (\mathbf{n} \cdot \mathbf{J}) dS,$$

соотношение (4.3.1) принимает следующий вид:

$$\int_V \left(\frac{\partial c}{\partial t} + \operatorname{div} \mathbf{J} \right) dV = 0.$$

Поскольку выбор объема V произволен, то последний интеграл равен нулю лишь при обращении в нуль подынтегрального выражения, что дает известное *уравнение непрерывности*:

$$\frac{\partial c}{\partial t} + \operatorname{div} \mathbf{J} = 0. \quad (4.3.2)$$

Следует иметь в виду, что при выводе уравнения (4.3.2) были исключены из рассмотрения процессы появления и исчезновения частиц внутри объема V как результат, например, протекания химических реакций с участием частиц рассматриваемого сорта. С учетом этого уравнение непрерывности (4.3.2) должно содержать в правой части дополнительные слагаемые, включающие скорости таких реакций.

Используем соотношение (1.18.3), связывающее плотность потока частиц с их концентрацией (индекс i для простоты опущен):

$$\mathbf{J}(\mathbf{r}, t) = -D \operatorname{grad} c(\mathbf{r}, t), \quad (4.3.3)$$

называемое *первым законом Фика* (см. сноску на с. 150).

Подстановкой (4.3.3) в уравнение (4.3.2) получаем *второй закон Фика*. Этот закон описывает изменение частиц во времени и в пространстве с помощью диффузионного уравнения,

$$\frac{\partial c(\mathbf{r}, t)}{\partial t} = D \nabla^2 c(\mathbf{r}, t), \quad (4.3.4)$$

где введен оператор Лапласа $\nabla^2 c \equiv \operatorname{div} \operatorname{grad} c$.

Коэффициент диффузии в уравнении (4.3.4) вынесен за знак оператора Лапласа, хотя в общем случае он может быть функцией координат. Поэтому более общей формой записи для правой части уравнения (4.3.4) является $\operatorname{div}(D \operatorname{grad} c)$.

Диффузионное уравнение (4.3.4) является дифференциальным уравнением в частных производных. Так как оно содержит производную по времени первого порядка и пространственные производные второго порядка, то для его решения необходимо задать одно начальное условие (при $t = 0$) и два граничных условия (в случае одномерной пространственной задачи).

При решении диффузионных задач обычно используют следующие два варианта границы:

- **отражающая граница** реализуется при отсутствии нормальной компоненты потока частиц через границу S , а именно

$$J_n(t) \equiv \mathbf{n} \cdot \mathbf{J}(t)|_S = -D \frac{\partial c(\mathbf{r}, t)}{\partial n} \Big|_S = 0; \quad (4.3.5)$$

- **связывающая граница** не требует отсутствия нормальной компоненты потока на границе S и реализуется путем задания поверхностной концентрации частиц $c_s(t)$ (в частном случае постоянной или равной нулю), а именно

$$c_s(t) = c(\mathbf{r}, t)|_S \quad (= \text{const или } 0). \quad (4.3.6)$$

При использовании модели полуграниченного тела одно из граничных условий (4.3.5) или (4.3.6) дополняют требованием поведения концентрации частиц на бесконечности, исходя из физической постановки задачи.

Среди практических задач, наиболее часто встречающихся в технологии электронных приборов, по виду начальных условий можно выделить следующие два типа:

- **задачи на удаление вещества** из твердого тела (обезгаживание деталей вакуумных приборов и геттерирование или испарение примесей из полупроводников):

$$c(\mathbf{r}, t)|_{t=0} = c_0(\mathbf{r}), \quad \text{когда при } t > 0 \quad c(\mathbf{r}, t) < c_0(\mathbf{r}); \quad (4.3.7)$$

- **задачи на введение вещества** в твердое тело (легирование полупроводников примесями и поглощение газов материалами):

$$c(\mathbf{r}, t)|_{t=0} = c_0(\mathbf{r}), \quad \text{когда при } t > 0 \quad c(\mathbf{r}, t) > c_0(\mathbf{r}), \quad (4.3.8)$$

где в частном случае может быть $c_0(\mathbf{r}) = 0$.

4.4. Диффузионные задачи на удаление вещества из твердого тела

Можно указать две практические задачи, когда целью технологической операции является удаление атомов вещества, растворенного в объеме твердого тела.

Во-первых, сюда относится задача обезгаживания элементов конструкции вакуумных приборов и устройств. Эту операцию обычно проводят перед сборкой прибора путем отжига деталей в вакууме или в атмосфере очищенного водорода. Окончательное обезгаживание всей арматуры достигается путем вакуумной обработки прибора на откачном посту.

Во-вторых, в технологии полупроводниковых материалов для очистки их от примесей используется метод геттерирования. Он заключается в нанесении на поверхность полупроводника материала, называемого геттером, который (обычно в расплавленном состоянии) эффективно экстрагирует растворенные в полупроводнике примеси. Так, для удаления быстро диффундирующих примесей меди, золота и железа, создающих нежелательные глубокие ловушки в кремнии, в качестве геттера используют фосфорно-силикатное стекло (см. п. 4.8).

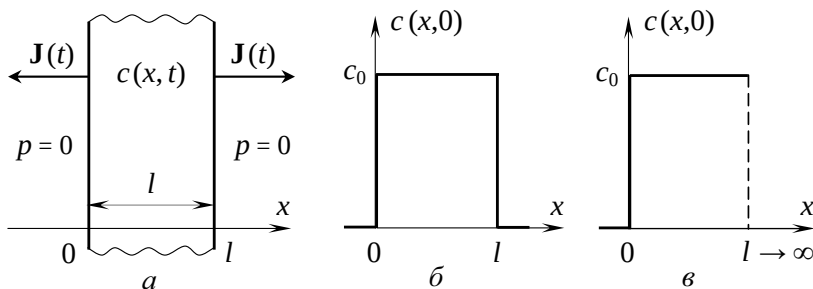


Рис. 4.4. К решению задачи диффузионного обезгаживания пластины, толщину l (а) с однородным начальным распределением растворенного газа (б), и полуограниченного тела (при $l \rightarrow \infty$) с однородным начальным распределением растворенного газа (в). Векторы $\mathbf{J}(t)$ показывают диффузионные потоки газовой выделения через границы пластины

Закономерности диффузионного удаления вещества из твердого тела будем изучать на примере вакуумного обезгаживания пластины толщиной l с неограниченными поперечными размерами, которая показана на рис. 4.4 а. В этом случае диффузионная задача становится одномерной и для нее второй закон Фика

в форме (4.3.4) принимает следующий вид:

$$\frac{\partial c(x, t)}{\partial t} = D \frac{\partial^2 c(x, t)}{\partial x^2}. \quad (4.4.1)$$

Величина $c(x, t)$, входящая в уравнение (4.4.1), представляет собой искомое распределение растворенного газа в пластине, изменяющееся во времени в нестационарных условиях.

Начальное распределение газа по толщине пластины принимаем однородным (рис. 4.4 б), так что начальное условие типа (4.3.7) в данном случае имеет вид

$$c(x, t)|_{t=0} = \begin{cases} 0 & \text{при } x < 0 \text{ и } x > l, \\ c_0 & \text{при } 0 \leq x \leq l. \end{cases} \quad (4.4.2)$$

Для записи граничных условий при $x = 0$ и $x = l$ воспользуемся законами растворения газов в твердом теле. Как известно (см. п. 2.6), эти законы, называемые законами Генри и Сивертса, справедливы лишь в равновесных условиях. К рассматриваемой неравновесной задаче они в общем случае неприменимы, если речь идет о внутренних точках обезгаживаемой пластины.

Однако равновесие между газом и поверхностными слоями твердого тела обычно устанавливается достаточно быстро. Это позволяет считать, что концентрация c_s растворенного газа на поверхности (отмечаемая нижним индексом s), входящая в граничное условие (4.3.6), подчиняется законам Генри и Сивертса, т. е. определяется парциальным давлением p этого газа (без индекса i):

$$c_s(T, p) = s(T)p^m. \quad (4.4.3)$$

Здесь $s(T)$ — температурно-зависимый коэффициент растворимости, m — показатель степени, равный 1 или 1/2 для молекулярного или атомарного механизма растворения двухатомных газов в соответствии с законами Генри или Сивертса.

Будем считать скорость диффузионного выделения газов много меньшей скорости их откачки вакуумным насосом, что обеспечивает в газовой среде, окружающей пластину, очень низкое давление (в пределе $p = 0$ — модель абсолютного вакуума). В этом случае граничное условие (4.3.6) с учетом (4.4.3) позволяет записать

$$c_s(t) \equiv c(x, t)|_{x=0} = c(x, t)|_{x=l} = 0 \quad \text{при } t > 0. \quad (4.4.4)$$

Уравнение в частных производных (4.4.1) при начальных и граничных условиях (4.4.2) и (4.4.4) будем решать по методу

Фурье путем разделения переменных. Для этого неизвестная функция двух переменных $c(x, t)$ представляется в виде произведения двух искоемых функций X и T , зависящих соответственно только от x и t :

$$c(x, t) = X(x) T(t). \quad (4.4.5)$$

Подстановка (4.4.5) в уравнение (4.4.1) с последующим делением на $X(x) T(t)$ дает

$$\frac{X''(x)}{X(x)} = \frac{1}{D} \frac{T'(t)}{T(t)} = -\lambda^2, \quad (4.4.6)$$

где производные обозначены штрихами.

Уравнение (4.4.6) приведено к форме с разделением переменных, поскольку его левая часть зависит только от x , а правая — только от t . Функции разных переменных могут быть равными друг другу, лишь будучи константой, обозначенной $-\lambda^2$ и называемой константой разделения, которая подлежит определению. Из уравнения (4.4.6) следуют два обыкновенных уравнения:

$$\frac{d^2 X(x)}{dx^2} + \lambda^2 X(x) = 0,$$

$$\frac{dT(t)}{dt} + \lambda^2 D T(t) = 0.$$

Общее решение этих уравнений может быть записано в следующем виде (здесь верхние штрихи не являются обозначением производных):

$$X(x) = A' \sin \lambda x + B' \cos \lambda x, \quad (4.4.7)$$

$$T(t) = C' \exp(-\lambda^2 D t). \quad (4.4.8)$$

Граничное условие (4.4.4) при $x = 0$ требует $X(0) = 0$, что после подстановки в равенство (4.4.7) дает $B' = 0$ при $A' \neq 0$. Другое граничное условие (4.4.4) при $x = l$, записанное в виде $X(l) = 0$, удовлетворяется при $\sin \lambda l = 0$. Отсюда получаем допустимые значения константы разделения:

$$\lambda_n = (2n + 1) \frac{\pi}{l}, \quad n = 0, 1, 2, \dots \quad (4.4.9)$$

При этом отбрасываем значения $\lambda_n = 2n\pi/l$, дающие решения, антисимметричные относительно середины пластины, как физически нереализуемые при одновременном обезгаживании пластины с двух сторон.

Подстановка полученных результатов в (4.4.7) и (4.4.8) позволяет записать общее решение путем суммирования слагаемых $c_n(x, t)$ типа (4.4.5) с константой разделения (4.4.9):

$$\begin{aligned} c(x, t) &= \sum_{n=0}^{\infty} A_n \sin \lambda_n x \exp(-\lambda_n^2 D t) \equiv \\ &\equiv \sum_{n=0}^{\infty} A_n \sin \left[(2n+1) \frac{\pi x}{l} \right] \exp \left\{ - \left[(2n+1) \frac{\pi}{l} \right]^2 D t \right\}. \end{aligned} \quad (4.4.10)$$

Неизвестная амплитудная постоянная $A_n \equiv A'_n C'_n$ находится из начального условия (4.4.2), записанного с помощью (4.4.10) при $t = 0$ в следующем виде:

$$c(x, 0) = \sum_{n=0}^{\infty} A_n \sin \lambda_n x = \begin{cases} 0 & \text{при } x < 0 \text{ и } x > l, \\ c_0 & \text{при } 0 \leq x \leq l. \end{cases} \quad (4.4.11)$$

Иными словами, амплитуды A_n являются фурье-коэффициентами для разложения по синусам прямоугольного пространственного импульса, изображенного на рис. 4.4 б. Тогда коэффициенты Фурье ряда (4.4.11) равны

$$A_n = \frac{2}{l} \int_0^l c_0 \sin \lambda_n x \, dx = \frac{4c_0}{\lambda_n l}. \quad (4.4.12)$$

Подстановка фурье-коэффициентов (4.4.12) в выражение (4.4.10) дает искомое нестационарное распределение концентрации в форме ряда

$$\begin{aligned} c(x, t) &= \frac{4c_0}{\pi} \sum_{n=0}^{\infty} \frac{1}{2n+1} \sin \left[(2n+1) \frac{\pi x}{l} \right] \times \\ &\times \exp \left\{ - \left[(2n+1) \frac{\pi}{l} \right]^2 D t \right\}. \end{aligned} \quad (4.4.13)$$

Скорость удельного газовыделения, измеряемая количеством газа, выделяющегося из толщи материала через единицу площади в единицу времени, определяется выражением (см. рис. 4.4)

$$J(t) = D \left. \frac{\partial c(x, t)}{\partial x} \right|_{x=0} = \frac{4Dc_0}{l} \sum_{n=0}^{\infty} \exp \left\{ - \left[(2n+1) \frac{\pi}{l} \right]^2 D t \right\}. \quad (4.4.14)$$

Пространственное распределение концентрации растворенного газа, построенное по формуле (4.4.13), показано на рис. 4.5

для пяти моментов времени. Среди этих кривых выделены жирными линиями две кривые, соответствующие моментам времени $t_1 = 0,0263 l^2/D$ и $t_2 = 0,045 l^2/D$, которые отделяют начальный и конечный этапы обезгаживания.

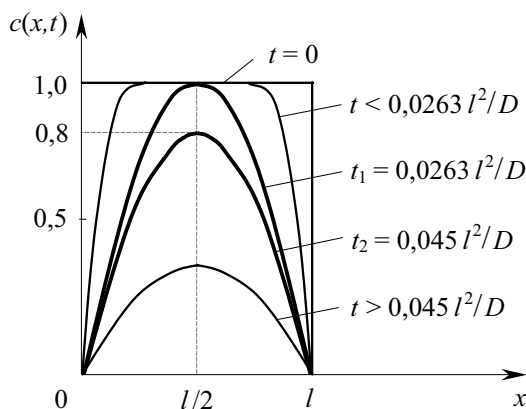


Рис. 4.5. Качественный вид распределения растворенного в пластине газа для разных моментов времени, среди которых значения $t_1 = 0,0263 l^2/D$ и $t_2 = 0,045 l^2/D$ разделяют начальный (при $t < t_1$) и конечный (при $t > t_2$) этапы обезгаживания

Ряды, входящие в формулы (4.4.13) и (4.4.14), содержат экспоненту, сильно затухающую с ростом n , что обеспечивает их быструю сходимость. Легко убедиться в том, что при условии $\pi^2 Dt/l^2 \geq 0,45$ можно ограничиться лишь первым членом ряда ($n=0$), делая при этом ошибку, не превышающую 1%. Эта ситуация соответствует конечному этапу обезгаживания, реализуемому по прошествии времени $t \geq t_2 = 0,045 l^2/D$.

С другой стороны, на начальном этапе обезгаживания (при $t < t_1$) вплоть до момента времени $t_1 = 0,0263 l^2/D$, когда начальная концентрация c_0 сохраняется только в центре пластины (см. рис. 4.5), каждая ее сторона обезгаживается независимо от другой. В этом случае работает модель полуограниченного тела, реализуемая для левой стороны пластины при $l \rightarrow \infty$, когда начальное распределение концентрации растворенных газов имеет вид, показанный на рис. 4.4 в.

Рассмотрим по отдельности эти два этапа обезгаживания.

1. Начальный этап обезгаживания имеет место при $t < t_1 = 0,0263 l^2/D$ и описывается моделью *полуограниченного тела*, для которой диффузионное уравнение (4.4.1) решается при

следующем начальном условии (ср. формулу (4.4.2))

$$c(x, t)|_{t=0} = \begin{cases} 0 & \text{при } x < 0, \\ c_0 & \text{при } 0 \leq x < \infty, \end{cases} \quad (4.4.15)$$

и граничных условиях вида (ср. формулу (4.4.4))

$$c(x, t)|_{x=0} = 0 \quad \text{и} \quad c(x, t)|_{x \rightarrow \infty} \rightarrow c_0 \quad \text{при } t > 0. \quad (4.4.16)$$

Начальное условие (4.4.15) соответствует профилю распределения растворенных газов, изображенному на рис. 4.4 в.

Искомое решение граничной задачи, поставленной уравнениями (4.4.1), (4.4.15) и (4.4.16), имеет вид

$$c(x, t) = c_0 \operatorname{erf} \frac{x}{2\sqrt{Dt}}. \quad (4.4.17)$$

Здесь использовано общепринятое обозначение для *функции ошибок*,

$$\operatorname{erf} z = \frac{2}{\sqrt{\pi}} \int_0^z \exp(-\xi^2) d\xi, \quad (4.4.18)$$

название которой происходит от английских слов *error function*, а график функции показан на рис. 4.6. Аргументом этой функции является комбинация координаты x и времени t в форме $z = x/2\sqrt{Dt}$, в то время как нормировочный множитель выбран таким, чтобы обеспечить условия $\operatorname{erf}(\pm\infty) = \pm 1$.

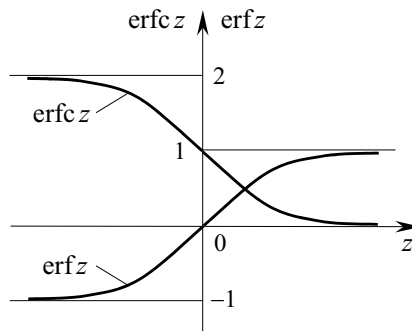


Рис. 4.6. Функция ошибок $\operatorname{erf} z$ и дополнительная функция ошибок $\operatorname{erfc} z$

В справедливости решения граничной задачи для модели полуграниченного тела в форме выражения (4.4.17) легко убедиться простой подстановкой этого выражения в уравнение (4.4.1).

При этом надо учитывать правила дифференцирования интеграла (4.4.18) по переменному верхнему пределу, так что

$$\begin{aligned}\frac{\partial}{\partial t}(\operatorname{erf} z) &= \frac{\partial}{\partial z}(\operatorname{erf} z) \frac{\partial z}{\partial t}, & \frac{\partial z}{\partial t} &= -\frac{x}{4t\sqrt{Dt}}; \\ \frac{\partial}{\partial x}(\operatorname{erf} z) &= \frac{\partial}{\partial z}(\operatorname{erf} z) \frac{\partial z}{\partial x}, & \frac{\partial z}{\partial x} &= \frac{1}{2\sqrt{Dt}}; \\ \frac{\partial}{\partial z}(\operatorname{erf} z) &= \frac{2}{\sqrt{\pi}} e^{-z^2}.\end{aligned}\quad (4.4.19)$$

Качественный вид решения (4.4.17) показан на рис. 4.7 для трех моментов времени, не превышающих значение $0,0263 l^2/D$, когда работает модель полуограниченного тела.

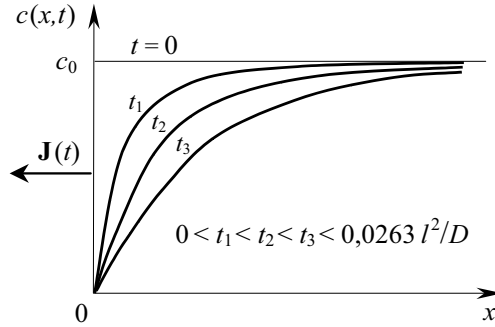


Рис. 4.7. Качественный вид распределения растворенного газа для начального этапа обезгаживания пластины в разные моменты времени

Скорость удельного газовыделения, полученная из выражения (4.4.17) с учетом (4.4.19), равна

$$J(t) = D \left. \frac{\partial c(x,t)}{\partial x} \right|_{x=0} = c_0 \sqrt{\frac{D}{\pi t}}. \quad (4.4.20)$$

Формулы (4.4.17) и (4.4.20) перестают работать, начиная с того момента времени $t_1 = 0,0263 l^2/D$, при котором происходит смыкание независимых профилей обезгаживания левой и правой стороны пластины. При промежуточных значениях времени t (когда $0,0263 l^2/D < t < 0,045 l^2/D$) для нахождения концентрации растворенного газа $c(x,t)$ и удельного газовыделения $J(t)$ необходимо применять общие формулы (4.4.13) и (4.4.14).

2. Конечный этап обезгаживания наступает при $t > t_2 = 0,045 l^2/D$, когда доминирующий вклад в общие выражения

(4.4.13) и (4.4.14) вносит первый член ряда с номером $n = 0$. В этом случае

$$c(x, t) = \frac{4c_0}{\pi} \sin\left(\frac{\pi x}{l}\right) \exp\left[-\frac{\pi^2 D}{l^2} (t - t_2)\right], \quad (4.4.21)$$

$$J(t) = D \left. \frac{\partial c(x, t)}{\partial x} \right|_{x=0} = \frac{4Dc_0}{l} \exp\left[-\frac{\pi^2 D}{l^2} (t - t_2)\right]. \quad (4.4.22)$$

Следует иметь в виду, что конечной целью обезгаживания является не достижение предельно низкой концентрации остаточных растворенных газов, а обеспечение малых газовых потоков $J_{\text{экс}}$ в эксплуатационном режиме работы прибора. Для этого вакуумная термическая обработка деталей прибора предварительно проводится в более жестком тепловом режиме, обеспечивающем за время обработки $t_{\text{обр}}$ достаточно малые газовые потоки, так что $J(t_{\text{обр}}) < J_{\text{экс}}$.

Зависимость удельного газовыделения от времени, подчиняющаяся формулам (4.4.14), (4.4.20) и (4.4.22), показана на рис. 4.8. Бесконечно большая скорость газоотделения, получаемая из (4.4.20) при $t \rightarrow 0$, вполне объяснима в рамках рассматриваемой модели. Действительно, начальное распределение (4.4.15) обеспечивает на поверхности пластины концентрацию растворенного газа, равную c_0 при $t = 0$, а сразу же при $t > 0$ граничное условие (4.4.16) требует $c_s = 0$ при $x = 0$. Именно это создает конечный скачок концентрации $\Delta c = c_0$ за бесконечно малый интервал времени $\Delta t \rightarrow 0$.

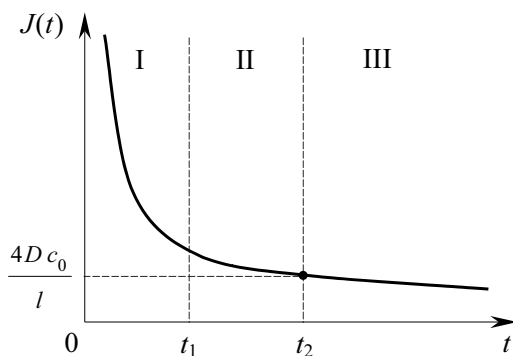


Рис. 4.8. Качественный ход удельного газовыделения для трех этапов обезгаживания пластины: I — начальный этап ($0 < t < t_1 = 0,0263 l^2/D$), описываемый формулой (4.4.20), II — промежуточный этап ($t_1 < t < t_2$), описываемый формулой (4.4.14), III — конечный этап ($t > t_2 = 0,045 l^2/D$), описываемый формулой (4.4.22)

Оценим численные значения интервала времени t_1 для металлических и силикатных материалов. Коэффициент диффузии D для газов в различных металлах лежит в широком диапазоне от 10^{-10} см²/с до 10^{-4} см²/с в зависимости от температуры и конкретной пары газ–металл. Для стекол и керамики аналогичные значения примерно в 10^5 – 10^6 раз меньше. Для пластины толщиной $l = 1$ мм расчет величины $t_1 = 0,0263 l^2/D$ дает для металлов $t_1 \approx 10$ мин при $D = 10^{-7}$ см²/с, а для силикатных материалов $t_1 \approx 20$ лет при $D = 10^{-13}$ см²/с. На основании этих численных оценок можно сделать следующие выводы.

В реальных условиях обезгаживания стекол и керамики газы удаляются из очень тонкого приповерхностного слоя, включая адсорбированные на поверхности атомы, в то время как растворенные в объеме газы практически не затрагиваются. Несмотря на это, для силикатных материалов даже короткое время их вакуумной обработки ($t_{\text{обр}} \ll t_1 \approx 20$ лет) обеспечивает, согласно (4.4.20), приемлемо малые газовые потоки с поверхности стеклянных и керамических оболочек из-за малого коэффициента диффузии растворенных газов: $J(t_{\text{обр}}) = c_0 \sqrt{D/\pi t_{\text{обр}}} \ll J_{\text{экс}}$.

В отличие от этого, для металлов вполне достижима конечная стадия обезгаживания за время вакуумной обработки $t_{\text{обр}} > t_2 = 0,045 l^2/D$. Как следует из формулы (4.4.22), времени t_2 соответствует газовый поток $J(t_2) = 4Dc_0/l$ (точка на кривой рис. 4.8), который может заметно превышать необходимую величину $J_{\text{экс}}$. Так как удельное газовыделение (4.4.22) с течением времени экспоненциально убывает, то требуемому соотношению,

$$J(t_{\text{обр}}) = J(t_2) \exp[-\pi^2 D(t_{\text{обр}} - t_2)/l^2] < J_{\text{экс}},$$

отвечают реально достижимые времена $t_{\text{обр}}$ термической обработки прибора.

Отсюда следует, что для металлических деталей, которые при обработке на откачном посту могут быть прогреты до высокой температуры (с помощью индукционного нагрева или электронной бомбардировки), нужно выбирать материалы с большим коэффициентом диффузии (такие как медь, никель или сплавы на их основе). Это обеспечивает быстрый экспоненциальный спад газового потока со временем. Наоборот, для внутренних деталей, которые трудно или недопустимо прогревать при высоких температурах, выгоднее использовать материалы с малым коэффициентом диффузии (типа нержавеющей стали и алюминированного железа). В этом случае требуемое условие $J(t_{\text{обр}}) < J_{\text{экс}}$, несмотря на медленный экспоненциальный спад газовыделения, может

быть обеспечено за счет малости величины $J(t_2) = 4Dc_0/l$. Такие детали, даже при недостаточном обезгаживании, будут иметь меньшее газовыделение, чем если бы они были изготовлены из материала с большим коэффициентом диффузии.

4.5. Принципы вакуумного обезгаживания материалов

Остаточная газовая среда в вакуумных приборах, формируемая как в процессе откачки, так и в результате газовыделения деталей и газопроницаемости оболочки в рабочих условиях, определяет срок службы прибора, в первую очередь термокатода. Для оксидного катода наиболее опасны, во-первых, кислород, хлор и сера, которые химически отравляют его путем связывания избыточного бария в составе окиси бария (что снижает ток эмиссии из-за повышения работы выхода электронов — см. п. 3.5) или приводят к спеканию оксидного покрытия из-за низкой температуры плавления образующихся соединений (например, хлоридов и сульфидов металлов) и, во-вторых, вещества типа углерода, которые окрашивают покрытие в темный цвет (что снижает ток эмиссии из-за понижения рабочей температуры катода — см. п. 3.5). Неблагоприятное состояние газовой среды сильно влияет также на работу газоразрядных приборов, например понижает мощность излучения газовых лазеров или даже приводит к выходу прибора из строя в результате газового пробоя. Поэтому обезгаживание внутренних деталей и вакуумной оболочки прибора имеет принципиально важное значение.

По характеру локализации в металлах и силикатных материалах различают следующие три группы газов.

I группа — поверхностный монослой адсорбированных газов, включая углеводороды и компоненты воздуха. Для стекол и полированных металлических поверхностей сорбционная емкость составляет $(1-5) \cdot 10^{-5}$ л·Торр/см² ¹⁾. Эта величина может в десятки раз возрастать при разрыхленной поверхности металла.

¹⁾ В вакуумной технике газосодержание и газовыделение часто измеряют в так называемых *PV*-единицах ($1 \text{ см}^3 \cdot \text{атм} = 0,76 \text{ л} \cdot \text{Торр}$), поскольку, согласно (1.12.1), число молей n при температуре T равняется $n = PV/RT$. Строго говоря, *PV*-единицы являются энергетическими единицами: так как $1 \text{ Дж} = 0,239 \text{ кал} = 7,5 \text{ л} \cdot \text{Торр}$, то $R = k_B N_A = 8,314 \text{ Дж}/(\text{моль} \cdot \text{К}) = 1,987 \text{ кал}/(\text{моль} \cdot \text{К}) = 62,335 \text{ л} \cdot \text{Торр}/(\text{моль} \cdot \text{К})$. При $T = 298 \text{ К}$ имеем число молей $n = 5,382 \cdot 10^{-5} PV [\text{л} \cdot \text{Торр}]$, так что при $PV = 1 \cdot 10^{-5} \text{ л} \cdot \text{Торр}/\text{см}^2$ получаем $n/\text{см}^2 = 5,382 \cdot 10^{-10} \text{ моль}/\text{см}^2 = 3,348 \cdot 10^{14} \text{ частиц}/\text{см}^2$. Полезно напомнить, что $1 \text{ Па} (\text{Н}/\text{м}^2) = 7,5 \cdot 10^{-3} \text{ Торр} (\text{мм. рт. ст.})$, см. с. 763.

Время десорбции таких газов равно 2–3 минуты для металлов при 300–500°C (с преимущественным выделением N_2 , O_2 , H_2O и углеводородов) и для стекол при 150–250°C (с преимущественным выделением H_2O и CO_2). Физические закономерности адсорбции и механизмы удаления поверхностных загрязнений будут рассмотрены в следующей главе.

II группа — газы, содержащиеся в тонком приповерхностном слое в виде окислов и других химических соединений. Вещества, растворенные в металлах (водород, углерод и др.), при нагреве диффундируют к поверхности, где вступают в химическое взаимодействие с поверхностными окислами и восстанавливают их с выделением H_2O , CO_2 , CO , CH_4 , H_2 и др. Общее количество таких газов, выделившихся за несколько часов прогрева при температурах 700–900°C, достигает $(5–100) \cdot 10^{-2}$ л · Торр/см².

В приповерхностных слоях стекол газосодержание составляет $(5–10) \cdot 10^{-4}$ л · Торр/см², что эквивалентно нескольким десяткам мономолекулярных слоев. Здесь преимущественно содержатся вода и небольшое количество углекислого газа, что вызвано химической неустойчивостью структурных связей в стеклах. Как известно (см. ниже п. 4.6), структура стекол включает, кроме основных стеклообразующих связей Si–O–Si, также связи типа Si–O–X (где X — ион примеси-модификатора, в частности щелочного или щелочно-земельного металла). Последние под воздействием воды разрушаются, давая новые связи Si–O–H (типа силикагеля) с выделением растворимых в воде соединений (типа NaOH). Углекислый газ CO_2 растворяется, например, в натриевых стеклах с образованием соды Na_2CO_3 , что способствует разрушению приповерхностного слоя. Удаление этих газов достигается прогревом стекла при температуре 300–400°C в течение нескольких часов путем восстановления связей Si–O–Si и Si–O–X с выделением главным образом паров воды и в меньшей степени (на два порядка ниже) углекислого газа.

III группа — газы, растворенные внутри материала, а именно: а) в металлах — преимущественно водород и в меньшей степени азот (последний главным образом в Mo, W, Fe и сталях), б) в стеклах — преимущественно пары воды и в меньшей степени другие газы, определяемые составом исходных для варки стекла материалов.

Количество таких газов в металлах составляет небольшую долю от газов II группы, в то время как в стеклах их содержание в сотни раз больше, чем в приповерхностных слоях. Однако это не создает особых затруднений в обеспечении вакуума в приборе,

если стекло (как и керамика) не подвергается электронной бомбардировке, способствующей разложению окислов. Действительно, ввиду медленной диффузии газов в стеклах полное их обезгаживание практически неосуществимо (см. п. 4.4) и обезгаживается лишь тонкий приповерхностный слой толщиной в несколько десятков микрон. При удалении из металлов газов III группы приемлемое газовыделение достигается за несколько часов при температуре прогрева 800–900°C.

Таким образом, предварительная обработка деталей вакуумных приборов включает, во-первых, *химические методы очистки* для удаления поверхностных загрязнений и газов I (и отчасти II) группы и, во-вторых, *термические методы очистки* с целью пиролиза остатков промывочных жидкостей и химических растворителей, восстановления или разложения окислов на поверхности металлов и удаления растворенных газов II и III группы. Рафинирующий термический отжиг как завершающий этап очистки металлических деталей является чрезвычайно важным, так как по эффективности превосходит предыдущие этапы, сглаживая технологические шероховатости различных химических обработок.

Основа обезгаживания состоит в достижении равновесия растворенного в металле газа с внешней средой. Согласно закону Сиверта (2.6.12), концентрация i -го двухатомного газа в металлах пропорциональна $p_i^{1/2}$, т.е. уменьшается с понижением парциального давления газа. Поэтому обезгаживание может вестись не только в вакууме, но и в любой газовой среде, не содержащей газ i -го сорта. В частности, для отжига металлов могут быть использованы чистые инертные газы (He, Ne, Ar), а для никеля и меди — также и азот N_2 , так как эти газы не растворяются в металлах.

Наиболее распространен рафинирующий отжиг металлических деталей в чистом водороде, который эквивалентен отжигу в вакууме для очистки от всех газов, кроме водорода. Водородный отжиг в открытой проточной системе (при $p_{H_2} = 1$ атм) дешевле и технически проще по сравнению с вакуумным отжигом и обеспечивает за приемлемое время (не выше 30–90 мин) эффективное удаление растворенных газов и восстановление поверхностных окислов. Однако при этом происходит «наводороживание» материала, что в ряде случаев (например, при отжиге Ti, Ta, Zr или графитовых и карбонизированных деталей) нежелательно, поскольку либо повышается его хрупкость в результате образования гидридов, либо наблюдается образование летучих

углеводородов. В этих случаях приемлемым остается только вакуумный отжиг.

Однако достигнуть желаемого эффекта лишь за счет понижения соответствующего парциального давления в окружающей среде не удастся, так как основным процессом при обезгаживании является диффузия, скорость которой по закону Аррениуса (1.18.16) возрастает с повышением температуры отжига. Но увеличение температуры также влияет и на растворимость газов в металлах (см. п. 2.6): для конструкционных металлов (типа Fe, Ni, Cu и др.) с ростом температуры растворимость газов возрастает, а для активных газопоглотителей (типа Ti, Zr, Ta и др.) — падает. Это различие принципиально меняет требование к скорости охлаждения деталей после водородного отжига (чего нет при отжиге в вакууме). Действительно, длительная выдержка при высокой температуре в атмосфере водорода при $p_{H_2} = 1$ атм насыщает растворенным водородом конструкционные металлы до больших концентраций, а газопоглотители — до меньших концентраций по сравнению с комнатной температурой.

Следовательно, чтобы обеспечить меньшее количество остаточного водорода в холодном состоянии, в первом случае (для конструкционных металлов) требуется малая скорость остывания с целью приближения к равновесному режиму охлаждения, а во втором случае — быстрое охлаждение для «замораживания» в объеме газопоглотителя меньшего количества водорода, равно го растворенному ранее при более высокой температуре отжига.

4.6. Роль диффузии в газопроницаемости вакуумных оболочек

Диффузионное проникновение газов через оболочку вакуумного прибора происходит в несколько последовательных этапов, таких как адсорбция газов из атмосферы на наружной поверхности оболочки, молекулярное или атомарное растворение в поверхностных слоях, собственно диффузия через толщину оболочки и газовыделение внутрь прибора.

В стационарных условиях левая часть диффузионного уравнения (4.4.1) обращается в нуль, тогда $\partial c / \partial x = \text{const}$. Это обеспечивает линейное распределение концентрации растворенного газа по толщине оболочки, тогда плотность диффузионного потока записывается в виде

$$J = D \frac{\partial c}{\partial x} = D \frac{c_{s1} - c_{s2}}{l}. \quad (4.6.1)$$

Здесь c_{s1} и c_{s2} — поверхностные концентрации растворенного газа, соответственно, на наружной и внутренней сторонах оболочки. Эти концентрации связаны с парциальными давлениями данного газа в атмосфере (p_1) и в приборе (p_2) соотношением (4.4.3), общим для законов Генри и Сивертса:

$$c_{s1,2} = s(T) p_{1,2}^m. \quad (4.6.2)$$

При растворении двухатомных газов показатель степени m равен 1 для силикатных оболочек (молекулярная растворимость по закону Генри) и $1/2$ — для металлических оболочек (атомарная растворимость по закону Сивертса). Подстановка (4.6.2) в формулу (4.6.1) дает плотность потока газонатекания в прибор через вакуумную оболочку:

$$J_{\text{газ}} = K_{\text{газ}} \frac{p_1^m - p_2^m}{l} \approx K_{\text{газ}} \frac{p_1^m}{l}, \quad (4.6.3)$$

где примерное равенство соответствует условию $p_1 \gg p_2$. Здесь введен *коэффициент газопроницаемости* $K_{\text{газ}} = sD$ в виде произведения коэффициентов растворимости и диффузии, который равен плотности газового потока при $p_1 = 1$ атм и единичной толщине стенки (обычно 1 мм). Как правило, энергия активации процесса диффузии превосходит теплоту растворения, поэтому с увеличением температуры функция $K_{\text{газ}}(T)$, подобно функции $D(T)$ в законе Аррениуса (1.18.16), возрастает. Поскольку плотность газового потока измеряют в PV -единицах, $\text{л} \cdot \text{Торр}/(\text{см}^2 \cdot \text{с})$ (см. сноску на с. 454), то, согласно (4.6.3), для коэффициента газопроницаемости $K_{\text{газ}}$ используют комбинированную единицу измерения — $(\text{л} \cdot \text{Торр}/(\text{см}^2 \cdot \text{с}))(\text{мм}/(\text{атм})^m)$.

Рассмотрим механизмы газопроницаемости металлических и силикатных оболочек, основываясь на составе атмосферного воздуха (в парциальных давлениях): N_2 — 595 Торр, O_2 — 159 Торр, Ar — 7 Торр, CO_2 — 0,25 Торр, Ne — $1,4 \cdot 10^{-2}$ Торр, He — $4 \cdot 10^{-3}$ Торр, H_2 — $4 \cdot 10^{-4}$ Торр, H_2O — до 12 Торр. Если исходить из этих данных, то наибольшую опасность газонатекания создают азот, кислород и аргон, как имеющие наибольшие парциальные давления. Однако это не соответствует действительности по следующим причинам.

Металлические оболочки проницаемы только для атомов (но не молекул) газа, так как процессу диффузии предшествует растворение газа в атомарной форме (см. п. 2.6). Возможность диффузионного перемещения растворенных атомов зависит от характера их химической связи с основными атомами и от того, насколько прочны такие связи. Азот и кислород образуют

прочные нитриды и оксиды на поверхности металлов, поэтому они слабо диффундируют сквозь металлические оболочки. Сложные газы типа H_2O , CO_2 и др., если и диффундируют, то в форме входящих в них атомов с последующим образованием молекул на поверхности металла. Благородные газы типа Ar , Ne и He практически не проникают через металлы, поскольку они не растворяются в них из-за невозможности химической адсорбции (см. ниже п. 5.9).

Единственно возможным для металлов остается проникновение атомарного водорода. Несмотря на то, что атомы водорода образуют с большинством металлов гидриды и твердые растворы, энергия их связи с решеткой невелика, что и определяет возможность диффузии. Более того, атомарный водород, растворенный в металле, легко теряет свой единственный электрон и превращается в протон, уменьшаясь в размере примерно в 10^5 раз. Именно это обстоятельство и объясняет наблюдаемую на практике высокую проникаемость металлов по водороду. Следовательно, предельное атмосферное натекание через металлические оболочки ограничивается величиной $p_{\text{H}_2} = 4 \cdot 10^{-4}$ Торр.

Силикатные оболочки изготавливаются из диэлектрических материалов (типа стекол и керамики), имеющих в своем составе двуокись кремния SiO_2 . Основную структурную группу в силикатах (являющихся солями кремниевой кислоты $\text{H}_2\text{SiO}_3 \equiv \text{SiO}_2 \cdot \text{H}_2\text{O}$) составляет комплексный ион SiO_4^{4-} в форме кислородного тетраэдра, окружающего атом Si . Каждый из четырех атомов кислорода в тетраэдре одной валентностью связывается с центральным атомом Si , а остающейся валентностью — с другим атомом кремния, расположенным в центре соседнего тетраэдра. Следовательно, между атомами кремния существует кислородный мостик, образующий разветвленную силикатную сетку с хаотически меняющимся углом мостиковой связи $\text{Si}-\text{O}-\text{Si}$. В результате этого возникает аморфное состояние структуры, сохраняющее лишь ближний порядок в расположении кремний-кислородных тетраэдров.

Примером аморфной силикатной структуры является чистое кварцевое стекло (плавленый кварц), представляющее собой затвердевшую переохлажденную жидкость. Она термодинамически неустойчива при высоких температурах (выше 1700°C), когда появляются включения кристаллических модификаций кварца («расстекловывание»).

Изменение свойств кварцевого стекла достигается путем введения дополнительных окислов — *стеклообразующих* (B_2O_3

и P_2O_5) и *структурообразующих* (PbO и Al_2O_3), а также примесей-модификаторов типа Na, K, Ca, Ba, Mg и др. Атомы B, P, Pb и Al, входящие в состав стеклообразующих и структурообразующих окислов, замещают атомы кремния в центре тетраэдров. Это формирует стекла со специфическими свойствами, которые называют по названию окисла, а именно: боросиликатное стекло $mB_2O_3 \cdot nSiO_2$ (БСС), фосфорно-силикатное стекло $mP_2O_5 \cdot nSiO_2$ (ФСС), свинцово-силикатное стекло $mPbO \times \times nSiO_2$ (ССС) и алюмосиликатное стекло $mAl_2O_3 \cdot nSiO_2$ (АСС) (последнее название применяют чаще к алюмосиликатной керамике). В отличие от этого, атомы примесей-модификаторов не участвуют в образовании структуры стекла, а располагаются в пустотах силикатной сетки, будучи окруженными ионами кислорода, что сильно изменяет термические, механические и электрические свойства стекол.

Приведенные краткие сведения о структуре стекла позволяют понять физический механизм проникновения атмосферных газов сквозь стеклянные оболочки. Этот механизм отличается от описанного выше механизма для металлов, где преобладает газонатекание по водороду. Стекла как структурно неупорядоченные материалы с пустотами силикатной сетки ведут себя по отношению к газам как «молекулярное сито»: чем больше размер газовой молекулы, тем вероятнее ее застревание в пустотах сетки и тем меньше таких молекул должно проходить сквозь «сито». В этом случае заполнение пустот атомами модификаторов должно снижать газопроницаемость стекол в сравнении с чистым кварцевым стеклом, что действительно подтверждается на опыте.

Молекулы атмосферных газов имеют следующие эффективные диаметры: He — $2,0 \text{ \AA}$, H_2 — $2,5 \text{ \AA}$, Ag — $3,1 \text{ \AA}$, O_2 — $3,2 \text{ \AA}$, N_2 — $3,4 \text{ \AA}$, H_2O — $4,6 \text{ \AA}$. Поэтому гелий как газ с наименьшим диаметром атома должен наиболее интенсивно проникать через стекла. Молекула водорода, имея также небольшой диаметр, тем не менее не проходит сквозь стекла, так как ее взаимодействие с кислородом силикатной сетки приводит к образованию большой по размерам молекулы воды. Следовательно, предельное атмосферное натекание через стеклянные оболочки ограничивается величиной $p_{He} = 4 \cdot 10^{-3}$ Торр.

Керамика является превосходным материалом для вакуумных оболочек, так как сочетает в себе положительные качества стекла и металла. Вакуумные керамические материалы, такие как алюмосиликатная керамика (АСК) и магнезиально-силикатная керамика (МСК), получают спеканием SiO_2 с тугоплавкими

окислами Al_2O_3 и MgO . Керамика является многофазной системой, в которой силикатная компонента, образующая стеклофазу, цементирует кристаллофазу, созданную окислом Al_2O_3 или MgO . Керамика на основе Al_2O_3 с предельно низким содержанием стеклофазы (до 1–10%) носит название алюмооксидной керамики (АОК).

По отношению к гелию керамика сходна с металлами, так как химически инертный гелий не имеет устойчивого иона и не может проникать в решетку кристаллофазы, которая в процентном содержании преобладает над стеклофазой. По отношению к водороду керамика ведет себя подобно стеклу, поскольку водород, имея сильное химическое сродство с окислами керамики, химически связывается в объеме как стеклофазы, так и кристаллофазы, и не может проникать сквозь керамическую оболочку.

Приведем ориентировочные значения коэффициента проницаемости по водороду K_{H_2} (для металлов и керамики) и по гелию K_{He} (для стекол и керамики), измеренные в комбинированных единицах — $(\text{л} \cdot \text{Торр}/(\text{см}^2 \cdot \text{с}))(\text{мм}/(\text{атм})^m)$:

- **Металлы** (298 К): $K_{\text{H}_2} \approx 10^{-6}$ (Pd), 10^{-9} (Fe), 10^{-11} (Ni), 10^{-14} (Cu), 10^{-20} (Al);
- **Стекла** (298 К): $K_{\text{He}} \approx 10^{-8}$ (кварцевое стекло), $10^{-10} - 10^{-14}$ (щелочные стекла), 10^{-12} (БСС), 10^{-13} (ФСС), 10^{-15} (ССС);
- **Керамика** (973 К): $K_{\text{H}_2} \approx 10^{-12}$ (АОК и МСК), $K_{\text{He}} \approx 10^{-11}$ (АОК), 10^{-13} (МСК).

Приведенные значения K_{H_2} и K_{He} соответствуют температуре, указанной в скобках, и экспоненциально возрастают с ростом температуры.

Как видно из этих данных, наибольшей проницаемостью по водороду обладает палладий, из которого по этой причине изготавливают мембраны для диффузионной очистки водорода. Из металлов, используемых в качестве элементов оболочек, малым значением K_{H_2} обладает медь, которая по своим вакуумным свойствам не уступает стеклам, что делает ее пригодной для корпусов мощных СВЧ-приборов. Алюминий, из-за предельно низкой газопроницаемости, применяется в качестве наружного покрытия вакуумных колб с целью предотвращения натекания атмосферных газов.

Данные по керамике свидетельствуют в пользу ее применения вместо стекол для создания металлокерамических оболочек

приборов. Газонатекание таких приборов в целом по воздуху превышает примерно на порядок приведенные значения, что вызвано капиллярным проникновением атмосферных газов в силу как естественной пористости керамики, так и пористости металлов, возникшей в результате их коррозионного и усталостного растрескивания.

4.7. Диффузионные задачи на введение вещества в твердое тело

Характерным практическим примером задачи такого типа является диффузионное легирование полупроводников с целью изменения электрофизических свойств при создании электронно-дырочных переходов и изоляции элементов интегральных схем.

Для анализа процесса легирования будем использовать диффузионное уравнение типа (4.4.1), а именно

$$\frac{\partial c(x, t)}{\partial t} = D \frac{\partial^2 c(x, t)}{\partial x^2}, \quad (4.7.1)$$

применяя его к модели *полугораниченного тела*, занимающего полупространство при $x > 0$.

Полупроводник считается изначально нелегированным, т. е. всегда $c(x > 0, t) = 0$ при $t \leq 0$, где момент времени $t = 0$ соответствует началу процесса легирования. Внешняя среда, содержащая легирующую примесь, располагается при $x < 0$.

Рассмотрим постановку граничной задачи и приведем ее окончательное решение (без вывода) для трех наиболее распространенных моделей диффузионного легирования.

1. Диффузия из постоянного источника. В этой модели предполагается, что среда, расположенная при $x < 0$, содержит диффузант (легирующую примесь), убыль которого в результате диффузии в полупространство $x > 0$ непрерывно восполняется за счет внешнего источника. Это обеспечивает на границе полупроводника при $x = 0$ неизменную во времени поверхностную концентрацию $c_s(t) = \text{const}$ (*связывающая граница* с граничным условием (4.3.6)). Практическая реализация модели постоянного источника описана в п. 4.8.

На этом основании начальное условие при $t = 0$ и граничные условия при $x = 0$ и $x \rightarrow \infty$ записываем в следующем виде:

$$c(x, t)|_{t=0} = 0 \quad \text{при } 0 < x < \infty, \quad (4.7.2)$$

$$c(x, t)|_{x=0} = c_s \quad \text{и} \quad c(x, t)|_{x \rightarrow \infty} \rightarrow 0 \quad \text{при } t > 0. \quad (4.7.3)$$

Искомое решение граничной задачи, поставленной уравнениями (4.7.1)–(4.7.3), имеет вид (ср. формулу (4.4.17))

$$c(x, t) = c_s \operatorname{erfc} \frac{x}{2\sqrt{Dt}}. \quad (4.7.4)$$

Здесь $\operatorname{erfc} z = 1 - \operatorname{erf} z$ — дополнительная функция ошибок (от англ. *error function complement*), график которой приведен на рис. 4.6; при этом erf -функция определена в форме (4.4.18). Качественный вид решения (4.7.4) показан на рис. 4.9.

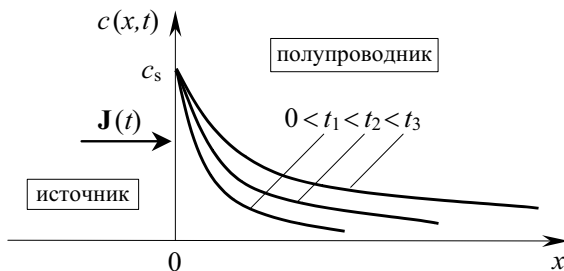


Рис. 4.9. Распределение легирующей примеси в полупроводнике в разные моменты времени при диффузии из постоянного источника с заданной поверхностной концентрацией $c_s = \text{const}$

Рассматриваемая диффузионная задача противоположна «вакуумной» задаче обезгаживания, поставленной уравнениями (4.4.1), (4.4.15) и (4.4.16), которая в «полупроводниковой» формулировке соответствует случаю испарения примеси из предварительно насыщенного полупроводника. В данном рассматриваемом случае полученное «полупроводниковое» решение (4.7.4) близко по структуре к «вакуумному» решению (4.4.17). В частности, плотность потока диффузанта, вводимого в полупроводник в момент времени t , вычисляется аналогично тому, как это было сделано при выводе формулы (4.4.20):

$$J(t) = -D \left. \frac{\partial c(x, t)}{\partial x} \right|_{x=0} = c_s \sqrt{\frac{D}{\pi t}}. \quad (4.7.5)$$

Общее количество примеси $Q(t) = \int_0^\infty c(x, t) dx$, введенной через единицу поверхности полупроводника за время t , легко найти интегрированием плотности диффузионного потока (4.7.5):

$$Q(t) = \int_0^t J(t') dt' = 2c_s \sqrt{\frac{Dt}{\pi}} \approx 1,13 c_s \sqrt{Dt}. \quad (4.7.6)$$

2. Диффузия из полуограниченного источника. В этой модели предполагается, что среда, расположенная при $x < 0$, изначально содержит однородно распределенную легирующую примесь с концентрацией c_0 , убыль которой при диффузии в полупространство $x > 0$ не восполняется внешним источником. Практическая реализация модели полуограниченного источника описана в п. 4.8.

На этом основании начальное условие при $t = 0$ и граничные условия на сингулярных краях $x \rightarrow \mp \infty$ записываются в следующем виде:

$$c(x, t)|_{t=0} = \begin{cases} c_0 & \text{при } -\infty < x < 0, \\ 0 & \text{при } 0 \leq x < \infty, \end{cases} \quad (4.7.7)$$

$$c(x, t)|_{x \rightarrow -\infty} = c_0 \quad \text{и} \quad c(x, t)|_{x \rightarrow \infty} \rightarrow 0 \quad \text{при } t > 0. \quad (4.7.8)$$

При $x = 0$ обеспечивается непрерывность диффузионного потока $D(\partial c / \partial x)$ в силу того, что на границе раздела областей нет стока или истока частиц.

Искомое решение граничной задачи, поставленной уравнениями (4.7.1), (4.7.7) и (4.7.8), имеет вид (ср. формулу (4.7.4))

$$c(x, t) = \frac{c_0}{2} \operatorname{erfc} \frac{x}{2\sqrt{Dt}}. \quad (4.7.9)$$

На рис. 4.10 качественно показана зависимость (4.7.9) для трех моментов времени. В данном случае распределение примеси в полупроводнике (при $x > 0$) полностью совпадает с аналогичными распределениями, полученными при диффузии из постоянного источника (рис. 4.9), только концентрация в плоскости $x = 0$ равна $c_0/2$ вместо c_s . Для каждой кривой площади выше ее (до значения c_0) при $x < 0$ и ниже ее при $x > 0$ равны.

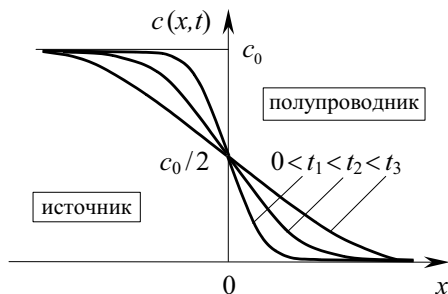


Рис. 4.10. Распределение легирующей примеси в источнике и полупроводнике в разные моменты времени при диффузии из полуограниченного источника в гомогенной системе

Представленные на рис. 4.10 кривые в большинстве случаев далеки от реального распределения примесей, так как они получены в предположении, что коэффициент диффузии примеси одинаков для материала источника и для полупроводника (*гомогенная система*). Чаще всего источник легирующей примеси и полупроводник представляют собой разные вещества, то есть имеем дело с *гетерогенной системой*.

В этом случае источник и полупроводник должны рассматриваться как две отдельные среды с искомыми распределениями диффузанта: $c_1(x, t)$ — в источнике (при $-\infty < x < 0$) и $c_2(x, t)$ — в полупроводнике (при $0 < x < \infty$). Для каждой из концентраций записывается свое диффузионное уравнение типа (4.7.1), а начальные и граничные условия (4.7.7) и (4.7.8), переписанные с введением индексов 1 и 2, дополняются следующими граничными условиями при $x = 0$, обеспечивающими:

- непрерывность плотности потока диффузанта из-за отсутствия накопления примеси на границе,

$$D_1 \frac{\partial c_1(x, t)}{\partial x} \Big|_{x=0} = D_2 \frac{\partial c_2(x, t)}{\partial x} \Big|_{x=0} \quad \text{при } t > 0; \quad (4.7.10)$$

- перераспределение примеси между средами из-за различия ее растворимости в материале источника и в полупроводнике, характеризуемого *коэффициентом распределения* (сегрегации) $K(T) = c_{s1}/c_{s2}$, зависящим только от температуры,

$$c_1(x, t) \Big|_{x=0} = K c_2(x, t) \Big|_{x=0} \quad \text{при } t > 0. \quad (4.7.11)$$

Здесь коэффициент распределения примеси между двумя твердыми фазами введен по аналогии с аналогичным коэффициентом (2.7.5) для жидкости и твердого тела, полученным ранее в условиях равновесия. В рассматриваемой неравновесной ситуации считаем, что между приповерхностными слоями двух смежных фаз равновесие наступает достаточно быстро. Это позволяет приравнять химические потенциалы легирующего компонента, взятые по разные стороны границы раздела сред:

$$\mu_1^\circ(T) + RT \ln c_{s1} = \mu_2^\circ(T) + RT \ln c_{s2}. \quad (4.7.12)$$

Поскольку для разных сред $\mu_1^\circ(T) \neq \mu_2^\circ(T)$, то из выражения (4.7.12) получаем $c_{s1}/c_{s2} = \exp[(\mu_2^\circ(T) - \mu_1^\circ(T))/RT] \equiv K(T)$.

Решение данной граничной задачи имеет следующий вид:

$$c_1(x, t) = \frac{c_0}{K+R} \left(K - R \operatorname{erf} \frac{x}{2\sqrt{D_1 t}} \right) \quad \text{при } -\infty < x \leq 0, \quad (4.7.13)$$

$$c_2(x, t) = \frac{c_0}{K+R} \operatorname{erfc} \frac{x}{2\sqrt{D_2 t}} \quad \text{при } 0 \leq x < \infty, \quad (4.7.14)$$

где введено обозначение $R = \sqrt{D_2/D_1}$.

Характер распределения диффузанта, описываемый формулами (4.7.13) и (4.7.14), качественно можно понять, если воспользоваться кривыми, показанными на рис. 4.10, и дополнительными граничными условиями (4.7.10) и (4.7.11). Согласно (4.7.11), отличный от единицы коэффициент распределения обеспечивает в смежных фазах различные поверхностные концентрации,

$$c_{s1} = \frac{c_0 K}{K+R} \quad \text{и} \quad c_{s2} = \frac{c_0}{K+R}, \quad (4.7.15)$$

как показано на рис. 4.11, где $K > 1$.

При разных коэффициентах диффузии ($D_1 \neq D_2$) требование (4.7.10), выражающее равенство диффузионных потоков на границе раздела сред, обеспечивает различный наклон кривых по разные стороны от границы, что видно из рис. 4.11, где $D_1 > D_2$. Поскольку на границе отсутствуют истоки или стоки частиц и нет их накопления, в любой момент времени $t > 0$ количество ушедших из источника и поступивших в полупроводник частиц одинаково. Следовательно, на рис. 4.11 площади выше кривой $c_1(x, t)$, до горизонтальной линии c_0 , и под кривой $c_2(x, t)$ должны быть равными, как и на рис. 4.10.

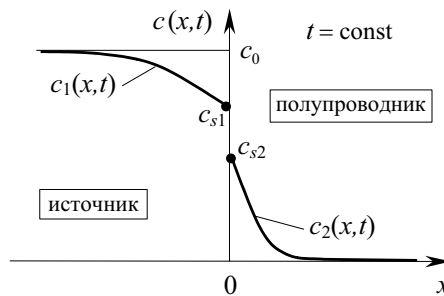


Рис. 4.11. Распределение легирующей примеси в источнике и полупроводнике при диффузии из полуограниченного источника в гетерогенной структуре

3. Диффузия из поверхностного источника с отражающей границей. В этой модели источником служит поверхностный слой толщиной d ($0 < x < d$), в котором изначально присутствует однородно распределенная примесь с концентрацией c_0 . Левая граница этого слоя предполагается отражающей с граничным условием (4.3.5) при $x = 0$, т. е. она не пропускает

частицы в область $x < 0$. Тогда в процессе диффузии легирующая примесь из слоя толщиной d может быть «разогнана» только в область $x > d$. На этом основании начальное и граничные условия записываются в следующем виде:

$$c(x, t)|_{t=0} = \begin{cases} c_0 & \text{при } 0 < x < d, \\ 0 & \text{при } d \leq x < \infty, \end{cases} \quad (4.7.16)$$

$$J(t)|_{x=0} = -D \frac{\partial c(x, t)}{\partial x} \bigg|_{x=0} = 0 \quad \text{и} \quad c(x, t)|_{x \rightarrow \infty} \rightarrow 0 \quad \text{при } t > 0. \quad (4.7.17)$$

Искомое решение граничной задачи, поставленной уравнениями (4.7.1), (4.7.16) и (4.7.17), имеет вид

$$c(x, t) = \frac{c_0}{2} \left(\operatorname{erf} \frac{x+d}{2\sqrt{Dt}} - \operatorname{erf} \frac{x-d}{2\sqrt{Dt}} \right). \quad (4.7.18)$$

При диффузии из слоя конечной толщины количество легирующей примеси ограничено исходной величиной $Q = c_0 d$. В процессе легирования происходит перераспределение примеси между слоями, называемое «разгонкой примеси». Естественно, что при $t \rightarrow \infty$ концентрация примеси, разогнанной на очень большое расстояние, стремится к нулю.

В пределе бесконечно тонкого слоя, когда сохраняется количество примеси $Q = c_0 d$ в слое при $d \rightarrow 0$, выражение (4.7.18) существенно упрощается. При этом начальное распределение примеси (4.7.16) представляется в виде единичной импульсной функции $\delta(x)$, для которой $\int_{-\infty}^{\infty} \delta(x) dx = 1$. Отсюда

$$c(x, t)|_{t=0} = Q\delta(x), \quad \text{так что} \quad \int_{-\infty}^{\infty} c(x, 0) dx = Q. \quad (4.7.19)$$

Применяем к выражению (4.7.18) предельный переход $d \rightarrow 0$ при условии $Q \equiv c_0 d = \text{const}$, с учетом того факта, что для erf-функции (4.4.18) разложение в ряд Тейлора в линейном приближении дает

$$\operatorname{erf}(z \pm u) \approx \operatorname{erf} z \pm \frac{2u}{\sqrt{\pi}} e^{-z^2} \quad \text{при } u \ll z.$$

Отсюда получаем распределение легирующей примеси в форме функции Гаусса:

$$c(x, t) = \frac{Q}{\sqrt{\pi Dt}} \exp\left(-\frac{x^2}{4Dt}\right). \quad (4.7.20)$$

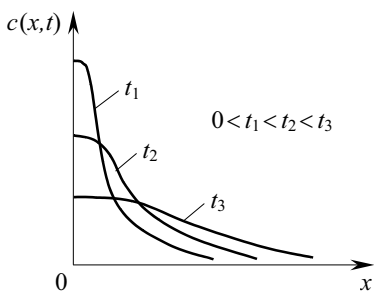


Рис. 4.12. Распределение легирующей примеси в полупроводнике для разных моментов времени при диффузионной разгонке примеси из бесконечно тонкого поверхностного источника с отражающей границей

Качественный вид зависимости (4.7.20) для разных моментов времени изображен на рис. 4.12. При этом площадь под каждой кривой одинакова и равна $Q = \text{const}$, а касательная, проведенная к кривой в точке $x = 0$, горизонтальна, поскольку граничное условие (4.7.17) для отражающей границы обеспечивает здесь $\partial c / \partial x = 0$.

В заключение полезно сопоставить две модели диффузии — из постоянного источника и из бесконечно тонкого слоя с отражающей границей, в которых проявляется разное поведение поверхностной концентрации c_s и количества введенной примеси Q (при этом $z \equiv x / 2\sqrt{Dt}$):

erfc-распределение (4.7.4) гауссово распределение (4.7.20)

$$c(x, t) = c_s \operatorname{erfc} z,$$

$$c(x, t) = c_s(t) \exp(-z^2),$$

$$c_s = \text{const}, \quad Q(t) = c_s \sqrt{\frac{4Dt}{\pi}}; \quad Q = \text{const}, \quad c_s(t) = \frac{Q}{\sqrt{\pi Dt}}.$$

Графическое сравнение этих распределений дано ниже на рис. 4.13 *a*, где на верхнем графике качественно показано распределение акцепторной примеси $N_A(x)$ при легировании из постоянного источника (сплошная кривая) и из бесконечно тонкого слоя с отражающей границей (штриховая кривая) для случая одинаковой поверхностной концентрации.

Применимость рассмотренных выше моделей диффузии к различным практическим способам легирования полупроводников с целью нахождения глубины залегания p – n -перехода обсуждаются в следующем параграфе для четырех частных случаев.

4.8. Принципы диффузионного легирования полупроводников

Рассмотрим некоторые вопросы практической реализации моделей диффузионного легирования полупроводников, изложенных в предыдущем параграфе.

По типу источника и способу введения примесей в полупроводник все современные методы легирования можно разделить на две большие группы.

1. Легирование из парогазовой фазы. В этом случае процесс диффузионного легирования обычно проводят одним из следующих способов:

- в запаянной кварцевой ампуле с необходимыми компонентами;
- в открытой проточной системе с потоком газа-носителя (H_2 , N_2 , Ar , He);
- в вакуумной камере с откачкой или с наполнением ее инертным газом (N_2 , Ar , He).

Необходимая для легирования примесь поставляется в газовую фазу либо путем подачи соответствующего газа, либо за счет испарения твердого или жидкого источника. Подбором технологических параметров (давления газа или температуры испаряемого источника) поддерживают постоянным нужное парциальное давление p легирующего газового компонента. Неизменная температура T в зоне легирования обеспечивает постоянство коэффициента растворения примеси $s(T)$. Согласно соотношению (4.4.3), это приводит к постоянной поверхностной концентрации c_s растворенной в полупроводнике примеси, а именно: $c_s(T) = s(T)p^m = \text{const}$ при $p = \text{const}$. Следовательно, при легировании из парогазовой фазы реализуется *модель постоянного источника* (см. формулу (4.7.4) и рис. 4.9).

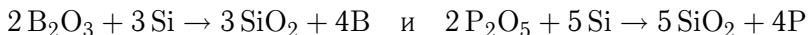
2. Легирование из поверхностных источников. В полупроводниковой технологии используют следующие способы легирования из поверхностных источников.

Легирование из напыленного металлического слоя проводят при температуре, превышающей температуру плавления предварительно напыленного металла-примеси. Расплав металла растворяет в себе атомы полупроводника, создавая жидкий раствор-расплав состава $c^{\text{ж}}$. Наряду с этим, в приповерхностном слое полупроводника в результате вхождения в его решетку атомов металла (донора или акцептора) возникает твердый раствор состава $c^{\text{т}}$. В соответствии с диаграммой плавкости системы металл–полупроводник, при данной температуре устанавливается вполне определенное значение коэффициента распределения примеси $K(T) = c^{\text{т}}/c^{\text{ж}}$. Условие $T = \text{const}$ обеспечивает постоянную поверхностную концентрацию c_s растворенного в полупроводнике металла-примеси, а именно: $c_s \equiv c^{\text{т}} = K(T)c^{\text{ж}} = \text{const}$. Следовательно, при легировании из напыленного металлического

слоя реализуется *модель постоянного источника* (см. формулу (4.7.4) и рис. 4.9).

Легирование из эпитаксиального слоя происходит за счет управляемой или неконтролируемой диффузии примеси, присутствующей в эпитаксиальном слое, в смежный с ним кристалл. Как правило, толщина слоя превышает диффузионную длину примеси, $L_{\text{диф}} = \sqrt{Dt}$, так что в данном случае работает *модель полуограниченного источника*. В многослойных эпитаксиальных структурах обычно подложка содержит наибольшее количество примесей, которые могут нежелательным образом в обратном порядке легировать эпитаксиальные слои. Закономерность легирования для гомоэпитаксиальных структур (в которых материал эпитаксиальных слоев и подложки один и тот же) подчиняется соотношению (4.7.9), показанному на рис. 4.10, в отличие от гетероэпитаксиальных структур, для которых применимы более сложные зависимости (4.7.13) и (4.7.14), качественно изображенные на рис. 4.11.

Легирование из поверхностных окислов в кремниевой технологии выполняют, используя в качестве источников бора В (акцептор) и фосфора Р (донор), соответственно, боросиликатное стекло $m\text{B}_2\text{O}_3 \cdot n\text{SiO}_2$ (БСС) и фосфорно-силикатное стекло $m\text{P}_2\text{O}_5 \cdot n\text{SiO}_2$ (ФСС). Окислы B_2O_3 и P_2O_5 взаимодействуют с кремнием по реакциям



с образованием свободных атомов бора и фосфора, легирующих приповерхностный слой кремния. Поскольку толщина слоев БСС и ФСС велика по сравнению с диффузионной длиной, $L_{\text{диф}} = \sqrt{Dt}$, для бора и фосфора в кремнии, то к этому случаю применима *модель полуограниченного источника* в гетероструктуре, описываемая зависимостями (4.7.13) и (4.7.14), качественно показанными на рис. 4.11.

Легирование из рекристаллизованного и ионно-имплантированного слоев проводят следующим образом. Рекристаллизованный слой возникает на поверхности полупроводника в результате расплавления предварительно напыленного металла или сплава, содержащего легирующую примесь, с последующей рекристаллизацией раствора-расплава. Возникающий на поверхности полупроводника легированный слой достаточно тонкий (измеряемый единицами микрон), и в дальнейшем требуется разгонка примеси в глубь полупроводника. Операция формирования рекристаллизованного слоя соответствует «загонке примеси».

Цель загонки может быть также достигнута применением ионной имплантации, заключающейся во внедрении в кристаллическую решетку полупроводника высокоэнергетических ионов примеси, бомбардирующих его поверхность. Толщина ионно-имплантированных слоев определяется средней длиной пробега ионов в твердом теле. Ее величина измеряется десятками и сотнями ангстрем в зависимости от энергии ионов, что на 1–2 порядка меньше толщины рекристаллизованных слоев.

Операция загонки примеси является первым этапом комбинированного процесса легирования. Заключительному этапу разгонки предшествует операция окисления кремния с целью создания на его поверхности окисла SiO_2 , моделирующего отражающую границу, которая должна предотвращать диффузию атомов примеси из кристалла. Модель отражающей границы может быть реализована только для тех примесей, коэффициент диффузии которых в двуокиси кремния существенно меньше, чем в кремнии. Ориентировочно отношения этих коэффициентов для основных легирующих примесей в кремнии при 1100°C составляют $D_{\text{SiO}_2}/D_{\text{Si}} = 2,5 \cdot 10^{-4}$ для бора и $D_{\text{SiO}_2}/D_{\text{Si}} = 1,2 \cdot 10^{-2}$ для фосфора. Однако не для всех примесей справедливы такие соотношения: например, галлий и алюминий диффундируют в оксиде кремния в 400–500 раз быстрее, чем в кремнии, что исключает их использование в качестве легирующих примесей.

Таким образом, после окисления поверхности кремния заключительная стадия разгонки примеси реализуется в рамках модели *поверхностного источника с отражающей границей*, при этом рекристаллизованный слой соответствует *источнику конечной толщины* (см. формулу (4.7.18)), а ионно-имплантированный слой — *бесконечно тонкому источнику* (см. формулу (4.7.20) и рис. 4.12).

Подведем итоги вышеизложенным сведениям.

1. Модель диффузии из постоянного источника, описываемая уравнениями (4.7.1)–(4.7.4) (см. рис. 4.9), реализуется при легировании из парогазовой фазы и напыленного слоя металла.

2. Модель диффузии из полуограниченного источника в гомогенной структуре, описываемая уравнениями (4.7.1) и (4.7.7)–(4.7.9) (см. рис. 4.10), реализуется для процессов взаимного легирования слоев в гомоэпитаксиальных структурах.

3. Модель диффузии из полуограниченного источника в гетерогенной структуре, описываемая уравнениями (4.7.1), (4.7.7), (4.7.8) и (4.7.10)–(4.7.14) (см. рис. 4.11), реализуется при легировании из гетероэпитаксиальных слоев и поверхностных окислов.

4. Модель диффузии из поверхностного источника конечной толщины с отражающей границей, описываемая уравнениями (4.7.1) и (4.7.16)–(4.7.18), реализуется при легировании из рекристаллизованных слоев.

5. Модель диффузии из бесконечно тонкого поверхностного источника с отражающей границей, описываемая уравнениями (4.7.1), (4.7.17), (4.7.19) и (4.7.20) (см. рис. 4.12), реализуется при легировании из ионно-имплантированных слоев.

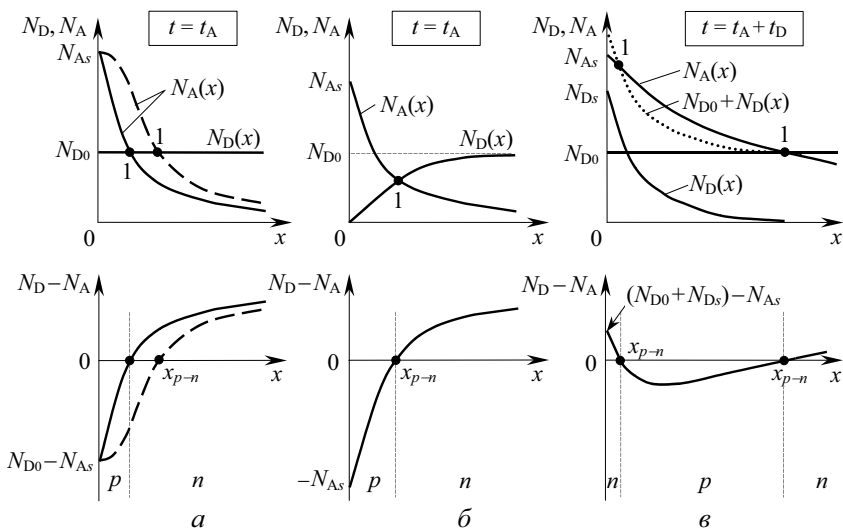


Рис. 4.13. Картины распределения донорной и акцепторной примесей для трех случаев легирования полупроводника n -типа с однородным распределением N_{D0} исходных донорных центров: *а* — легирование акцепторами из постоянного источника (сплошные кривые) и из бесконечно тонкого поверхностного источника с отражающей границей (штриховые кривые), *б* — легирование акцепторами из постоянного источника с одновременным испарением исходных доноров, *в* — комбинированное легирование акцепторами и донорами из постоянного источника

Кратко рассмотрим схемы создания диффузионных $p-n$ -переходов на основе модели легирования из постоянного источника, часто реализуемой в полупроводниковой технологии. На рис. 4.13 показаны распределения донорной и акцепторной примесей для трех случаев: *а* — диффузия акцепторной примеси в кристалл с однородной концентрацией N_{D0} исходных доноров; *б* — то же с одновременным испарением исходных доноров из кристалла; *в* — комбинированная диффузия акцепторной и донорной примесей в однородно легированный полупроводник n -типа.

Для сравнения на рис. 4.13 *а* штриховыми кривыми показан случай легирования из бесконечно тонкого источника с отражающей границей. Точки пересечения кривых на верхних графиках, обозначенные цифрами 1, дают положение $p-n$ -перехода x_{p-n} , показанное на нижних графиках, изображающих ход разности концентраций $N_D(x) - N_A(x)$. Для вычисления глубины залегания $p-n$ -перехода воспользуемся формулами предыдущего параграфа.

Случай 1. *Легирование полупроводника с однородной концентрацией доноров N_{D0} акцепторной примесью из постоянного источника* (сплошные кривые на рис. 4.13 *а*). В этом случае применение формулы (4.7.4) при $c_s = N_{As}$ позволяет записать условие образования $p-n$ -перехода в виде

$$N_A(x_{p-n}, t_A) \equiv N_{As} \operatorname{erfc} \frac{x_{p-n}}{2\sqrt{D_A t_A}} = N_{D0}.$$

Здесь D_A и t_A — коэффициент диффузии и время введения акцепторной примеси. Тогда глубина залегания $p-n$ -перехода

$$x_{p-n} = 2z_0 L_A, \quad \text{при этом} \quad \operatorname{erf} z_0 = 1 - \frac{N_{D0}}{N_{As}}, \quad (4.8.1)$$

где $L_A = \sqrt{D_A t_A}$ — длина диффузии для акцепторной примеси, а величина z_0 находится как корень второго уравнения (4.8.1).

Случай 2. *Легирование полупроводника с однородной концентрацией доноров N_{D0} акцепторной примесью из бесконечно тонкого поверхностного источника с отражающей границей* (штриховые кривые на рис. 4.13 *а*). В этом случае применение формулы (4.7.20) позволяет записать условие образования $p-n$ -перехода в виде

$$N_A(x_{p-n}, t_A) \equiv \frac{Q_A}{\sqrt{\pi D_A t_A}} \exp\left(-\frac{x_{p-n}^2}{4D_A t_A}\right) = N_{D0},$$

где Q_A — количество акцепторов на единицу площади, введенных в полупроводник на этапе загонки примеси. Отсюда получаем уравнение для нахождения глубины залегания $p-n$ -перехода:

$$x_{p-n}^2 = 4D_A t_A \ln \frac{Q_A}{N_{D0} \sqrt{\pi D_A t_A}}. \quad (4.8.2)$$

Из выражения (4.8.2) следует, что с ростом времени t_A величина x_{p-n} сначала возрастает, а затем начинает убывать, когда максимум функции Гаусса снижается до значений, близких

к величине N_{D0} . Нетрудно найти максимальное удаление $p-n$ -перехода от поверхности полупроводника:

$$x_{p-n}^{\max} = \sqrt{2D_A t_A^{\max}} \equiv \sqrt{2} L_A^{\max},$$

при этом

$$t_A^{\max} = \frac{1}{2,72 \pi D_A} \left(\frac{Q_A}{N_{D0}} \right)^2 = \frac{0,117}{D_A} \left(\frac{Q_A}{N_{D0}} \right)^2.$$

Случай 3. *Легирование полупроводника акцепторной примесью из постоянного источника с одновременным испарением исходной донорной примеси* (рис. 4.13 б). В этом случае применение формул (4.4.17) и (4.7.4) при $c_0 = N_{D0}$ и $c_s = N_{As}$ позволяет записать условие образования $p-n$ -перехода в следующем виде:

$$\begin{aligned} N_A(x_{p-n}, t_A) &\equiv N_{As} \operatorname{erfc} \frac{x_{p-n}}{2\sqrt{D_A t_A}} = \\ &= N_{D0} \operatorname{erf} \frac{x_{p-n}}{2\sqrt{D_D t_A}} \equiv N_D(x_{p-n}, t_A). \end{aligned}$$

Здесь D_A и D_D — коэффициенты диффузии акцепторов и доноров, t_A — время диффузионного введения акцепторов и одновременного испарения исходной донорной примеси. Отсюда получаем глубину залегания $p-n$ -перехода:

$$x_{p-n} = 2z_0 L_A, \quad \text{при этом} \quad \operatorname{erf} z_0 = 1 - \frac{N_{D0}}{N_{As}} \operatorname{erf} R z_0, \quad (4.8.3)$$

где $R = \sqrt{D_A/D_D} \equiv L_A/L_D$ — отношение диффузионных длин для акцепторов и доноров, а величина z_0 находится как корень второго уравнения (4.8.3).

Случай 4. *Комбинированное легирование полупроводника в виде последовательных циклов введения акцепторов и доноров из постоянного источника.* В подобных случаях обычно вместо диффузионной длины $L_{\text{диф}} = \sqrt{D t}$ вводят в рассмотрение эффективную диффузионную длину

$$L_{\text{диф}}^{\text{эф}} = \sqrt{D_1(T_1)t_1 + D_2(T_2)t_2 + \dots}.$$

Здесь $D_1(T_1)$, $D_2(T_2)$, ... — значения коэффициента диффузии для акцепторов или доноров при температурах нагрева T_1 , T_2 , ..., а t_1 , t_2 , ... — продолжительность циклов нагрева.

В данном случае (рис. 4.13 в) полупроводник n -типа имеет однородную концентрацию N_{D0} исходной донорной примеси. При проведении диффузии сначала вводятся акцепторы, а затем

доноры со следующими параметрами процесса: $D_1(T_1) \equiv D_A(T_A)$, $t_1 \equiv t_A$ — для первого («акцепторного») цикла и $D_2(T_2) \equiv D_D(T_D)$, $t_2 \equiv t_D$ — для второго («донорного») цикла. Тогда применение формулы (4.7.4) позволяет записать условие образования $p-n$ -перехода в следующем виде:

$$N_{D0} + N_{Ds} \operatorname{erfc} \frac{x_{p-n}}{2\sqrt{D_D(T_D)t_D}} = \\ = N_{As} \operatorname{erfc} \frac{x_{p-n}}{2\sqrt{D_A(T_A)t_A + D_A(T_D)t_D}}.$$

Здесь N_{Ds} и N_{As} — концентрация доноров и акцепторов на поверхности при $x = 0$ (см. рис. 4.13 в), $D_A(T_D)$ — коэффициент диффузии акцепторов при температуре $T_2 \equiv T_D$ для второго («донорного») цикла легирования за время $t_2 \equiv t_D$. Отсюда получаем глубину залегания $p-n$ -перехода:

$$x_{p-n} = 2z_0 L_D, \text{ при этом } \operatorname{erf} z_0 = \left(1 - \frac{N_{As} - N_{D0}}{N_{Ds}}\right) + \frac{N_{As}}{N_{Ds}} \operatorname{erf} R z_0, \quad (4.8.4)$$

где $R = \sqrt{D_D(T_D)t_D / [D_A(T_A)t_A + D_A(T_D)t_D]}$, $L_D = \sqrt{D_D(T_D)t_D}$ — диффузионная длина для донорной примеси, а величина z_0 находится как корень второго уравнения (4.8.4).

Таким образом, комбинированная диффузия примесей при определенных параметрах процесса легирования позволяет создавать транзисторные структуры типа $n-p-n$ или $p-n-p$.

4.9. Диффузионная и химическая кинетика гетерогенных процессов

До сих пор рассматривалась лишь диффузионная кинетика в задачах удаления из твердого тела и введения в него необходимых веществ. В технологической практике широкое распространение получили процессы, в которых одновременно с диффузионным массопереносом принципиальную роль играют гетерогенные химические реакции. Эти реакции происходят на поверхности твердого тела, граничащего со смежной (газовой или жидкой) фазой, движущейся относительно этого тела и подпитывающей компонентами реакционную поверхность.

Примером гетерогенного процесса могут служить газотранспортные химические реакции (см. п. 2.16), служащие для химического осаждения на подложке твердофазных (в том числе эпитаксиальных) слоев путем транспорта веществ в газовой фазе.

Наряду с осаждением, при помощи таких процессов решаются и прямо противоположные задачи удаления твердых веществ путем химического травления в газовой или жидкой фазе.

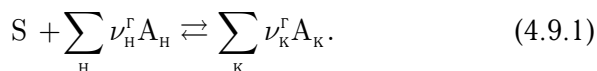
Сложные гетерогенные процессы, как правило, протекают в несколько последовательных стадий (этапов), среди которых:

- диффузионная доставка исходных химических реагентов из объема питающей фазы (газовой или жидкой) к реакционной поверхности;
- адсорбция молекул и атомов реагентов на поверхности;
- диффузия адсорбированных частиц (атомов и молекул) по поверхности;
- химические взаимодействия между частицами на реакционной поверхности (собственно химическая реакция);
- десорбция продуктов реакции с поверхности;
- диффузионный отвод продуктов реакции от поверхности вглубь внешней фазы (газовой или жидкой).

Для многостадийных процессов их результирующая скорость обычно лимитируется скоростью наиболее медленной стадии. Принято говорить, что процесс протекает **в диффузионной области**, если медленной лимитирующей стадией является диффузионный массоперенос. Наоборот, процесс протекает **в кинетической области**, если скорость процесса в целом определяется скоростью собственно химической реакции на границе раздела фаз как самой медленной стадии процесса.

Как правило, сорбционные процессы на поверхности твердого тела, рассмотренные ниже в гл. 5, являются достаточно быстрыми в сравнении с диффузионным массопереносом и химическими взаимодействиями. Поэтому в данный момент исключаем их из рассмотрения как мало существенные, поскольку они не лимитируют общую скорость гетерогенного процесса. Более того, в стационарных условиях поверхностные процессы (адсорбция, поверхностная диффузия, десорбция) неотделимы от собственно химической реакции и вместе с ней формируют кинетическую стадию процесса.

Для определенности будем рассматривать гетерогенные процессы с участием газовой фазы. По аналогии с химическими реакциями (2.12.13) и (2.16.1), запишем реакцию взаимодействия между твердым веществом S и многокомпонентной газовой средой в обобщенной форме (приведенной к одному молю твердого компонента S):



Здесь по-прежнему нижние индексы «н» и «к» соответствуют начальным и конечным участникам реакции, а верхний индекс «г» отображает их газообразное состояние.

Можно говорить, что реакция (4.9.1) формирует на реакционной поверхности парциальные *химические потоки* $J_{A_n}^{\text{хим}}$ и $J_{A_k}^{\text{хим}}$, порожденные компонентами A_n и A_k . Будем считать данные потоки отрицательными для исчезающих реагентов ($J_{A_n}^{\text{хим}} < 0$ при прямой реакции и $J_{A_k}^{\text{хим}} < 0$ при обратной реакции) и положительными для возникающих реагентов ($J_{A_k}^{\text{хим}} > 0$ при прямой реакции и $J_{A_n}^{\text{хим}} > 0$ при обратной реакции).

В стационарном режиме парциальные химические потоки жестко связаны между собой условием стехиометрии обобщенной химической реакции (4.9.1), записанным в форме

$$\frac{J_{A_n}^{\text{хим}}}{\nu_n^{\text{г}}} = -\frac{J_{A_k}^{\text{хим}}}{\nu_k^{\text{г}}} = J_S^{\text{хим}}, \quad (4.9.2)$$

где знак минус отражает противоположные знаки потоков для исчезающих и возникающих участников реакции. Соотношение (4.9.2) содержит химический поток $J_S^{\text{хим}}$ твердого компонента S, который определяет скорость травления (при $J_S^{\text{хим}} < 0$) или осаждения (при $J_S^{\text{хим}} > 0$) вещества S, получаемую для конкретной ситуации с использованием основного постулата химической кинетики (1.18.13).

Следуя формулам (4.1.1) и (4.1.2), представим результирующую скорость обратимой химической реакции (4.9.1) в виде разности скоростей прямой (v_1) и обратной (v_2) реакции (ср. формулы (1.23.28)–(1.23.29)):

$$\begin{aligned} v &= v_1 - v_2 = k_1(T) \prod_n p_n^{\nu_n^{\text{г}}} - k_2(T) \prod_k p_k^{\nu_k^{\text{г}}} = \\ &= k_1(T) \prod_n p_n^{\nu_n^{\text{г}}} \left(1 - K_p^{-1} \frac{\prod_k p_k^{\nu_k^{\text{г}}}}{\prod_n p_n^{\nu_n^{\text{г}}}} \right) = k_2(T) \prod_k p_k^{\nu_k^{\text{г}}} \left(K_p \frac{\prod_n p_n^{\nu_n^{\text{г}}}}{\prod_k p_k^{\nu_k^{\text{г}}}} - 1 \right), \end{aligned} \quad (4.9.3)$$

где, в соответствии с формулой (4.1.5), использована константа равновесия $K_p = k_1/k_2$.

Здесь по сравнению с выражениями (4.1.2) вместо концентраций c_i введены парциальные давления $p_i = c_i RT$, а соответствующие температурные зависимости включены в константы скорости $k_1(T)$ и $k_2(T)$.

Для преобразования формулы (4.9.3) воспользуемся уравнением (2.13.5) для изотермы Вант-Гоффа, переписанным в следующем виде:

$$\Delta G = RT \ln \left(\frac{1}{K_p} \frac{\prod_{\text{к}} p_{\text{к}}^{\nu_{\text{к}}^r}}{\prod_{\text{н}} p_{\text{н}}^{\nu_{\text{н}}^r}} \right). \quad (4.9.4)$$

Нетрудно убедиться в том, что подстановка (4.9.4) в выражение (4.9.3) дает искомые соотношения (ср. формулу (1.23.34)):

$$\begin{aligned} v &= k_1(T) \left[1 - \exp \left(\frac{\Delta G}{RT} \right) \right] \prod_{\text{н}} p_{\text{н}}^{\nu_{\text{н}}^r} = \\ &= k_2(T) \left[\exp \left(-\frac{\Delta G}{RT} \right) - 1 \right] \prod_{\text{к}} p_{\text{к}}^{\nu_{\text{к}}^r}. \end{aligned} \quad (4.9.5)$$

Вышеизложенное позволяет отождествить плотность химического потока $J_{\text{S}}^{\text{хим}}$ (моль/(м²·с)) со скоростью гомогенной химической реакции v (моль/(м³·с)). На основании равенства (4.9.5) запишем для гетерогенной реакции (4.9.1) следующие выражения для скорости химического травления (реакция смещена вправо) и осаждения (реакция смещена влево) вещества S:

$$\begin{aligned} |J_{\text{S}}^{\text{хим}}|_{\text{тр}} &= k_{\text{тр}}(T) \prod_{\text{н}} p_{\text{н}}^{\nu_{\text{н}}^r}, \\ k_{\text{тр}}(T) &= k_1(T) \left[1 - \exp \left(\frac{\Delta G_{\text{тр}}}{RT} \right) \right] \approx k_1(T) \frac{|\Delta G_{\text{тр}}|}{RT}, \end{aligned} \quad (4.9.6)$$

$$\begin{aligned} |J_{\text{S}}^{\text{хим}}|_{\text{ос}} &= k_{\text{ос}}(T) \prod_{\text{к}} p_{\text{к}}^{\nu_{\text{к}}^r}, \\ k_{\text{ос}}(T) &= k_2(T) \left[1 - \exp \left(-\frac{\Delta G_{\text{ос}}}{RT} \right) \right] \approx k_2(T) \frac{\Delta G_{\text{ос}}}{RT}. \end{aligned} \quad (4.9.7)$$

Здесь введены эффективные константы скорости травления ($k_{\text{тр}}$) и скорости осаждения ($k_{\text{ос}}$), для которых приближенные равенства соответствуют малому химическому пересыщению, такому что $|\Delta G| \ll RT$ (ср. формулу (1.23.35)). При этом $\Delta G_{\text{тр}} = -|\Delta G_{\text{тр}}| < 0$ для процесса травления (когда реакция (4.9.1) протекает слева направо) и $\Delta G_{\text{ос}} > 0$ для процесса осаждения (когда реакция (4.9.1) протекает справа налево). Следует иметь в виду, что величина ΔG , входящая в выражения (4.9.6) и (4.9.7), не является параметром системы, а сама зависит от состава газовой фазы и подлежит определению в каждом конкретном случае.

Как уже отмечалось, большинство гетерогенных процессов сопровождается движением газовой (или жидкой) фазы вдоль границы раздела с твердым телом. Если диффузионный поток,

вызванный неоднородной концентрацией компонентов, происходит на фоне общего массового движения вещества, вызванного конвекцией, то такую диффузию принято называть *конвективной диффузией*.

В этом случае к диффузионному потоку (1.18.3) для i -го компонента должна быть добавлена конвективная составляющая в форме $\mathbf{J}_i^{\text{кон}} = c_i \mathbf{v}$. Тогда полная плотность потока частиц записывается в виде суммы диффузионного ($\mathbf{J}_i^{\text{диф}}$) и конвективного ($\mathbf{J}_i^{\text{кон}}$) потоков (ср. формулу (1.22.18)):

$$\mathbf{J}_i^{\Sigma} = \mathbf{J}_i^{\text{диф}} + \mathbf{J}_i^{\text{кон}} = -D_i \text{grad } c_i + c_i \mathbf{v}. \quad (4.9.8)$$

Здесь $\mathbf{v}$ — общая скорость движения частиц газовой (или жидкой) среды (см. формулу (1.22.1)), которая вызвана, например, перепадом полного давления в проточной системе или разностью температур в условиях естественной конвекции.

Подобная ситуация имеет место при конвективном потоке газовой (или жидкой) среды вдоль границы химически активного твердого тела S , потребляющего или рождающего частицы этой среды. Неоднородная концентрация таких частиц в направлении нормали к границе изменяется по закону диффузии, формируя диффузионный поток $\mathbf{J}_i^{\text{диф}}$ в этом направлении, а конвективный массоперенос со скоростью $\mathbf{v}$ направлен по касательной к границе раздела фаз.

Картина неоднородного распределения концентрации i -го компонента в газовой фазе в районе границы твердого тела S качественно изображена на рис. 4.14, где введена *эффективная толщина диффузионного слоя* δ_i . В этом случае вместо формулы (1.18.3), выражающей дифференциальную связь между плотностью диффузионного потока $\mathbf{J}_i^{\text{диф}} = \mathbf{e}_x J_i^{\text{диф}}$ и концентрацией частиц $c_i(x)$, на границе раздела фаз используют следующее алгебраическое соотношение:

$$J_i^{\text{диф}} \equiv -D_i \frac{\partial c_i(x)}{\partial x} \Big|_{x=0} = D_i \frac{c_{si} - c_{0i}}{\delta_i} \equiv \beta_i (c_{si} - c_{0i}), \quad (4.9.9)$$

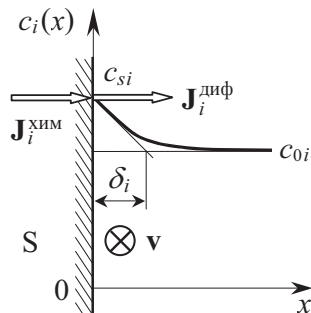


Рис. 4.14. К определению эффективной толщины δ_i диффузионного слоя для i -го компонента в газовой (или жидкой) фазе. Кругок $\otimes$ обозначает конвективный поток газа (или жидкости) со скоростью течения $\mathbf{v}$ вдоль границы твердого тела S

где c_{si} и c_{0i} — значения концентрации на границе твердой фазы и внутри объема газовой фазы. Здесь введена величина $\beta_i = D_i/\delta_i$, называемая *коэффициентом диффузионного массопереноса* (или массоотдачи) и имеющая размерность м/с.

Применение алгебраического равенства (4.9.9) вместо дифференциального соотношения (1.18.3) существенно упрощает анализ неравновесных систем с одновременным участием диффузионных и химических процессов. При этом коэффициент массопереноса β_i считается феноменологически заданной величиной.

На рис. 4.14 в дополнение к диффузионному потоку i -го компонента показан также химический поток $J_i^{\text{хим}} = \mathbf{e}_x J_i^{\text{хим}}$, формируемый гетерогенной реакцией (4.9.1). В стационарных условиях естественно потребовать равенства этих потоков,

$$J_i^{\text{хим}} = J_i^{\text{диф}}, \quad (4.9.10)$$

где номер i соответствует любому из газовых компонентов A_n или A_k , участвующих в реакции (4.9.1). Картина на рис. 4.14 справедлива для конечных компонентов A_k при протекании реакции слева направо или для начальных компонентов A_n при обратном протекании реакции.

Простейшая связь типа (4.9.9) между плотностью диффузионного потока $J_i^{\text{диф}}$ и перепадом концентрации ($c_{si} - c_{0i}$) существенно осложняется в тех случаях, когда гетерогенная химическая реакция (4.9.1) на поверхности твердой фазы S протекает с изменением числа молей в газовой фазе, определяемым как $\Delta\nu^r = \sum_k \nu_k^r - \sum_n \nu_n^r$ (см. формулу (2.12.8)).

Покажем, что при $\Delta\nu^r \neq 0$ для реакции (4.9.1) в системе возникает дополнительный газовый поток по нормали к реакционной поверхности, называемый *стефановским потоком*, направление которого определяется, как увидим ниже, знаком $\Delta\nu^r$. Для этого записываем x -ю компоненту плотности потока (4.9.8) с учетом искомого стефановского потока $J_i^{\text{ст}}$ в виде

$$J_i \equiv J_i^{\text{диф}} + J_i^{\text{ст}} = -D_i \frac{\partial c_i}{\partial x} + c_i v_{\text{ст}} \equiv -\frac{D_i}{RT} \frac{\partial p_i}{\partial x} + \frac{p_i}{RT} v_{\text{ст}}. \quad (4.9.11)$$

Здесь $v_{\text{ст}}$ — стефановская скорость движения газовой смеси по нормали к реакционной поверхности раздела фаз, которую необходимо найти. Суммирование по всем компонентам смеси дает общий газовый поток:

$$J \equiv \sum_i J_i = -\frac{D}{RT} \frac{\partial P}{\partial x} + \frac{P}{RT} v_{\text{ст}}, \quad (4.9.12)$$

где введен результирующий коэффициент диффузии D газовой смеси и общее давление P в системе.

Обычно гидродинамическое сопротивление движению газа очень мало, так что можно положить $\partial P / \partial x = 0$. Тогда из формулы (4.9.12) получаем выражение для скорости стефановского потока:

$$v_{\text{ст}} = \frac{RT}{P} J. \quad (4.9.13)$$

Как и следовало ожидать, стефановская скорость обращается в нуль в отсутствие результирующего потока по направлению нормали к поверхности (при $J = 0$).

Следовательно, с использованием выражения (4.9.13) плотность потока (4.9.11) для интересующего нас i -го компонента принимает следующий вид:

$$J_i = -\frac{D_i}{RT} \frac{\partial p_i}{\partial x} + \frac{p_i}{P} J. \quad (4.9.14)$$

Для нахождения общего газового потока J воспользуемся равенством (4.9.10), в котором при наличии стефановского потока величина $J_i^{\text{диф}}$ должна быть заменена на J_i . Тогда из (4.9.10) получаем $J_i^{\text{хим}} = J_i \equiv J_i^{\text{диф}} + J_i^{\text{ст}}$.

В качестве i -го компонента возьмем для определенности любой из начальных реагентов, стоящих в левой части реакции (4.9.1) (например, транспортный агент X для газотранспортной химической реакции типа (2.16.1)). Таким образом, соотношение (4.9.2) между парциальными химическими потоками переписываем в следующем виде:

$$\frac{J_i^{\text{хим}}}{\nu_i^{\Gamma}} = \frac{J_{\text{Ан}}^{\text{хим}}}{\nu_{\text{н}}^{\Gamma}} = -\frac{J_{\text{Ак}}^{\text{хим}}}{\nu_{\text{к}}^{\Gamma}}. \quad (4.9.15)$$

Используем соотношения (4.9.15) для вычисления общего газового потока:

$$\begin{aligned} J &= \sum_i J_i = \sum_i J_i^{\text{хим}} \equiv \sum_{\text{н}} J_{\text{Ан}}^{\text{хим}} + \sum_{\text{к}} J_{\text{Ак}}^{\text{хим}} = \\ &= \left(\sum_{\text{н}} \nu_{\text{н}}^{\Gamma} - \sum_{\text{к}} \nu_{\text{к}}^{\Gamma} \right) \frac{J_i^{\text{хим}}}{\nu_i^{\Gamma}} = -\frac{\Delta \nu^{\Gamma}}{\nu_i^{\Gamma}} J_i, \end{aligned} \quad (4.9.16)$$

где $\Delta \nu^{\Gamma} = \sum_{\text{к}} \nu_{\text{к}}^{\Gamma} - \sum_{\text{н}} \nu_{\text{н}}^{\Gamma}$ — изменение числа молей в газовой фазе при одном пробеге реакции (4.9.1) в прямом направлении (слева направо).

Подстановка (4.9.16) в формулу (4.9.13) дает искомое выражение для скорости стефановского потока:

$$v_{\text{ст}} = \frac{RT}{P} J = -\frac{RT}{P} \frac{\Delta\nu^\Gamma}{\nu_i^\Gamma} J_i. \quad (4.9.17)$$

Поскольку рассматриваемый i -й компонент был выбран из числа начальных реагентов A_n , исчезающих при пробеге реакции слева направо, то $J_i < 0$ в формуле (4.9.17). Это означает, что стефановский поток направлен от реакционной поверхности ($v_{\text{ст}} > 0$) при протекании реакции с увеличением газового объема ($\Delta\nu^\Gamma > 0$) или к поверхности ($v_{\text{ст}} < 0$) при уменьшении объема ($\Delta\nu^\Gamma < 0$).

Подставляя поток J в форме (4.9.16) в уравнение (4.9.14), получаем следующее выражение для плотности суммарного (диффузионного и стефановского) потока i -го компонента:

$$J_i \equiv J_i^{\text{диф}} + J_i^{\text{ст}} = -\frac{D_i}{1 + (\Delta\nu^\Gamma/\nu_i^\Gamma)(RT/P)} c_i \frac{\partial c_i}{\partial x}. \quad (4.9.18)$$

Используем полученное выражение (4.9.18) для определения эффективной толщины δ_i диффузионного слоя и коэффициента массопереноса $\beta_i = D_i/\delta_i$ при наличии стефановского потока. Для этого по аналогии с (4.9.9) записываем для границы раздела фаз следующее соотношение (см. рис. 4.14):

$$\begin{aligned} J_i &= -\frac{D_i}{b_i} \frac{\partial \ln[1 + b_i c_i(x)]}{\partial x} \Big|_{x=0} = \\ &= \frac{D_i}{b_i} \frac{\ln(1 + b_i c_{si}) - \ln(1 + b_i c_{0i})}{\delta_i} \equiv \frac{\beta_i}{b_i} \ln \frac{1 + b_i c_{si}}{1 + b_i c_{0i}}, \end{aligned} \quad (4.9.19)$$

где введено обозначение $b_i = (\Delta\nu^\Gamma/\nu_i^\Gamma)(RT/P)$.

Легко убедиться в том, что для разбавленной газовой смеси при $p_i \ll P$ в случае, когда изменение числа газовых молей невелико, так что $b_i c_i = (\Delta\nu^\Gamma/\nu_i^\Gamma)(p_i/P) \ll 1$ и $\ln(1 + b_i c_i) \approx b_i c_i$, вклад стефановского потока становится пренебрежимо малым. Тогда выражение (4.9.19) принимает вид (4.9.9). Именно это выражение будет использовано в дальнейшем.

Следует иметь в виду, что при анализе гетерогенных процессов на реакционной границе раздела фаз парциальные давления p_n и p_k (или концентрации $c_n = p_n/RT$ и $c_k = p_k/RT$), входящие в формулы (4.9.6) и (4.9.7), соответствуют их поверхностным значениям $p_{si} = c_{si}RT$ ($i = n, k$), взятым на границе раздела фаз при $x = 0$ (см. рис. 4.14).

Строго говоря, рассуждения, приведшие к введению диффузионного слоя соотношениями (4.9.9) и (4.9.19), справедливы лишь в пренебрежении вязкостью в газовой (или жидкой) среде. Ее учет приводит к появлению вязкостного пограничного слоя, называемого *динамическим слоем*. В пределах этого слоя касательная компонента скорости плавно нарастает от нуля на стенке до установившейся величины в потоке. Толщина динамического слоя изменяется в зависимости от характера течения вязкой среды. При относительно небольших скоростях потока имеет место *ламинарное течение* с упорядоченным движением отдельных слоев по отношению друг к другу. Возрастание силы трения между слоями с увеличением скорости приводит к хаотическим пульсациям при обмене поперечным импульсом между слоями. Это обеспечивает неупорядоченное движение частиц, называемое *турбулентным течением*. Ламинарный и турбулентный режимы течения влияют на толщину как динамического, так и диффузионного слоя. Однако для последующего рассмотрения принципиальным является учет неоднородности в распределении лишь концентрации частиц, характеризуемой толщиной диффузионного слоя δ_i и коэффициентом массопереноса $\beta_i = D_i/\delta_i$. Эти величины будем считать феноменологически заданными, не вдаваясь в детали гидродинамического течения вязкой среды.

4.10. Маскирующие свойства слоев двуокиси кремния

Диэлектрические слои имеют различное функциональное назначение в полупроводниковых приборах и интегральных микросхемах, играя как активную, так и пассивную роль. *Активную функцию* эти слои выполняют в МДП-структурах в качестве подзатворного диэлектрика, а также в качестве изолятора активных и пассивных элементов ИМС. В кремниевой технологии для этой цели используют аморфные пленки двуокиси кремния SiO_2 и нитрида кремния Si_3N_4 . *Пассивная функция* диэлектрических слоев реализуется при их использовании в качестве защитных покрытий для электрической стабилизации, пассивации и защиты поверхности приборов (особенно в местах выхода $p-n$ -переходов) от внешних атмосферных воздействий, в первую очередь, от кислорода и паров воды. Для этой цели, кроме пленок SiO_2 , Si_3N_4 и Al_2O_3 , успешно используют легкоплавкие силикатные или халькогенидные стекла, а также кремнийорганические полимерные пленки, влагостойкие лаки и компаунды.

Для технологического использования диэлектрические пленки представляют интерес в качестве *маскирующих покрытий*,

локально защищающих поверхность полупроводника от технологических воздействий, например при легировании методом диффузии или ионной имплантации. Исключительно важное значение маскирующая функция диэлектрических покрытий приобретает в планарной технологии, где локальность и селективность технологических воздействий играет первостепенную роль.

Из полупроводников, широко используемых в современной электронике, только кремний имеет химически стабильный и структурно плотный собственный окисел SiO_2 . На поверхности германия, арсенида галлия и других полупроводников собственные окислы термически нестабильны и склонны к гидролизу. По этой причине будем анализировать маскирующую функцию диэлектрических слоев на примере пары Si-SiO_2 , а в следующем параграфе рассмотрим физико-технологические особенности окисления кремния.

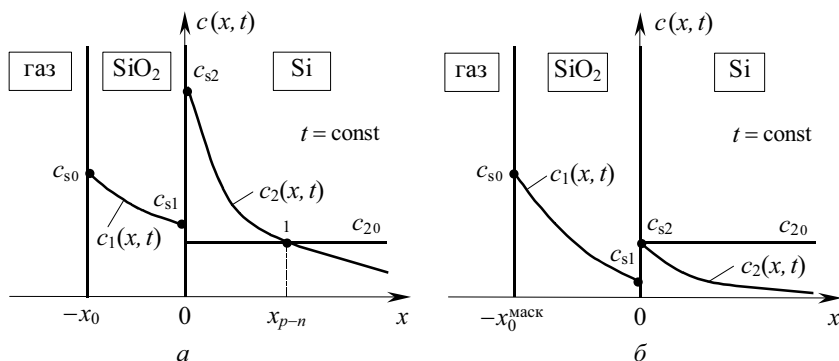


Рис. 4.15. Распределение легирующей примеси в двуокиси кремния $c_1(x, t)$ и в кремнии $c_2(x, t)$ при диффузии ее из газовой фазы как постоянного источника через тонкий слой SiO_2 с образованием $p-n$ -перехода (а) и через слой SiO_2 маскирующей толщины $x_0^{\text{маск}}$ (б)

Под маскирующей функцией двуокиси кремния понимается способность предотвращать диффузионное проникновение примесей сквозь нее при проведении высокотемпературных локальных процессов, в частности, при локальном диффузионном легировании кремния через окна, вскрытые в поверхностном слое маскирующего покрытия SiO_2 .

Рассмотрим диффузионную задачу проникновения легирующей примеси (например, фосфора или бора) в кремний через слой SiO_2 (толщиной x_0) с образованием $p-n$ -перехода в точке 1 на расстоянии x_{p-n} от границы раздела $\text{SiO}_2\text{-Si}$, как показано на рис. 4.15 а. Это означает, что вводимая примесь (донор или

акцептор) с искомой концентрацией $c_1(x, t)$ противоположна по типу исходной (акцепторной или донорной) примеси с заданной концентрацией $c_{20}(x) = \text{const}$.

Пусть легирование производится из газовой фазы, содержащей легирующую примесь с парциальным давлением p_0 . В этом случае, как отмечалось в п. 4.8, работает модель постоянного источника, обеспечивающего неизменную поверхностную концентрацию растворенной в SiO_2 примеси, равную $c_{s0} = s(T)p_0$.

Для решения поставленной диффузионной задачи надо использовать следующие соотношения (рис. 4.15 а):

- уравнение диффузии

$$\frac{\partial c_{1,2}(x, t)}{\partial t} = D_{1,2} \frac{\partial^2 c_{1,2}(x, t)}{\partial x^2}, \quad (4.10.1)$$

где $D_{1,2}$ — коэффициенты диффузии атомов примеси в двуокиси кремния (D_1) и в кремнии (D_2);

- начальные условия

$$c_1(x, t)|_{t=0} = 0 \quad \text{при} \quad -x_0 < x < 0, \quad (4.10.2, \text{а})$$

$$c_2(x, t)|_{t=0} = 0 \quad \text{при} \quad 0 < x < \infty, \quad (4.10.2, \text{б})$$

выражающие отсутствие примеси в SiO_2 и Si при $t \leq 0$;

- граничные условия

$$c_1(x, t)|_{x=-x_0} = c_{s0}, \quad (4.10.3, \text{а})$$

$$c_2(x, t)|_{x \rightarrow \infty} \rightarrow 0, \quad (4.10.3, \text{б})$$

$$D_1 \frac{\partial c_1(x, t)}{\partial x} \Big|_{x=0} = D_2 \frac{\partial c_2(x, t)}{\partial x} \Big|_{x=0}, \quad (4.10.3, \text{в})$$

$$c_1(x, t)|_{x=0} = K c_2(x, t)|_{x=0}, \quad (4.10.3, \text{г})$$

выражающие: а — постоянство поверхностной концентрации на границе с питающей газовой фазой; б — отсутствие рассматриваемой примеси в Si при бесконечном удалении от границы; в — непрерывность потока из-за отсутствия накопления примеси на границе SiO_2 – Si ; г — перераспределение примеси между SiO_2 и Si из-за различия ее растворимости, выражаемого температурно-зависимым коэффициентом распределения примеси (см. обоснование в формуле (4.7.12)): $K(T) = c_{\text{SiO}_2}/c_{\text{Si}} \equiv c_{s1}/c_{s2}$, где $c_{s1} = c_1(0, t)$ и $c_{s2} = c_2(0, t)$.

Внутри области, занятой кремнием (при $0 < x < x_{\text{max}}$), решение поставленной задачи имеет следующий вид:

$$c_2(x, t) \approx c_{s0} \frac{2}{K+R} \operatorname{erfc} \left(\frac{x_0}{2\sqrt{D_1 t}} + \frac{x}{2\sqrt{D_2 t}} \right), \quad (4.10.4)$$

где введен температурно-зависимый параметр

$$R(T) = \sqrt{D_2(T)/D_1(T)} \equiv \sqrt{D_{\text{Si}}(T)/D_{\text{SiO}_2}(T)}.$$

Приблизительное равенство в (4.10.4) связано с тем фактом, что область применимости этого выражения не распространяется на бесконечно большие расстояния x , а ограничена максимальным значением $x_{\max} = x_0 R \equiv x_0 \sqrt{D_{\text{Si}}/D_{\text{SiO}_2}}$. Так как обычно $D_{\text{SiO}_2} \ll D_{\text{Si}}$, то $R \gg 1$, что дает достаточно большие и практически приемлемые значения $x_{\max} \gg x_0$.

Именно этой ситуации соответствует картина, качественно изображенная на рис. 4.15 а.

Условие образования $p-n$ -перехода в виде $c_2(x_{p-n}, t) = c_{20}$ записывается с использованием выражения (4.10.4):

$$\frac{2c_{s0}}{K+R} \operatorname{erfc} \left(\frac{x_0}{2\sqrt{D_1 t}} + \frac{x_{p-n}}{2\sqrt{D_2 t}} \right) = c_{20}. \quad (4.10.5)$$

Равенство (4.10.5) является трансцендентным уравнением для нахождения глубины x_{p-n} залегания $p-n$ -перехода, графическое решение которого показано точкой 1 на рис. 4.15 а.

Выражение (4.10.4) позволяет записать поверхностные значения c_{s1} и c_{s2} , так как $c_{s2} = c_2(0, t)$ и $c_{s1} = Kc_{s2} = Kc_2(0, t)$. Отсюда получаем

$$\begin{aligned} c_{s1} &= c_{s0} \frac{2K}{K+R} \operatorname{erfc} \frac{x_0}{2\sqrt{D_1 t}}, \\ c_{s2} &= c_{s0} \frac{2}{K+R} \operatorname{erfc} \frac{x_0}{2\sqrt{D_1 t}}. \end{aligned} \quad (4.10.6)$$

Эти формулы позволяют выяснить поведение величины x_{p-n} с увеличением толщины x_0 слоя SiO_2 при $t = \text{const}$, $T = \text{const}$ и $c_{s0} = \text{const}$ (за счет неизменности состава газовой фазы), что обеспечивает постоянство значений $K(T)$, $R(T)$ и $D_1(T)$.

Как видно из рис. 4.6, функция $\operatorname{erfc} z$ при $z = x_0/(2\sqrt{D_1 t})$ уменьшается с ростом x_0 . Это, согласно (4.10.6), понижает поверхностные значения c_{s1} и c_{s2} . Следовательно, кривая $c_2(x, t)$ на рис. 4.15 а опускается вниз, а точка 1 смещается влево, что приводит к уменьшению величины x_{p-n} . Дальнейшим увеличением толщины x_0 можно придти к ситуации, при которой $c_{s2} = c_{20}$ и $x_{p-n} = 0$, так что $p-n$ -переход возникает как раз на границе раздела $\text{SiO}_2\text{-Si}$.

Соответствующая описанной ситуации толщина слоя двуокиси кремния называется *маскирующей толщиной*. Картина, качественно изображенная на рис. 4.15 б, соответствует слою двуокиси кремния маскирующей толщины.

Из формулы (4.10.5) при $x_0 = x_0^{\text{маск}}$ и $x_{p-n} = 0$ получаем уравнение для нахождения маскирующей толщины $x_0^{\text{маск}}$ следующего вида:

$$\operatorname{erfc} \frac{x_0^{\text{маск}}}{2\sqrt{D_1 t}} = \frac{c_{20}}{2c_{s0}} (K + R). \quad (4.10.7)$$

Графическое решение трансцендентного уравнения (4.10.7) показано на рис. 4.16. Здесь функция $\operatorname{erfc} z$ при $z = x_0/(2\sqrt{D_1 t})$ пересекается горизонтальной прямой $(c_{20}/2c_{s0})(K + R)$, представляющей собой правую часть уравнения (4.10.7). Точка их пересечения дает искомую величину $z^{\text{маск}} = x_0^{\text{маск}}/(2\sqrt{D_1 t})$.

Эффективность маскирования считается тем выше, чем меньше величина $x_0^{\text{маск}}$. Как следует из графического решения уравнения (4.10.7), при постоянстве c_{20} и c_{s0} , это достигается за счет уменьшения коэффициента диффузии $D_1 \equiv D_{\text{SiO}_2}$ и увеличения как коэффициента распределения $K = c_{\text{SiO}_2}/c_{\text{Si}}$, так и параметра $R = \sqrt{D_{\text{Si}}/D_{\text{SiO}_2}}$. Следовательно, для повышения маскирующих свойств надо не просто выбирать диэлектрики с минимальным коэффициентом диффузии D_1 . Следует подбирать такие пары диэлектрик–полупроводник, в которых для диэлектрика коэффициент диффузии примеси D_1 был бы много меньше, а ее растворимость c_1 много больше по сравнению с аналогичными величинами D_2 и c_2 для полупроводника.

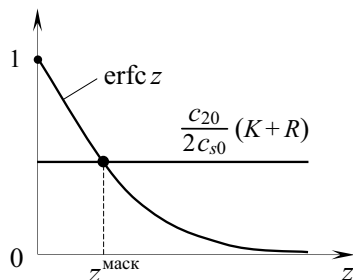


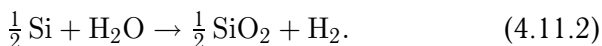
Рис. 4.16. Графическое нахождение маскирующей толщины $x_0^{\text{маск}}$ для двуокиси кремния

4.11. Кинетика термического окисления кремния

Как уже было отмечено, кремний — единственный из полупроводников, который имеет стабильный собственный окисел SiO_2 , пригодный для технологического использования в качестве маскирующего покрытия при проведении высокотемпературных локальных процессов. В свое время именно отработка надежного технологического процесса нанесения маскирующих слоев SiO_2

на поверхность кремния дала импульс развитию планарной технологии, являющейся основой создания современных интегральных микросхем.

В настоящее время окисление поверхности кремния осуществляют с помощью следующих химических реакций:



Процесс проводят при температуре в интервале $900\text{--}1200^\circ\text{C}$ либо в сухом кислороде по реакции (4.11.1), либо в парах воды по реакции (4.11.2), либо в увлажненном кислороде с использованием обеих реакций. На практике нередко применяют комбинированное окисление: сначала в сухом кислороде формируют совершенную границу раздела при толщине оксида до $0,1\text{ мкм}$, а затем в увлажненном кислороде увеличивают толщину слоя до $0,4\text{--}0,8\text{ мкм}$.

Согласно данным радиоизотопного анализа, окисляющий агент (O_2 или H_2O) диффундирует через растущий слой оксида, т. е. процесс окисления происходит на внутренней границе раздела $\text{SiO}_2\text{--Si}$. Так как мольный объем двуокиси кремния больше, чем кремния, то примерно половина образовавшегося слоя SiO_2 замещает по толщине окисленный кремний, а другая половина растет наружу.

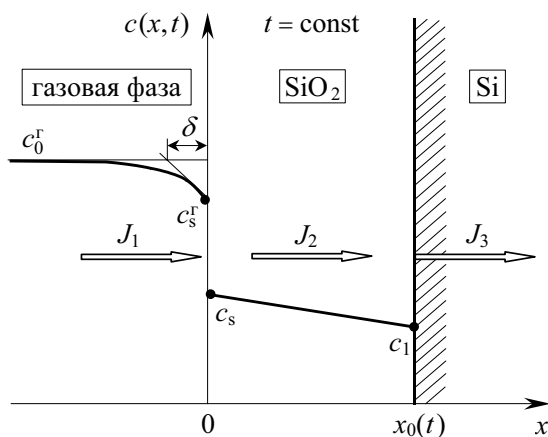


Рис. 4.17. Распределение концентрации молекул окислителя в газовой фазе и в растущем слое SiO_2 для процесса окисления кремния

Схематически процесс окисления изображен на рис. 4.17, где показано распределение концентрации молекул окислителя

$c(x, t)$ вдоль оси x , перпендикулярной плоскости пластины кремния. Толщина растущего слоя окисла как функция времени t обозначена через $x_0(t)$. Границе окисла с газовой фазой, содержащей молекулы окислителя, соответствует координата $x = 0$. На рис. 4.17 введены следующие обозначения для концентрации молекул окислителя: c_0^r — в объеме газовой фазы; c_s^r и c_s — на границе раздела фаз (при $x = 0$), соответственно, в газовой фазе и в двуокиси кремния; c_1 — на границе между SiO_2 и Si .

По современным представлениям процесс роста окисла на поверхности кремния происходит в три стадии:

- диффузионный транспорт молекул окислителя из объема газовой фазы к поверхности SiO_2 , характеризуемый потоком J_1 ;
- диффузионный транспорт растворенных в оксиде молекул окислителя через слой SiO_2 к границе $\text{SiO}_2\text{--Si}$, характеризуемый потоком J_2 ;
- химическая реакция окисления кремния, протекающая на границе $\text{SiO}_2\text{--Si}$ и характеризующаяся потоком J_3 .

Целью изучения кинетики процесса окисления является установление временной зависимости толщины $x_0(t)$ слоя окисла и выяснение влияния на нее управляемых технологических параметров, таких как температура T и парциальное давление p_0 окислителя в объеме газовой фазы, определяющее c_0^r . Толщину $x_0(t)$ можно связать с величинами T и c_0^r , используя выражения для трех потоков J_1 , J_2 и J_3 , которые в стационарных условиях должны быть равными друг другу, т. е.

$$J_1 = J_2 = J_3 \equiv J. \quad (4.11.3)$$

Запишем выражения для потоков молекул окислителя, соответствующих трем этапам процесса окисления

Диффузионный поток J_1 определяет подвод молекул окислителя из объема внешней газовой фазы к наружной поверхности растущего слоя двуокиси кремния. Внутри газовой фазы концентрация молекул окислителя c_0^r соответствует парциальному давлению газовых молекул:

$$p_0 = c_0^r k_B T. \quad (4.11.4)$$

Поскольку слой окисла растет непрерывно, то происходит непрерывное растворение молекул окислителя в двуокиси кремния с обеднением приповерхностной области газа этими молекулами. По этой причине их концентрация c_s^r на границе раздела фаз при $x = 0$ всегда меньше, чем в объеме, т. е. $c_s^r < c_0^r$ (рис. 4.17). Следовательно, в приповерхностном диффузионном слое толщиной δ

возникает градиент концентрации молекул окислителя, который создает диффузионный поток по направлению к границе, равный (см. формулу (4.9.9))

$$J_1 = -D^r \frac{\partial c^r(x)}{\partial x} \bigg|_{x=0} = D^r \frac{c_0^r - c_s^r}{\delta} \equiv \beta(c_0^r - c_s^r). \quad (4.11.5)$$

Здесь β — коэффициент массопереноса в газовой фазе, связанный соотношением $\beta = D^r/\delta$ с коэффициентом диффузии D^r молекул окислителя. Толщина диффузионного слоя δ является феноменологическим параметром, который определяется экспериментально или рассчитывается методами газодинамики.

Плотность потока J_1 можно выразить через поверхностную концентрацию c_s молекул окислителя, растворенных в SiO_2 . Процесс растворения происходит в молекулярной форме и подчиняется закону Генри:

$$c_s = sp_s. \quad (4.11.6)$$

Здесь s — коэффициент растворимости молекул окислителя в двуокиси кремния, а p_s — парциальное давление окислителя на межфазной границе при $x = 0$, записанное по аналогии с выражением (4.11.4) как $p_s = c_s^r k_B T$. Тогда равенство (4.11.6) принимает следующий вид:

$$c_s = c_s^r sk_B T \equiv sp_s, \quad \text{откуда} \quad c_s^r = \frac{c_s}{sk_B T}. \quad (4.11.7)$$

Аналогично этому, для объемной концентрации c_0^r можно ввести величину

$$c_0 = c_0^r sk_B T \equiv sp_0, \quad \text{откуда} \quad c_0^r = \frac{c_0}{sk_B T}. \quad (4.11.8)$$

Величина c_0 определяет концентрацию растворенных в SiO_2 молекул окислителя, если в законе Генри (4.11.6) вместо давления p_s подставить парциальное давление p_0 окислителя в объеме газовой фазы, равное (4.11.4). С учетом равенств (4.11.7) и (4.11.8), выражение (4.11.5) для плотности диффузионного потока принимает вид

$$J_1 = \bar{\beta}(c_0 - c_s). \quad (4.11.9)$$

где $\bar{\beta} = \beta/sk_B T \equiv D^r/\delta sk_B T$ — коэффициент массопереноса молекул окислителя в газовой фазе, приведенный к твердой фазе (к окислу SiO_2).

Диффузионный поток J_2 определяет транспорт молекул окислителя через слой SiO_2 за счет диффузии, характеризуемой коэффициентом D . Поскольку в стационарных условиях левая часть диффузионного уравнения (4.4.1) обращается в нуль, то

$\partial^2 c / \partial x^2 = 0$. Это обеспечивает линейное распределение растворенных в SiO_2 молекул окислителя, показанное на рис. 4.17. Тогда плотность диффузионного потока равна

$$J_2 = -D \frac{\partial c}{\partial x} = D \frac{c_s - c_1}{x_0}. \quad (4.11.10)$$

Химический поток J_3 следует понимать как потерю молекул окислителя вследствие превращения их в молекулы SiO_2 . Этот поток, в соответствии с основным постулатом химической кинетики (1.18.13), пропорционален концентрации c_1 молекул окислителя на границе SiO_2 – Si , так что для реакции (4.11.1) или (4.11.2) можно записать

$$J_3 = k c_1, \quad (4.11.11)$$

где k — константа скорости реакции окисления кремния.

В стационарном режиме справедливо равенство (4.11.3), которое с учетом (4.11.9)–(4.11.11) принимает вид

$$\bar{\beta}(c_0 - c_s) = \frac{D}{x_0} (c_s - c_1) = k c_1.$$

Решение этой системы уравнений относительно неизвестных c_s и c_1 дает

$$c_s = \frac{(1 + k x_0 / D) c_0}{1 + k x_0 / D + k / \bar{\beta}}, \quad (4.11.12)$$

$$c_1 = \frac{c_0}{1 + k x_0 / D + k / \bar{\beta}}. \quad (4.11.13)$$

Скорость роста слоя SiO_2 определяется плотностью химического потока (4.11.11), а значит и концентрацией c_1 , даваемой формулой (4.11.13):

$$J = k c_1 = \frac{k c_0}{1 + k x_0 / D + k / \bar{\beta}}. \quad (4.11.14)$$

За время dt на единицу площади поверхности кремния поступает $J dt$ молекул окислителя (O_2 или H_2O). В результате этого толщина слоя SiO_2 увеличивается на величину dx_0 , так что

$$J dt = N_0 dx_0. \quad (4.11.15)$$

Здесь N_0 — число молекул окислителя, необходимое для построения единицы объема SiO_2 . Величину N_0 легко рассчитать, зная число молекул SiO_2 в единице объема:

$$N_{\text{SiO}_2} = \frac{\rho_{\text{SiO}_2} \mathcal{N}_A}{M_{\text{SiO}_2}}. \quad (4.11.16)$$

Здесь $\rho_{\text{SiO}_2} = 2,2 \text{ г/см}^3$ — массовая плотность SiO_2 , $M_{\text{SiO}_2} = 60,1 \text{ г/моль}$ — молярная масса SiO_2 , $N_A = 6,022 \cdot 10^{23} \text{ моль}^{-1}$ — число Авогадро. По формуле (4.11.16) находим значение $N_{\text{SiO}_2} \approx 2,2 \cdot 10^{22} \text{ см}^{-3}$, тогда $N_0 = N_{\text{SiO}_2} \approx 2,2 \cdot 10^{22} \text{ см}^{-3}$ для реакции (4.11.1) и $N_0 = 2N_{\text{SiO}_2} \approx 4,4 \cdot 10^{22} \text{ см}^{-3}$ для реакции (4.11.2).

Из выражений (4.11.14) и (4.11.15) следует дифференциальное уравнение для нахождения функции $x_0(t)$:

$$\frac{dx_0}{dt} = \frac{kc_0/N_0}{1 + kx_0/D + k/\bar{\beta}}. \quad (4.11.17)$$

Начальное условие для решения уравнения (4.11.17), означающее отсутствие окисла при $t \leq 0$, имеет вид $x_0(0) = 0$. Интегрирование этого уравнения методом разделения переменных дает следующий результат:

$$x_0^2 + Ax_0 = Bt, \quad (4.11.18)$$

где введены обозначения

$$A = 2D \frac{k + \bar{\beta}}{k\bar{\beta}} \quad \text{и} \quad B = 2D \frac{c_0}{N_0} \equiv 2D \frac{sp_0}{N_0}, \quad (4.11.19)$$

имеющие размерность м и м²/с, соответственно.

Следовательно, для отыскания функции $x_0(t)$ необходимо решить квадратное уравнение (4.11.18). Это решение в качественной форме графически изображено на рис. 4.18. Не записывая здесь общее решение уравнения (4.11.18), проведем анализ двух предельных случаев.

1. Начальная стадия окисления реализуется при такой длительности процесса, когда $x_0(t) \ll A$, и описывается решением уравнения (4.11.18) в форме

$$x_0(t) \approx C_{\text{л}} t. \quad (4.11.20)$$

Этой стадии соответствует так называемый *линейный рост*, поскольку зависимость $x_0(t)$ описывается линейной функцией (4.11.20). Константа линейного роста,

$$C_{\text{л}} \equiv \frac{B}{A} = \frac{k\bar{\beta}}{k + \bar{\beta}} \frac{sp_0}{N_0}, \quad (4.11.21)$$

во-первых, пропорциональна парциальному давлению p_0 окислителя в газовой фазе и коэффициенту растворимости s молекул окислителя в двуокиси кремния и, во-вторых, не содержит коэффициента диффузии D . Это справедливо только для слоев малой толщины, когда диффузия молекул окислителя через тонкий

слой SiO_2 протекает достаточно быстро. В этом случае медленной стадией, лимитирующей рост двуокиси кремния, является химическая реакция, так как обычно $k \ll \bar{\beta}$. Тогда константа линейного роста (4.11.21) принимает форму $C_{\text{л}} \approx k(sp_0/N_0)$, содержащую только константу скорости реакции k .

Это означает, что процесс окисления кремния на начальном этапе протекает в *кинетической области* и скорость роста линейно возрастает с увеличением парциального давления p_0 окислителя в газовой фазе. Кинетической области окисления кремния соответствует область I на рис. 4.18.

Влияние температуры на процесс линейного роста SiO_2 определяется противоборством температурных зависимостей для константы скорости $k(T)$ и коэффициента растворимости $s(T)$. Несмотря на то, что, согласно (2.6.9), при молекулярном механизме растворения величина $s(T)$ уменьшается с увеличением температуры (см. рис. 2.9 б), произведение $k(T)s(T)$, стоящее в числителе константы линейного роста (4.11.21), тем не менее растет. Это происходит из-за преобладающего вклада закона Аррениуса (1.18.17) для константы скорости $k(T)$. Следовательно, с возрастанием температуры всегда увеличивается константа линейного роста $C_{\text{л}}(T) \approx k(T)s(T)p_0/N_0$.

2. Конечная стадия окисления реализуется при такой длительности процесса, когда $x_0(t) \gg A$, и описывается решением уравнения (4.11.18) в форме

$$x_0(t) \approx C_{\text{п}} t^{1/2}. \quad (4.11.22)$$

Этой стадии соответствует так называемый *параболический рост*, поскольку зависимость $x_0(t)$ описывается функцией вида (4.11.22). Константа параболического роста,

$$C_{\text{п}} \equiv B^{1/2} = \sqrt{\frac{2K_{\text{газ}}}{N_0} p_0}, \quad (4.11.23)$$

содержит коэффициент газопроницаемости $K_{\text{газ}} = sD$, введенный равенством (4.6.3), который в данном случае соответствует газопроницаемости двуокиси кремния для молекул окислителя.

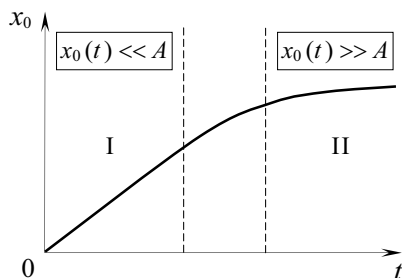


Рис. 4.18. Качественный ход зависимости толщины растущего слоя SiO_2 от времени

Температурная зависимость коэффициента газопроницаемости $K_{\text{газ}}(T)$ определяет влияние температуры на процесс параболического роста двуокиси кремния. Поскольку энергия активации процесса диффузии превосходит теплоту молекулярного растворения, то вклад в зависимость $K_{\text{газ}}(T)$ от закона Аррениуса (1.18.16) для коэффициента диффузии $D(T)$ преобладает над температурной зависимостью (2.6.9) коэффициента растворимости $s(T)$. Следовательно, с увеличением температуры коэффициент газопроницаемости $K_{\text{газ}}(T)$ и константа параболического роста $C_{\text{п}}(T)$ всегда возрастают.

Так как константа параболического роста (4.11.23) не включает константу скорости $k(T)$ для реакции (4.11.1) или (4.11.2), то в этом случае процесс окисления не управляется химической реакцией, протекающей несравненно быстрее, чем диффузия.

Следовательно, на конечном этапе процесс роста двуокиси кремния протекает в *диффузионной области*, т. е. управляется исключительно диффузионным массопереносом, являющимся лимитирующей стадией. Действительно, через толстый слой SiO_2 диффузия молекул окислителя происходит гораздо медленнее по сравнению с химической реакцией. При этом с увеличением парциального давления p_0 константа параболического роста $C_{\text{п}}$ возрастает как $p_0^{1/2}$. Диффузионной области процесса окисления кремния соответствует область II на рис. 4.18.

В промежуточном диффузионно-кинетическом режиме, когда скорости диффузионного массопереноса и химической реакции сравнимы, ход кривой $x_0(t)$ на рис. 4.18 между областями I и II описывается полным решением уравнения (4.11.18).

Выполненный выше анализ не учитывает наличие естественного окисла, всегда присутствующего на поверхности кремния, толщиной около $x_i = 3-5$ нм. Для его учета начальное условие при $t = 0$ должно быть записано в виде $x_0(0) = x_i$. В этом случае изменяется лишь правая часть уравнения (4.11.18), принимая форму $B(t + \tau)$, где $\tau = (x_i^2 + Ax_i)/B$ — временной сдвиг, вызванный наличием естественного окисла SiO_2 . Этот сдвиг проявляется лишь на начальной стадии роста, описываемой уравнением (4.11.20), в правой части которого теперь вместо t стоит $(t + \tau)$.

В заключение следует отметить, что в настоящее время, кроме описанного способа высокотемпературного окисления кремния по реакциям (4.11.1) и (4.11.2), для получения совершенных слоев SiO_2 применяют также низкотемпературное ионно-плазменное окисление в потоке кислорода.

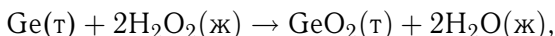
4.12. Кинетика химического травления полупроводников

Процессы химического травления находят широкое применение как в вакуумной, так и в полупроводниковой технологии на различных этапах изготовления дискретных приборов и интегральных микросхем. С их помощью решаются различные технологические задачи, например, такие как:

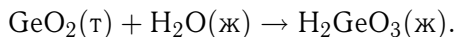
- межоперационная подготовка поверхности и очистка ее от посторонних загрязнений с удалением механически нарушенных поверхностных слоев;
- контролируемое удаление материала с целью локального профилирования поверхности;
- селективное травление с целью выявления неоднородностей структуры и дефектов кристалла (неоднородностей в распределении примесей, выходов дислокаций, границ зерен, блоков, двойников и $p-n$ -переходов).

Применимость химических методов существенно расширяется за счет обратных травлению процессов химического осаждения материалов с целью управления физико-технологическими свойствами поверхности.

По современным воззрениям жидкостное травление рассматривается как процесс окисления поверхности полупроводника с образованием растворимых в травителе продуктов окисления¹⁾. Например, при травлении германия водным раствором перекиси водорода H_2O_2 по реакции



возникающая двуокись германия $GeO_2(t)$ растворяется в воде с образованием метагерманиевой кислоты:



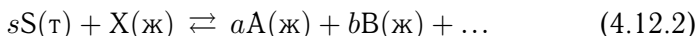
Результирующая реакция окисления германия с образованием жидкофазных продуктов записывается в виде



¹⁾ Наряду с химическим механизмом травления, для объяснения некоторых особенностей поведения полупроводников в травящем растворе, в ряде случаев лучше работает электрохимический механизм. В основе этого механизма лежат сопряженные реакции анодного окисления полупроводника и катодного восстановления окислителя, которые протекают на различных микроанодных и микрокатодных участках поверхности полупроводниковой пластины.

В тех случаях, когда образующийся окисел нерастворим в воде, как в случае SiO_2 , в состав травителя, наряду с окислителем, вводят также растворитель окисла, например фтористоводородную кислоту HF . Она растворяет SiO_2 с образованием жидкофазного соединения $\text{H}_2\text{SiF}_6(\text{ж})$, в результате чего атомы кремния переходят в водный раствор в форме устойчивых комплексных ионов SiF_6^{2-} , возникающих за счет гидролитической диссоциации соединения по реакции $\text{H}_2\text{SiF}_6 \rightleftharpoons 2\text{H}^+ + \text{SiF}_6^{2-}$.

Для выяснения общих закономерностей кинетики жидкостного травления полупроводников, запишем химическую реакцию типа (4.12.1) как обратимую в обобщенной форме (ср. формулу (2.16.1)):



В отличие от (2.16.1) и (4.12.1), эта реакция приведена к одному молю жидкого травителя X (окислителя H_2O_2 в реакции (4.12.1)): в результате травления s молей твердого вещества S дают $a, b, \dots$ молей жидкофазных продуктов $\text{A}, \text{B}, \dots$

Найдем скорость процесса химического травления в рамках упрощенной модели, описываемой реакцией (4.12.2). В соответствии с тем, что было рассмотрено в п. 4.9, процесс травления протекает в три последовательные стадии:

- подвод химического травителя X к поверхности S (*диффузионный поток* $J_{\text{X}}^{\text{диф}}$);
- гетерогенная химическая реакция на поверхности S (*химические потоки* $J_{\text{S}}^{\text{хим}}, J_{\text{X}}^{\text{хим}}, J_{\text{A}}^{\text{хим}}, J_{\text{B}}^{\text{хим}}, \dots$ для всех участников реакции $\text{S}, \text{X}, \text{A}, \text{B}, \dots$);
- отвод продуктов реакции $\text{A}, \text{B}, \dots$ от поверхности S (*диффузионные потоки* $J_{\text{A}}^{\text{диф}}, J_{\text{B}}^{\text{диф}}, \dots$).

Как было отмечено в п. 4.9, относительно быстрые сорбционные процессы на поверхности (адсорбция, поверхностная диффузия, десорбция) включены в стадию гетерогенного химического взаимодействия.

Качественная картина распределения концентрации травителя X и продуктов травления A и B приведена на рис. 4.19, при этом $c_{\text{X}0}, c_{\text{A}0}$ и $c_{\text{B}0}$ обозначают заданные концентрации в объеме приготовленного травящего раствора, в то время как $c_{\text{X}}, c_{\text{A}}$ и c_{B} — искомые концентрации на реакционной поверхности S .

Для каждого жидкофазного участника реакции диффузионный поток представляется в соответствии с формулой (4.9.9) в виде

$$J_i^{\text{диф}} = \beta_i |c_i - c_{i0}|, \quad i = \text{X}, \text{A}, \text{B}, \dots, \quad (4.12.3)$$

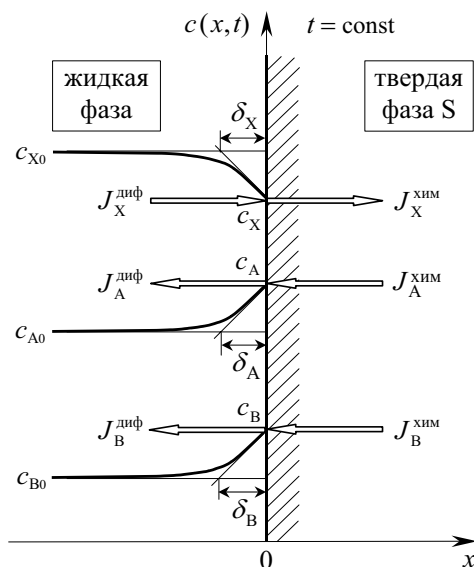


Рис. 4.19. Распределение в жидкой фазе концентраций травящего компонента X и продуктов A и B химической реакции (4.12.2) при травлении твердого вещества S

где $\beta_i = D_i/\delta_i$ — коэффициент диффузионного массопереноса i -го компонента в жидкой фазе.

В стационарном режиме травления, как уже было отмечено формулой (4.9.10), диффузионный поток должен быть равен химическому потоку для каждого компонента. Тогда с использованием выражения (4.12.3) можно записать следующие равенства:

$$\begin{aligned} J_X &\equiv J_X^{\text{хим}} = J_X^{\text{диф}} = \beta_X(c_{X0} - c_X), \\ J_A &\equiv J_A^{\text{хим}} = J_A^{\text{диф}} = \beta_A(c_A - c_{A0}), \\ J_B &\equiv J_B^{\text{хим}} = J_B^{\text{диф}} = \beta_B(c_B - c_{B0}). \end{aligned} \quad (4.12.4)$$

В отличие от п. 4.9, диффузионные и химические потоки, введенные равенствами (4.12.3) и (4.12.4), являются положительными, так как их направления учтены стрелками на рис. 4.19.

Условие стехиометрии, записанное для реакции (4.12.2) по аналогии с выражением (4.9.2), устанавливает следующее соотношение между химическими потоками:

$$J_X^{\text{хим}} = \frac{J_A^{\text{хим}}}{a} = \frac{J_B^{\text{хим}}}{b} = \dots = \frac{J_S^{\text{хим}}}{s}. \quad (4.12.5)$$

В соответствии с основным постулатом химической кинетики для реакции травления (4.12.2) по аналогии с (4.9.6) поток равен

$$J_X^{\text{хим}} = k_{\text{тр}} c_X, \quad (4.12.6)$$

где $k_{\text{тр}}$ — константа скорости травления, определенная для обратимой реакции в форме (4.9.6), включающей пересыщение травления $|\Delta G_{\text{тр}}|$.

Подстановкой (4.12.6) в первое равенство (4.12.4) получаем концентрацию травителя X на реакционной поверхности S:

$$c_X = \frac{\beta_X}{\beta_X + k_{\text{тр}}} c_{X0}. \quad (4.12.7)$$

Формулы (4.12.5)–(4.12.7) дают окончательное выражение для плотности потока травления твердого вещества S:

$$J_S^{\text{хим}} = s \frac{\beta_X k_{\text{тр}}}{\beta_X + k_{\text{тр}}} c_{X0} \equiv k_{\text{эф}} c_{X0}, \quad (4.12.8)$$

где введена эффективная константа травления $k_{\text{эф}}$, такая что

$$\frac{1}{k_{\text{эф}}} = \frac{1}{s k_{\text{тр}}} + \frac{1}{s \beta_X}. \quad (4.12.9)$$

Чтобы найти «геометрическую» скорость травления, определяемую как $v_{\text{тр}} = dx/dt$ (мкм/мин), надо использовать соотношение, аналогичное равенству (4.11.15):

$$J_S^{\text{хим}} = N_S \frac{dx}{dt} \equiv N_S v_{\text{тр}}, \quad (4.12.10)$$

где $N_S = \rho_S \mathcal{N}_A / M_S$ — число молекул в единице объема твердого вещества S (см. формулу (4.11.16)). В частности, для кремния $\rho_{\text{Si}} = 2,33 \text{ г/см}^3$ и $M_{\text{Si}} = 28,08 \text{ г/моль}$, тогда $N_{\text{Si}} = 5 \cdot 10^{22} \text{ см}^{-3}$. Из формул (4.12.8) и (4.12.10) получаем скорость травления

$$v_{\text{тр}} \equiv \frac{1}{N_S} J_S^{\text{хим}} = \frac{1}{N_S} k_{\text{эф}} c_{X0}. \quad (4.12.11)$$

Возвращаясь к соотношению (4.12.8), удобно воспользоваться аналогией с электрическими цепями, записав его с помощью (4.12.9) в виде

$$J_S \equiv J_S^{\text{хим}} = \frac{c_{X0}}{R_{\text{диф}} + R_{\text{хим}}}. \quad (4.12.12)$$

Подобно току в электрической цепи, протекающему через сопротивление под действием электродвижущей силы, поток, $J_S = s J_X$, вещества S создается заданной концентрацией травителя

c_{X0} в объеме жидкой фазы, которая играет роль «движущей силы» процесса травления. Согласно выражению (4.12.12), система оказывает сопротивление протеканию потока J_S , складывающееся из *диффузионного сопротивления* $R_{\text{диф}} = 1/s\beta_X$ и *химического сопротивления* $R_{\text{хим}} = 1/sk_{\text{тр}}$. В зависимости от соотношения между этими сопротивлениями различают диффузионную и кинетическую области процесса травления.

1. Диффузионная область травления осуществляется при $R_{\text{хим}} \ll R_{\text{диф}}$, или $\beta_X \ll k_{\text{тр}}$. В этом случае из двух последовательных стадий процесса — диффузионного массопереноса и собственно химической реакции — наиболее медленной (лимитирующей) стадией является диффузия. Именно она управляет процессом травления в целом, при этом $k_{\text{эф}} \approx s\beta_X$ и $v_{\text{тр}} \approx (s\beta_X/N_S)c_{X0}$. Следовательно, скорость травления ограничивается скоростью диффузионной доставки молекул травителя X к реакционной поверхности S, т. е. процесс травления протекает в диффузионной области. Поскольку диффузионная константа β_X для жидкой фазы является изотропной, то все кристаллографические грани и дефекты кристалла травятся с одинаковой скоростью, определяемой диффузионной доставкой компонента X к поверхности полупроводника. Независимость от кристаллографической ориентации и дефектности границы обеспечивает достаточно гладкую зеркально-полированную поверхность пластины.

Иными словами, для диффузионной области характерен режим *полирующего травления* поверхности. Увеличение полирующих свойств травителя может быть достигнуто за счет усиления неравенства $\beta_X \equiv D_X/\delta_X \ll k_{\text{тр}}$. Следовательно, введение вязких добавок в травящий раствор и понижение температуры должно улучшать полирующее действие в результате снижения коэффициента диффузии D_X . В то же время интенсивное перемешивание раствора может оказать обратное действие в результате уменьшения эффективной толщины диффузионного слоя δ_X .

2. Кинетическая область травления осуществляется при $R_{\text{диф}} \ll R_{\text{хим}}$, или $k_{\text{тр}} \ll \beta_X$. В этом случае из двух стадий процесса наиболее медленной (лимитирующей) стадией является собственно химическая реакция, при этом $k_{\text{эф}} \approx sk_{\text{тр}}$ и $v_{\text{тр}} \approx (sk_{\text{тр}}/N_S)c_{X0}$. Следовательно, процесс травления протекает в кинетической области, так как его скорость ограничивается скоростью химической реакции, а быстрая диффузионная доставка молекул травителя X обеспечивает практически постоянную их концентрацию на поверхности S, равную c_{X0} . Поскольку константа травления $k_{\text{тр}}$ чувствительна к дефектности поверхности

и имеет различные значения для разных граней кристалла, то для кинетической области процесса характерно *селективное и анизотропное травление*.

Селективность травления проявляется в преимущественном вытравливании дефектных мест на поверхности в форме ямок травления (например, в местах выхода дислокаций). В районе этих мест атомы находятся в возбужденном состоянии, что понижает активационный барьер для константы $k_{тр}$ и увеличивает локальную скорость травления.

Анизотропность травления связана с различной скоростью травления на разных кристаллографических гранях, характеризуемых разными индексами Миллера. Обычно (например, для кремния Si и арсенида галлия GaAs) выполняются неравенства $v_{\{100\}} > v_{\{110\}} > v_{\{111\}}$. В результате этого даже на бездефектных поверхностях, ориентированных по совокупности плоскостей $\{100\}$ или $\{110\}$, при локальном травлении через специальные окна в защитной маске можно получить профилированные углубления, ограниченные плоскостями, которые характеризуются малыми скоростями травления.

Подбором химического состава специально создают полирующие или селективные травители для полупроводников. Нередко применяют *химико-динамическое травление* путем перемещения травителя по поверхности пластины при помощи механического вращения сосуда с жидкостью. Фактор движения травителя не влияет на константу травления $k_{тр}$, но изменяет диффузионную константу $\beta_X \equiv D_X/\delta_X$ из-за изменения эффективной толщины диффузионного слоя δ_X .

Следовательно, подбором температуры, скорости вращения и вязкости раствора можно обеспечить необходимый тип и качество полировки пластин. Более того, при достаточно высокой скорости вращения, когда разрушается диффузионный слой (так что $\delta_X \rightarrow 0$), принципиально возможно обеспечить неравенство $\beta_X \gg k_{тр}$. Это позволяет перевести процесс травления из диффузионной области в кинетическую, осуществляя переход от полирующего режима травления к селективному.

В заключение отметим, что в настоящее время, наряду с жидкостным травлением, широкое распространение получили методы так называемого «сухого травления», основанные на использовании реактивных парогазовых смесей и ионно-плазменных источников. Парогазовое травление обеспечивает высокое качество и чистоту поверхности и легко совмещается с другими процессами планарно-эпитаксиальной технологии. В дополнение к этому, ионно-плазменное травление обладает более высокой

разрешающей способностью и пригодно для изготовления элементов субмикронных размеров. Естественно, что рассмотренные здесь основные закономерности кинетики жидкостного химического травления в общих чертах применимы и к травлению с применением газофазных реакций.

4.13. Принципы выращивания монокристаллических слоев методами эпитаксии

Планарно-эпитаксиальная технология составляет основу производства современных интегральных микросхем разнообразного функционального назначения. Эпитаксиальные слои полупроводников в структурном отношении превосходят объемные монокристаллы и обладают более высокими электрофизическими параметрами. Это вызвано существенно меньшим содержанием неконтролируемых загрязнений при практически идеальной однородности распределения легирующих примесей. Именно благодаря эпитаксиальной технологии за последние годы резко повысилась надежность микроэлектронных изделий и возрос процент выхода годных вплоть до 98%. Развитие новых эпитаксиальных методов и технологий, в частности молекулярно-лучевой эпитаксии, позволило создать многослойные гетероструктуры и сверхрешетки с управляемыми геометрическими параметрами и уникальными физическими свойствами.

Эпитаксиальные процессы, применяемые в современной твердотельной технологии, можно разделить на три большие группы в соответствии с агрегатным состоянием исходной (питающей) фазы — процессы жидкофазной эпитаксии (ЖФЭ), процессы газофазной эпитаксии (ГФЭ) и процессы молекулярно-лучевой эпитаксии (МЛЭ). Кратко рассмотрим особенности практической реализации этих технологических процессов.

1. Процессы жидкофазной эпитаксии осуществляют путем кристаллизации на поверхности подложки монокристаллического вещества из жидкой фазы, представляющей собой либо расплав этого вещества, либо его раствор (или раствор-расплав) в другом жидком (или расплавленном) веществе как растворителе. При этом роль растворителя может выполнять либо один из компонентов твердого соединения (например, раствор арсенида галлия GaAs в жидком галлии Ga), либо постороннее вещество (например, раствор железо-иттриевого граната $Y_3Fe_5O_{12}$ — перспективного материала ферритовой техники — в расплаве окиси свинца PbO).

Понятно, что в идейном плане методы ЖФЭ мало чем отличаются от выращивания объемных монокристаллов из жидкой фазы — методов направленной кристаллизации и зонной перекристаллизации. Термодинамическую основу этих методов составляют диаграммы плавкости, рассмотренные в гл. 2.

Выращивание из расплавов применимо только для элементарных веществ и химических соединений, плавящихся конгруэнтно. Пересыщение в системе достигается принудительным охлаждением расплава ниже температуры плавления $T_{\text{пл}}$, что вызывает осаждение полупроводника на подложке с выделением теплоты кристаллизации. Если не предпринять меры по отводу этой теплоты, то повышение температуры может прекратить кристаллизацию при достижении $T_{\text{пл}}$. Для элементарных полупроводников (например, для кремния с $T_{\text{пл}} = 1417^\circ\text{C}$) кривые принудительного охлаждения аналогичны изображенным под номерами 1 и 5 на рис. 2.14 б. Различие между методами выращивания монокристаллов из расплава состоит главным образом в способе отвода теплоты кристаллизации.

Выращивание из растворов-расплавов позволяет выбором растворителя и состава раствора вблизи линии ликвидуса проводить кристаллизацию вещества при температурах на сотни градусов ниже его температуры плавления. Это применимо, в том числе, и к сложным (бинарным и тройным) соединениям с инконгруэнтным плавлением. Понижение температуры кристаллизации снижает концентрацию структурных дефектов (тепловых вакансий) и уменьшает опасность загрязнения кристалла побочными примесями, для большинства которых коэффициент распределения (2.7.5) уменьшается с понижением температуры. Преимущества кристаллизации из нестехиометрических растворов-расплавов особенно ярко проявляются при выращивании полупроводниковых соединений с высокими температурами плавления и разлагающихся соединений со значительным давлением паров летучих компонентов, входящих в состав кристалла. В последнем случае в процессе роста надо обеспечивать подпитку раствора летучим компонентом из парогазовой фазы.

В бинарной системе А–В при кристаллизации соединений из нестехиометрических растворов-расплавов пересыщение может быть достигнуто тремя способами:

- путем *принудительного охлаждения*, осуществляемого в неизотермических условиях (как и при кристаллизации из расплава): $T = \text{var}$ при неизменном составе раствора ($c_{\text{В}}^{\text{ж}} = \text{const}$), соответствующем исходной жидкой фазе;

- путем *подпитки раствора-расплава из газовой фазы*, осуществляемой в изотермических условиях: $T = \text{const}$ при изменяющемся составе раствора ($c_B^{\text{ж}} = \text{var}$);
- путем *подпитки раствора-расплава из твердой фазы*, осуществляемой в условиях температурного градиента.

Рассмотрим особенности трех способов пересыщения жидкой фазы на примере бинарной системы Ga–As, фазовая диаграмма которой ранее была приведена на рис. 2.22 а вместе с кривыми принудительного охлаждения (рис. 2.22 б) и неоднородного распределения температуры (рис. 2.22 в).

Принудительное охлаждение изначально насыщенного мышьяком расплава галлия в точке 1 на рис. 2.22 а, осуществляется в соответствии с кривой охлаждения, изображенной на рис. 2.22 б. Согласно правилу фаз (1.15.3), для бинарной системы при $P = \text{const}$ число степеней свободы равно $C = 3 - \Phi$. Поэтому с понижением температуры при достижении линии ликвидуса, когда начинается кристаллизация GaAs, система становится трехфазной ($\text{ж} + \text{As}^{\text{п}} + \text{GaAs}$) и неинвариантной ($C = 0$). Такому состоянию отвечает постоянная температура T_1 на кривой охлаждения в форме площадки неинвариантности (при условии отвода теплоты кристаллизации). При этом в процессе кристаллизации исходный состав раствора, соответствующий точке 1 на линии ликвидуса, остается неизменным.

Подпитка раствора-расплава из газовой фазы осуществляется для исходного расплава $\text{Ga}^{\text{ж}}$, насыщенного мышьяком, что соответствует верхней точке 1' на рис. 2.22 а. При охлаждении до температуры T_1 раствор-расплавы будет ненасыщенным, поскольку нижняя точка 1' не совпадает с линией ликвидуса. Сохраняя неизменной температуру $T_1 = \text{const}$, можно повышением давления паров мышьяка $\text{As}^{\text{п}}$ (путем нагрева его источника, расположенного в специальной температурной зоне) насыщать $\text{Ga}^{\text{ж}}$ мышьяком, т. е. смещать точку 1' вправо до совпадения ее с линией ликвидуса в точке 1. С этого момента начинается изотермическая кристаллизация арсенида галлия.

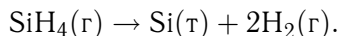
Подпитка раствора-расплава из твердой фазы реализуется в условиях существования температурного градиента в жидкой фазе, расположенной между двумя пластинами одного и того же вещества, в данном случае GaAs. Профиль распределения температуры показан на рис. 2.22 в, где перепад $\Delta T = T_{\text{max}} - T_{\text{min}}$ обеспечивает температурный градиент $\Delta T / \Delta x$, на практике равный $(10 - 15)^\circ\text{C}/\text{см}$. Пластина GaAs с температурой T_{max} , растворяясь, служит источником подпитки мышьяком жидкой

фазы. Другая пластина с температурой $T_{\min}$ является подложкой, на которой кристаллизуется GaAs из раствора-расплава, заполняющего зазор между пластинами. Действительно, температуре $T_{\max}$ соответствует точка $1_{\max}$ на линии ликвидуса (рис. 2.22 а), в которой раствор более обогащен мышьяком по сравнению с точкой $1_{\min}$ при температуре $T_{\min}$. Избыточный мышьяк диффундирует через слой расплава галлия из высокотемпературной зоны с $T_{\max}$ в низкотемпературную зону с $T_{\min}$, что показано стрелками в верхней части рис. 2.22 в. Это создает на поверхностях пластин GaAs пересыщения, противоположные по знаку и обеспечивающие растворение источника при $T_{\max}$ и осаждение GaAs на подложке при $T_{\min}$. Изложенная идея составляет основу градиентной зонной перекристаллизации слитков, а также применяется при выращивании эпитаксиальных слоев методом жидкофазной эпитаксии.

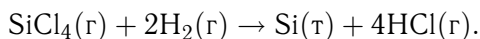
2. Процессы газофазной эпитаксии осуществляют путем кристаллизации на поверхности подложки монокристаллического вещества из парогазовой фазы. Эта фаза содержит атомы или молекулы, идентичные по составу осаждаемому веществу, либо имеющие более сложный состав, включающий компоненты кристаллизуемого материала. В этой связи все используемые методы ГФЭ могут быть отнесены к одной из следующих групп.

Метод диссоциации и восстановления химических соединений в газовой фазе связан с нарушением условий гетерогенного химического равновесия, в результате чего на подложке осаждается твердотельный продукт реакции. Источником материала для роста кристаллов обычно служат легколетучие химические соединения (чаще гидриды и галогениды). Эти соединения на реакционной поверхности подвергаются термической диссоциации или восстановлению газообразным восстановителем, которым чаще всего является водород. В качестве примера использования таких реакций можно привести силановый и хлоридный методы эпитаксии кремния, осуществляемые в проточных системах.

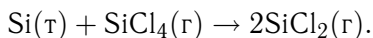
Силановый метод основан на реакции термического разложения моносилана кремния SiH_4 , подаваемого в смеси с водородом в реакционную камеру, где при температурах 950–1100°C протекает реакция пиролиза:



Хлоридный метод использует гетерогенную реакцию водородного восстановления тетрахлорида кремния SiCl_4 :



В реальных условиях при температурах 1150–1250°C эта реакция сопровождается аналогичными газофазными реакциями, в правой части которых вместо $\text{Si}(\text{г})$ стоят газообразные продукты — дихлорид SiCl_2 и трихлорсилан SiHCl_3 . Это существенно осложняет термодинамический и кинетический анализ процесса. В частности, наряду с основной реакцией осаждения кремния, надо также учитывать конкурирующую реакцию его газового травления:

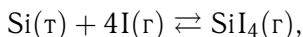


Как показывают эксперименты и кинетический анализ, конкурирующая реакция начинает заметно преобладать при достаточно большом количестве вводимого в систему тетрахлорида кремния. Тогда скорость осаждения, пройдя через максимум, начинает уменьшаться и изменяет знак, т. е. кристаллизация кремния сменяется его травлением.

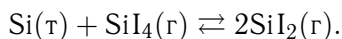
Метод газотранспортных химических реакций использует химический транспорт труднолетучих компонентов, которые с помощью обратимой гетерогенной реакции в зоне источника переводятся из твердой фазы в газовую, а затем обратно в твердую фазу на поверхности подложки. Термодинамика ГХР, представленной в обобщенной форме (2.16.1), изучена в п. 2.16. Химические транспортные реакции осуществляют разными способами в зависимости от выбора типа реакционной системы, например,

- закрытая система типа запаянной кварцевой ампулы,
- проточная система типа открытой кварцевой трубы (см. ниже рис. 4.23),
- сэндвич-система типа параллельных пластин с малым (не более 1 мм) зазором (см. ниже рис. 4.24).

Примером применения ГХР в запаянной ампуле является йодидный метод эпитаксии кремния, в котором транспортным агентом являются пары йода $\text{I}(\text{г})$ или его газообразное соединение с кремнием — тетрайодид кремния $\text{SiI}_4(\text{г})$. Как показывает термодинамический анализ, при низких давлениях паров йода в ампуле (ниже 1 Торр) химический транспорт кремния осуществляется с помощью реакции



в то время как при высоких давлениях (выше 10 Торр) преобладает реакция диспропорционирования:



Так как первая реакция экзотермическая ($\Delta H < 0$), а вторая эндотермическая ($\Delta H > 0$), то при одной и той же температуре

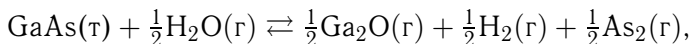
зоны источника ($T_{\text{и}} = 1100^\circ\text{C}$), содержащей поликристаллический кремний, температура подложки для этих реакций различна: $T_{\text{п}} \approx T_{\text{и}} \pm 150^\circ\text{C}$ для первой и второй реакции.

Примером применения ГХР в открытой проточной камере является хлоридный метод эпитаксии арсенида галлия в системе $\text{Ga}-\text{AsCl}_3-\text{H}_2$. В реакционную камеру подают газ-носитель H_2 , насыщенный парами AsCl_3 . Взаимодействие между этими газами во входной зоне при $T_{\text{вх}} = 400-450^\circ\text{C}$ дает газообразные продукты $\text{HCl}(\text{г})$ и $\text{As}_4(\text{г})$, которые поступают в зону источника, содержащую лодочку с расплавленным галлием $\text{Ga}(\text{ж})$. Здесь при $T_{\text{и}} = 800-900^\circ\text{C}$ газообразный мышьяк $\text{As}_4(\text{г})$ насыщает расплав $\text{Ga}(\text{ж})$ с образованием на его поверхности твердого слоя GaAs . Хлористый водород, образовавшийся во входной зоне, выполняет функцию транспортного агента X в обобщенной ГХР (2.16.1). В данном случае транспортная химическая реакция имеет вид



В зоне источника при $T_{\text{и}} = 800-900^\circ\text{C}$ эта реакция смещена вправо в сторону образования газообразных продуктов. Поскольку реакция эндотермическая ($\Delta H > 0$, см. рис. 2.23), то кристаллизация арсенида галлия происходит при более низкой температуре подложки ($T_{\text{п}} \approx T_{\text{и}} - 50^\circ\text{C}$), на поверхности которой реакция смещена влево.

Сэндвич-метод, называемый иногда методом близкого переноса, используется для изучения химического транспорта газообразных веществ в лабораторных условиях. В частности, он применяется в системе $\text{GaAs}-\text{H}_2\text{O}-\text{H}_2$ для переноса арсенида галлия парами воды. В этом случае роль газотранспортной химической реакции играет реакция

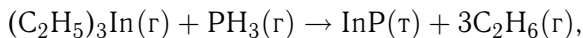
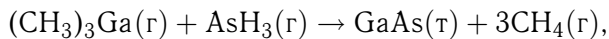


с образованием газообразного окисла галлия $\text{Ga}_2\text{O}(\text{г})$. Как и в хлоридном методе, водород выполняет функцию газа-носителя, т. е. имеет наибольшее парциальное давление $p_{\text{H}_2} \approx 1$ атм. Термодинамический анализ этой системы с учетом конкурирующих реакций образования жидкого галлия $\text{Ga}(\text{ж})$ и твердого окисла $\text{Ga}_2\text{O}_3(\text{т})$, препятствующих процессу кристаллизации GaAs , выполнен в приложении А.2.

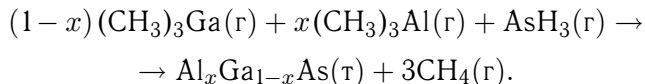
Метод МОС-гидридной эпитаксии, называемый также методом МОС ГФЭ (MOVPE, сокращение от англ. *Metal Organic Vapor Phase Epitaxy*), основан на взаимодействии металлоорганических соединений (МОС) с гидридами элементов, такими как

газы — арсин AsH_3 , фосфин PH_3 и др. МОС относятся к классу веществ (обычно в жидком состоянии при комнатной температуре), имеющих химические связи типа металл–углерод, например метиловые или этиловые соединения, такие как триметилгаллий $(\text{CH}_3)_3\text{Ga}$, триметилалюминий $(\text{CH}_3)_3\text{Al}$, триэтилиндий $(\text{C}_2\text{H}_5)_3\text{In}$, тетраэтилсвинец $(\text{C}_2\text{H}_5)_4\text{Pb}$ и др. (см. с. 609).

При температурах $600\text{--}800^\circ\text{C}$ взаимодействие МОС с гидридами протекает по химическим реакциям,



с образованием полупроводникового соединения (арсенида галлия GaAs или фосфида индия InP) и газообразного продукта (метана CH_4 или этана C_2H_6). Комбинацией исходных парогазовых компонентов синтезируют тройные и четверные соединения, например, твердый раствор $\text{Al}_x\text{Ga}_{1-x}\text{As}$ по реакции



Обычно реакцию проводят в проточной системе, на вход которой подают парогазовую смесь МОС, гидрида и водорода как газа-носителя. Кристаллизация происходит на поверхности нагретой подложки в условиях, достаточно близких к термодинамическому равновесию. Понижение давления и увеличение скорости газового потока позволяет получать более однородные пленки с контролем толщины до нескольких ангстрем. Добавлением в парогазовую фазу легколетучих примесей осуществляют легирование эпитаксиальных слоев.

Методом МОС ГФЭ создают сверхрешетки и многослойные гетероструктуры с толщиной слоев в несколько межатомных расстояний и резкими гетерограницами, используемые в полупроводниковых лазерах на квантовых ямах, нитях и точках.

Методы химической сборки поверхностных наноструктур относятся к циклично-дискретным процессам. Они основаны на многократном чередовании моноатомных структурных единиц, последовательно формирующих химическое (чаще бинарное) соединение путем хемосорбции компонентов из газовой фазы или из атомно-молекулярных пучков. При этом толщина создаваемого слоя определяется не временем синтеза или интенсивностью потоков, а только числом повторяющихся реакционных циклов.

Малая энергия активации поверхностных реакций позволяет проводить химическую наносборку поверхности при достаточно

низкой температуре (обычно не превышающей 400°C). Это снижает влияние диффузионных процессов и обеспечивает резкие гетерограницы в многослойных структурах.

На практике для формирования поверхностных наноструктур применяют две разновидности химической сборки:

- метод молекулярного наслаивания,
- метод атомно-слоевой эпитаксии.

Молекулярное наслаивание осуществляют с обязательным участием *функциональных групп*, в которых один из элементов является химической составляющей синтезируемого соединения. Для синтеза оксидных слоев используют OH-группы, нитридных слоев — NH-группы, сульфидных слоев — SH-группы, создаваемые путем обработки поверхности соответствующими газами — H₂O, NH₃, H₂S. Вторая составляющая бинарного соединения входит в состав молекулы *прекурсора* (от англ. *precursor* «предшественник»); так называют химическое вещество, участвующее в промежуточной реакции при синтезе другого вещества.

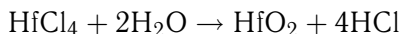
Химическая сборка состоит из ряда последовательных химических превращений с участием только функциональных групп, не затрагивающих поверхностные атомы подложки. Целенаправленное формирование нужного набора функциональных групп на подложке дает возможность наращивать новый слой, связанный с подложкой прочной химической связью. Это достигается созданием условий для взаимодействия функциональных групп с молекулами выбранного прекурсора. С помощью такого взаимодействия поверхность достраивается монослоем новых структурных единиц, которые совместно с соответствующим элементом функциональной группы (O, N, S) создают синтезируемое соединение (оксид, нитрид, сульфид).

Отличительным признаком молекулярного наслаивания является саморегуляция процесса, выражающаяся в остановке монослойного роста после завершения одного цикла процесса синтеза. Осуществляя необходимое число циклов, можно синтезировать наноструктуры заданной толщины с точностью до одного монослоя. Использование на разных стадиях различных прекурсоров позволяет наносить на поверхность несколько монослоев структурных единиц в заданной последовательности со структурными единицами другой химической природы. Это обеспечивает расположение по заданной программе химически связанных между собой слоев различных атомов.

Для синтеза оксидных наноструктур в качестве прекурсора обычно применяют хлориды металлов (Ti, Ta, Hf, Cr, Al и др.),

молекулы которых на гидроксيليрированной (покрытой группами OH) поверхности нагретой подложки взаимодействуют с OH.

В качестве примера рассмотрим формирование диоксида гафния HfO_2 , используемого в качестве подзатворного диэлектрика МОП-транзисторов в современных процессорах, выполненных по технологии High-K/Metal-Gate (см. с. 39). На рис. 4.20 схематически изображены два цикла процесса молекулярного насаживания. Здесь роль прекурсора выполняет хлорид гафния HfCl_4 , взаимодействие которого с парами воды по реакции



дает молекулы HfO_2 на поверхности кремния без участия последнего в химической реакции.

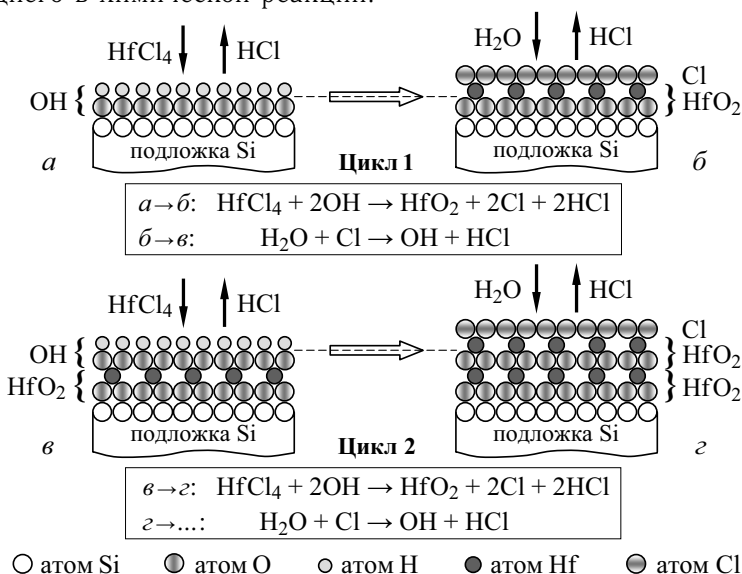


Рис. 4.20. Последовательность реакций при химической сборке диоксида гафния на поверхности кремния методом молекулярного насаживания

Первый цикл начинается с подачи на предварительно гидроксيليрированную кремниевую поверхность хлорида HfCl_4 , молекулы которого, взаимодействуя с функциональными группами OH (рис. 4.20 а) по реакции $a \rightarrow b$, заменяют атомы водорода атомами гафния (по стрелке вдоль штриховой линии). Это создает первый монослой диоксида гафния HfO_2 , который сверху покрыт химически активными атомами хлора (рис. 4.20 б). Тем самым поверхность становится подготовленной к последующей операции гидроксيليрирования. С этой целью в реактор напускают молекулы

воды, которые, взаимодействуя с атомарным хлором по реакции $b \rightarrow v$, заменяют атомы хлора гидроксильными группами. Этим завершается первый цикл и возникает структура, показанная на рис. 4.20 в, которая содержит на поверхности функциональные группы OH, как и исходная структура на рис. 4.20 а.

Второй цикл, изображенный на рис. 4.20 в и г, в точности повторяет все последовательные операции первого цикла и завершается формированием второго монослоя HfO_2 , покрытого атомами хлора, с последующим гидроксилированием для перехода к третьему циклу (не показанному на рис. 4.20). В ходе каждой реакции возникают молекулы HCl , которые следует быстро удалять во избежание нежелательного обратного травления.

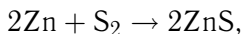
Многократным повторением циклов осуществляют синтез наноструктур необходимой толщины. Привлекательность метода молекулярного наслаивания состоит в том, что простой сменой прекурсора можно, не прерывая процесс синтеза, по желанию создавать любые гетероструктуры с нанослоями моноатомной толщины и резкими границами по составу и легированию.

Атомно-слоевая эпитаксия осуществляется за счет реакций присоединения элементов к координационно ненасыщенным атомам поверхности. Этот метод по своей сути существенно проще метода молекулярного наслаивания, поскольку не требует создания функциональных групп на поверхности подложки.

При атомно-слоевой эпитаксии (АСЭ) поверхностная химическая сборка бинарного соединения АВ происходит в результате послынной хемосорбции компонентов А и В из атомно-молекулярных пучков, получаемых испарением соответствующих источников в техническом вакууме. Цикличность процесса обеспечивается поочередным открыванием испарительных ячеек А и В. Толщина каждого слоя, как и при молекулярном наслаивании, определяется только числом циклов обработки поверхности потоками компонентов и также контролируется с точностью до одного монослоя.

Образование химического соединения АВ и формирование пленки происходят за счет реакции между компонентами А и В в адсорбционных слоях. Поэтому важно, чтобы хемосорбция было мономолекулярной и однослойной, что определяется выбором температуры подложки, обеспечивающей рост одного монослоя за один цикл химической сборки методом АСЭ.

В качестве примера рассмотрим формирование слоев сульфида цинка ZnS в соответствии с результирующей химической реакцией,



циклично протекающей на поверхности подложки путем поочередной ее обработки парами серы и цинка, как изображено на рис. 4.21. Здесь, как и ранее на рис. 4.20, показаны только два последовательных цикла, формирующих два монослоя ZnS.

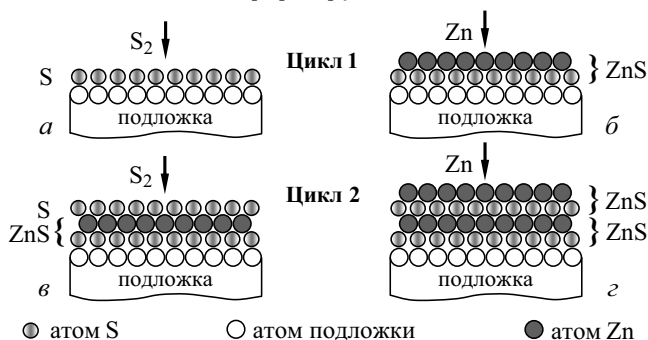


Рис. 4.21. Последовательность реакций при химической сборке сульфида цинка методом атомно-слоевой эпитаксии

Общим недостатком всех методов химической сборки в настоящее время являются ограниченный набор достаточно эффективных летучих реагентов и сложность создания однородно активированных поверхностей для реализации монослойного механизма роста наноструктур.

3. Молекулярно-лучевая эпитаксия (МЛЭ) представляет собой аппаратное усовершенствование метода атомно-слоевой эпитаксии (АСЭ), изложенного выше. Главное отличие состоит в использовании при МЛЭ сверхвысокого вакуума (с давлением остаточных газов не более 10^{-9} мм рт. ст.)

В основе обоих методов лежит обычный способ конденсации вещества (в молекулярной или атомарной форме), получаемого термическим испарением в вакууме из соответствующих источников. Изначально термовакuumное испарение применяли для напыления металлических пленок в обычном техническом вакууме. Лишь после разработки соответствующего вакуумного оборудования в 70-х годах прошлого века начался новый этап в использовании

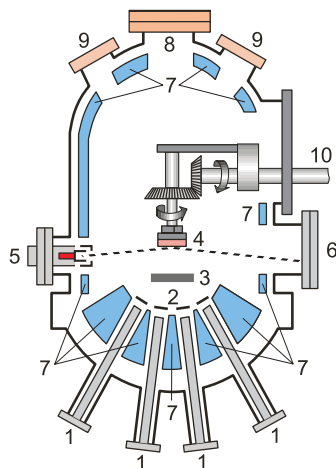


Рис. 4.22. Схема технологической камеры для МЛЭ

вакуумной конденсации из молекулярно-атомарных пучков. Это заложило основы современной МЛЭ для выращивания монокристаллических пленок полупроводниковых соединений.

Технологическая камера установки МЛЭ схематически изображена на рис. 4.22, где: 1 — эффузионные испарители с заслонками 2; 3 — основная заслонка; 4 — подложка на вращающемся держателе; 5 — электронная пушка дифрактометра с флуоресцентным экраном 6 (пунктиром показан ход электронного луча); 7 — экраны с азотным охлаждением; 8 — вакуумный шлюз для смены образцов; 9 — смотровые окна; 10 — ось двигателя для вращения подложки 4 с подачей питания на ее нагреватель.

Кроме использования сверхвысоковакуумной откачной системы на основе современных ионных и криогенных насосов, для уменьшения давления остаточных газов вся свободная поверхность камеры окружена экранами 7, охлаждаемыми жидким азотом. В дополнение к этому, в установках МЛЭ используют сверхчистые источники испаряемых материалов, точный контроль температур подложки 4 и испарителей 1, а также различные методы диагностики *in situ* (лат. «в месте нахождения») в сочетании с компьютерным управлением процессами роста и диагностики. В ходе МЛЭ осуществляют контроль как состава газовой фазы (методом масс-спектрометрии), так и параметров растущего слоя, таких как толщина (методом эллипсометрии), химический состав (методом оже-спектроскопии) и кристаллическая структура (методом дифракции быстрых или медленных электронов — см. позиции 5 и 6 на рис. 4.22),

Для обеспечения однородности состава и структуры по поверхности растущей пленки предусмотрена возможность вращения держателя подложки. Наличие вакуумного шлюза позволяет использовать разные подложки без разгерметизации вакуумной камеры, а наблюдение за процессом осуществляют через смотровые окна (см. позиции 8, 9 и 10).

Атомарные или молекулярные пучки создаются при помощи эффузионных испарителей 1 (называемых также ячейками Кнудсена), внутрь которых раздельно помещены испаряемые вещества — компоненты соединения и элементы легирующих примесей. Индивидуальный нагрев каждой ячейки расплавляет содержащееся в ней вещество и создает достаточно высокое давление насыщенного пара, атомы/молекулы которого истекают в вакуум через малое отверстие (такой процесс называют *эффузией*). Образующиеся таким образом атомарные/молекулярные пучки при открытых заслонках 2 и 3 поступают на поверхность подложки 4 со скоростями, существенно превышающими скорость,

соответствующую равновесному давлению пара при температуре подложки (более низкой по сравнению с температурами ячеек).

Управление составом эпитаксиального слоя и скоростью его роста достигается регулировкой моментов открывания/закрывания заслонок и температурами испарителей и подложки. Баллистический транспорт осаждаемых компонентов в сверхвысоком вакууме при достаточно малой скорости их поступления на поверхность подложки (около $10^{14} - 10^{15}$ атомов в секунду) обеспечивает эпитаксиальный рост. В этом случае имеет место послойное заполнение растущей поверхности с исключительно точным управлением профилями как химического состава, так и легирования. В основе этого лежат следующие элементарные процессы:

- адсорбция падающих на подложку частиц,
- термическая десорбция частиц с подложки,
- поверхностная диффузия адсорбированных атомов,
- встраивание адсорбированных атомов в моноатомный слой,
- образование и дальнейший рост поверхностных (двумерных) зародышей кристалла.

Ионно-ковалентный характер и большая энергия химических связей в полупроводниковых соединениях типа $A^{III}B^V$ и $A^{II}B^{VI}$ приводят к тому, что за время формирования одного монослоя (порядка одной секунды) адсорбированные атомы А и В в результате поверхностной диффузии совершают несколько тысяч диффузионных прыжков. Это продолжается до тех пор, пока они не займут подходящее место в кристаллической решетке, то есть происходит своеобразная самоорганизация растущей структуры.

Существуют две модификации стандартной методики МЛЭ:

- твердофазная лучевая эпитаксия (ТЛЭ),
- химическая лучевая эпитаксия (ХЛЭ).

Твердофазная лучевая эпитаксия состоит из двух этапов. Сначала по обычной методике МЛЭ, только проводимой при пониженной температуре, на подложке создают аморфную пленку осаждаемого материала, который затем кристаллизуется при более высоких температурах. Например, для кремния осаждение проводят при комнатной температуре, а отжиг при $500 - 600^\circ\text{C}$.

Метод ТЛЭ, несмотря на более низкое структурное совершенство пленок, имеет преимущество перед обычной МЛЭ в том, что позволяет получать так называемые *дельта-легированные слои* с резким профилем распределения легирующей примеси по толщине полупроводниковой пленки. Это связано с тем, что при типичных температурах и длительностях отжига, используемых

при ТЛЭ, диффузионная длина атомов примеси очень мала, порядка одного межатомного расстояния.

Химическая лучевая эпитаксия является модификацией метода МОС-гидридной эпитаксии (см. с. 506). Обычно ее проводят путем подачи компонентов III группы в виде металлоорганических соединений, а компонентов IV и V групп в виде гидридов при относительно высоких давлениях (порядка 1–760 Торр).

В отличие от традиционной МОС-гидридной технологии, процесс ХЛЭ протекает в условиях сверхвысокого вакуума (СВВ) внутри камеры МЛЭ, куда подают требуемые химические реагенты (МОС и гидриды) через специальную систему дозированного напуска газов. Из-за большой длины свободного пробега молекул в СВВ химические взаимодействия в газовой фазе практически исключены и не влияют на рост пленок, который определяется только реакциями на поверхности подложки.

Метод ХЛЭ, несмотря на большую сложность и повышенные температуры, обеспечивает более высокие скорости роста пленок при сохранении их структурного совершенства. Кроме того, ХЛЭ дает возможность использовать материалы, которые мало пригодны при обычной МЛЭ, например, высокотемпературные металлы (типа W, Nb и др.) с очень низким давлением паров, что затрудняет их испарение из эффузионных ячеек Кнудсена. Методы МЛЭ и ХЛЭ могут быть объединены в едином технологическом процессе, когда одни компоненты испаряются из ячеек Кнудсена, а другие подаются в форме газообразных соединений.

В настоящее время молекулярно-лучевая эпитаксия (как и ее модификации в виде ТЛЭ и ХЛЭ) позволяет выращивать самые совершенные монокристаллические пленки толщиной в несколько периодов решетки с уникальными физическими и геометрическими свойствами. В частности, сюда относятся полупроводниковые сверхрешетки на основе двойных, тройных и четверных твердых растворов, а также полупроводниковые гетероструктуры с использованием квантовых ям, проволок и точек как базовых элементов оптоэлектроники. Эпитаксиальные структуры сложного состава со слоями нанометровой толщины составляют основу разнообразных направлений современной наноэлектроники.

4.14. Кинетика химического транспорта веществ в проточных системах

Рассмотрим проточную систему, схематически изображенную на рис. 4.23, имеющую две температурные зоны. В этих зонах

протекает обратимая газотранспортная химическая реакция типа (2.16.1). В первой зоне с температурой $T_{\text{и}}$ располагается лодочка с твердым веществом S , служащим источником для роста эпитаксиальной пленки на подложке, расположенной в другой зоне с температурой $T_{\text{п}}$. Поскольку $T_{\text{и}} > T_{\text{п}}$ на рис. 4.23, то распределение температуры соответствует эндотермической химической реакции (см. п. 2.16).

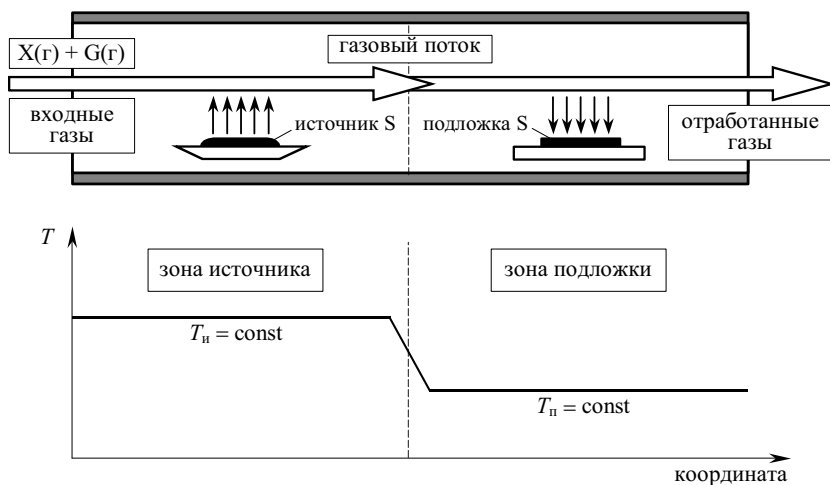


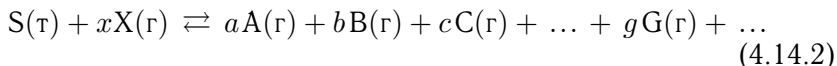
Рис. 4.23. Схематическое изображение газотранспортной проточной системы и распределение температуры в зонах источника с температурой $T_{\text{и}}$ и подложки с температурой $T_{\text{п}}$. Вертикальными стрелками в этих зонах показано направление движения газовых компонентов, содержащих элементы, входящие в твердое вещество S

На вход системы подается газовая смесь, в которую входит транспортный агент $X(r)$ (переносчик элементов, входящих в состав твердого вещества S) и газ-носитель $G(r)$, каковым наиболее часто является водород. Пусть в единицу времени на вход поступает n_X^0 молей газа-переносчика и n_G^0 молей газа-носителя. Эти величины при комнатной температуре T_0 связаны с входным давлением P_0 (обычно $P_0 = 1$ атм) соотношением

$$P_0 V_0 = (n_X^0 + n_G^0) R T_0. \quad (4.14.1)$$

Здесь V_0 (л/с) означает объемную скорость подачи входной газовой смеси, равную объему $(n_X^0 + n_G^0)$ молей, вводимых в реактор за единицу времени (не путать верхний нулевой индекс «0» с индексом «о», ранее введенным для стандартного состояния).

Перепишем ГХР типа (2.16.1) в форме, содержащей в явном виде газ-носитель в правой части уравнения:



Если степень превращения газового травителя X равна ξ , то на основании стехиометрии реакции (4.14.2) можно записать, во-первых, число существующих газовых молей:

$$n_X = (1 - \xi)n_X^0, \quad n_A = \frac{a}{x} \xi n_X^0, \\ n_B = \frac{b}{x} \xi n_X^0, \dots, \quad n_G = \xi n_G^0 + \frac{g}{x} \xi n_X^0, \dots$$

и, во-вторых, число молей твердого вещества S, вступивших в реакцию за единицу времени:

$$n_S \equiv \frac{1}{x} \xi n_X^0 = \frac{n_A}{a} = \frac{n_B}{b} = \frac{n_C}{c} = \dots = \frac{n_G - n_G^0}{g} = \dots \quad (4.14.3)$$

Из этих равенств вытекают следующие соотношения:

$$n_X = n_X^0 - \frac{x}{a} n_A, \quad n_G = n_G^0 + \frac{g}{a} n_A, \quad n_B = \frac{b}{a} n_A \quad \text{и т. д.}, \quad (4.14.4)$$

которые выражают величины n_X, n_G, n_B и т. д. через n_A .

Записывая с помощью (4.14.3) выражения для n_S в зоне источника, $n_S(T_{\text{и}})$, и в зоне подложки, $n_S(T_{\text{п}})$, можно найти число молей Δn_S вещества S, кристаллизующихся на подложке в единицу времени, в виде разности

$$\Delta n_S \equiv n_S(T_{\text{и}}) - n_S(T_{\text{п}}) = \frac{1}{a} [n_A(T_{\text{и}}) - n_A(T_{\text{п}})]. \quad (4.14.5)$$

Здесь $n_A(T_{\text{и}})$ и $n_A(T_{\text{п}})$ — число молей компонента A при температурах источника ($T_{\text{и}}$) и подложки ($T_{\text{п}}$). Для их нахождения необходимо установить связь между парциальным давлением p_A компонента A и полным давлением P в соответствующей температурной зоне.

По аналогии с равенством (4.14.1), для полного давления P при температуре T записываем соотношение

$$PV = (n_X + n_A + n_B + n_C + \dots + n_G + \dots)RT. \quad (4.14.6)$$

С помощью равенств (4.14.4) соотношение (4.14.6) преобразуется к виду

$$PV = (n_X^0 + n_G^0)RT + \left[1 + \frac{(b + c + \dots + g + \dots) - x}{a} \right] n_A RT \equiv$$

$$\equiv (n_X^0 + n_G^0)RT + \frac{\Delta\nu^r}{a} n_A RT. \quad (4.14.7)$$

Здесь $\Delta\nu^r \equiv \sum_K \nu_K^r - \sum_H \nu_H^r = (a + b + c + \dots + g + \dots) - x$ — изменение числа молей в газовой фазе при одном пробеге реакции (4.14.2) слева направо.

Из формулы (4.14.6) следует, что $p_A V = n_A RT$, тогда с учетом выражения (4.14.7) легко получить отношение давлений

$$\frac{p_A}{P} = \frac{n_A / (n_X^0 + n_G^0)}{1 + (\Delta\nu^r/a) [n_A / (n_X^0 + n_G^0)]}.$$

Отсюда находим выражение для числа молей компонента А,

$$n_A = (n_X^0 + n_G^0) \frac{p_A}{P_0 - (\Delta\nu^r/a) p_A}, \quad (4.14.8)$$

в котором давление P в реакционной камере принято равным входному давлению P_0 (возможно, атмосферному). Так как обычно $p_A \ll P_0$, то второе слагаемое в знаменателе (4.14.8) может быть опущено, что дает приближенное выражение

$$n_A \approx \frac{V_0}{RT_0} p_A, \quad (4.14.9)$$

где использовано соотношение (4.14.1), содержащее объемную скорость подачи газовой смеси V_0 (л/с).

Подстановка выражения (4.14.9) в формулу (4.14.5) дает окончательное выражение для числа молей вещества S, которые осаждаются на подложке за единицу времени:

$$\Delta n_S = \frac{V_0}{RT_0} \frac{p_A(T_{\text{и}}) - p_A(T_{\text{п}})}{a} \quad (\text{моль/с}), \quad (4.14.10)$$

где $p_A(T_{\text{и}})$ и $p_A(T_{\text{п}})$ — парциальные давления компонента А в зонах источника и подложки, соответственно.

Для вычисления «геометрической» скорости кристаллизации надо воспользоваться выражением (ср. формулу (4.12.11))

$$v_{\text{кр}} = \Delta n_S \frac{\mathcal{N}_A}{N_S A_{\text{п}}} \equiv \Delta n_S \frac{M_S}{\rho_S A_{\text{п}}} \quad (\text{см/с}),$$

где $\mathcal{N}_A$ — число Авогадро (моль⁻¹), N_S — число молекул S в единице объема твердого вещества (см⁻³), ρ_S и M_S — плотность (г/см³) и молярная масса (г/моль) вещества S, $A_{\text{п}}$ — площадь поверхности подложки.

Как видно из (4.14.10), для нахождения скорости эпитаксиального роста надо знать парциальное давление компонента А

в обеих температурных зонах. С ним жестко связаны давления других компонентов, в соответствии с равенствами (4.14.3):

$$p_B = \frac{b}{a} p_A, \quad p_C = \frac{c}{a} p_A \quad \text{и т. д.} \quad (4.14.11)$$

Давления транспортного агента p_X и газа-носителя p_G с помощью формул (4.14.4) и (4.14.9) также могут быть выражены через p_A :

$$p_X = \frac{n_X^0 - (x/a)n_A}{n_A} p_A = \left(n_X^0 \frac{RT_0}{p_A V_0} - \frac{x}{a} \right) p_A = p_X^0 - \frac{x}{a} p_A, \quad (4.14.12)$$

$$p_G = \frac{n_G^0 + (g/a)n_A}{n_A} p_A = \left(n_G^0 \frac{RT_0}{p_A V_0} + \frac{g}{a} \right) p_A = p_G^0 + \frac{g}{a} p_A, \quad (4.14.13)$$

где $p_X^0 = n_X^0 RT_0 / V_0$ и $p_G^0 = n_G^0 RT_0 / V_0$ — парциальные давления во входной газовой смеси, такие что $p_X^0 + p_G^0 = P_0$; обычно $p_X^0 \ll \ll p_G^0 \approx P_0 = 1$ атм.

Для нахождения давления p_A воспользуемся законом действующих масс для реакции (4.14.2),

$$K_p(T) = \frac{p_A^a p_B^b p_C^c \cdots p_G^g \cdots}{p_X^x},$$

который после подстановки в него равенств (4.14.11)–(4.14.13) принимает следующий вид:

$$K_p(T) = p_A^{a+b+c+\dots} \frac{(b/a)^b (c/a)^c \dots}{[p_X^0 - (x/a)p_A]^x}. \quad (4.14.14)$$

Здесь учтено, что в силу неравенства $p_A \ll p_G^0$ для давления газа-носителя (4.14.13) можно принять $p_G \approx p_G^0 \approx P_0 = 1$ атм.

Выражение (4.14.14) позволяет найти значения $p_A(T)$, соответствующие температурам источника ($T_{\text{и}}$) и подложки ($T_{\text{п}}$), которые входят в выражение (4.14.10) для скорости кристаллизации. Простейший вариант возникает в том случае, если $p_A \ll p_X^0$, тогда в знаменателе (4.14.14) можно опустить второе слагаемое.

В этом случае температурная зависимость парциального давления компонента А принимает вид

$$p_A(T) = \left[K_p(T) (p_X^0)^x \left(\frac{a}{b} \right)^b \left(\frac{a}{c} \right)^c \dots \right]^{\frac{1}{a+b+c+\dots}}. \quad (4.14.15)$$

Подстановка в выражение (4.14.15) значений констант равновесия в зонах источника, $K_p(T_{\text{и}})$, и подложки, $K_p(T_{\text{п}})$, дает

величины $p_A(T_{\text{и}})$ и $p_A(T_{\text{п}})$, которые, согласно формуле (4.14.10), определяют скорость эпитаксиального роста в открытых проточных системах.

4.15. Кинетика химического транспорта веществ в сэндвич-системах

Рассмотрим сэндвич-систему, схематически изображенную на рис. 4.24 а в виде двух параллельных пластин твердого вещества S, на поверхности которых при температурах T_1 и T_2 протекает обратимая газотранспортная химическая реакция типа (2.16.1).

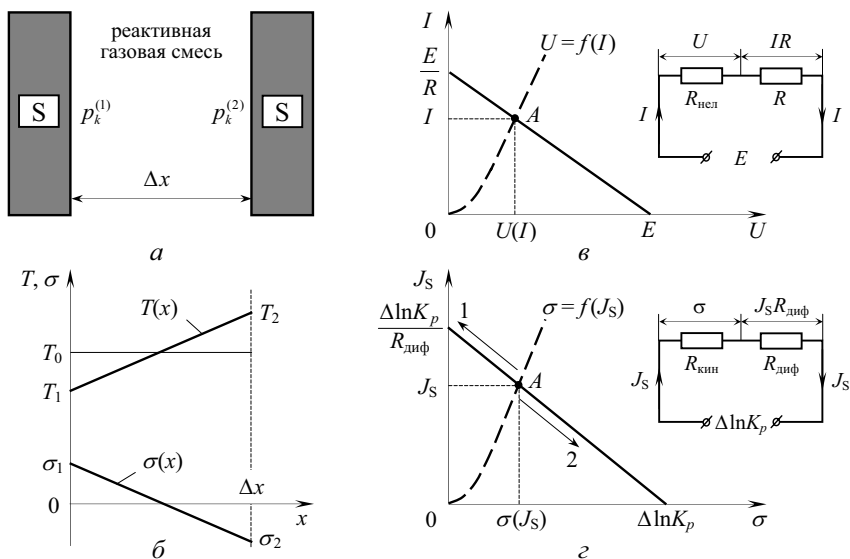


Рис. 4.24. Схематическое изображение газотранспортной сэндвич-системы (а), распределение температуры $T(x)$ и пересыщения $\sigma(x)$ в системе (б), электрический аналог транспортных процессов (в), скорость роста кристалла $J_s(\sigma)$ как функция пересыщения и эквивалентная схема химического транспорта вещества S (г)

Физические процессы в рассматриваемой системе определяются диффузионным массопереносом через многокомпонентную газовую смесь, заполняющую промежуток между пластинами. В обычных условиях в процессах рассеяния частиц в газовой смеси преобладают парные столкновения. По этой причине бинарный коэффициент диффузии D_{ij} для каждой пары компонентов, входящих в *многокомпонентную* смесь, обычно полагают равным аналогичной величине для *двухкомпонентной* ($i-j$) смеси.

Простейшая модель парного рассеяния представляет частицы в форме абсолютно упругих жестких шаров с массами m_i и m_j и с эффективными радиусами a_i и a_j . Эта модель дает следующее выражение для бинарного коэффициента диффузии как функции температуры T и давления P :

$$D_{ij}(T, P) = \frac{1}{3} \sqrt{\frac{8}{\pi m_{ij}}} \frac{(k_B T)^{3/2}}{\sigma_{ij} P}. \quad (4.15.1)$$

Здесь $m_{ij} = m_i m_j / (m_i + m_j)$ — приведенная масса двух частиц, $\sigma_{ij} = \pi(a_i + a_j)^2$ — сечение рассеяния, $P = (c_i + c_j)RT$ — полное давление в бинарной смеси газов с мольно-объемными концентрациями c_i и c_j .

Используя формулу (4.9.11), запишем полную плотность потока для i -го и j -го компонентов с учетом вклада стефановского потока, возникающего при изменении числа молей в газовой фазе:

$$\begin{aligned} J_i &= J_i^{\text{диф}} + J_i^{\text{ст}} = -\frac{D_{ij}}{RT} \frac{\partial p_i}{\partial x} + \frac{p_i}{RT} v_{\text{ст}}, \\ J_j &= J_j^{\text{диф}} + J_j^{\text{ст}} = -\frac{D_{ji}}{RT} \frac{\partial p_j}{\partial x} + \frac{p_j}{RT} v_{\text{ст}}. \end{aligned}$$

Из этих выражений с учетом того, что $D_{ij} = D_{ji}$, составляем комбинацию

$$p_i J_j - p_j J_i = \frac{D_{ij}}{RT} \left(p_j \frac{\partial p_i}{\partial x} - p_i \frac{\partial p_j}{\partial x} \right), \quad (4.15.2)$$

которая явно не включает стефановские вклады, так как соответствующие слагаемые в предыдущих равенствах скомпенсировали друг друга. После умножения выражения (4.15.2) на RT/D_{ij} и суммирования по j получаем

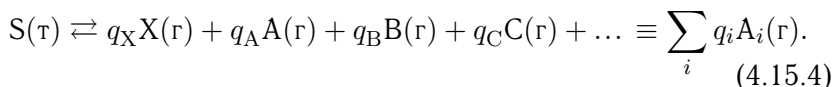
$$RT \sum_j \frac{p_i J_j - p_j J_i}{D_{ij}} = \sum_j \left(p_j \frac{\partial p_i}{\partial x} - p_i \frac{\partial p_j}{\partial x} \right) = P \frac{\partial p_i}{\partial x} - p_i \frac{\partial P}{\partial x}.$$

Полагая, что $P \equiv \sum_j p_j = cRT$ и $\partial P/\partial x = 0$, получаем общее уравнение, справедливое для многокомпонентной газовой смеси с учетом стефановских потоков (входящих в J_i и J_j),

$$\frac{\partial p_i}{\partial x} = \sum_j \frac{p_i J_j - p_j J_i}{c D_{ij}}, \quad (4.15.3)$$

которое называется **уравнением Кертисса–Гиришфельдера**. Стоящая в знаменателе величина $c D_{ij}$, как видно из выражения (4.15.1), пропорциональна $T^{1/2}$ и явно не зависит от давления P .

Для последующего анализа удобно переписать газотранспортную химическую реакцию, представленную в обобщенной форме (2.16.1), таким образом, чтобы в ее левой части осталось только твердое вещество S:



Введенные здесь новые стехиометрические коэффициенты q_i могут быть положительными, отрицательными и нулевыми (для инертных газов), в частности, $q_X = -x$, $q_A = a$, $q_B = b$ и т. д.

В стационарных условиях газовые потоки J_i , созданные ГХР, связаны с потоком J_S твердого вещества S равенством (4.9.2), переписанным для реакции (4.15.4) в форме

$$J_i = -q_i J_S. \quad (4.15.5)$$

Подстановка (4.15.5) в уравнение (4.15.3) приводит его к следующему виду:

$$\frac{\partial p_i}{\partial x} = J_S \sum_j \frac{q_i p_j - q_j p_i}{c D_{ij}}. \quad (4.15.6)$$

Запишем уравнение изотермы Вант-Гоффа в форме (2.13.5) для газотранспортной реакции (4.15.4):

$$\Delta G(T) = RT \left[\sum_i q_i \ln p_i - \ln K_p(T) \right], \quad (4.15.7)$$

где p_i — парциальные давления газообразных компонентов $A_i(\tau)$ в неравновесных условиях.

Считаем, что для сэндвич-системы с распределением температуры $T(x)$, показанным на рис. 4.24 б, при средней температуре $T_0 = (T_1 + T_2)/2$ реакция (4.15.4) находится в равновесии. На поверхности пластин с температурами T_1 и T_2 она смещена вправо над травящейся поверхностью ($\Delta G < 0$) и влево над растущей поверхностью ($\Delta G > 0$).

Используя выражение (4.15.7), запишем нормированные пересыщения для температур T_1 и T_2 :

$$\sigma_1 \equiv \frac{\Delta G(T_1)}{RT_1} = \sum_i q_i \ln p_i^{(1)} - \ln K_p(T_1),$$

$$\sigma_2 \equiv \frac{\Delta G(T_2)}{RT_2} = \sum_i q_i \ln p_i^{(2)} - \ln K_p(T_2).$$

Составляем разностное соотношение вычитанием первого уравнения из второго, после деления его на Δx получаем

$$\frac{\Delta \ln K_p}{\Delta x} = \frac{\sigma_1 - \sigma_2}{\Delta x} + \sum_i q_i \frac{\Delta \ln p_i}{\Delta x}, \quad (4.15.8)$$

где

$$\Delta \ln K_p = \ln K_p(T_2) - \ln K_p(T_1). \quad (4.15.9)$$

Как увидим ниже, величина $\Delta \ln K_p$, определенная равенством (4.15.9), является движущей силой процесса химического транспорта. Эта величина находится из уравнения изобары Вант-Гоффа типа (2.14.3) путем интегрирования его по T в пределах от T_1 до T_2 :

$$\Delta \ln K_p = \frac{\Delta H \Delta T}{RT_0^2}, \quad (4.15.10)$$

где ΔH — энтальпия химической реакции (4.15.4) и $\Delta T = T_2 - T_1$, при этом обычно $\Delta T \ll T_0 = (T_1 + T_2)/2$.

Для преобразования суммы в правой части (4.15.8) используем уравнение (4.15.6), в котором производная $\partial p_i / \partial x$ заменена отношением конечных приращений $\Delta p_i / \Delta x$, в силу малости величины Δx . Итак,

$$\begin{aligned} \sum_i q_i \frac{\Delta \ln p_i}{\Delta x} &= \sum_i \frac{q_i}{p_i} \frac{\Delta p_i}{\Delta x} = J_S \sum_i \sum_j \frac{q_i}{p_i} \frac{q_i p_j - q_j p_i}{c D_{ij}} = \\ &= J_S \sum_i \sum_j \frac{1}{2 c D_{ij}} \left(q_i^2 \frac{p_j}{p_i} - 2 q_i q_j + q_j^2 \frac{p_i}{p_j} \right) = \\ &= J_S \sum_i \sum_j \frac{1}{2 c D_{ij}} \frac{(q_i p_j - q_j p_i)^2}{p_i p_j}. \end{aligned} \quad (4.15.11)$$

Подставляя (4.15.11) в выражение (4.15.8), записываем его в следующем виде:

$$\Delta \ln K_p = \sigma + J_S R_{\text{диф}}, \quad \text{откуда} \quad J_S = \frac{\Delta \ln K_p - \sigma}{R_{\text{диф}}}, \quad (4.15.12)$$

где $\sigma = \sigma_1 - \sigma_2 = \sigma_1 + |\sigma_2| > 0$ — результирующее пересыщение в системе, соответствующее распределению $\sigma(x)$, показанному на рис. 4.24 б.

Нетрудно убедиться в том, что приведенные на этом рисунке распределения $T(x)$ и $\sigma(x)$ соответствуют эндотермической реакции, при которой левая пластина S является подложкой

($\sigma_1 > 0$: реакция (4.15.4) смещена влево), а правая пластина S — источником ($\sigma_2 < 0$: реакция (4.15.4) смещена вправо).

Как увидим ниже, введенная в формуле (4.15.12) величина

$$R_{\text{диф}} = \Delta x \sum_i \sum_{j>i} \frac{\Phi_{ij}}{cD_{ij}}, \quad \text{где} \quad \Phi_{ij} = \frac{(q_i p_j - q_j p_i)^2}{p_i p_j}, \quad (4.15.13)$$

имеет физический смысл *диффузионного сопротивления* процессу химического транспорта.

Для каждой конкретной реакции величина $R_{\text{диф}}$ рассчитывается с использованием формулы (4.15.1) при параметрах газовой фазы, соответствующих средней температуре T_0 . Вклад каждой пары молекулярных форм учитывается коэффициентом Φ_{ij} , который всегда положителен или равен нулю, в частности $\Phi_{ii} = 0$, а также

а) при $p_i/q_i = p_j/q_j$ и $q_i q_j > 0$ (коэффициенты q_i и q_j одного знака) имеем

$$\Phi_{ij} = 0; \quad (4.15.14)$$

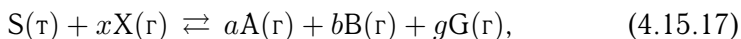
б) при $p_i/q_i = -p_j/q_j$ и $q_i q_j < 0$ (коэффициенты q_i и q_j разных знаков) имеем

$$\Phi_{ij} = -4q_i q_j; \quad (4.15.15)$$

в) для газа i -го сорта, являющегося либо инертным газом ($q_i = 0$), либо газом-носителем (при $p_i \equiv p_G \gg$ всех p_j), либо любым компонентом газовой смеси с парциальным давлением $p_i \gg p_j |q_i/q_j|$, имеем

$$\Phi_{ij} = q_j^2 \frac{p_i}{p_j}. \quad (4.15.16)$$

Найдем диффузионное сопротивление $R_{\text{диф}}$ для широко распространенных ГХР типа (4.14.2),



содержащих четыре газовых компонента, включая транспортный агент X и газ-носитель G. Эти газы, как и в предыдущем параграфе, подаются в систему с парциальными давлениями p_X^0 и p_G^0 . Для реакции (4.15.17) имеем $q_X = -x$, $q_A = a$, $q_B = b$, $q_G = g$, а парциальные давления связаны между собой соотношениями (4.14.11)–(4.14.13), т. е.

$$p_B = \frac{b}{a} p_A,$$

$$p_X = p_X^0 - \frac{x}{a} p_A \approx p_X^0 \gg p_A \text{ и } p_B,$$

$$p_G = p_G^0 + \frac{g}{a} p_A \approx p_G^0 \gg p_X^0 \gg p_A \text{ и } p_B.$$

С учетом этих соотношений, из формул (4.15.14) и (4.15.16) получаем равенства:

$$\begin{aligned}\Phi_{AB} &= 0, & \Phi_{XA} &= ab \Phi_{XB} = a^2 \frac{p_X^0}{p_A}, \\ \Phi_{GX} &= x^2 \frac{p_G^0}{p_X^0}, & \Phi_{GA} &= \frac{a^3}{b} \Phi_{GB} = a^2 \frac{p_G^0}{p_A}.\end{aligned}\quad (4.15.18)$$

Диффузионное сопротивление (4.15.13) для реакции (4.15.17) принимает следующий вид:

$$\begin{aligned}R_{\text{диф}} &= \frac{\Delta x}{c} \left(\frac{\Phi_{XA}}{D_{XA}} + \frac{\Phi_{XB}}{D_{XB}} + \frac{\Phi_{GA}}{D_{GA}} + \frac{\Phi_{GB}}{D_{GB}} + \frac{\Phi_{GX}}{D_{GX}} \right) = \\ &= \frac{\Delta x}{c} \left[\frac{x^2}{D_{GX}} \frac{p_G^0}{p_X^0} + \left(\frac{a^2}{D_{GA}} + \frac{b/a}{D_{GB}} \right) \frac{p_G^0}{p_A} + \left(\frac{a^2}{D_{XA}} + \frac{a/b}{D_{XB}} \right) \frac{p_X^0}{p_A} \right],\end{aligned}\quad (4.15.19)$$

где при переходе ко второму равенству были использованы выражения (4.15.18).

Следовательно, для вычисления $R_{\text{диф}}$ по формуле (4.15.19) необходимо знать бинарные коэффициенты диффузии (4.15.1) и состав газовой фазы (в частности, как и в п. 4.14, парциальное давление p_A), найденный из термодинамического расчета при температуре T_0 .

Для физической интерпретации формулы (4.15.12) полезно, как и ранее в п. 4.12, обратиться к электрической аналогии. Как известно, в последовательной электрической цепи, содержащей нелинейный элемент (вакуумный или полупроводниковый диод) и его нагрузочное сопротивление R , электродвижущая сила E делится между ними за счет протекания тока I . Этот ток создает падение напряжения IR на нагрузке и напряжение на нелинейном элементе $U(I)$, определяемое его вольт-амперной характеристикой (ВАХ) вида $U = f(I)$, тогда

$$E = U(I) + IR, \quad \text{или} \quad I = \frac{E - U(I)}{R}. \quad (4.15.20)$$

Выражение (4.15.20) является нелинейным уравнением для нахождения тока в цепи при известной ВАХ нелинейного элемента. В координатах $I - U$ оно описывает прямую линию с отрицательным наклоном, которая отсекает на осях отрезки E/R и E . Эта линия, называемая «нагрузочной прямой», изображена сплошной линией на рис. 4.24 в. Нелинейная ВАХ элемента

показана штриховой кривой $U=f(I)$. Графическое решение уравнения (4.15.20) дается точкой A , в которой пересекаются штриховая и сплошная линии. Координаты точки A дают ток I и напряжение $U(I)$ в рабочем режиме нелинейной цепи.

На рис. 4.24 *в* показана также эквивалентная схема, в которой нелинейный элемент изображен в виде сопротивления $R_{\text{нел}} = U(I)/I$, так что

$$I = \frac{E}{R + R_{\text{нел}}}. \quad (4.15.21)$$

Сравнение «электрического» уравнения (4.15.20) с аналогичным «химическим» уравнением (4.15.12) позволяет провести между ними следующую параллель. отождествляя электрический ток I с потоком J_S твердого вещества S , видим, что роль движущей силы для процесса химического транспорта играет величина $\Delta \ln K_p$ (аналог ЭДС E). Тогда пересыщение σ исполняет в химической системе ту же роль, которую в электрической цепи исполняет напряжение на нелинейном элементе $U = f(I)$. Следовательно, поверхностные процессы травления или роста кристаллов формируют в общем случае нелинейную зависимость $\sigma = f(J_S)$, задаваемую данным типом кристалла. Именно эта зависимость показана штриховой кривой на рис. 4.24 *г*.

Уравнение (4.15.12) в координатах $J_S - \sigma$ изображается прямой линией с отрицательным наклоном, определяемым величиной диффузионного сопротивления $R_{\text{диф}}$, которая отсекает на осях отрезки $\Delta \ln K_p / R_{\text{диф}}$ и $\Delta \ln K_p$. Следовательно, эта линия выполняет роль «нагрузочной прямой» для нелинейных поверхностных процессов, описываемых кривой $\sigma = f(J_S)$. Нахождение той или иной конкретной зависимости $\sigma = f(J_S)$ составляет основную задачу теории роста кристаллов.

По аналогии с рис. 4.24 *в*, точка A пересечения двух линий на рис. 4.24 *г* дает графическое решение нелинейного уравнения (4.15.12). Координаты этой точки соответствуют потоку J_S и пересыщению $\sigma(J_S)$ в рабочем режиме.

Этот режим определяется как массопереносом в газовой фазе, характеризуемым *диффузионным сопротивлением* $R_{\text{диф}}$, так и кинетическими процессами травления и кристаллизации. Для этих процессов, по аналогии с нелинейным электрическим сопротивлением $R_{\text{нел}} = U(I)/I$, можно ввести *кинетическое сопротивление* $R_{\text{кин}} = \sigma(J_S)/J_S$. Эта аналогия позволяет изобразить эквивалентную схему для процесса химического транспорта вещества S , показанную на рис. 4.24 *з*, где движущей силой является $\Delta \ln K_p$.

В полном соответствии с эквивалентной схемой на рис. 4.24 *з* и с формулой (4.15.21) можно записать следующее выражение:

$$J_S = \frac{\Delta \ln K_p}{R_{\text{диф}} + R_{\text{кин}}} \equiv \frac{\Delta \ln K_p}{R_{\text{тр}} + R_{\text{диф}} + R_{\text{кр}}}. \quad (4.15.22)$$

Здесь кинетическое сопротивление $R_{\text{кин}}$ представлено в виде двух слагаемых, обязанных процессам травления (*сопротивление травления* $R_{\text{тр}}$) и кристаллизации (*сопротивление кристаллизации* $R_{\text{кр}}$). Строго говоря, экспериментальному определению доступна лишь суммарная величина $R_{\text{кин}} = R_{\text{тр}} + R_{\text{кр}}$. Найти по отдельности вклады $R_{\text{кин}}$ и $R_{\text{диф}}$ можно благодаря тому, что от величины зазора Δx зависит лишь $R_{\text{диф}}$ (см. формулу (4.15.19)). Тогда, измеряя кривую $J_S(\Delta x)$, удастся разделить эти вклады и построить нелинейную «ростовую» зависимость $\sigma = f(J_S)$, показанную штриховой линией на рис. 4.24 *з*.

В заключение можно сделать следующие выводы на основании уравнения (4.15.22).

1. Диффузионная область процесса химического транспорта, для которой $R_{\text{диф}} \gg R_{\text{кин}}$, контролируется медленной стадией диффузионного массопереноса, при этом

$$J_S \approx \frac{\Delta \ln K_p}{R_{\text{диф}}} \quad \text{и} \quad R_{\text{кин}} \equiv \frac{\sigma}{J_S} = \frac{\sigma}{\Delta \ln K_p} R_{\text{диф}} \ll R_{\text{диф}}.$$

Отсюда следует, что $\sigma \ll \Delta \ln K_p$, т. е. в диффузионной области рабочая точка *A* на рис. 4.24 *з* смещается вверх вдоль «нагрузочной прямой» в направлении стрелки 1. Следовательно, скорость роста кристалла J_S достигает максимально возможного значения $\Delta \ln K_p / R_{\text{диф}}$ при фиксированной величине $\Delta \ln K_p$. Это значение не зависит от кристаллографической ориентации и дефектности поверхности.

2. Кинетическая область процесса химического транспорта, для которой $R_{\text{кин}} \gg R_{\text{диф}}$, контролируется медленной стадией гетерогенной химической реакции, при этом

$$J_S \approx \frac{\Delta \ln K_p}{R_{\text{кин}}} = \frac{\Delta \ln K_p}{\sigma} J_S.$$

Отсюда следует, что $\sigma \approx \Delta \ln K_p$, т. е. в кинетической области рабочая точка *A* на рис. 4.24 *з* смещается вниз вдоль «нагрузочной прямой» в направлении стрелки 2. Следовательно, скорость роста кристалла в кинетической области много меньше, чем в диффузионной, и сильно зависит от дефектности, кристаллографической ориентации и свойств поверхности роста.

Поскольку величина диффузионного сопротивления $R_{\text{диф}}$ линейно зависит от расстояния Δx между пластинами, то изменением его можно регулировать соотношение между $R_{\text{диф}}$ и $R_{\text{кин}}$, переводя процесс кристаллизации в диффузионную область при достаточно большом зазоре. Этот зазор (расстояние между пластинами) в применяемых на практике сэндвич-системах обычно выбирают в пределах от 0,2 до 0,5 мм.

4.16. Кинетика химического транспорта веществ при реактивном распылении

Реактивное распыление металлов в плазме инертного газа (обычно аргона) с добавлением химически активных газов (чаще всего кислорода, азота и т. п.) широко применяется как способ пленочного осаждения химических соединений металлов (таких как оксиды, нитриды и т. п.). В последние годы круг практического применения подобных пленок (в первую очередь, оксидов переходных металлов типа титана, тантала и т. п.) расширился от микроэлектронных и оптоэлектронных устройств до биомедицинских и экологических приложений.

Многочисленные исследования привели к заключению о том, что распыление металлов бомбардировкой мишени ионами аргона в присутствии реактивного газа является существенно нелинейным процессом и проявляет гистерезисный эффект, показанный на рис. 4.25.

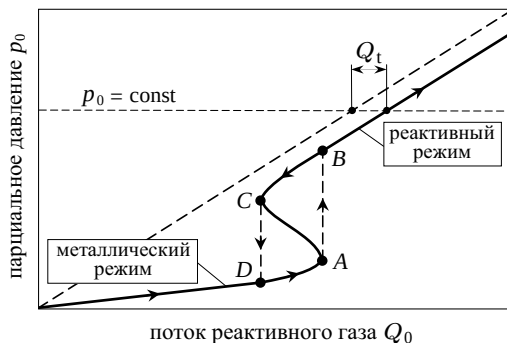


Рис. 4.25. Качественная зависимость парциального давления p_0 реактивного газа от его потока Q_0 ; штриховой линией показана аналогичная зависимость в отсутствие распыления металла с наклоном, определяемым скоростью откачки

Как видно из рисунка, на кривой имеются две ветви распыления, названные *металлическим режимом* и *реактивным режимом*, которые разделены петлей гистерезиса $ABCD$. Этому

можно дать следующее качественное объяснение, которое ниже будет теоретически подтверждено на основе численных расчетов.

На начальном этапе, когда поток Q_0 реактивного газа (например, кислорода) достаточно мал, все поверхности рабочей камеры (стенок, подложки, мишени) почти не покрыты химическим соединением (оксидом) и активно потребляют кислород для процесса окисления. Поэтому в газовой фазе остается мало молекул кислорода, создающих низкое парциальное давление p_0 .

По мере окисления металла давление p_0 постепенно нарастает вдоль нижней («металлической») ветви вплоть до некоторой критической точки A . В этой точке из-за нелинейности химических процессов возникает неустойчивость и скачкообразный переход в точку B на верхней ветви, названной реактивным (оксидным) режимом. В этом режиме металлические поверхности (мишень, подложка и стенки) почти полностью (в той или иной степени) покрыты химическим соединением (оксидом), что и дает название данному режиму. При этом давление кислорода в камере близко к той величине, которая создается вакуумным насосом в отсутствие процесса распыления (штриховая линия на рис. 4.25).

При $p_0 = \text{const}$ расстояние между сплошной и штриховой линиями (имеющими одинаковый наклон, формируемый режимом откачки) равно той части Q_t реактивного газового потока Q_0 , которая потребляется мишенью. Действительно, в реактивном (оксидном) режиме только окисленная поверхность мишени теряет оксид в результате его распыления ионами аргона Ar^+ .

Обратное движение вдоль верхней («реактивной») ветви от точки B в точку C путем уменьшения потока Q_0 снова приводит к неустойчивости процесса распыления и скачкообразному возвращению системы в металлический режим в точке D .

Подводя итог вышесказанному, отметим основные особенности двух режимов реактивного распыления металлов и на качественном уровне объясним природу гистерезисного эффекта.

- В *металлическом режиме* мишень практически свободна от продуктов плазмо-химической реакции, а с ее поверхности распыляются (с достаточно большой скоростью) атомы металла, осаждающиеся на подложке и стенках вакуумной камеры, которые эффективно потребляют реактивный газ.

- В *реактивном режиме* мишень практически полностью покрыта химическим соединением металла с реактивным газом, которое и распыляется (с относительно небольшой скоростью) ионной бомбардировкой мишени, в результате чего реактивный газ потребляется лишь для восполнения распыленного соединения.

Таким образом, на поверхности мишени работают два конкурирующих процесса: формирование химического соединения в результате поверхностной реакции и его распыление ускоренными ионами аргона. При малом вводимом потоке Q_0 реактивного газа скорость химической реакции меньше скорости ионного распыления продуктов реакции. Поэтому распыленный поток преимущественно содержит атомы металла (металлический режим), осаждение которых на стенках камеры способствует эффективному потреблению реактивного газа, что снижает давление p_0 .

С увеличением Q_0 скорость химической реакции становится больше скорости распыления химического соединения, которое начинает покрывать поверхность мишени. В результате этого доля атомов металла в потоке распыленных частиц снижается, уменьшается потребление реактивного газа стенками камеры, растет его парциальное давление в камере, что еще более увеличивает скорость химической реакции. Следовательно, в распылительной системе возникает положительная обратная связь, которая и приводит к лавинообразному переходу из металлического режима в реактивный (из точки *A* в точку *B* на рис. 4.25).

Ниже рассмотрим в деталях кинетику реактивного распыления металлов, следуя ранее опубликованной работе авторов¹⁾.

1. Кинетика поверхностной химической реакции. Как уже отмечалось в п. 4.9, химическому взаимодействию молекул реактивного газа с поверхностью твердого тела всегда предшествует сорбционный процесс, протекающий в два этапа (см. п. 5.9).

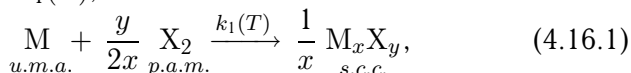
Первый этап, известный как физическая адсорбция (физадсорбция) реализуется за счет сил межмолекулярного взаимодействия (диполь-дипольных сил Ван-дер-Ваальса, см. пп. 3.1 и 5.9). Химическая адсорбция (хемосорбция), как второй этап, вступает в действие после термической диссоциации физически адсорбированных молекул на атомы. Каждый из атомов удерживается на поверхности межатомными силами химической природы. На этом основании разумно включить хемосорбцию в состав поверхностной химической реакции в качестве ее начальной стадии. Поэтому поверхностному химическому взаимодействию теперь считаем предшествующей лишь физическую адсорбцию, кинетика которой будет изучена ниже (см. п. 5.10).

Рассмотрим взаимодействие двухатомных газовых молекул X_2 (кислород O_2 , азот N_2 и т. п.), физически адсорбированных

¹⁾ Barybin A. A., Shapovalov V. I. "Nonisothermal chemical model of reactive sputtering", J. Appl. Phys., vol. 101, p. 054905, 2007.

на поверхности металла, с его атомами М. Результатом такого взаимодействия является химическое соединение M_xX_y (оксид M_xO_y , нитрид M_xN_y и т. п.) в виде тонкого слоя на поверхности металла.

Опишем взаимодействие частиц на поверхности металла в виде односторонней химической реакции, характеризующейся константой скорости $k_1(T)$,



где сокращения *u.m.a.*, *p.a.m.* и *s.c.c.* от англ. *unoccupied metal atom*, *physically adsorbed molecule* и *surface chemical compound*.

На поверхности металла будем различать центры физадсорбции с концентрацией N_Φ и центры химической реакции с концентрацией N_x . Первые удерживают газовые молекулы X_2 , поэтому $N_\Phi \approx N_M$ (число атомов М на единице площади поверхности). Каждый химический центр содержит одну молекулу соединения M_xX_y , возникающую по реакции (4.16.1). Это означает, что x соседних атомов М на поверхности связаны друг с другом при помощи y атомов Х, расположенных *над поверхностью*, образуя единый центр химической реакции. Поэтому можно считать, что $N_x = N_\Phi/x \approx N_M/x$. Степень заполнения этих центров будем характеризовать следующими величинами:

$$\begin{aligned} \theta &= \frac{[X_2]}{N_\Phi} && \text{— для адсорбционных центров,} \\ \Theta &= \frac{[M_xX_y]}{N_x} && \text{— для химических центров,} \end{aligned} \quad (4.16.2)$$

где [...] означает поверхностную концентрацию соответствующих частиц (м^{-2}). Отсюда концентрация незанятых атомов металла (*u.m.a.*), которые могут участвовать в реакции (4.16.1), равна

$$[M] = xN_x(1 - \Theta). \quad (4.16.3)$$

Определим скорость поверхностной реакции (4.16.1) с учетом ее стехиометрии как (с размерностью $\text{м}^{-2} \cdot \text{с}^{-1}$)

$$v \equiv -\frac{d[M]}{dt} = -\frac{1}{y/2x} \frac{d[X_2]}{dt} = \frac{1}{1/x} \frac{d[M_xX_y]}{dt}. \quad (4.16.4)$$

Подстановкой (4.16.3) в формулу (4.16.4) выражаем скорость реакции через скорость заполнения химических центров:

$$v \equiv -\frac{d[M]}{dt} = xN_x \frac{d\Theta}{dt}. \quad (4.16.5)$$

В соответствии с основным постулатом химической кинетики (1.18.13) для односторонней реакции (4.16.1), ее скорость (4.16.5) при температуре T может быть представлена в виде произведения концентраций исходных реагентов, взятых в степенях, равных их стехиометрическим коэффициентам:

$$v(T) = k_1(T) [M][X_2]^{y/2x}. \quad (4.16.6)$$

В дальнейшем будем рассматривать неизотермическую ситуацию, когда мишень, подложка и стенки камеры имеют разные температуры T_t , T_s и T_w (индексы t, s и w от англ. *target*, *substrate* и *wall*). Температуру газовой среды внутри распылительной камеры обозначаем T_0 и считаем, что $T_0 = T_w$.

Приравнивая формулы (4.16.5) и (4.16.6) с учетом выражений (4.16.2) и (4.16.3) получаем уравнение для скорости заполнения химических центров на поверхностях мишени, подложки и стенок камеры (соответственно, индекс $i = t, s, w$):

$$\left(\frac{d\Theta_i}{dt} \right)_{\text{ch}} \equiv \frac{v(T_i)}{x N_x} = \frac{k(T_i)}{N_x} \theta_i^{y/2x} (1 - \Theta_i), \quad (4.16.7)$$

где $k(T_i) \equiv k_1(T_i) N_x N_\phi^{y/2x}$ — результирующая константа скорости поверхностной реакции (4.16.1) (с размерностью $\text{м}^{-2} \cdot \text{с}^{-1}$). Здесь нижний индекс ch (от англ. *chemical*) отражает вклад только химической реакции. В дополнение к этому вкладу ниже появятся вклады от процессов *распыления* мишени (индекс sp от англ. *sputtering*) и *осаждения* на подложке и стенках камеры (индекс dep от англ. *deposition*).

Кинетика заполнения физадсорбционных центров на i -ой поверхности ($i = t, s, w$), характеризующая изменением во времени степени заполнения $\theta_i(t)$, рассмотрена ниже (см. п. 5.10).

2. Потоки реактивного газа и распыленных частиц. Поток Q_0 (с размерностью с^{-1}) реактивного газа X_2 поступает в вакуумную камеру при температуре T_0 , создавая парциальное давление $p_0 = c_0 k_B T_0$. Эти величины позволяют найти число молекул X_2 , падающих на единицу площади в единицу времени (плотность падающего потока, $\text{м}^{-2} \cdot \text{с}^{-1}$), в виде (см. формулу (5.10.18))

$$F_0 = \frac{p_0}{\sqrt{2\pi m k_B T_0}}, \quad (4.16.8)$$

где m — масса молекулы X_2 , а индекс 0 отмечает газовую среду.

Поток Q_0 частично потребляется мишенью (Q_t), подложкой (Q_s) и стенками (Q_w) благодаря химической реакции (4.16.1), протекающей на них при соответствующих температурах T_t , T_s

и $T_w = T_0$. Остаток неиспользованного реактивного газа откачивается вакуумным насосом, создающим поток откачки,

$$Q_p = c_0 Y_p = \frac{p_0}{k_B T_0} Y_p, \quad (4.16.9)$$

где Y_p — скорость откачки ($\text{м}^3/\text{с}$) вакуумного насоса (индекс p от англ. *pump*). На основании вышесказанного можем записать

$$Q_0 = Q_t + Q_s + Q_w + Q_p. \quad (4.16.10)$$

Для записи парциальных газовых потоков Q_i ($i = t, s, w$), потребляемых химической реакцией на мишени, подложке и стенках камеры, надо воспользоваться стехиометрическим соотношением (4.16.4) и формулой (4.16.7):

$$\begin{aligned} Q_i &\equiv \left| \frac{d[X_2]}{dt} \right|_i A_i = \frac{y}{2x} v(T_i) A_i = \frac{y}{2} N_x \left(\frac{d\Theta_i}{dt} \right)_{\text{ch}} A_i = \\ &= \frac{y}{2} k(T_i) \theta_i^{y/2x} (1 - \Theta_i) A_i. \end{aligned} \quad (4.16.11)$$

Здесь A_i — *эффективные* (для процессов распыления и осаждения) *площади* мишени ($i = t$), подложки ($i = s$) и стенок ($i = w$).

Теперь рассмотрим потоки частиц, распыляемых с поверхности мишени в результате бомбардировки ионами аргона Ar^+ . Поскольку на поверхности присутствуют как атомы металла M , так и молекулы химического соединения $M_x X_y$, то мишень характеризуется двумя *коэффициентами распыления*: Y_m — для металла и Y_c — для соединения (индексы m и c от англ. *metal* и *compound*).

Эффективные площади поверхности мишени, покрытые химическим соединением $M_x X_y$ и металлом M , соответственно, равны $\Theta_t A_t$ и $(1 - \Theta_t) A_t$.

Итак, выражения для потоков распыленных частиц таковы:

- для молекул $M_x X_y$ (отмеченных верхним индексом c)

$$F_{sp}^c = \frac{J}{e} Y_c \Theta_t A_t \quad (c^{-1}), \quad (4.16.12)$$

- для атомов M (отмеченных верхним индексом m)

$$F_{sp}^m = \frac{J}{e} Y_m (1 - \Theta_t) A_t \quad (c^{-1}), \quad (4.16.13)$$

где J — плотность тока, создаваемого ионами аргона при распылении мишени.

Распыленные частицы, переносимые потоками (4.16.12) и (4.16.13), осаждаются на поверхности подложки и стенках камеры. Предполагаем, что распределение потоков F_{sp}^c и F_{sp}^m по всей

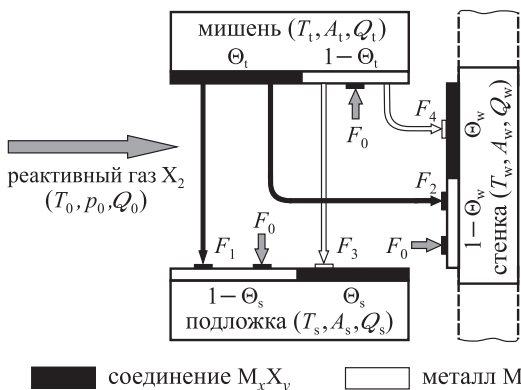


Рис. 4.26. Схема разделения потоков частиц, распыленных с мишени, между подложкой и стенками камеры: $F_1 - F_4$ — потоки молекул $M_x X_y$ и атомов M , даваемые формулами (4.16.14)–(4.16.17); F_0 — плотность потока реактивного газа X_2 , подаваемого в распылительную камеру

площади равномерное. Тогда их можно разделить между подложкой (части потоков F_1 и F_3) и стенками камеры (части потоков F_2 и F_4) пропорционально эффективным площадям A_s и A_w . Такое разделение потоков схематически показано на рис. 4.26.

Следовательно, потоки распыленных с мишени частиц, которые осаждаются на подложке и стенках камеры, следующие:

- поток молекул $M_x X_y$, поступающий на ту часть подложки площадью $(1 - \Theta_s)A_s$, которая не покрыта соединением $M_x X_y$ (это означает увеличение степени заполнения Θ_s),

$$F_1 = F_{sp}^c \frac{(1 - \Theta_s)A_s}{A_s + A_w} = \frac{J}{e} Y_c \Theta_t (1 - \Theta_s) \frac{A_t A_s}{A_s + A_w}; \quad (4.16.14)$$

- поток молекул $M_x X_y$, поступающий на ту часть стенок площадью $(1 - \Theta_w)A_w$, которая не покрыта соединением $M_x X_y$ (это означает увеличение степени заполнения Θ_w),

$$F_2 = F_{sp}^c \frac{(1 - \Theta_w)A_w}{A_s + A_w} = \frac{J}{e} Y_c \Theta_t (1 - \Theta_w) \frac{A_t A_w}{A_s + A_w}; \quad (4.16.15)$$

- поток атомов M , поступающий на ту часть подложки площадью $\Theta_s A_s$, которая покрыта соединением $M_x X_y$ (это означает уменьшение степени заполнения Θ_s),

$$F_3 = F_{sp}^m \frac{\Theta_s A_s}{A_s + A_w} = \frac{J}{e} Y_m \Theta_s (1 - \Theta_t) \frac{A_t A_s}{A_s + A_w}; \quad (4.16.16)$$

- поток атомов M , поступающий на ту часть стенок площадью $\Theta_w A_w$, которая покрыта соединением $M_x X_y$ (это означает

уменьшение степени заполнения Θ_w),

$$F_4 = F_{sp}^m \frac{\Theta_w A_w}{A_s + A_w} = \frac{J}{e} Y_m \Theta_w (1 - \Theta_t) \frac{A_t A_w}{A_s + A_w}. \quad (4.16.17)$$

Здесь опущены поток молекул $M_x X_y$ на поверхность соединения $M_x X_y$ и поток атомов M на поверхность металла M , которые не изменяют степени заполнения химических центров, Θ_s и Θ_w .

3. Кинетика распыления соединения $M_x X_y$ с поверхности мишени. Как уже отмечалось, кинетика процессов на мишени определяются конкуренцией между образованием химического соединения $M_x X_y$ и его ионным распылением. Скорость химического процесса дается уравнением (4.16.7) при $i = t$.

Скорость изменения поверхностной концентрации $[M_x X_y]$, вызванного ионным распылением, дается уравнением (4.16.12):

$$\left(\frac{d[M_x X_y]}{dt} \right)_{sp} \equiv - \frac{F_{sp}^c}{A_t} = - \frac{J}{e} Y_c \Theta_t. \quad (4.16.18)$$

Поскольку, согласно (4.16.2), $[M_x X_y] = \Theta_t N_x$, то формула (4.16.18) дает скорость изменения (уменьшения) величины Θ_t в результате ионного распыления:

$$\left(\frac{d\Theta_t}{dt} \right)_{sp} = - \frac{1}{N_x} \frac{J}{e} Y_c \Theta_t. \quad (4.16.19)$$

Результирующая скорость изменения степени заполнения химических центров на мишени получается суммированием химического вклада (4.16.7) при $i = t$ и вклада (4.16.19) от распыления, что приводит к искомому кинетическому уравнению

$$\begin{aligned} \frac{d\Theta_t}{dt} &\equiv \left(\frac{d\Theta_t}{dt} \right)_{ch} + \left(\frac{d\Theta_t}{dt} \right)_{sp} = \\ &= \frac{1}{N_x} \left[k(T_t) \theta_t^{y/2x} (1 - \Theta_t) - \frac{J}{e} Y_c \Theta_t \right]. \end{aligned} \quad (4.16.20)$$

Нетрудно убедиться в том, что уравнение (4.16.20) можно привести к релаксационной форме,

$$\frac{d}{dt} \left(\Theta(t) - \frac{\tau_{\Phi}}{\tau_{ch}} \right) = - \frac{1}{\tau_{\Phi}} \left(\Theta(t) - \frac{\tau_{\Phi}}{\tau_{ch}} \right), \quad (4.16.21)$$

структура которой совпадает с кинетическим уравнением (5.10.23) для степени заполнения θ физадсорбционных центров.

В кинетическом уравнении (4.16.21) введено эффективное время релаксации для процессов на мишени (ср. (5.10.26)),

$$\tau_{\text{эф}} = \left(\frac{1}{\tau_{\text{ch}}} + \frac{1}{\tau_{\text{sp}}} \right)^{-1}, \quad (4.16.22)$$

построенное из «химического» (τ_{ch}) и «распылительного» (τ_{sp}) времен релаксации, введенных равенствами

$$\frac{1}{\tau_{\text{ch}}} = \frac{k(T_t)}{N_x} \theta_t^{y/2x} \quad \text{и} \quad \frac{1}{\tau_{\text{sp}}} = \frac{1}{N_x} \frac{J}{e} Y_c. \quad (4.16.23)$$

Решение кинетического уравнения (4.16.21) при начальном условии $\Theta_t(0) = 0$ имеет вид (ср. уравнение (5.10.27))

$$\Theta_t(t) = \Theta_{t0} \left[1 - \exp(-t/\tau_{\text{эф}}) \right], \quad (4.16.24)$$

где стационарная степень заполнения, достигаемая при $t \rightarrow \infty$, дается выражением

$$\Theta_{t0} \equiv \frac{\tau_{\text{эф}}}{\tau_{\text{ch}}} = \frac{k(T_t) \theta_t^{y/2x}}{k(T_t) \theta_t^{y/2x} + (J/e) Y_c}. \quad (4.16.25)$$

Из вышеизложенного ясно, что полученные выражения справедливы, если для физической адсорбции брать стационарную степень заполнения центров ($\theta_t = \theta_{t0}$), даваемую формулой (5.10.28), а именно

$$\theta_{t0} = \frac{b(T_t, T_0) p_0}{1 + b(T_t, T_0) p_0} \equiv \frac{b_t p_0}{1 + b_t p_0}, \quad (4.16.26)$$

где величина $b_t \equiv b(T_t, T_0)$ определена в форме (5.10.25) при $T_a = T_t$. Это действительно допустимо, так как процесс физадсорбции с временами релаксации порядка микросекунд протекает существенно быстрее, чем химические процессы.

4. Кинетика осаждения соединения M_xX_y на подложке и стенках камеры. Здесь скорость образования соединения M_xX_y в результате химической реакции (4.16.1) описывается уравнением (4.16.7) при $i = s$ (подложка) и $i = w$ (стенки камеры).

Анализ осаждения молекул M_xX_y и атомов M начнем с рассмотрения процессов на поверхности подложки.

Скорость *увеличения* поверхностной концентрации $[M_xX_y]$ вызвана потоком F_1 , доставляющим молекулы M_xX_y на поверхность подложки (см. формулу (4.16.14) и рис. 4.26):

$$\left(\frac{d[M_xX_y]}{dt} \right)_{\text{dep}} \equiv \frac{F_1}{A_s} = \frac{J}{e} Y_c \Theta_t (1 - \Theta_s) \frac{A_t}{A_s + A_w}. \quad (4.16.27)$$

Скорость *уменьшения* поверхностной концентрации $[M_x X_y]$ вызвана потоком F_3 , доставляющим атомы М на поверхность подложки (см. формулу (4.16.16) и рис. 4.26):

$$\left(\frac{d[M]}{dt}\right)_{\text{dep}} \equiv \frac{F_3}{A_s} = \frac{J}{e} Y_m \Theta_s (1 - \Theta_t) \frac{A_t}{A_s + A_w}. \quad (4.16.28)$$

Из формул (4.16.2) и (4.16.3) получаем равенства

$$\frac{d[M_x X_y]}{dt} = -\frac{1}{x} \frac{d[M]}{dt} = N_x \frac{d\Theta_s}{dt}. \quad (4.16.29)$$

Подстановка соответствующих равенств (4.16.29) в левую часть формул (4.16.27) и (4.16.28) с последующим сложением результатов дает скорость изменения величины Θ_s в результате осаждения распыленных частиц на поверхности подложки:

$$\left(\frac{d\Theta_s}{dt}\right)_{\text{dep}} = \frac{1}{N_x} \frac{J}{e} \frac{A_t}{A_s + A_w} \left[Y_c \Theta_t (1 - \Theta_s) - \frac{Y_m}{x} \Theta_s (1 - \Theta_t) \right]. \quad (4.16.30)$$

Результирующая скорость изменения степени заполнения химических центров на подложке получается суммированием химического вклада (4.16.7) при $i = s$ и вклада (4.16.30) от осаждения, что приводит к искомому кинетическому уравнению:

$$\begin{aligned} \frac{d\Theta_s}{dt} \equiv \left(\frac{d\Theta_s}{dt}\right)_{\text{ch}} + \left(\frac{d\Theta_s}{dt}\right)_{\text{dep}} &= \frac{1}{N_x} \left\{ k(T_s) \theta_s^{y/2x} (1 - \Theta_s) + \right. \\ &\left. + \frac{J}{e} \frac{A_t}{A_s + A_w} \left[Y_c \Theta_t (1 - \Theta_s) - \frac{Y_m}{x} \Theta_s (1 - \Theta_t) \right] \right\}. \end{aligned} \quad (4.16.31)$$

Для стенок камеры, когда работающими являются потоки F_2 и F_4 (см. формулы (4.16.15), (4.16.17) и рис. 4.26), формулы, аналогичные вышеприведенным для подложки, получаются при взаимной замене нижних индексов $s \rightleftharpoons w$. Тогда из уравнения (4.16.31) получаем кинетическое уравнение для стенок камеры:

$$\begin{aligned} \frac{d\Theta_w}{dt} \equiv \left(\frac{d\Theta_w}{dt}\right)_{\text{ch}} + \left(\frac{d\Theta_w}{dt}\right)_{\text{dep}} &= \frac{1}{N_x} \left\{ k(T_w) \theta_w^{y/2x} (1 - \Theta_w) + \right. \\ &\left. + \frac{J}{e} \frac{A_t}{A_s + A_w} \left[Y_c \Theta_t (1 - \Theta_w) - \frac{Y_m}{x} \Theta_w (1 - \Theta_t) \right] \right\}. \end{aligned} \quad (4.16.32)$$

Кинетические уравнения (4.16.31) и (4.16.32) для подложки и стенок камеры имеют более сложную форму в сравнении с кинетическим уравнением (4.16.20) для мишени. В частности, они

не могут быть приведены к релаксационной форме, аналогичной уравнению (4.16.21).

5. Анализ стационарного режима распыления. В стационарном состоянии, когда для всех $i = t, s, w$ в кинетических уравнениях полагаем $d\Theta_i/dt = 0$, степени заполнения химических и физадсорбционных центров ($\Theta_i(t)$ и $\theta_i(t)$) принимают постоянные значения (Θ_{i0} и θ_{i0}).

В частности, для мишени (при $i = t$) справедливы стационарные выражения (4.16.25) и (4.16.26), которые перепишем в виде (опуская для упрощения записи нижний нулевой индекс):

$$\Theta_t = \frac{k(T_t) \theta_t^{y/2x}}{k(T_t) \theta_t^{y/2x} + (J/e) Y_c}, \quad (4.16.33)$$

$$\theta_t = \frac{b_t p_0}{1 + b_t p_0}, \quad (4.16.34)$$

где формула (5.10.25) для мишени при $T_a = T_t$ переписана как

$$b_t \equiv b(T_t, T_0) = \frac{\alpha_0}{N_\phi} \frac{\tau_0 \exp(Q_\phi/RT_t)}{\sqrt{2\pi m k_B T_0}}. \quad (4.16.35)$$

При этом теплота физической адсорбции $Q_\phi < 10$ ккал/моль ≈ 0.435 эВ, $\tau_0 \approx 10^{-13}$ с (см. формулу (5.9.1)), а для коэффициента конденсации обычно берут $\alpha_0 = 1$.

Для подложки и стенок камеры их кинетические уравнения (4.16.31) и (4.16.32) при нулевой левой части позволяют связать стационарные степени заполнения Θ_s и Θ_w с аналогичной величиной Θ_t для мишени следующими соотношениями:

$$\Theta_s = \left[1 + \frac{(J/e)(Y_m/x)(1 - \Theta_t) a_t}{k(T_s) \theta_s^{y/2x} + (J/e) Y_c \Theta_t a_t} \right]^{-1}, \quad (4.16.36)$$

$$\Theta_w = \left[1 + \frac{(J/e)(Y_m/x)(1 - \Theta_t) a_t}{k(T_w) \theta_w^{y/2x} + (J/e) Y_c \Theta_t a_t} \right]^{-1}. \quad (4.16.37)$$

Здесь $a_t \equiv A_t/(A_s + A_w)$ — доля эффективной площади мишени в общей эффективной площади ($A_s + A_w$) внутри камеры, на которую осаждаются распыленные частицы. Адсорбционные степени заполнения θ_s и θ_w также описываются формулами (4.16.34)–(4.16.35), но с заменой температур $T_t \rightarrow T_s$ и $T_t \rightarrow T_w = T_0$.

При численных расчетах уравнения (4.16.33)–(4.16.37) должны быть дополнены уравнениями (4.16.9)–(4.16.11) для потоков реактивного газа X_2 . Расчеты были выполнены для реактивного

окисления тантала в неізотермической системе $\text{Ta}-\text{O}_2-\text{Ta}_2\text{O}_5$ ($x = 2$, $y = 5$) при следующих физических параметрах:

$$Q_\phi = 5 \text{ ккал/моль}, \quad N_\phi = 1 \cdot 10^{15} \text{ см}^{-2}, \quad N_x = N_\phi/x = 5 \cdot 10^{14} \text{ см}^{-2},$$

$$\alpha_0 = 1, \quad \tau_0 = 10^{-13} \text{ с}, \quad T_t = 600 \text{ К}, \quad T_s = 450 \text{ К}, \quad T_w = T_0 = 300 \text{ К},$$

$$A_s + A_w = 300 \text{ см}^2, \quad a_t \equiv \frac{A_t}{A_s + A_w} = 0,075, \quad a_s \equiv \frac{A_s}{A_s + A_w} = 0,035,$$

$$Y_m = 0,6, \quad Y_c = 0,025, \quad Y_p = 8 \text{ л/с}, \quad J = 0,05 \text{ А/см}^2.$$

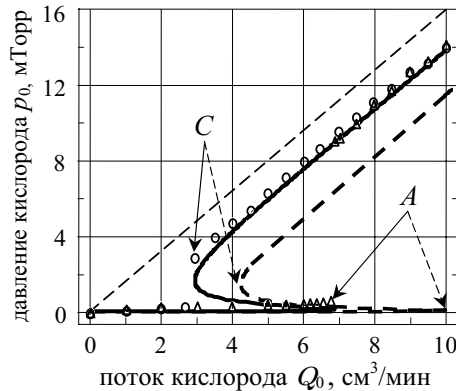


Рис. 4.27. Зависимость стационарного давления кислорода p_0 от его потока Q_0 , подаваемого в реактивную систему $\text{Ta}-\text{O}_2-\text{Ta}_2\text{O}_5$: сплошная кривая — расчет для неіотермических условий ($T_t = 600 \text{ К}$, $T_s = 450 \text{ К}$, $T_w = T_0 = 300 \text{ К}$), штриховая кривая — для изотермических условий ($T_t = T_s = T_w = T_0 = 300 \text{ К}$). Значки Δ и $\circ$ соответствуют экспериментальным точкам, полученным в ходе увеличения Q_0 (металлический режим) и уменьшения Q_0 (оксидный режим)

Результаты расчета показаны на рис. 4.27 в форме сплошной кривой, которая по форме совпадает с аналогичной кривой, качественно изображенной на рис. 4.25.

Различие между рисунками заключается в разной размерности газового потока Q_0 : с^{-1} (число молекул за секунду) — на рис. 4.25 и во всех формулах; $\text{см}^3/\text{мин}$ (объем газа за минуту) — на рис. 4.27. Последнюю величину в англоязычной литературе принято обозначать «sccm» (*standard cubic centimeter per minute*). В качестве *стандартных условий* берут, как обычно, $T^\circ = 298 \text{ К}$ и $P^\circ = 1 \text{ атм}$. Отсюда объем, приходящийся на одну молекулу, равен $V_1^\circ = k_B T^\circ / P^\circ = 4,06 \cdot 10^{-20} \text{ см}^3$. Таким образом, объемный расход газа вычисляется следующим образом:

$$Q (\text{см}^3/\text{мин}) \equiv Q (\text{sccm}) = 60 V_1^\circ Q (\text{с}^{-1}) = 2,436 \cdot 10^{-18} Q (\text{с}^{-1}).$$

Следовательно, на рис. 4.27: $Q_0 = 10 \text{ см}^3/\text{мин} \approx 4,1 \cdot 10^{18} \text{ с}^{-1}$.

На рис. 4.27 также приведена *штриховая кривая*, полученная расчетом для изотермических условий ($T_t = T_s = T_w = T_0 = 300 \text{ K}$) по формулам модели, основанной только на процессе адсорбции *без учета химической реакции*, которая широко используется в литературе большинством авторов. Сравнение этой кривой с экспериментальными точками (Δ — при увеличении потока Q_0 и $\circ$ — при уменьшении потока Q_0) показывает их существенное несовпадение. Так, точкам неустойчивости A и C (границы петли гистерезиса на рис. 4.25) соответствуют следующие значения потока кислорода (см. рис. 4.27):

$$\begin{cases} Q_0(A) = 10 \text{ см}^3/\text{мин} & \text{— штриховая кривая} \\ Q_0(C) = 4,2 \text{ см}^3/\text{мин} & \text{(штриховые стрелки);} \\ Q_0(A) = 6,7 \text{ см}^3/\text{мин} & \text{— эксперимент. точки} \\ Q_0(C) = 2,9 \text{ см}^3/\text{мин} & \text{(сплошные стрелки).} \end{cases}$$

Это свидетельствует о некорректности *изотермической адсорбционной модели* (широко распространенной в литературе по реактивному распылению металлов) в применении к реальным условиям эксперимента.

Во-первых, в распылительных системах температуры мишени, подложки и стенок камеры всегда различные. Во-вторых, учет только адсорбционных процессов на реактивных поверхностях в пренебрежении химической реакцией (как делают многие авторы публикаций) слишком упрощенно описывает физические процессы при реактивном распылении.

Расчет по вышеприведенным формулам *неизотермической химической модели* при указанных выше численных значениях параметров проводился таким образом, чтобы обеспечить по возможности более близкое совпадение результатов расчета (сплошная кривая на рис. 4.27) с экспериментальными данными. Это удалось получить при температурной зависимости константы скорости (закон Аррениуса) для химической реакции (4.16.1) в системе $\text{Ta}-\text{O}_2-\text{Ta}_2\text{O}_5$,

$$k(T) = k_0 \exp(-\mathcal{E}/k_B T), \quad (4.16.38)$$

со следующими значениями параметров: $k_0 = 1,9 \cdot 10^{29} \text{ см}^{-2} \cdot \text{с}^{-1}$ и $\mathcal{E} = 0,2 \text{ эВ}$. При этом полученное выражение (4.16.38) применимо для мишени, подложки и стенок камеры с различными температурами T_i ($i = t, s, w$).

Найденное из расчета положенное точек неустойчивости A и C вдоль оси Q_0 на рис. 4.27 практически совпадает с экспериментальными результатами.

Довольно малая величина энергии химической активации (0,2 эВ), полученная из расчета, может быть объяснена тем фактом, что в действительности «распылительная» реакция является плазмохимической, а не химической, как рассматривалось выше. Участие ионов газовой плазмы в реальных явлениях на поверхности вполне понятным образом понижает *эффективную энергию активации* результирующего процесса, который описывался нами на языке атомарно-молекулярных (а не ионных!) взаимодействий. Последнее естественно упрощает физическую картину явлений, но дает существенное продвижение по сравнению с широко используемой изотермической адсорбционной моделью.

4.17. Кинетика процессов в электрохимических системах

Для технологических применений интерес представляют процессы электрохимического травления и осаждения вещества в режиме электролиза, когда материал анода растворяется, а на катоде выделяется осаждаемое вещество. Отличительная особенность неравновесных процессов в таких системах связана с протеканием электрического тока, вызванного подключением внешнего источника к электродам электролитической ячейки.

Согласно первому закону Фарадея, масса выделившегося (или удаленного) вещества прямо пропорциональна количеству электричества, прошедшего через электролит. Поэтому кинетика ионных процессов и скорость электрохимических реакций на электродах определяются током в цепи электролитической ячейки. Следовательно, изучение неравновесных процессов сводится к анализу так называемой поляризационной кривой отдельных электродов и вольт-амперной характеристики (ВАХ) электролитической ячейки в целом.

Термодинамические особенности электрохимических систем и равновесные электродные процессы были ранее изучены (см. пп. 2.18–2.20). Как известно, при бесконечно малом токе реализуется электрохимическое равновесие. В этом случае анод и катод в растворе электролита принимают, в соответствии с уравнением Нернста, определенные электродные потенциалы, разность которых дает обратимую ЭДС ΔE элемента. Следовательно, в отсутствие тока внешняя ЭДС $E_{\text{вн}}$ уравнивает внутреннюю ЭДС, так что $E_{\text{вн}} = \Delta E$. Тот факт, что для обеспечения бесконечно малого (в пределе нулевого) тока требуется конечная величина ЭДС $E_{\text{вн}}$, трактуется как *химическая поляризация* электродов с ЭДС ΔE , имеющая место даже в равновесных условиях.

Протекание тока конечной силы I изменяет равновесные электродные потенциалы на некоторую величину η , вызванную неравновесной поляризацией электрода и называемую *электродным перенапряжением*: для анода $\eta_a > 0$ (потенциал электрода выше равновесного), для катода $\eta_k < 0$ (потенциал электрода ниже равновесного). Зависимость электродного перенапряжения от тока в форме $\eta = f(I)$ (или наоборот $I = f(\eta)$) называется *поляризационной кривой* электрода. Суммарная электродная поляризация $E_{\text{пол}}(I) = \eta_a(I) - \eta_k(I) \equiv \eta_a(I) + |\eta_k(I)|$ (неравновесная) вместе с химической поляризацией ΔE (равновесной) и падением напряжения IR на сопротивлении R электролита и электрической цепи уравниваются внешней ЭДС $E_{\text{вн}}$, тогда

$$E \equiv E_{\text{вн}} - \Delta E = E_{\text{пол}}(I) + IR \quad \text{или} \quad I = \frac{E - E_{\text{пол}}(I)}{R}. \quad (4.17.1)$$

Здесь $E = E_{\text{вн}} - \Delta E$ — внешняя ЭДС, отсчитанная от обратимой ЭДС (постоянной для рассматриваемой электролитической ячейки), так что в отсутствие тока $E = 0$.

Сравнение (4.17.1) с аналогичной формулой (4.15.20) показывает их полную идентичность при условии, что $E_{\text{пол}}(I) \equiv U(I)$. Следовательно, для анализа электролитической цепи применима эквивалентная схема, изображенная на рис. 4.24 в, при этом нелинейное сопротивление в схеме $R_{\text{нел}} = U(I)/I \equiv E_{\text{пол}}(I)/I$ отражает нелинейность физического процесса поляризации электродов. Суммарная поляризационная кривая $E_{\text{пол}} = f(I)$ качественно изображена на рис. 4.24 в пунктирной линией $U = f(I)$.

Таким образом, для понимания механизма электрохимической кинетики надо знать поляризационные кривые электродов $\eta_a(i)$ и $\eta_k(i)$, которые обычно измеряют в зависимости от плотности тока $i = I/A$, где A — площадь поверхности электрода.

В основе возникновения электродного перенапряжения, вызванного поляризацией электрода, лежит кинетическая замедленность той или иной стадии электрохимического процесса, среди которых:

- *стадия диффузионной кинетики*, обеспечивающая за счет диффузии (электрического поля или конвекции) транспорт реагирующих частиц к поверхности электрода и отвод от нее продуктов реакции;
- *стадия электрохимической кинетики*, связанная с переносом электронов через границу раздела электролит–электрод при разряде или ионизации частиц на поверхности электрода, называемая также *стадией разряда–ионизации*;

- *стадия фазового превращения*, определяющая зарождение и развитие новой фазы в условиях электролиза.

Та из стадий, которая является наиболее замедленной, то есть лимитирует процесс в целом, и определяет его результирующую скорость, является непосредственной причиной поляризации электрода. Именно медленная стадия и вызывает электродное перенапряжение, названное в соответствии с вышеуказанными стадиями *диффузионным, электрохимическим и фазовым перенапряжением*.

Выведем уравнения, описывающие поляризационную кривую электрода, вызванную стадиями диффузионной и электрохимической кинетики.

1. Диффузионная кинетика управляет электрохимическим процессом в том случае, если лимитирующей стадией является транспортировка участников электродной реакции. Это означает, что все частицы, которые за счет диффузии медленно подводятся к электроду, очень быстро там разряжаются или ионизируются, так что их концентрация на поверхности электрода отличается от ее равновесного значения внутри объема электролита. По этой причине поляризацию, вызванную торможением на стадии диффузионного транспорта, часто называют *концентрационной поляризацией*.

Распределение частиц в районе электрода имеет тот же качественный вид, что и на рис. 4.14. Изображенный ход кривой соответствует анодному процессу, так как растворение материала анода обеспечивает неравенство $c_{si} > c_{0i}$, где c_{si} и c_{0i} — концентрации ионов i -го вещества на поверхности электрода и в объеме раствора. Для процесса катодного осаждения концентрация ионов у электрода меньше, чем вдали от него, то есть $c_{si} < c_{0i}$ (индекс i для упрощения записи ниже опускаем).

Плотность стационарного диффузионного потока J на поверхности электрода дается формулой (4.9.9), тогда плотность электрического тока, $i \equiv qJ$, равна

$$i = z\mathcal{F}\beta(c_s - c_0) = i_0 \left(\frac{c_s}{c_0} - 1 \right), \quad (4.17.2)$$

где $\beta = D/\delta$ — феноменологический коэффициент диффузионного массопереноса для рассматриваемых частиц, $q = z\mathcal{F} \equiv zeN_A$ — заряд одного моля частиц с зарядовым числом z ; при этом концентрация измеряется в мольно-объемных единицах. В последнем равенстве (4.17.2) введена величина $i_0 = z\mathcal{F}\beta c_0$, называемая *предельным током диффузии*.

Даже в условиях медленного подвода и отвода вещества, нарушающего концентрационное равновесие в растворе, можно считать выполненным электродное равновесие, связанное с быстрыми переходами электронов через границу раздела фаз. Это означает, что для определения концентрационной поляризации можно воспользоваться равновесным уравнением Нернста.

В частности, можно применять формулу Нернста (2.19.12) для металлического электрода, процесс разряда–ионизации на котором описывается реакцией (2.19.7). Тогда для электрода, обратимого относительно катиона, равновесный потенциал (помечаемый нижним индексом 0) определяется следующим соотношением (в условиях, когда можно оперировать концентрациями вместо активностей):

$$E_0 = E^\circ + \frac{RT}{z\mathcal{F}} \ln c_0. \quad (4.17.3)$$

Здесь c_0 — концентрация потенциалобразующих ионов в объеме раствора, E° — стандартный потенциал данной системы. При прохождении тока концентрация потенциалобразующих ионов вблизи поверхности электрода принимает значение c_s , отличное от c_0 , что дает новое (неравновесное) значение электродного потенциала (без нижнего индекса 0):

$$E = E^\circ + \frac{RT}{z\mathcal{F}} \ln c_s. \quad (4.17.4)$$

Таким образом, поляризация электрода (электродное перенапряжение) на стадии диффузионной кинетики, определяемая разностью величин (4.17.4) и (4.17.3), выражается в виде

$$\eta \equiv E - E_0 = \frac{RT}{z\mathcal{F}} \ln \frac{c_s}{c_0}. \quad (4.17.5)$$

Подстановка в формулу (4.17.2) отношения c_s/c_0 , найденного из равенства (4.17.5), дает искомое уравнение для поляризационной кривой:

$$i(\eta) = i_0 \left(\exp \frac{ze\eta}{k_B T} - 1 \right). \quad (4.17.6)$$

Это уравнение справедливо для металлического электрода, обратимого относительно катиона, на котором в диффузионном режиме происходит разряд–ионизация собственных ионов металла, например, ионов Ag^+ в растворе AgNO_3 .

На рис. 4.28 изображены две ветви поляризационной кривой (4.17.6), соответствующие разряду ионов (катодная ветвь $i_k(\eta)$ при $\eta < 0$) и ионизации атомов (анодная ветвь $i_a(\eta)$ при $\eta > 0$)

на одноименном металле в условиях лимитирующей диффузии в растворе электролита.

Отрицательным значениям перенапряжения $\eta < 0$ соответствует процесс катодного осаждения вещества, при котором область у электрода обедняется ионами (вплоть до $c_s \rightarrow 0$) при большой отрицательной поляризации электрода, когда $z|\eta| \gg \gg k_B T/e \approx 26$ мВ при $T = 298$ К. В этом пределе экспонентой в формуле (4.17.6) можно пренебречь по сравнению с единицей, так что катодный ток (отрицательный) равен предельному току диффузии, $i_0 = z\mathcal{F}\beta c_0$.

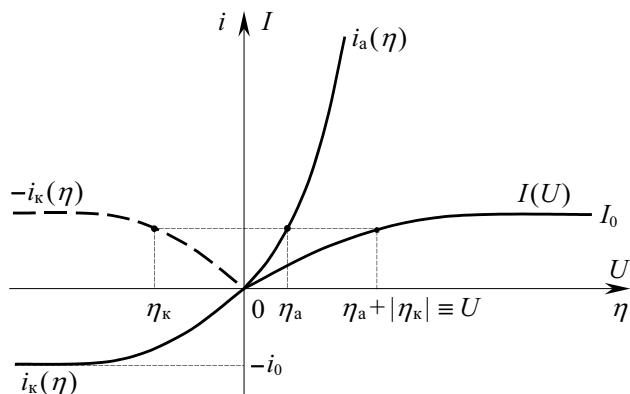


Рис. 4.28. Катодная (i_k) и анодная (i_a) ветви поляризационной кривой $i(\eta)$, соответствующие разряду ионов и ионизации атомов на одноименном металлическом электроде в условиях замедленной стадии диффузионного массопереноса. Показана схема построения ВАХ $I(U)$ для электролитической ячейки с одинаковыми металлическими электродами

Положительные перенапряжения $\eta > 0$ соответствуют анодному процессу, при этом плотность тока, а значит, и скорость растворения материала, экспоненциально возрастает с ростом η . Это вызвано тем, что по мере растворения вещества концентрация ионов металла у поверхности растворяемого электрода в принципе может возрастать до довольно высоких значений, однако, не превышающих произведение растворимости соли этого металла. При достижении этого значения выпадает осадок соли, который экранирует электрод, и анодный ток падает. С ростом концентрации ионов возможна смена лимитирующей стадии электрохимического процесса, так что его скорость уже не будет определяться диффузией.

Более того, в электролитической ячейке с одинаковыми металлическими электродами протекающий ток принципиально не

может превышать предельный ток диффузии. Действительно, в последовательной цепи анод – электролит – катод течет ток одного направления, совпадающий, например, с током анода $I = i_a A$ (считаем площади анода и катода одинаковыми и равными A).

В этом случае надо изменить знак тока катода, поскольку $i_k = -i_a$, что графически сделано на рис. 4.28 в виде штриховой кривой $-i_k(\eta)$, зеркально симметричной по отношению к катодной ветви $i_k(\eta)$. Как видно из графического построения на рис. 4.28, плотности тока i_a соответствуют разные катодная ($\eta_k < 0$) и анодная ($\eta_a > 0$) поляризации. Суммарная электродная поляризация $E_{\text{пол}}(i_a) = \eta_a(i_a) - \eta_k(i_a) \equiv \eta_a(i_a) + |\eta_k(i_a)|$ дает полное падение напряжения U на электролитической ячейке (при одинаковых электродах обратимая ЭДС $\Delta E = 0$). Построенная таким способом ВАХ ячейки имеет форму кривой $I(U)$, изображенной на рис. 4.28, с током насыщения, равным предельному току диффузии, $I_0 = i_0 A = zF\beta c_0 A$.

Полезно отметить, что уравнение поляризационной кривой (4.17.6) по форме в точности совпадает с ВАХ полупроводникового диода на $p-n$ -переходе, что неудивительно, так как в обоих случаях транспортные процессы управляются диффузией.

2. Электрохимическая кинетика управляет процессом электролиза в случае, если лимитирующей стадией является замедленная стадия разряда-ионизации, вызванная медленным переходом заряженных частиц через границу между электродом и раствором электролита.

Рассмотрим кинетику редокс-процесса на электроде, когда в прямом направлении окисленная форма Ox принимает z электронов и превращается в восстановленную форму Red по окислительно-восстановительной реакции (2.19.5):



Так как в рассматриваемом случае предполагается, что электронные процессы на электроде протекают намного медленнее транспортных процессов в объеме электролита, то последние обеспечивают на поверхности электрода концентрацию реагирующих частиц c_{Ox} и c_{Red} , близкую к объемным значениям.

По аналогии с уравнениями (4.1.2) записываем скорости электрохимической реакции (4.17.7), протекающей в прямом направлении (v_1) и в обратном направлении (v_2), в следующей форме (в единицах плотности тока):

$$i_1 \equiv qv_1 = zFk_1 c_{\text{Ox}} \quad \text{и} \quad i_2 \equiv qv_2 = zFk_2 c_{\text{Red}}. \quad (4.17.8)$$

Здесь электрохимические константы скорости k_1 и k_2 , в отличие от обычных химических реакций, зависят от неравновесного потенциала электрода, $E = E_0 + \eta$, где η — электрохимическое перенапряжение, называемое также *перенапряжением разряда*.

Если, как показано на рис. 4.2, при $E = 0$ энергии активации для прямой и обратной реакций, соответственно, равны U'_1 и U'_2 , то ненулевой потенциал электрода ($E \neq 0$) изменяет высоту активационных барьеров для частицы с зарядом $q = z\mathcal{F}$ на величину, пропорциональную энергии, $qE = z\mathcal{F}E$. На этом основании можно записать закон Аррениуса (1.18.17) (в котором не путать энергию активации $\mathcal{E}$ с потенциалом электрода E) для констант скорости прямой и обратной реакции в следующей форме:

$$k_1 = k_{10} \exp\left(-\frac{U'_1 + \alpha z\mathcal{F}E}{RT}\right) \equiv k_1^\circ \exp\left(-\alpha \frac{z\mathcal{F}E}{RT}\right), \quad (4.17.9)$$

$$k_2 = k_{20} \exp\left(-\frac{U'_2 - \beta z\mathcal{F}E}{RT}\right) \equiv k_2^\circ \exp\left(\beta \frac{z\mathcal{F}E}{RT}\right). \quad (4.17.10)$$

Разные знаки перед коэффициентами пропорциональности α и β вызваны тем, что по реакции (4.17.7) заряд окисленной формы Ox уменьшается на величину ze , а восстановленной формы Red увеличивается на ту же величину.

В условиях электрохимического равновесия скорости прямой и обратной реакции становятся одинаковыми ($i_1 = i_2 \equiv i_0$ при $E = E_0$), давая с учетом формул (4.17.8)–(4.17.10) так называемый *ток обмена*:

$$\begin{aligned} i_0 &\equiv z\mathcal{F}k_1^\circ c_{\text{Ox}} \exp\left(-\alpha \frac{z\mathcal{F}E_0}{RT}\right) = \\ &= z\mathcal{F}k_2^\circ c_{\text{Red}} \exp\left(\beta \frac{z\mathcal{F}E_0}{RT}\right). \end{aligned} \quad (4.17.11)$$

Решение уравнения (4.17.11) относительно равновесного потенциала E_0 дает величину

$$E_0 = \frac{RT}{(\alpha + \beta)z\mathcal{F}} \ln K_c - \frac{RT}{(\alpha + \beta)z\mathcal{F}} \ln \frac{c_{\text{Red}}}{c_{\text{Ox}}}, \quad (4.17.12)$$

где, согласно (4.1.4), введена константа равновесия $K_c = k_1^\circ/k_2^\circ$. С другой стороны, по уравнению Нернста для редокс-процесса (4.17.7) имеем

$$E_0 = E^\circ - \frac{RT}{z\mathcal{F}} \ln \frac{c_{\text{Red}}}{c_{\text{Ox}}}. \quad (4.17.13)$$

Из сравнения (4.17.12) и (4.17.13) следует, что $\alpha + \beta = 1$ и $E^\circ = (RT/z\mathcal{F}) \ln K_c$. Подставляя выражения (4.17.9) и (4.17.10)

в равенства (4.17.8) с учетом (4.17.11), получаем следующие выражения для плотности прямого и обратного токов:

$$i_1 = i_0 \exp\left(-\alpha \frac{z\mathcal{F}(E-E_0)}{RT}\right) \equiv i_0 \exp\left(-\alpha \frac{ze\eta}{k_B T}\right), \quad (4.17.14)$$

$$i_2 = i_0 \exp\left((1-\alpha) \frac{z\mathcal{F}(E-E_0)}{RT}\right) \equiv i_0 \exp\left((1-\alpha) \frac{ze\eta}{k_B T}\right). \quad (4.17.15)$$

Определяя результирующий ток в форме $i \equiv i_2 - i_1$, делаящей его положительным для анодного процесса окисления, из выражений (4.17.14) и (4.17.15) получаем искомое уравнение поляризационной кривой $i = f(\eta)$ для редокс-электрода, работающего в режиме замедленного разряда-ионизации:

$$i(\eta) = i_0 \left[\exp\left((1-\alpha) \frac{ze\eta}{k_B T}\right) - \exp\left(-\alpha \frac{ze\eta}{k_B T}\right) \right]. \quad (4.17.16)$$

Коэффициент α , называемый *коэффициентом электронного переноса*, рассматривается, как и ток обмена i_0 , в качестве феноменологического параметра задачи.

Графики поляризационных кривых, рассчитанных по формуле (4.17.16), в нормированном виде приведены на рис. 4.29 для шести значений коэффициента α , взятых в диапазоне от 0 до 0,5. Значениям α от 0,5 до 1 отвечают значения $(1 - \alpha)$, лежащие в диапазоне от 0,5 до 0, т. е. соответствующие им кривые получаются инверсией кривых, изображенных на рис. 4.29 (путем выполнения следующих замен: $i \rightarrow -i$ и $\eta \rightarrow -\eta$).

Рассмотрим отдельные частные случаи формулы (4.17.16) и соответствующие им кривые на рис. 4.29.

1. При нулевом коэффициенте электронного переноса ($\alpha = 0$) уравнение (4.17.16) принимает форму,

$$i(\eta) = i_0 \left(\exp \frac{ze\eta}{k_B T} - 1 \right), \quad (4.17.17)$$

совпадающую по виду с уравнением (4.17.6) для стадии диффузионного массопереноса. Соответствующая этому случаю кривая под номером 0 на рис. 4.29 имеет наибольшую асимметрию относительно равновесной точки $\eta = 0$, при этом обратная (катодная) ветвь при $z|\eta| \gg k_B T/e$ обеспечивает ток насыщения, равный току обмена i_0 .

2. При $\alpha = 0,5$ уравнение (4.17.16) преобразуется к следующему виду:

$$i(\eta) = 2i_0 \operatorname{sh} \frac{ze\eta}{2k_B T}, \quad (4.17.18)$$

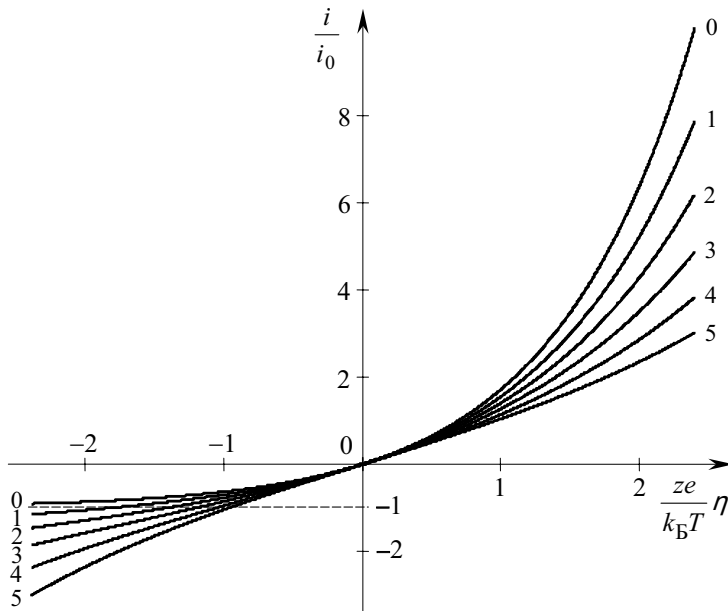


Рис. 4.29. Поляризационные кривые для редокс-электрода в условиях замедленной стадии разряда-ионизации, соответствующие разным значениям коэффициента электронного переноса: $\alpha = 0$ (кривая 0), $\alpha = 0,1$ (кривая 1), $\alpha = 0,2$ (кривая 2), $\alpha = 0,3$ (кривая 3), $\alpha = 0,4$ (кривая 4), $\alpha = 0,5$ (кривая 5)

которому соответствует кривая 5 на рис. 4.29, симметрично расположенная относительно начала координат при $\eta = 0$.

3. При малых перенапряжениях, когда $z|\eta| \ll k_B T/e = 26$ мВ, экспоненты в уравнении (4.17.16) могут быть разложены в ряд Тейлора. Тогда в линейном приближении

$$\eta(i) = \frac{k_B T/ze}{i_0} i \equiv R_e i, \quad (4.17.19)$$

где $R_e = (k_B T/ze)/i_0$ — сопротивление электролита при электрохимической поляризации электрода.

4. При большом отклонении от состояния равновесия, когда $z|\eta| \gg k_B T/e = 26$ мВ, одна из экспонент в уравнении (4.17.16) становится пренебрежимо малой, в частности,

$$i(\eta) = i_0 \exp\left((1-\alpha)\frac{ze\eta}{k_B T}\right) \quad \text{или} \quad \eta(i) = a + b \lg i. \quad (4.17.20)$$

Последнее выражение в (4.17.20) известно как *формула Тафеля* с эмпирическими параметрами $a = -[k_B T/(1-\alpha)ze] \ln i_0$

и $b = 2,3[k_B T / (1 - \alpha)ze]$, которые позволяют находить величины i_0 и α из экспериментальных кривых $\eta(i)$.

Особенностью стадии разряда–ионизации является тот факт, что скорость перехода электронов через границу раздела фаз зависит от строения двойного электрического слоя на этой границе, что существенно усложняет анализ этой стадии электрохимического процесса. Еще большей сложностью отличается смешанная кинетика процесса, протекающего при сравнимых скоростях диффузионной и электрохимической стадий, усугубленная влиянием фазового перенапряжения.

4.18. Контрольные вопросы

1. Вспомните общие условия фазового равновесия и поясните, что происходит в случае их нарушения.
2. Вспомните общее условие химического равновесия и поясните, к чему приводит его нарушение.
3. Сформулируйте основной постулат химической кинетики и приведите обоснование закона Аррениуса для обратимой химической реакции.
4. Как связаны константы скорости прямой и обратной реакции с константой равновесия?
5. Опишите основные механизмы диффузии атомов в твердом теле и дайте качественное обоснование закона Аррениуса для коэффициента диффузии.
6. Сформулируйте диффузионные законы Фика и граничные условия для задач диффузии.
7. Обоснуйте физическую постановку задачи диффузионного удаления растворенного газа из твердого тела.
8. В чем заключается конечная цель вакуумного обезгаживания металлической пластины и стеклянной оболочки.
9. Опишите практические принципы и приемы вакуумного обезгаживания материалов.
10. Какие механизмы газопроницаемости реализуются для металлических и силикатных вакуумных оболочек?
11. Обсудите физическую постановку и конечное решение задачи легирования полупроводника в рамках модели диффузии из постоянного источника.

12. Обсудите физическую постановку и конечное решение задачи легирования полупроводника в рамках модели диффузии из полуограниченного источника.
13. Обсудите физическую постановку и конечное решение задачи легирования полупроводника в рамках модели диффузии из поверхностного источника с отражающей границей.
14. Опишите практические принципы и приемы диффузионного легирования полупроводников.
15. Какие физические принципы положены в основу определения диффузионной и кинетической областей протекания многостадийного гетерогенного процесса?
16. Дайте обоснование введению диффузионного слоя, коэффициента массопереноса и стефановского потока на химически реакционной границе раздела фаз.
17. Какова физическая причина появления стефановского потока в газовой фазе?
18. В чем выражается маскирующая функция двуокиси кремния в планарной кремниевой технологии?
19. Каков физический смысл маскирующей толщины слоя двуокиси кремния?
20. Перечислите последовательные стадии термического окисления кремния и укажите их лимитирующее влияние при осуществлении процесса окисления в кинетической и диффузионной областях.
21. Перечислите последовательные стадии химического травления полупроводников и укажите их лимитирующее влияние при осуществлении процесса травления в кинетической и диффузионной областях.
22. Какая из областей процесса травления (диффузионная или кинетическая) обеспечивает режим полирующего травления и режим селективного анизотропного травления?
23. Опишите принципы выращивания монокристаллических слоев методами жидкофазной эпитаксии.
24. Опишите принципы выращивания монокристаллических слоев методами газофазной эпитаксии.
25. Опишите особенности и области применения химической сборки поверхностных наноструктур.

26. Опишите особенности и области применения молекулярно-лучевой эпитаксии, включая методы ТЛЭ и ХЛЭ.
27. Каковы особенности кинетики химического транспорта веществ в проточных системах?
28. Каковы особенности кинетики химического транспорта веществ в сэндвич-системах?
29. Поясните эквивалентную схему газотранспортного процесса в сэндвич-системе, изображенную на рис. 4.24 *з*, и обоснуйте с ее помощью появление диффузионной и кинетической области процесса кристаллизации.
30. Объясните физическую причину появления гистерезисного эффекта при реактивном распылении металлов.
31. Каковы особенности металлического и реактивного режимов ионного распыления металлов?
32. Какие физические процессы определяют кинетику реактивного распыления?
33. Перечислите последовательные стадии кинетики процессов в электрохимических системах.
34. Укажите особенности процессов при диффузионной и электрохимической кинетике.
35. Какова природа поляризации электродов и электродного перенапряжения в электрохимических системах?

Глава 5

УПРАВЛЕНИЕ СВОЙСТВАМИ ПОВЕРХНОСТИ, МЕЖФАЗНЫМИ ВЗАИМОДЕЙСТВИЯМИ И ФОРМИРОВАНИЕМ НАНООБЪЕКТОВ

В предыдущих главах любая граница раздела фаз рассматривалась как чисто геометрический объект, не обладающий самостоятельными термодинамическими свойствами, в частности, не имеющий собственного запаса энергии. Именно по этой причине к макроскопическим системам было применено понятие квази-замкнутости, порождающее аддитивность физических характеристик, таких как внутренняя энергия, энтропия, энтальпия и др.

Действительно, взаимодействие любой макросистемы со своим внешним окружением осуществляется с преимущественным участием тех частиц, которые находятся в ее приповерхностных слоях. Обычно количество таких частиц, которые будем называть *поверхностными*, сравнительно невелико по отношению к числу *объемных* частиц, т. е. расположенных внутри объема, занятого системой. Тогда относительный вклад поверхностных частиц в общую энергию системы также невелик и быстро уменьшается с увеличением ее размеров. Следовательно, для системы больших размеров энергия ее взаимодействия с внешним окружением пренебрежимо мала по сравнению с внутренней энергией. Именно это служит основанием для того, чтобы полагать энергетические характеристики системы аддитивно составленными из вкладов входящих в нее подсистем.

На основании вышесказанного можно заключить, что свойство аддитивности экстенсивных характеристик системы начинает нарушаться при уменьшении размеров отдельных ее частей, когда становятся сравнимыми количества объемных и поверхностных частиц. Межфазной границей обычно называют слой вещества, разделяющий две смежные объемные фазы, толщиной в несколько межатомных расстояний.

В течение многих лет после разработки основ квантовой теории традиционными объектами физического изучения являлись, с одной стороны, объекты классической физики — *макроскопические тела* (в разных агрегатных и структурных состояниях) с размерами милли-, санти- и метрового диапазона, а с другой стороны, объекты квантовой физики — *микрочастицы* (атомы, ионы, молекулы) с размерами в десятые доли и единицы ангстрема ($1 \text{ \AA} = 0,1 \text{ нм} = 10^{-4} \text{ мкм} = 10^{-8} \text{ см}$).

Современная физика заполняет нишу между макроскопическими и микроскопическими состояниями вещества так называемым *мезоскопическим состоянием*. Соответствующий размерный диапазон (от долей миллиметра до единиц нанометра) охватывает огромное количество разнообразных по своим проявлениям физических процессов и явлений. Поэтому принято *условно* разделять мезоскопическую область на следующие диапазоны:

- *субмиллиметровый диапазон* — от 1 мм до 0,1 мм;
- *микрометровый диапазон* — от 100 мкм до 1 мкм;
- *субмикрометровый диапазон* — от 1 мкм до 0,1 мкм;
- *нанометровый диапазон* — от 100 нм до 1 нм, который при-
мыкает вплотную к молекулярно-атомному диапазону с мас-
штабами в доли нанометра (см. рис. 1 во введении).

Условность классификации по размерному признаку связана с выбором масштаба измеряемых явлений, каждое из которых имеет свои *характерные длины* (такие как длина свободного пробега частиц/квазичастиц, диффузионная длина, длина волны де Бройля для микрочастиц и т. п.). Более того, характерная длина одного и того же процесса в зависимости от физических условий может изменяться в очень широких пределах, что существенно видоизменяет физическую картину явления и способы его описания. Так, длина свободного пробега газовых молекул изменяется примерно от 50 нм в атмосферных условиях до 50 км в условиях сверхвысокого вакуума (при давлении ниже 10^{-9} мм рт.ст.).

Современная физическая триада «макро–мезо–микро» (от греческих слов *большой–средний–малый*) становится объединяющей и общей для ранее независимо развивающихся направлений естествознания (таких как физика, химия, биология), в силу схожести основополагающего поведения наноразмерных объектов различной природы (физической, химической, биологической).

Мезоскопические объекты микрометрового и, особенно, нанометрового диапазона существенно отличаются по характеру проявления физических и химических свойств как от своих макроскопических аналогов, так и от микроскопических прародителей.

Для макрообъектов, состоящих из колоссального количества образующих их частиц, физические свойства проявляются в форме непрерывных (плавных) характеристик как результат *статистического усреднения*. Они описывают свойства объектов в целом и практически не зависят от их макроразмеров.

В отличие от этого, квантовая физика считает объекты одного типа (электроны, атомы одного элемента, идентичные молекулы) абсолютно одинаковыми и статистически неразличимыми. На этом основании физически наблюдаемые характеристики микроскопических объектов обычно проявляются в форме дискретных изменений той или иной величины как результат *квантования*.

Мезообъекты, которые занимают промежуточное место между макро- и микрообъектами, *во-первых*, могут наследовать, хотя бы частично, физико-химические свойства тех и других, которые теперь уже должны зависеть от количества входящих в них микрочастиц. *Во-вторых*, этого количества может оказаться недостаточно для статистического формирования коллективных свойств мезообъекта, но в то же время избыточно для проявления индивидуальных свойств микрочастиц. *Первая* особенность служит основанием для применения существующих методов описания макро- и микроявлений к миру наноразмерного масштаба; тогда *вторая* свидетельствует о неполноте подобного описания и о необходимости разработки адекватных альтернативных подходов, отличных от существующих.

На основании вышесказанного, к объектам наноэлектроники будем относить любые твердотельные структуры, размер которых хотя бы в одном направлении соответствует принятому нанометровому диапазону (от 1 нм до 100 нм):

- *2D-объекты* (двумерные) нанометровой толщины (нанослой, нанопленки, квантовые ямы);
- *1D-объекты* (одномерные) нанометрового сечения (нанотрубки, нановолокна, нанопроволоки, квантовые нити);
- *0D-объекты* (нульмерные) с размерами до 100 нм по всем направлениям (наночастицы, фуллерены, квантовые точки). При этом наночастицы могут состоять из более мелких частей с размерами до 10 нм, называемых *кластерами* (от англ. *cluster* — скопление взаимодействующих друг с другом однородных частиц атомно-молекулярных размеров). Отличительной чертой кластеров является немонотонная (дискретная) зависимость свойств от количества (до 10^3) атомов или молекул, входящих в их состав.

Объекты, имеющие линейные размеры по всем направлениям свыше 100 нм, принято называть *3D-объектами* (трехмерными).

По сути дела они попадают в субмикрометровый/микрометровый диапазон и являются объектами современной микроэлектроники.

В последние годы характерные линейные размеры элементов микроэлектроники приблизились к субмикрометровому диапазону и лежат в пределах от единиц до десятых долей микрона. Однако уже первые попытки практического применения структур с размерами менее 0,1 мкм (или 100 нм) показали, что их поведение и свойства качественно отличаются от поведения структур больших размеров. Особенно ярко это проявляется, когда размеры структурных элементов приближаются к десяти нанометрам. Для подобного рода систем, содержащих, например зародыши новой фазы, адсорбированные поверхностью частицы, ультрадисперсные частицы в коллоидных средах и т. п., основные положения термодинамики подлежат пересмотру с целью учета энергетического вклада поверхности и межфазных границ.

Задача настоящей главы — дать общие представления о механизмах формирования нанообъектов и наноструктурированных материалов на их основе с учетом термодинамических и кинетических особенностей тех поверхностных и межфазных взаимодействий, которые играют принципиальную роль в технологических проблемах микро- и нанoeлектроники.

5.1. Квантовые особенности нанообъектов пониженной размерности

Физические и химические свойства массивных (объемных) материалов, состоящих из огромного числа микрочастиц, проявляются в виде плавно меняющихся (в пределах одного агрегатного состояния) зависимостей их свойств от разнообразных внешних воздействий (температуры, давления, электрического и магнитного полей и др., в том числе и от химического состава), которые, однако, не зависят от размеров макрообразца. Такая картина сохраняется неизменной при уменьшении размеров лишь до некоторых минимальных значений, сравнимых с длинами, характерными для того или иного физического явления (например, с длиной волны де Бройля для электронных явлений). Начиная с таких размеров (обычно нанометрового или даже микрометрового масштаба), возникают так называемые *размерные эффекты*.

Размерные эффекты проявляются в том, что для образцов, содержащих еще достаточно большое число составляющих их частиц (вплоть до 10^3 – 10^6 – 10^9), ранее плавно меняющиеся кривые приобретают скачкообразный или даже дискретный характер. Как известно, последнее в принципе присуще частицам

микромира (электронам, атомам, молекулам) и обычно наблюдаемо в форме разного рода так называемых линейчатых спектров.

Качественное обоснование размерных эффектов в форме дискретного (или скачкообразного) характера физических свойств нанокластеров наиболее просто и понятно можно дать, исходя из двух противоположных позиций (они, как правило, лежат в основе большинства наблюдаемых свойств нанообъектов):

1) из рассмотрения размерных эффектов *электронной природы* путем анализа плотности электронных состояний в структурах разной размерности ($3D$, $2D$, $1D$, $0D$);

2) из рассмотрения размерных эффектов *атомной природы* путем структурно-геометрического анализа роли поверхности в нанокластерах сферической и сфероидальной формы.

Первые принято называть *квантово-размерными эффектами*, а вторые по аналогии назовем *структурно-размерными эффектами*. Ниже в этом параграфе проанализируем квантово-размерные эффекты для электронных явлений, а структурно-геометрическое обоснование масс-спектров нанокластеров и описание структурно-размерных эффектов будут выполнены в следующем параграфе.

Как известно, физические свойства любой электронной системы определяются двумя факторами: а) функцией статистического распределения электронов по квантовым состояниям (функцией Ферми–Дирака $f(\varepsilon)$); б) плотностью этих состояний в пространстве энергии ε (энергетической плотностью $g(\varepsilon)$).

В силу универсальности функции $f(\varepsilon)$ электронные свойства конкретной структуры зависят лишь от специфики плотности состояний $g(\varepsilon)$. Она определяется законом дисперсии $\varepsilon(p)$, связывающим энергию электрона ε с его импульсом/квазиимпульсом, $p \equiv \hbar k$ (где k — волновое число электронной волны, в частности волны де Бройля для свободного электрона).

Рассмотрим влияние пространственной локализации электрона с эффективной массой m^* на плотность электронных состояний для простейшего случая изотропного параболического закона дисперсии поочередно в $3D$ -, $2D$ -, $1D$ - и $0D$ -структурах.

1. 3D-структура (объемный образец). В этом случае *отсутствует* пространственная локализация электрона, поскольку считаем, что он находится в образце с размерами (L_x, L_y, L_z) , существенно превышающими длину волны де Бройля, т. е. ведет себя как квазисвободная частица с законом дисперсии

$$\varepsilon(k) = \frac{\hbar^2 k^2}{2m^*} \equiv \frac{\hbar^2}{2m^*} (k_x^2 + k_y^2 + k_z^2). \quad (5.1.1)$$

Отсюда получаем необходимые для дальнейшего равенства:

$$k = \frac{1}{\hbar} \sqrt{2m^*\varepsilon} \quad \text{и} \quad dk = \frac{m^*}{\hbar} \frac{d\varepsilon}{\sqrt{2m^*\varepsilon}}. \quad (5.1.2)$$

Формула для $k \equiv 2\pi/\lambda$ устанавливает связь между длиной волны де Бройля и энергией свободного электрона (при $m^* = m$):

$$\lambda = \frac{2\pi\hbar}{\sqrt{2m\varepsilon}} \approx \frac{1,226}{\sqrt{\varepsilon[\text{эВ}]} } [\text{нм}]. \quad (5.1.3)$$

Известно, что, в соответствии с принципом неопределенности, для размера L_x вдоль направления x каждому разрешенному энергетическому состоянию в $\mathbf{k}$ -пространстве соответствует элемент длины $2\pi/L_x$; аналогично для двух других направлений имеем $2\pi/L_y$ и $2\pi/L_z$. Согласно принципу Паули, в каждом из таких состояний находится по два электрона с противоположно направленными спинами.

Найдем плотность электронных состояний для куба со сторонами $L_x = L_y = L_z \equiv L$, которому в трехмерном $\mathbf{k}$ -пространстве соответствует ячейка объемом $(2\pi/L)^3$, занятая двумя электронами. Объем сферического слоя радиусом k и толщиной dk равен $4\pi k^2 dk$. Отсюда число состояний dG_3 в этом слое получаем делением его объема на объем, приходящийся на одно состояние, с учетом двух разрешенных направлений спина:

$$dG_3 = 2 \frac{4\pi k^2 dk}{(2\pi/L)^3} = \frac{L^3}{2\pi^2} \left(\frac{2m^*}{\hbar^2} \right)^{3/2} \sqrt{\varepsilon} d\varepsilon. \quad (5.1.4)$$

Здесь переход ко второму равенству осуществлен с использованием выражений (5.1.2).

Трехмерная плотность электронных состояний вводится в расчете на единицу объема образца. Тогда из (5.1.4) следует искомое трехмерное выражение (с размерностью $\text{эВ}^{-1} \cdot \text{м}^{-3}$):

$$g_3(\varepsilon) \equiv \frac{1}{L^3} \frac{dG_3}{d\varepsilon} = \frac{1}{2\pi^2} \left(\frac{2m^*}{\hbar^2} \right)^{3/2} \sqrt{\varepsilon}. \quad (5.1.5)$$

Это хорошо известный результат для свободного электронного газа с плотностью состояний, пропорциональной квадратному корню из энергии, как показано штриховой линией на рис. 5.1 а.

2. 2D-структура (квантовая яма). В этом случае имеет место пространственная локализация электрона только в одном измерении (пусть вдоль направления оси z), так как считаем, что в двух других измерениях размеры образца (L_x, L_y) по-прежнему

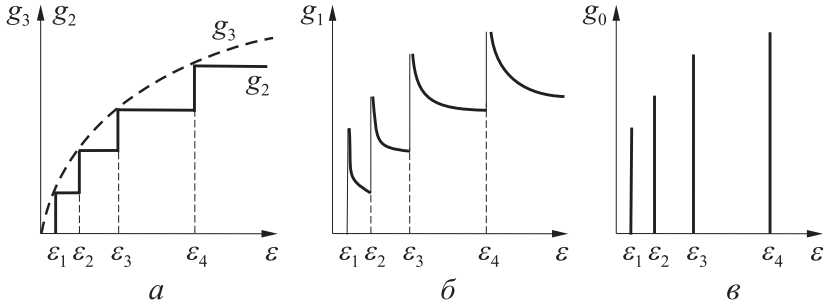


Рис. 5.1. Энергетическая зависимость плотности электронных состояний: а — 3D объем (штриховая плавная кривая) и 2D квантовая яма (сплошная ступенчатая кривая); б — 1D квантовая нить; в — 0D квантовая точка

существенно превышают длину волны де Бройля. Тогда, исходя из выражения (5.1.1), закон дисперсии записываем в виде

$$\varepsilon(k) = \varepsilon_n + \frac{\hbar^2 k^2}{2m^*} \equiv \varepsilon_n + \frac{\hbar^2}{2m^*} (k_x^2 + k_y^2). \quad (5.1.6)$$

Здесь учтено, что вдоль направления локализации имеет место квантование энергии ε_n ($n = 1, 2, \dots$). Отсюда для волнового числа $k \equiv (k_x^2 + k_y^2)^{1/2}$ следуют выражения, аналогичные (5.1.2):

$$k = \frac{1}{\hbar} \sqrt{2m^*(\varepsilon - \varepsilon_n)} \quad \text{и} \quad dk = \frac{m^*}{\hbar} \frac{d\varepsilon}{\sqrt{2m^*(\varepsilon - \varepsilon_n)}}. \quad (5.1.7)$$

В данном случае подсчет числа состояний проводим для кольцевого слоя радиусом k и толщиной dk , имеющего площадь $2\pi k dk$. В двумерном $\mathbf{k}$ -пространстве каждое разрешенное энергетическое состояние занимает площадь $(2\pi/L)^2$. Тогда число состояний dG_2 в кольцевом слое, с учетом выражений (5.1.7), равно

$$dG_2 = 2 \frac{2\pi k dk}{(2\pi/L)^2} = L^2 \frac{m^*}{\pi \hbar^2} d\varepsilon. \quad (5.1.8)$$

Двумерная плотность электронных состояний вводится в расчете на единицу площади квантовой ямы с суммированием по всем дискретным уровням ($n = 1, 2, \dots$), энергия которых (ε_n) лежит ниже данного значения ε . Тогда из формулы (5.1.8) следует искомое двумерное выражение (с размерностью эВ $^{-1} \cdot \text{м}^{-2}$):

$$g_2(\varepsilon) \equiv \frac{1}{L^2} \sum_n \frac{dG_2}{d\varepsilon} H(\varepsilon - \varepsilon_n) = \frac{m^*}{\pi \hbar^2} \sum_n H(\varepsilon - \varepsilon_n), \quad (5.1.9)$$

где использована функция Хевисайда, определенная как

$$H(\varepsilon - \varepsilon_n) = \begin{cases} 0 & \text{при } \varepsilon < \varepsilon_n, \\ 1 & \text{при } \varepsilon > \varepsilon_n. \end{cases}$$

Таким образом, в квантовой яме (с одномерной локализацией электронов) плотность состояний для каждого квантового уровня ε_n не зависит от энергии, так что результирующая кривая $g_2(\varepsilon)$ имеет ступенчатую форму, показанную сплошной линией на рис. 5.1 а.

3. 1D-структура (квантовая нить). В этом случае имеет место пространственная локализация электрона в двух измерениях (вдоль направлений y и z). Вдоль оси нити по координате x его движение остается свободным, так как размер L_x существенно превышает длину волны де Бройля. Тогда, по аналогии с выражением (5.1.6), закон дисперсии записываем в виде

$$\varepsilon(k) = \varepsilon_n + \frac{\hbar^2 k^2}{2m^*} \equiv \varepsilon_n + \frac{\hbar^2}{2m^*} k_x^2. \quad (5.1.10)$$

Как и в 2D-структуре, здесь учтено, что в поперечном сечении нити имеет место квантование энергии ε_n ($n = 1, 2, \dots$). Тогда для волнового числа $k \equiv k_x$ справедливы выражения (5.1.7).

В одномерном случае подсчет разрешенных состояний ведется в двух интервалах: от $-(k + dk)$ до $-k$ и от k до $(k + dk)$. Следовательно, результирующий одномерный слой имеет длину $2dk$, а каждое разрешенное энергетическое состояние занимает длину $2\pi/L$. Отсюда число состояний dG_1 в одномерном слое, с учетом выражений (5.1.7), равно

$$dG_1 = 2 \frac{2dk}{2\pi/L} = L \frac{2m^*}{\pi\hbar} \frac{d\varepsilon}{\sqrt{2m^*(\varepsilon - \varepsilon_n)}}. \quad (5.1.11)$$

Одномерная плотность электронных состояний вводится в расчете на единицу длины нити с суммированием по всем дискретным уровням ($n = 1, 2, \dots$), энергия которых (ε_n) лежит ниже данного значения ε . Тогда из формулы (5.1.11) следует искомое одномерное выражение (с размерностью $\text{эВ}^{-1} \cdot \text{м}^{-1}$):

$$g_1(\varepsilon) \equiv \frac{1}{L} \sum_n \frac{dG_1}{d\varepsilon} H(\varepsilon - \varepsilon_n) = \frac{(2m^*)^{1/2}}{\pi\hbar} \sum_n \frac{H(\varepsilon - \varepsilon_n)}{\sqrt{\varepsilon - \varepsilon_n}}, \quad (5.1.12)$$

где также введена функция Хевисайда $H(\varepsilon - \varepsilon_n)$.

Таким образом, в квантовой нити (с двумерной локализацией электронов) плотность состояний для каждого квантового уровня

ня ε_n неограниченно возрастает при приближении к величине ε_n со стороны больших значений, так что результирующая кривая $g_1(\varepsilon)$ принимает разрывный вид, показанный на рис. 5.1 б.

4. 0D-структура (квантовая точка). В нульмерном случае имеет место пространственная локализация электрона в трех измерениях, т. е. все значения волнового числа строго квантованы ($k = k_n$, $n = 1, 2, \dots$). Следовательно, разрешенные состояния существуют только при дискретных значениях энергии ε_n . Тогда *нульмерная плотность* электронных состояний представляется в форме совокупности дельта-функций (с размерностью эВ⁻¹):

$$g_0(\varepsilon) = \sum_n \delta(\varepsilon - \varepsilon_n). \quad (5.1.13)$$

Отсюда видно, что одномерная плотность электронных состояний $g_1(\varepsilon)$ в форме (5.1.12) превращается для квантовой точки в дискретную сетку энергетических уровней $g_0(\varepsilon_n)$, показанную на рис. 5.1 в. Таким образом, нульмерные системы типа квантовых точек (содержащие достаточно большое число реальных частиц) могут рассматриваться как искусственные атомы, имеющие дискретный спектр энергетических уровней и конечное число электронов, занимающих эти уровни.

Вышеприведенное рассмотрение наглядно демонстрирует (см. рис. 5.1) переход от монотонной зависимости $g_3(\varepsilon)$ для объемных структур 3D-размерности к другим зависимостям — ступенчатой формы $g_2(\varepsilon)$, разрывно-монотонной формы $g_1(\varepsilon)$ и дискретной формы $g_0(\varepsilon)$ для квантовых структур пониженной размерности, соответственно, для квантовых ям 2D-размерности, квантовых нитей 1D-размерности и квантовых точек 0D-размерности.

5.2. Структурно-геометрические особенности и квазикристаллическое состояние нанокластеров

По мере уменьшения размеров частиц вещества и перехода в наноразмерный диапазон происходит изменение их физических и физико-химических свойств. Как отмечалось выше, эти изменения связаны с размерными эффектами как квантовой электронной природы (КРЭ), рассмотренные в предыдущем параграфе), так структурно-атомной природы, которые будем называть *структурными размерными эффектами* (СРЭ). Причина появления СРЭ связана с возрастающим вкладом поверхностных атомов в энергию частицы при уменьшении ее геометрических размеров с переходом в нанодиапазон.

Качественную картину относительного вклада поверхностных и объемных атомов в объем частицы демонстрирует рис. 5.2.

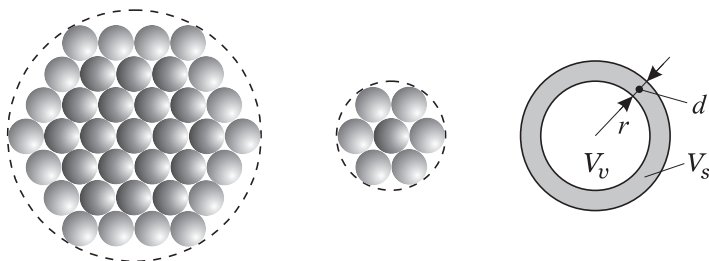


Рис. 5.2. Картина, качественно показывающая доминирующий вклад поверхностных атомов (светлые кружки) с уменьшением размера частицы, и модель частицы в форме двухслойной сферы

Для двухслойной сферы, изображенной на этом же рисунке, найдем отношение объема V_s , занимаемого поверхностным слоем толщиной d , и объема V_v внутренней сферы радиусом r :

$$\frac{V_s}{V_v} = 3 \frac{d}{r} \left(1 + \frac{d}{r} + \frac{1}{3} \frac{d^2}{r^2} \right). \quad (5.2.1)$$

Отсюда видно, что $V_s = V_v$ при $d/r = \sqrt[3]{2} - 1 \approx 0,26$. Тогда, полагая толщину d равной эффективному диаметру атома, получаем, что частица с наружным диаметром $D = 2(r + d)$ составлена из 5 атомных слоев. Этот результат близок к приведенному ниже в табл. 5.1 (см. строку, выделенную полужирным шрифтом), если считать количество объемных и поверхностных атомов, N_v и N_s , пропорциональным, соответственно, объемам V_v и V_s , а именно: в поверхностном слое находится примерно половина всех атомов сферической частицы, состоящей из 5 атомных слоев.

Состояние атомов, находящихся в поверхностном слое или вблизи него, строго говоря, отличаются от состояния таких же атомов, расположенных в объеме материала. Отличия вызваны оборванными химическими связями, механическими напряжениями решетки, адсорбированными атомами и т. п., что изменяет свойства поверхности и приводит к ее структурной перестройке (подробно эти вопросы рассмотрены в п. 5.3).

Чем больше доля атомов, находящихся на поверхности, тем больше избыточная энергия, накопленная в поверхностном слое, которая в общем случае пропорциональна площади поверхности. Отсюда стремление сократить площадь поверхности (A) при постоянном объеме ($V = \text{const}$) путем изменения формы частицы.

Для сферы $A/V = 6/D$ минимально, в то время как площади поверхности куба, октаэдра и икосаэдра больше таковой для равновеликой сферы в 1,241, 1,183 и 1,099 раза, соответственно.

Таким образом, с целью снижения избыточной поверхностной энергии наночастицы стремятся приобрести по возможности более симметричную форму с максимальной плотностью упаковки атомов. Таковой формой является икосаэдр, изображенный во введении на рис. 11 б. Он представляет собой 20-гранник с правильными треугольными гранями и 12 вершинами, в каждой из которых сходятся по пять граней. Такой многогранник обладает поворотными осями 5-го порядка, которые отсутствуют в обычных кристаллических структурах, описанных в п. 3.3.

Структура элементарной ячейки большинства кристаллов основана на таких правильных многоугольниках (телах Платона), как куб, тетраэдр и октаэдр (см. рис. 3.4), обеспечивающих как трансляционный, так и ориентационный дальний порядок. В отличие от таких многоугольников, икосаэдр (как и додекаэдр), имея симметрию 5-го порядка, не в состоянии путем трансляций плотно (без пустот) заполнить все пространство.

Более наглядной является двумерная картина плоской мозаики, которая не может быть выложена из одинаковых правильных пятиугольников. Однако, если за основу взять два ромба — с углами 36° и 144° («тонкий» ромб 1) и с углами 72° и 108° («толстый» ромб 2), то плоскость оказывается заполненной ими без зазоров, образуя так называемую *мозаику Пенроуза*, показанную на рис. 5.3. Поскольку $72^\circ = 360^\circ : 5$, то мозаика Пенроуза имеет ось вращения 5-го порядка, что обеспечивает дальний ориентационный порядок. Однако из-за отсут-

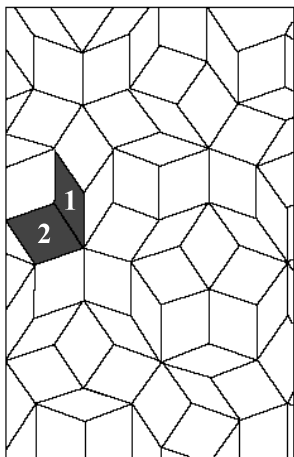


Рис. 5.3. Мозаика Пенроуза

ствия трансляционной симметрии такая двумерная структура не является периодической (см. ниже).

Трехмерное обобщение мозаики Пенроуза, составленное из двух ромбоэдров, носит название *сети Аммана–Маккея*. Трехмерная структура, подобно двумерной, не является периодической, но обладает **икосаэдрической симметрией**, обеспечивающей дальний ориентационный порядок. В отличие от обычных кристаллов (имеющих *два типа симметрии* — трансляционную

и ориентационную), структуры с икосаэдрической симметрией, строго запрещенной в традиционной кристаллографии, называют **квазикристаллами**. Квазикристаллическое состояние (с одним типом дальнего порядка — ориентационным) занимает промежуточное положение между кристаллическим (с двумя типами порядка) и аморфным (без дальнего порядка) состояниями.

Не вдаваясь в детали, заметим, что в основе квазикристаллических структур, как показал математический анализ, лежит знаменитое число $\tau = (1 + \sqrt{5})/2 \approx 1,618$, которое принято называть «золотым сечением»¹⁾.

В применении к мозаике Пенроуза следует отметить, что отношение числа «толстых» ромбов к числу «тонких» при бесконечном заполнении мозаики стремится к «золотому числу» $\tau = 1,618$. Так как это число иррациональное, то нельзя выделить элементарную ячейку, содержащую целое число ромбов каждого типа. Поэтому мозаика Пенроуза не переходит сама в себя ни при каких трансляциях, то есть отсутствует периодичность. Однако любая конечная часть мозаики встречается бесконечно много раз, что обуславливает наличие дальнего порядка.

Впервые икосаэдрическая симметрия экспериментально была обнаружена в 1984 году (Д. Шехтман и др.) при наблюдении дифракции электронов на ленте сплава $\text{Al}_{0,86}\text{Mn}_{0,14}$, полученного при сверхбыстром охлаждении расплава со скоростью до 10^6 К/с. Ярко выраженная дифракционная картина свидетельствовала о наличии дальнего порядка в расположении атомов, но противоречила фундаментальным представлениям кристаллографии из-за симметрии пятого порядка в расположении дифракционных максимумов. Сначала это было воспринято как парадокс, вызванный многократным двойникованием (см. рис. 3.16 б) в кубических кристаллах с разориентацией соседних микрокристаллов на 72° . Позднее обнаружение квазикристаллического состояния не только в быстро охлажденных сплавах, но и в структурах, близких к равновесным, таких как сплавы Al-Li-Cu , Al-Li-Cu-Mg и т. п.,

¹⁾ Напомним, что термин «золотое сечение» связан со следующим геометрическим построением: поперечным сечением отрезка длиной l получить две его части l_1 и l_2 (пусть $l_1 > l_2$) таким образом, чтобы $l/l_1 = l_1/l_2 \equiv \tau$. Отсюда следует квадратное уравнение $\tau^2 - \tau - 1 = 0$, корень которого $\tau = (1 + \sqrt{5})/2 \approx 1,618$ вошел в историю живописи и архитектуры как *золотое сечение*. При этом обратное число $\phi = 1/\tau$ таково, что $\phi = \tau - 1 \approx 0,618$. Эти два числа лежат в основе *ряда Фибоначчи* — числовой последовательности, в которой каждое число является суммой двух предыдущих и так до бесконечности. Результат деления друг на друга двух соседних чисел ряда по мере их роста стремится либо к τ (большее на меньшее), либо к ϕ (меньшее на большее).

привело к пересмотру фундаментальных положений традиционной кристаллографии и к отказу от требования строгой периодичности как основного свойства кристаллов.

Особенности структуры нанокластеров обычно исследуются при помощи масс-спектроскопических измерений. Масс-спектры характеризуются набором острых пиков, указывающих на повышенную стабильность кластеров, содержащих вполне определенное количество атомов. Числа, соответствующие пикам в распределении кластеров по массам, называют *магическими числами*.

Установлено, что для наночастиц, макроскопическая фаза которых имеет гранецентрированную кубическую (ГЦК) решетку (Au, Pt, Pd, Ag, Al, Cu и др.), при размерах менее 10 нм наиболее стабильной является оболочечная структура на основе *икосаэдра*, построенного вокруг центрального атома (см. рис. 5.4 а). Если к поверхности частицы, показанной на этом рисунке, добавить следующую оболочку, состоящую из 42 атомов, то получится частица той же формы, содержащая 55 атомов. Увеличивая число оболочек (n), можно получать все бóльшие наночастицы. Они образуют ряд устойчивых кластеров с суммарным количеством атомов $N_n = 1, 13, 55, 147, 309, 561, \dots$, которые и являются магическими числами. Оболочечная модель кластера на основе икосаэдра дает выражение, связывающее магическое число N_n с числом n полностью заполненных атомных оболочек,

$$N_n = \frac{1}{3} (10n^3 - 15n^2 + 11n - 3).$$

Наружная оболочка, содержащая $(N_n - N_{n-1})$ атомов, образует поверхностный слой частицы с числом атомов

$$N_s = 10n^2 - 20n + 12.$$

Нетрудно видеть, что диаметр наночастицы, состоящей из n атомных оболочек, равен

$$D_n = (2n - 1)d,$$

где d — диаметр атома ($d = 2,88 \text{ \AA}, 2,86 \text{ \AA}, 2,75 \text{ \AA}, 2,56 \text{ \AA}, 2,48 \text{ \AA}$, соответственно, для Au, Al, Pd, Cu, Fe).

Результаты расчета по этим формулам приведены в табл. 5.1. Отсюда видно, что при уменьшении размеров частицы (снизу вверх) количество поверхностных атомов (N_s) примерно сравнивается с числом объемных ($N_v = N_{n-1}$) при пяти атомных оболочках (строка $n = 5$, выделенная полужирным шрифтом). Этот результат, как ни удивительно, совпадает с упрощенной геометрической оценкой, выполненной выше по формуле (5.2.1).

Таблица 5.1. Магические числа N_n для наночастиц с оболочечной атомной структурой на основе икосаэдра

Число оболочек n	Количество атомов в наночастице		
	$N_n = N_v + N_s$	N_s	$N_s/N_n, \%$
1	1	1	100
2	13	12	92,3
3	55	42	76,4
4	147	92	62,6
5	309	162	52,4
6	561	252	44,9
7	923	362	39,2
8	1415	492	34,8
9	2057	642	31,2
10	2869	812	28,3
25	$4,09 \cdot 10^4$	$5,76 \cdot 10^3$	11,7
50	$4,04 \cdot 10^5$	$2,40 \cdot 10^4$	5,9
75	$1,38 \cdot 10^6$	$5,48 \cdot 10^4$	4,0
100	$3,28 \cdot 10^6$	$9,80 \cdot 10^4$	3,0

Таким образом, нанокластеры металлов с ГЦК-решеткой, состоящие из 150–300 атомов (в 4–5 слоях), должны иметь икосаэдрическую форму. Экспериментально было подтверждено, что кластеры золота, платины и палладия с размерами 4–10 нм действительно имеют симметрию пятого порядка, характерную для оболочечной икосаэдрической модели. Аналогичные структуры наблюдаются также у наночастиц металлических сплавов типа Al–Cu, Au–Cu и др.

По мере увеличения числа атомов в наночастице возрастает вклад упругих напряжений, запаасающих энергию. Этот вклад может превысить уменьшение поверхностной энергии, вызванное увеличением объема. По этой причине стабильность икосаэдрических частиц с размерами более 10 нм начинает снижаться.

Как показывают эксперименты, вышеупомянутые наночастицы золота, платины и палладия с размерами 15–100 нм приобретают новую структуру *кубооктаэдра*, показанного на рис. 5.4 б. Для наночастиц веществ, макроскопическая фаза которых имеет гексагональную плотноупакованную (ГПУ) решетку, таких как

Mg, Zn, Ti, Cd, Hf и др., наибольшей стабильностью обладает структура в форме *гексагонального кубооктаэдра* (рис. 5.4 в).

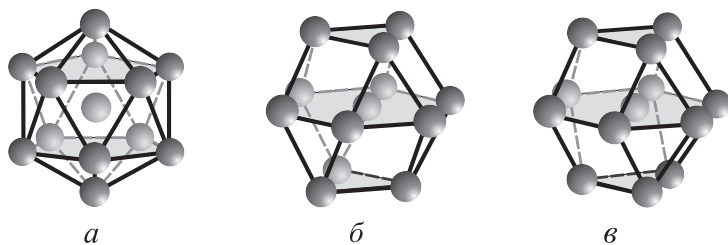


Рис. 5.4. Три формы 13-атомных наночастиц: *а* — икосаэдр, *б* — кубооктаэдр, *в* — гексагональный кубооктаэдр

Все три полиэдра, изображенные на рис. 5.4, являются центрированными, то есть в них 12 атомов, расположенных по вершинам, окружают центральный атом.

Центрированный икосаэдр (рис. 5.4 *а*) получается из исходного икосаэдра, состоящего из 20 треугольных граней (рис. 11 *б*) помещением центрального атома между двумя пятиугольниками (слегка затемненными для наглядности). Эта картина соответствует двум заполненным оболочкам ($n = 1, 2$) икосаэдрической модели нанокластера. Заполнение последующих оболочек ($n = 3, 4, 5, \dots$) дает последовательность магических чисел ($N_n = 1, 13, 55, 147, 309, 561, \dots$), указанных в табл. 5.1, что характерно для структур с поворотной симметрией пятого порядка.

Оба кубооктаэдра, изображенные на рис. 5.4 *б* и *в*, не имеют осей пятого порядка, так как представляют собой 14-гранник, составленный из 6 квадратных и 8 треугольных граней. Половина этих граней расположена выше, а другая — ниже центрального шестиугольника (затемненного для ясности). Эти кубооктаэдры отличаются разным соединением между собой верхних и нижних граней вдоль сторон центрального шестиугольника.

В кубооктаэдре (рис. 5.4 *б*) треугольники соединены с квадратами и квадраты с треугольниками, а в гексагональном кубооктаэдре (рис. 5.4 *в*) — треугольники с треугольниками и квадраты с квадратами. Такое различие порождено разным взаимным расположением *базисных треугольников* — верхнего и нижнего (также изображенных слегка затемненными): на рис. 5.4 *в* они параллельны, а на рис. 5.4 *б* — антипараллельны, что достигается простым поворотом одного из них на 60° . В основе этого лежит разный порядок заполнения слоев при гексагональной и кубической плотнейших упаковках, подробно рассмотренных ранее в п. 3.3 (см. рис. 3.5 и относящийся к нему текст).

Поскольку между двумя кубооктаэдрами нет принципиального различия, то они имеют один и тот же набор магических чисел 1, 13, 57, 153, 321, 581, ..., слегка отличающийся от аналогичного набора для икосаэдрических нанокластеров.

Рассмотренные выше магические числа получаются при минимизации объема наночастицы и максимализации плотности упаковки атомов в ней с формой, близкой к сферической. Так как речь идет о пространственной атомной структуре наночастиц, то такие магические числа нередко называют *структурными*, чтобы отличить их от *электронных* магических чисел.

Электронные магические числа возникают при анализе электронной структуры (энергетических уровней) кластеров, состоящих, в первую очередь, из щелочных металлов. Как известно (см. п. 3.1), электронные свойства металлов создаются делокализацией валентных электронов каждого атома с образованием коллектива обобществленных электронов, движущихся на фоне малоподвижных положительных ионов.

Такое же модельное представление распространяется и на металлические кластеры, содержащие от единиц до многих тысяч атомов. В основу анализа положена так называемая *модель желе*: обобществленные валентные электроны движутся в однородном сферически симметричном поле ионного остова (без учета его дискретной атомной структуры).

Решение уравнения Шредингера для модели желе в одноэлектронном приближении, подобно одиночному атому, характеризуется тремя квантовыми числами — главным n , орбитальным l и магнитным m (см. сноску на с. 336). Так как электронные состояния с одинаковым значением n образуют оболочку, то кластеры, как и атомы, имеют оболочечную электронную структуру. Последовательность полностью заполненных электронных оболочек проявляется в экспериментах как *электронные магические числа*. Так, кластеры щелочных металлов обладают наибольшей устойчивостью при числе атомов $N = 8, 18, 20, 34, 40, \dots$

Наличие у нанокластеров магических чисел (структурных и электронных) является объяснением многих вновь экспериментально наблюдаемых физических и химических свойств, проявляющихся в их немонотонной зависимости от размеров частиц, что характеризуется как структурные размерные эффекты.

При изучении наночастиц и консолидированных наноматериалов большое значение приобретает понимание разнообразных процессов, происходящих на поверхности и в приповерхностных слоях твердых тел, включая специфику возникновения атомных поверхностных структур, отличных от объемных фаз.

5.3. Атомная структура, релаксация и реконструкция поверхностей

Систематическое изучение поверхности твердых тел началось около 50 лет назад после того, как были надежно отработаны:

- а) техника получения сверхвысокого вакуума (СВВ) с давлением остаточных газов, не превышающим 10^{-9} мм рт. ст. (вакуумно-чистые материалы, вакуумные насосы, вентили и манометры);
- б) практические приемы приготовления атомарно-чистых поверхностей (скол кристаллов в СВВ с последующим термическим прогревом, химическим травлением и ионной бомбардировкой);
- в) методы диагностики поверхности (электронная, ионная и зондовая микроскопия, электронная спектроскопия и дифрактометрия, масс-спектропия, см. краткое описание на с. 65).

За эти годы, кроме экспериментальной техники и методик исследования, были также разработаны на их основе новые модельные представления об атомной структуре поверхности.

Поверхностью твердого тела принято называть приповерхностную область толщиной в несколько (до 4–5) атомных слоев, структура которой (как атомная, так и электронная) существенно отличается от таковой для объемного кристалла. Поверхность, даже приготовленная самым тщательным образом путем скола в СВВ, всегда отличается от *идеальной поверхности*, мысленно образованной путем разрыва связей между атомами с сохранением их расположения в решетке. Идеализация нужна для того, чтобы по отношению к ней фиксировать реально наблюдаемые положения атомов, возникшие в результате двух процессов, называемых *релаксацией* и *реконструкцией* поверхности. Прежде чем описывать эти процессы, укажем принятую систему обозначения поверхностной структуры атомов.

Идеальную поверхность, как и трехмерный идеальный кристалл, характеризуют кристаллической решеткой, только двумерной, с двумя трансляционными периодами a , b и углом γ между ними. Все многообразие двумерных решеток описывают пятью основными решетками Браве (вспомним, что трехмерные кристаллы имеют всего 14 решеток Браве, см. рис. 3.1):

- квадратная решетка $a = b, \gamma = 90^\circ$,
- прямоугольная решетка $a \neq b, \gamma = 90^\circ$,
- прямоугольная центрированная решетка $a \neq b, \gamma = 90^\circ$,
- гексагональная решетка $a = b, \gamma = 120^\circ$,
- косоугольная решетка $a \neq b, \gamma \neq 90^\circ$.

Так как указанные двумерные решетки соответствуют идеальной поверхности, то элементарные поверхностные трансляции a и b совпадают с аналогичными периодами идентичности исходной трехмерной решетки. Действительное положение атомов даже на атомарно-чистой поверхности отличается от идеального, как результат релаксации и реконструкции, и характеризуется другими поверхностными трансляциями a_s и b_s .

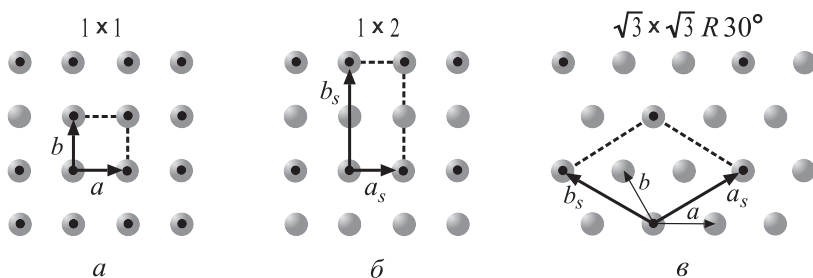


Рис. 5.5. Примеры обозначения поверхностных структур: a — идеальная квадратная решетка 1×1 (поверхностная и объемная ячейки совпадают); $б$ — прямоугольная решетка 1×2 с периодами $a_s = a$, $b_s = 2b$; $в$ — гексагональная решетка $\sqrt{3} \times \sqrt{3} R 30^\circ$ с периодами $a_s = \sqrt{3} a$, $b_s = \sqrt{3} b$ и с поворотом на 30° относительно идеальной решетки с периодами a и b (тонкие стрелки). Здесь атомы идеальной поверхности (подложки) обозначены светло-серыми шарами, а атомы поверхностной структуры — черными точками

1. Индексация поверхностной структуры построена следующим образом. Если на кристаллографической плоскости с индексами Миллера (hkl) объемного кристалла (подложки) X расположена двумерная решетка с периодами $a_s = ma$ и $b_s = nb$, которая к тому же повернута на угол θ° (в градусах) по отношению к положению трансляций a и b на идеальной поверхности (или на подложке), то такую структуру поверхности записывают как

$$\begin{aligned} X(hkl) m \times n R \theta^\circ & \text{ при } \theta^\circ \neq 0, \\ X(hkl) m \times n & \text{ при } \theta^\circ = 0. \end{aligned} \quad (5.3.1)$$

Примеры обозначения поверхностных структур приведены на рис. 5.5. Так, например, на кремнии ($X = \text{Si}$) встречаются следующие поверхностные структуры:

$$\text{Si}(111)7 \times 7; \quad \text{Si}(100)c 4 \times 2; \quad \text{Si}(111)\sqrt{3} \times \sqrt{3} R 30^\circ\text{-}3\text{Bi}.$$

Первая структура наиболее исследована и будет далее обсуждаться (см. рис. 5.10); вторая соответствует прямоугольной центрированной решетке, что отмечено буквой c ; третья отражает,

кроме поворота на угол 30° , также присутствие атомов адсорбата (три атома висмута Bi на элементарную ячейку).

Кроме наиболее распространенного типа обозначений (5.3.1), используют также матричную запись, менее наглядную, но более универсальную, применимую, например, когда обе решетки не принадлежат к одному типу решеток Браве. В этом случае квадратная матрица, описывающая поверхностную структуру, связывает между собой вектор-столбцы $\{a_s, b_s\}$ и $\{a, b\}$.

Перейдем к описанию атомных процессов релаксации и реконструкции, формирующих поверхностную структуру.

Идеальной атомарно-чистой поверхностью называют поверхность, возникающую в результате разрыва химических связей и удаления одной из частей кристалла с сохранением расположения атомов в оставшихся слоях. Она может быть представлена гипотетически лишь при температуре абсолютного нуля.

В реальных условиях наблюдается перестройка кристаллической структуры атомарно-чистой поверхности на глубину в несколько атомных слоев. Движущая сила такой перестройки связана со стремлением к термодинамическому равновесию путем понижения свободной энергии системы конденсированная фаза — поверхность. Равновесие устанавливается в результате смещения приповерхностных атомов, условно разделенного на два этапа — *релаксация* и *реконструкция* атомных слоев.

2. Релаксация поверхности подразумевает смещение атомного слоя (одного или нескольких) в целом без изменения межатомных расстояний и структуры этого слоя. Однако *по нормали* к поверхности расстояния между соседними слоями могут изменяться. Поэтому такую релаксацию, схематически показанную на рис. 5.6 *a*, называют *нормальной*.

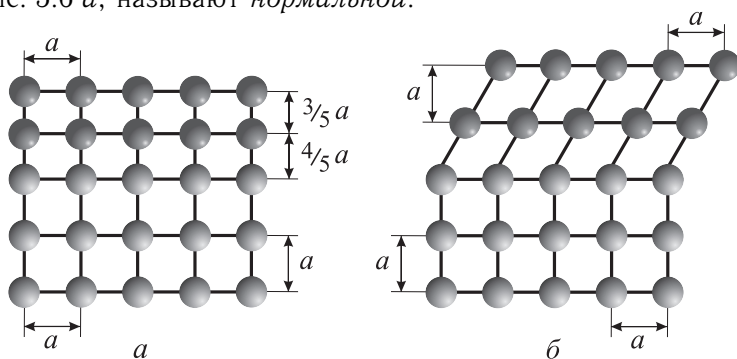


Рис. 5.6. Примеры релаксации поверхностных атомных слоев (темные шары): *a* — нормальная релаксация, *б* — латеральная релаксация

Таким образом, нормальная релаксация¹⁾ сохраняет симметрию в расположении атомов параллельно поверхности, но изменяет расстояния между кристаллографическими плоскостями. В чистом виде нормальная релаксация наблюдается, как правило, для низкоиндексных плоскостей в некоторых ГЦК- и ОЦК-металлах типа Al, Ni, Fe и др. Изменение межслойного расстояния может достигать 25% от его объемного значения, уменьшаясь в глубь кристалла, причем нередко с осцилляциями. Так, на поверхности Al(110) первое межслойное расстояние сжато на $-8,6\%$, второе растянуто на $+5\%$, а третье снова сжато, хотя и незначительно, на $-1,6\%$.

В ряде случаев атомы верхнего слоя (одного или нескольких) смещаются в плоскости, параллельной поверхности, относительно своего положения в объеме кристалла с сохранением межатомных расстояний и структуры этого слоя. Такое смещение, схематически показанное на рис. 5.6 б, принято называть тангенциальной или *латеральной* (боковой от англ. *lateral*) релаксацией. Латеральная релаксация наблюдается обычно в комбинации с нормальной релаксацией на высокоиндексных плоскостях с низкой концентрацией атомов. Латеральные смещения атомов всегда меньше нормальных и составляют от долей до единиц процента. Например, на поверхности Ni(311) три верхних слоя имеют нормальные смещения $-15,9\%$, $+4,1\%$, $-1,6\%$ (осцилляторные) и латеральные смещения $0,8\%$, $1,4\%$, $0,5\%$ (одного направления).

Приведенные нормальные и латеральные смещения измерены методом ДМЭ — дифракции медленных электронов (см. с. 65).

3. Реконструкция поверхности означает перестройку атомного слоя (одного или нескольких) в структуру, где симметрия и периодичность расположения атомов отличаются от таковых для плоскостей в объеме кристалла. Реконструкция влияет на все структурно-чувствительные свойства поверхности (физические, химические, оптические и др.). В зависимости от того, сохраняется или нет число атомов в поверхностных слоях, различают два вида реконструкции — консервативную и неконсервативную.

При *консервативной реконструкции* атомы из поверхностного слоя не исчезают, а лишь смещаются из узлов кристаллической решетки, образуя при этом разнообразные поверхностные структуры. Простейший пример консервативной реконструкции,

¹⁾ Введенный здесь термин «релаксация» в узко специализированном применении к атомным слоям следует отличать от его общефизического смысла (см. с. 80) как самопроизвольно развивающегося во времени процесса возвращения системы в равновесное состояние после снятия внешних воздействий.

сопровождающийся формированием на поверхности димерных структур (сдвоенных атомов), показан на рис. 5.7 *а*. После возникновения димеров поверхность совместима сама с собой при сдвиге на расстояние между центрами соседних димеров, которое вдвое больше объемного периода решетки. Такую реконструкцию обозначают символом 2×1 , показывающим, что период элементарной ячейки поверхностной двумерной решетки в одном направлении увеличивается вдвое, а в другом остается прежним.

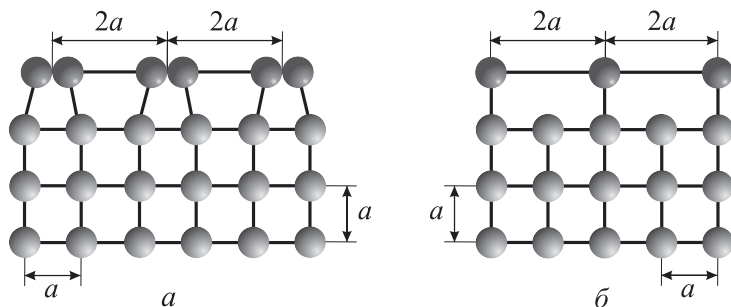


Рис. 5.7. Примеры реконструированной поверхности 2×1 : *а* — консервативная реконструкция, *б* — неконсервативная реконструкция

При *неконсервативной реконструкции* число атомов в реконструированном слое меньше их количества в кристаллографических плоскостях объемного кристалла. Примером может служить структура 2×1 , показанная на рис. 5.7 *б*. В ней отсутствует каждый атомный ряд через один, так что период двумерной решетки в этом направлении удваивается, подобно поверхностной димерной структуре на рис. 5.7 *а*.

Реконструкция (консервативная и неконсервативная) может охватывать несколько верхних слоев. Кроме того, процесс реконструкции обычно сопровождается смещением атомов по нормали к поверхности, т. е. релаксацией, не отраженной на рис. 5.7.

Как правило, изменения решетки в процессе реконструкции сопровождаются изменением ее точечной симметрии (понижением по сравнению с идеальной поверхностью). Поэтому в настоящее время принято рассматривать реконструкцию как своеобразный фазовый переход, приводящий к изменению ряда поверхностных характеристик, например, работы выхода электронов, параметров поверхностных фононов и др. По завершении реконструкции возникает новая «поверхностная фаза» с присущим ей характером поверхностных энергетических зон. В частности, из-за образования димеров наблюдается появление поверхностной

запрещенной зоны в полупроводниках. Ее ширина определяется требованием минимизации энергии электронов на поверхности, при этом уровень Ферми располагается в запрещенной зоне, т. е. поверхность сохраняет полупроводниковые свойства.

Имеются и другие причины усложнения строения поверхности. При обычных температурах из-за малой скорости протекания процессов реконструкции и релаксации, поверхность твердого тела имеет не равновесную, а метастабильную форму, которая определяется предысторией (процессами зародышеобразования, условиями последующего роста, обработкой поверхности и т. п.).

Как ранее было отмечено, поверхности большинства металлов нереконструированы, т. е. претерпевают лишь релаксацию. Только в некоторых благородных ГЦК-металлах (Au, Pt, Ir) и переходных ОЦК-металлах (W, Mo) проявляется реконструкция поверхности, наблюдаемая при помощи СТМ — сканирующего туннельного микроскопа (см. с. 57–58).

В наибольшей степени реконструкции подвержены поверхности полупроводников, среди которых более исследованы поверхности кремния Si(100) и Si(111), которые и рассмотрим ниже.

Особенность поверхностной реконструкции полупроводников, в частности кремния, связана с направленностью ковалентных связей, которая отсутствует в металлах (см. п. 3.1). Обрыв электронных связей между атомами на идеальной поверхности лишает их соседей, которые должны были бы находиться в вышележащих слоях. Это делает такие связи *ненасыщенными*, что повышает свободную энергию кристалла с поверхностью.

Для понижения энергии у неспаренных электронов остается единственная возможность замкнуть связи (ставшие оборванными) между соседними атомами в своем слое и в нижележащих слоях. Замыкание связей приводит к формированию на поверхности определенных структурных единиц в виде *димеров* и *адатомов*¹⁾. В результате этого число оборванных (ненасыщенных) связей на реконструированной поверхности сокращается по сравнению с идеальной поверхностью.

¹⁾ Отметим, что термин «адатом» (от англ. *adatom*) обычно расшифровывается как «адсорбированный атом» (от англ. *adsorbed atom*) и применяется по отношению к чужеродным атомам на поверхности конденсированной фазы. При описании реконструкции атомарно-чистой поверхности речь, однако, идет о *собственных атомах* кристалла, например кремния, в отсутствие чужеродных частиц. Тем не менее однотипные атомы (того же кремния) в самом верхнем слое реконструкции, расположенном выше идеальной поверхности, принято также называть *адатомами*. В данном случае такой аббревиатуре разумнее дать расшифровку как «дополнительный атом» (от англ. *additional atom*).

Как известно, кремний имеет структуру алмаза (рис. 3.9 *a*) с тетраэдрической координацией атомов: каждый атом располагается в центре правильного тетраэдра и имеет четыре ближайших соседа в его вершинах (см. рис. 3.4 при $z = 4$).

Реконструкция поверхности Si(100) схематически показана на рис. 5.8. Идеальная поверхность Si(100)1×1 представляет собой квадратную решетку, образованную верхними атомами кремния, каждый из которых связан с двумя атомами нижележащего слоя и имеет две ненасыщенные связи (рис. 5.8 *a*).

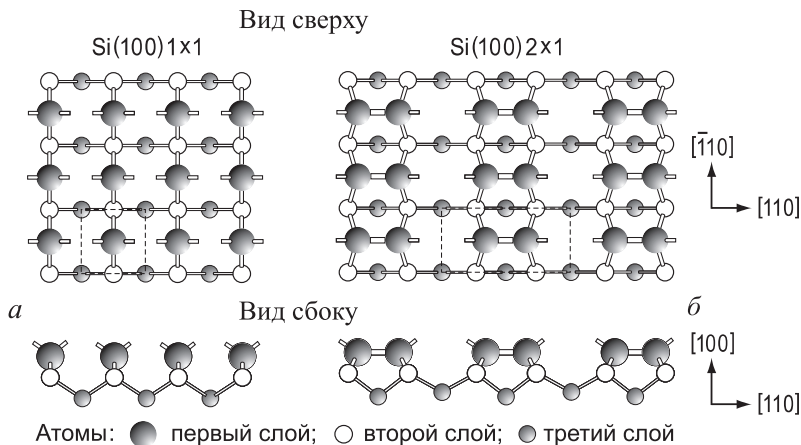


Рис. 5.8. Схематическая диаграмма, иллюстрирующая реконструкцию поверхности кремния: *a* — идеальная нереконструированная поверхность Si(100)1×1; *б* — реконструированная поверхность Si(100)2×1 (димерная реконструкция). Пунктирными линиями выделены элементарные ячейки

На реконструированной поверхности Si(100)2×1 атомы кремния спариваются, образуя димеры. В результате этого период решетки увеличивается, а число ненасыщенных связей уменьшается вдвое, что и отражено на рис. 5.8 *б*.

В этом случае модификация поверхности сводится лишь к возникновению димеров (без появления адатомов). Описанная модель димерной реконструкции Si(100)2×1 впервые получила экспериментальное подтверждение (методом ДМЭ) в 1959 году. Длина димера оказалась равной 2,28 Å при межатомной связи $d_{\text{Si-Si}} = 2,35 \text{ Å}$ (период решетки кремния $a = 5,43 \text{ Å}$, см. с. 361).

Реконструкция поверхности Si(111) происходит более сложным и многовариантным образом, хотя бы потому, что каждый атом на идеальной поверхности имеет всего одну ненасыщенную ковалентную связь, а не две, как в случае Si(100). Такие

оборванные связи, изображенные на рис. 5.9, называют «болтающимися связями» (от англ. *dangling bonds*). Соответствующее им состояние поверхности является энергетически невыгодным, и она перестраивается путем замыкания болтающихся связей в структуру, обладающую меньшим числом ненасыщенных связей и меньшей свободной энергией.

На атомарно-чистой поверхности Si(111) наблюдаются различные варианты реконструкции, среди которых 2×1 , 3×3 , 5×5 , 7×7 , 9×9 и т. п. Самой минимальной энергией обладает поверхностная структура Si(111) 7×7 , устойчивая до температуры около 850°C , где она превращается в структуру 1×1 .

Впервые поверхность Si(111) 7×7 была наблюдаема методом ДМЭ в 1959 году. В течение 25 лет она оставалась загадкой для исследователей, пока не была предложена модель Такаянаги, ставшая ныне общепризнанной. В ее основу положено предположение об образовании в реконструированном слое димеров, адатомов и дефектов упаковки, что дало ей название «DAS-модель» (от англ. *Dimer Adatom Staking-fault model*). На сегодняшний день эта модель получила подтверждение методами электронной дифракции на просвет и сканирующей туннельной микроскопии.

DAS-модель поверхностной структуры Si(111) 7×7 схематично изображена на рис. 5.10 в двух видах — сверху и сбоку.

Элементарная ячейка 7×7 реконструированной поверхности ограничена пунктирной линией в форме ромба (на виде сверху) с диагоналями длиной $46,6 \text{ \AA}$ (горизонтальная) и $26,9 \text{ \AA}$ (вертикальная). Она состоит из 49 элементарных ячеек 1×1 идеальной поверхностной грани Si(111). Каждый атом этой грани, согласно рис. 5.9, является центральным атомом тетраэдра, потерявшим соседа в верхней вершине (см. рис. 3.4 при $z = 4$). Следовательно, нереконструированная ячейка 7×7 имеет 49 ненасыщенных ковалентных связей. Покажем, что после реконструкции число таких связей сократится до 19 в соответствии с картиной на рис. 5.10.

До реконструкции верхний слой ячейки 7×7 (состоящий из светло-серых шаров), считается 1-м слоем. После реконструкции выше его располагается 0-й слой, содержащий так называемые адатомы (см. сноску на с. 573) в виде темно-серых шаров большего диаметра. Количество таких адатомов 12, и каждый из них

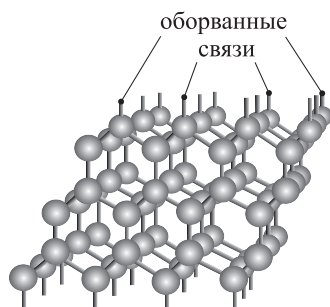


Рис. 5.9. Схематическое изображение оборванных (ненасыщенных) связей на поверхности Si(111)

связан с тремя атомами нижележащего слоя. Следовательно, химические связи 36 атомов из 1-го слоя оказываются насыщенными связью с адатомами, которые сами имеют по одной «болтающейся» связи. Оставшиеся 6 атомов 1-го слоя, над которыми нет адатомов в 0-м слое, сохраняют свою изначально ненасыщенную связь. Они называются *рест-атомами* («оставшимися атомами» от англ. *rest atoms*) и показаны светло-серыми шарами с крестом.

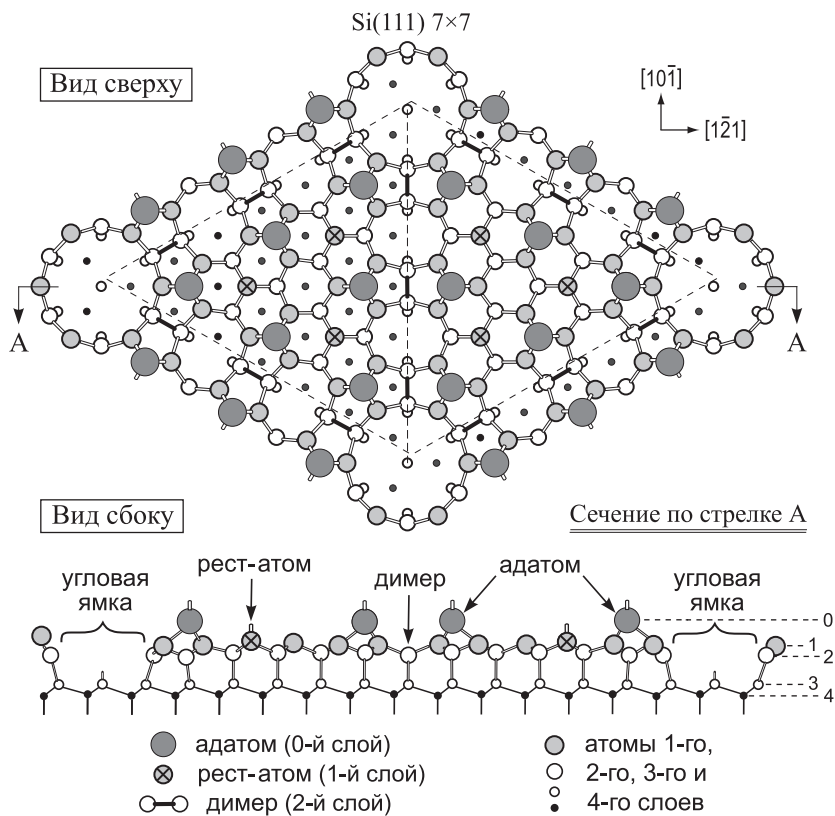


Рис. 5.10. DAS-модель реконструкции поверхностной структуры Si(111) 7×7 с элементарной ячейкой в форме ромба, выделенного пунктирной линией

В четырех углах ячейки располагаются *угловые ямки*, окруженные 12 атомами из 1-го и 2-го слоев, на дне которых (глубиной 2 Å) имеется один атом 3-го слоя с ненасыщенной связью. Поскольку угловая ямка принадлежит одновременно четырём примыкающим ячейкам 7×7 , то на каждую из них приходится по одной ненасыщенной связи в этой ямке.

Таким образом, DAS-структура $\text{Si}(111)7 \times 7$ содержит 19 насыщенных связей на элементарную ячейку, из которых 12 приходится на адатомы, 6 на рест-атомы и одна на угловые ямки.

Кроме того, во втором слое располагаются димеры: 12 вдоль границ ромбической ячейки 7×7 и три по ее вертикальной диагонали, так что на одну ячейку приходится девять димеров, полностью насыщающих межатомные связи. Нижеследующие два слоя (3-й и 4-й), образуют треугольные подъячейки (разделенные вертикальной диагональю ромба с тремя димерами), которые попеременно содержат и не содержат *дефекты упаковки*. Это отражено на рисунке как наличие и отсутствие черных точек в левом и правом треугольниках, ограниченных пунктирными линиями.

Многочисленные исследования показывают, что на реконструкцию поверхности полупроводников влияют многие факторы, в частности физическая и химическая адсорбция, подробно рассмотренная в п. 5.9. Адсорбированные атомы и молекулы, взаимодействуя с атомами поверхностного слоя, снижают свободную поверхностную энергию конденсированной фазы и суммарную энергию системы. Это является одной из стимулирующих причин реконструкции, интенсивность которой увеличивается с ростом сил взаимодействия. В этом случае описание структуры поверхности проводят отдельно для атомов конденсированной фазы и адсорбированных атомов. Реконструкция наиболее ярко проявляется при хемосорбции, которая сопровождается заметной деформацией поверхности твердой фазы, поэтому ее часто называют реконструктивной хемосорбцией.

5.4. Структурные дефекты реальных поверхностей

Рассмотренные выше процессы релаксации и реконструкции формируют упорядоченную поверхность с полной трансляционной симметрией. Однако в реальности любая поверхность обязательно имеет структурные дефекты, которые играют существенную роль в процессах адсорбции, поверхностной диффузии, зародышеобразования и формирования новой фазы.

Дефекты поверхности принято описывать в рамках простейшей модели ТСИ (Терраса–Ступень–Излом), в которой отдельные атомы изображаются в виде кубика. Базовые элементы модели отмечены на рис. 5.11 цифрами: ① — *террасы* (плоские поверхности кристалла); ② — *ступени* (скачки при переходе с террасы на террасу чаще моноатомной высоты); ③ — *изломы* (уступы ступени чаще моноатомной ширины). На рисунке показаны атомы, принадлежащие террасе, ступени и излому ступени.

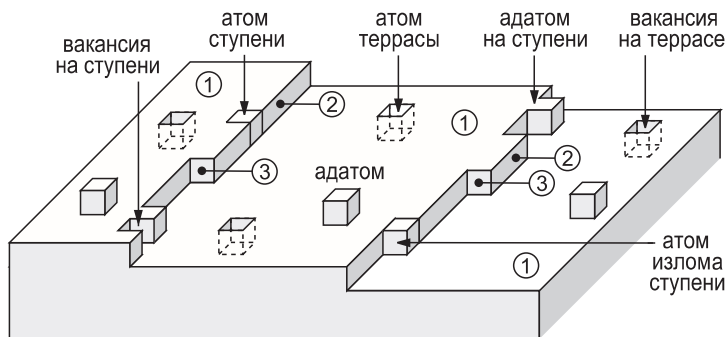


Рис. 5.11. Модель ТСИ поверхности (100) простого кубического кристалла: ① — террасы, ② — ступени, ③ — изломы ступени

К дефектам модели ТСИ относятся, во-первых, *адатомы* — одиночные атомы (собственные и чужеродные, см. сноску на с. 573) на поверхности террасы и на грани ступени и, во-вторых, *вакансии* — отсутствующие атомы на террасах и ступенях.

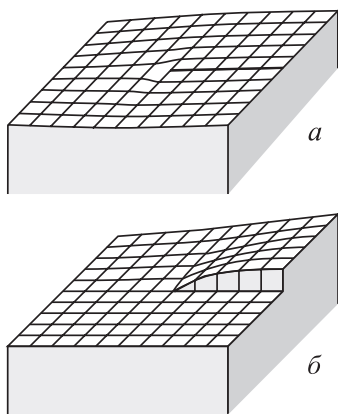


Рис. 5.12. Выходы дислокаций: а — краевой, б — винтовой

Дефектами поверхности являются также *точки выхода* краевых и винтовых *дислокаций* (рис. 3.15), образующих в модели ТСИ картины расположения атомов, схематически изображенные на рис. 5.12 а и б. Отсюда видно, что в точках выхода дислокаций упорядоченное расположение атомов нарушено, а при винтовой дислокации образуется атомная ступень, один конец которой закреплен на оси дислокации. Выход винтовой дислокации оказывает влияние на рост кристалла путем присоединения атомов к незарастающей атомной ступени, вращающейся по спирали вокруг оси дислокации.

Модель ТСИ не учитывает электронную структуру, тепловые колебания решетки и другие практически важные особенности кристаллов, но в большинстве случаев правильно отражает качественную картину реальной поверхности.

Адаомы и вакансии на террасах и ступенях и изломы ступеней являются термодинамическими устойчивыми дефектами. Их относительное количество зависит от температуры и энергии образования дефекта каждого типа. В модели ТСИ энергия

образования дефекта определяется через энергию связи атома, зависящую от числа его ближайших соседей в различных положениях. Например, для поверхности (100) простого кубического кристалла (рис. 5.11) количество ближайших соседей равно числу граней кубика (одиночного атома), контактирующих со своим окружением, а именно: шесть *для атома* в объеме, пять в террасе, четыре в ступени, три в изломе ступени и, наконец, два и один *для адатома* на ступени и на террасе, соответственно.

Для других типов кристаллических структур ситуация аналогичная с учетом лишь того, что изменяется число ближайших соседей (координационное число) для объема кристалла: $z = 12$ для ГЦК- и ГПУ-структур, $z = 8$ — для ОЦК-структуры и $z = 4$ — для структуры алмаза (см. табл. 3.2).

Положение атома в изломе ступени является энергетически особым положением. Дело в том, что число связей у атома в изломе равно половине аналогичных связей внутри объема, независимо от типа кристалла. Отсюда энергия атома в изломе всегда равна половине его энергии в объеме, а последняя как раз и определяет энергию сублимации кристалла. Действительно, для расчета последней следует работу, затраченную на вырывание атома из кристалла, разделить пополам, так как одна связь образована двумя атомами, один из которых сублимирует, а другой остается на поверхности.

Таким образом, изменение внутренней энергии кристалла, вызванное удалением атома из излома, равно теплоте сублимации, которая может быть легко измерена. Это дает возможность рассчитывать энергию образования и концентрацию поверхностных дефектов. Например, образование собственного адатома может быть представлено как перемещение атома из излома на террасу, а возникновение вакансии — как перенос атома террасы в излом ступени. Тогда концентрации адатомов и вакансий записывают следующим образом (ср. формулу (3.10.1)):

$$N_A = N \exp\left(-\frac{\Delta F_A}{k_B T}\right) \quad \text{и} \quad N_V = N \exp\left(-\frac{\Delta F_V}{k_B T}\right).$$

Здесь N — полное число положений атома на единицу площади, $\Delta F_A = \mathcal{E}_{\text{ад}} - \mathcal{E}_{\text{изл}}$ и $\Delta F_V = \mathcal{E}_{\text{изл}} - \mathcal{E}_{\text{тер}}$ — свободные энергии образования адатомов и вакансий; при этом $\mathcal{E}_{\text{ад}}$, $\mathcal{E}_{\text{изл}}$ и $\mathcal{E}_{\text{тер}}$ означают энергию связи атома в положениях адатома, излома и террасы.

Для более точного определения энергии дефектов необходимо учитывать не только ближайших соседей, но и соседей второго и более высоких порядков, а также релаксацию поверхности в районе дефектов.

Согласно модели ТСИ, ступени и изломы ступеней на поверхности формируются искусственно в результате ее разориентации относительно основных кристаллографических граней кристалла (с малыми индексами Миллера). По этому признаку поверхности идеальных кристаллов принято подразделять на три типа: *сингулярные* (атомарно-гладкие) поверхности (рис. 5.13 а), *вицинальные* (ступенчатые) поверхности (рис. 5.13 б и в), *несингулярные* (диффузные) поверхности (рис. 5.14 а).

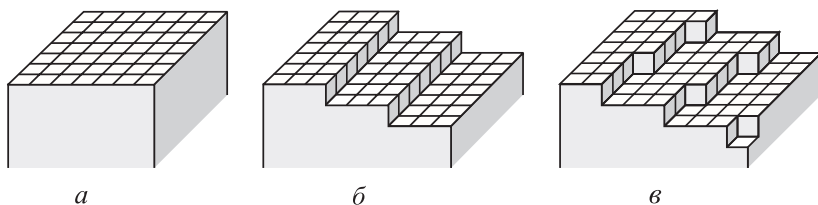


Рис. 5.13. Типы поверхностей в модели ТСИ: а — сингулярная поверхность, б — вицинальная поверхность, разориентированная по отношению к сингулярной грани в одном направлении, в — вицинальная поверхность, разориентированная по отношению к сингулярной грани в двух направлениях

Сингулярными (от лат. «единственный») **гранями** называют атомарно-гладкие поверхности, не имеющие ступеней, например грань (100) кубической решетки (рис. 5.13 а). По сравнению с другими поверхностями сингулярные грани обладают наименьшей поверхностной энергией, наибольшей плотностью упаковки атомов и характеризуются малыми индексами Миллера.

Вицинальными (от лат. «близкий») **гранями** (или просто *вициналями*) принято называть побочные грани кристалла, слабо отклоненные от какой-либо сингулярной грани на угол менее 5° . При отклонении от сингулярной грани в одном направлении вицинальные поверхности образуют «лестницу», состоящую из террас в виде участков основной грани и ступеней монокристаллической высоты (рис. 5.13 б). Ширина террас определяется углом наклона. При отклонении в двух направлениях ступени содержат регулярно повторяющиеся изломы (рис. 5.13 в). В процессе роста на одной грани кристалла может возникать несколько вициналей. Из-за существования ступеней и изломов вицинальные поверхности характеризуются более высокой поверхностной энергией, чем сингулярные грани.

Несингулярные грани, называемые также диффузными, составляют достаточно большие углы с сингулярными гранями (рис. 5.14 а). Они имеют большую концентрацию ступеней и изломов, что сильно повышает их поверхностную энергию.

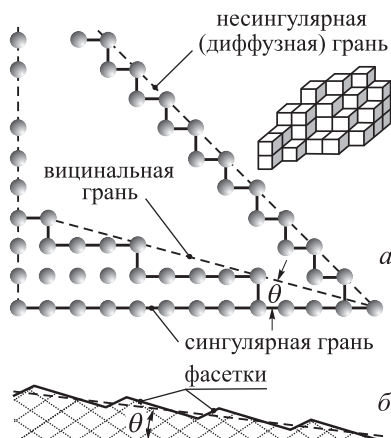


Рис. 5.14. а — схематическое изображение различных граней кубического кристалла, б — фасетированная вицинальная поверхность

Качественная картина расположения атомов для трех типов граней — сингулярной, вицинальной и несингулярной (диффузной), схематически изображена на рис. 5.14 а.

С ростом угла θ (наклона вицинали) увеличивается энергия, запасаемая атомами вицинальной поверхности, что приводит к ее неустойчивости. В этом случае энергетически выгоднее образование на месте вицинали двух (или более) микрограней, называемых *фасетками* (от фр. *facette* «скошенная грань»), как показано на рис. 5.14 б. Хотя результирующая площадь поверхности возрастает, общая энергия кристалла снижается из-за меньшей удельной поверхностной энергии образовавшихся фасеток.

Методом СТМ наблюдалось фасетирование вицинальной поверхности Si(111), разориентированной в направлении $[11\bar{2}]$ на угол 10° . При 620°C возникали фасетки (331) с углом наклона 22° и широкие фасетки (111), покрытые реконструкцией 7×7 .

Как будет показано ниже (см. п. 5.15), фасетированные поверхности представляют большой интерес для нанотехнологий, поскольку на них наблюдается спонтанное формирование упорядоченных монослойных и многослойных гетероструктур.

5.5. Термодинамика поверхностных слоев и межфазных границ

Предыдущие параграфы данной главы показали продуктивность *атомно-молекулярного подхода* к описанию свойств поверхности, по крайней мере, на уровне качественного понимания

поверхностных свойств. Дальнейшее развитие такого подхода требует существенной детализации атомно-молекулярной структуры поверхности на основе методов квантовой химии. Они позволяют рассчитывать атомные и молекулярные орбитали, характеризовать структуру молекулярных уровней в наночастицах и тем самым определять их электрические, магнитные, оптические, тепловые и другие свойства. Однако подобная квантово-механическая детализация в описании поверхностных явлений чрезвычайно сложна. Базируясь на анализе частных моделей, она, к сожалению, не может дать общих представлений, применимых к широкому кругу физических процессов на поверхности.

В качестве единственной возможности в этом направлении остается *термодинамический подход*. К счастью, такой подход сохраняет свою применимость и к наночастицам, содержащим достаточно большое число атомов и молекул, измеряемых по крайней мере тысячами. Термодинамический подход к анализу свойств и синтезу наночастиц позволяет определять закономерности их образования и роста, фазовый состав и свойства, структурные и физико-химические изменения в ходе фазовых превращений и многое другое.

1. Общая характеристика поверхности. *Поверхностью* называют границу конденсированной фазы (жидкой или твердой) с парогазовой средой, при этом атомарно-чистая поверхность достижима лишь при сверхвысоком вакууме с давлением остаточных газов менее 10^{-9} мм рт. ст. Границу раздела двух конденсированных фаз (кристалл–кристалл, кристалл–жидкость, жидкость–жидкость) принято называть *межфазной границей*.

Состояние частицы (молекулы, атома, иона), находящейся в приповерхностном слое конденсированной фазы, отличается от состояния той же частицы внутри объема этой фазы. Такое отличие качественно демонстрирует рис. 5.15.

Внутри твердого тела атомы, ионы или молекулы, образующие его структуру, располагаются в пространстве в соответствии с кристаллической решеткой, характерной для данного вещества. Взаимодействие частицы со своим окружением уравновешено по всем направлениям (рис. 5.15 *а*), так что внутренние области вещества отвечают требованию минимума свободной энергии.

В противоположность этому, для частицы приповерхностного слоя взаимодействия с ближайшими соседями не уравновешены, так как различны силы, действующие со стороны внутренних слоев вещества и со стороны смежной фазы. При этом их равнодействующая направлена внутрь объема твердого тела

(рис. 5.15 б). Это приводит к тому, что в приповерхностном (межфазном) слое возникает дополнительное *поле межмолекулярных* (межатомных или межионных) *сил*, изменяющее термодинамические свойства этого слоя по отношению к объемным свойствам смежных фаз. Поэтому приповерхностные слои и межфазные границы являются особыми областями, где атомные и молекулярные конфигурации, электронная структура и локальные физические характеристики иные, чем в объеме вещества. Возникающие при этом изменения термодинамически проявляются двояким образом:

- во-первых, в дополнительном силовом поле запасается избыточная поверхностная (или межфазная) энергия, характеризующаяся **поверхностным натяжением** σ ;
- во-вторых, это поле изменяет концентрацию частиц в поверхностном слое, что характеризуется так называемой **гиббсовской адсорбцией** Γ_i для частиц i -го сорта.

Термодинамические основы поверхностных явлений восходят ко временам Дж. Гиббса (см. сноску на с. 74). Поэтому ниже будем следовать анализу, впервые изложенному Гиббсом.

2. Элементы термодинамики поверхности. Для выяснения физического смысла термодинамических величин σ и Γ_i , вновь введенных для описания поверхностных свойств, Гиббс предложил рассматривать межфазный переходный слой, показанный на рис. 5.16, в качестве самостоятельной фазы. Термодинамические свойства переходного слоя отличаются от свойств смежных объемных фаз I и II. В частности, отличаются внутренняя энергия U , свободная энергия F , энтропия S и число молей n_i каждого i -го компонента ($i = 1, 2, \dots, K$). Эти отличия принято задавать в виде избытков соответствующих величин:

$$\begin{aligned} \Delta U &= U - (U^I + U^{II}), & \Delta F &= F - (F^I + F^{II}), \\ \Delta S &= S - (S^I + S^{II}), & \Delta n_i &= n_i - (n_i^I + n_i^{II}). \end{aligned} \quad (5.5.1)$$

Здесь верхние индексы I и II характеризуют значения величин в объемах V^I и V^{II} , занимаемых фазами I и II до возникновения на их месте межфазного переходного слоя (см. рис. 5.16).

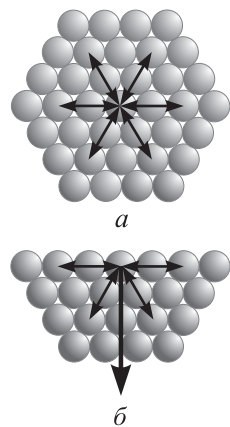


Рис. 5.15. Схема сил взаимодействия частиц в объеме (а) и на поверхности (б)

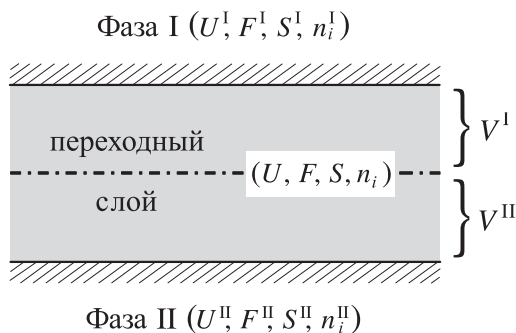


Рис. 5.16. К выводу поверхностных термодинамических соотношений. Штрих-пунктирная линия соответствует гипотетической резкой границе между смежными фазами в отсутствие переходного слоя

Следовательно, объем переходного слоя равен $V = V^I + V^{II}$, тогда

$$\Delta V = V - (V^I + V^{II}) = 0. \quad (5.5.2)$$

Согласно (5.5.2), образование переходного слоя происходит в условиях неизменности объема, занимаемого исходными фазами. Поэтому термодинамической функцией, наиболее подходящей для описания поверхностных явлений, является *изохорный потенциал* F (см. п. 1.8).

Поверхностное натяжение вводится как *удельная поверхностная энергия*, т. е. избыточная свободная энергия Гельмгольца, приходящаяся на единицу площади поверхности:

$$\sigma = \frac{\Delta F}{\Delta A} \left(\frac{\text{Дж}}{\text{м}^2} \right), \quad (5.5.3)$$

где ΔA — площадь элемента поверхности границы раздела.

Следовательно, термодинамический смысл поверхностного натяжения состоит в том, что оно выражает работу внешних сил, затрачиваемую на увеличение границы раздела фаз на единицу площади. Поверхностное натяжение определяет энергию перехода частиц из объема в поверхностный слой и поэтому является одной из стимулирующих причин для процессов релаксации и реконструкции поверхности твердых тел (см. п. 5.3).

С другой стороны, поверхностное натяжение имеет и механическую интерпретацию. Действительно, энергетическая размерность σ (Дж/м²) может быть записана в силовых единицах как Н/м. Тогда поверхностное натяжение выражает силу, которую надо приложить к единице длины линии, ограничивающей поверхность раздела фаз, для увеличения площади этой поверхности на единицу (см. ниже рис. 5.18 б и 5.48).

Гиббсовская адсорбция вводится, по аналогии с (5.5.3), как *удельный избыток молей* i -го компонента в поверхностном слое:

$$\Gamma_i = \frac{\Delta n_i}{\Delta A} \left(\frac{\text{моль}}{\text{м}^2} \right). \quad (5.5.4)$$

Следует подчеркнуть, что силовое поле, возникающее в переходном слое, всегда *запасает* избыточную энергию, что однозначно обеспечивает $\sigma > 0$. В то же время это поле, в зависимости от природы частиц и их взаимодействия между собой, может либо втягивать частицы внутрь слоя, либо выталкивать из него, обеспечивая алгебраические избытки Δn_i разных знаков. Этому соответствует $\Gamma_i > 0$ или $\Gamma_i < 0$ при увеличении или уменьшении числа частиц i -го сорта в поверхностном слое по сравнению с объемными фазами.

Поверхностная энтропия может быть также определена как *удельная избыточная энтропия* поверхностного слоя:

$$S_s = \frac{\Delta S}{\Delta A} \left(\frac{\text{Дж}}{\text{м}^2 \cdot \text{К}} \right). \quad (5.5.5)$$

Введенные поверхностные физические величины (5.5.3)–(5.5.5), будучи отнесенными к единице площади переходного слоя, никоим образом не учитывают его толщину. Это позволило Гиббсу рассматривать переходный слой между двумя фазами как *разделяющую поверхность нулевой толщины*. Такой разделяющей поверхности приписываются физические свойства, характеризующиеся поверхностным натяжением (5.5.3), гиббсовской адсорбцией (5.5.4) и поверхностной энтропией (5.5.5).

В модели переходного слоя нулевой толщины условие (5.5.2) уже не является определяющим, так как $V \equiv 0$. Поэтому можно воспользоваться разными внешними условиями, наложенными на систему, для определения поверхностного натяжения. Можно записать σ в виде частных производных термодинамических потенциалов по площади поверхности при надлежащем выборе в качестве постоянных соответствующих физических величин:

$$\sigma = \left(\frac{\partial U}{\partial A} \right)_{S, V, n_i} = \left(\frac{\partial H}{\partial A} \right)_{S, P, n_i} = \left(\frac{\partial F}{\partial A} \right)_{T, V, n_i} = \left(\frac{\partial G}{\partial A} \right)_{T, P, n_i}.$$

Далее воспользуемся изохорно-изотермическими условиями и запишем результирующий избыток изохорного потенциала (ΔF_Σ) для поверхностного слоя в виде двух слагаемых:

$$\Delta F_\Sigma = \sigma \Delta A + \sum_i \mu_i \Delta n_i. \quad (5.5.6)$$

Первое слагаемое учитывает увеличение свободной энергии за счет сил поверхностного натяжения ($\sigma > 0$). Второе слагаемое порождено изменением энергии за счет алгебраического избытка молей в поверхностном слое (Δn_i больше или меньше нуля), при этом μ_i — химический потенциал частиц i -го сорта.

Разделив левую и правую части равенства (5.5.6) на площадь элемента границы раздела фаз (ΔA), получим выражение для удельной свободной энергии поверхностного слоя $F_s \equiv \Delta F_\Sigma / \Delta A$:

$$F_s = \sigma + \sum_i \mu_i \Gamma_i. \quad (5.5.7)$$

Поверхностное натяжение σ и гиббсовская адсорбция Γ_i обязательно должны быть связаны между собой, поскольку их образование вызвано общей физической причиной, а именно полем поверхностных молекулярных сил.

Для установления этой связи рассмотрим межфазную поверхность площадью A . Используя выражение (5.5.7), запишем для межфазного слоя (нулевой толщины) полный дифференциал всей свободной поверхностной (избыточной) энергии $F_s A$:

$$\begin{aligned} d(F_s A) &= d(\sigma A) + d \sum_i \mu_i (\Gamma_i A) = \\ &= \sigma dA + A d\sigma + \sum_i \mu_i d(\Gamma_i A) + \sum_i (\Gamma_i A) d\mu_i. \end{aligned} \quad (5.5.8)$$

Выразим дифференциал $d(F_s A)$ другим способом, исходя из выражения (1.10.3) для объемной фазы, записанного в виде

$$dF = -SdT + (-PdV) + \sum_i \mu_i dn_i. \quad (5.5.9)$$

Здесь второе слагаемое, $(-PdV) > 0$, отражает работу, производимую над системой при ее сжатии ($dV < 0$), что увеличивает энергию системы.

По аналогии с объемным выражением (5.5.9) для межфазного слоя, делая замены $S \rightarrow S_s A$, $-P \rightarrow \sigma$, $V \rightarrow A$, $n_i \rightarrow \Gamma_i A$, записываем дифференциал свободной поверхностной энергии как

$$d(F_s A) = -(S_s A)dT + \sigma dA + \sum_i \mu_i d(\Gamma_i A). \quad (5.5.10)$$

Поскольку поверхностное натяжение всегда положительно, то работа $\sigma dA > 0$, производимая при увеличении площади поверхностного слоя, приводит к росту запасенной в нем энергии, подобно работе $(-PdV)$ в равенстве (5.5.9), совершаемой внешней средой против давления P в объемной фазе.

Приравнивая правые части формул (5.5.8) и (5.5.10), получаем известную **адсорбционную формулу Гиббса**:

$$d\sigma = -S_s dT - \sum_i \Gamma_i d\mu_i. \quad (5.5.11)$$

Из выражения (5.5.11), как полного дифференциала, следует связь между гиббсовской адсорбцией i -го компонента и поверхностным натяжением в виде дифференциального соотношения

$$\Gamma_i = - \left(\frac{\partial \sigma}{\partial \mu_i} \right)_{T, \mu_j \neq \mu_i}. \quad (5.5.12)$$

Аналогично из адсорбционной формулы (5.5.11) получается выражение для полной поверхностной энтропии $S = S_s A$:

$$S = -A \left(\frac{\partial \sigma}{\partial T} \right)_{\mu_i}. \quad (5.5.13)$$

Рассмотрим обратимый процесс изотермического изменения площади межфазного слоя от значения A_1 до A_2 . Тогда, учитывая соотношение (5.5.13), получаем в соответствии с формулой (1.6.4) количество теплоты, выделяемой/поглощаемой в ходе рассматриваемого процесса:

$$Q = T(S_2 - S_1) = T(A_1 - A_2) \left(\frac{\partial \sigma}{\partial T} \right)_{\mu_i}. \quad (5.5.14)$$

Следовательно, увеличение площади поверхности сопровождается поглощением теплоты ($Q > 0$), а уменьшение — выделением тепла ($Q < 0$), так как обычно поверхностное натяжение уменьшается с повышением температуры ($\partial \sigma / \partial T < 0$).

Вышеприведенные термодинамические соотношения получены в предположении, что процесс образования новой поверхности не сопровождается изменениями структуры конденсированной фазы. В действительности такое предположение справедливо лишь для жидкостей, когда изменениям подвергается та их часть, которая переходит в поверхностный слой. При этом жидкости, будучи изотропными, имеют одно и то же поверхностное натяжение σ в любой точке поверхности. В случае твердых тел изменения, происходящие при образовании границы, обычно затрагивают также и объемные фазы.

Атомы кристаллической фазы, выходя на поверхность, начинают изменять свое положение таким образом, чтобы уравновесить силы взаимодействия с атомами в объеме и максимально снизить избыточную поверхностную энергию. Этот процесс сопровождается деформацией поверхности и совершением работы.

Согласно *принципу Гиббса–Кюри*, термодинамически более устойчивой является та форма кристалла, которая обладает минимальной свободной поверхностной энергией, то есть

$$F = \sum_k \sigma_k A_k = \min \quad \text{при } V = \text{const.} \quad (5.5.15)$$

Здесь σ_k и A_k — поверхностное натяжение и площадь k -й грани кристалла. Поскольку для жидкостей в силу изотропности их свойств всегда $\sigma_k = \text{const}$ во всех точках поверхности, то принцип Гиббса–Кюри требует минимизации площади поверхности: $A \equiv \sum_k A_k = \min$ при $V = \text{const}$.

Таким образом, термодинамически устойчивая форма жидкости имеет минимальную поверхность. Поэтому любая жидкость в свободном состоянии (в невесомости) принимает форму шара.

Кристаллические тела принимают в равновесии более сложную форму, поскольку кристаллографические грани имеют разную поверхностную энергию. В кристаллах при формировании поверхности важную роль играет ее ориентация по отношению к силам связи между атомами по разным направлениям. Минимальная поверхностная энергия будет у тех поверхностей, на которых располагается наибольшее количество сильных связей и которые имеют минимальное число пересечений с ними. Например, для гранецентрированной кубической решетки плотноупакованная грань (111) обладает наименьшим поверхностным натяжением (минимальной запасенной энергией).

Следовательно, в анизотропных телах минимум поверхностной энергии достигается при такой форме, которая обеспечивает больший вклад кристаллографических граней с меньшей поверхностной энергией при одновременной минимизации суммарной площади поверхности. Для определения термодинамически устойчивой формы монокристаллов существует геометрическое *построение Вульфа*, применимое к выпуклым многогранникам.

Для этого из некоторой точки, принимаемой за центр кристалла, откладывают по нормали к возможной k -й грани отрезок длиной l_k , пропорциональный соответствующему значению σ_k , и замыкают этот отрезок перпендикулярной ему плоскостью. Совокупность плоских граней с площадями A_k , возникающих при взаимном пересечении таких плоскостей, дает равновесную форму кристалла. Плоскости, наиболее удаленные от центра и не пересекающиеся с другими плоскостями, соответствуют возможным, но реально не существующим (неустойчивым) граням.

Термодинамически устойчивая равновесная форма монокристалла удовлетворяет соотношению (*правило Кюри–Вульфа*):

$$\frac{\sigma_1}{l_1} = \frac{\sigma_2}{l_2} = \dots = \frac{\sigma_k}{l_k} = \text{const.} \quad (5.5.16)$$

Отсюда следует, что чем больше поверхностное натяжение σ_k на k -й грани, тем больше расстояние l_k этой грани от центра кристалла и меньше ее площадь A_k . По мере увеличения σ_k данная грань приближается к своему пределу неустойчивости ($A_k \rightarrow 0$).

Следовательно, в условиях неустойчивости, когда многогранный кристалл продолжает расти, кристаллизация происходит в первую очередь за счет неустойчивых граней. Скорость роста таких граней пропорциональна их удельной поверхностной энергии σ_k , при этом разрастание стабильных граней отстает. Как результат этого, в конечном итоге исчезают неустойчивые грани, а стабильные увеличиваются по площади. На заключительной стадии роста кристалл ограничен устойчивыми гранями с наименьшими скоростями роста, то есть с минимальными σ_k , удовлетворяющими равновесным условиям (5.5.16).

Если одна из граней кристалла имеет поверхностную энергию намного меньшую, чем другие грани, то почти вся поверхность кристалла (за исключением небольших закруглений вблизи ребер) будет образована именно такими гранями. Для модели ТСИ кубического кристалла равновесной формой является куб, образованный сингулярными плоскостями. Примером являются ионные кристаллы поваренной соли, у которых все шесть граней образованы одинаковыми плоскостями $\{100\}$, обладающими наименьшей поверхностной энергией.

Таким образом, грани с наименьшим поверхностным натяжением (имеющие минимальную скорость роста) развиты сильнее других граней и доминируют при формообразовании кристаллов. Для неполярных (ковалентных) кристаллов такими плоскостями являются плоскости: $\{111\}$ в ГЦК-решетке и решетке алмаза, $\{110\}$ в ОЦК-решетке и $\{0001\}$ в ГПУ-решетке.

5.6. Поверхностно-активные вещества, термодинамические свойства и технологии. Пленки Ленгмюра–Блоджетт

1. Общая характеристика поверхностно-активных и ин-активных веществ. Воспользуемся формулой (5.5.12), связывающей гиббсовскую адсорбцию i -го компонента (Γ_i) с поверхностным натяжением раствора (σ), для того чтобы качественно проанализировать поведение растворенных веществ в воде и выяснить их влияние на поверхностные свойства растворов.

Пусть водный раствор имеет достаточно низкую концентрацию c_i растворенного вещества i -го сорта, так что для его химического потенциала можно использовать формулу (1.13.2) для идеального раствора, а именно

$$\mu_i(T, c_i) = \mu_i^\circ(T) + RT \ln c_i. \quad (5.6.1)$$

С учетом производной $\partial\mu_i/\partial c_i = RT/c_i$, получаемой из формулы (5.6.1), перепишем выражение (5.5.12) в следующей форме:

$$\Gamma_i \equiv -\frac{\partial\sigma}{\partial\mu_i} = -\frac{\partial\sigma}{\partial c_i} \frac{\partial c_i}{\partial\mu_i} = -\frac{c_i}{RT} \frac{\partial\sigma}{\partial c_i}, \quad (5.6.2)$$

называемой *уравнением Гиббса*. Производная $\partial\sigma/\partial c_i$ служит мерой *поверхностной активности*, т. е. способности растворенного вещества изменять поверхностное натяжение раствора.

Рассмотрим термодинамические следствия, вытекающие из соотношения (5.6.2), в применении к двум типам растворенного вещества — к *поверхностно-активным веществам* (ПАВ) и к *поверхностно-инактивным веществам* (ПИВ).

Поверхностно-активные вещества, введенные в растворитель (воду), ведут себя таким образом, что с увеличением их концентрации c_i *снижается* поверхностное натяжение раствора ($\partial\sigma/\partial c_i < 0$), как показано на рис. 5.17 а. По формуле (5.6.2) гиббсовская адсорбция такого вещества всегда положительна: $\Gamma_i > 0$ (рис. 5.17 б). Согласно выражениям (5.5.1) для Δn_i и (5.5.4) для Γ_i , получаем $n_i > (n_i^I + n_i^{II})$. Следовательно, концентрация вещества в поверхностном слое превышает его концентрацию в растворе, то есть такие вещества накапливаются на поверхности, что дает им название ПАВ.

Для ПАВ на рис. 5.17 а и б приведены два типа наблюдаемых концентрационных зависимостей σ и Γ_i , изображенные сплошной и штриховой линиями. К растворам с подобными свойствами относятся растворы с положительным отклонением от закона Рауля. Для них процесс растворения является *эндотермическим* ($\Delta H > 0$ — см. п. 2.5). Это свойство является следствием того, что взаимодействие между одинаковыми молекулами растворителя сильнее, чем между разнородными молекулами растворителя и растворенного вещества.

На этом основании энергетически невыгодно вводить в объем растворителя молекулы поверхностно-активного вещества, и они преимущественно вытесняются из раствора на поверхность раствора ($\Gamma_i > 0$). Вследствие накопления ПАВ на поверхности молекулярное взаимодействие в поверхностном слое уменьшается и поверхностное натяжение падает.

По отношению к воде поверхностно-активными веществами являются органические соединения, например сложные углеводороды в виде солей высших жирных кислот и спиртов типа мыла, водные растворы которых используют для очистки поверхностей от загрязнений. Молекулы таких соединений в форме длинного углеводородного радикала («хвоста»), несмачиваемого водой, имеют на одном конце полярную группу (типа COOH , OH , NH_2 и др., см. рис. 5.20), обладающую дипольным моментом. Эта группа эффективно взаимодействует с полярными молекулами воды. Молекулы ПАВ, стремясь уменьшить энергию системы, самопроизвольно образуют ориентированный монослой на поверхности раздела фаз. Полярные группы ПАВ располагаются в полярной (водной) фазе, а несмачиваемые водой радикалы вытесняются из ее среды и переходят в менее полярную фазу (воздух), располагаясь на поверхности раствора вертикально «хвостами» вверх (см. ниже рис. 5.21). Механизмы очистки путем удаления поверхностных загрязнений с помощью ПАВ рассмотрены ниже.

В последнее время усилился интерес к ПАВ в связи с возможностью формирования структурно-упорядоченных мономолекулярных пленок путем перенесения адсорбционных монослоев на твердую поверхность по *методу Ленгмюра–Блоджетт*¹⁾, описанная ниже, относительно проста и позволяет воспроизводимо получать двумерные моно- и мультимолекулярные слои на твердых подложках.

Поверхностно-инактивные вещества, введенные в растворитель (воду), ведут себя таким образом, что с увеличением концентрации c_i *возрастает* поверхностное натяжение раствора

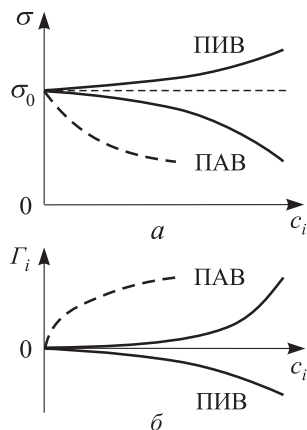


Рис. 5.17. Качественный ход концентрационных зависимостей поверхностного натяжения (а) и гиббсовской адсорбции (б)

¹⁾ Ленгмюр (Langmuir) Ирвинг (1881–1957) — американский ученый, работал в области поверхностных и адсорбционных явлений, вывел уравнение изотермы адсорбции, совместно со своей ассистенткой К. Блоджетт показал, что разреженные мономолекулярные слои на поверхности жидкостей обладают свойствами двумерного газа. Исследуя электрические разряды в газах, ввел для ионизованного газа термин «плазма», разработал теорию и способ измерения электронной температуры в плазме. Нобелевская премия по химии (1932) «за открытия и исследования в области химии поверхностных явлений».

($\partial\sigma/\partial c_i > 0$), как показано на рис. 5.17 а. По формуле (5.6.2) гиббсовская адсорбция такого вещества всегда отрицательна: $\Gamma_i < 0$ (рис. 5.17 б). Согласно выражениям (5.5.1) для Δn_i и (5.5.4) для Γ_i , получаем $n_i < (n_i^I + n_i^{II})$. Следовательно, концентрация вещества в поверхностном слое меньше, чем в растворе, то есть такие вещества вытесняются с поверхности и распределяются внутри объема, что дает им название ПИВ.

Для ПИВ характерно отрицательное отклонение от закона Рауля из-за экзотермичности процесса растворения ($\Delta H < 0$). Более сильное взаимодействие между разнородными молекулами (ионами) растворителя и растворенного вещества способствует интенсивному растворению с преимущественным размещением растворенных частиц внутри раствора. Повышение поверхностного натяжения с ростом концентрации ПИВ обычно невелико и становится заметным лишь в области высоких концентраций.

По отношению к воде поверхностно-инактивными веществами являются обычные неорганические кислоты, щелочи и соли, которые при растворении в воде образуют сильные электролиты. В них растворенные вещества диссоциированы на ионы, преимущественно присутствующие в объеме электролита, да и то в гидратированной форме.

2. Роль ПАВ в удалении поверхностных загрязнений.

Современные методы очистки поверхности могут быть разделены на две большие группы, использующие:

а) химически активные травители (типа кислот и щелочей) и электрохимические процессы;

б) нейтральные жидкие растворители (типа трихлорэтилена C_2HCl_3 , четыреххлористого углерода CCl_4 , этилового спирта C_2H_5OH , дистиллированной и деионизованной воды).

В первом случае химическое травление разрушает молекулы поверхностных загрязнений одновременно с частичным удалением приповерхностных слоев материала. Это делает очищаемую поверхность активированной, т. е. способной к повторному взаимодействию с загрязнениями из окружающей среды.

По этой причине более предпочтительными (особенно на заключительной стадии) оказываются методы очистки, которые приводят не к активации, а к пассивации очищенной поверхности путем сохранения на ней мономолекулярного слоя, состоящего из молекул растворителя. Этот пассивирующий монослой предохраняет поверхность от повторных загрязнителей и легко удаляется при последующем термическом отжиге металлических деталей в вакууме или в атмосфере водорода.

Механизм удаления поверхностных загрязнений с помощью химически нейтральных растворителей (обычно называемых в этом случае *моющими жидкостями*) сначала рассмотрим без участия молекул ПАВ, а затем качественно опишем их роль.

Начнем с выяснения термодинамических условий для реализации моющей способности жидкости путем вытеснения ее молекулами молекул жирового загрязнения, покрывающего поверхность подложки. На рис. 5.18 *а* показана качественная картина положения частицы загрязнения в прилипшем к поверхности состоянии I и в оторвавшемся от поверхности состоянии II.

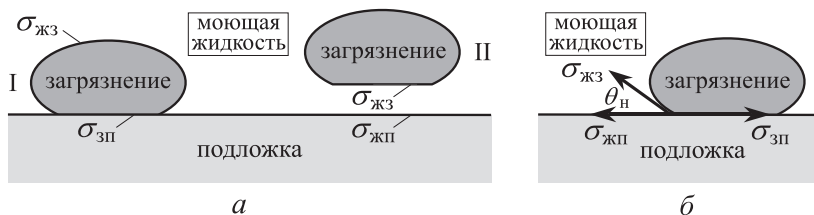


Рис. 5.18. К объяснению механизма удаления поверхностных загрязнений моющей жидкостью: *а* — поверхностные натяжения на границах загрязнение–подложка ($\sigma_{зп}$), жидкость–подложка ($\sigma_{жп}$), жидкость–загрязнение ($\sigma_{жз}$); *б* — диаграмма сил поверхностного натяжения для несмачиваемого (жирового) загрязнения с краевым углом несмачивания θ_n

В начальном состоянии (прилипшая частица I) граница раздела загрязнение–подложка площадью A запасает избыточную свободную энергию $G_n = \sigma_{зп}A$. В конечном состоянии (оторвавшаяся частица II) исходная граница заменяется двумя вновь возникшими границами: жидкость–подложка и жидкость–загрязнение, в которых запасена энергия $G_k = (\sigma_{жп} + \sigma_{жз})A$.

Таким образом, изменение свободной поверхностной энергии системы моющая жидкость–загрязненная поверхность, происходящее при отрыве частицы загрязнения от поверхности, равно

$$\Delta G \equiv G_k - G_n = (\sigma_{жп} + \sigma_{жз} - \sigma_{зп})A. \quad (5.6.3)$$

Самопроизвольный отрыв частицы возможен при условии, что $\Delta G < 0$. С учетом (5.6.3), это накладывает требование

$$\sigma_{зп} > \sigma_{жп} + \sigma_{жз}. \quad (5.6.4)$$

Для капли жидкости (масла), несмачивающей твердую поверхность, диаграмма сил поверхностного натяжения изображена на рис. 5.18 *б*. При угле несмачивания θ_n из условия равновесия сил получаем

$$\sigma_{зп} = \sigma_{жп} + \sigma_{жз} \cos \theta_n. \quad (5.6.5)$$

Поскольку в реальных условиях $\theta_n \geq 0$, то равенство (5.6.5), подставленное в формулу (5.6.3), обеспечивает условие $\Delta G \geq 0$. Оно гарантирует устойчивое существование капли на поверхности вплоть до состояния полной несмачиваемости при $\theta_n = 0$. Требование отрыва капли от поверхности в форме (5.6.4) математически может быть реализовано при мнимом угле $\theta_n = 0 + i\theta'_n$, когда $\cos \theta_n = \operatorname{ch} \theta'_n > 1$. Физически это означает неустойчивость границы раздела загрязнение–подложка и, как результат этого, возникновение двух новых границ раздела жидкость–подложка и жидкость–загрязнение, происходящее при отрыве частицы.

Следовательно, при выполнении условия (5.6.4) молекулы моющей жидкости в силу своей большей сорбционной активности вытесняют молекулы загрязнений с адсорбционных центров. Последние удерживаются на поверхности за счет сил физической адсорбции (см. ниже п. 5.9) и на энергетической диаграмме размещаются в потенциальной яме глубиной U_Φ , показанной на рис. 5.19 а (см. также рис. 5.32 и соответствующий ему текст).

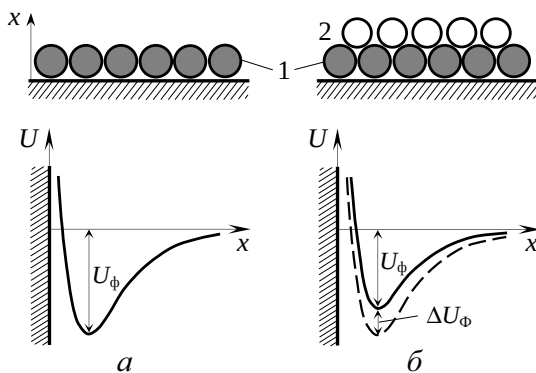


Рис. 5.19. Энергетическая диаграмма для физически адсорбированных молекул загрязнения 1 на поверхности тела (а) и то же (б) при наличии молекул моющей жидкости 2, уменьшающих глубину потенциальной ямы для молекул загрязнения на величину ΔU_Φ

Каждая молекул живет в адсорбированном состоянии в среднем довольно короткое время τ_a (порядка нескольких микросекунд), даваемое ниже формулой (5.9.1). После отрыва молекулы загрязнения от поверхности ее место занимает молекула растворителя. Поскольку последних много больше (особенно при наличии потока моющей жидкости), то поверхность должна постепенно очищаться от загрязнений. Однако простое замещение одних молекул (загрязнения) другими молекулами (растворителя) не может обеспечить эффективную очистку.

Дело в том, что, согласно распределению Больцмана, число десорбируемых молекул ($N_{\text{дес}}$) с энергией, превышающей энергию десорбции $U_{\text{ф}}$, очень мало по отношению к полному числу адсорбированных молекул ($N_{\text{адс}}$). Например, при комнатной температуре и $U_{\text{ф}} = 0,43$ эВ (что соответствует теплоте адсорбции $Q_{\text{ф}} = 10$ ккал/моль) получаем $N_{\text{дес}}/N_{\text{адс}} = \exp(-U_{\text{ф}}/k_{\text{В}}T) \approx 10^{-7}$, т. е. число энергичных молекул ничтожно мало.

Более значительным фактором, влияющим на эффективность очистки, является уменьшение глубины потенциальной ямы $U_{\text{ф}}$ в результате взаимодействия между молекулами загрязнения и растворителя, что отражено на рис. 5.19 б.

Возникновение новой межмолекулярной связи ослабляет существующую сорбционную связь молекул загрязнения с поверхностью. В результате этого $U_{\text{ф}}$ уменьшается на величину $\Delta U_{\text{ф}}$, что резко сокращает время десорбции $\tau_{\text{а}} = \tau_0 \exp(U_{\text{ф}}/k_{\text{В}}T)$. Поэтому любая нейтральная жидкость, в том числе вода, может играть и играет роль моющей жидкости. Для повышения моющего эффекта загрязненная поверхность должна смачиваться жидкостью, что усиливает взаимодействие молекул растворителя с молекулами загрязнения и достигается использованием мыла.

Отрыв молекул загрязнения от поверхности является первым этапом процесса очистки. Следом за ним идет стадия транспорта оторвавшихся частиц от поверхности в глубь раствора, предотвращающая их обратное осаждение. В этом существенная роль отводится диффузии даже при наличии потока моющей жидкости. Дело в том, что на твердой поверхности, омываемой вязкой жидкостью, отсутствуют тангенциальные скорости движения жидкости. Поэтому единственной физической причиной транспорта частиц по нормали вблизи поверхности в те слои, где скорость тангенциального потока становится существенной, является диффузия.

Оторвавшиеся от поверхности молекулы загрязнения образуют коллоидную среду (см. ниже), состоящую из так называемых *мицелл* (от лат. «крошка») — сложных молекулярных комплексов в форме плотноупакованного роя перестраивающихся молекул. Поэтому, говоря о диффузионном движении частиц, надо иметь в виду эффективную диффузию именно мицелл.

Для качественного рассмотрения воспользуемся известной формулой Стокса $D = k_{\text{В}}T/6\pi\eta a$ (см. ниже уравнение (5.7.5)), дающей коэффициент диффузии частицы радиусом a (мицеллы) в жидкости с вязкостью η . Из нее видно, что с ростом температуры коэффициент диффузии возрастает из-за снижения вязкости, так как $\eta \propto \exp(\mathcal{E}_{\text{в}}/k_{\text{В}}T)$, где $\mathcal{E}_{\text{в}}$ — энергия активации.

Одновременно с этим повышение интенсивности теплового движения частиц увеличивает вероятность распада мицелл на составляющие ее молекулы. Полагая эффективный радиус мицелл равным a , приходим к заключению, что процесс их распада (сопровождающийся уменьшением a) увеличивает скорость диффузионного транспорта (за счет роста эффективного коэффициента диффузии D). Отсюда повышение температуры моющей жидкости должно увеличивать результативность очистки.

Однако, как ясно из ранее сказанного, с ростом температуры снижается эффективность предшествующего этапа отрыва молекул загрязнения из-за ослабления их связи с молекулами моющей жидкости. Конкуренция этих двух противоборствующих факторов приводит к некоторой оптимальной температуре промывки.

Поверхностно-активные вещества нередко добавляются в

моющую жидкость для повышения эффективности очистки. Дело в том, что условие отрыва частицы загрязнения от поверхности в форме (5.6.4) наиболее эффективно реализуется при снижении обоих поверхностных натяжений $\sigma_{жп}$ и $\sigma_{жз}$. Для этого моющая жидкость должна достаточно хорошо смачивать как очищаемую подложку (для понижения $\sigma_{жп}$), так и загрязнение (для понижения $\sigma_{жз}$).

Именно это достигается введением в растворитель поверхностно-активного вещества, одновременно адсорбирующегося на границах раздела жидкость–подложка и жидкость–загрязнение с понижением их поверхностных натяжений.

Как уже говорилось, ПАВ относятся к классу высокомолекулярных углеводородных соединений (таких как соли высших жирных кислот типа мыла и т. п.).

Рассмотрим структуру ПАВ на примере молекулы стеариновой кислоты, схематически изображенной на рис. 5.20.

Молекула стеариновой кислоты $\text{CH}_3(\text{CH}_2)_{16}\text{COOH}$ состоит из углеводородной 16-звенной цепочки $-\text{CH}_2-\dots-\text{CH}_2-$, заканчивающейся на одном (верхнем) конце группой CH_3 , а на другом (нижнем) конце — карбоксильной группой COOH (головкой). Углеводородный радикал $\text{CH}_3-\text{CH}_2-\dots-\text{CH}_2$ (хвост) является *гидрофобным* (несмачиваемым водой), в то время как другая

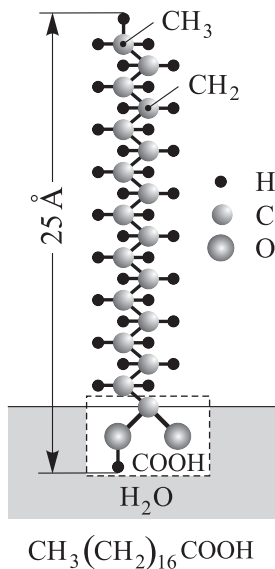
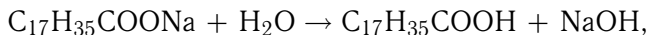


Рис. 5.20. Схематическое изображение молекулы стеариновой кислоты

концевая группа (в данном случае карбоксильная COOH), будучи полярной, является *гидрофильной* (смачиваемой водой)¹⁾.

Таким образом, поверхностная активность молекулы ПАВ обусловлена гидрофобным (неполярным) радикалом, а растворимость в воде — гидрофильной (полярной) группой.

Следует заметить, что в состав любого мыла входят молекулы не самой стеариновой кислоты (и ей подобных), а ее натриевой соли $\text{CH}_3(\text{CH}_2)_{16}\text{COONa}$ (см. с. 608), получаемой заменой атома водорода натрием, которая сохраняет поверхностную активность. В водных растворах мыло гидролизуется (то есть соединяется с водой) по следующей реакции:



в результате чего раствор приобретает щелочной характер.

Для объяснения механизма действия молекул ПАВ используем рис. 5.21, где полярная группа (головка ПАВ) изображена кружочком, а углеводородный радикал (хвост ПАВ) — палочкой. Полярная группа электрически взаимодействует с дипольными молекулами воды, в то время как углеводородный хвост, являясь гидрофобным, не смачивается водой.

Поэтому вода стремится вытеснить хвосты молекул ПАВ из своей среды, и они располагаются на поверхности раствора вертикально наружу, образуя так называемый «частько́л Ленгмюра», обозначенный цифрой 4 на рис. 5.21. По той же причине молекулы ПАВ, оставшиеся в объеме жидкости, стремясь экранировать свои гидрофобные хвосты от молекул воды, адсорбируются на разных поверхностях (загрязнениях и газовых пузырьках) таким образом, чтобы их полярные группы были обращены наружу, как показано цифрами 1, 2 и 3 на рис. 5.21.

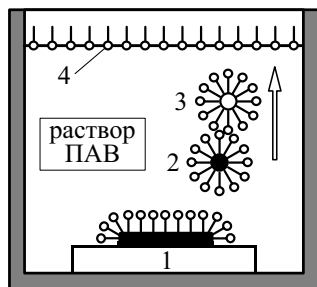


Рис. 5.21. Схематическая иллюстрация действия молекул ПАВ в моющей жидкости, применяемой для очистки поверхности

¹⁾ Термины *гидро-*, *-фильность*, *-фобность* (имея корнями греческие слова «вода», «любовь», «страх») характеризуют способность частиц к межмолекулярному взаимодействию с водой. «Нелюбовь к воде» (*гидрофобность*) является синонимом «любви к жиру», что с греческого переводится как *липофильность*. Молекулы, обладающие одновременно обоими свойствами (то есть химическим сродством и к воде, и к жирам) принято называть *амфифильными* или *дифильными* (от греч. *амфи-* и лат. *ди-*, означающих «двойной».)

Взаимодействие ПАВ с адсорбентом осуществляется через посредство сил Ван-дер-Ваальса и водородной связи. Полярные молекулы воды оказывают тяговое действие на полярные группы молекул ПАВ, которое через радикалы передается поверхностным загрязнениям, покрывающим очищаемую деталь 1. Это способствует диспергированию (размельчению) и отрыву частиц загрязнения от поверхности. Оторвавшаяся частица 2 окружается молекулами ПАВ, которые создают оболочку в форме мицеллы, что препятствует обратному осаждению частицы.

Поверхностно-активные вещества добавляются также в химически активные растворители. Возникающие в процессе реакции газовые пузырьки 3 обволакиваются молекулами ПАВ, ориентированными полярной группой наружу. Эти полярные группы взаимодействуют с такими же группами молекул ПАВ, адсорбированных на частицах загрязнения 2. В результате этого всплывающие газовые пузырьки транспортируют за собой частицы загрязнения в направлении, указанном стрелкой.

Для интенсификации процессов очистки и промывки, в том числе в дистиллированной и деионизованной воде, нередко используют *ультразвук*. В сильном ультразвуковом поле возникает так называемая *кавитация* в форме газовых пузырьков, образующихся в результате разрыва сплошности среды под действием интенсивных ультразвуковых колебаний частиц жидкости. Кавитационные пузырьки, схлопываясь на очищаемой поверхности, создают сильные гидродинамические микроудары с давлением до тысяч атмосфер. Это способствует диспергированию и отрыву поверхностных загрязнений.

Кроме того, в такой сильно полярной жидкости, как вода, стенки кавитационных пузырьков заряжаются до высоких потенциалов. Между пузырьками с разноименными зарядами могут возникать электрические разряды, сопровождающиеся диссоциацией воды и рядом цепных реакций с образованием химически активных радикалов и молекул типа OH , H_2O_2 , HNO_3 (последняя за счет растворенного в воде азота). Следовательно, в ультразвуковом поле даже нейтральная вода приобретает химически активные свойства, способствующие очистке поверхности.

3. Технология получения наноразмерных органических пленок методом Ленгмюра–Блоджетт. Метод послойного нанесения тонких пленок органических материалов на поверхность подложек, носящий имя И. Ленгмюра и К. Блоджетт (ЛБ-метод), был разработан ими в 30-х годах прошлого века. Однако только в 80-х годах возник практический интерес к этому методу как

открывающему возможности нанотехнологической реализации мультислойных молекулярных структур.

Идея ЛБ-метода достаточно проста и заключается в формировании на поверхности воды или водного раствора (называемого в этом случае *субфазой*) мономолекулярного слоя органического вещества с последовательным переносом упорядоченных монослоев на твердую подложку в целях формирования на ней многослойной пленки (ЛБ-пленки). Молекулы таких веществ должны удовлетворять двум противоречивым требованиям: с одной стороны, их взаимодействие между собой должно быть более слабым, чем с молекулами воды, а с другой стороны, их растворимость в воде должна быть по возможности минимальной.

Подобным требованиям отвечают молекулы ПАВ, структура которых схематически изображена на рис. 5.20. Такие молекулы имеют длинный углеводородный хвост, *гидрофобность* которого, во-первых, снижает растворимость молекул в воде и, во-вторых, сообщает им поверхностную активность. Последнее проявляется в вытеснении молекул из объема воды на ее поверхность с ориентацией вниз полярных групп, *гидрофильность* которых обеспечивает сильное взаимодействие с молекулами воды.

Описанная картина с образованием «частокола Ленгмюра» на поверхности схематически изображена на рис. 5.21. Однако такая картина с вертикальным расположением гидрофобных хвостов в воздухе возникает только при достаточно большой поверхностной концентрации молекул ПАВ. Подобная ситуация реализуется, как будет описано ниже, на окончательном этапе формирования монослоя путем его постепенного сжатия.

На начальном этапе ПАВ наносят на поверхность воды в виде раствора его в легколетучем растворителе, не смешивающемся с водой. После испарения растворителя на поверхности остается монослой ПАВ с малой концентрацией молекул, которые ведут себя, как обосновал Ленгмюр, подобно двумерному газу.

Состояние двумерного газа принято характеризовать поверхностным давлением $\pi = \sigma_0 - \sigma$. Оно определено как разность между поверхностным натяжением σ_0 чистой воды (при 25°C $\sigma_0 \approx 0,072$ Н/м) и поверхностным натяжением σ воды, покрытой монослоем ПАВ ($\sigma < \sigma_0$). Поверхностное давление характеризует степень сопротивления монослоя механическому сжатию.

Пусть в монослое площадью A содержится N_s молекул, находящихся достаточно далеко друг от друга. Тогда молекулы можно считать не взаимодействующими между собой с произвольным расположением хвостов в воздухе и совершающими хаотическое тепловое движение, подобно идеальному газу (см. п. 1.12).

На этом основании к разреженному газу молекул ПАВ в монослое с поверхностным давлением π можно применить двумерный аналог уравнения газового состояния (см. уравнение Клапейрона–Менделеева в форме (1.12.1)):

$$\begin{aligned} PV &= Nk_B T & \text{— трехмерный идеальный газ,} \\ \pi A &= N_s k_B T & \text{— двумерный идеальный газ.} \end{aligned} \quad (5.6.6)$$

Состояние системы монослой–вода принято характеризовать *изотермой сжатия*, построенной в координатах $\pi - \omega$, то есть зависимостью поверхностного давления от площади $\omega \equiv A/N_s$, приходящейся на одну молекулу в монослое. Типичный вид изотермы сжатия качественно показан на рис. 5.22.

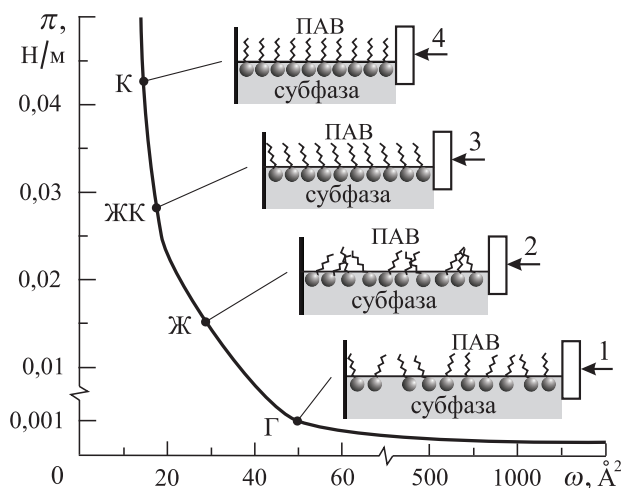


Рис. 5.22. Качественный характер изотермы сжатия двумерного слоя ПАВ, $\pi = f(A_1)$, и картины положения поршня, сжимающего монослой: 1 — газовое состояние (Г), 2 — жидкое состояние (Ж), 3 — жидко-кристаллическое состояние (ЖК), 4 — кристаллическое состояние (К)

Процесс сжатия монослоя технологически выполняется в специальной кювете, называемой ванной Ленгмюра, впервые проводившего эксперименты подобного рода с монослоями ПАВ на поверхности воды. Она состоит из двух секций, разделенных барьером (подвижным поршнем). Одна из секций заполнена деионизованной водой ($\rho \approx 20$ МОм·см и $\text{pH} \approx 5,5$), а другая — водой, покрытой монослоем. Сжатие монослоя осуществляется путем прецизионного смещения поршня с электронным контролем поверхностного давления при помощи так называемых весов Вильгельми с точностью до 10^{-5} Н/м.

На рис. 5.22 справа качественно показаны картины ориентации молекул ПАВ на поверхности субфазы для четырех этапов последовательного сжатия монослоя.

На начальном этапе 1 молекулы находятся в газообразном состоянии (Г), а изотерма сжатия описывается уравнением (5.6.6) в форме $\pi = k_B T / \omega$. При этом поверхностное давление π не превышает 10^{-3} Н/м, а площадь ω , приходящаяся на одну молекулу, составляет сотни и тысячи квадратных ангстрем, что в десятки и сотни раз превышает поперечный размер молекулы.

При сжатии двумерного газа происходит фазовый переход газ $\rightarrow$ жидкость, проявляющийся на изотерме в форме горизонтального участка (не показанного на рис. 5.22), так как плотность молекул в жидком монослое становится намного больше. Последующий рост поверхностного давления в жидкости (Ж) на этапе 2 начинается примерно с площади $50 \text{ \AA}^2$ на молекулу.

Дальнейшее сжатие жидкого монослоя приводит (при $\omega \approx 20 \text{ \AA}^2$ и $\pi \approx 0,02$ Н/м на рис. 5.22) к началу этапа 3 с упорядочением молекул в жидко-кристаллическом состоянии (ЖК). Процесс завершается на этапе 4 плотной упаковкой вертикально ориентированных молекул в форме монослойного кристалла (К).

Именно в такой кристаллической форме монослой, сформированный на поверхности субфазы (воды), переносится на твердую подложку. Перенос осуществляется методом вертикального лифта, то есть вертикальным погружением и извлечением подложки. В процессе переноса площадь, занимаемая монослоем, сокращается, поэтому согласованным движением поршня поддерживают неизменным поверхностное давление ($\pi = \text{const}$).

Существуют три разновидности синтеза мультислойных пленок Ленгмюра–Блоджетт, называемых пленками X-типа, Y-типа и Z-типа со структурой, схематически показанной на рис. 5.23 в. Как видно из рисунка, эти структуры отличаются друг от друга,

- *во-первых*, разной ориентацией по отношению к поверхности подложки амфифильной молекулы в 1-м слое:
 - а) либо на гидрофобной подложке (ГФб-подложке) размещается гидрофобный радикал (X-тип и Y-тип),
 - б) либо на гидрофильной подложке (ГФл-подложке) размещается гидрофильная группа (Z-тип);
- *во-вторых*, взаимной ориентацией по отношению друг к другу амфифильных молекул в соседних слоях:
 - а) либо последовательное чередование слоев (X-тип и Z-тип),
 - б) либо встречное соединение слоев (Y-тип).

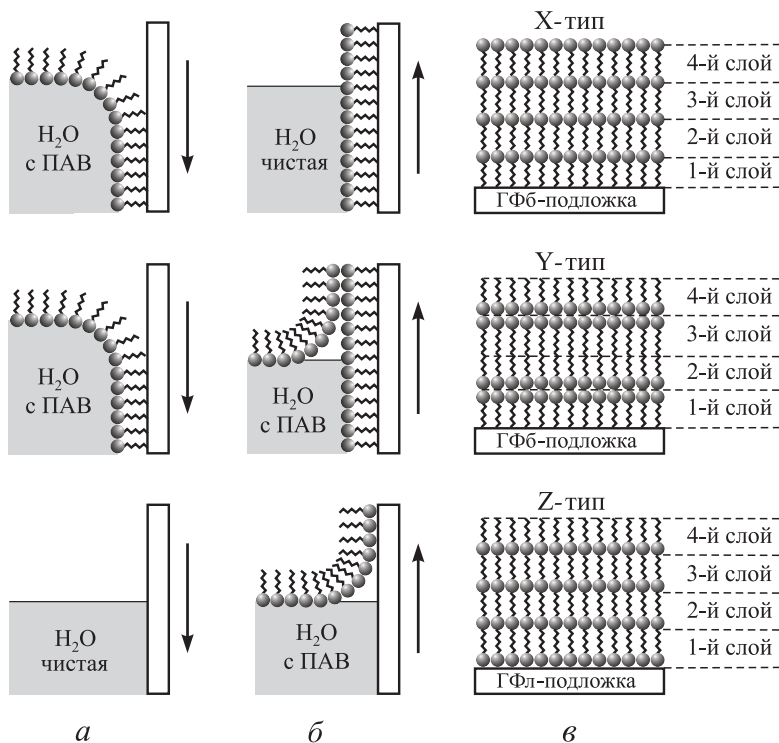


Рис. 5.23. Три варианта погружения (а) и извлечения (б) подложек при послойном формировании мультислойных пленок Ленгмюра–Блоджетт трех типов (в)

Синтез каждого типа структуры определяется, как ясно из вышесказанного, *во-первых*, выбором материала подложки — либо гидрофобная подложка (ГФб-подложка), либо гидрофильная подложка (ГФл-подложка); *во-вторых*, способом погружения и извлечения подложки, что демонстрирует рис. 5.23 а и б.

При синтезе пленки X-типа (в которой гидрофобные хвосты молекул ориентированы к подложке) надо брать ГФб-подложку и погружать ее в секцию кюветы с водой, покрытой сформированным кристаллическим монослоем, как показано на верхней позиции рис. 5.23 а. Из-за двойной гидрофобности контактный угол между подложкой и поверхностью субфазы является тупым. Поэтому поверхность в зоне контакта изгибается вниз, так что изначально обращенные вверх гидрофобные хвосты разворачиваются в сторону подложки, формируя на ней мономолекулярный слой. Извлечение подложки с монослоем проводится из секции с чистой водой, как показано на верхней позиции рис. 5.23 б.

При синтезе пленки *Y*-типа (в которой гидрофобные хвосты молекул ориентированы к подложке) погружение ГФб-подложки выполняется в точности также, как и в случае пленки *X*-типа, что видно из средней позиции на рис. 5.23 *а*. В результате этого на подложке формируется 1-й монослой той же структуры, что и в пленке *X*-типа. Извлечение осуществляется из той же секции с водой, покрытой сформированным кристаллическим монослоем, как показано на средней позиции рис. 5.23 *б*. В этом случае контактируют две гидрофильные поверхности, так что между ними угол острый. В результате этого полярные головки молекул ПАВ разворачиваются в сторону движущейся вверх подложки, формируя на 1-м слое 2-й монослой («головка к головке»), как показано на средней позиции рис. 5.23 *б*.

При синтезе пленки *Z*-типа (в которой гидрофильные головки молекул ориентированы к подложке) нужно теперь брать ГФл-подложку и погружать ее в ячейку с чистой водой, как показано на нижней позиции рис. 5.23 *а*. Перенос сформированного кристаллического монослоя на подложку осуществляется при ее извлечении из секции с субфазой (вода с монослоем на поверхности). В этом случае картина процесса, изображенная на нижней позиции рис. 5.23 *б*, в точности повторяет процесс извлечения пленки *Y*-типа на средней позиции того же рисунка.

Для формирования мультислойных структур каждого из трех типов, изображенных на рис. 5.23 *в*, описанные выше проходы (погружение + извлечение) многократно повторяются. При этом за один проход для структур *X*-типа и *Z*-типа синтезируется один монослой, а для структур *Y*-типа — сразу два монослоя, что, естественно, более технологично.

Следует обратить внимание на то, что для упрощения картины на рис. 5.23 *а* и *б* была отображена только одна левая сторона подложки. В действительности же монослой симметрично переносится и на правую ее сторону, как изображено на рис. 5.24 для двух проходов при синтезе структуры *X*-типа.

Технология Ленгмюра–Блоджетт позволяет синтезировать не только однокомпонентные, но и многокомпонентные мультислойные гетероструктуры, содержащие в своем составе молекулы различных амфифильных веществ. Структура слоев, входящих в состав гетероструктур, определяется соотношением концентрацией и строением различных молекул, в частности соотношением длин их главных осей.

Например, для молекул, имеющих одинаковые длины главных осей и близкое строение длинноцепочечных фрагментов, при определенном соотношении их концентраций можно получать

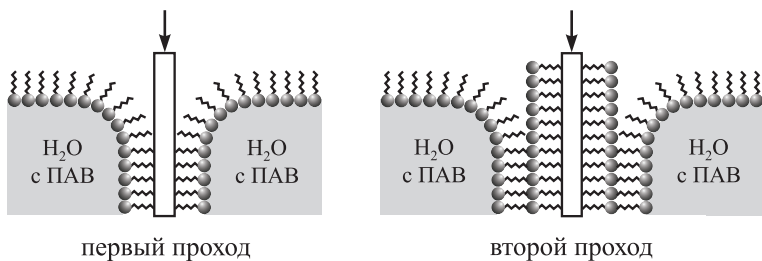


Рис. 5.24. Двухсторонняя картина переноса монослоя на подложку для двух проходов при формировании структуры Х-типа

практически равномерно перемешанные по составу слои. При существенно разных длинах цепочек молекулы каждого сорта будут собираться в отдельные области, образуя нечто подобное доменной структуре.

ЛБ-метод позволяет также путем изменения рН субфазы и растворения в ней солей металлов (с последующей их диссоциацией) синтезировать слои, содержащие атомы, внедренные в органическую матрицу. Растворением солей редкоземельных элементов (например, гадолиния) получают структуры с прослойками из магнитного материала. Есть сообщения об органических ЛБ-пленках с внедренными в них молекулами фуллерена C_{60} .

Структурной организацией и составом монослоев возможно управлять также изменением состава газовой фазы над поверхностью субфазы. Если растворить в жидкости соль какого-либо металла (например, $PbSO_4$), а в газовой фазе создать избыточное содержание нужного газа (например, H_2S), то при его растворении с последующей диссоциацией субфаза обогатится анионами серы. Последние, химически взаимодействуя с катионами свинца, образуют кристаллы PbS . Выбором соответствующих условий можно добиться того, что активные группы молекул монослоя вблизи поверхности раздела создадут решетку с параметрами, близкими к решетке сульфида. В результате кристаллы сульфида свинца будут образовывать эпитаксиальную пленку, кристаллографическая ориентация которой определяется структурой как исходной матрицы, так и самого сульфида.

Таким образом, привлекательность ЛБ-метода определяется не только относительной простотой его технологической реализации, не требующей создания высокого вакуума и специальных температурных условий. Важна также практическая возможность управлять структурой и химическим составом пленок, создавая тем самым мультислойные и многокомпонентные структуры с необходимыми физическими и химическими свойствами.

Последнее обстоятельство существенно расширяет области практического применения ЛБ-пленок, среди которых можно отметить резисты для сверхразрешающей литографии, длинноволновые рентгеновские дифракционные решетки и просветляющие покрытия, интерференционные и поляризационные светофильтры, химические и биологические сенсоры и др.

5.7. Основы золь-гель технологии наноструктурированных материалов

Из вышеизложенного понятно, что технология Ленгмюра–Блоджетт базируется на применении водных растворов органических веществ с амфифильными свойствами молекул, обеспечивающими их поверхностную активность. К такому «органико-растворному» направлению можно отнести и другую технологию, называемую *золь-гель технологией*, в которой используемые органические вещества не обязательно должны быть поверхностно-активными. Ниже кратко изложим основные идеи этой технологии как составляющей конкуренцию ЛБ-технологии.

Оба типа «органико-растворной» технологии (Ленгмюра–Блоджетт и золь-гель) относятся к так называемой «мягкой химии» (англ. *mild chemistry*), что отражает низкотемпературность процессов и их гомогенность на молекулярном уровне. Именно это делает оба метода перспективными для получения наноструктурированных пленок с разнообразными физическими и физико-химическими свойствами.

Золь-гель технология представляет собой метод управляемого синтеза материалов путем нанесения на подложку коллоидного раствора твердых частиц в жидкости (называемого *золем*) с последующим превращением его в твердообразное структурированное состояние (называемое *гелем*). Термины «золь» и «гель» издавна принадлежат коллоидной химии, однако словосочетание «золь-гель» как технологический метод появилось только в конце 80-х годов прошлого столетия.

Прежде чем излагать физико-химическую сущность золь-гель технологии, дадим краткую расшифровку терминов, обозначений и понятий из органической и коллоидной химии, широко используемых не только в золь-гель тематике, но и в других направлениях химической технологии.

1. Терминология и некоторые понятия органической химии. Органическая химия имеет дело с химическими соединениями на основе *углеводородов*, содержащих атомы углерода С

и водорода Н (связанные ковалентными связями с образованием углеводородных цепей — незамкнутых или замкнутых в кольца), а также неуглеродные атомы, такие как О, N, S и др.

Если в составе углеводородного каркаса, кроме атомов С и Н, имеются также другие элементы (металлы или неметаллы), связанные непосредственно с атомами углерода, то такие соединения называются *элементоорганическими*, например металлоорганические соединения, кремнийорганические соединения, фосфорорганические соединения (*фосфины*), мышьякорганические соединения (*арсины*) и другие. В тех же случаях, когда инородные атомы (гетероатомы) находятся в составе замкнутого углеводородного циклического кольца, органические соединения называют *гетероциклическими*. Органические вещества, имеющие одинаковый элементный состав (одну и ту же молекулярную формулу), но различное пространственно-структурное строение и обладающие разными свойствами, называются *изомерами*.

Довольно сложная классификация органических веществ становится более логически понятной с позиции представления молекул в виде комбинации так называемых радикалов. *Радикалом* называется любой химически активный (с ненасыщенными связями) фрагмент той или иной молекулы. В применении к парной связи углерод–водород различают следующие простые радикалы:

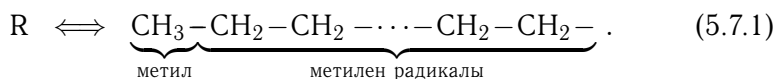
- одновалентный радикал *метил* — CH_3 ,
- двухвалентный радикал *метилен* = CH_2 ,
- трехвалентный радикал *метин* $\equiv \text{CH}$.

Из таких простых радикалов обычно построены углеводородные цепи. Если в составе цепи радикалы подобраны так, что соседние атомы углерода связаны между собой простой (одинарной) связью, а все остальные валентности насыщены атомами водорода, то подобные соединения получили название ***предельных*** (или *насыщенных*) ***углеводородов***. Одинарная ковалентная связь С–С, называемая *σ-связью* (сигма-связью), осуществляется обобществлением пары электронов от соседних атомов путем перекрытия электронных оболочек по прямой, соединяющей эти атомы. Предельные углеводороды реализуются только на основе метила (CH_3) и метилена (CH_2) без участия метина (CH).

Органические соединения, построенные на двойных и тройных связях между атомами углерода, называют ***непредельными*** (или *ненасыщенными*) ***углеводородами***. Одна связь в составе двойной или тройной (с участием 4 или 6 валентных электронов) обязательно является *σ-связью* (с двумя электронами), а оставшиеся одна или две связи — *π-связью* (пи-связью). Это название

носит связь, образованная за счет перекрытия электронных оболочек, направленных перпендикулярно плоскости, в которой расположена паров атомов и ориентирована их σ -связь.

Предельные углеводороды содержат *одновалентный* углеводородный радикал, состоящий из одного концевго метила CH_3 и произвольного числа (начиная с нуля) метиленовых групп CH_2 . Такие углеводородные радикалы, обычно обозначаемые буквой R, имеют общее название «*алкилы*» и следующую структуру:



Одной черточкой принято обозначать валентную связь между углеводородными радикалами, образованную *двумя* обобществленными валентными электронами.

Графическое изображение алкильного радикала (5.7.1) ранее было приведено на рис. 5.20 для частного случая, соответствующего стеариновой кислоте $\text{CH}_3(\text{CH}_2)_{16}\text{COOH}$.

Валентность алкила R, или углеводородного радикала (5.7.1) (в крайнем правом метиле CH_3), может быть насыщена двояко:

- либо атомом водорода H, что создает так называемые *алканы*:

метан	CH_4 ,	или	$\text{CH}_3 - \text{H}$,
этан	C_2H_6 ,	или	$\text{CH}_3 - \text{CH}_2 - \text{H}$,
пропан	C_3H_8 ,	или	$\text{CH}_3 - \text{CH}_2 - \text{CH}_2 - \text{H}$,
бутан	C_4H_{10} ,	или	$\text{CH}_3 - \text{CH}_2 - \text{CH}_2 - \text{CH}_2 - \text{H}$;

- либо концевыми активными *функциональными группами*, среди которых наиболее важные перечислены ниже.

1. **Гидроксильная группа** (гидроксил $-\text{OH}$) в комбинации с алкильным радикалом R создает *спирты* ROH :

метиловый спирт (метанол)	CH_3OH ,
этиловый спирт (этанол)	$\text{C}_2\text{H}_5\text{OH}$,
пропиловый спирт	$\text{C}_3\text{H}_7\text{OH}$,
бутиловый спирт	$\text{C}_4\text{H}_9\text{OH}$.

2. **Карбоксильная группа** (карбоксил $-\text{COOH}$) в комбинации с алкильной группой R создает *карбоновые кислоты* RCOOH :

муравьиная кислота	HCOOH ,
уксусная кислота	CH_3COOH ,
пропионовая кислота	$\text{C}_2\text{H}_5\text{COOH}$,
масляная кислота	$\text{C}_3\text{H}_7\text{COOH}$;

высшие карбоновые (или жирные) кислоты — пальмитиновая $C_{15}H_{31}COOH$ и стеариновая $C_{17}H_{35}COOH$ входят в форме своих натриевых солей $CH_3(CH_2)_{14}COONa$ и $CH_3(CH_2)_{16}COONa$ в состав *мыл*, которые также содержат аналогичные соли непредельных жирных кислот, например олеиновой кислоты $C_{17}H_{33}COOH$.

3. *Альдегидная группа* ($-COH$) создает *альдегиды* $RCOH$:

муравьиный альдегид	$HCOH$,
уксусный альдегид	CH_3COH ,
пропионовый альдегид	C_2H_5COH ,
масляный альдегид	C_3H_7COH .

4. *Аминогруппы* ($-NH_2$ и $-NH-$ как радикалы аммиака NH_3) в комбинации с алкилами создают *амины* RNH_2 и $RNHR'$:

метиламин	CH_3NH_2 ,
этиламин	$C_2H_5NH_2$,
метилэтиламин	C_3H_8NH ,
диэтиламин	$(C_2H_5)_2NH$.

5. *Аминогруппа* ($-NH_2$) создает *амиды кислот* $RCONH_2$ путем замены гидроксила OH в карбоксильной группе $COOH$:

амид муравьиной кислоты	$HCONH_2$,
амид уксусной кислоты	CH_3CONH_2 ,
амид пропионовой кислоты	$C_2H_5CONH_2$,
амид масляной кислоты	$C_3H_7CONH_2$.

6. *Карбонильная группа* ($-CO-$) создает *кетоны* $RCOR'$ путем связи с двумя углеводородными радикалами R и R' :

диметилкетон (ацетон)	CH_3COCH_3 ,
метилэтилкетон	$CH_3COC_2H_5$.

7. *Кислородный атом* ($-O-$) в комбинации с двумя алкилами создает *простые эфиры* (ROR'):

диэтиловый эфир	$C_2H_5OC_2H_5$,
метилэтиловый эфир	$CH_3OC_2H_5$.

8. *Алкильный радикал* ($-R'$) создает *сложные эфиры* $RCOOR'$ карбоновых кислот при замене атома H в карбоксиле $COOH$:

муравьиноэтиловый эфир	$HCOOC_2H_5$,
уксусноэтиловый эфир	$CH_3COOC_2H_5$.

9. **Нитрогруппа** ($-\text{NO}_2$) в комбинации с алкильным радикалом создает **углеводородные нитросоединения** (RNO_2):

нитрометан CH_3NO_2 ,

нитроэтан $\text{C}_2\text{H}_5\text{NO}_2$.

Металлоорганические соединения содержат атомы металла вместо одной из вышеперечисленных функциональных групп:

метилнатрий CH_3Na ,

диметилцинк $(\text{CH}_3)_2\text{Zn}$,

триметилгаллий $(\text{CH}_3)_3\text{Ga}$,

тетраэтилсвинец $(\text{C}_2\text{H}_5)_4\text{Pb}$.

Кремнийорганические соединения содержат атомы кремния вместо части атомов углерода в углеводородных соединениях:

монометилсилан CH_3SiH_3 ,

тетраметилсилан $(\text{CH}_3)_4\text{Si}$,

триэтилхлорсилан $(\text{C}_2\text{H}_5)_3\text{SiCl}$,

тетраэтоксисилан $(\text{C}_2\text{H}_5\text{O})_4\text{Si}$.

Силанами называются кремневодороды (гидриды кремния), прямые аналоги углеводородов, с ковалентными связями Si-Si , Si-H и общей формулой $\text{Si}_n\text{H}_{2n+2}$. Простейшие силаны: моносилан SiH_4 (аналог метана), дисилан Si_2H_6 (аналог этана) и т. д.

Наибольший интерес для золь-гель технологии представляет **тетраэтоксисилан** (ТЭОС). Он относится к классу алкоксисоединений, или алкоксидов металлов типа M(OR)_m , где M — металл (замененный на Si), а OR — алкокси-группа (например, метокси-группа OCH_3 , этокси-группа OC_2H_5 и т. п.).

Алкокси-группа (являющаяся основой сложных эфиров, получаемых из карбоновых кислот заменой атома H в карбоксиле COOH алкильным радикалом R), называется также **спиртовым гидроксидом** OR (или **алкоксидом**, т. е. остатком спирта ROH в форме RO- , потерявшей атом водорода). Название дано по аналогии с водным гидроксидом OH , в котором водород H заменен одновалентным алкидом R в форме (5.7.1).

Для алкоксисоединений кремния встречаются две равносильные формы их записи: Si(OR)_4 или $(\text{RO})_4\text{Si}$, что в применении к ТЭОС принимает вид $\text{Si(OC}_2\text{H}_5)_4$ или $(\text{C}_2\text{H}_5\text{O})_4\text{Si}$. Следует отметить, что в литературе часто ТЭОС называют этиловым эфиром ортокремниевой кислоты ($\text{H}_4\text{SiO}_4 \equiv \text{Si(OH)}_4$).

Непределенные углеводороды содержат углеводородные радикалы, которые, в отличие от алкила R в форме (5.7.1), состоят

не только из одновалентного *метила* CH_3 и двухвалентного *метилена* CH_2 , но и трехвалентного *метина* CH . Такое отличие сообщает радикалам двойную и тройную валентности, что способствует образованию разветвленных углеводородных цепей.

Непредельные углеводороды, как и предельные, образованы комбинацией двухвалентного или трехвалентного радикала со многими из ранее перечисленных функциональных групп.

В частности: а) гидроксил OH образует *непредельные спирты* (виниловый $\text{CH}_2=\text{CH}-\text{OH}$, аллиловый $\text{CH}_2=\text{CH}-\text{CH}_2-\text{OH}$); б) карбоксил COOH образует *непредельные кислоты* (акриловая $\text{CH}_2=\text{CH}-\text{COOH}$, кротоновая $\text{CH}_3-\text{CH}=\text{CH}-\text{COOH}$, олеиновая $\text{CH}_3-(\text{CH}_2)_7-\text{CH}=\text{CH}-(\text{CH}_2)_7-\text{COOH}$) и др.

До сих пор речь шла об открытых (незамкнутых) углеводородных цепях (неразветвленных и разветвленных), которые при- сущи органическим соединениям (предельным и непредельным), имеющим два равносильных названия, отражающих разные стороны структуры — *ациклические соединения* (без замкнутых циклических колец) и *алифатические*, или *жирные соединения* (включающие жиры, так как *aliphatic* означает лат. «жирный»). Следует обратить внимание, что алифатические предельные углеводороды (с алкильными радикалами (5.7.1)), называемые, как уже говорилось, «алканами», имеют также имя *парафины* (от лат. *parum* «мало» и *affinitas* «сродство»), что означает их малую склонность к химическим реакциям с другими веществами.

Циклические соединения содержат замкнутые кольца (циклы), состоящие только из атомов углерода (*карбоциклические*) или же с включением других элементов (N, S, O и др. — *гетероциклические*). Первые принято подразделять на две группы:

а) *алициклические углеводороды*, сходные по многим свойствам с аналогичными алифатическими (отсюда приставка *али*);

б) *ароматические углеводороды*, состоящие из специфического кольца в форме гексагона с шестью атомами углерода в его вершинах, называемого *бензольным кольцом*.

Алициклические углеводороды, содержащие двухвалентные *метилены* CH_2 , замкнутые в углеродное кольцо, имеют общую формулу C_nH_{2n} , что на два атома водорода меньше, чем у соответствующих алканов, и называются *циклопарафинами*, например, циклопропан $\text{C}_3\text{H}_6 \equiv (\text{CH}_2)_3$ и циклобутан $\text{C}_4\text{H}_8 \equiv (\text{CH}_2)_4$.

Ароматические углеводороды (сокращенно *арены*) имеют родоначальником бензол C_6H_6 , в вершинах кольца которого расположены шесть трехвалентных *метинов* CH с длиной связи $\text{C}-\text{C}$, равной $1,4 \text{ \AA}$ (каждая сторона гексагона). Любой атом углерода

связан тремя σ -связями с двумя атомами С и атомом Н в плоскости кольца, а над и под кольцом шесть π -связей образуют единые p -электронные облака (круговое сопряжение молекулы бензола).

Производные бензола образуются путем замещения атомов водорода в бензольном кольце перечисленными выше функциональными группами, называемыми в этом случае *заместителями*. Роль заместителей могут исполнять как отдельные атомы, например галоиды, так и предельные алкильные радикалы R в форме (5.7.1), создающие арены с боковыми цепями. Остатки (радикалы) бензольного кольца (возможно с присоединенными алкилами R) носят общее название *арилы* и обозначаются как Ar, например простейшие арилы: $-\text{C}_6\text{H}_5$ (без одного атома водорода) и $=\text{C}_6\text{H}_4$ (без двух атомов водорода).

Приведем примеры образования производных ароматических углеводов с использованием некоторых заместителей.

1. Алкильные радикалы R вместе с арилами Ar создают соединения ArR_i , например метилбензол (толуол) ArR_0 ($\text{C}_6\text{H}_5\text{CH}_3$) и диметилбензол (ксилол) Ar(R)_2 ($\text{C}_6\text{H}_4(\text{CH}_3)_2$).

2. Гидроксильный радикал OH при замещении H в бензольном кольце создает *фенолы*, например простой фенол $\text{C}_6\text{H}_5\text{OH}$, а при размещении в боковой цепи арила образует *ароматические спирты*, например бензиловый спирт $\text{C}_6\text{H}_5-\text{CH}_2\text{OH}$.

3. Карбоксильный радикал COOH создает *ароматические кислоты* ArCOOH , например бензойную кислоту $\text{C}_6\text{H}_5\text{COOH}$.

4. Альдегидная группа CONH создает *ароматические альдегиды* ArCONH , например бензойный альдегид $\text{C}_6\text{H}_5\text{CONH}$.

5. Аминогруппа NH_2 создает *ароматические амины* ArNH_2 , например аминобензол (анилин) $\text{C}_6\text{H}_5\text{NH}_2$.

6. Карбонильная группа CO создает *ароматические кетоны*: *чисто ароматические* кетоны — при связи двух ароматических радикалов (двух арилов) $\text{Ar}-\text{CO}-\text{Ar}$ и *жирно-ароматические* кетоны — при связи ароматического и алифатического радикалов (арила и алкила) $\text{Ar}-\text{CO}-\text{R}$.

Молекулы органических веществ, подобно неорганическим молекулам, могут взаимодействовать друг с другом. Однотипные простые молекулы (*мономеры*), соединяясь между собой, образуют более сложные молекулы, называемые *олигомерами*¹⁾ (димеры, тримеры и т. д.), и в конечном итоге макромолекулы,

¹⁾ Термины *моно-*, *олиго-*, *поли-*, *-меры* имеют корнями греческие слова «один», «мало», «много», «части» и характеризуют способность углеводородных молекул к образованию повторяющихся составных звеньев.

входящие в состав *полимеров* (высокомолекулярных соединений с числом молекул, достигающим десятков и сотен тысяч).

Полимеры образуют линейные или пространственно разветвленные цепи и при разных типах мономеров в цепи называются *сополимерами* или *привитыми сополимерами* (если в последнем случае основная цепь составлена из одного мономера, а ответвления — из звеньев другого мономера).

Принято различать два типа процессов, положенных в основу синтеза полимеров, — полимеризацию и поликонденсацию.

Полимеризация происходит путем последовательного присоединения мономеров к активному центру на конце растущей цепи. В зависимости от природы активного центра различают *радикальную полимеризацию* (активный центр — свободный радикал) и *ионную полимеризацию* (активный центр — анион или катион). В ряде случаев присоединению мономера предшествует образование координационного комплекса с противоионом, что создает *координационно-ионную полимеризацию*. В подобной ситуации регулирующее действие противоиона способствует образованию полимера с высокой степенью пространственной упорядоченности (стереорегулярные полимеры). По количеству участвующих в полимеризации типов мономера различают *гомополимеризацию* (один мономер) и *сополимеризацию* (минимум два мономера).

Поликонденсация заключается в образовании полимера в результате взаимодействия функциональных групп, входящих в состав мономеров, с выделением побочных низкомолекулярных веществ (воды, спирта, аммиака, галогеноводорода и др.). В реакцию поликонденсации могут вступать как один мономер с двумя различными функциональными группами (*гомополиконденсация*), так и два мономера, несущие различные функциональные группы (*сополиконденсация*, или гетерополиконденсация).

Процессы полимеризации и поликонденсации обычно протекают с участием вспомогательных веществ, называемых *инициаторами* или *катализаторами* в зависимости от того, входят или не входят их частицы в состав образующегося полимера как активные центры растущей полимерной цепи.

По агрегатному состоянию исходных мономеров, взятых для участия в синтезе полимеров, различают газофазные и жидкофазные процессы полимеризации и поликонденсации.

2. Терминология и основные понятия коллоидной химии.

Коллоидная химия изучает физико-химические свойства так называемых *дисперсных систем*, состоящих по меньшей мере из двух фаз, одна из которых находится в сильно диспергированном

(измельченном) состоянии. Совокупность частиц диспергированного вещества образует *дисперсную фазу*, а окружающее их однородное вещество (твердое, жидкое, газообразное) принято называть *дисперсионной средой*.

Степень дисперсности (измельченности) частиц существенным образом влияет на физико-химические свойства дисперсных систем. Это происходит, в первую очередь, потому, что с уменьшением размеров частиц поверхность раздела дисперсной фазы и дисперсионной среды достигает огромной площади, в которой запасается большая избыточная энергия. Например, размельчение исходного куба с длиной ребра 1 см и площадью граней 6 см^2 на маленькие кубики с длиной ребра 1 мкм (числом 10^{12}) и 1 нм (числом 10^{21}) приводит к увеличению общей поверхности, соответственно, в 10^4 и 10^7 раз.

Дисперсные системы принято классифицировать по двум признакам — по степени дисперсности (среднему размеру частиц) и по агрегатному состоянию дисперсной фазы в дисперсионной среде. По первому признаку их подразделяют на грубодисперсные, тонкодисперсные и ультрадисперсные системы.

Грубодисперсные системы с частицами микрометрового размера (1–100 мкм), называемые также *микроретерогенными*, по отношению к жидкой дисперсионной среде подразделяются на:

- *суспензии* с твердофазными дисперсными частицами,
- *эмульсии* с жидкофазными дисперсными частицами,
- *пены* с газообразными дисперсными частицами.

Тонкодисперсные системы с твердыми частицами нанометрового размера (1–100 нм), называемые также *коллоидными растворами*, сохраняют микроретерогенность и образуют **золи** в жидких дисперсионных средах (*гидрозоли* в воде и *алкозоли* в спиртах) или *аэрозоли* в газовых средах (пыль, дым и туман, последний с жидкими микрокаплями). Именно золь служит исходной основой для образования **геля** — твердообразного (студенистого) нетекучего вещества, обладающего упругостью и пластичностью из-за наличия пространственно непрерывной структурной сетки, образованной частицами дисперсной фазы¹⁾.

Ультрадисперсные системы с частицами ангстремного размера, подобно *истинным растворам* (с молекулами, атомами и ионами), являются гомогенными, так как по существу исчезает дисперсная фаза в силу молекулярной степени дисперсности.

¹⁾ Термины *коллоид*, *золь* и *гель* порождены, соответственно, корнями слов «клей» (греч.), «растворять» (лат.) и «застывать» (лат.).

Так как коллоидные системы имеют огромную поверхность раздела фаз, то они обладают большим избытком поверхностной энергии, что делает их термодинамически неустойчивыми. Стремление понизить эту энергию приводит к самопроизвольному протеканию двух процессов, разрушающих дисперсную систему, под названием *коагуляция* (от лат. «слипание») и *седиментация* (от лат. «оседание»).

В процессе коагуляции происходит самопроизвольное укрупнение частиц дисперсной фазы путем их слипания, что дает уменьшение их суммарной поверхности и понижение избыточной свободной энергии системы в целом.

В процессе седиментации под действием силы тяжести происходит выделение из раствора дисперсных частиц в результате их оседания на дне или всплывания на поверхность в зависимости от соотношения массовых плотностей частиц (ρ) и дисперсионной среды (ρ_0). Известно, что вес шарика радиусом a , находящегося внутри жидкости, с учетом выталкивающей силы с ее стороны (называемой подъемной силой Архимеда) равен

$$P = \frac{4\pi a^3}{3} g (\rho - \rho_0), \quad (5.7.2)$$

где g — ускорение силы тяжести. Отсюда при $\rho > \rho_0$ получаем $P > 0$, т. е. шарик опускается на дно, а при обратном неравенстве имеем $P < 0$, т. е. шарик всплывает вверх. При движении со скоростью v шарик в среде с динамической вязкостью η испытывает силу трения $F_{\text{тр}}$, даваемую формулой Стокса:

$$F_{\text{тр}} = 6\pi\eta av. \quad (5.7.3)$$

При установившемся движении $P = F_{\text{тр}}$, что дает скорость седиментации:

$$v_c = \frac{2g}{9} \frac{a^2}{\eta} (\rho - \rho_0). \quad (5.7.4)$$

Отсюда видно, что коагуляция повышает скорость седиментации дисперсных частиц из-за увеличения их размеров ($v_c \propto a^2$).

Процессу седиментации противодействует так называемое броуновское движение частиц, происходящее под действием ударов молекул дисперсионной среды, совершающих тепловое хаотическое движение. Как известно, средний квадрат смещения частицы, $\overline{\Delta x^2}$, за время Δt в результате броуновского движения равен $2D\Delta t$, где коэффициент диффузии определен для вязкой среды в виде

$$D = k_B T \frac{v}{F_{\text{тр}}} = \frac{k_B T}{6\pi\eta a}. \quad (5.7.5)$$

Отношение скорости к силе, $v/F_{\text{тр}} \equiv \mu$, называют подвижностью частиц радиусом a в среде с вязкостью η , тогда $\mu = 1/6\pi\eta a$. Именно подвижность определяет скорость седиментации частиц в поле силы тяжести, $v_c = \mu P$ (см. формулы (5.7.2) и (5.7.4)).

Из сравнения формул (5.7.4) и (5.7.5) следует, что с уменьшением размеров частиц скорость их седиментации снижается ($v_c \propto a^2$), а коэффициент диффузии растет ($D \propto a^{-1}$). Поэтому, в отличие от суспензий, коллоидные растворы с наноразмерными частицами (золи) являются *седиментационно устойчивыми*, так как в них хаотическое броуновское движение препятствует седиментации частиц под действием силы тяжести.

Интенсивное диффузионное движение частиц малых размеров препятствует процессу коагуляции, так как уменьшает их вероятность соединиться за счет сил Ван-дер-Ваальса в агрегаты больших размеров, что повышает *агрегативную устойчивость* зольей (способность противостоять укрупнению частиц). Процесс, обратный коагуляции, то есть разъединение слипшихся частиц на агрегаты малых размеров, называется *пептизацией* (корень слова — от фермента «пепсин», ответственного за переваривание белков в желудочно-кишечном тракте). Повышению пептизации может способствовать введение в дисперсионную среду поверхностно-активных веществ (ПАВ) и специальных электролитов.

По современным представлениям строение коллоидных систем описывается в рамках *мицеллярной модели*. Согласно этой модели, всякий золь состоит из мицелл (дисперсных частиц) и интермицеллярной жидкости (дисперсионной среды). Ядро мицеллы образовано нерастворимыми в дисперсионной среде веществами и окружено адсорбционным слоем ионов того или иного знака, что в совокупности создает заряженную *гранулу*. Именно гранулы движутся под действием электрического поля в процессе массопереноса и осаждения материала мицелл на электродах, называемом *электрофорезом* (термин *форез* от греч. «перенос»).

Электронеutrальность мицеллы в целом обеспечивается наружными ионами противоположного знака (противоионами), создающими подвижный диффузный слой (слой Гуи), окружающий гранулу и нейтрализующий заряд адсорбированных ионов.

Таким образом, твердое ядро мицеллы окружено двойным заряженным слоем, состоящим из неподвижного адсорбционного и подвижного диффузного слоев. По границе между этими слоями, называемой поверхностью скольжения, происходит разрыв двойного слоя при броуновском движении частицы. Поверхность скольжения имеет по отношению к объему раствора потенциал, называемый *электрокинетическим* или ζ -потенциалом.

Дзета-потенциал однозначно определяет скорость движения заряженной гранулы при электрофорезе в электрическом поле с напряженностью E . Как известно, одиночная сфера радиусом a , имеющая потенциал поверхности $\varphi_s = \zeta$, несет заряд $q = C\varphi_s = C\zeta$, где $C = 4\pi\epsilon a$ — емкость сферы в среде с диэлектрической проницаемостью ϵ . На этот заряд со стороны поля действует электрическая сила $F_{\text{эл}} = qE = 4\pi a\zeta\epsilon E$. Из равенства этой силы силе вязкостного трения (5.7.3) в стационарных условиях ($F_{\text{эл}} = F_{\text{тр}}$) следует средняя скорость электрофоретического движения частиц:

$$v_{\text{эл}} = \mu F_{\text{эл}} = \frac{2}{3} \frac{\zeta}{\eta} \epsilon E. \quad (5.7.6)$$

Здесь использована ранее введенная подвижность частиц радиусом a в среде с вязкостью η : $\mu = 1/6\pi\eta a$.

Электрофорез широко применяется для разделения взвешенных смесей на фракции и при нанесении поверхностных покрытий, например оксидной массы на катод (см. приложение А.1). При этом массовая скорость осаждения частиц ($\text{кг}/(\text{м}^2 \cdot \text{с})$), имеющих среднюю массу M и концентрацию N в растворе, определяется величиной $MNv_{\text{эл}}$, т. е. согласно (5.7.6), она растет с увеличением ζ -потенциала и понижением вязкости раствора η .

Величина ζ -потенциала (обычно равная 30–70 мВ) влияет также на агрегативную устойчивость золей. Последняя определяется равновесием между ван-дер-ваальсовыми силами притяжения частиц и их отталкиванием из-за одноименного электрического заряда $q = C\zeta$ у всех частиц данного золя. Поэтому снижение ζ -потенциала способствует коагуляции, а повышение — пептизации золей. Коагуляцию вызывает, как правило, добавление ионов, имеющих заряд, противоположный по знаку зарядам гранул золя. Добавление молекул ПАВ, обволакивающих гранулы, способствует экранировке частиц друг от друга и повышает пептизацию золей.

Среди коллоидных растворов особое место занимают растворы полимеров (высокомолекулярных соединений). Они проявляют качества, свойственные как коллоидным золям, так и истинным растворам. Подобно золям, полимерные частицы обладают достаточно малым коэффициентом диффузии (из-за большого эффективного размера a в формуле (5.7.5)), способны коагулировать под действием внешних факторов и неспособны проходить через полупроницаемые мембраны.

Вместе с тем, растворы полимеров с макромолекулами, способными диссоциировать с образованием макроионов, приближаются по свойствам к гомогенным истинным растворам. Так,

белковые молекулы, имея в своем составе *основные* (NH_2) и *кислотные* (COOH) функциональные группы, проявляют амфотерные свойства. В зависимости от pH среды они либо диссоциируют по кислотному типу, отщепляя H^+ с образованием *макроанионов* (при $\text{pH} > 7$), либо, как щелочи, посылают в раствор ионы OH^- и образуют *макрोकатионы* (при $\text{pH} < 7$). Подобно истинным растворам, растворы полимеров долго сохраняют агрегативную устойчивость, т. е. находятся в устойчивом термодинамическом равновесии.

Особенностью растворения полимеров, отличающей их от низкомолекулярных веществ, является то, что процесс начинается с самопроизвольного набухания полимерных цепей, поглощающих малые молекулы жидкого растворителя с увеличением объема и массы полимера. По мере насыщения полимер переходит в раствор, образуя однофазную гомогенную систему.

Таким образом, растворы полимеров занимают промежуточное место между истинными (коллоидными) золями и истинными растворами низкомолекулярных неорганических веществ. Поэтому в современной литературе принято различать *коллоидные золи* (с дисперсной фазой из твердых частиц, образующих ядро мицеллы) и *полимерные золи* (с дисперсной фазой, сформированной из разветвленных макромолекул).

3. Сущность и реализация золь-гель технологии. Как уже отмечалось, в основе любой золь-гель технологии лежит физико-химический процесс превращения бесструктурного жидкообразного текучего золя в устойчивую структурированную систему в форме твердообразного нетекучего геля.

Процесс гелеобразования в коллоидных и полимерных золях протекает по разным механизмам.

В коллоидных золях частицы дисперсной фазы (мицеллы) при коагуляционных контактах за счет сил Ван-дер-Ваальса образуют агрегаты, состоящие из большого числа частиц, из которых впоследствии создается остов неорганического полимера.

В полимерных золях гель образуется в результате полимеризации и поликонденсации макромолекул (мономеров и олигомеров), находящихся в растворе, с участием ковалентных связей между ними, что создает в итоге единую полимерную макроцепь.

Нередко гели, получаемые из водных растворов солей, а затем гидрозолей неорганических веществ, называют *физическими гелями*, чтобы отличить их от *химических гелей*, синтезируемых из полимерных золь-гелей путем реакций гидролиза и поликонденсации алкоксидов металлов и неметаллов.

В состоянии золя коллоидный раствор имеет одну единственную непрерывную (сплошную) среду — жидкий растворитель, в котором хаотически распределены частицы дисперсной фазы, практически не взаимодействующие между собой.

В состоянии геля сохраняется непрерывность жидкой дисперсионной среды, но внутри нее вместо совокупности одиночно разбросанных дисперсных частиц возникает новая непрерывная среда в форме единого структурного каркаса. Этот твердый каркас (будучи стохастическим или фрактально-организованным) составлен из взаимодействующих (сцепленных друг с другом) частиц золя и образует огромный кластер, пронизывающий (без разрыва) весь исходный объем. Иными словами, на начальном этапе гелеобразования возникает совокупность двух непрерывных взаимопроникающих фаз (жидкообразной и твердообразной), которую принято называть *мокрым гелем*.

В состоянии *мокрого геля* золь полностью теряет свою изначальную текучесть и превращается в студнеобразное вещество, принимающее форму исходного сосуда. Однако на этом формирование геля не заканчивается: продолжается присоединение оставшихся в жидкой фазе мелких твердотельных кластеров к единой кластерной сетке мокрого геля. В полимерных золях этот процесс может сопровождаться реакциями полимеризации и поликонденсации мономеров и олигомеров.

Описанный процесс, называемый *старением геля*, сопровождается усадкой (уменьшением объема) и деформацией структурной сетки с удалением из нее жидкости при искусственной сушке. Полностью высушенный гель называется *ксерогелем* (*ксеро* от греч. «сухой»), который уменьшается в объеме до 5–10 раз по сравнению с мокрым гелем. В режиме суперкритической сушки (при высоком давлении и температуре, более низкой, чем при сушке в атмосферных условиях) можно сохранить структуру мокрого геля с заменой жидкости в порах воздухом, что создает так называемый *аэрогель* с объемом пор до 98%.

Заключительный этап формирования гелевой структуры (порошкообразной, стеклообразной, керамической или пленочной) завершается *высокотемпературным отжигом* (при 500–800°C): окончательно удаляются промежуточные продукты физико-химических превращений и стабилизируются межмолекулярные связи, в том числе с участием процессов спекания и кристаллизации.

Вышеописанная последовательность процессов, происходящих в золь-гель системах в применении к технологии создания пленочных структур на основе химического полимерного геля, схематически изображена на рис. 5.25.

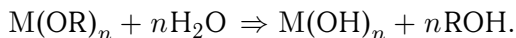
При этом исходными веществами являются алкоксиды металлов М (Al, Ti, Zr, Sn и др.) или неметаллов (типа Si), т. е. их химические соединения с алкоксилем OR (спиртовым гидроксилом, остатком спирта ROH), имеющие формулу $M(OR)_n$, где R — любая алкильная группа типа $R_0 = CH_3$, $R_1 = C_2H_5$, $R_2 = C_3H_7$, $R_3 = C_4H_9$ и т. д.

Подобные вещества называются *прекурсорами* (от англ. *precursor* «предшественник»), то есть являются золеобразующими (способствуют образованию золя). Это, как правило, жидкости, хорошо растворимые в спиртах, что создает исходный однородный раствор, содержащий также и другие необходимые вещества (органические и неорганические соли, связующие добавки и т. п.).

В дальнейшем процесс развивается по нижеследующей схеме:

раствор $\Rightarrow$ золь $\Rightarrow$ мокрый гель $\Rightarrow$ ксерогель.

Добавление воды в приготовленный спиртовой раствор алкоксида приводит к его гидролизу путем замещения алкоксидной группы OR на гидроксильную группу OH по реакции



Таким образом, гидролиз жидкого алкоксида $M(OR)_n$ приводит к образованию мономеров гидроксида $M(OH)_n$, которые выступают в роли активных центров в последующих реакциях поликонденсации. Обычно гидролиз и поликонденсация алкоксидов протекают одновременно, что приводит к образованию олигомеров (димеров, тримеров и т. д.). В результате формируется седиментационно устойчивая дисперсная система (золь) с очень мелкими частицами размером 2–4 нм, из которых впоследствии будет строиться структурная сетка мокрого геля на подложке. Сформированный таким образом золь центрифугированием или другим способом наносится на подложку.

На поверхности подложки по мере испарения растворителя возрастает концентрация дисперсной фазы. Это приводит к началу процесса структурирования, в котором основную роль играет реакция поликонденсации гидроксида металла,



Рис. 5.25. Последовательность операций при создании пленочных структур методами золь-гель технологии

с образованием металл-кислородного каркаса типа $\text{MO}_{n/2}$ в виде трехмерной разветвленной сетки мокрого геля.

В действительности процесс является многостадийным и реализуется через большое число промежуточных реакций. Степень полноты их протекания зависит от многих технологических факторов — концентрации исходного алкоксида металла, воды, растворителя, pH раствора, температуры и т. д.

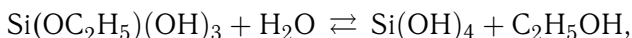
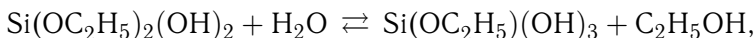
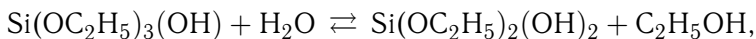
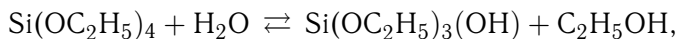
По мере удаления жидкого растворителя в процессе высушивания мокрого геля он превращается в ксерогель — твердое тонкопористое вещество с конденсационно-кристаллизационной основой. Последующим отжигом ксерогеля окончательно формируется пленочная оксидная структура на поверхности подложки.

В качестве примера рассмотрим процесс формирования геля диоксида кремния из тетраэтоксисилана $\text{Si}(\text{OC}_2\text{H}_5)_4$ (ТЭОС).

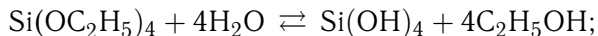
На первом этапе, смешивая ТЭОС с этиловым спиртом (исходным растворителем) и водой, получают золь диоксида кремния SiO_2 . После фильтрации золь наносят на подложку и выдерживают при температуре в диапазоне $100\text{--}200^\circ\text{C}$.

Под воздействием температуры в золе, нанесенном на поверхность подложки, протекают следующие химические реакции:

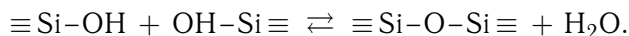
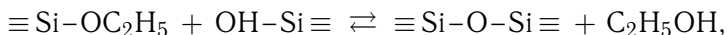
- Постадийные реакции гидролиза тетраэтоксисилана:



образующие суммарную реакцию гидролиза ТЭОС в форме



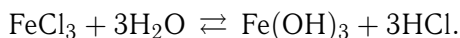
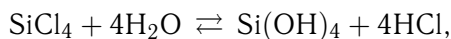
- Реакции поликонденсации:



Эти реакции приводят к образованию мокрого геля, который после испарения растворителя в процессе сушки создает ксерогель в виде пористой ультратонкой сетки диоксида кремния, обычно называемой кремний-кислородным каркасом, с размерами пор в единицы нанометров. В процессе последующего отжига при температуре в диапазоне $500\text{--}800^\circ\text{C}$ происходит заметное

уплотнение ксерогеля с уменьшением его пористости до 10%, что окончательно формирует сплошную пленку SiO_2 на подложке.

Часто при получении золь-гелей труднорастворимых веществ используют физические методы, основанные на конденсации или изменении растворимости с температурой. В качестве исходных веществ используются хлориды:



Осаждение из коллоидных растворов позволяет формировать наночастицы смешанного состава, т. е. синтезировать нанокристаллические гетероструктуры. В этом случае ядро и оболочку смешанной наночастицы создают из полупроводниковых веществ с разным типом проводимости. Образование гетероструктур, например CdSe/ZnS , ZnS/CdS , HgS/CdS , происходит в результате контролируемого осаждения молекул полупроводника одного типа на предварительно синтезированные наночастицы из полупроводника другого типа.

В настоящее время на основе золь-гель технологии неорганических материалов получены не только оксидные структуры, исключительно важные для химического катализа, газочувствительной сенсорики и волоконной оптики, но и новые виды тонкой керамики и разнообразные нанокомпозиты, в том числе органо-неорганические, полимерные и использующие добавки в виде фуллеренов и углеродных нанотрубок.

5.8. Термодинамика искривленных поверхностей

Плоские поверхности межфазных границ, термодинамика которых была рассмотрена в п. 5.5, представляют собой идеализированный случай, в чистом виде достаточно редко встречающийся на практике. Стремление минимизировать поверхностную энергию границ раздела приводит к искривлению поверхности и появлению ряда эффектов, порожденных такой кривизной. Особенно ярко это проявляется в ультрадисперсных и высокопористых структурах. Здесь именно кривизна поверхности является тем фактором, который определяет многие важнейшие характеристики термодинамической системы (капиллярное давление, давление насыщенных паров, температура плавления и др.).

1. Капиллярное давление (формула Лапласа). Общие соотношения, описывающие влияние кривизны поверхности на условия равновесия границы раздела фаз жидкость – газ впервые

были получены Гиббсом. Возможность их применения к описанию разнообразных систем с границами конденсированная фаза–парогазовая фаза позднее была обсуждена многими авторами.

Из сравнения формул (5.5.9) и (5.5.10), выражающих дифференциал свободной энергии для объемной фазы и для межфазного слоя, видно, что отличие между ними заключается только в форме записи второго слагаемого. Это слагаемое выражает механическую работу, совершаемую системой при изменении ее размеров (объема или поверхности).

На этом основании следует ожидать изменения условия механического равновесия, полученного для объемных фаз с плоскими границами в виде равенства давлений (см. формулу (1.14.2)):

$$P^I = P^{II} = \dots \equiv P = \text{const.}$$

При этом условие теплового равновесия (см. формулу (1.14.1)),

$$T^I = T^{II} = \dots \equiv T = \text{const.},$$

и условие материального равновесия (см. формулы (1.14.3)),

$$\mu_i^I = \mu_i^{II} = \dots \equiv \mu_i = \text{const} \quad (i = 1, 2, \dots, K),$$

должны оставаться неизменными.

Рассмотрим механическое равновесие в системе, состоящей из двух объемных фаз I и II, разделенных переходным слоем (рис. 5.16). Полная свободная энергия такой системы равна

$$F = F^I + F^{II} + F_s A. \quad (5.8.1)$$

Используя формулы (5.5.9) и (5.5.10) для изотермических условий ($T = \text{const}$), запишем дифференциал выражения (5.8.1):

$$dF = (-P^I dV^I - P^{II} dV^{II} + \sigma dA) + \sum_i [\mu_i^I dn_i^I + \mu_i^{II} dn_i^{II} + \mu_i d(\Gamma_i A)]. \quad (5.8.2)$$

С учетом того, что: а) в условиях фазового равновесия общая энергия системы не меняется ($dF = 0$); б) суммарное число частиц каждого сорта остается постоянным ($n_i^I + n_i^{II} + \Gamma_i A = \text{const}$); в) выполняется условие материального равновесия $\mu_i^I = \mu_i^{II} = \mu_i$, из формулы (5.8.2) получаем $P^I dV^I + P^{II} dV^{II} = \sigma dA$.

Принимая гипотезу Гиббса о нулевой толщине переходного слоя ($V = 0$), из равенства (5.5.2) получаем $dV^I = -dV^{II}$ (т. е. увеличение объема одной фазы происходит за счет уменьшения объема другой). Поэтому

$$P^I - P^{II} = \sigma \frac{dA}{dV^I}. \quad (5.8.3)$$

Отношение dA/dV^I следует понимать как локальную величину в точке поверхности, в окрестности которой определены элемент площади поверхности dA и элемент объема dV^I . Из дифференциальной геометрии известно, что такая величина определяет *локальную кривизну поверхности* в произвольной точке:

$$H \equiv \frac{1}{2} \frac{dA}{dV^I} = \frac{1}{2} \left(\frac{1}{r_1} + \frac{1}{r_2} \right), \quad (5.8.4)$$

где r_1 и r_2 — *главные радиусы кривизны* поверхности¹⁾.

Обозначая перепад давлений $\Delta P_L \equiv P^I - P^{II}$, приводим выражение (5.8.3) с учетом кривизны поверхности (5.8.4) к форме

$$\Delta P_L = 2\sigma H \equiv \sigma \left(\frac{1}{r_1} + \frac{1}{r_2} \right), \quad (5.8.5)$$

известной как **формула Лапласа**²⁾ (реже в литературе встречаются также и другие ее названия — *формула Юнга–Лапласа* или *формула Кельвина*, см. сноску на с. 631).

Из общей формулы (5.8.5) следуют три частных случая:

- для *плоской поверхности* ($r_1 = \infty$, $r_2 = \infty$)

$$\Delta P_L = 0; \quad (5.8.6)$$

- для *цилиндрической поверхности* радиусом r ($r_1 = r$, $r_2 = \infty$) (см. рис. 5.26 а)

$$\Delta P_L = \frac{\sigma}{r}; \quad (5.8.7)$$

- для *сферической поверхности* радиусом r ($r_1 = r_2 = r$) (см. рис. 5.26 а)

$$\Delta P_L = \frac{2\sigma}{r}. \quad (5.8.8)$$

Выражение (5.8.8) для сферической границы раздела фаз непосредственно вытекает из исходной формулы (5.8.3), в которой

¹⁾ Для нахождения главных радиусов кривизны r_1 и r_2 в некоторой точке произвольно искривленной поверхности следует пересечь ее плоскостью, содержащей нормаль к поверхности в этой точке. Поворотом плоскости вокруг нормали находим кривую пересечения с поверхностью, имеющую минимальный радиус кривизны r_1 . Тогда другая нормальная плоскость, перпендикулярная первой, дает вторую кривую пересечения с главным радиусом кривизны r_2 .

²⁾ Лаплас (Laplace) Пьер Симон (1749–1827) — французский астроном, математик и физик, известен работами в области небесной механики, теории вероятностей и дифференциальных уравнений математической физики, а также физическими исследованиями в молекулярной физике, акустике и гидродинамике, где им разработана теория капиллярности (формула Лапласа), выведены барометрическая формула с учетом влажности воздуха и формула для скорости распространения звука с поправкой на адиабатичность.

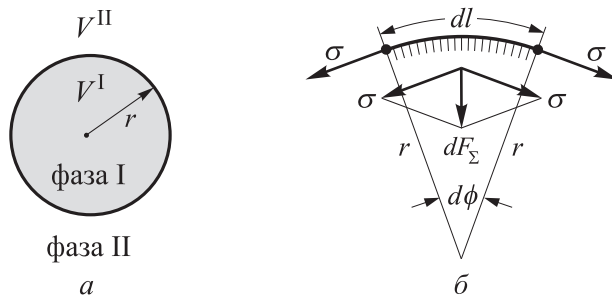


Рис. 5.26. *а* — поперечное сечение цилиндра и шара радиусом r (фаза I всегда внутренняя), *б* — геометрическое обоснование давления Лапласа на цилиндрической поверхности, разделяющей две смежные фазы

объем и площадь поверхности шара следует положить равными $V^I = 4\pi r^3/3$ и $A = 4\pi r^2$. Тогда формула (5.8.3) дает

$$\Delta P_L \equiv P^I - P^{II} = \sigma \frac{dA/dr}{dV^I/dr} = \frac{2\sigma}{r}.$$

Перепад давлений $\Delta P_L \equiv P^I - P^{II}$ между смежными фазами, разделенными искривленной границей, в литературе называют по-разному: либо *капиллярным давлением*, либо *поверхностным давлением Лапласа* (с добавлением имени Лапласа для того, чтобы отличить по названию от *поверхностного давления* $\pi = \sigma_0 - \sigma$, входящего в уравнение газового состояния (5.6.6) для двумерного газа).

Построение, приведенное на рис. 5.26 б, геометрически обосновывает возникновение давления Лапласа на цилиндрической поверхности. Здесь выделен элемент dl единичной длины вдоль оси цилиндра (по нормали к плоскости рисунка). В соответствии с силовой интерпретацией поверхностного натяжения, данной после формулы (5.5.3), две касательные силы σ удерживают в равновесии элемент $dl = r d\phi$. Результат сложения этих сил направлен по радиусу к центру окружности и равен (ср. (5.8.7))

$$dF_\Sigma = 2\sigma \sin(d\phi/2) \approx \sigma d\phi = \frac{\sigma}{r} dl, \quad \text{откуда} \quad \Delta P_L = \frac{\sigma}{r}.$$

Следовательно, капиллярное давление ΔP_L на искривленной поверхности всегда направлено к центру ее кривизны. Согласно вышеприведенным формулам и рис. 5.26 а, с вогнутой стороны границы раздела сред располагается фаза I. Поэтому внутри этой фазы давление P^I всегда превышает наружное давление P^{II} на величину давления Лапласа $\Delta P_L > 0$ (так как $\sigma > 0$), независимо от агрегатного состояния веществ, составляющих фазы I и II.

Дополнительное давление Лапласа (или капиллярное давление) ΔP_L , порожденное кривизной границы раздела фаз, служит причиной появления различных капиллярных явлений. Среди них: вертикальный подъем по стенкам капилляра смачивающей жидкости, ее движение в неоднородном капилляре под действием разности капиллярных давлений, стягивание ими эластичных стенок капилляра (так называемая капиллярная контракция), капиллярная конденсация жидкости из пара в тонких порах и др. Ниже дадим краткое описание вышеуказанных явлений.

Капиллярные явления начнем изучать с самопроизвольного поднятия/опускания жидкости по стенкам цилиндрического однородного (с постоянным сечением) капилляра (рис. 5.27 *a* и *б*).

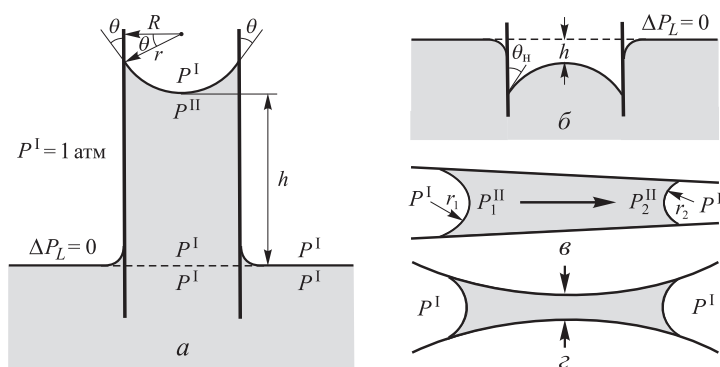


Рис. 5.27. Капиллярные явления: *a* — капиллярное поднятие жидкости, смачивающей стенки капилляра; *б* — капиллярное опускание жидкости, не смачивающей стенки капилляра; *в* — движение жидкости в капилляре (по стрелке) под действием разности капиллярных давлений ($P_1^{II} > P_2^{II}$); *г* — стягивающее действие отрицательного капиллярного давления на эластичные стенки капилляра

Капиллярное поднятие жидкости происходит при ее смачивании стенок капилляра, когда угол смачивания θ острый, так что возникает вогнутый мениск, показанный на рис. 5.27 *a*. Согласно принятым обозначениям, над вогнутым мениском всегда располагается среда с давлением P^I , которое в данном случае является атмосферным давлением ($P^I = 1 \text{ атм}$). Под мениском (внутри жидкости) давление P^{II} меньше на величину капиллярного давления ΔP_L , так что снизу сферической поверхности

$$P^{II} = P^I - \Delta P_L = 1 \text{ атм} - \frac{2\sigma}{r}. \quad (5.8.9)$$

Добавка к атмосферному давлению в виде $-\Delta P_L = -2\sigma/r < 0$ часто трактуется как *отрицательное капиллярное давление*,

которое втягивает жидкость внутрь капилляра, заставляя ее подниматься вверх. С ростом высоты столба жидкости (h) увеличивается его гидростатическое давление $P_r = g\rho h$ на плоскую поверхность в сосуде (вдоль пунктирной линии, продолжающей наружную сплошную линию, где $\Delta P_L = 0$ и $P^{\text{II}} = P^{\text{I}}$). Равновесие наступит, когда гидростатическое давление P_r скомпенсирует отрицательное капиллярное давление $-\Delta P_L$, так что на пунктирной линии

$$P^{\text{II}} + P_r = P^{\text{I}} \equiv 1 \text{ атм.} \quad (5.8.10)$$

Из равенств (5.8.9) и (5.8.10) получаем условие равновесия капиллярного столба жидкости в виде

$$\Delta P_L = P_r, \quad \text{или} \quad \frac{2\sigma}{r} = g\rho h, \quad (5.8.11)$$

где g — ускорение силы тяжести и ρ — плотность жидкости.

Из рис. 5.27 а следует, что угол смачивания θ связывает радиусы сферы и цилиндра соотношением $R = r \cos \theta$. Тогда из (5.8.11) получаем высоту капиллярного подъема жидкости:

$$h = \frac{2\sigma \cos \theta}{g\rho R}. \quad (5.8.12)$$

Как и следовало ожидать, подъем жидкости в капилляре возрастает с ростом поверхностного натяжения σ и с уменьшением как угла смачивания ($\theta \rightarrow 0$), так и радиуса капилляра R .

Капиллярное опускание жидкости имеет место, когда жидкость не смачивает стенки капилляра (при остром угле несмачивания $\theta_n \equiv 180^\circ - \theta < 90^\circ$), что формирует выпуклый мениск, показанный на рис. 5.27 б (ср. рис. 5.18 б). В этом случае ситуация противоположна предыдущей: давление под мениском (внутри жидкости) превышает атмосферное на величину капиллярного давления ΔP_L . В результате этого жидкость вытесняется из капилляра в сосуд, что вызывает ее опускание ниже уровня свободной поверхности (ниже горизонтальной сплошной и пунктирной линии на рис. 5.27 б). Формула (5.8.12) остается справедливой и при несмачивании, когда $\theta > 90^\circ$, что дает $h < 0$ — отрицательный капиллярный подъем.

Капиллярное движение жидкости в неоднородном капилляре (с плавно меняющимся поперечным сечением) демонстрирует рис. 5.27 в для жидкости, смачивающей стенки капилляра. Движение вызвано различием капиллярных давлений ($P_1^{\text{II}} > P_2^{\text{II}}$) на противоположных концах неоднородного капилляра, заполненного жидкостью. Различие давлений вызвано разной кривизной поверхности жидкости на концах капилляра ($r_1 > r_2$).

Действительно, в соответствии с формулой (5.8.9), имеем

$$P_1^{\text{II}} \equiv P^{\text{I}} - \frac{2\sigma}{r_1} > P^{\text{I}} - \frac{2\sigma}{r_2} \equiv P_2^{\text{II}}.$$

Следовательно, поток смачивающей жидкости всегда направлен в сторону мениска с меньшим радиусом кривизны, как показано стрелкой на рисунке. Нетрудно видеть, что для несмачивающей жидкости аналогичный поток направлен в противоположную сторону, но всегда от большего давления к меньшему.

Капиллярная контракция (от англ. *contraction*) является результатом сжимающего действия отрицательного капиллярного давления (см. формулу (5.8.9)) на эластичные стенки капилляра, как показано на рис. 5.27 *г* для смачивающей жидкости. Этот эффект приводит к деформации высокодисперсных и пористых материалов, в частности к усадке ксерогеля при высушивании.

Капиллярная конденсация будет изложена после вывода формулы Гиббса–Томсона. Однако прежде, чем это делать, рассмотрим вопрос, напрямую не связанный с капиллярными явлениями, но проявляющийся зачастую в близких практических ситуациях. Речь пойдет о так называемом осмотическом давлении.

Осмотическое давление в разбавленных растворах возникает при самопроизвольном проникновении молекул чистого жидкого растворителя (воды) в водный раствор через так называемую полупроницаемую перегородку (мембрану), задерживающую растворенные вещества. Такой процесс принято называть *осмосом* (от греч. «толчок, давление»). Роль полупроницаемой мембраны играют как ткани природного происхождения (стенки сосудов, клеток и т. п.), так и искусственно созданные материалы (пленки целлофана, синтетических полимеров и т. п.).

Схематическое изображение устройства для наблюдения осмоса приведено на рис. 5.28. Стрелкой указано направление движения растворителя через полупроницаемую мембрану М из сосуда ① в сосуд ②, заполненный разбавленным раствором на основе этого растворителя. Если бы мембрана была проницаемой для всех компонентов (или бы просто отсутствовала), то растворенные вещества из сосуда ② в результате диффузии равномерно распределились бы по всему суммарному объему.

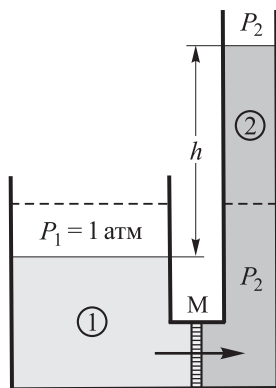


Рис. 5.28. Схема наблюдения осмотического давления при $P_2 = 1$ атм

В основе подобного процесса лежит стремление системы уменьшить свободную энергию Гиббса за счет смешивания компонентов ($\Delta G_{\text{см}} < 0$) путем понижения концентрации (а значит и химического потенциала) растворенных веществ, обеспечивая при этом возрастание энтропии на величину энтропии смешения ($\Delta S_{\text{см}} > 0$, см. пп. 1.12 и 2.4).

Полупроницаемая мембрана препятствует растворенным веществам понижать свою концентрацию путем диффузионного перераспределения по всему объему с целью уменьшения свободной энергии системы. Эта цель (а именно $\Delta G < 0$) теперь достигается другим способом — путем проникновения растворителя из сосуда ① в раствор ②, что также снижает концентрацию (а значит и химический потенциал) растворенных веществ и обеспечивает $\Delta G < 0$ как результат не диффузии, а осмоса.

Молекулы растворенного вещества в сосудах ②, будучи не в состоянии пройти сквозь мембрану, подобно газовым молекулам под поршнем, давят на нее, стремясь сместить ее силами, создающими дополнительное давление π , называемое *осмотическим*.

До начала осмоса уровень жидкостей в обоих сосудах был одинаковым (как показано пунктирными линиями на рисунке), что соответствует постоянному внешнему давлению $P_1 = 1$ атм. По мере проникновения чистого растворителя в раствор уровень жидкости в сосудах ① понижается, а в сосудах ② повышается. Так продолжается до тех пор, пока возникшее гидростатическое давление столба жидкости высотой h не окажется равным осмотическому давлению. Описанная ситуация реализуется при давлении P_2 над раствором в сосудах ②, равном $P_1 = 1$ атм.

Можно не допустить смещения уровней жидкости в сосудах ① и ②, если приложить к раствору внешнее давление $P_2 > P_1 = 1$ атм. Понятно, что то значение P_2 , которое требуется для сохранения уровня жидкостей на исходной пунктирной линии, равняется осмотическому давлению. Более того, если и далее повышать P_2 , то можно получить *обратный осмос*. При этом растворитель переходит из раствора ② в сосуд ①, то есть из более концентрированного раствора в менее концентрированный.

Обратный осмос применяют в различных технологиях очистки воды от примесей, в том числе для опреснения морской воды. Этот процесс требует давления P_2 около 2–17 атм для питьевой (солонатовой) воды и не менее 40–70 атм для морской воды, которая имеет естественное осмотическое давление около 25 атм.

Термодинамика осмоса основана на поведении химического потенциала растворителя, который (точнее его неоднородность) является движущей силой процесса.

Рассмотрим исходный (до начала осмоса) бинарный раствор при температуре T и внешнем давлении $P_1 = 1$ атм. Пусть такой раствор содержит n_1 молей растворителя (с индексом 1) и n молей растворенного вещества (без индекса). Тогда мольная доля растворенного вещества равна $x \equiv n/(n + n_1) \approx n/n_1 \ll 1$ (последнее при сильном разбавлении).

Как известно, в предельно разбавленных растворах растворитель подчиняется модели идеального раствора. Тогда его химический потенциал имеет вид (см. формулу (2.4.29))

$$\begin{aligned}\mu_1(T, P_1, x) &= \mu_1^\circ(T, P_1) + RT \ln(1 - x) \approx \\ &\approx \mu_1^\circ(T, P_1) - xRT.\end{aligned}\quad (5.8.13)$$

Здесь стандартный химический потенциал $\mu_1^\circ(T, P_1)$ соответствует чистому растворителю в сосуде (1), где $x = 0$. Из уравнения (5.8.13) видно, что $\mu_1^\circ(T, P_1) > \mu_1(T, P_1, x)$, и это обеспечивает начало процесса осмоса.

В процессе осмоса число молей растворенного вещества не меняется ($n = \text{const}$), а число молей n_1 растворителя в растворе возрастает, пусть на величину Δn_1 . Тогда мольная доля растворенного вещества уменьшается до значения

$$x' = \frac{n}{n_1 + \Delta n_1} \approx \frac{x}{1 + \Delta n_1/n_1} \approx x \quad \text{при } \Delta n_1 \ll n_1. \quad (5.8.14)$$

Так как давление в растворе при осмосе возрастает на величину осмотического давления π , то химический потенциал растворителя (5.8.13) с учетом (5.8.14) принимает вид

$$\begin{aligned}\mu_1(T, P_1 + \pi, x') &= \mu_1^\circ(T, P_1 + \pi) - x'RT \approx \\ &\approx \mu_1^\circ(T, P_1) + \left(\frac{\partial \mu_1^\circ(T, P)}{\partial P} \right)_{T, P_1} \pi - x'RT \approx \\ &\approx \mu_1^\circ(T, P_1) + \pi v_1 - xRT.\end{aligned}\quad (5.8.15)$$

Здесь введен молярный объем растворителя $v_1 \equiv (\partial \mu_1^\circ / \partial P)_{T, P_1}$, так что исходный объем раствора $V = v_1 n_1$.

Равновесие осмоса наступает при равенстве химических потенциалов чистого растворителя и растворителя в растворе:

$$\mu_1^\circ(T, P_1) = \mu_1(T, P_1 + \pi, x'). \quad (5.8.16)$$

Подставляя (5.8.15) в (5.8.16), получаем выражение для осмотического давления в разбавленном растворе (при $x' \approx x$):

$$\pi = \frac{x}{v_1} RT, \quad \text{или} \quad \pi V = nRT. \quad (5.8.17)$$

Уравнение (5.8.17) называют **законом Вант-Гоффа** (см. сноску на с. 269), который впервые получил его в 1887 году.

Сравнение закона Вант-Гоффа с уравнением Клапейрона–Менделеева (1.12.1), описывающим состояние идеального газа, говорит о их полной тождественности. Следовательно, осмотическое давление, создаваемое растворенным веществом в разбавленном растворе, равно давлению которое оно производило бы в газообразном состоянии в том же объеме и при той же температуре. Подобно газам, один моль любого вещества, растворенный в 22,4 л, создает при 0°C осмотическое давление в 1 атм.

Это вполне объяснимо тем, что при сильном разбавлении частицы растворенного вещества практически не взаимодействуют между собой, подобно молекулам идеального газа. Более того, закон Вант-Гоффа, как и уравнение Клапейрона–Менделеева, является универсальным в применении к достаточно разбавленным растворам, поскольку не зависит от типа растворителя, растворенного вещества и материала полупроницаемой перегородки.

Осмоз играет важнейшую роль в процессах жизнедеятельности животных и растений, поскольку клеточная плазматическая мембрана является полупроницаемой. Осмотическое давление в плазме крови человека поддерживается постоянным на уровне 7,7 атм. Осмос обеспечивает поднятие воды по стеблю растений, а также прочность и упругость клеточных оболочек с осмотическим давлением внутриклеточного сока до 5–20 атм.

2. Давление насыщенного пара частиц малых размеров (формула Гиббса–Томсона). Выясним вопрос о парциальном давлении p_r собственного пара любого вещества над его искривленной поверхностью (с радиусом кривизны r) по сравнению с давлением p_∞ насыщенного пара того же вещества над плоской поверхностью (при $r = \infty$), отмечая индексами « r » и « ∞ » конечный и бесконечный радиусы кривизны. Кроме того, конденсат (жидкий или твердый) будем отмечать верхним индексом « k » («ж» или «т»), а паровое состояние — индексом « p », а именно p_r^k и p_r^p . Следует обратить внимание на то обстоятельство, что речь пойдет не о полном давлении P (как это было выше), а о парциальном давлении компонента, которое ранее обозначали малой буквой p (опуская при этом за ненадобностью нижний индекс, характеризующий компонент).

Для определенности будем рассматривать сферическую границу раздела фаз. Тогда на основании выражения (5.8.8) для давления Лапласа ($\Delta p_L = 2\sigma/r > 0$) связываем давления в конденсате и паре следующим равенством:

$$p_r^k = p_r^\pi \pm \Delta p_L = p_r^\pi \pm \frac{2\sigma}{r}. \quad (5.8.18)$$

Верхний знак в формуле (5.8.18) соответствует сферическому зародышу конденсата в паровой среде, а нижний знак — пузырьку пара в среде конденсата.

Ниже отойдем от традиционного подхода к выводу формулы Гиббса–Томсона (см., например, первое издание данной книги ¹⁾), в основу которого положено линейное приближение при разложении химического потенциала конденсата, $\mu^k(p_r^k, T)$, в ряд Тейлора по приращению $\Delta p_r^k \equiv p_r^k - p_\infty$ (см. там формулу (6.22)).

Основной целью последующего рассмотрения является *вывод дифференциального уравнения*, решение которое должно дать классическую формулу Гиббса–Томсона ²⁾.

Исходным для вывода является условие фазового равновесия на искривленной границе раздела (жидкий или твердый) конденсат–пара, записанное в форме равенства соответствующих химических потенциалов:

$$\mu^k(p_r^k, T) = \mu^\pi(p_r^\pi, T). \quad (5.8.19)$$

Входящие сюда давления (внутри конденсата p_r^k и в паре p_r^π) при искривленной поверхности оба отличаются от давления p_∞ насыщенного пара над плоской границей. Для их нахождения продифференцируем левую и правую части (5.8.19) по r , тогда

$$v^k \frac{\partial p_r^k}{\partial r} = v^\pi \frac{\partial p_r^\pi}{\partial r}. \quad (5.8.20)$$

Здесь введены мольные объемы конденсата и пара, определенные как

$$v^k = \frac{\partial \mu^k}{\partial p_r^k} \quad \text{и} \quad v^\pi = \frac{\partial \mu^\pi}{\partial p_r^\pi}. \quad (5.8.21)$$

Мольный объем для конденсата вычисляется как отношение его молярной массы M к плотности ρ^k , то есть $v^k = M/\rho^k$, а для пара находится из уравнения Клапейрона–Менделеева (1.12.1) в виде $v^\pi = RT/p_r^\pi$.

¹⁾ Барыбин А. А. Электроника и микроэлектроника. Физико-технологические основы. — М.: Физматлит, 2006.

²⁾ Томсон (Thomson) Уильям (1824–1907) — английский физик, получил титул лорда Кельвина (1892) за научные достижения в области электромагнетизма, гидродинамики, теории упругости, вихревых и поверхностных явлений (уравнение Кельвина). Наиболее известен работами по термодинамике, где он ввел понятие диссипации энергии, дал одно из определений второго начала термодинамики, ввел абсолютную шкалу температур (шкалу Кельвина), открыл эффект понижения температуры газа при расширении в пустоту (эффект Джоуля–Томсона) и один из термоэлектрических эффектов (эффект Томсона).

Подставляем $v^\Pi = RT/p_r^\Pi$ в правую часть формулы (5.8.20), тогда

$$v^K \frac{\partial p_r^K}{\partial r} = RT \frac{\partial \ln p_r^\Pi}{\partial r}. \quad (5.8.22)$$

Дифференцированием равенства (5.8.18) по r получаем

$$\frac{\partial p_r^K}{\partial r} = \frac{\partial p_r^\Pi}{\partial r} \mp \frac{2\sigma}{r^2}. \quad (5.8.23)$$

Подстановка (5.8.23) в левую часть (5.8.22) дает равенство

$$(v^\Pi - v^K) \frac{\partial p_r^\Pi}{\partial r} = \mp \frac{2\sigma v^K}{r^2},$$

откуда, с учетом того, что $v^K \ll v^\Pi = RT/p_r^\Pi$, вытекает искомое дифференциальное уравнение для нахождения p_r^Π :

$$\frac{\partial \ln p_r^\Pi}{\partial r} = \mp \frac{v^K}{RT} \frac{2\sigma}{r^2}. \quad (5.8.24)$$

Интегрирование уравнения (5.8.24) путем разделения переменных приводит к следующему результату:

$$\int_{p_\infty}^{p_r^\Pi} d \ln p_r^\Pi = \mp \frac{2\sigma v^K}{RT} \int_{\infty}^r \frac{dr}{r^2}, \quad \text{или} \quad \ln \frac{p_r^\Pi}{p_\infty} = \pm \frac{v^K}{RT} \frac{2\sigma}{r}. \quad (5.8.25)$$

Отсюда получаем классическую **формулу Гиббса–Томсона**:

$$p_r^\Pi = p_\infty \exp\left(\pm \frac{v^K}{RT} \frac{2\sigma}{r}\right), \quad (5.8.26)$$

где $v^K = M/\rho^K$. Двойные знаки в (5.8.26) и в предыдущих формулах соответствуют аналогичным знакам в равенстве (5.8.18).

Следует отметить, что вывод формулы Гиббса–Томсона был сделан вне рамок каких-либо ограничений (типа указанного выше линейного приближения при разложении в ряд Тейлора). Поэтому чисто математически она допускает предельный переход $r \rightarrow 0$.

Формула (5.8.26) характеризует зависимость давления собственного пара p_r^Π над сферической поверхностью конденсата от радиуса сферы r , что показано на рис. 5.29 сплошными линиями. Штриховые линии изображают аналогичные зависимости

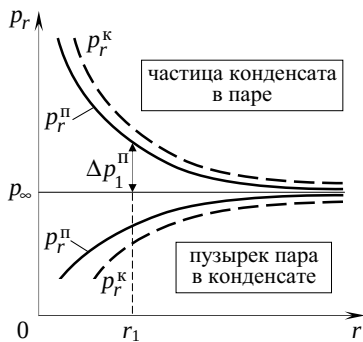


Рис. 5.29. Зависимости давления внутри конденсата p_r^K и давления пара p_r^Π над поверхностью сферической частицы от ее радиуса r

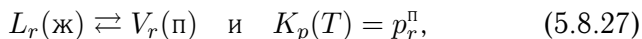
для давления p_r^k внутри конденсированной фазы, равновесной с паром над искривленной поверхностью. Согласно равенству (5.8.18), штриховые кривые смещены относительно сплошных на величину поверхностного давления Лапласа $\Delta p_L = 2\sigma/r$. Расположение этих кривых по отношению друг к другу таково, что внутри сферы давление всегда больше, независимо от того, какая фаза (конденсат или пар) заполняет внутренность сферы. С увеличением размеров частицы все кривые смыкаются и приближаются к давлению p_∞ насыщенного пара над плоской границей.

Кривые выше и ниже горизонтальной линии $p_\infty = \text{const}$ соответствуют верхнему и нижнему знакам в формулах (5.8.18) и (5.8.26), то есть они характеризуют зародыш конденсата в паре и пузырек пара в конденсате. Отсюда видно, что равновесное давление пара над искривленной поверхностью зависит от того, внутри или вне сферы находится паровая фаза.

Проанализируем поведение кривых на рис. 5.29 в применении к соответствующим физическим ситуациям.

1. *Верхние кривые* на рис. 5.29 применимы не только к замкнутой сфере, для которой был выполнен анализ, но и к любым *выпуклым поверхностям* конденсата, над которыми имеется равновесный с ними пар, имеющий давление $p_r^n > p_\infty$.

Повышенное давление пара над мелкими каплями и маленькими кристалликами приводит к тому, что они являются нестабильными образованиями. Для пояснения этого запишем, следуя (2.15.2), обратимый процесс испарение $\rightleftharpoons$ конденсация сферической жидкой частицы радиусом r в виде квазихимической реакции совместно с соответствующим ей законом действующих масс:



где L_r и V_r обозначают жидкую частицу (*Liquid*) и пар (*Vapour*) над нею. Здесь p_r^n означает равновесное давление пара над поверхностью сферического конденсата при температуре $T = \text{const}$. При этой температуре испарение жидкости (например воды) с плоских ее поверхностей формирует постоянное давление $p_\infty(T) = \text{const}$, которое играет роль реального (неравновесного) давления по отношению к сферическому конденсату.

В соответствии с уравнением изотермы Вант-Гоффа (2.13.5) записываем изобарный потенциал реакции (5.8.27):

$$\Delta G_r = RT \left[\ln p_\infty - \ln K_p(T) \right] = RT \ln \frac{p_\infty}{p_r^n}. \quad (5.8.28)$$

Поскольку $p_r^n > p_\infty$, то из (5.8.28) получаем $\Delta G_r < 0$, то есть реакция (5.8.27) смещена в сторону испарения. Следовательно,

в условиях, когда для объектов больших размеров пар при $T = \text{const}$ является насыщенным с давлением p_∞ , мелкие капли все еще продолжают испаряться. По мере испарения капель их радиус r уменьшается, давление p_r^Π растет, а пересыщение (5.8.28) увеличивается по модулю (оставаясь отрицательным), что увеличивает скорость испарения. Полное термодинамическое равновесие в системе наступает после испарения всех капель и их конденсации на плоской поверхности жидкости.

Однако при конденсации из парогазовой фазы *пересыщенный пар* (с давлением $p_r^\Pi > p_\infty$) может сохраняться, если отсутствуют центры конденсации (мелкодисперсные частицы, пылинки, ионы в камере Вильсона и т. п.). Так, для паров воды, имеющей $v^{\text{ж}} = 1,8 \cdot 10^{-5} \text{ м}^3$ и $\sigma = 0,072 \text{ Дж/м}^2$ при $T = 298 \text{ К}$, получаем

$$\ln \frac{p_r^\Pi}{p_\infty} = \frac{v^{\text{ж}}}{RT} \frac{2\sigma}{r} = \frac{1,05}{r [\text{нм}]}.$$

Отсюда $p_r^\Pi/p_\infty = 1,001, 1,011, 1,111$ и $2,86$, соответственно, при радиусах $r = 1 \text{ мкм}, 0,1 \text{ мкм}$ (капли тумана), 10 нм и 1 нм .

2. *Нижние кривые* на рис. 5.29 применимы не только к замкнутой сфере, для которой был выполнен анализ, но и к любым *вогнутым поверхностям* конденсата, над которыми имеется равновесный с ними пар, имеющий давление $p_r^\Pi < p_\infty$.

Пониженное давление над вогнутым мениском жидкости возникает в капилляре, стенки которого смачиваются этой жидкостью, как изображено на рис. 5.27 а. Изобарный потенциал для обратимого процесса испарение $\rightleftharpoons$ конденсация в форме (5.8.27) представлен уравнением (5.8.28). Поскольку $p_r^\Pi < p_\infty$, то отсюда получаем $\Delta G_r > 0$, то есть квазихимическая реакция (5.8.27) смещена влево, в сторону конденсации. Следовательно, заполнение капилляров жидкостью начинается при более низких давлениях, чем того требует плоская поверхность. Именно это и составляет сущность **капиллярной конденсации**. По мере заполнения капилляра радиус кривизны мениска увеличивается, так что объем жидкости, сконденсировавшейся в капиллярах, достигает предельной величины при давлении p_∞ .

Капиллярная конденсация играет важную роль при адсорбции паров пористыми материалами и более детально будет рассмотрена ниже (см. п. 5.11).

3. *Растворимость твердых частиц* в жидкостях может быть проанализирована применением вышеизложенного подхода к равновесию между жидким раствором и сферическими частицами твердого вещества, растворимость которого задается величиной

c_∞ , характерной для частиц больших размеров. Получаемый результат аналогичен тому, что дает формула (5.8.26) с верхним знаком при замене давления p_r^Π растворимостью c_r , которая теперь зависит от радиуса частицы r , согласно соотношению

$$c_r = c_\infty \exp\left(\frac{v^T}{RT} \frac{2\sigma}{r}\right). \quad (5.8.29)$$

Здесь σ — поверхностное натяжение на границе между твердой фазой и жидким раствором, а $v^T = M/\rho^T$ — мольный объем твердого вещества с молярной массой M и плотностью ρ^T . Следовательно, мелкие частицы твердого вещества обладают большей растворимостью в жидкости, чем крупные.

Поведение частиц малых размеров в растворе описывается уравнением изотермы Вант-Гоффа (2.13.13), записанным по аналогии с выражением (5.8.28) как

$$\Delta F_r = RT \left[\ln c_\infty - \ln K_c(T) \right] = RT \ln \frac{c_\infty}{c_r}. \quad (5.8.30)$$

В силу того, что $c_r > c_\infty$, из выражения (5.8.30) получаем, что зародыши новой фазы и кластеры малых размеров (как и малые частицы в газе) нестабильны из-за большей растворимости в жидкой фазе. Действительно, раствор, являющийся насыщенным для крупных частиц, для малых остается ненасыщенным ($\Delta F_r < 0$), и они продолжают растворяться в этом растворе.

Значения c_r (как и p_r^Π в газах) различны для частиц разных размеров и для участков поверхности, имеющих впадины и выступы разной кривизны. Направление переноса вещества в процессе перехода системы к равновесию однозначно определяется размерами неоднородностей. Действительно, химический потенциал частиц в растворе, $\mu(r) = \mu^\circ(T) + RT \ln c_r$, имеет большие значения, как показывает формула (5.8.29), для частиц меньших размеров, то есть $\mu(r_2) > \mu(r_1)$ при $r_2 < r_1$. Следовательно, крупные частицы (или капли) растут за счет растворения (или испарения) более мелких, а неровные поверхности сглаживаются за счет растворения выступов и заполнения впадин.

Оба полученных соотношения (5.8.26) и (5.8.29) в литературе называют *формулами Гиббса–Томсона* (встречается также название формулы (5.8.26) уравнением Кельвина).

3. Температура плавления частиц малых размеров. Нестабильность малых частиц проявляется не только в отношении испарения и растворимости, но и приводит к понижению их температуры плавления. Этот вывод является прямым следствием формулы Гиббса–Томсона (5.8.26).

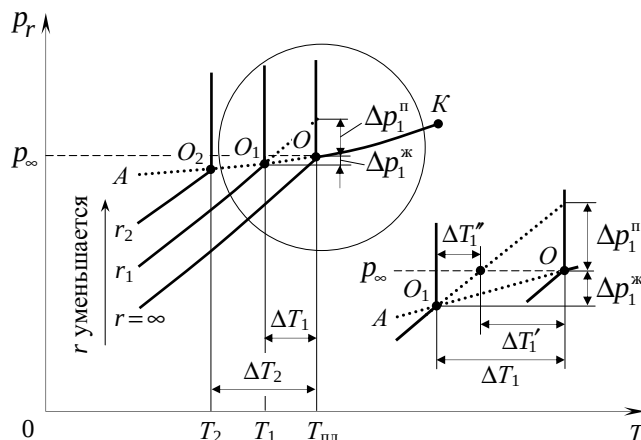


Рис. 5.30. Качественный вид фазовой диаграммы с учетом конечного радиуса r твердых частиц, сдвигающего кривые сублимации. Положение тройных точек O , O_1 и O_2 определяет температуры плавления $T_{пл} \equiv T_\infty$, $T_1 \equiv T(r_1)$ и $T_2 \equiv T(r_2)$, соответственно, для безграничной среды (при $r = \infty$) и для частиц с характерными радиусами r_1 и r_2 . Участок диаграммы внутри круга показан в правой части рисунка в увеличенном масштабе и использован для расчета понижения температуры плавления ΔT_1

Сначала рассмотрим качественную сторону этого явления, основываясь на P – T диаграмме фазового состояния чистого вещества, приведенной на рис. 2.1. Изображенные там кривые сублимации, испарения и плавления перенесены на рис. 5.30 в форме кривых, пересекающихся в тройной точке O . Пунктирная линия OA является продолжением кривой испарения OK . Кривая сублимации, выражающая температурную зависимость давления насыщенного пара над твердой фазой, показана в трех вариантах: для безграничной среды (при $r \rightarrow \infty$) и для твердых частиц с характерными радиусами r_1 и r_2 . Поскольку с уменьшением размеров частицы давление насыщенного пара увеличивается (см. кривую $p_r^п$ на рис. 5.29, лежащую выше горизонтальной линии p_∞), то кривые сублимации смещаются вверх, что показано вертикальной стрелкой на рис. 5.30.

Как было отмечено в п. 2.7, тройная точка отвечает температуре плавления, когда равновесные между собой жидкость и твердое тело формируют общую паровую фазу, давление которой теперь определяется не только температурой, но и размерами твердых частиц. Поэтому точки O_1 и O_2 на рис. 5.30, в которых пересекаются линии сублимации с пунктирной линией испарения OA , дают соответствующие температуры плавления $T_1 \equiv T(r_1)$ и $T_2 \equiv T(r_2)$. Следовательно, для частиц малых размеров их

температура плавления всегда ниже точки плавления $T_{\text{пл}} \equiv T_{\infty}$ для безграничной твердой фазы.

Для получения количественных соотношений используем формулу Гиббса–Томсона (5.8.26) с верхним знаком, в которой полагаем $\Delta p_r^{\text{п}} \equiv p_r^{\text{п}} - p_{\infty} \ll p_{\infty}$ и раскладываем экспоненту в ряд Тейлора. Тогда в линейном приближении получаем

$$\frac{\Delta p_r^{\text{п}}}{p_{\infty}} \approx \frac{v^{\text{к}}}{RT} \frac{2\sigma}{r}. \quad (5.8.31)$$

Конечной целью настоящего анализа является вывод формулы, дающей понижение температуры плавления (ΔT) для частиц конечных размеров относительно точки плавления вещества $T_{\text{пл}}$. Для расчета приращения температуры ΔT при изменении давления на величину Δp будем использовать следующее соотношение:

$$\Delta T = \frac{dT}{dp} \Delta p = \left(\frac{d \ln p(T)}{dT} \right)^{-1} \frac{\Delta p}{p}. \quad (5.8.32)$$

Входящая сюда температурная зависимость давления насыщенного пара $p(T)$ для кривых сублимации и испарения дается уравнением (2.2.6), переписанным следующим образом:

$$\left(\frac{d \ln p}{dT} \right)_{\text{суб}} = \frac{\Delta H_{\text{суб}}}{RT^2} \quad \text{и} \quad \left(\frac{d \ln p}{dT} \right)_{\text{исп}} = \frac{\Delta H_{\text{исп}}}{RT^2}. \quad (5.8.33)$$

Здесь теплоты сублимации и испарения ($\Delta H_{\text{суб}}$ и $\Delta H_{\text{исп}}$) взяты в тройной точке O , т. е. при температуре $T_{\text{пл}}$, которой соответствует давление насыщенного пара p_{∞} (см. рис. 5.30).

Геометрические построения для участка фазовой диаграммы внутри круга на рис. 5.30, выполненные в правой его части, показывают, что для частиц с характерным радиусом r_1 величина ΔT_1 содержит два вклада, обязанных наклону кривой сублимации $(dp/dT)_{\text{суб}}$:

$$\Delta T_1 = \Delta T_1' + \Delta T_1''. \quad (5.8.34)$$

Первый вклад порожден приращением давления пара на величину $\Delta p_1^{\text{п}}$, которая показана на рис. 5.29 и 5.30 для частиц радиусом r_1 . Применение формулы (5.8.32) позволяет записать

$$\Delta T_1' = \left(\frac{d \ln p(T)}{dT} \right)^{-1}_{\text{суб}} \frac{\Delta p_1^{\text{п}}}{p_{\infty}}. \quad (5.8.35)$$

Второй вклад связан с изменением давления пара на величину $\Delta p_1^{\text{ж}}$, вызванным понижением давления над жидкостью при переходе вдоль кривой испарения из точки O в точку O_1 . Тогда

по аналогии с выражением (5.8.35) записываем

$$\Delta T_1'' = \left(\frac{d \ln p}{dT} \right)_{\text{суб}}^{-1} \frac{\Delta p_1^*}{p_\infty}. \quad (5.8.36)$$

Величина Δp_1^* определяется наклоном кривой испарения $(dp/dT)_{\text{исп}}$ и, в соответствии с рис. 5.30, для нее справедливо следующее равенство:

$$\Delta p_1^* = \left(\frac{dp}{dT} \right)_{\text{исп}} \Delta T_1 \quad \text{или} \quad \frac{\Delta p_1^*}{p_\infty} = \left(\frac{d \ln p}{dT} \right)_{\text{исп}} (\Delta T_1' + \Delta T_1''). \quad (5.8.37)$$

Подстановка равенства (5.8.37) в выражение (5.8.36) позволяет выразить $\Delta T_1''$ и ΔT_1 через $\Delta T_1'$:

$$\Delta T_1'' = \frac{\alpha}{1 - \alpha} \Delta T_1' \quad \text{и} \quad \Delta T_1 \equiv \Delta T_1' + \Delta T_1'' = \frac{1}{1 - \alpha} \Delta T_1', \quad (5.8.38)$$

где введено обозначение

$$\alpha \equiv \frac{(d \ln p / dT)_{\text{исп}}}{(d \ln p / dT)_{\text{суб}}} = \frac{\Delta H_{\text{исп}}}{\Delta H_{\text{суб}}}, \quad (5.8.39)$$

при этом последнее равенство получено с использованием уравнений (5.8.33).

Подстановка (5.8.31) и (5.8.33) в равенство (5.8.35) дает

$$\frac{\Delta T_1'}{T_{\text{пл}}} = \frac{v^{\text{к}}}{\Delta H_{\text{суб}}} \frac{2\sigma}{r_1}. \quad (5.8.40)$$

Тогда из формул (5.8.38) и (5.8.40) с учетом (5.8.39) получаем искомое выражение (опуская везде нижний индекс 1, характеризующий радиус r_1 , и вводя обозначение $\Delta T_r \equiv \Delta T(r)$):

$$\frac{\Delta T_r}{T_{\text{пл}}} = \frac{v^{\text{к}}}{\Delta H_{\text{суб}} - \Delta H_{\text{исп}}} \frac{2\sigma}{r}. \quad (5.8.41)$$

Поскольку по закону Гесса теплоты сублимации, испарения и плавления связаны между собой соотношением $\Delta H_{\text{суб}} = \Delta H_{\text{исп}} + \Delta H_{\text{пл}}$, то выражение (5.8.41) принимает привычную форму

$$\frac{\Delta T_r}{T_{\text{пл}}} = \frac{v^{\text{к}}}{\Delta H_{\text{пл}}} \frac{2\sigma}{r}. \quad (5.8.42)$$

Здесь понижение температуры плавления частиц конечных размеров определено как $\Delta T_r = T_{\text{пл}} - T(r) \equiv T_\infty - T_r$.

Таким образом, уравнение (5.8.42), полученное еще Томсоном, предсказывает уменьшение температуры плавления частиц, обратно пропорциональное их радиусу r . Однако позже, при

экспериментальном исследовании тепловых свойств *наночастиц*, было обнаружено, что коэффициент пропорциональности в уравнении (5.8.42) обеспечивает завышенные значения ΔT_r . Оказалось, что уравнение Томсона дает приемлемое совпадение с экспериментом *только для микрочастиц* радиусом более 0,1 мкм, когда понижение температуры плавления (ΔT_r) не превышает сотых долей температуры плавления вещества ($T_{\text{пл}} \equiv T_{\infty}$).

Полученный результат становится вполне объяснимым, если вспомнить, что на пути к уравнению Томсона (5.8.42) было использовано в уравнении (5.8.31) требование малости, $\Delta p_r^n \ll p_{\infty}$, при разложении в ряд Тейлора экспоненты в формуле Гиббса – Томсона (5.8.26) в линейном приближении.

Более того, в основу вывода уравнения Томсона (5.8.42) заложена физическая некорректность. Дело в том, что оно было получено в предположении фазового равновесия между паровой фазой, сформированной безграничным жидким расплавом, и твердыми частицами конечных размеров. Именно это отражено на рис. 5.30, где температуры плавления ($T_1, T_2, \dots$) для частиц разного радиуса ($r_1, r_2, \dots$) находятся как точки пересечения ($O_1, O_2, \dots$) кривых сублимации, зависящих от r , с кривой OA , характеризующей испарение с плоской границы жидкой фазы.

В действительности при температуре плавления сферического кластера радиусом r жидкий расплав отсутствует, поскольку в силу $T_r < T_{\text{пл}}$ при нагревании наночастица должна расплавиться раньше, чем появится расплав массивного твердого тела. В этом случае жидкая фаза существует лишь в форме одиночных жидких (расплавленных) частиц малого размера, которые формируют свое давление насыщенного пара, большее чем $p_{\infty} \equiv p(T_{\text{пл}})$.

На этом основании надо более строго рассмотреть задачу о температуре плавления частиц, взяв за основу подход, использованный выше при выводе дифференциального уравнения (5.8.24), решением которого явилась формула Гиббса – Томсона (5.8.26).

4. Дифференциальное уравнение для температуры плавления нанокластеров. Определим температуру плавления твердой сферической частицы радиусом r^{T} как температуру T_r , при которой она находится в термодинамическом равновесии с жидкой (расплавленной) частицей радиусом $r^{\text{Ж}}$ и массой, равной исходной.

Условие сохранения числа молей при плавлении частицы дает связь между отношением радиусов частицы в жидком и твердом состояниях и отношением мольных объемов (в общем случае зависящих от r):

$$\frac{r^{\text{Ж}}}{r^{\text{T}}} = \left(\frac{v_r^{\text{Ж}}}{v_r^{\text{T}}} \right)^{1/3}. \quad (5.8.43)$$

При этом молярные объемы твердой и жидкой частицы произвольных размеров (что отмечено индексом r) определяем, по аналогии с их объемными величинами, следующим образом:

$$v_r^\tau = \frac{\partial \mu_r^\tau}{\partial p_r^\tau} \quad \text{и} \quad v_r^{\text{ж}} = \frac{\partial \mu_r^{\text{ж}}}{\partial p_r^{\text{ж}}}, \quad (5.8.44)$$

Здесь $\mu_r^{\tau(\text{ж})}$ и $p_r^{\tau(\text{ж})}$ — химический потенциал и давление *внутри* твердой (жидкой) частицы радиусом r . Это давление при помощи формулы (5.8.18) связано с давлением паров p_r^Π *над частицей*, которое подчиняется уравнению Гиббса–Томсона типа (5.8.26) с верхним знаком (плюс).

Аналогично равенствам (5.8.44), вводят молярные энтропии твердой и жидкой частицы произвольных размеров:

$$S_r^\tau = -\frac{\partial \mu_r^\tau}{\partial T_r} \quad \text{и} \quad S_r^{\text{ж}} = -\frac{\partial \mu_r^{\text{ж}}}{\partial T_r}. \quad (5.8.45)$$

Исходным для вывода дифференциального уравнения является равенство химических потенциалов (ср. уравнение (5.8.19)):

$$\mu_r^\tau(p_r^\tau, T_r) = \mu_r^{\text{ж}}(p_r^{\text{ж}}, T_r). \quad (5.8.46)$$

Дифференцируем по r^τ обе части равенства (5.8.46):

$$\frac{\partial \mu_r^\tau(p_r^\tau, T_r)}{\partial r^\tau} = \frac{\partial \mu_r^{\text{ж}}(p_r^{\text{ж}}, T_r)}{\partial r^\tau}. \quad (5.8.47)$$

Сначала преобразуем левую часть формулы (5.8.47) для твердой частицы, учитывая равенства (5.8.44) и (5.8.45):

$$\frac{\partial \mu_r^\tau(p_r^\tau, T_r)}{\partial r^\tau} = \frac{\partial \mu_r^\tau}{\partial p_r^\tau} \frac{\partial p_r^\tau}{\partial r^\tau} + \frac{\partial \mu_r^\tau}{\partial T_r} \frac{\partial T_r}{\partial r^\tau} = v_r^\tau \frac{\partial p_r^\tau}{\partial r^\tau} - S_r^\tau \frac{\partial T_r}{\partial r^\tau}. \quad (5.8.48)$$

На основании формул (5.8.22) и (5.8.24) записываем

$$v_r^\tau \frac{\partial p_r^\tau}{\partial r^\tau} = RT_r \frac{\partial \ln p_r^\Pi}{\partial r^\tau} = -\frac{2\sigma_r^\tau v_r^\tau}{(r^\tau)^2}. \quad (5.8.49)$$

Подстановка (5.8.49) в формулу (5.8.48) дает окончательное выражение для левой части равенства (5.8.47), а именно

$$\frac{\partial \mu_r^\tau(p_r^\tau, T_r)}{\partial r^\tau} = -\frac{2\sigma_r^\tau v_r^\tau}{(r^\tau)^2} - T_r S_r^\tau \frac{\partial \ln T_r}{\partial r^\tau}. \quad (5.8.50)$$

Правая часть формулы (5.8.47) для жидкой частицы, с учетом соотношения (5.8.43), принимает следующий вид:

$$\frac{\partial \mu_r^{\text{ж}}(p_r^{\text{ж}}, T_r)}{\partial r^\tau} = \frac{\partial \mu_r^{\text{ж}}(p_r^{\text{ж}}, T_r)}{\partial r^{\text{ж}}} \frac{\partial r^{\text{ж}}}{\partial r^\tau}, \quad \text{где} \quad \frac{\partial r^{\text{ж}}}{\partial r^\tau} = \left(\frac{v_r^{\text{ж}}}{v_r^\tau} \right)^{1/3}. \quad (5.8.51)$$

По аналогии с выражением (5.8.50), путем замены всех верхних индексов «т → ж», записываем для жидкой частицы

$$\frac{\partial \mu_r^{\text{ж}}(p_r^{\text{ж}}, T_r)}{\partial r^{\text{ж}}} = -\frac{2\sigma_r^{\text{ж}} v_r^{\text{ж}}}{(r^{\text{ж}})^2} - T_r S_r^{\text{ж}} \frac{\partial \ln T_r}{\partial r^{\text{ж}}}. \quad (5.8.52)$$

Подстановка (5.8.52) в формулу (5.8.51) дает окончательное выражение для правой части равенства (5.8.47), а именно

$$\frac{\partial \mu_r^{\text{ж}}(p_r^{\text{ж}}, T_r)}{\partial r^{\text{т}}} = -\frac{2\sigma_r^{\text{ж}} v_r^{\text{ж}}}{r^{\text{т}} r^{\text{ж}}} - T_r S_r^{\text{ж}} \frac{\partial \ln T_r}{\partial r^{\text{т}}}. \quad (5.8.53)$$

В соответствии с (5.8.47), приравниваем друг другу выражения (5.8.50) и (5.8.53) с учетом соотношения (5.8.43) и того факта, что мольная теплота плавления равна $\Delta H_{\text{пл}}(r) = T_r(S_r^{\text{ж}} - S_r^{\text{т}})$. В итоге получаем *дифференциальное уравнение* для нахождения $T_r \equiv T_{\text{пл}}(r)$:

$$\frac{\partial \ln T_r}{\partial r^{\text{т}}} = \frac{v_r^{\text{т}}}{\Delta H_{\text{пл}}(r)} \frac{2\sigma_{\text{эфф}}(r)}{(r^{\text{т}})^2}. \quad (5.8.54)$$

Здесь введено *эффективное поверхностное натяжение*,

$$\sigma_{\text{эфф}}(r) = \sigma_r^{\text{т}} - \sigma_r^{\text{ж}} \left(\frac{v_r^{\text{ж}}}{v_r^{\text{т}}} \right)^{2/3}, \quad (5.8.55)$$

учитывающее совместный вклад границ раздела твердой и жидкой фаз с собственным паром и в общем случае зависящее от r .

Уравнение (5.8.54) для T_r по структуре совпадает с уравнением (5.8.24), дающим давление $p_r^{\text{т}}$ в виде формулы Гиббса–Томсона (5.8.26). Следовательно, решение уравнения (5.8.54) (в пренебрежении зависимостью от r величин, стоящих в правой части) строится по той же схеме (5.8.25) и принимает форму:

$$T_r = T_{\text{пл}} \exp\left(-\frac{v^{\text{т}}}{\Delta H_{\text{пл}}} \frac{2\sigma_{\text{эфф}}}{r^{\text{т}}}\right) \equiv T_{\text{пл}} e^{-r_0/r^{\text{т}}}. \quad (5.8.56)$$

Здесь введен параметр длины r_0 , не зависящий от r и имеющий вид

$$r_0 = \sigma_{\text{эфф}} \frac{2v^{\text{т}}}{\Delta H_{\text{пл}}}. \quad (5.8.57)$$

Для сферических частиц достаточно большого радиуса, когда $r^{\text{т}} \gg r_0$, выражение (5.8.56) принимает привычную форму,

$$\frac{\Delta T_r}{T_{\text{пл}}} = \frac{v^{\text{т}}}{\Delta H_{\text{пл}}} \frac{2\sigma_{\text{эфф}}}{r^{\text{т}}} \equiv \frac{v^{\text{т}}}{\Delta H_{\text{пл}}} \frac{2\sigma^{\text{т}}}{r^{\text{т}}} \left[1 - \frac{\sigma^{\text{ж}}}{\sigma^{\text{т}}} \left(\frac{v^{\text{ж}}}{v^{\text{т}}} \right)^{2/3} \right]. \quad (5.8.58)$$

При $\sigma_{\text{эфф}} = \sigma^{\text{т}}$ (то есть в пренебрежении вкладом жидкой фазы при $\sigma^{\text{ж}} = 0$) выражение (5.8.58) полностью совпадает с уравнением Томсона (5.8.42).

Таким образом, линейная зависимость ΔT_r от $1/r^T$, присущая уравнению Томсона (5.8.42), в том числе и в более общем случае при замене $\sigma^T \rightarrow \sigma_{\text{эф}}$, реализуется только для частиц достаточно большого размера. Для частицы радиусом менее $5r_0$ надо учитывать зависимость ΔT_r от $1/r^T$ в форме экспоненты (5.8.56).

Физический смысл полученных формул становится вполне понятным, если параметр r_0 представить, с учетом (5.8.55), как

$$r_0 = (\sigma^T a^T - \sigma^J a^J) \frac{2(v^T)^{1/3}}{\Delta H_{\text{пл}}}. \quad (5.8.59)$$

Здесь введены величины (имеющие размерность $\text{м}^2/\text{моль}$),

$$a^T = (v^T)^{2/3} \quad \text{и} \quad a^J = (v^J)^{2/3}, \quad (5.8.60)$$

пропорциональные площади поверхности, приходящейся на один моль в твердом состоянии (a^T) и в жидком состоянии (a^J). Следовательно, взятая со знаком минус скобка в (5.8.59) (то есть $(\sigma^J a^J - \sigma^T a^T)$) представляет собой изменение поверхностной энергии, происходящее при плавлении твердой частицы, когда ее площадь поверхности a^T заменяется площадью a^J .

Установим связь между давлениями насыщенного пара над частицей в жидком состоянии ($p_r^{\text{п/ж}}$) и в твердом состоянии ($p_r^{\text{п/т}}$), которые термодинамически равновесны друг с другом при температуре плавления T_r . По формуле Гиббса–Томсона (5.8.26) эти давления, соответственно, равны

$$p_r^{\text{п/ж}} = p_\infty \exp\left(\frac{v^J}{RT_r} \frac{2\sigma^J}{r^J}\right) \quad \text{и} \quad p_r^{\text{п/т}} = p_\infty \exp\left(\frac{v^T}{RT_r} \frac{2\sigma^T}{r^T}\right).$$

Отсюда, учитывая соотношение (5.8.43) между радиусами жидкой и твердой частицы и вводя эффективное поверхностное натяжение (5.8.55), нетрудно найти отношение давлений:

$$\frac{p_r^{\text{п/ж}}}{p_r^{\text{п/т}}} = \exp\left(-\frac{v^T}{RT_r} \frac{2\sigma_{\text{эф}}}{r^T}\right) \equiv \exp\left(-\frac{\Delta H_{\text{пл}}}{RT_r} \frac{r_0}{r^T}\right). \quad (5.8.61)$$

Выражения (5.8.56) и (5.8.61) дают искомую связь между отношением давлений и температурой плавления частицы:

$$\frac{p_r^{\text{п/ж}}}{p_r^{\text{п/т}}} = \left(\frac{T_r}{T_{\text{пл}}}\right)^{\Delta H_{\text{пл}}/RT_r}. \quad (5.8.62)$$

Если взять давление паров над плоской границей при температуре T_r плавления частицы, равное $p_\infty(T_r)$, то в силу неравенства $T_r < T_{\text{пл}}$ из формулы (5.8.62) следует, что всегда

$$p_\infty(T_r) < p_r^{\text{п/ж}} < p_r^{\text{п/т}}. \quad (5.8.63)$$

Формирование давлений паров над жидкой и твердой частицами ($p_r^{\text{п/ж}}$ и $p_r^{\text{п/т}}$) при температуре T_r запишем, подобно (5.8.27), в виде квазихимических реакций испарение $\rightleftharpoons$ конденсация и сублимация $\rightleftharpoons$ кристаллизация совместно с соответствующими им законами действующих масс:

$$\begin{aligned} L_r(\text{ж}) \rightleftharpoons V_r^{\text{ж}}(\text{п}) \quad \text{и} \quad K_p^{\text{ж}}(T_r) = p_r^{\text{п/ж}}, \\ S_r(\text{т}) \rightleftharpoons V_r^{\text{т}}(\text{п}) \quad \text{и} \quad K_p^{\text{т}}(T_r) = p_r^{\text{п/т}}. \end{aligned} \quad (5.8.64)$$

По аналогии с выражением (5.8.28) записываем соотношение между изобарными потенциалами двух реакций (5.8.64):

$$\Delta G_r^{\text{ж}} \equiv RT_r \ln \frac{p_{\infty}(T_r)}{p_r^{\text{п/ж}}} > RT_r \ln \frac{p_{\infty}(T_r)}{p_r^{\text{п/т}}} \equiv \Delta G_r^{\text{т}}, \quad (5.8.65)$$

где знак неравенства следует из второго неравенства в формуле (5.8.63). При этом, согласно первому неравенству, оба изобарных потенциала в (5.8.65) отрицательные, а тогда $|\Delta G_r^{\text{ж}}| < |\Delta G_r^{\text{т}}|$. Это означает, что по отношению к газовой фазе твердые частицы менее устойчивы, чем жидкие, то есть равновесие между ними смещено в сторону последних. Твердые частицы быстрее исчезают, что создает в конечном итоге единый расплав с $r \rightarrow \infty$.

Металл	Al	Cu	Ni
$T_{\text{пл}}, \text{ К}$	934	1358	1728
$\Delta H_{\text{пл}}, \frac{\text{кДж}}{\text{моль}}$	10,70	13,05	17,47
$\sigma^{\text{т}}, \frac{\text{Дж}}{\text{м}^2}$	1,032	1,592	2,104
$\sigma^{\text{ж}}, \frac{\text{Дж}}{\text{м}^2}$	0,865	1,310	1,750
$v^{\text{т}}, \frac{\text{см}^3}{\text{моль}}$	10,80	7,58	7,14
$v^{\text{ж}}, \frac{\text{см}^3}{\text{моль}}$	11,19	8,00	7,41
$\sigma_{\text{эфф}}, \frac{\text{Дж}}{\text{м}^2}$	0,146	0,234	0,310
$r_0, \text{ нм}$	0,295	0,272	0,253

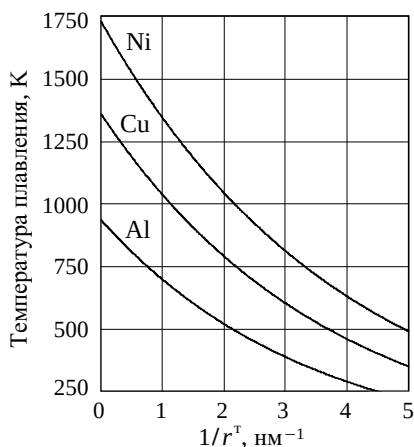


Рис. 5.31. Численные значения физических параметров для Al, Cu и Ni, использованные при расчете зависимости температуры плавления наночастиц от их размеров по формулам (5.8.55)–(5.8.57), и соответствующие кривые

Результаты расчета зависимости температуры плавления наночастиц от их обратного радиуса по формуле (5.8.56) показаны на рис. 5.31 для Al, Cu и Ni. Физические параметры металлов

даны на этом же рисунке в виде таблицы. Здесь также приведены значения эффективного поверхностного натяжения $\sigma_{\text{эф}}$ и параметра длины r_0 , рассчитанные по формулам (5.8.55) и (5.8.57).

Из рис. 5.31 видно, что для достаточно малых частиц понижение температуры плавления может достигать сотен градусов, что подтверждается экспериментом. Заметное уменьшение температуры наблюдается для частиц с размерами в единицы и доли нанометра, где справедлив экспоненциальный закон (5.8.56). Как уже отмечалось, для частиц достаточно большого радиуса (при $r^T > 5r_0$) можно использовать упрощенную формулу (5.8.58). В отличие от уравнения Томсона (5.8.42), она учитывает влияние свойств расплавленной частицы с помощью $\sigma_{\text{эф}}$.

В приведенном выше численном расчете в качестве физических параметров металлов были использованы величины, присущие массивным материалам (без индекса r). Это естественным образом вносит погрешность в случае нанокластеров достаточно малых размеров. В частности, когда характерный размер частиц соизмерим с радиусом сферы межмолекулярного взаимодействия, проявляется размерная зависимость поверхностного натяжения $\sigma_r \equiv \sigma(r)$. Существующие на сегодняшний день теории поверхностных явлений не дают точного вида подобной зависимости. Предложено несколько феноменологических выражений, из которых наиболее часто используется формула $\sigma(r) = \sigma_\infty(1 - \delta/r)$, справедливая при $\delta/r \ll 1$, где δ — эмпирический параметр порядка толщины межфазного переходного слоя. Экспериментально установлено, что для нанокластеров Ag, Au и Pb размером 3–5 нм понижение поверхностного натяжения достигает 30–40%.

5.9. Адсорбционные явления на поверхности твердых тел

1. Общая характеристика адсорбционных процессов. Ранее отмечалось, что в силу развитой поверхности наноструктурированные и высокопористые вещества относятся к классу сильно неравновесных систем. Стремление системы к самопроизвольному уменьшению избыточной поверхностной энергии сопровождается не только рассмотренными выше процессами релаксации и реконструкции поверхности, но и адсорбционными явлениями.

Адсорбцией называется процесс концентрирования вещества на границе раздела двух смежных фаз. Вещество, поверхность которого удерживает частицы, называют *адсорбентом*, а вещество, поглощаемое из соседней объемной фазы, — *адсорбатом*. Процесс, обратный адсорбции, называется *десорбцией*.

Адсорбция является частным случаем *сорбции*, включающей также *абсорбцию* — поглощение вещества в *объеме* конденсированной среды, например в компактированных ультрадисперсных и высокопористых материалах.

Адсорбированные частицы взаимодействуют с молекулами или атомами поверхности и понижают как поверхностную свободную энергию, так и суммарную энергию системы в целом ($\Delta G < 0$). Это свидетельствует о самопроизвольности процесса адсорбции. При этом частицы лишаются, по крайней мере, одной из поступательных степеней свободы (по нормали к поверхности), и энтропия системы понижается ($\Delta S < 0$). В результате этого тепловой эффект $\Delta H = \Delta G + T\Delta S < 0$, то есть процесс адсорбции всегда является *экзотермическим*.

Во многих случаях адсорбционные явления определяют характер физико-химических процессов, протекающих на поверхности. Например, в наноструктурированных материалах адсорбционные слои на поверхности частиц дисперсной фазы могут существенно изменить условия их взаимодействия между собой и тем самым повлиять на свойства системы в целом.

В адсорбционных слоях на поверхности твердого тела возникает избыточное поверхностное давление, которое обычно приводит к деформации тела. Кроме того, адсорбция изменяет величину поверхностных напряжений и вызывает реконструкцию поверхности. Конечный результат определяется характером взаимодействия на границе адсорбат–адсорбент: одни атомы могут вызывать растяжение, а другие — сжатие. Например, адсорбция электроотрицательных атомов на металлах сопровождается сжатием поверхности, а электроположительных — ее растяжением.

Увеличение поверхностных напряжений часто приводит к перестройке и изменению поверхностной структуры. Различают два типа такой перестройки. В первом случае происходит поверхностный фазовый переход, вызванный изменением равновесного положения атомов, — так называемый *фазовый переход смещения*. Во втором случае изменяется тип структуры, то есть происходит *структурный фазовый переход*. Тип элементарной ячейки в поверхностном слое определяется конкуренцией между процессами образования химических связей адатомов с подложкой и между собой. Наличие указанных фазовых переходов подтверждено многочисленными экспериментальными исследованиями. Тем не менее на сегодняшний день отсутствует достаточно строгая и последовательная теория, позволяющая предсказывать величину и форму элементарной поверхностной ячейки для произвольной комбинации «адатом–подложка».

При адсорбции молекул ПАВ или электролитов часто наблюдается явление, связанное с развитием микродефектов структуры твердого тела, которое называется адсорбционным понижением прочности. В результате нелокализованной адсорбции молекулы или ионы стремятся раздвинуть микрощели и этим способствуют диспергированию под действием внешних деформирующих сил. Облегчение диспергирования под влиянием адсорбции получило название эффекта Ребиндера.

В последнее время стимулирующим фактором интенсивного изучения адсорбционных процессов является развитие нанотехнологий и материаловедения наноструктурированных веществ. Особый интерес представляет исследование влияния адсорбции на электронные характеристики полупроводников и металлов в связи с возможностью создания металлических нанокластеров, сверхрешеток и многослойных полупроводниковых гетероструктур. Поскольку эти вопросы детально освещены в многочисленной литературе, кратко отметим лишь наиболее важные из известных результатов.

Так, адсорбция на металлах сопровождается сильным изменением работы выхода электронов (вплоть до 1 эВ). Это вызвано формированием на поверхности металла двойного электрического слоя, возникающего либо из-за поляризации адсорбированных молекул, либо из-за обмена заряженными частицами. При этом в силу высокой плотности электронов в металлах возмущению подвергается лишь тонкий слой (в доли нм), прилегающий к поверхности. Когда адатом приближается к поверхности, может возникнуть новая электронная структура, поскольку сильное взаимодействие между ним и поверхностью твердого тела изменяет электронные уровни обоих.

В полупроводниках плотность носителей заряда существенно меньше, чем в металлах. Поэтому адсорбированные атомы приводят к формированию достаточно протяженного приповерхностного слоя (вплоть до сотен нм). В результате адсорбции электрически активных атомов/молекул происходит либо инжекция носителей в приповерхностный слой, либо их перенос на адсорбированные частицы. Из-за этого приповерхностный слой обогащается электронами или дырками, а в зонной структуре появляется поверхностный изгиб энергетических зон, который необходимо учитывать при создании устройств микро- и наноэлектроники.

Многие оптические характеристики материалов, в первую очередь спектральные свойства коэффициентов отражения и пропускания, также зависят от состояния адсорбционных или конденсационных покрытий на поверхности твердых тел.

Все вышеприведенные случаи служат обоснованием важности изучения закономерностей образования, строения и свойств адсорбционных слоев.

2. Физическая и химическая адсорбция. При адсорбции атомы или молекулы удерживаются поверхностью адсорбента. Вероятность удержания частиц в поверхностном слое определяется соотношением между потенциальной энергией их притяжения к поверхности и кинетической энергией теплового движения. В тех случаях, когда первое превалирует над вторым, частицы можно считать адсорбированными. При этом они лишены, по крайней мере, одной из поступательных степеней свободы (по нормали к поверхности), свойственных частицам внешней (газовой) среды.

В зависимости от характера силовых взаимодействий между молекулами адсорбата и поверхностью адсорбента различают физическую адсорбцию (*физадсорбцию*) и химическую адсорбцию (*хемосорбцию*), представляющих собой два последовательных этапа единого сорбционного процесса.

Физическая адсорбция обусловлена силами межмолекулярного взаимодействия (силами Ван-дер-Ваальса) и силами водородной связи (см. п. 3.1). Межмолекулярные взаимодействия не сопровождаются обобществлением валентных электронов и по своей природе относятся к классу диполь-дипольных взаимодействий, в основе которых лежат дисперсионные, индукционные и ориентационные силы Ван-дер-Ваальса.

Дисперсионные силы возникают вследствие флуктуации дипольных моментов в нейтральных атомах или молекулах. Они порождены синхронными осцилляциями взаимно индуцированных мгновенных диполей соседних молекул. Дело в том, что мгновенное положение центра электронного облака не совпадает с центром положительных зарядов, то есть в молекуле имеется отличный от нуля флуктуирующий дипольный момент. При сближении двух молекул мгновенный дипольный момент одной из них создает в центре другой электрическое поле и поляризует ее. Последняя аналогичным образом наводит дипольный момент в первой молекуле. В результате образуется пара притягивающихся взаимно индуцированных и синхронно флуктуирующих диполей. Дисперсионные силы проявляются при адсорбции неполярных молекул (N_2 , O_2 , H_2 и инертных газов) на неполярных адсорбентах типа графита.

Индукционные силы порождены взаимодействием дипольной молекулы с другим наведенным ею диполем в нейтральной молекуле. Они характерны для адсорбции неполярных молекул (N_2 ,

O₂, H₂ и инертных газов) на полярных адсорбентах с ионной связью (типа NaCl, BaO и др.) и для адсорбции полярных молекул (H₂O, CO₂, NH₃ и др.) на неполярных адсорбентах и металлах.

Ориентационные силы порождены взаимной ориентацией диполей, присущих молекулам с постоянными дипольными моментами. Сюда относится обширная группа межмолекулярных взаимодействий, при которых происходит ориентационное электростатическое взаимодействие молекул, содержащих ионы, полярные группы или π -электроны. Эти силы характерны для адсорбции полярных молекул (H₂O, CO₂, NH₃ и др.) на полярных адсорбентах с ионной связью (типа NaCl, BaO и др.).

Водородная связь обеспечивает адсорбцию полярных молекул (H₂O, CO₂, NH₃ и др.) на поверхности адсорбентов с гидроксильными группами OH (типа стекол, силикагелей и цеолитов). Распределение электронной плотности в гидроксильных группах, прочно связанных с атомами кремний-кислородного остова, таково, что отрицательный заряд в них смещен к атому кислорода. Следовательно, гидроксильная группа OH представляет собой электрический диполь с положительным зарядом у атома водорода, выдвинутого наружу. Этот положительный заряд обеспечивает связь с отрицательными зарядами других молекул, например с молекулами воды. В этих случаях образуются молекулярные комплексы с водородной связью при одновременном сохранении дисперсионного и индукционного взаимодействий. Именно этим обусловлены высокие сорбционные свойства по отношению к воде, характерные для стекол, силикагелей (гелей кремниевой кислоты H₂SiO₃) и цеолитов (кристаллических алюмосиликатов щелочных и щелочноземельных металлов).

При физической адсорбции адсорбированная молекула сохраняет свою структуру, не теряя индивидуальности. Следовательно, молекула и адсорбент представляют собой независимые подсистемы, связанные слабыми силами Ван-дер-Ваальса. Поскольку прочность межмолекулярных связей невелика по сравнению с энергией связи между атомами адсорбента, то физически адсорбированная молекула может быть относительно легко удалена (десорбирована) с поверхности.

Энергетическое состояние адсорбированной частицы принято изображать в виде диаграммы $U(x)$, показывающей зависимость потенциальной энергии взаимодействия частицы с поверхностью твердого тела от расстояния до этой поверхности. На энергетической диаграмме, приведенной на рис. 5.32, устойчивому состоянию физической адсорбции соответствует положение молекулы в потенциальной яме глубиной $U_{\text{ф}}$, расположенной на расстоянии

x_{ϕ} от поверхности. Величина U_{ϕ} определяется силой взаимодействия молекулы с поверхностью и называется *энергией физической адсорбции*. Процесс физического сорбирования всегда сопровождается выделением энергии в виде *теплоты физической адсорбции* $Q_{\phi} = U_{\phi} N_A$, где N_A — число Авогадро.

Адсорбированная молекула, удерживаемая силами Ван-дер-Ваальса в потенциальной яме, совершает тепловые колебания по нормали к поверхности. Для нахождения частоты этих колебаний разложим потенциальную энергию $U(x)$ в ряд Тейлора по малому смещению, $u = x - x_{\phi}$, частицы из потенциального минимума, ограничиваясь квадратичным слагаемым:

$$U(x) = U(x_{\phi}) + \frac{1}{2} \frac{\partial^2 U(x)}{\partial x^2} \bigg|_{x=x_{\phi}} (x - x_{\phi})^2 + \dots \equiv U_{\phi} + \frac{1}{2} f u^2 + \dots$$

Вторая производная, обозначенная как $f = \partial^2 U / \partial x^2|_{x=x_{\phi}}$, характеризует кривизну потенциального минимума и называется коэффициентом упругости. В соответствии со вторым законом механики $m d^2 x / dt^2 = F$, сила $F \equiv -\partial U / \partial x = -f u$, действующая на частицу, приводит к уравнению ее движения:

$$m \frac{d^2 u}{dt^2} = -f u.$$

Уравнение такого вида описывает гармонические колебания типа $u(t) = A \sin \omega_0 t$, частота которых ($\omega_0 = \sqrt{f/m}$) определяется упругостью связи молекулы с поверхностью.

Тепловые колебания в направлении нормали к поверхности, совершаемые молекулами в потенциальной яме, приводят к тому, что наиболее энергичные из них покидают поверхность. Среднее время τ_{ϕ} жизни молекулы в физадсорбированном состоянии зависит от *энергии десорбции* $\mathcal{E}_{\text{дес}}$, которая при физадсорбции равна глубине потенциальной ямы U_{ϕ} .

Среднее время пребывания физадсорбированной молекулы на поверхности определяется соотношением, аналогичным формуле (4.2.1). Она была получена для диффузионных скачков частицы

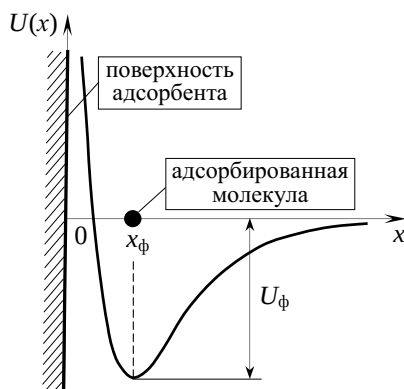


Рис. 5.32. Энергетическая диаграмма для процесса физической адсорбции на поверхности твердого тела

в периодическом поле кристаллической решетки с энергией активации $\mathcal{E}$, роль которой в данном случае играет $\mathcal{E}_{\text{дес}} = U_{\text{ф}}$. Тогда

$$\tau_{\text{ф}} = \tau_0 \exp\left(\frac{\mathcal{E}_{\text{дес}}}{k_{\text{В}}T}\right) \equiv \tau_0 \exp\left(\frac{Q_{\text{ф}}}{RT}\right). \quad (5.9.1)$$

Здесь время τ_0 по порядку величины близко к периоду колебаний молекулы в потенциальной яме, определяемому упругостью связи молекулы с поверхностью, и его можно положить равным $2\pi/\omega_0$, где ω_0 — введенная ранее частота колебаний. Так как для разных адсорбционных пар упругость связи не сильно меняется, то принято считать, что $\tau_0 = 10^{-13} - 10^{-12}$ с.

Теплота физической адсорбции $Q_{\text{ф}}$, как правило, не превышает 10 ккал/моль (примерно 42 кДж/моль), что близко к 0,5 эВ на одну молекулу. В этом случае время $\tau_{\text{ф}}$, вычисленное по формуле (5.9.1), при комнатной температуре измеряется микросекундами. Исключение составляют предельные углеводороды $\text{C}_n\text{H}_{2n+2}$ (так называемые алканы: метан CH_4 , этан C_2H_6 и т. д. — см. с. 607). Для них теплота адсорбции $Q_{\text{ф}}$ зависит от числа атомов углерода в молекуле и может достигать 30 ккал/моль, при этом величина $\tau_{\text{ф}}$ приближается к 30 годам.

Понижение температуры существенно увеличивает $\tau_{\text{ф}}$ даже в случае слабой адсорбции. Так, для инертного газа гелия, имеющего $Q_{\text{ф}} = 0,6$ кДж/моль, снижение температуры от комнатной до 77 К увеличивает $\tau_{\text{ф}}$ всего в два раза, а при 4 К величина $\tau_{\text{ф}}$ возрастает на семь с половиной порядков.

Для инертных газов, атомы которых имеют полностью застроенную электронную конфигурацию, сорбционный процесс начинается и заканчивается физической адсорбцией, так как они способны удерживаться у поверхности лишь межмолекулярными силами Ван-дер-Ваальса без участия валентных электронов.

Химическая адсорбция обусловлена силами взаимодействия химической природы типа гомеопольярной и гетеропольярной связи между атомами. Поэтому молекулы неинертных газов и паров металлов, пройдя этап физадсорбции, способны к дальнейшему более сильному взаимодействию с поверхностью твердого тела. Поскольку такая связь возможна лишь на атомарном уровне, то физически адсорбированная молекула предварительно диссоциирует на атомы, которые и обеспечивают межатомное взаимодействие, лежащее в основе процесса хемосорбции.

Хемосорбцию нельзя считать химической реакцией, поскольку при этом не образуется каких-либо новых веществ. Она является лишь преддверием гетерогенной химической реакции, также как физическая адсорбция предшествует процессу хемосорбции.

Энергетическая диаграмма сорбционного процесса, включающего физическую адсорбцию и химическую адсорбцию как два последовательных этапа, изображена на рис. 5.33. Пунктирная кривая 1 описывает процесс физической адсорбции и аналогична кривой на рис. 5.32. Кривая 2 соответствует процессу химической адсорбции двух атомов, возникших в результате диссоциации исходной молекулы, где $\mathcal{E}_{\text{дис}}$ — энергия диссоциации. Сплошной кривой 3 описывается результирующий процесс адсорбции молекулы, протекающий следующим образом.

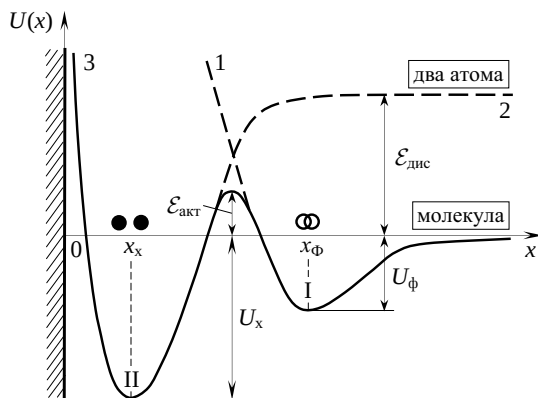


Рис. 5.33. Энергетическая диаграмма для последовательных этапов физической и химической адсорбции двухатомной молекулы на поверхности твердого тела

Приближающаяся к поверхности молекула самопроизвольно физически адсорбируется, то есть скатывается в потенциальную яму I глубиной U_f и располагается на расстоянии x_f от поверхности. Дальнейший переход молекулы в хемосорбционную потенциальную яму II глубиной U_x (в виде двух атомов, расположенных на расстоянии x_x) требует преодоления потенциального барьера высотой $\mathcal{E}_{\text{акт}}$, разделяющего ямы I и II. На это способны лишь энергичные молекулы, обладающие достаточным запасом энергии, превышающим величину $\mathcal{E}_{\text{акт}}$. Высота $\mathcal{E}_{\text{акт}}$ хемосорбционного потенциального барьера называется *энергией активации процесса хемосорбции*. В этом смысле хемосорбция, подобно диффузии, является термоактивационным процессом. Поэтому ее часто называют *активационной* (активированной) *адсорбцией*.

Энергия десорбции хемосорбированных молекул определяется суммарной высотой барьера, препятствующего выходу частиц из ямы II: $\mathcal{E}_{\text{дес}} = U_x + \mathcal{E}_{\text{акт}}$. По аналогии с (5.9.1), среднее время жизни молекулы в хемосорбированном состоянии определяем как

$$\tau_x = \tau_0 \exp\left(\frac{\mathcal{E}_{\text{дес}}}{k_B T}\right) \equiv \tau_0 \exp\left(\frac{Q_x + \mathcal{E}_{\text{акт}} \mathcal{N}_A}{RT}\right). \quad (5.9.2)$$

Поскольку силы межатомного взаимодействия в 5–10 раз превышают силы Ван-дер-Ваальса, то *теплота хемосорбции* $Q_x = U_x \mathcal{N}_A$ достигает 50–100 ккал/моль (около 2–4 эВ/частица). Это обеспечивает, согласно (5.9.2), времена жизни τ_x , измеряемые часами, сутками и годами в зависимости от T и $\mathcal{E}_{\text{дес}}$.

Приведенная на рис. 5.33 энергетическая диаграмма не является единственно возможной для процесса хемосорбции. Возможна ситуация, при которой максимум, разделяющий две потенциальные ямы I и II, оказывается расположенным ниже линии $U = 0$. Эта нулевая линия соответствует кинетической энергии газовой молекулы, приближающейся к поверхности адсорбента с бесконечно малой скоростью. В этом случае максимум (типа $\mathcal{E}_{\text{акт}}$ на рис. 5.33, но лежащий при $U < 0$) *не является* энергией активации. Такая хемосорбция называется *безактивационной*.

Два типа энергетических диаграмм, характерные для активационной и безактивационной хемосорбции, соответствуют *ковалентной связи* между атомами адсорбата и адсорбента, устанавливаемой путем обобществления пары валентных электронов.

Однако хемосорбция атомов щелочного металла (Na, K, Cs) на поверхности другого металла (W, Mo, Ta) реализуется по типу *ионной связи* путем полного перехода валентного электрона к атомам адсорбента. В этом случае энергетическая диаграмма соответствует той, которая характерна для физической адсорбции и показана на рис. 5.32. Подобное сходство породило другую терминологию в отношении физической и химической адсорбции.

Адсорбцию (как физическую, так и химическую), происходящую путем закрепления на адсорбенте частицы внешней фазы (молекулы или атома) без предварительной диссоциации ее на составляющие части, принято называть *недиссоциативной* адсорбцией. В остальных случаях хемосорбция является *диссоциативной*, при этом промежуточное (физадсорбционное) состояние молекулы, предшествующее хемосорбированию, называют *прекурсорным состоянием* (от англ. *precursor* «предшественник»).

Прекурсорное состояние может быть и в случае недиссоциативной адсорбции, если слабо связанная с адсорбционными центрами частица мигрирует по поверхности до тех пор, пока не находит положение с большей энергией связи. Такой вид адсорбции принято называть также *нелокализованной* адсорбцией, в отличие от *локализованной*, при которой частица сразу закрепляется на адсорбционном центре. Понятно, что число поступательных

степеней свободы газовой молекулы уменьшается на единицу при нелокализованной и на три — при локализованной адсорбции.

Среднее время жизни частицы в адсорбированном состоянии, введенное формулами (5.9.1) и (5.9.2), является одним из основных параметров, определяющих постоянную времени адсорбции.

Кинетической характеристикой процесса адсорбции служит степень заполнения адсорбционных центров, определяемая как $\theta(t) = N_a(t)/N_{\text{ц}}$, где $N_a(t)$ — концентрация занятых адсорбционных центров (т. е. адсорбированных частиц) в момент времени t , а $N_{\text{ц}}$ — общее число адсорбционных центров на единице площади поверхности. Кинетика заполнения адсорбционных центров молекулами газовой фазы, описываемая функцией $\theta(t)$, и стационарная степень заполнения $\theta_0 \equiv N_a^0/N_{\text{ц}}$ в форме изотермы Ленгмюра будут рассмотрены в следующем параграфе.

Энергетическая диаграмма на рис. 5.33 позволяет дать следующее объяснение температурного хода стационарной степени заполнения адсорбционных центров θ_0 , качественно показанного на рис. 5.34, который подтверждается экспериментом.

В области достаточно низких температур энергия тепловых колебаний физически адсорбированных молекул недостаточна для отрыва их от поверхности. В результате этого возникает адсорбционный монослой со степенью заполнения θ_0 , близкой к единице. С повышением температуры величина θ_0 уменьшается из-за увеличения вероятности выброса молекул из потенциальной ямы I наружу. По мере роста температуры вступает в действие конкурирующий процесс переброса молекулы через активационный барьер $\mathcal{E}_{\text{акт}}$ в яму II в форме двух атомов. Следовательно, при достаточно большой температуре процесс превращения физически адсорбированных молекул в хемосорбированные атомы начинает преобладать над процессом теплового выброса первых наружу из потенциальной ямы I. Это объясняет возрастание стационарной степени заполнения центров адсорбции, θ_0 , начинающийся на рис. 5.34 с некоторой температуры T_0 .

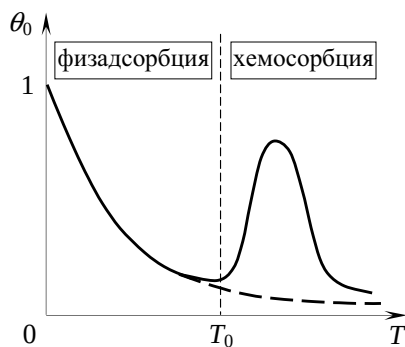


Рис. 5.34. Температурная зависимость стационарной степени заполнения адсорбционных центров на поверхности твердого тела

Однако при еще более высокой температуре хемосорбированные атомы приобретают настолько большую энергию тепловых колебаний, что выбрасываются наружу из потенциальной ямы II глубиной ($U_x + \mathcal{E}_{\text{акт}}$) и, соединившись в молекулы, испаряются с поверхности. В результате этого появляется хемосорбционный максимум на кривой рис. 5.34. Для физической адсорбции количественные соотношения, подтверждающие температурную зависимость $\theta_0(T)$, получены в следующем параграфе.

Термоактивационная природа хемосорбции, в отличие от физической адсорбции, делает процесс химического сорбирования необратимым, а именно: изменение температуры от T_1 до T_2 с последующим возвратом к исходной температуре T_1 не оставляет неизменным количество адсорбированного газа, что проявляется в форме гистерезиса.

Из-за существенного различия в энергии адсорбции хемосорбированные газы высвобождаются значительно труднее, чем физически адсорбированные инертные газы, и обычно выделяются в измененном виде. Например, кислород с вольфрама удаляется в виде WO_3 , а с графита — в виде CO и CO_2 , поскольку прочность связей в этих молекулах превышает силы когезии между атомами W или C . На многих металлах такие газы, как N_2 , O_2 , CO_2 , сорбируются сильнее, чем H_2 , так как имеют большую теплоту адсорбции. Это приводит к тому, что они сорбционно вытесняют водород с поверхности этих металлов. Все это усложняет процессы сорбции, так что теоретически полученные закономерности в большинстве своем носят качественный характер.

Многочисленные экспериментальные данные, собранные различными авторами, свидетельствуют о том, что наблюдаемые изотермы адсорбции отличаются большим разнообразием в зависимости от пар адсорбент–адсорбат и условий эксперимента. Характерные типы изотерм адсорбции приведены на рис. 5.35.

Понятно, что дать физическое описание таких сильно отличающихся закономерностей адсорбции, приведенных на рис. 5.35, в рамках какой-либо одной модели практически нереально. Разными авторами были предложены адсорбционные модели, среди которых наиболее разработанными являются следующие:

- **модель адсорбции Ленгмюра** при монослойной локализованной адсорбции ($\theta < 1$), дающая изотермы I типа;
- **модель двумерной конденсации** при монослойной нелокализованной адсорбции ($\theta < 1$) с латеральным (от англ. *lateral*) взаимодействием, дающая при температурах выше критической изотермы II типа;

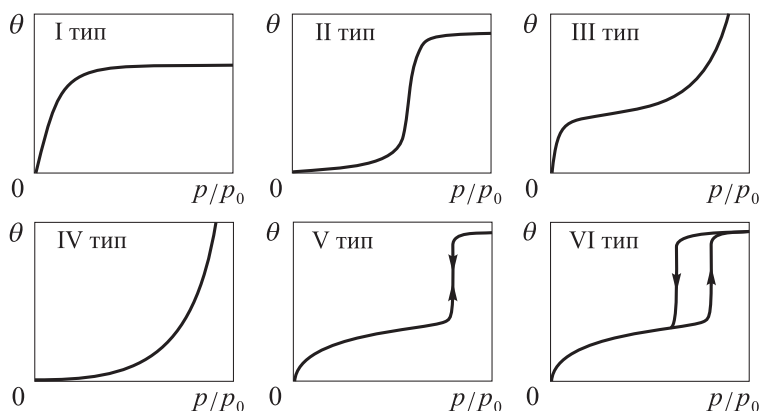


Рис. 5.35. Основные типы экспериментально наблюдаемых изотерм адсорбции (p_0 — давление насыщенного пара адсорбата в газовой фазе)

• **модель адсорбции Брунауэра–Эммета–Теллера (БЭТ)** при многослойной (полимолекулярной) локализованной адсорбции ($\theta > 1$), дающая изотермы III и IV типов;

• **модель капиллярной конденсации** в пористых адсорбентах, дающая в зависимости от формы и структуры пор изотермы либо V типа (без гистерезиса), либо VI типа (с гистерезисом).

Ниже дадим детальное рассмотрение особенностей каждого из четырех видов адсорбции.

5.10. Кинетика мономолекулярной адсорбции и двумерная конденсация

Как уже ранее было сказано, кинетической характеристикой процесса адсорбции является степень заполнения адсорбционных центров $\theta(t) = N_a(t)/N_{\text{ц}}$, меняющаяся во времени в нестационарном состоянии. Здесь $N_{\text{ц}}$ — общая концентрация адсорбционных центров, а $N_a(t)$ — концентрация адсорбционных центров, занятых адсорбированными частицами (молекулами или атомами) в момент времени t . В равновесном состоянии эти величины при необходимости будем отмечать индексами 0: $\theta_0 = N_a^0/N_{\text{ц}}$.

1. Модель адсорбции Ленгмюра. Эта первая и простейшая модель описания адсорбции была разработана еще в 1918 году. В основу ее положены следующие упрощающие предположения.

1. Поверхность адсорбента является *плоской и однородной*, т. е. однотипные адсорбционные центры расположены на поверхности с постоянной плотностью, и все они имеют одинаковую

энергию адсорбции/десорбции, что обеспечивает равновероятность их заполнения частицами адсорбента.

2. Каждый адсорбционный центр может удерживать только *одну частицу* (молекулу или атом) адсорбата, что по сути отражает существование силы отталкивания падающей молекулы от занятого центра, действующей лишь при непосредственном соприкосновении молекул.

3. Адсорбция происходит *локализованно* на незанятых участках поверхности, что вместе с предыдущим допущением равносильно утверждению о *монослойной адсорбции* ($\theta \leq 1$), и тем самым исключается возможность полимолекулярной адсорбции.

4. Притяжение между соседними адсорбированными частицами, называемое *латеральным взаимодействием*, отсутствует, что исключает возможность двумерной конденсации частиц адсорбированного газа при $\theta \rightarrow 1$.

Прежде чем рассматривать кинетику монослойного заполнения поверхности адсорбента, дадим определение скоростей адсорбции и десорбции частиц и установим их связь со степенью заполнения адсорбционных центров, называемой также степенью покрытия поверхности или просто покрытием.

Скорость адсорбции, вводимую как плотность потока адсорбируемых частиц, dN_a^+/dt , представляют в следующей форме:

$$v_{\text{адс}} \equiv \frac{dN_a^+}{dt} = \alpha F. \quad (5.10.1)$$

Здесь величина F представляет собой поток молекул, падающих на единицу площади поверхности из парогазовой фазы. Как известно из молекулярно-кинетической теории газов, при температуре T для молекул с массой m и парциальным давлением p плотность падающего потока равна ($\text{м}^{-2} \cdot \text{с}^{-1}$)

$$F = \frac{p}{\sqrt{2\pi m k_B T}}. \quad (5.10.2)$$

Множитель α в формуле (5.10.1), называемый *коэффициентом прилипания* (нередко обозначаемый как s от англ. *sticking*), учитывает вероятность того, что часть падающих молекул при ударе о поверхность отражается, а другая часть находит подходящие адсорбционные центры и закрепляется на них.

Коэффициент прилипания α , как функцию температуры T и степени покрытия θ , принято записывать следующим образом:

$$\alpha(T, \theta) = \bar{\alpha}(T) f_a(\theta), \quad \text{где} \quad \bar{\alpha}(T) = \alpha_0 \exp\left(\mp \frac{\mathcal{E}_{\text{акт}}}{k_B T}\right). \quad (5.10.3)$$

Температурная зависимость $\alpha(T)$ учитывается множителем $\bar{\alpha}(T) \equiv \alpha(T, 0)$ (ибо $f_a(\theta) = 1$ при $\theta = 0$, см. рис. 5.36), который содержит, во-первых, коэффициент конденсации α_0 , равный доле сконденсировавшихся на поверхности молекул, и во-вторых, экспоненту, отражающую температурную зависимость, присущую хемосорбции (активационной и безактивационной).

Верхний знак в показателе экспоненты (5.10.3) характерен для *активационной адсорбции* с энергией активации $\mathcal{E}_{\text{акт}}$, показанной на рис. 5.33. Нижний знак соответствует *безактивационной адсорбции* (не изображенной на этом рисунке), при которой максимум, разделяющий потенциальные ямы I и II, расположен ниже горизонтальной оси x и имеет значение $-\mathcal{E}_{\text{акт}} < 0$.

Следовательно, с ростом температуры коэффициент прилипания (5.10.3) увеличивается для активационной адсорбции и уменьшается для безактивационной адсорбции.

Зависимость коэффициента прилипания α в форме (5.10.3) от степени покрытия θ описывается при помощи функции $f_a(\theta)$, учитывающей вероятность для молекулы найти подходящее адсорбционное состояние, что определяется характером адсорбции.

В рамках модели Ленгмюра для диссоциативной адсорбции молекулы, состоящей из n фрагментов, требуется одновременное наличие n свободных адсорбционных центров, так что функция $f_a(\theta)$ в этом случае имеет следующий общий вид:

$$f_a(\theta) = (1 - \theta)^n, \quad (5.10.4)$$

где n называют *порядком кинетики адсорбции Ленгмюра*.

В простейших частных случаях функция (5.10.4) принимает следующие формы (см. рис. 5.36):

- *недиссоциативная адсорбция* ($n = 1$):

$$f_a(\theta) = 1 - \theta; \quad (5.10.5)$$

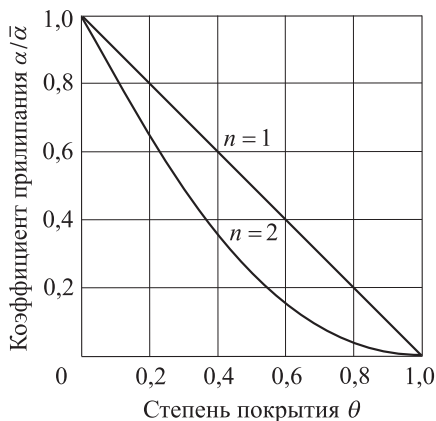


Рис. 5.36. Нормированный коэффициент прилипания $\alpha/\bar{\alpha}$ при недиссоциативной адсорбции ($n = 1$) и диссоциативной адсорбции ($n = 2$)

- *диссоциативная адсорбция* двухатомных молекул ($n = 2$):

$$f_a(\theta) = (1 - \theta)^2 \quad \text{— нелокализованная,} \quad (5.10.6)$$

$$f_a(\theta) = \frac{z}{z - \theta} (1 - \theta)^2 \quad \text{— локализованная,} \quad (5.10.7)$$

где z означает число ближайших соседних мест. Различие между выражениями (5.10.6) и (5.10.7) невелико, особенно при малой степени заполнения адсорбционных центров ($\theta \ll z$).

Нередко на экспериментальных зависимостях $\alpha(\theta)$ наблюдаются два участка, соответствующие кинетике второго и первого порядка. Это означает, что двухатомные молекулы сначала адсорбируются диссоциативно ($n = 2$), а затем в молекулярной форме ($n = 1$).

Скорость десорбции, вводимую как плотность потока десорбируемых частиц, dN_a^-/dt , представляют в следующей форме:

$$v_{\text{дес}} \equiv \frac{dN_a^-}{dt} = \frac{N_a}{\tau} f_d(\theta) \equiv \frac{N_{\text{ц}}}{\tau} \theta f_d(\theta), \quad (5.10.8)$$

где τ — среднее время пребывания частицы в адсорбированном состоянии ($\tau_{\text{ф}}$ или $\tau_{\text{х}}$), даваемое формулами (5.9.1) или (5.9.2). Согласно вышесказанному, для хемосорбции входящая в формулу (5.9.2) энергия десорбции равна (см. рис. 5.33)

$$\mathcal{E}_{\text{дес}} = \begin{cases} U_{\text{адс}} + \mathcal{E}_{\text{акт}} & \text{— активационная хемосорбция,} \\ U_{\text{адс}} - \mathcal{E}_{\text{акт}} & \text{— безактивационная хемосорбция.} \end{cases} \quad (5.10.9)$$

Здесь $U_{\text{адс}}$ — энергия адсорбционной связи, равная глубине потенциальной ямы $U_{\text{х}}$ на рис. 5.33; при этом для физадсорбции в формуле (5.9.1) всегда $\mathcal{E}_{\text{дес}} = U_{\text{адс}} \equiv U_{\text{ф}}$ (см. рис. 5.32).

Функция $f_d(\theta)$, входящая в выражение (5.10.8), учитывает зависимость скорости десорбции от степени покрытия и в простейших случаях принимает следующий вид:

$$f_d(\theta) = \begin{cases} 1 & \text{— неассоциативная десорбция,} \\ \theta & \text{— ассоциативная десорбция.} \end{cases} \quad (5.10.10)$$

Неассоциативная десорбция соответствует наиболее простому случаю кинетики первого порядка ($n = 1$). В этом случае отдельные адсорбированные частицы независимо друг от друга термически десорбируют из своих адсорбционных мест, имеющих концентрацию $N_a = \theta N_{\text{ц}}$, так что $v_{\text{дес}} \propto \theta$.

При *ассоциативной десорбции* имеет место кинетика второго порядка ($n = 2$). В этом случае десорбирующаяся молекула образуется из двух атомов, изначально занимающих отдельные адсорбционные центры, в результате чего $v_{\text{дес}} \propto \theta^2$.

Адсорбционное равновесие имеет *динамический характер*: числа адсорбируемых и десорбируемых частиц в единицу времени на единице площади в среднем равны друг другу в стационарных условиях.

Приравнивая скорость адсорбции (5.10.1) к скорости десорбции (5.10.8) с учетом (5.9.2) и (5.10.2)–(5.10.3), получаем

$$p \frac{\alpha_0 f_a(\theta)}{\sqrt{2\pi m k_B T}} \exp\left(\mp \frac{\mathcal{E}_{\text{акт}}}{k_B T}\right) = \frac{N_{\text{ц}}}{\tau_0} \theta f_d(\theta) \exp\left(-\frac{\mathcal{E}_{\text{дес}}}{k_B T}\right).$$

Отсюда, учитывая соотношения (5.10.9), находим в общей форме *уравнение изотермы адсорбции* (здесь $\theta \equiv \theta_0$ — равновесное покрытие):

$$p = \frac{1}{b(T)} \frac{\theta f_d(\theta)}{f_a(\theta)}, \quad (5.10.11)$$

при этом температурно-зависимая константа имеет вид

$$b(T) = \frac{\alpha_0}{N_{\text{ц}}} \frac{\tau_a(T)}{\sqrt{2\pi m k_B T}}, \quad \text{где } \tau_a(T) = \tau_0 e^{Q_a/RT}. \quad (5.10.12)$$

Выражение (5.10.12) для $\tau_a(T)$, являющееся одинаковым для всех видов адсорбции (физической, безактивационной химической и активационной химической), содержит соответствующую им теплоту адсорбции $Q_a = U_{\text{адс}} \mathcal{N}_d$, определяемую глубиной потенциальной ямы $U_{\text{адс}}$ (равной $U_{\text{ф}}$ или $U_{\text{х}}$ на рис. 5.32 или 5.33).

При использовании выражений (5.10.5)–(5.10.7) и (5.10.10) для $f_a(\theta)$ и $f_d(\theta)$ уравнение (5.10.11) устанавливает связь между давлением p газа при постоянной температуре T и равновесным покрытием θ_0 поверхности адсорбента.

2. Уравнения для изотермы Генри и изотермы Ленгмюра.

Сначала представим процесс адсорбции в форме квазихимической реакции заполнения адсорбционных центров $A_{\text{ц}}$ газовыми молекулами $M(\text{г})$, имеющими парциальное давление $p = ck_B T$:



Здесь $A_{\text{ц}}^*$ обозначает любой из заполненных адсорбционных центров с равновесной концентрацией N_a^0 , при этом концентрация свободных центров $A_{\text{ц}}$ равна $N_{\text{св}} = N_{\text{ц}} - N_a^0$.

Закон действующих масс для реакции (5.10.13), записанный в виде (2.12.6), имеет следующий вид (здесь $N_a \equiv N_a^0$ и $\theta \equiv \theta_0$ — равновесные величины, для которых верхний и нижний нулевые индексы ради упрощения записи опущены):

$$K_c(T) = \frac{N_a}{c N_{\text{св}}} = \frac{N_a}{c(N_{\text{ц}} - N_a)} = \frac{\theta}{c(1 - \theta)}. \quad (5.10.14)$$

В частном случае малого заполнения адсорбционных центров (при $\theta \ll 1$) из (5.10.14) следует **уравнение изотермы Генри**,

$$\theta(T, p) = K(T)p. \quad (5.10.15)$$

Согласно уравнению (5.10.15), степень заполнения адсорбционных центров прямо пропорциональна давлению газа, что аналогично закону Генри (2.6.4) для молекулярной растворимости газов в объеме материала. Следовательно, в рамках модели Генри любая соударяющаяся с поверхностью молекула адсорбируется на ней, независимо от наличия других молекул на поверхности.

В общем случае из закона действия масс (5.10.14) в рамках модели Ленгмюра получаем стационарную (при температуре T) степень заполнения адсорбционных центров в форме

$$\theta(T, p) = \frac{b(T)p}{1 + b(T)p}, \quad \text{или} \quad b(T)p(\theta) = \frac{\theta}{1 - \theta}, \quad (5.10.16)$$

называемой **уравнением изотермы Ленгмюра**, где введено обозначение $b(T) \equiv K_c(T)/k_B T$ и $\theta \equiv \theta_0$ — равновесное покрытие.

Из вышеприведенного вывода следует, что рассмотренная ситуация соответствует случаю кинетики первого порядка с недиссоциативной адсорбцией и неассоциативной десорбцией. В соответствии с выражениями (5.10.5) и (5.10.10), в общем уравнении изотермы адсорбции (5.10.11) полагаем $f_a = 1 - \theta$ и $f_d = 1$. Отсюда следует уравнение изотермы Ленгмюра в форме (5.10.16).

3. Уравнение кинетики недиссоциативной адсорбции в неизотермических условиях. Будем изучать кинетику мономолекулярной адсорбции, при которой $N_a(t) \leq N_{\text{ц}}$, так что $\theta(t) \leq 1$. Скорость изменения концентрации адсорбированных молекул равна разности между плотностями потока адсорбируемых частиц dN_a^+/dt и потока десорбируемых частиц dN_a^-/dt :

$$\frac{dN_a}{dt} = \frac{dN_a^+}{dt} - \frac{dN_a^-}{dt}. \quad (5.10.17)$$

В отличие от изотермической модели Ленгмюра, рассмотрим наиболее общий случай неизотермической адсорбции, когда температуры адсорбента (T_a) и газовой среды (T_0) различные. Именно такая ситуация реализуется в процессе реактивного распыления металлов в плазме инертного газа, кинетика которого была изучена в п. 4.16.

Следовательно, плотность (5.10.2) падающего потока газовых молекул, имеющих массу m и парциальное давление p , определяется температурой газовой среды T_0 , а именно

$$F = \frac{p}{\sqrt{2\pi mk_B T_0}}. \quad (5.10.18)$$

Часть этих молекул отражается от поверхности, а другая часть адсорбируется ею. Феноменологически это учитывается с помощью коэффициента прилипания $\alpha(T, \theta)$ в форме (5.10.3), характеризующего вероятность закрепления падающей молекулы на поверхности адсорбента.

Будем рассматривать недиссоциативную адсорбцию, тогда у коэффициента прилипания $\alpha(T, \theta)$ в форме (5.10.3) исчезает температурная зависимость.

Согласно (5.10.5), зависимость $\alpha(\theta)$ от степени заполнения адсорбционных центров принимает вид $\alpha(\theta) = \alpha_0(1 - \theta)$. Следовательно, $\alpha = \alpha_0 < 1$ для свободной поверхности (при $\theta = 0$) и $\alpha = 0$ для полностью заполненного монослоя (при $\theta = 1$). Тогда плотность потока адсорбируемых частиц равна

$$\frac{dN_a^+}{dt} = \alpha(\theta)F = \alpha_0 F(1 - \theta). \quad (5.10.19)$$

В рассматриваемой неизотермической ситуации кинетическая энергия падающих газовых молекул отличается от энергии тепловых колебаний адсорбированных частиц, находящихся в тепловом равновесии с адсорбентом при температуре T_a . Как правило (за исключением очень легких молекул с энергией, соответствующей температуре свыше 6000 К), падающая молекула после закрепления на поверхности адсорбента приходит с ним в энергетическое равновесие в течение нескольких периодов тепловых колебаний решетки ($\tau_0 \approx 10^{-13}$ с).

Таким образом, вышеописанный процесс, называемый *термической аккомодацией*, можно считать протекающим практически мгновенно. Это позволяет считать температуру десорбируемых частиц равной температуре адсорбента T_a . Именно она и определяет среднее время пребывания частицы в адсорбированном состоянии, даваемое формулой (5.10.12):

$$\tau_a = \tau_0 \exp\left(\frac{Q_a}{RT_a}\right). \quad (5.10.20)$$

Число молекул, десорбируемых с единицы площади в единицу времени (т. е. плотность потока десорбции dN_a^-/dt), прямо пропорционально концентрации молекул $N_a(t) = N_{\text{ц}}\theta(t)$, адсорбированных в данный момент времени на поверхности, и обратно пропорционально среднему времени τ_a их жизни в адсорбированном состоянии, даваемому формулой (5.10.20):

$$\frac{dN_a^-}{dt} = \frac{N_{\text{ц}}\theta}{\tau_a}. \quad (5.10.21)$$

Подстановка выражений (5.10.19) и (5.10.21) в равенство (5.10.17) дает следующее уравнение для нахождения $\theta(t)$:

$$\frac{d\theta(t)}{dt} = -\left(\frac{1}{\tau_a} + \frac{\alpha_0 F}{N_{\text{ц}}}\right)\theta(t) + \frac{\alpha_0 F}{N_{\text{ц}}}. \quad (5.10.22)$$

Использование выражения (5.10.18) для падающего потока F позволяет привести уравнение (5.10.22) к окончательной форме

$$\frac{d}{dt}\left(\theta(t) - \frac{\tau_{\text{эф}}}{\tau_p}\right) = -\frac{1}{\tau_{\text{эф}}}\left(\theta(t) - \frac{\tau_{\text{эф}}}{\tau_p}\right), \quad (5.10.23)$$

которая имеет привычный вид релаксационного уравнения, описывающего кинетику процесса недиссоциативной адсорбции.

Уравнение (5.10.23) содержит два характерных времени:

а) постоянную времени τ_p , зависящую от давления p в форме

$$\tau_p \equiv \frac{N_{\text{ц}}}{\alpha_0 F(T_0)} = \frac{N_{\text{ц}}}{\alpha_0} \frac{\sqrt{2\pi m k_B T_0}}{p} \equiv \frac{\tau_a}{b(T_a, T_0)p}, \quad (5.10.24)$$

где с учетом формулы (5.10.20) введена температурно-зависимая константа

$$b(T_a, T_0) = \frac{\alpha_0}{N_{\text{ц}}} \frac{\tau_a(T_a)}{\sqrt{2\pi m k_B T_0}} = \frac{\alpha_0}{N_{\text{ц}}} \frac{\tau_0 \exp(Q_a/RT_a)}{\sqrt{2\pi m k_B T_0}}; \quad (5.10.25)$$

б) эффективное время релаксации для процесса недиссоциативной адсорбции,

$$\tau_{\text{эф}} = \left(\frac{1}{\tau_a} + \frac{1}{\tau_p}\right)^{-1} = \frac{\tau_a}{1 + bp}. \quad (5.10.26)$$

Решение кинетического уравнения (5.10.23) при начальном условии $\theta(0) = 0$ имеет следующий вид:

$$\theta(t) = \theta_0 \left(1 - e^{-t/\tau_{\text{эф}}}\right). \quad (5.10.27)$$

Стационарная степень заполнения θ_0 , отмечаемая индексом 0 и достигаемая при $t \rightarrow \infty$, дается выражением

$$\theta_0(T_a, T_0, p) \equiv \frac{\tau_{\text{эф}}}{\tau_p} = \frac{b(T_a, T_0)p}{1 + b(T_a, T_0)p}, \quad (5.10.28)$$

которое является функцией двух температур T_a, T_0 и давления p .

В изотермическом случае при $T_a = T_0 \equiv T$ формула (5.10.25) для $b(T_a, T_0)$ принимает форму (5.10.12) для $b(T)$. При этом

выражение (5.10.28) превращается в обычное *уравнение изотермы Ленгмюра*, даваемое формулой (5.10.16).

Зависимость постоянных времени, рассчитанных по формулам (5.10.24) и (5.10.26) в нормированном виде τ_p/τ_a и $\tau_{эф}/\tau_a$, от параметра bp показана на рис. 5.37. Отсюда очевидно, что с ростом давления обе постоянные уменьшаются. Влияние же температуры проявляется более сложным образом, поскольку нормировочная величина τ_a в виде (5.10.20) при $T_a = T$ сама зависит от температуры. Согласно выражению (5.10.24), возрастающая зависимость $\tau_p(T) \propto (k_B T)^{1/2}$

незначительна во всем температурном диапазоне. Как видно из (5.10.26), для эффективного времени релаксации ситуация иная: а) $\tau_{эф}(T) \approx \tau_p(T) \propto (k_B T)^{1/2}$ слабо возрастает в области низких температур, где $\tau_p \ll \tau_a$; б) $\tau_{эф}(T) \approx \tau_a(T) \propto \exp(Q_a/RT)$ экспоненциально убывает при высоких температурах, где $\tau_a \ll \tau_p$.

Зависимость θ_0 от параметра bp , рассчитанная по формуле (5.10.16), приведена на рис. 5.37. Таким образом, с увеличением давления p в газовой фазе стационарная степень заполнения θ_0 возрастает, стремясь к единице. В то же время эффективное время релаксации $\tau_{эф}$ уменьшается от значения τ_a , приближаясь к нулю. Следовательно, при более высоких давлениях адсорбционные центры заполняются до больших значений θ_0 за более короткое время. Именно это и отображает рис. 5.38, на котором показано

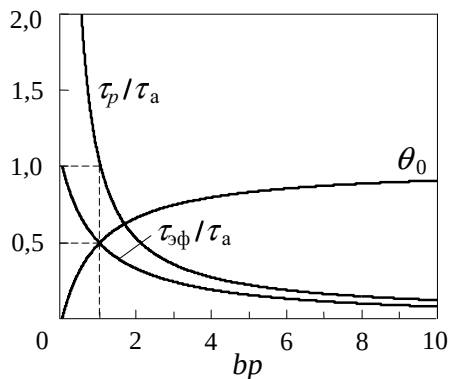


Рис. 5.37. Зависимости стационарной степени заполнения адсорбционных центров θ_0 и нормированных значений постоянной времени τ_p/τ_a и эффективного времени релаксации $\tau_{эф}/\tau_a$ от параметра bp

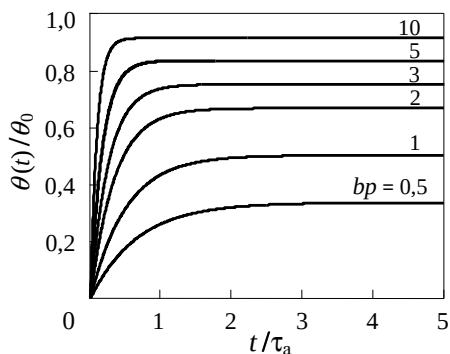


Рис. 5.38. Временная зависимость нестационарной степени заполнения адсорбционных центров $\theta(t)$ при различных значениях параметра bp

изменение во времени нестационарной степени заполнения адсорбционных центров в нормированном виде $\theta(t)/\theta_0$, рассчитанное по формулам (5.10.26) и (5.10.27) при разных значениях параметра bp . Из кривых рис. 5.38 видно, что с увеличением давления адсорбционные центры быстрее заполняются до максимально возможного значения $\theta_0(p)$, которое также возрастает.

В соответствии с изотермой Ленгмюра (5.10.16), стационарная зависимость $\theta_0(T)$ определяется температурной зависимостью параметра $b(T)$, введенного формулой (5.10.12). Величина $b(T) \propto \tau_a(T)$ и с повышением температуры уменьшается от бесконечно больших значений до нуля. Тогда соотношение (5.10.16) при $p = \text{const}$ обеспечивает экспоненциальное уменьшение θ_0 от единицы (при $T = 0$ К) до нуля, наблюдаемое в эксперименте и показанное на рис. 5.34. Физическая трактовка этого была дана в предыдущем параграфе, включая качественное объяснение хемосорбционного максимума.

Таким образом, теория Ленгмюра описывает *мономолекулярную локализованную адсорбцию* на однородной поверхности без учета латерального взаимодействия адсорбат–адсорбат. Из сравнения зависимости θ_0 от параметра bp , рассчитанной по формуле Ленгмюра (5.10.16) и изображенной на рис. 5.37, с экспериментально наблюдаемыми кривыми на рис. 5.35 следует, что модели Ленгмюра соответствует I тип изотермы адсорбции.

4. Латеральное взаимодействие и двумерная конденсация. По мере монослойного заполнения поверхности молекулами адсорбата расстояние между ближайшими соседями сокращается. Это приводит к поверхностному (латеральному) взаимодействию между молекулами адсорбата. Для учета такого взаимодействия необходимо модифицировать модель Ленгмюра, допуская, подобно молекулам в газе, тепловое хаотическое движение молекул вдоль поверхности, т. е. *нелокализованную адсорбцию*.

Простейший способ описания взаимодействий между молекулами в неидеальных газах был предложен Ван-дер-Ваальсом¹⁾. Для этого уравнение состояния идеального газа в форме уравнения Клапейрона–Менделеева, записанное для одного моля как

$$p = RT/v, \quad (5.10.29)$$

¹⁾ Ван-дер-Ваальс (Van-der-Waals) Ян Дидерик (1837–1923) — голландский физик-теоретик, известен работами в области молекулярной физики, исследовал поведение молекул в газах и жидкостях, открыл силы взаимодействия между молекулами (силы Ван-дер-Ваальса), с учетом которых вывел уравнение состояния реальных газов (уравнение Ван-дер-Ваальса). Нобелевская премия по физике (1910) «за работу над уравнением состояния газов и жидкостей».

было представлено для реальных газов в следующей форме:

$$p = \frac{RT}{v - b} - \frac{a}{v^2}. \quad (5.10.30)$$

Поскольку модель идеального газа пренебрегает как межмолекулярным взаимодействием, так и конечными размерами молекул (представляемых материальными точками), то уравнение (5.10.30), предложенное Ван-дер-Ваальсом, с помощью феноменологически введенных параметров $a > 0$ и $b > 0$ учитывает два физических явления, характерных для неидеальных газов:

во-первых, притяжение между молекулами, действие которого понижает давление, оказываемое молекулами на стенки сосуда, на величину $\Delta p = a/v^2$;

во-вторых, уменьшение объема $v = V/n$, занимаемого одним молекул, на величину b , равную суммарному объему, занимаемому всеми реальными молекулами в одном моле (числом N_A).

Исходя из уравнения Ван-дер-Ваальса для трехмерного газового состояния, представленного в форме (5.10.30), можно записать аналогичное уравнение для двумерного состояния нелокализованного адсорбционного монослоя.

Уравнение состояния идеального двумерного газа, записанное для одной молекулы, имеет вид (5.6.6), а именно

$$\pi = k_B T / \omega, \quad (5.10.31)$$

где $\omega = 1/N_a^0$ — площадь, занимаемая одной адсорбированной молекулой при стационарной концентрации заполненных центров, равной $N_a^0 = \theta_0 N_{\text{ц}}$ (оба нулевые индексы будут ниже опущены).

Поверхностное давление двумерного газа внутри монослоя определено в виде (в единицах Н/м)

$$\pi = \sigma_0 - \sigma, \quad (5.10.32)$$

где σ_0 и σ — поверхностные натяжения чистого адсорбента и адсорбента, покрытого монослоем адсорбата.

На основании сравнения формул (5.10.29) и (5.10.31) записываем уравнение состояния неидеального двумерного газа по аналогии с уравнением Ван-дер-Ваальса (5.10.30):

$$\pi = \frac{k_B T}{\omega - b_s} - \frac{a_s}{\omega^2}. \quad (5.10.33)$$

Величина a_s (в единицах Дж·м²), как и величина a в трехмерном случае, учитывает взаимодействие между молекулами адсорбата. Величина b_s (в единицах м²) представляет собой площадь, занимаемую молекулой в плотном монослое, когда все адсорбционные центры заполнены, т. е. при $\theta = 1$, так что $b_s = 1/N_{\text{ц}}$.

Для дальнейшего преобразования уравнения (5.10.33) воспользуемся уравнением адсорбции Гиббса (5.6.2), в котором $\Gamma_i = N_a \equiv 1/\omega$. В изотермических условиях, с учетом того, что из (5.10.32) следует $d\pi = -d\sigma$, переписываем уравнение (5.6.2) в следующем виде (опуская далее индекс i у $p_i = c_i k_B T$):

$$\omega = \frac{1}{c_i} \frac{\partial p_i}{\partial \pi} \equiv k_B T \frac{\partial \ln p}{\partial \pi}. \quad (5.10.34)$$

Интегрируя уравнение (5.10.34) при $T = \text{const}$, приходим к выражению

$$\ln \frac{p}{\hat{p}} = \frac{1}{k_B T} \int \omega d\pi, \quad (5.10.35)$$

где использована константа интегрирования $C(T) = \ln \hat{p}(T)$.

Из уравнения (5.10.33) находим дифференциал поверхностного давления,

$$d\pi = -k_B T \frac{d\omega}{(\omega - b_s)^2} + 2a_s \frac{d\omega}{\omega^3},$$

подстановкой которого в (5.10.35) после интегрирования получаем

$$\begin{aligned} \ln \frac{p}{\hat{p}} &= -\ln \left(\frac{\omega}{b_s} - 1 \right) + \frac{1}{\omega/b_s - 1} - \frac{2a_s/b_s}{k_B T} \frac{b_s}{\omega} = \\ &= \ln \left(\frac{\theta}{1 - \theta} \right) + \frac{\theta}{1 - \theta} - \frac{2a_s/b_s}{k_B T} \theta. \end{aligned} \quad (5.10.36)$$

В последнем равенстве сделана подстановка $b_s/\omega = N_a/N_u \equiv \theta$.

Из выражения (5.10.36) следует искомое **уравнение изотермы нелокализованной адсорбции**,

$$p(\theta) = \hat{p}(T) \left(\frac{\theta}{1 - \theta} \right) \exp \left[\left(\frac{\theta}{1 - \theta} \right) - \frac{2a_s/b_s}{k_B T} \theta \right], \quad (5.10.37)$$

учитывающее *латеральное взаимодействие* адсорбат–адсорбат.

Подстановкой $\omega = b_s/\theta$ в формулу (5.10.33) получаем уравнение для поверхностного давления двумерного газа адсорбата,

$$\pi(\theta) = \hat{\pi}(T) \left[\left(\frac{\theta}{1 - \theta} \right) - \frac{a_s/b_s}{k_B T} \theta^2 \right], \quad (5.10.38)$$

где введено обозначение $\hat{\pi}(T) = k_B T/b_s \equiv N_u k_B T$.

В сравнении с уравнением Ленгмюра (5.10.16) полученное уравнение (5.10.37) имеет два дополнительных экспоненциальных множителя: один отвечает за энергетику латерального взаимодействия (параметр a_s), а другой обусловлен исключением площади, занятой молекулами адсорбата, из общей площади

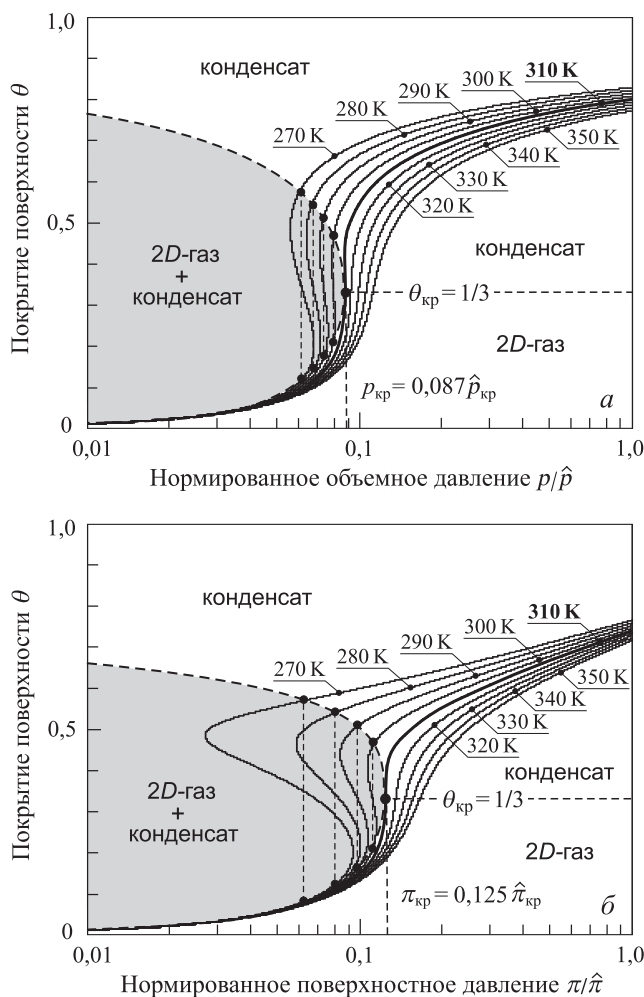


Рис. 5.39. Демонстрация двумерной конденсации в адсорбционном монослое: *a* — изотермы нелокализованной адсорбции с учетом взаимодействия молекул адсорбата, рассчитанные по уравнению (5.10.37) с параметром $a_s/b_s = 90$ мэВ; *б* — кривые, рассчитанные по формуле (5.10.38) и показывающие зависимость поверхностного давления π от степени заполнения адсорбционных центров θ

адсорбента (параметр $b_s \equiv 1/N_{ц}$). Действие обоих этих факторов порождает **двумерную конденсацию** — фазовый переход первого рода от двумерного газа (2D-газа) к двумерной конденсированной фазе адсорбата. Явление двумерной конденсации во многом подобно конденсации газа в трехмерном случае, описываемом уравнением Ван-дер-Ваальса (5.10.30).

Изотермы адсорбции, построенные по уравнению (5.10.37), приведены на рис. 5.39 а в безразмерных координатах $\theta - p/\hat{p}$, где $\hat{p}(T)$ — феноменологическая температурно-зависимая константа. На этом рисунке область сосуществования фаз (в которой $2D$ -газ находится в равновесии с двумерной конденсированной фазой) показана затемненной. Внутри такой области изотермы имеют участок с отрицательным наклоном, который не реализуется на практике, как и при трехмерной конденсации, из-за неустойчивости системы. Поэтому реальному двухфазному состоянию адсорбционного монослоя соответствуют вертикальные пунктирные прямые, соединяющие точки верхней и нижней границ затемненной области, которые показаны штриховой линией. Каждая пунктирная прямая (проведенная так, чтобы были равными ограниченные изотермой площади слева и справа от нее) по сути своей выражает давление насыщенного пара p над поверхностью двумерного конденсата при данной температуре T .

Аналогичным образом выглядят кривые, построенные по формуле (5.10.38) и приведенные на рис. 5.39 б. Здесь в безразмерных координатах $\theta - \pi/\hat{\pi}$ (где $\hat{\pi} = N_{\text{ц}}k_{\text{B}}T$) показана зависимость поверхностного давления π от степени покрытия θ . Как и на рис. 5.39 а, вертикальные пунктирные прямые в затемненной области характеризуют давление π двумерного газа ($2D$ -газа), равновесного с двумерным конденсатом при данной температуре T .

Точка смыкания верхней и нижней границы затемненной области соответствует *критическому состоянию* системы. Изотерма, проходящая через эту точку и соответствующая критической температуре $T_{\text{кр}}$, имеет здесь вертикальную касательную и точку перегиба. На рис. 5.39 такая изотерма изображена толстой линией и выделена полужирным шрифтом для $T_{\text{кр}} = 310$ К (что соответствует $a_s/b_s = 90$ мэВ). При температурах $T < T_{\text{кр}}$ малые покрытия (ниже нижней границы затемненной области) соответствуют $2D$ -газу, а большие покрытия (выше верхней границы затемненной области) — двумерному конденсату. При температурах $T > T_{\text{кр}}$, как и в трехмерном случае, различие между двумерным газом и его конденсатом исчезает.

Найдем критические параметры двумерного адсорбционного монослоя, используя уравнения (5.10.37) и (5.10.38). Как уже отмечалось, изотерма, записанная как функция $p = F_1(\theta)$ в форме (5.10.37), имеет в критической точке горизонтальную касательную и перегиб кривой, чему соответствуют равенства

$$\left(\frac{\partial p}{\partial \theta}\right)_T = 0 \quad \text{и} \quad \left(\frac{\partial^2 p}{\partial \theta^2}\right)_T = 0. \quad (5.10.39)$$

Легко убедиться в том, что первое равенство (5.10.39), определяющее положение точек экстремума (минимума и максимума) кривой $p = F_1(\theta)$, реализуется при

$$(1 - \theta)^2 = \frac{1}{\alpha\theta}, \quad \text{где} \quad \alpha = \frac{2a_s/b_s}{k_B T}. \quad (5.10.40)$$

Равенство (5.10.40) обеспечивает также требование $\partial\pi/\partial\theta = 0$. Это означает, что положения экстремальных точек $\theta_{\min}$ и $\theta_{\max}$ на кривых $p = F_1(\theta)$ (рис. 5.39 а) и $\pi = F_2(\theta)$ (рис. 5.39 б) при одной и той же температуре совпадают, как и должно быть.

Графическое решение уравнения (5.10.40) качественно показано на рис. 5.40, при этом $f_1(\theta) = (1 - \theta)^2$ и $f_2(\theta) = 1/\alpha\theta$. При $T < T_{\text{кр}}$ кривая $f_2(\theta)$ имеет положение 1, так что точки пересечения с кривой $f_1(\theta)$ суть точки экстремума $\theta_{\min}$ и $\theta_{\max}$. При $T > T_{\text{кр}}$ кривая $f_2(\theta)$ занимает положение 3, когда отсутствуют вещественные корни уравнения (5.10.40). Понятно, что промежуточный случай 2, когда кривые $f_1(\theta)$ и $f_2(\theta)$ касаются друг друга, соответствует *критическому режиму*. Поэтому вместо решения второго уравнения (5.10.39) необходимо обеспечить равенство производных от функций $f_1(\theta)$ и $f_2(\theta)$:

$$\frac{\partial}{\partial\theta}(1 - \theta)^2 = \frac{\partial}{\partial\theta} \frac{1}{\alpha\theta}, \quad \text{откуда} \quad 2(1 - \theta) = \frac{1}{\alpha\theta^2}.$$

Подставляя сюда величину $1/\alpha = \theta(1 - \theta)^2$, взятую из уравнения (5.10.40), получаем *критическую степень покрытия*:

$$\theta_{\text{кр}} = \frac{1}{3}. \quad (5.10.41)$$

Отсюда, с учетом обозначения (5.10.40) для α , получаем

$$\alpha_{\text{кр}} = \frac{1}{\theta_{\text{кр}}(1 - \theta_{\text{кр}})^2} = \frac{27}{4}, \quad (5.10.42)$$

$$k_B T_{\text{кр}} = \frac{2a/b}{\alpha_{\text{кр}}} = \frac{8}{27} \frac{a_s}{b_s}. \quad (5.10.43)$$

При величине $a_s/b_s = 90$ мЭВ, использованной ранее в расчетах, формула (5.10.43) дает значение критической температуры $T_{\text{кр}} \approx 310$ К, приведенное на рис. 5.39.

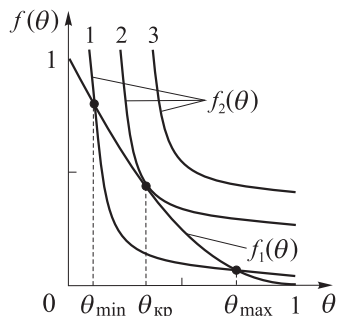


Рис. 5.40. Графическое нахождение критических параметров

Подстановка критических параметров (5.10.41)–(5.10.43) в формулы (5.10.37) и (5.10.38) дает, соответственно, *критическое объемное давление* в газовой фазе (при этом $\hat{p}_{\text{кр}} = \hat{p}(T_{\text{кр}})$),

$$p_{\text{кр}} = \hat{p}_{\text{кр}} \left(\frac{\theta_{\text{кр}}}{1 - \theta_{\text{кр}}} \right) \exp \left[\left(\frac{\theta_{\text{кр}}}{1 - \theta_{\text{кр}}} \right) - \alpha_{\text{кр}} \theta_{\text{кр}} \right] = \\ = \frac{\hat{p}_{\text{кр}}}{2} e^{-7/4} \approx 0,087 \hat{p}_{\text{кр}}, \quad (5.10.44)$$

и *критическое поверхностное давление* двумерного газа в адсорбционном монослое (при этом $\hat{\pi}_{\text{кр}} = N_{\text{ц}} k_B T_{\text{кр}}$),

$$\pi_{\text{кр}} = \hat{\pi}_{\text{кр}} \left[\left(\frac{\theta_{\text{кр}}}{1 - \theta_{\text{кр}}} \right) - \frac{\alpha_{\text{кр}}}{2} \theta_{\text{кр}}^2 \right] = \frac{\hat{\pi}_{\text{кр}}}{8} \equiv \frac{N_{\text{ц}}}{27} \frac{a_s}{b_s}. \quad (5.10.45)$$

Найденные критические параметры (5.10.41)–(5.10.45) являются универсальными, в частности $\theta_{\text{кр}} = 1/3$ и $\alpha_{\text{кр}} = 27/4$. Все они применимы в рамках ван-дер-ваальсовой модели латерального взаимодействия к различным случаям *мономолекулярной не локализованной адсорбции*. Специфику взаимодействий учитывают феноменологические константы: температурная константа $\hat{p}(T)$ для взаимодействия адсорбат–адсорбент и энергетическая константа a_s/b_s для взаимодействия адсорбат–адсорбат.

Сравнение изотерм на рис. 5.39 а, соответствующих температурам $T > T_{\text{кр}}$, с экспериментально наблюдаемыми кривыми на рис. 5.35 подтверждает, что в рамках модели двумерной конденсации реализуется, в частности, II тип изотермы адсорбции. В зависимости от соотношения констант a_s/b_s и $\hat{p}(T)$ форма изотермы (5.10.37) сильно изменяется от вогнутой во всем диапазоне давлений до S-образной с вертикальным разрывом при температурах $T < T_{\text{кр}}$, что указывает на двумерную конденсацию.

5. Латеральное взаимодействие при локализованной монослойной адсорбции. К ситуации с локализованной адсорбцией, когда частицы прочно удерживаются на адсорбционных центрах, например силами химической связи, нельзя применять использованную выше концепцию двумерного газа и конденсата.

В этом случае латеральное (вдоль поверхности) взаимодействие между частицами рассматривают в рамках квазихимического описания адсорбции с помощью реакции (5.10.13). Учет притяжения адсорбат–адсорбат выполняется путем замены концентрации N_a адсорбированных частиц на их активность $a = \gamma_a N_a$. Коэффициент активности принято феноменологически принимать зависящим от степени θ покрытия поверхности в виде $\gamma_a = \exp(-k\theta)$.

Закон действующих масс для реакции (5.10.13) принимает следующий вид (ср. уравнение (5.10.14)):

$$K_c(T) = \frac{\gamma_a N_a}{c N_{св}} = \frac{\theta e^{-k\theta}}{c(1-\theta)}. \quad (5.10.46)$$

Из формулы (5.10.46), с учетом соотношения $p = ck_B T$, вытекает **уравнение изотермы локализованной адсорбции** с латеральным взаимодействием адсорбат–адсорбат:

$$p(\theta) = \frac{\theta}{K(1-\theta)e^{k\theta}} \approx \frac{\theta}{K(1-\theta)(1+k\theta)}, \quad (5.10.47)$$

где $K \equiv K_c/k_B T$ и приближенное равенство верно при $k\theta \ll 1$.

Уравнение (5.10.47) по сравнению с уравнением Ленгмюра (5.10.16) хорошо описывает изотермы адсорбции (наблюдаемые на поверхности графитированной сажи, но не показанные на рис. 5.35), которые имеют в области малых θ (где выполняется линейный закон Генри) выпуклость, обращенную к оси p , и последующую точку перегиба с переходом к изотерме Ленгмюра.

При помощи двух уравнений адсорбции, учитывающих латеральное притяжение адсорбат–адсорбат: (5.10.37) для нелокализованной адсорбции и (5.10.47) для локализованной адсорбции, можно приближенно описать разнообразные изотермы мономолекулярной адсорбции на однородных (непористых) поверхностях, наблюдаемые в условиях эксперимента.

5.11. Полимолекулярная адсорбция и капиллярная конденсация

Рассмотренная в предыдущем параграфе локализованная монослойная адсорбция с насыщением типа изотермы Ленгмюра надежно реализуется при химической адсорбции. В этом случае сильное взаимодействие с поверхностью адсорбента, во-первых, прочно удерживает частицы на центрах адсорбции и, во-вторых, исключает образование второго слоя в силу различия в свойствах поверхностных атомов адсорбента (по отношению к частицам первого монослоя) и атомов хемосорбционного покрытия (по отношению к частицам последующего второго слоя).

При физической адсорбции молекул, удерживаемых силами Ван-дер-Ваальса (преимущественно дисперсионного типа), нет большого различия в энергиях связи первого монослоя с подложкой и последующих слоев друг с другом. Поэтому часто наблюдается многослойное (полимолекулярное) заполнение поверхности адсорбента молекулами адсорбата.

1. Модель адсорбции БЭТ. Свойство многослойности адсорбции заложено в основу теории Брунауэра–Эммета–Теллера, являющейся развитием теории Ленгмюра. Модель БЭТ снимает лишь требование адсорбции на свободных местах однородной поверхности, оставляя в силе другие допущения Ленгмюра, в частности отсутствие латерального взаимодействия адсорбат–адсорбат. Особенность многослойной модели БЭТ состоит в том, что процесс протекает не путем поочередного заполнения адсорбционных слоев, а допускается закрепление молекул адсорбата не только на поверхности адсорбента, но и на ранее адсорбированных молекулах первого, второго и последующих слоев.

Строго говоря, теплота адсорбции, будучи наибольшей для первого слоя, с ростом номера слоя l снижается ($Q_1 > Q_2 > \dots$) и при $l \rightarrow \infty$ приближается к теплоте Q_L конденсации адсорбата. Упрощающее предположение теории БЭТ состоит в том, что во всех слоях выше первого теплота адсорбции принимается одинаковой и равной теплоте конденсации ($Q_2 = Q_3 = \dots = Q_L$), т. е. учтены только два крайних значения теплоты: Q_1 и Q_L .

Отсюда следует, что среднее время пребывания молекулы в адсорбированном состоянии, даваемое формулой (5.10.20), для первого слоя (τ_1) больше, чем для остальных слоев ($\tau_l < \tau_1$), где оно одинаково, а именно

$$\tau_1 = \tau_0 e^{Q_1/RT} \quad \text{и} \quad \tau_l = \tau_0 e^{Q_L/RT}, \quad l = 2, 3, \dots \quad (5.11.1)$$

Таким образом, на поверхности адсорбента образуются вертикально расположенные молекулярные комплексы, состоящие из одной, двух, трех и т. д. молекул адсорбата, как показано на рис. 5.41. Кроме того, имеются свободные места, незанятые молекулами, доля поверхности которых равна

$$\theta_0 = 1 - (\theta_1 + \theta_2 + \theta_3 + \dots), \quad (5.11.2)$$

при этом $\theta_1, \theta_2, \theta_3, \dots$ — доли поверхности, покрытые единичными, двойными, тройными и т. д. вертикальными столбиками.

С учетом кратности каждого молекулярного комплекса общая степень покрытия, определяемая как $\theta = N_a/N_{\text{ц}}$, равна

$$\theta = \theta_1 + 2\theta_2 + 3\theta_3 + \dots \quad (5.11.3)$$

Для вывода уравнения изотермы БЭТ воспользуемся кинетическим подходом, примененным в теории Ленгмюра (см. п. 5.10). С этой целью для каждого l -го слоя, характеризуемого коэффициентом прилипания α_l , вводим скорости адсорбции (dN_l^+/dt) и десорбции (dN_l^-/dt), записанные по аналогии с выражениями (5.10.19) и (5.10.21).

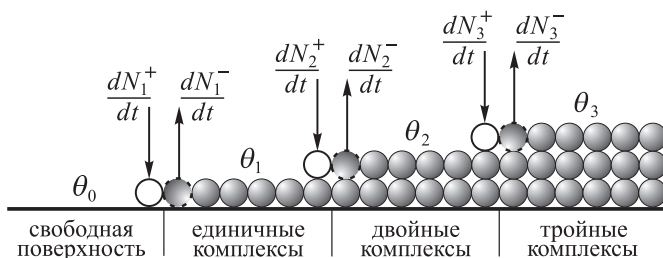


Рис. 5.41. Схема заполнения слоев в модели БЭТ вертикальными молекулярными комплексами, состоящими из одной, двух, трех и т. д. молекул и имеющими покрытия θ_1 , θ_2 , θ_3 и т. д., при этом свободной поверхности соответствует θ_0

На рис. 5.41 направления адсорбции и десорбции для трех слоев показаны встречными стрелками, при этом адсорбируемая молекула изображена белой, а десорбируемая выделена пунктирным контуром. Используя формулы (5.10.19) и (5.10.21), запишем скорости адсорбции и десорбции молекул адсорбата

- для первого слоя ($l = 1$) на поверхности адсорбента:

$$\frac{dN_1^+}{dt} = \alpha_1 \theta_0 F \quad \text{и} \quad \frac{dN_1^-}{dt} = \frac{N_u \theta_1}{\tau_1}; \quad (5.11.4)$$

- для l -го слоя на поверхности предыдущего слоя ($l = 2, 3, \dots$):

$$\frac{dN_l^+}{dt} = \alpha_l \theta_{l-1} F \quad \text{и} \quad \frac{dN_l^-}{dt} = \frac{N_u \theta_l}{\tau_l}. \quad (5.11.5)$$

Здесь плотность F падающего потока молекул адсорбата определяется их парциальным давлением p в газовой фазе ($F \propto p$) и дается формулой (5.10.2).

Поскольку в стационарном (равновесном) состоянии скорости адсорбции и десорбции равны, то приравниванием их в формулах (5.11.4)–(5.11.5) получаем отношения покрытий в соседних слоях:

$$\begin{aligned} \theta_1/\theta_0 &= \alpha_1 \tau_1 \frac{F}{N_u} \equiv K_1(T)p, \\ \theta_2/\theta_1 &= \alpha_2 \tau_2 \frac{F}{N_u} \equiv K_2(T)p, \\ \theta_3/\theta_2 &= \alpha_3 \tau_3 \frac{F}{N_u} \equiv K_3(T)p \quad \text{и т. д.} \end{aligned} \quad (5.11.6)$$

Принятая в модели БЭТ идентичность адсорбционных слоев, начиная со второго, позволяет, с учетом выражений (5.11.1), записать температурную зависимость констант, введенных в равенствах (5.11.6), следующим образом:

$$K_1(T) = K_1^\circ e^{Q_1/RT}, \quad (5.11.7)$$

$$K_2(T) = K_3(T) = \dots = K_L(T) = K_L^\circ e^{Q_L/RT}. \quad (5.11.8)$$

На поверхности последнего слоя (в пределе при $l \rightarrow \infty$) имеет место обратимый процесс конденсация $\rightleftharpoons$ испарения. Запишем этот процесс в виде квазихимической реакции совместно с соответствующим ей законом действующих масс:

$$V(\text{п}) \rightleftharpoons L(\text{ж}) \quad \text{и} \quad K_L(T) = 1/p_0. \quad (5.11.9)$$

Здесь p_0 означает равновесное давление пара адсорбата, однозначно определяемое константой конденсации $K_L(T)$, что видно из формулы (5.11.9).

С помощью равенств (5.11.7)–(5.11.9) можно ввести новую температурно-зависимую константу теории БЭТ:

$$C(T) = K_1(T)p_0 \equiv \frac{K_1(T)}{K_L(T)} = C_0 e^{(Q_1 - Q_L)/RT}. \quad (5.11.10)$$

Равенства (5.11.6) могут быть переписаны с использованием константы (5.11.10) следующим образом:

$$\begin{aligned} \theta_1 &= \theta_0 K_1(T) p = \theta_0 C(T) \frac{p}{p_0}, \\ \theta_2 &= \theta_1 K_L(T) p = \theta_0 C(T) \left(\frac{p}{p_0} \right)^2, \\ \theta_3 &= \theta_2 K_L(T) p = \theta_0 C(T) \left(\frac{p}{p_0} \right)^3 \quad \text{и т. д.} \end{aligned} \quad (5.11.11)$$

Суммирование всех равенств (5.11.11) вплоть до бесконечности дает следующий результат (где $C \equiv C(T)$):

$$\begin{aligned} \theta_1 + \theta_2 + \theta_3 + \dots &= C \frac{p}{p_0} \left[1 + \frac{p}{p_0} + \left(\frac{p}{p_0} \right)^2 + \left(\frac{p}{p_0} \right)^3 + \dots \right] \theta_0 = \\ &= \frac{C p/p_0}{1 - p/p_0} \theta_0. \end{aligned} \quad (5.11.12)$$

Здесь ряд в квадратных скобках вычислен как сумма убывающей геометрической прогрессии (при $x = p/p_0 < 1$):

$$1 + x + x^2 + x^3 + \dots = \frac{1}{1 - x}. \quad (5.11.13)$$

Подстановка (5.11.12) в формулу (5.11.2) позволяет найти долю поверхности адсорбента, не занятой молекулами адсорбата:

$$\theta_0 = \frac{1 - p/p_0}{1 + (C - 1)p/p_0}. \quad (5.11.14)$$

Аналогичная подстановка в формулу (5.11.3) дает результирующую степень покрытия поверхности адсорбента:

$$\theta = C \frac{p}{p_0} \left[1 + 2 \frac{p}{p_0} + 3 \left(\frac{p}{p_0} \right)^2 + \dots \right] \theta_0. \quad (5.11.15)$$

Дифференцированием геометрической прогрессии (5.11.13) по x с заменой $x \rightarrow p/p_0$ легко убедиться в том, что

$$1 + 2(p/p_0) + 3(p/p_0)^2 + \dots = \frac{1}{(1 - p/p_0)^2}. \quad (5.11.16)$$

Подставляя (5.11.14) и (5.11.16) в формулу (5.11.15), получаем **уравнение изотермы полимолекулярной адсорбции БЭТ**:

$$\theta(T, p) = \frac{C(T)p/p_0}{(1 - p/p_0)[1 + (C(T) - 1)p/p_0]}. \quad (5.11.17)$$

Результаты численного расчета по формуле (5.11.17) при различных значениях константы $C(T)$ приведены на рис. 5.42.

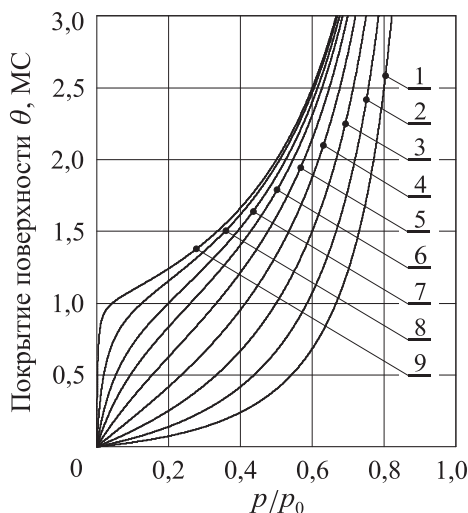


Рис. 5.42. Изотермы полимолекулярной адсорбции БЭТ, рассчитанные по формуле (5.11.17) при различных значениях константы $C(T)$: 0,25 (кривая 1), 0,5 (кривая 2), 1 (кривая 3), 2 (кривая 4), 4 (кривая 5), 8 (кривая 6), 16 (кривая 7), 48 (кривая 8), 480 (кривая 9); покрытие измеряется числом монослоев (МС)

Уравнение изотермы БЭТ содержит единственную константу C , характеризующую отношение времен жизни молекул в первом слое (τ_1) и в последующих слоях ($\tau_l = \tau_L$). Согласно (5.11.10), зависимость $C(T)$ определяется превышением теплоты адсорбции

в первом слое над теплотой конденсации, равным $(Q_1 - Q_L)$ и называемым *чистой теплотой адсорбции*.

При малом давлении паров адсорбата $(p/p_0 \ll 1)$ и больших значениях константы теории БЭТ $(C \gg 1)$, что справедливо при $Q_1 \gg Q_L$ уравнение (5.11.17) превращается в уравнение изотермы Ленгмюра типа (5.10.16),

$$\theta(T, p) = \frac{K_1(T)p}{1 + K_1(T)p}, \quad (5.11.18)$$

включающее только константу адсорбции $K_1(T)$ в первом слое.

Из сравнения рис. 5.42 с рис. 5.35 видно что, теория БЭТ вполне применима для описания экспериментальных изотерм III типа (с перегибом) при $C \gg 1$ и IV типа (вогнутых) при $C \ll 1$.

В дополнение к традиционной форме (5.11.17), в литературе имеются модификации уравнения БЭТ, приближающие расчетные кривые к экспериментальным изотермам III типа путем введения дополнительных (к константе C) параметров. Если считать, что полимолекулярная адсорбция ограничивается l слоями (второй параметр в теории БЭТ, где было $l \rightarrow \infty$), то уравнение изотермы адсорбции принимает следующий вид ($x = p/p_0 < 1$):

$$\theta = \frac{Cx}{1-x} \frac{1 - (l+1)x^l + lx^{l+1}}{1 + (-1)x - Cx^{l+1}}. \quad (5.11.19)$$

Следует отметить, что для одного слоя (при $l = 1$) уравнение (5.11.19) полностью (при всех возможных значениях $x = p/p_0$ и $C = K_1 p_0$) превращается в уравнение Ленгмюра (5.11.18).

Расчет по формуле (5.11.19), показывает, что при небольшом числе слоев ($l < 10$) степень покрытия поверхности для высоких значений p/p_0 заметно уменьшается. Встречаются и другие модификации уравнения БЭТ, в частности, учитывающие латеральное взаимодействие адсорбат-адсорбат при локализованной адсорбции по методике, описанной в связи с уравнением (5.10.47).

Путем соответствующего выбора дополнительных параметров модели БЭТ, например изменением конечного числа l адсорбционных слоев, удастся получить лучшее согласование с экспериментальными данными в более широком диапазоне давлений p/p_0 , чем дает традиционное уравнение БЭТ вида (5.11.17). Отмеченное ограничение числа слоев может реализоваться в достаточно малых порах из-за близости противоположных стенок, что также способствует капиллярной конденсации адсорбата.

2. Капиллярная конденсация в мезопористых адсорбентах. Морфология реальных пористых структур сложна и разнообразна, что создает определенные трудности при установлении

взаимосвязи между их морфологическими особенностями и физическими свойствами. Действительно, пористые материалы имеют нерегулярное строение, при котором пространство пор образовано не индивидуальными порами той или иной конкретной формы, а представляет собой сложный лабиринт из пересекающихся пор и капилляров разных размеров и формы.

По структуре пористые тела подразделяют на два класса — *корпускулярные* и *губчатые*. Корпускулярные пористые тела образованы из конгломератов сросшихся частиц произвольной формы и размеров, промежутки между которыми образуют поры. Типичными представителями являются силикагели, состоящие из сферических частиц, разделенных порами. В губчатых телах поры представляют собой сеть каналов (капилляров) различной формы и размеров, которые образуются, например при удалении какого-либо компонента из многофазной системы.

Вне зависимости от структуры пористые системы принято условно классифицировать по размерам пор следующим образом:

- *макропоры* с размерами более 50 нм и удельной поверхностью, не превышающей $0,5 \text{ м}^2/\text{г}$;
- *мезопоры* с размерами от 2 до 50 нм и удельной поверхностью, достигающей $200 \text{ м}^2/\text{г}$;
- *микропоры* с размерами менее 2 нм и удельной поверхностью, достигающей $1000 \text{ м}^2/\text{г}$.

Несмотря на использованную триаду «макро–мезо–микро», приведенные размеры пор соответствуют нанометровому диапазону (менее 100 нм). Поэтому представляющие практический интерес вещества являются *нанопористыми материалами*.

В основу вышеприведенной классификации положены характерные особенности адсорбционных свойств, присущие данному диапазону размеров и отличающие его от двух других.

Макропоры имеют средний диаметр, намного превышающий размеры адсорбируемых молекул. Поэтому их стенки можно рассматривать как почти плоские и ровные поверхности, на которых происходит обратимая моно- или полимолекулярная адсорбция, описываемая теориями Ленгмюра и БЭТ. В основном они выполняют функцию транспортных каналов для доставки адсорбата в глубь сорбента к поверхности адсорбирующих микро- и мезопор.

Микропоры имеют средний диаметр, соизмеримый с размерами адсорбируемых молекул, вследствие чего микропористые адсорбенты часто называют *молекулярными ситами*. В ряде случаев микропоры дополнительно подразделяют на *ультрамикропоры* (с размерами менее 1 нм, что характерно, например для

цеолитов и активированных углей) и *супермикропоры* (с размерами от 1 до 2 нм, например в некоторых силикагелях).

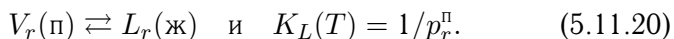
Малое расстояние между противоположными стенками ультрамикропор, сравнимое с размерами адсорбированных молекул, способствует их непосредственному взаимодействию с окружающими молекулами на стенках адсорбента. В этом случае имеет место сразу полное заполнение пор молекулами, которому для супермикропор предшествует образование адсорбционной пленки.

Мезопоры представляют наибольший физический интерес для дальнейшего рассмотрения. Им присуще послойное заполнение поверхности физадсорбируемыми или хемосорбируемыми молекулами, завершающееся объемным заполнением пор по механизму *капиллярной конденсации*.

Как уже отмечено (см. с. 634), конденсация пара на вогнутой поверхности капилляра в виде жидкости, смачивающей эту поверхность, происходит при меньшем давлении пара, чем на плоской поверхности, т. е. при $p_r^n < p_\infty$ (нижние кривые на рис. 5.29).

Воспользуемся аналогией с уравнением (5.11.9), использованным для описания объемной конденсации. В этом уравнении закон действующих масс, $K_L(T) = 1/p_0$, содержит давление p_0 , по сути своей означающее равновесное давление пара над плоской поверхностью жидкости, то есть $p_0 \equiv p_\infty$.

По аналогии с вышесказанным, представляем процесс конденсации жидкости из пара на вогнутой поверхности с радиусом кривизны r в форме квазихимической реакции, записанной совместно с законом действующих масс следующим образом:



Если реальное давление внутри капилляра $p \geq p_0 \equiv p_\infty$, то, в соответствии с уравнением изотермы Вант-Гоффа (2.13.5), изобарный потенциал для реакции (5.11.20) принимает следующий вид (ср. уравнение (5.8.28)):

$$\Delta G_r = RT \left[\ln \frac{1}{p} - \ln K_L(T) \right] = RT \ln \frac{p_r^n}{p}. \quad (5.11.21)$$

Отсюда видно, что $\Delta G_r > 0$ при невысоких давлениях пара ($p < p_r^n$), таких что реакция (5.11.20) смещена влево, в сторону испарения. С ростом давления достигается условие $p \geq p_r^n$, при котором $\Delta G_r \leq 0$ (смещение реакции вправо) и начинается конденсация жидкости на стенках капилляра.

Рассмотрим особенности процессов адсорбции/десорбции и конденсации пара на поверхности трех капилляров конусообразной и цилиндрической формы, изображенных на рис. 5.43.

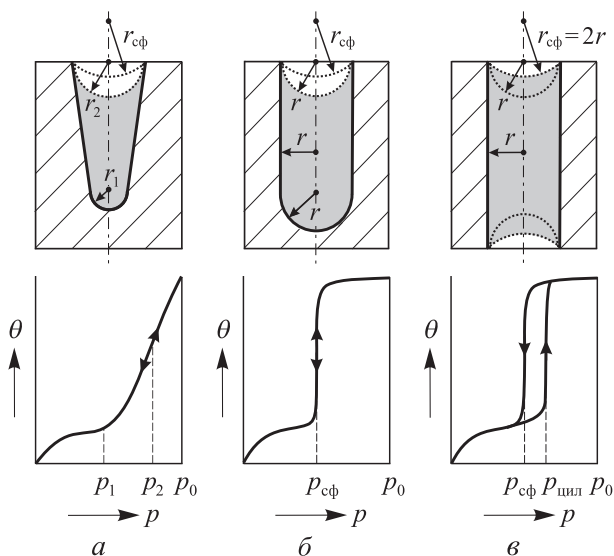


Рис. 5.43. Схема капиллярной конденсации в трех порах простейшей формы: а — конусообразная пора; б — цилиндрическая пора, открытая с одного конца; в — цилиндрическая пора, открытая с обоих концов

Для дальнейшего анализа удобно воспользоваться формулой Гиббса–Томсона (5.8.26) для вогнутой поверхности (со знаком минус), которую переписываем в следующем виде:

$$p_r^n = p_\infty \exp\left(-\frac{v^{\text{ж}} \Delta p_L}{RT}\right). \quad (5.11.22)$$

Здесь введено капиллярное давление Лапласа Δp_L , даваемое формулами (5.8.7) и (5.8.8) для цилиндрической поверхности (радиусом $r_{\text{цил}}$) и для сферической поверхности (радиусом $r_{\text{сф}}$):

$$\Delta p_L^{\text{цил}} = \frac{\sigma^{\text{ж}}}{r_{\text{цил}}} \quad \text{и} \quad \Delta p_L^{\text{сф}} = \frac{2\sigma^{\text{ж}}}{r_{\text{сф}}}. \quad (5.11.23)$$

Использование формулы Гиббса–Томсона (5.11.22) позволяет записать выражение (5.11.21) для изобарного потенциала процесса адсорбции–конденсации следующим образом (при $p < p_0$):

$$\Delta G_r = RT \ln \frac{p_r^n / p_\infty}{p / p_0} = RT \left| \ln \frac{p}{p_0} \right| - v^{\text{ж}} \Delta p_L \geq 0. \quad (5.11.24)$$

Здесь верхнее неравенство ($\Delta G_r > 0$) соответствует процессу адсорбции при $p < p_r^n$, а нижнее неравенство ($\Delta G_r < 0$) — процессу конденсации при $p > p_r^n$. При $\Delta G_r = 0$ имеет место равновесие жидкости и насыщенного пара с давлением $p = p_r^n$.

Критический режим перехода от адсорбции к конденсации ($\Delta G_r = 0$ при $p_{кр} = p_r^n$) описывается следующим уравнением:

$$\left| \ln \frac{p_{кр}}{p_0} \right| = \frac{v^ж \Delta p_L}{RT} = \begin{cases} \frac{v^ж}{RT} \frac{\sigma^ж}{r_{цил}} & \text{— цилиндр,} \\ \frac{v^ж}{RT} \frac{2\sigma^ж}{r_{сф}} & \text{— сфера.} \end{cases} \quad (5.11.25)$$

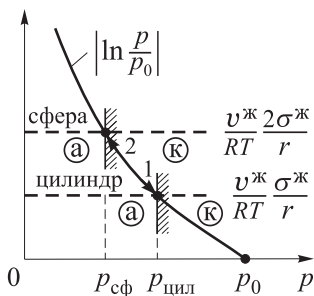


Рис. 5.44. К объяснению капиллярно-конденсационного гистерезиса

Отсюда следует, что при увеличении давления пара внутри капилляра ($p < p_0$) конденсация жидкости на цилиндрических и сферических поверхностях начинается при разных его значениях. Именно с этим связано появление *капиллярно-конденсационного гистерезиса*.

Графическая интерпретация уравнений (5.11.24) и (5.11.25) приведена на рис. 5.44. Здесь сплошной линией изображена левая часть уравнения (5.11.25), а штриховыми линиями — его правая часть для сферической и цилиндрической поверхностей, имеющих одинаковый радиус r . Точки пересечения таких кривых дают критические давления для этих поверхностей: $p_{кр} = p_{сф}$ и $p_{кр} = p_{цил}$, так что $p_{сф} < p_{цил}$ при $r_{сф} = r_{цил} = r$. По разные стороны от таких точек располагаются области адсорбции (слева) и конденсации (справа), отмеченные как (а) и (к), соответственно. Стрелками вдоль кривой $|\ln(p/p_0)|$ показано направление перехода (уравнение (5.11.24)) из одной области в другую при конденсации (стрелка 1) и испарении (стрелка 2).

В конусообразной поре, изображенной на рис. 5.43 а, конденсация начинается на нижней сферической поверхности с минимальным радиусом r_1 , которому соответствует критическое давление $p_1 = p_0 \exp(-2\sigma^ж v^ж / r_1 RT)$. Для продвижения жидкости в более широкую часть поры надо повышать давление вплоть до значения p_2 , соответствующего верхнему мениску с радиусом r_2 . Область между полусферами r_1 и r_2 , заполненная жидкостью, показана затемненной на рис. 5.43 а. Выше этой области окончательное заполнение поры происходит в результате постепенного увеличения радиуса сферического мениска ($r_{сф} > r_2$) вплоть до полного заполнения поры при $r_{сф} \rightarrow \infty$, чему соответствует давление $p_\infty \equiv p_0$. При десорбции процесс протекает тем же путем в обратном направлении, то есть не наблюдается гистерезиса.

В цилиндрической поре, закрытой с одного конца, как показано на рис. 5.43 б, конденсация начинается именно на этом конце при давлении $p_{\text{сф}} = p_0 \exp(-2\sigma^{\text{ж}}v^{\text{ж}}/rRT)$ (см. рис. 5.44) и далее протекает аналогично предыдущему случаю. Различие состоит лишь в том, что пора из-за постоянного радиуса цилиндра ($r = \text{const}$) вплоть до верхнего мениска заполняется жидкостью (затемненная область) при неизменном давлении $p_{\text{сф}}$. Поэтому на этом участке изотерма капиллярной конденсации идет вертикально вверх. Заключительный процесс полного заполнения поры и десорбция (при понижении давления) протекают также, как и в конусообразной поре, то есть обратимо и без гистерезиса.

В цилиндрической поре, открытой с обоих концов, как показано на рис. 5.43 в, конденсация протекает другим способом. Отсутствие сферической поверхности приводит к конденсации на цилиндрическом мениске адсорбционной пленки, покрывающей стенки капилляра. Она начинается при более высоком давлении $p_{\text{цил}} = p_0 \exp(-\sigma^{\text{ж}}v^{\text{ж}}/rRT)$ (см. стрелку 1 на рис. 5.44) по сравнению с предыдущим случаем. При этом заполнение поры происходит одновременно по всей ее высоте путем уменьшения радиуса. В момент полного заполнения сечения капилляра на его концах образуются сферические мениски, ограничивающие затемненную область, заполненную жидкостью. Так как это происходит при постоянном давлении $p_{\text{цил}} > p_{\text{сф}}$, то радиус этих менисков равен удвоенному радиусу цилиндра ($r_{\text{сф}} = 2r$). При дальнейшем повышении давления ($p > p_{\text{цил}}$) радиус менисков возрастает вплоть до $r_{\text{сф}} \rightarrow \infty$ при полном заполнении капилляра.

Процесс десорбции (при снижении давления) протекает путем испарения жидкости с открытых концов капилляра. В результате этого образуются сферические мениски, радиус которых по мере испарения жидкости уменьшается от $r_{\text{сф}} = \infty$ до $r_{\text{сф}} = r$ (последние изображены пунктирной линией внутри затемненной области на рис. 5.43 в). Следовательно, при давлении $p_{\text{сф}} = p_0 \exp(-2\sigma^{\text{ж}}v^{\text{ж}}/rRT)$ радиус концевых сферических менисков становится равным радиусу адсорбционной пленки на стенках цилиндра, и вся оставшаяся в капилляре жидкость полностью испаряется (см. стрелку 2 на рис. 5.44). Десорбционная ветвь на рис. 5.43 в при давлении $p_{\text{сф}} < p_{\text{цил}}$ опускается вертикально вниз до изотермы полимолекулярной адсорбции, так что возникает капиллярно-конденсационная петля гистерезиса.

Однако реальные адсорбенты не обладают однородной пористой структурой. Поэтому капилляры заполняются и освобождаются стохастически, что приводит к наклону ветвей гистерезиса.

В целом процесс адсорбции в реальных пористых структурах, включающих капилляры различных типов и размеров, достаточно сложен и разнообразен. В общих чертах он протекает во всех порах одновременно, начиная с заполнения микропор, и далее продолжается в мезопорах и макропорах по механизму моно- или полимолекулярного сорбирования. В мезопорах адсорбция завершается капиллярной конденсацией при продолжающейся полимолекулярной адсорбции в крупных макропорах.

5.12. Роль адсорбции, растворения и диффузии в газопоглощении материалов

Процесс газопоглощения заключается в проникновении газовых молекул внутрь твердого тела с образованием твердых растворов или химических соединений в объеме этого тела. Молекулы газа, пройдя предварительные этапы физической и химической адсорбции на поверхности твердого тела, растворяются в его поверхностных слоях путем вхождения в кристаллическую решетку с последующим диффузионным распределением по всему объему твердого тела. Таким образом, физадсорбция, хемосорбция, растворение и диффузия представляют собой последовательные стадии единого процесса поглощения газов твердым телом. Исключение какой-либо одной из этих стадий в такой последовательности делает невозможным газопоглощение в целом. Следовательно, отдельные газы, не хемосорбируемые поверхностью металлов, не могут растворяться в них. Как пример, все благородные газы и азот не хемосорбируются такими металлами, как Cu, Ni, Al, Ag, Au, а золото также не хемосорбирует водород и кислород, что и делает его благородным металлом.

Так как растворение и диффузия завершают предшествующие им этапы адсорбции, то возможно построение общей энергетической диаграммы для процесса газопоглощения в целом, которая дана на рис. 5.45. Правая часть этой диаграммы при $x > 0$ повторяет энергетическую диаграмму для физической и химической адсорбции, изображенную на рис. 5.33. Левая часть диаграммы при $x < 0$ соответствует кристаллической решетке твердого тела.

Растворение, будучи продолжением процесса хемосорбции, требует более сильного сближения хемосорбированных атомов с поверхностью твердого тела с целью их последующего проникновения в кристаллическую решетку. Однако этому препятствуют силы отталкивания со стороны атомов (ионов) решетки. Для преодоления этих сил необходима дополнительная энергия ϵ_p , называемая *энергией активации процесса растворения*.

На энергетической диаграмме энергия активации отображена в форме потенциального барьера высотой $\mathcal{E}_p$, разделяющего газовую фазу (при $x > 0$) и твердое тело (при $x < 0$). Хемосорбированные атомы, обладающие достаточной энергией, преодолевают барьер $\mathcal{E}_p$ и попадают в кристаллическую решетку. Далее они распределяются в минимумах потенциального рельефа решетки, перемещаясь путем диффузионных скачков с преодолением активационных барьеров высотой $\mathcal{E}_d$, равной энергии активации процесса диффузии. Механизмы диффузии атомов в твердом теле были ранее рассмотрены в п. 4.2.

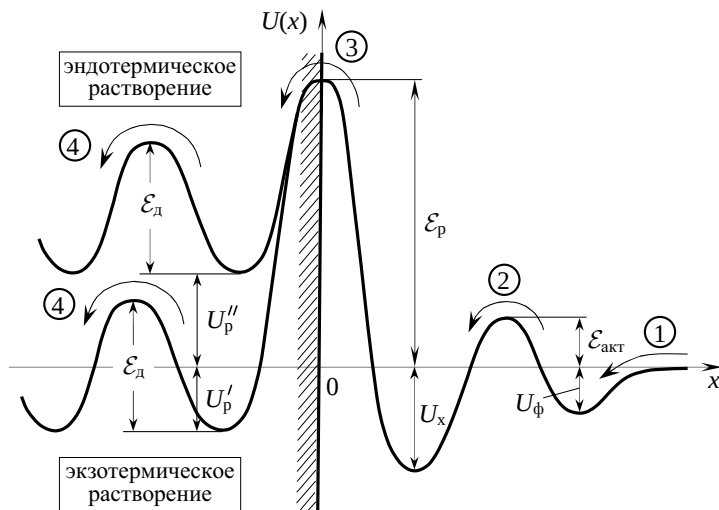


Рис. 5.45. Энергетическая диаграмма поглощения газов твердым телом, включающая последовательные этапы физической адсорбции (стрелка ①), хемосорбции (стрелка ②), растворения (стрелка ③) и диффузии (стрелки ④)

Последовательность прохождения молекулой этапов физадсорбции, хемосорбции, растворения и диффузии показана на рис. 5.45 соответствующими цифрами ①, ②, ③ и ④.

Растворение газов в конденсированных средах подчиняется закономерностям, изученным в п. 2.6. Диаграмма, показанная на рис. 5.45, позволяет понять энергетическую природу эндотермического и экзотермического растворения газов в металлах.

Как известно, при поглощении двухатомных газов (типа H_2 , N_2 и др.) металлами концентрация s атомов газа, растворенных в поверхностных слоях металла, связана с парциальным давлением p этого газа законом Сивертса (2.6.12):

$$c(T, p) = s(T) \sqrt{p}. \quad (5.12.1)$$

Коэффициент растворимости $s(T)$ зависит от температуры, согласно формуле (2.6.16),

$$s(T) = s_0 \exp\left(-\frac{\Delta H_{\text{раст}}}{2RT}\right). \quad (5.12.2)$$

Здесь $\Delta H_{\text{раст}} = \pm Q_p$ — тепловой эффект процесса растворения, происходящего с поглощением (при $\Delta H_{\text{раст}} > 0$) или с выделением (при $\Delta H_{\text{раст}} < 0$) *теплоты растворения* $Q_p = U_p \mathcal{N}_A$. Пропорциональность концентрации растворенного газа квадратному корню из давления ($c \propto \sqrt{p}$) в формуле (5.12.1) и множитель $1/2$ в показателе экспоненты (5.12.2) вызваны атомарным механизмом растворения двухатомных газов с диссоциацией молекулы на два атома.

Как уже отмечалось в п. 2.3, в зависимости от характера связи растворенных газовых атомов с кристаллической решеткой принципиально возможны два случая температурного хода растворимости (см. рис. 5.45).

1. Если энергия связи атомов с решеткой велика, то процесс растворения *экзотермический*, т. е. сопровождается выделением теплоты ($\Delta H_{\text{раст}} = -Q_p \equiv -U'_p \mathcal{N}_A < 0$). Энергетическая диаграмма отображает этот факт тем, что минимумы потенциального рельефа решетки располагаются, как и в случае процесса адсорбции, ниже нулевого уровня ($U = 0$) на величину U'_p . Такая ситуация реализуется, например при поглощении газов активными газопоглощающими материалами типа Ti, Zr, Ta, Th, Nb и др., в которых растворение газов идет с образованием химических соединений в твердой фазе (гидридов, нитридов, оксидов и т. д.). В этом случае формула (5.12.2) показывает, что коэффициент растворимости $s(T)$ уменьшается с ростом температуры.

2. Если энергия связи атомов с решеткой относительно невелика, то процесс растворения *эндотермический*, т. е. сопровождается поглощением теплоты ($\Delta H_{\text{раст}} = Q_p \equiv U''_p \mathcal{N}_A > 0$). В этом случае минимумы потенциального рельефа решетки располагаются выше нулевого уровня ($U = 0$) на величину U''_p . Такая ситуация реализуется при поглощении газов обычными конструкционными металлами типа Fe, Ni, Cu, W, Mo и их сплавами, в которых растворение газов идет в основном с образованием твердых растворов. В противоположность предыдущему случаю, коэффициент растворимости (5.12.2) увеличивается с возрастанием температуры.

Газопоглощающая способность активных газопоглотителей намного превосходит такую же способность обычных металлов.

Например, растворимость водорода в титане при 0°C примерно в 5000 раз больше, чем в никеле при 800°C.

Как известно, наряду с атомарным растворением газа, возможен механизм «молекулярного сита», реализуемый, например в цеолитах и силикатных материалах типа стекол и керамики путем удержания газовых молекул (без диссоциации на атомы) в порах своей структуры. В этом случае справедлив закон Генри (2.6.4) в форме

$$c(T, p) = s(T)p. \quad (5.12.3)$$

Температурная зависимость коэффициента растворимости, входящего в выражение (5.12.3), дается формулой (2.6.9), из которой следует, что $s(T)$ уменьшается с возрастанием температуры, так как молекулярное растворение является экзотермическим процессом ($\Delta H_{\text{раст}} = -Q_p \equiv -U'_p \mathcal{N}_d < 0$).

5.13. Движущие силы и термодинамика гомогенного зародышеобразования

На сегодняшний день ростовые технологии являются базовыми при формировании пленочных структур в современной микро- и нанoeлектронике. Рост тонких пленок на подложках при кристаллизации из жидкой или парогазовой фазы обычно начинается со стадии зародышеобразования. Технологические условия и характер образования новой фазы, механизмы формирования и роста пленок в конечном итоге определяют все практически значимые характеристики синтезируемых твердотельных структур.

Здесь и в последующих параграфах будут рассмотрены основные физико-технологические особенности формирования новой (вторичной) фазы из маточной (первичной) фазы как внутри ее при **гомогенной кристаллизации**, так и на поверхности третьей фазы (подложки) при **гетерогенной кристаллизации**. В последнем случае существенная роль отводится свойствам поверхности подложки и межфазной границы в целом.

Прежде чем начинать изучение термодинамических условий гомогенного зародышеобразования, рассмотрим, как создавать объемное пересыщение внутри первичной (жидкой или парогазовой) фазы, приводящее к возникновению так называемых *критических зародышей* как центров кристаллизации.

1. Объемное пересыщение в первичной фазе. Будем рассматривать изобарные процессы, протекающие при суммарном давлении $P = \text{const}$, для описания которых следует использовать изобарный потенциал G_v . Здесь введен ранее отсутствовавший

индекс v (от англ. *volume*), который отражает только объемные свойства первичной фазы. Ниже объемный потенциал G_v будет дополнен поверхностным вкладом G_s , отмечаемым индексом s (от англ. *surface*).

Как известно (см. п. 1.8 и рис. 1.6), неравновесные изобарно-изотермические процессы самопроизвольно протекают в направлении уменьшения изобарного потенциала ($\Delta G_v < 0$). Они идут до тех пор, пока не наступает термодинамическое равновесие (фазовое или химическое), при котором G_v достигает минимального при данных условиях значения, обеспечивая $\Delta G_v = 0$.

Таким образом, *движущей силой* фазовых и химических превращений в изобарно-изотермических условиях является изменение объемного изобарного потенциала ΔG_v , навязанное системе внешними условиями и называемое *объемным пересыщением*. Реальными параметрами, с помощью которых можно управлять процессами в технологических системах, служат температура T , объемная концентрация c_i частиц i -го компонента в жидкой фазе и парциальное давление $p_i = c_i k_B T$ в парогазовой фазе.

Все методы осаждения твердых веществ на подложках можно разделить на две группы (см. п. 4.13):

1) *физические методы осаждения*, основанные на принципе нарушения условий фазового равновесия при кристаллизации:

- из расплава осаждаемого вещества,
- из раствора-расплава, содержащего осаждаемое вещество,
- из парогазовой фазы, содержащей осаждаемое вещество;

2) *химические методы осаждения*, основанные на принципе нарушения условий химического равновесия при кристаллизации из химически реактивной газовой фазы.

Покажем, каким образом в этих случаях величину ΔG_v можно связать с управляемыми параметрами процесса кристаллизации — переохлаждением ΔT , концентрационным пересыщением Δc_i и избыточным парциальным давлением Δp_i , а также с составом химически реактивной газовой фазы.

Кристаллизация из расплава вещества с температурой кристаллизации $T_{кр}$ возможна при переохлаждении исходной жидкой фазы на величину $\Delta T = T_{кр} - T > 0$. Температурный ход объемных изобарных потенциалов $G_{кр}$ для кристалла и $G_{расп}$ для расплава качественно показан на рис. 5.46, построенном в полном соответствии с рис. 2.2, поскольку для чистого вещества $G_v = \mu N$. При температуре $T_{кр}$ обе фазы находятся в равновесии, тогда $\Delta G_v \equiv G_{кр} - G_{расп} = \Delta H_{кр} - T_{кр} \Delta S_{кр} = 0$. Так как кристаллизация всегда сопровождается выделением тепла,

то $\Delta H_{\text{кр}} = T_{\text{кр}} \Delta S_{\text{кр}} < 0$. Охлаждение расплава до температуры $T < T_{\text{кр}}$ обеспечивает $\Delta G_v = \Delta H - T \Delta S < 0$, т. е. делает возможным процесс кристаллизации. При достаточно малом температурном интервале ΔT можно считать, что $\Delta H \approx \Delta H_{\text{кр}}$ и $\Delta S \approx \Delta S_{\text{кр}} = \Delta H_{\text{кр}}/T_{\text{кр}}$.

Вышесказанное дает возможность связать величину объемного пересыщения ΔG_v с теплотой кристаллизации $Q_{\text{кр}} \equiv |\Delta H_{\text{кр}}| = -\Delta H_{\text{кр}}$ нижеследующим соотношением:

$$\Delta G_v = \Delta H - T \Delta S \approx \Delta H_{\text{кр}} - T \Delta S_{\text{кр}} = -Q_{\text{кр}} \frac{\Delta T}{T_{\text{кр}}} < 0. \quad (5.13.1)$$

Как видно из выражения (5.13.1), табличные величины $T_{\text{кр}}$ и $Q_{\text{кр}}$, известные для осаждаемого вещества, обеспечивают пропорциональность между расчетным пересыщением ΔG_v и измеряемым переохлаждением ΔT .

Кристаллизация из раствора-расплава вещества i -го сорта в изотермических условиях возможна в том случае, если обеспечено положительное концентрационное пересыщение $\Delta c_i = c_i - \bar{c}_i > 0$ по отношению к равновесной концентрации $\bar{c}_i$ этого вещества в растворе.

В равновесных условиях химический потенциал i -го компонента одинаков в растворе и в кристалле, т. е. $\mu_i^{\text{кр}} = \mu_i^{\text{раст}}$. Тогда, записывая химический потенциал $\mu_i^{\text{раст}}$ для модели идеального раствора, получаем

$$\mu_i^{\text{кр}} = \mu_i^{\circ}(T) + k_B T \ln \bar{c}_i. \quad (5.13.2)$$

В неравновесных условиях, когда в растворе концентрация i -го компонента $c_i > \bar{c}_i$, его химический потенциал определяется этой концентрацией и равен

$$\mu_i^{\text{раст}} = \mu_i^{\circ}(T) + k_B T \ln c_i. \quad (5.13.3)$$

Считаем, что отклонение концентрации от равновесного значения в жидкой фазе практически не оказывает влияния на термодинамические свойства кристалла, т. е. в неравновесных условиях остается справедливым выражение (5.13.2) для $\mu_i^{\text{кр}}$.

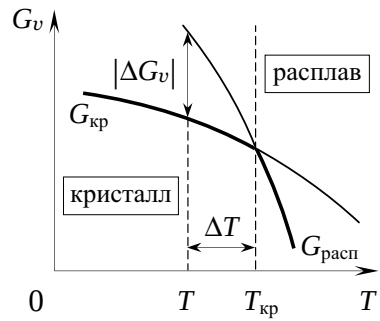


Рис. 5.46. Температурная зависимость изобарных потенциалов $G_{\text{кр}}$ для кристалла и $G_{\text{расп}}$ для расплава, переохлажденного на величину $|\Delta G_v|$ при температуре T

Если из раствора в кристаллическую фазу переходит N_i частиц, то изобарный потенциал системы изменяется на величину

$$\begin{aligned}\Delta G_v &\equiv G_{\text{кр}} - G_{\text{раст}} = N_i \Delta \mu_i = N_i (\mu_i^{\text{кр}} - \mu_i^{\text{раст}}) = -N_i k_B T \ln \frac{c_i}{\bar{c}_i} = \\ &= -N_i k_B T \ln \left(1 + \frac{\Delta c_i}{\bar{c}_i} \right) \approx -N_i k_B T \frac{\Delta c_i}{\bar{c}_i} < 0.\end{aligned}\quad (5.13.4)$$

При выводе выражения (5.13.4) были использованы формулы (5.13.2) и (5.13.3), при этом последнее приближенное равенство справедливо для малых пересыщений по i -му компоненту, когда $\Delta c_i \ll \bar{c}_i$.

Как видно из выражения (5.13.4), положительное концентрационное пересыщение в растворе ($\Delta c_i = c_i - \bar{c}_i > 0$) однозначно определяет объемное пересыщение в системе ($\Delta G_v < 0$), обеспечивающее кристаллизацию i -го вещества из раствора-расплава.

Кристаллизация из парогазовой фазы вещества i -го сорта в изотермических условиях возможна, если парциальное давление p_i обеспечивает в газовой фазе положительный избыток давления $\Delta p_i = p_i - \bar{p}_i > 0$ по отношению к равновесному парциальному давлению $\bar{p}_i$.

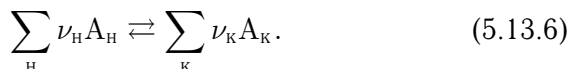
Последующий ход рассуждений в точности повторяет предыдущий случай с заменами $c_i \rightarrow p_i$ и $\bar{c}_i \rightarrow \bar{p}_i$. Если из парогазовой фазы в кристаллическую переходит N_i частиц, то изменение изобарного потенциала системы записываем, по аналогии с формулой (5.13.4), следующим образом:

$$\begin{aligned}\Delta G_v &\equiv G_{\text{кр}} - G_{\text{пар}} = N_i \Delta \mu_i = N_i (\mu_i^{\text{кр}} - \mu_i^{\text{пар}}) = -N_i k_B T \ln \frac{p_i}{\bar{p}_i} = \\ &= -N_i k_B T \ln \left(1 + \frac{\Delta p_i}{\bar{p}_i} \right) \approx -N_i k_B T \frac{\Delta p_i}{\bar{p}_i} < 0.\end{aligned}\quad (5.13.5)$$

Здесь последнее приближенное равенство справедливо при малых пересыщениях по i -му компоненту, когда $\Delta p_i \ll \bar{p}_i$.

Как видно из выражения (5.13.5), избыточное парциальное давление в газовой фазе ($\Delta p_i = p_i - \bar{p}_i > 0$) однозначно определяет объемное пересыщение в системе ($\Delta G_v < 0$), обеспечивающее кристаллизацию i -го вещества из парогазовой фазы.

Химическое осаждение из газовой фазы осуществляется с помощью обратимой гетерогенной химической реакции, записанной в обобщенной форме как (см. формулу (2.12.1))



Здесь по-прежнему A_n и A_k обозначают химические символы начальных (н) и конечных (к) реагентов, а ν_n и ν_k — соответствующие им стехиометрические коэффициенты.

Химическое пересыщение в неравновесной системе, создаваемое реакцией (5.13.6), записывается в форме (2.13.5), выражающей уравнение изотермы Вант-Гоффа:

$$\Delta G_v = RT \left[\left(\sum_k \nu_k \ln p_k - \sum_n \nu_n \ln p_n \right) - \ln K_p(T) \right]. \quad (5.13.7)$$

Здесь $K_p(T)$ — константа равновесия реакции, p_n и p_k — реальные парциальные давления начальных и конечных газообразных участников реакции, отличные от соответствующих равновесных величин $\bar{p}_n$ и $\bar{p}_k$. С помощью давлений p_n и p_k можно обеспечить нужный знак ΔG_v в равенстве (5.13.7), соответствующий смещению химического равновесия в сторону осаждения необходимого твердофазного продукта, входящего в реакцию (5.13.6).

Как ясно видно из полученных выше выражений (5.13.1), (5.13.4), (5.13.5) и (5.13.7) для объемного пересыщения ΔG_v , надлежащим выбором внешних условий всегда термодинамически возможно обеспечить процесс кристаллизации.

2. Понятие о критическом зародыше. Исходя из чисто объемных свойств маточной фазы, кристаллизация должна возникать даже при бесконечно малом пересыщении. Однако на практике для начала кристаллизации необходимо создавать конечные пересыщения в системе. Иными словами, имеется определенное *критическое пересыщение*, ниже которого существует метастабильная пересыщенная (или переохлажденная) первичная фаза, а выше — начинают образовываться зародыши новой (вторичной) фазы. Появление критического пересыщения и связанной с ним работы образования зародышей вызвано тем, что кристаллизация не является чисто объемным процессом. Она определяется в значительной степени свойствами границ раздела кристаллического зародыша с подложкой и с окружающей первичной фазой.

Как уже отмечалось в п. 5.5 (см. рис. 5.15), состояние частиц в поверхностных слоях отличается от их состояния внутри объема конденсированной фазы. Химические связи любой внутренней частицы насыщены валентными связями ее соседей. Для поверхностных частиц пространственная симметрия силовых взаимодействий нарушается из-за различия свойств смежных фаз и насыщение связей уже не является всесторонним. Это отражается в появлении избыточной поверхностной энергии ΔG_s , вводимой аналогично записанному в выражении (5.5.3). Следовательно,

поверхностный вклад в свободную энергию (ΔG_s) пропорционален площади ΔA поверхности раздела фаз:

$$\Delta G_s = \sigma \Delta A. \quad (5.13.8)$$

Здесь коэффициент пропорциональности равен поверхностному натяжению, $\sigma > 0$.

Поверхностное натяжение (удельная поверхностная энергия) имеет размерность Дж/м² $\equiv$ Н/м и выражает работу, которую необходимо совершить для увеличения границы раздела фаз на единицу площади. Выраженное в силовых единицах (ньютон на метр) поверхностное натяжение представляет собой силу, которую нужно приложить к единице длины линии, ограничивающей поверхность раздела фаз, для увеличения площади этой поверхности на единицу. Эта сила направлена по касательной к соответствующей границе раздела (см. ниже рис. 5.48).

Таким образом, образование кристаллического зародыша новой фазы на поверхности подложки, с одной стороны, способствует понижению энергии системы в силу уменьшения объемной свободной энергии ($\Delta G_v < 0$), а с другой стороны, требует затраты дополнительной энергии на создание границ раздела, поскольку поверхностная свободная энергия (5.13.8) всегда положительна ($\Delta G_s > 0$).

Следовательно, знак суммарного изобарного потенциала зародышеобразования,

$$\Delta G = \Delta G_v + \Delta G_s, \quad (5.13.9)$$

определяется соотношением двух вкладов — объемного $\Delta G_v < 0$ и поверхностного $\Delta G_s > 0$. Этот знак должен зависеть от размера зародыша, так как величины $|\Delta G_v|$ и ΔG_s пропорциональны, соответственно, объему зародыша ($V_{\text{зар}} \propto r^3$) и площади его поверхности ($A_{\text{зар}} \propto r^2$), где r — характерный линейный размер зародыша, уточняемый ниже в рамках рассматриваемой модели зародышеобразования. Разная степенная зависимость этих величин обеспечивает различную скорость их изменения с ростом r , а именно: при малых r быстрее нарастает квадратичная зависимость ΔG_s , а при больших r — кубическая зависимость $|\Delta G_v|$.

На рис. 5.47 качественно показаны зависимости ΔG_v , ΔG_s и ΔG от размера зародыша r . При малых r преобладает влияние поверхностного вклада, $\Delta G_s \propto r^2$, а при больших r — объемного вклада, $|\Delta G_v| \propto r^3$. Из-за разных знаков ΔG_v и ΔG_s суммарная кривая, $\Delta G = \Delta G_v + \Delta G_s$, проходит через максимум высотой ΔG^* при размере зародыша $r = r^*$. Зародыши таких размеров называют *критическими*.

Докритические зародыши с размерами $r < r^*$ неустойчивы и самопроизвольно растворяются в первичной фазе, поскольку их рост вызывает повышение изобарного потенциала в направлении к максимуму ΔG^* . Напротив, *закритические зародыши* с размерами $r > r^*$ способны к дальнейшему росту, так как при удалении вправо от максимума величина ΔG уменьшается. Иными словами, отрыв хотя бы одного атома от критического зародыша способствует его распаду, а присоединение атома приводит к дальнейшему росту зародыша, что показано стрелками на рис. 5.47.

Конкретные выражения для ΔG^* и r^* будут получены ниже в рамках двух моделей — трехмерного (куполообразного) зародыша и двумерного (дискообразного) зародыша.

Процесс возникновения критических зародышей связан с преодолением потенциального барьера зародышеобразования высотой ΔG^* , т. е. относится к классу термоактивационных процессов. Высота термоактивационного барьера ΔG^* трактуется как *работа образования критического зародыша*.

Появление критических зародышей, являющихся центрами кристаллизации, носит флуктуационный характер. Статистические флуктуации являются характерной чертой всех термодинамических систем (см. п. 1.19). Действительно, всякое равновесное состояние можно рассматривать как динамическое равновесие процессов переноса массы и энергии в двух противоположных направлениях. Флуктуации плотности массы и энергии являются исходной причиной возникновения зародышей в первичной фазе.

Иначе говоря, зародыши представляют собой флуктуационные отклонения от равновесного состояния гомогенной фазы. Вероятность таких отклонений, а следовательно, и вероятность образования критических зародышей, подчиняется формуле Больцмана (1.18.15), в которой энергия активации $\mathcal{E}$ заменена на работу образования критического зародыша ΔG^* :

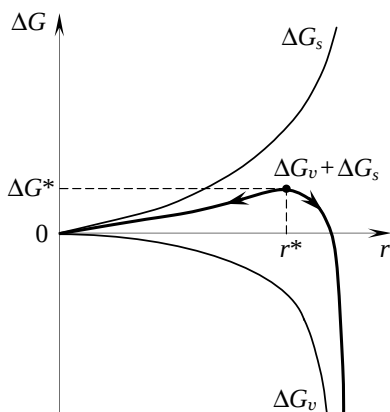


Рис. 5.47. Качественный вид зависимостей объемного (ΔG_v) и поверхностного (ΔG_s) вкладов в суммарный изобарный потенциал образования зародыша ($\Delta G = \Delta G_v + \Delta G_s$) от его характерного размера r

$$w = w_0 \exp\left(-\frac{\Delta G^*}{k_B T}\right). \quad (5.13.10)$$

Из формулы (5.13.10) следует, что в конкурентной борьбе между механизмами зародышеобразования побеждает тот механизм, для которого меньше величина $\Delta G^* > 0$.

К настоящему времени процессы кристаллизации и зародышеобразования изучались многими учеными в рамках различных подходов и модельных представлений, которые могут быть отнесены к одной из следующих теорий:

- термодинамическая теория,
- атомистическая теория,
- статистико–кинетическая теория.

Атомистическая теория рассматривает зародыши как кластеры, состоящие из дискретных частиц (молекул, атомов, ионов), взаимодействующих с подложкой и между собой. В этом случае зародыш характеризуется не термодинамическими свойствами осаждаемого материала, а потенциальной энергией, исходя из той или иной модели взаимодействия частиц. Задача атомистической теории состоит в определении числа частиц в критическом зародыше и скорости его образования, а не в нахождении критических величин r^* и ΔG^* , как в термодинамической теории.

Экспериментально установлено, что при достаточно больших пересыщениях зародыши могут состоять всего из нескольких атомов. В этом случае процесс зародышеобразования строго описать можно только в рамках статистической физики и кинетической теории, когда вместо свободных энергий используются статистические суммы и потенциальная энергия взаимодействия отдельных частиц.

Статистико–кинетические теории изучают развитие во времени процесса образования кластеров с различным числом частиц, учитывая при этом конкуренцию между элементарными актами адсорбции, десорбции, поверхностной диффузии, закрепления адатомов на дефектах подложки и т. п. Детализация кинетического описания сильно усложняет анализ процесса зародышеобразования. Более того, в большинстве случаев для этого требуется применение численного счета, особенно при моделировании методом Монте-Карло. Естественно, это выходит за рамки нашего изложения, поэтому дальнейшее рассмотрение будет ограничено термодинамическими рамками.

Термодинамическая модель рассматривает зародыши как частицы сплошной среды только малых размеров и приписывает им физические свойства массивного материала. Такая модель

применима лишь к макроскопическим зародышам, состоящим из достаточно большого числа частиц (молекул, атомов, ионов).

Задача термодинамической теории заключается в нахождении критических величин — r^* (размера критического зародыша) и ΔG^* (работы образования критического зародыша) — для *барьера зародышеобразования* в форме максимума на кривой зависимости суммарного изобарного потенциала, $\Delta G = \Delta G_v + \Delta G_s$, от характерного размера r зародыша (см. рис. 5.47).

3. Термодинамика гомогенного зародышеобразования. До того, как решать вышеуказанную задачу, обратимся к общим соотношениям, вытекающим из формулы Гиббса–Томсона (5.8.26) и выражения (5.13.5) для объемного пересыщения в парогазовой фазе. Кристаллические зародыши, спонтанно возникающие в объеме гомогенной фазы, имеют форму сферы малого радиуса, даваемого формулой Гиббса–Томсона (5.8.26):

$$r = \Omega \frac{2\sigma}{k_B T} \left[\ln(p_r/p_\infty) \right]^{-1}. \quad (5.13.11)$$

Здесь $\Omega = v^k/\mathcal{N}_A$ — объем кристаллического зародыша, приходящийся на одну его частицу.

Поскольку давление пара над зародышем (p_r) всегда больше, чем давление насыщенного пара (p_∞), то логарифм их отношения, стоящий в формуле (5.13.11), определяет пересыщение в парогазовой среде. Такое пересыщение для i -го компонента в форме ΔG_v дается формулой (5.13.5), откуда

$$\ln \frac{p_i}{\bar{p}_i} = \frac{|\Delta G_v|}{N_i k_B T} = \frac{|\Delta \mu_i|}{k_B T} \equiv \frac{\mu_i^{\text{пар}} - \mu_i^{\text{кр}}}{k_B T}. \quad (5.13.12)$$

По физическому смыслу имеет место соответствие между давлениями $p_i \rightarrow p_r$ (реальное давление над зародышем) и $\bar{p}_i \rightarrow p_\infty$ (равновесное давление в парогазовой среде). Поэтому, делая такие замены в (5.13.12) и подставляя результат в (5.13.11), получаем радиус сферического зародыша, возникающего в парогазовой среде при пересыщении (5.13.12):

$$r = \frac{2\sigma}{\Delta g_v}. \quad (5.13.13)$$

Здесь введено *удельное объемное пересыщение* для i -го компонента в парогазовой среде,

$$\Delta g_v = \frac{|\Delta \mu_i|}{\Omega_i} = \frac{k_B T}{\Omega_i} \ln \frac{p_i}{\bar{p}_i}, \quad (5.13.14)$$

соответствующее единичному объему конденсата, содержащему частицы i -го сорта в количестве $\Omega_i^{-1} = \mathcal{N}_A/v_i^k$.

Ниже покажем, что радиус (5.13.13) в точности соответствует критическому размеру $r_{\text{гом}}^*$ сферического зародыша при гомогенной кристаллизации.

Будем обозначать первичную (питающую) фазу цифрой 1, а вторичную фазу (конденсат) — цифрой 2. Разделяющая их граница в этих обозначениях имеет поверхностное натяжение σ_{12} .

Изобарный потенциал (5.13.9) образования сферического зародыша в гомогенной первичной фазе содержит два вклада:

- объемное пересыщение, выраженное через удельное пересыщение типа (5.13.14),

$$\Delta G_v = -V_{\text{зар}} \Delta g_v = -V_{\text{сф}} \Delta g_v \equiv -\frac{4\pi}{3} r^3 \Delta g_v; \quad (5.13.15)$$

- поверхностный вклад типа (5.13.8),

$$\Delta G_s = A_{\text{зар}} \sigma_{12} = A_{\text{сф}} \sigma_{12} \equiv 4\pi r^2 \sigma_{12}. \quad (5.13.16)$$

Суммируя выражения (5.13.15) и (5.13.16), получаем изобарный потенциал гомогенного зародышеобразования:

$$\Delta G_{\text{гом}} = -\frac{4\pi}{3} r^3 \Delta g_v + 4\pi r^2 \sigma_{12}. \quad (5.13.17)$$

Из условия $(\partial \Delta G_{\text{гом}} / \partial r)|_{r=r_{\text{гом}}^*} = 0$ находим радиус критического зародыша при гомогенном зародышеобразовании:

$$r_{\text{гом}}^* = \frac{2\sigma_{12}}{\Delta g_v}. \quad (5.13.18)$$

Подстановкой выражения (5.13.18) в формулу (5.13.17) получаем работу образования критического зародыша:

$$\Delta G_{\text{гом}}^* = \frac{16\pi}{3} \frac{\sigma_{12}^3}{\Delta g_v^2} \equiv \begin{cases} \frac{1}{2} V_{\text{зар}}^* \Delta g_v = \frac{1}{2} |\Delta G_v^*|, \\ \frac{1}{3} A_{\text{зар}}^* \sigma_{12} = \frac{1}{3} \Delta G_s^*. \end{cases} \quad (5.13.19)$$

Из соотношений (5.13.18) и (5.13.19), впервые полученных еще Гиббсом, видно, что критический радиус зародыша обратно пропорционален удельному пересыщению ($r_{\text{гом}}^* \propto \Delta g_v^{-1}$), а работа его образования ($\Delta G_{\text{гом}}^*$) пропорциональна площади поверхности критического зародыша, $A_{\text{зар}}^* = 4\pi r_{\text{гом}}^{*2} \propto \Delta g_v^{-2}$. Таким образом, с ростом пересыщения в системе снижается линейно критический размер зародыша и квадратично — работа его образования.

Сравнение формул (5.13.13) и (5.13.18) говорит о полной их тождественности. Следовательно, радиус r сферического конденсата, входящий в формулу Гиббса–Томсона (5.8.26), есть не что иное, как критический радиус (5.13.18) сферического зародыша, соответствующий пересыщению (5.13.12).

С технологических позиций наибольший интерес представляет не гомогенное, а гетерогенное зародышеобразование. Оно имеет место при эпитаксиальном росте совершенных монокристаллических пленок на подложках, выращивании монокристаллов на затравках, осаждении веществ на нерастворимых твердых частицах в расплаве и т. п.

5.14. Термодинамика и кинетика гетерогенного зародышеобразования

При анализе процессов зародышеобразования на поверхности подложек необходимо учитывать силы межатомного взаимодействия как с подложкой, так и в объеме первичной фазы. Если взаимодействие между частицами зародыша и подложки сильнее, чем внутри питающей (маточной) фазы, то гетерогенная кристаллизация энергетически более выгодна по сравнению с гомогенной, поскольку при этом достигается уменьшение свободной энергии Гиббса. Следовательно, гетерогенное зародышеобразование имеет место при выполнении условия $\Delta G_{\text{гет}}^* < \Delta G_{\text{гом}}^*$. Ниже будет показано, когда и как реализуется это условие.

Рассмотрим элементы термодинамической теории гетерогенного зародышеобразования в рамках модели трехмерного зародыша в форме сферического купола, цилиндра, прямоугольного параллелепипеда, гексагональной призмы и модели двумерного зародыша в форме плоского диска, квадрата и шестиугольника.

1. Модель куполообразного сферического зародыша применима к кристаллам с изотропной или со слабо анизотропной поверхностной энергией, зародыш которых имеет форму купола с радиусом r и с *контактным* (краевым) *углом* θ . Значение контактного угла определяется соотношением поверхностных натяжений на границах раздела трех фаз: первичная фаза–подложка ($\sigma_{1п}$), вторичная фаза–подложка ($\sigma_{2п}$) и первичная фаза–вторичная фаза (σ_{12}). Силовая трактовка, данная в п. 5.13, позволяет изобразить поверхностные натяжения в виде векторов силы, как это сделано на рис. 5.48.

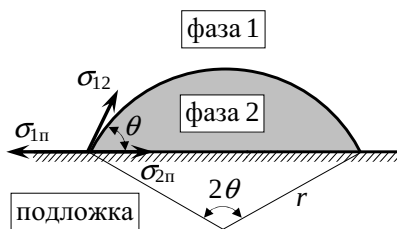


Рис. 5.48. Модель куполообразного зародыша с радиусом r и контактным углом θ

Из условия силового равновесия поверхностных натяжений следует уравнение Юнга для контактного угла:

$$\cos \theta = \frac{\sigma_{1п} - \sigma_{2п}}{\sigma_{12}}. \quad (5.14.1)$$

Объем куполообразного зародыша (рис. 5.48) равен

$$V_{\text{зар}} = V_{\text{сф}} \Phi(\theta) \equiv \frac{4\pi r^3}{3} \Phi(\theta), \quad (5.14.2)$$

где введена функция контактного угла θ следующего вида:

$$\Phi(\theta) = \frac{1}{4} (2 - 3 \cos \theta + \cos^3 \theta), \quad (5.14.3)$$

Функция $\Phi(\theta)$, изображенная на рис. 5.49, учитывает зависимость работы образования трехмерного критического зародыша (ΔG_3^*) от поверхностных натяжений на границах фаз.

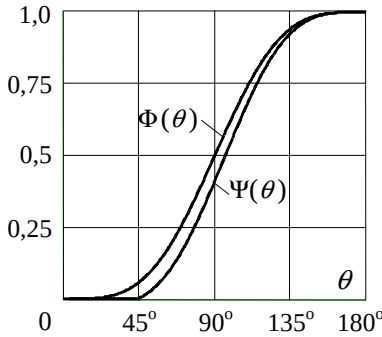


Рис. 5.49. Две функции контактного угла для трехмерного зародыша: $\Phi(\theta)$ на плоской поверхности и $\Psi(\theta)$ в двугранном угле (рис. 5.51)

По аналогии с (5.13.15) объемное пересыщение в первичной фазе записываем в форме

$$\begin{aligned} \Delta G_v &= -\Delta g_v V_{\text{зар}} = \\ &= -\frac{4\pi r^3}{3} \Phi(\theta) \Delta g_v. \end{aligned} \quad (5.14.4)$$

Здесь, как и ранее в формулах (5.13.13) и (5.13.14), введено удельное объемное пересыщение Δg_v , равное изобарному потенциалу, отнесенному к единице объема вторичной фазы. Если зародыш объемом $V_{\text{зар}}$, даваемым формулой (5.14.2), содержит

N_i частиц, тогда удельное объемное пересыщение равно

$$\Delta g_v \equiv \frac{|\Delta G_v|}{V_{\text{зар}}} = \frac{|\Delta G_v|/N_i}{V_{\text{зар}}/N_i} = \frac{|\Delta \mu_i|}{\Omega_i}, \quad (5.14.5)$$

где $\Omega_i = V_{\text{зар}}/N_i = v_i^k/\mathcal{N}_A \equiv M_i/(\rho_i^k \mathcal{N}_A)$ — объем конденсата, приходящийся на одну частицу i -го сорта.

В отличие от формулы (5.13.14) для парогазовой фазы, когда $\Delta g_v = (k_B T/\Omega_i) \ln(p_i/\bar{p}_i)$, выражение (5.14.5) применимо ко всем случаям объемного пересыщения, рассмотренным в п. 5.13.

Поверхностную составляющую изобарного потенциала образования трехмерного зародыша записываем в виде

$$\begin{aligned} \Delta G_s &= 2\pi r^2 (1 - \cos \theta) \sigma_{12} + \pi (r \sin \theta)^2 (\sigma_{2п} - \sigma_{1п}) = \\ &= 4\pi r^2 \Phi(\theta) \sigma_{12} \equiv A_{\text{зар}} \sigma_{12}. \end{aligned} \quad (5.14.6)$$

Здесь первое слагаемое связано с образованием новой границы раздела (характеризуемой σ_{12}) между первичной фазой и зародышем в форме купола с площадью поверхности, равной $2\pi r^2(1 - \cos\theta)$. Второе слагаемое вызвано заменой старой границы раздела подложка–фаза 1 (характеризуемой $\sigma_{1п}$) на новую границу подложка–зародыш (характеризуемую $\sigma_{2п}$) в форме круга радиусом, равным $r \sin\theta$. Появление функции $\Phi(\theta)$ в выражении (5.14.6) обусловлено уравнением Юнга (5.14.1).

Суммируя равенства (5.14.4) и (5.14.6), получаем искомое выражение для изобарного потенциала образования *трехмерного зародыша* (отмечаемого здесь и далее нижним индексом 3):

$$\Delta G_3(r) = \left(-\frac{4\pi r^3}{3} \Delta g_v + 4\pi r^2 \sigma_{12} \right) \Phi(\theta). \quad (5.14.7)$$

Из условия $(\partial \Delta G_3 / \partial r)|_{r=r^*} = 0$ находим радиус кривизны трехмерного сферического зародыша критических размеров:

$$r_3^* = \frac{2\sigma_{12}}{\Delta g_v}. \quad (5.14.8)$$

Тогда работа его образования ΔG_3^* вычисляется путем подстановки (5.14.8) в выражение (5.14.7) (ср. формулу (5.13.19)):

$$\Delta G_3^* = \frac{16\pi}{3} \frac{\sigma_{12}^3}{\Delta g_v^2} \Phi(\theta) \equiv \begin{cases} \frac{1}{2} V_{\text{зар}}^* \Delta g_v = \frac{1}{2} |\Delta G_v^*|, \\ \frac{1}{3} A_{\text{зар}}^* \sigma_{12} = \frac{1}{3} \Delta G_s^*. \end{cases} \quad (5.14.9)$$

Здесь тождественные равенства содержат объем ($V_{\text{зар}}^*$) и площадь поверхности ($A_{\text{зар}}^*$) куполообразного зародыша критических размеров, введенные на основе выражений (5.14.4) и (5.14.6):

$$V_{\text{зар}}^* = \frac{4\pi r_3^{*3}}{3} \Phi(\theta) \quad \text{и} \quad A_{\text{зар}}^* = 4\pi r_3^{*2} \Phi(\theta). \quad (5.14.10)$$

Для последующего обсуждения удобно, используя выражения (5.14.8) и (5.14.9) для r_3^* и ΔG_3^* , представить изобарный потенциал зародышеобразования (5.14.7) в следующей форме:

$$\Delta G_3(\bar{r}) = \Delta G_3^* (3 - 2\bar{r}) \bar{r}^2, \quad (5.14.11)$$

где введен нормированный радиус зародыша $\bar{r} = r/r_3^*$.

Сравнение формул (5.13.18) и (5.14.8) показывает, что критический радиус зародыша является одинаковым при гомогенном и гетерогенном зародышеобразовании ($r_{\text{гом}}^* = r_3^*$). Таким образом, условия взаимодействия подложки и зародышевой фазы, характеризующие краевым углом θ , не влияют на критический радиус.

Однако работа образования критического зародыша ΔG_3^* зависит от краевого угла через функцию $\Phi(\theta)$ и, согласно формулам (5.13.19) и (5.14.9), связана с $\Delta G_{\text{ГОМ}}^*$ соотношением:

$$\Delta G_3^* = \Delta G_{\text{ГОМ}}^* \Phi(\theta) \leq \Delta G_{\text{ГОМ}}^*. \quad (5.14.12)$$

На рис. 5.50 показана зависимость контактного угла θ от разности поверхностных натяжений $(\sigma_{1\text{п}} - \sigma_{2\text{п}})$ на границе подложки с первичной и зародышевой фазами, построенная по формуле Юнга (5.14.1). Здесь с использованием «жидкостной» терминологии, заимствованной из капиллярной теории, выделены области полного несмачивания (при $\theta = 180^\circ$) и полного смачивания (при $\theta = 0^\circ$). В верхней части рисунка изображены формы зародышевой фазы на поверхности подложки в этих областях, включая промежуточную область с краевыми углами, лежащими в диапазоне $0^\circ < \theta < 180^\circ$.

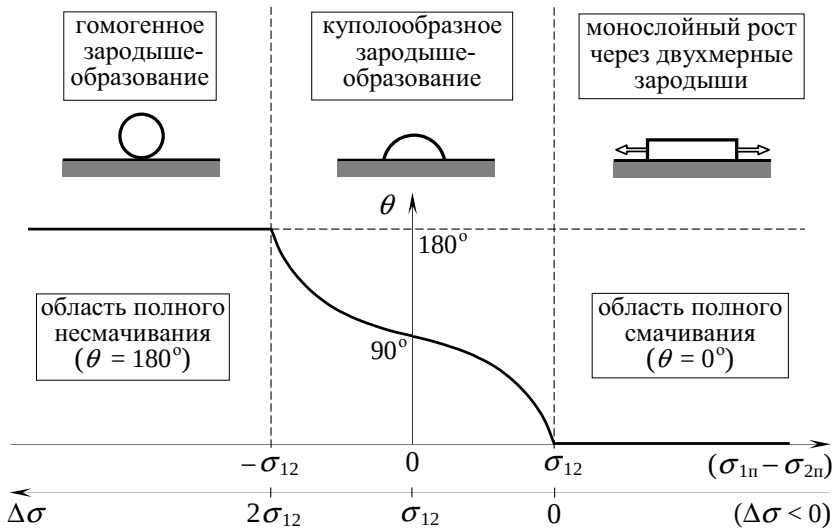


Рис. 5.50. Зависимость контактного угла θ от разности поверхностных натяжений $(\sigma_{1\text{п}} - \sigma_{2\text{п}})$ на границе подложки с первичной и вторичной фазами с указанием различных областей зародышеобразования

Как следует из рис. 5.49, при $\theta = 180^\circ$ (в области полного несмачивания, где $\sigma_{2\text{п}} \geq \sigma_{1\text{п}} + \sigma_{12}$) имеем $\Phi(\theta) = 1$. Это обеспечивает максимальную (при данном пересыщении) работу образования критического зародыша $\Delta G_{3,\text{max}}^* = \Delta G_{\text{ГОМ}}^*$. В этом случае подложка не оказывает влияния на рост зародыша, т. е. ситуация аналогична гомогенному зародышеобразованию, происходящему

в объеме первичной фазы. Как и там, критический зародыш на подложке имеет форму сферы, объем которой, согласно (5.14.10), максимален и равен $V_{\text{зар,мах}}^* = 4\pi r_3^{*3}/3$.

В области промежуточных контактных углов $0^\circ < \theta < 180^\circ$ (когда $|\sigma_{1п} - \sigma_{2п}| < \sigma_{12}$) имеем $\Phi(\theta) < 1$. Это понижает работу образования критического зародыша по сравнению с гомогенным зародышеобразованием, поскольку, согласно формуле (5.14.12), $\Delta G_3^* = \Delta G_{\text{гом}}^* \Phi(\theta) < \Delta G_{\text{гом}}^* = \Delta G_{3,\text{мах}}^*$. Следовательно, подложка оказывает каталитическое воздействие на рост трехмерных зародышей, обеспечивая гетерогенное зародышеобразование.

Каталитическое влияние оказывают также поверхностные несовершенства подложки в форме ступеней (рис. 5.51). Как показывает анализ, работа образования критического зародыша в 90-градусной ступени дается, как и ранее, той же формулой (5.14.9) с заменой функции $\Phi(\theta)$ на новую функцию $\Psi(\theta)$, изображенную на рис. 5.49. Отсюда видно, что $\Psi(\theta) < \Phi(\theta)$ при всех краевых углах, а при $\theta < 45^\circ$ имеем $\Psi(\theta) \equiv 0$. Это означает, что величина $\Delta G_3^* = \Delta G_{\text{гом}}^* \Psi(\theta)$ для ступени всегда меньше, чем величина $\Delta G_3^* = \Delta G_{\text{гом}}^* \Phi(\theta)$ для плоской поверхности. Более того, при краевых углах менее 45° вообще отсутствует энергетический барьер для роста зародышей в угле ступени.

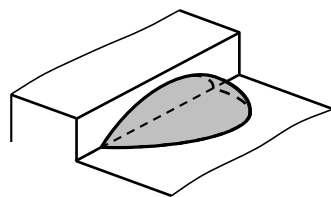


Рис. 5.51. Куполообразный зародыш в угле ступени на поверхности подложки

Следовательно, ступень более благоприятна для роста зародышей, нежели плоская граница. Этот эффект используется для так называемого *химического декорирования* с целью визуального наблюдения поверхностных несовершенств. Для этого применяют, например водный раствор AgNO_3 , из которого на поверхности кристалла в районе дефектов типа ступеней, границ блоков и выходов дислокаций осаждается серебро в форме макроагрегатов, видимых в обычном микроскопе.

Как следует из рис. 5.49, при $\theta = 0^\circ$ (в области полного смачивания, где $\sigma_{1п} \geq \sigma_{2п} + \sigma_{12}$) функция $\Phi(\theta)$, а вместе с ней и $\Delta G_3^* = \Delta G_{\text{гом}}^* \Phi(\theta)$ обращаются в нуль. Это соответствует полной «растекаемости» зародышевой фазы по подложке в виде монослоя, образование которого может происходить только по механизму двумерного роста. Рассматриваемые ниже особенности двумерного зародышеобразования создают основы для последующего анализа трехмерных многослойных зародышей.

2. Модель монослойного зародыша в форме плоского диска, квадрата и гексагона (шестиугольника), изображенных на рис. 5.52, применима при учете кристаллографической анизотропии растущего слоя. Анизотропия проявляется в различии поверхностного натяжения в плоскостях, параллельной и перпендикулярной подложке ($\sigma_{12} \neq \sigma_{\perp}$). Площади поверхностей зародыша, соответствующих этим плоскостям, и его объем обозначены на рис. 5.52 следующим образом:

$$A_{\parallel} = \alpha r^2, \quad A_{\perp} = \beta r a, \quad V_{\text{зар}} = \alpha r^2 a. \quad (5.14.13)$$

где a и r — моноатомная высота и характерный размер зародыша, равный радиусу диска и полудиagonали квадрата или гексагона. Факторы формы α и β для этих фигур даны в таблице рис. 5.52.

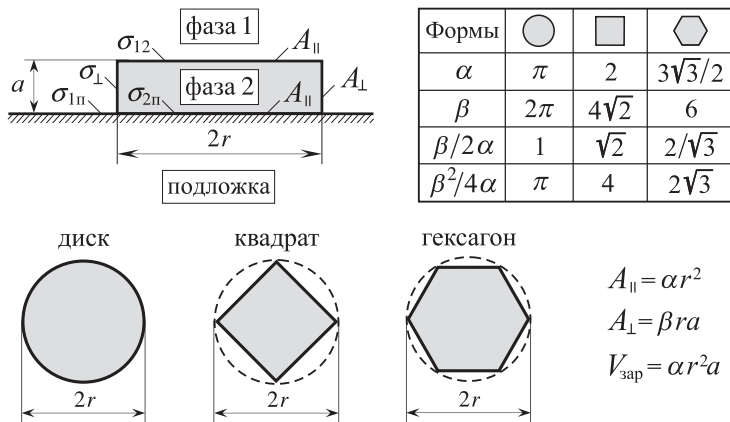


Рис. 5.52. Двумерные зародыши моноатомной высоты a с характерным размером $2r$ в форме диска, квадрата и гексагона. Факторы формы для вычисления площади поверхности и объема зародыша приведены в таблице

Выражения для объемной и поверхностной энергии образования двумерного монослойного зародыша имеют следующий вид:

$$\Delta G_v = -\Delta g_v V_{\text{зар}}, \quad (5.14.14)$$

$$\Delta G_s = \sigma_{12} A_{\parallel} + (\sigma_{2\text{п}} - \sigma_{1\text{п}}) A_{\parallel} + \sigma_{\perp} A_{\perp}. \quad (5.14.15)$$

Суммируя равенства (5.14.14) и (5.14.15) с учетом (5.14.13), получаем искомое выражение для изобарного потенциала образования *двумерного зародыша* (отмечаемого здесь и далее нижним индексом 2):

$$\Delta G_2(r) = -\alpha(\Delta g_v a - \Delta\sigma)r^2 + \beta\eta r. \quad (5.14.16)$$

Здесь $\eta = \sigma_{\perp} a$ — удельная свободная энергия моноатомной ступени (Дж/м), а также введено обозначение

$$\Delta\sigma = (\sigma_{12} + \sigma_2) - \sigma_{1\text{п}}. \quad (5.14.17)$$

Величина $\Delta\sigma$ представляет собой избыточную (над $\sigma_{1\text{п}}$) поверхностную энергию, запасаемую в верхней и нижней гранях зародыша, параллельных плоскости подложки ($A_{\parallel}$). На рис. 5.50 $\Delta\sigma$ взята в качестве горизонтальной оси, направленной противоположно по отношению к исходной оси ($\sigma_1 - \sigma_2$). Значение $\Delta\sigma = 0$ соответствует эпитаксиальному зародышу, который по структуре полностью совпадает с материалом подложки, так что $\sigma_{12} = \sigma_{1\text{п}}$ и $\sigma_{2\text{п}} = 0$ (см. рис. 5.53 а). Значения $\Delta\sigma < 0$ соответствуют области полного смачивания на рис. 5.50, когда подложка облегчает образование монослойного зародыша, так как $\Delta g_{va} - \Delta\sigma = \Delta g_{va} + |\Delta\sigma|$. При $\Delta\sigma > 0$ зародыш находится в области частичного смачивания и для его образования надо повышать пересыщение Δg_{va} , чтобы скомпенсировать вклад $\Delta\sigma$.

Из условия $(\partial\Delta G_2/\partial r)|_{r=r^*} = 0$ и выражения (5.14.16) находим критический размер и работу образования *двумерного зародыша*:

$$r_2^* = \frac{\beta}{2\alpha} \frac{\eta}{\Delta g_{va} - \Delta\sigma}, \quad (5.14.18)$$

$$\Delta G_2^* = \frac{\beta^2}{4\alpha} \frac{\eta^2}{\Delta g_{va} - \Delta\sigma}. \quad (5.14.19)$$

Для последующего обсуждения удобно, используя выражения (5.14.18) и (5.14.19) для r_2^* и ΔG_2^* , представить изобарный потенциал зародышеобразования (5.14.16) в следующей форме:

$$\Delta G_2(\bar{r}) = \Delta G_2^* (2 - \bar{r}) \bar{r}, \quad (5.14.20)$$

где введен нормированный радиус зародыша $\bar{r} = r/r_2^*$.

Основные различия между формулами (5.14.9) и (5.14.19), выражающими работу образования трехмерного и двумерного зародышей, заключаются в следующем: а) разная зависимость от пересыщения Δg_v ($\Delta G_3^* \propto 1/\Delta g_v^2$ и $\Delta G_2^* \propto 1/\Delta g_v$); б) наличие в знаменателе формулы (5.14.19) величины $\Delta\sigma$ вида (5.14.17), которая в зависимости от знака может увеличивать или уменьшать работу образования двумерного зародыша.

Монослойный рост через двумерные зародыши реализуется при так называемом *автоэпитаксиальном* (или гомоэпитаксиальном) *наращивании* монокристаллической пленки на ориентированной подложке того же материала, например кремния на кремнии (кубическая сингония), как изображено на рис. 5.53.

Рассмотрим частные случаи формулы (5.14.19) для когерентного и некогерентного срастания зародыша кремния с кристаллической решеткой подложки.

При *когерентном срастании* (рис. 5.53 а) атомы зародыша на границе раздела совершенно точно подходят к аналогичным атомам подложки и растет ориентированный зародыш, строго продолжающий кристаллографическую ориентацию подложки. В этом случае граница между зародышем и подложкой, как таковая, исчезает (что изображено пунктиром на рис. 5.53 а), в результате чего $\sigma_{2п} = 0$ и $\sigma_{12} = \sigma_{1п}$. Тогда работа образования *ориентированного зародыша* равна (при $\beta^2/4\alpha = 4$):

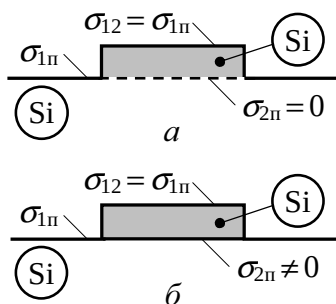


Рис. 5.53. Схема когерентного (а) и некогерентного (б) срастания Si-зародыша с кремниевой подложкой при гомоэпитаксии

$$\Delta G_{2, \text{ориен}}^* = \frac{4\eta^2}{\Delta g_v a}. \quad (5.14.21)$$

При *некогерентном срастании* (рис. 5.53 б) на границе раздела возникает небольшая кристаллографическая разориентация зародыша по отношению к подложке. В этом случае $\sigma_{2п} \neq 0$, но по-прежнему $\sigma_{12} = \sigma_{1п}$, поскольку внешняя фаза практически не чувствует малую разориентацию кристалла. Тогда работа образования *разориентированного зародыша* равна

$$\Delta G_{2, \text{разор}}^* = \frac{4\eta^2}{\Delta g_v a - \sigma_{2п}}. \quad (5.14.22)$$

Из сравнения формул (5.14.9), (5.14.21) и (5.14.22) можно сделать следующие выводы, основываясь на выражении (5.13.10) для вероятности образования критического зародыша, которая возрастает с уменьшением критического пересыщения ΔG^* .

1. При относительно небольших пересыщениях Δg_v , но таких, что $\Delta g_v a > \sigma_{2п}$, знаменатель в выражении (5.14.22) меньше, чем в выражении (5.14.21). Следовательно, рост разориентированных зародышей менее вероятен, так как $\Delta G_{2, \text{разор}}^* > \Delta G_{2, \text{ориен}}^*$. Более того, с уменьшением пересыщения возможно обратное неравенство, $\Delta g_v a < \sigma_{2п}$, так что величина $\Delta G_{2, \text{разор}}^*$ становится отрицательной. Это означает, что при малых пересыщениях разориентированные зародыши растворяются в первичной фазе, а на подложке кристаллизуются только ориентированные зародыши, разрастающиеся в *монокристаллическую пленку*.

2. Увеличение пересыщения может обеспечить неравенство $\Delta g_v a \gg \sigma_{2п}$, при котором $\Delta G_{2,разор}^* \approx \Delta G_{2,ориен}^*$. Следовательно, образование как ориентированных, так и разориентированных зародышей становится равновероятным. Это нарушает процесс эпитаксиального роста и вместо монокристаллической пленки образуется *поликристаллический* слой.

3. При еще большем пересыщении возможна смена двумерного механизма зародышеобразования на трехмерный. Действительно, как уже было отмечено, для этих механизмов наблюдается разная степенная зависимость от пересыщения Δg_v ($\Delta G_3^* \propto 1/\Delta g_v^2$ и $\Delta G_2^* \propto 1/\Delta g_v$), так что с ростом Δg_v возможно осуществление неравенства $\Delta G_3^* < \Delta G_2^*$. Обычно трехмерный рост реализуется при так называемом *гетероэпитаксиальном наращивании* монокристаллической пленки на ориентированной подложке из другого материала, например кремния на сапфире.

Следует отметить, что на величину ΔG^* существенным образом влияет состояние поверхности подложки — наличие ступеней, изломов ступеней, выходов дислокаций, адсорбированных атомов примесей. Несовершенства поверхности (ступени и дислокации) обычно являются катализаторами зародышеобразования, а адсорбированные примеси могут как активировать, так и отравлять центры кристаллизации.

Если в основе двумерного монослойного зародышеобразования лежит механизм мономолекулярной адсорбции Ленгмюра, то полимолекулярная адсорбция по механизму БЭТ порождает многослойное зародышеобразование на поверхности подложки.

3. Модель многослойного зародыша в форме цилиндра, прямоугольного параллелепипеда (с квадратным основанием) и гексагональной призмы изображена на рис. 5.54. Здесь высота h зародыша определяется числом слоев l ($h = la$, $l = 1, 2, \dots$).

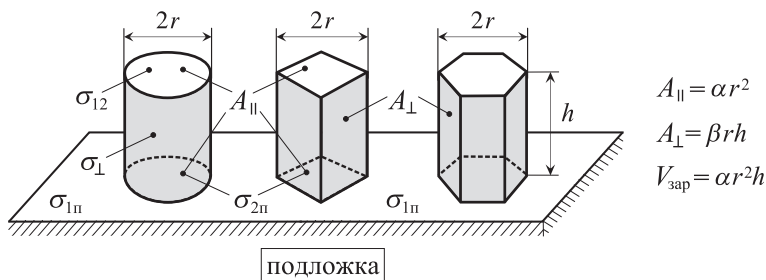


Рис. 5.54. Многослойные зародыши в форме цилиндра, прямоугольного параллелепипеда и гексагональной призмы высотой h с характерным размером $2r$ поперечного сечения. Факторы формы α и β приведены в таблице на рис. 5.52

На этом основании площади поверхности и объем зародыша вычисляются по формулам (5.14.13) с заменой $a \rightarrow h = la$:

$$A_{\parallel} = \alpha r^2, \quad A_{\perp} = \beta r h, \quad V_{\text{зар}} = \alpha r^2 h. \quad (5.14.23)$$

Свободная энергия образования многослойного зародыша, подобно выражениям (5.14.14) и (5.14.15), содержит три вклада:

- *объемная энергия*, вызванная пересыщением в системе,

$$\Delta G_v = -\Delta g_v V_{\text{зар}} = -\alpha \Delta g_v r^2 h; \quad (5.14.24)$$

- *поверхностная энергия*, связанная с верхней и нижней гранями зародыша,

$$\Delta G_{s\parallel} = \sigma_{12} A_{\parallel} + (\sigma_{2\text{п}} - \sigma_{1\text{п}}) A_{\parallel} = \alpha \Delta \sigma r^2; \quad (5.14.25)$$

- *поверхностная энергия*, связанная с боковой поверхностью зародыша,

$$\Delta G_{s\perp} = \sigma_{\perp} A_{\perp} = \beta \sigma_{\perp} r h. \quad (5.14.26)$$

Здесь были использованы использованы выражения (5.14.23), а также формула (5.14.17), то есть $\Delta \sigma = (\sigma_{12} + \sigma_2) - \sigma_{1\text{п}}$.

Суммируя равенства (5.14.24)–(5.14.26), получаем искомое выражение для *изобарного потенциала образования многослойного зародыша*:

$$\Delta G(r, h) = -\alpha \Delta g_v r^2 h + \alpha \Delta \sigma r^2 + \beta \sigma_{\perp} r h. \quad (5.14.27)$$

Вычисление частных производных функции (5.14.27) дает

$$\frac{\partial \Delta G(r, h)}{\partial r} = -2\alpha (\Delta g_v h - \Delta \sigma) r + \beta \sigma_{\perp} h, \quad (5.14.28)$$

$$\frac{\partial \Delta G(r, h)}{\partial h} = -(\alpha \Delta g_v r - \beta \sigma_{\perp}) r. \quad (5.14.29)$$

Из равенства нулю выражения (5.14.28) получаем

$$r_h = \frac{\beta}{2\alpha} \frac{\sigma_{\perp} h}{\Delta g_v h - \Delta \sigma}, \quad (5.14.30)$$

где нижний индекс h означает условие $h = \text{const}$, при котором вычислена производная (5.14.28).

Нетрудно видеть, что при $h = a$ формула (5.14.30) приводит к выражению (5.14.18) для критического радиуса r_2^* двумерного монослойного зародыша, то есть $r_a = r_2^*$, так как $\sigma_{\perp} a = \eta$.

Подстановка значения (5.14.30) в производную (5.14.29) дает

$$\frac{\partial \Delta G(r_h, h)}{\partial h} = \beta \frac{\sigma_{\perp} r_h}{2} \frac{\Delta g_v h - 2\Delta \sigma}{\Delta g_v h - \Delta \sigma}. \quad (5.14.31)$$

Обращение в нуль производной (5.14.31) приводит к выражению для *высоты критического зародыша*:

$$h^* = \frac{2\Delta\sigma}{\Delta g_v}. \quad (5.14.32)$$

Формула (5.14.30) при $h = h^*$ дает характерный *поперечный размер критического зародыша*, то есть $r^* = r_{h^*}$, а именно

$$r^* = \frac{\beta}{\alpha} \frac{\sigma_{\perp}}{\Delta g_v}. \quad (5.14.33)$$

Подстановкой (5.14.32) и (5.14.33) в уравнение (5.14.27) получаем *работу образования критического зародыша*:

$$\Delta G^* \equiv \Delta G(r^*, h^*) = \frac{\beta^2}{\alpha} \left(\frac{\sigma_{\perp}}{\Delta g_v} \right)^2 \Delta\sigma. \quad (5.14.34)$$

Исходя из формул (5.14.24)–(5.14.26) и (5.14.32)–(5.14.34), легко убедиться в том, что

$$\Delta G^* \equiv \alpha \Delta\sigma r^{*2} = \Delta G_{s\parallel}(r^*), \quad (5.14.35)$$

$$\Delta G_v(r^*, h^*) + \Delta G_{s\perp}(r^*, h^*) = 0. \quad (5.14.36)$$

Таким образом, активационный барьер зародышеобразования ΔG^* , даваемый формулами (5.14.34) и (5.14.35), определяется для критического зародыша только вкладом его верхнего и нижнего оснований ($\Delta G_{s\parallel}(r^*, h^*)$). Согласно (5.14.36), вклад боковой поверхности критического зародыша ($\Delta G_{s\perp}(r^*, h^*)$) полностью скомпенсирован объемным пересыщением ($\Delta G_v(r^*, h^*)$).

На основании соотношений (5.14.34)–(5.14.36) можно двояко выразить работу образования критического зародыша, а именно

$$\Delta G^* \equiv \frac{\beta^2}{\alpha} \left(\frac{\sigma_{\perp}}{\Delta g_v} \right)^2 \Delta\sigma = \begin{cases} \frac{1}{2} |\Delta G_v^*|, \\ \frac{1}{3} \Delta G_s^*, \end{cases} \quad (5.14.37)$$

где $\Delta G_v^* = \Delta G_v(r^*, h^*)$ и $\Delta G_s^* = \Delta G_{s\parallel}(r^*) + \Delta G_{s\perp}(r^*, h^*)$.

Из сравнения формулы (5.14.37) с аналогичными выражениями (5.13.19) и (5.14.9), ранее полученными, соответственно, для сферического зародыша и куполообразного зародыша, следует общее заключение: работа образования критического зародыша любой формы (по крайней мере из числа анализируемых) *всегда* равна половине модуля объемного пересыщения в первичной фазе ($|\Delta G_v^*|/2$) и одной трети свободной энергии, запасенной поверхностью критического зародыша ($\Delta G_s^*/3$).

Из выражений (5.14.32) и (5.14.33) следует, что отношение высоты и поперечного размера критического зародыша,

$$H \equiv \frac{h^*}{r^*} = \frac{\alpha}{\beta} \frac{2\Delta\sigma}{\sigma_{\perp}}, \quad (5.14.38)$$

не зависит от пересыщения Δg_v и для любой формы зародыша определяется соотношением поверхностных натяжений $\Delta\sigma$ и $\sigma_{\perp}$.

Для анализа разрастания *закритического зародыша* удобно ввести его нормированные размеры ($\bar{h} = h/h^* \geq 1$ и $\bar{r} = r/r^* \geq 1$) и переписать выражение (5.14.27) для изобарного потенциала образования многослойного зародыша в следующем виде:

$$\Delta G(\bar{r}, \bar{h}) = \Delta G^*(\bar{r} - 2\bar{r}\bar{h} + 2\bar{h})\bar{r}. \quad (5.14.39)$$

Поверхностные энергии, запасенные верхней и нижней гранями зародыша и его боковой поверхностью, даваемые формулами (5.14.25) и (5.14.26), принимают следующий вид:

$$\Delta G_{s\parallel} = \Delta\sigma A_{\parallel} = \Delta G^* \bar{r}^2, \quad (5.14.40)$$

$$\Delta G_{s\perp} = \sigma_{\perp} A_{\perp} = 2\Delta G^* \bar{r} \bar{h}. \quad (5.14.41)$$

В процессе разрастания закритического зародыша отношение его высоты к характерному поперечному размеру равно

$$\frac{h}{r} = H \frac{\bar{h}}{\bar{r}}, \quad (5.14.42)$$

где величина H была введена формулой (5.14.38).

Возможны три варианта разрастания закритического зародыша путем увеличения размеров $h \geq h^*$ и/или $r \geq r^*$, показанные на рис. 5.55 и анализируемые ниже.

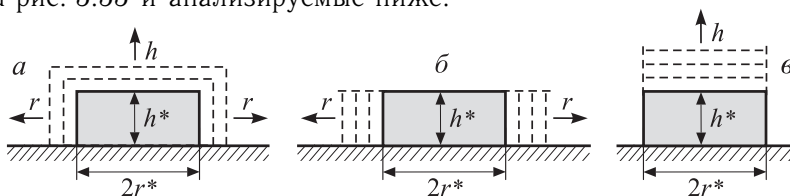


Рис. 5.55. Варианты разрастания закритического зародыша: а — трехмерное (объемное) разрастание, б — двумерное (латеральное) разрастание, в — одномерное (вертикальное) разрастание

Трехмерное (объемное) разрастание (рис. 5.55 а) имеет место, когда *сохраняется* изначальная форма критического зародыша, задаваемая отношением h^*/r^* в форме (5.14.38), то есть

$$\frac{h}{r} = \frac{h^*}{r^*} \equiv H = \text{const}. \quad (5.14.43)$$

Из сравнения формул (5.14.42) и (5.14.43) получаем основное свойство трехмерного (объемного) разрастания:

$$\bar{h} = \bar{r}. \quad (5.14.44)$$

Равенство (5.14.44) приводит к следующему соотношению между поверхностными энергиями (5.14.40) и (5.14.41) объемно разрастающегося трехмерного зародыша:

$$\Delta G_{s\perp} = 2\Delta G_{s\parallel}. \quad (5.14.45)$$

Подстановка (5.14.44) в формулу (5.14.39) дает выражение для свободной энергии *трехмерного зародыша* (отмечаемого индексом 3),

$$\Delta G_3(\bar{r}) = \Delta G^*(3 - 2\bar{r})\bar{r}^2, \quad (5.14.46)$$

в точности совпадающее с аналогичным выражением (5.14.11), полученным ранее для куполообразного сферического зародыша.

Двумерное (латеральное) разрастание (рис. 5.55 б) имеет место, когда размеры критического зародыша таковы, что

$$h^* \ll r^* \quad \text{или} \quad H \ll 1, \quad (5.14.47)$$

и по мере разрастания зародыша *сохраняется* изначальная критическая высота, то есть (ср. формулу (5.14.44))

$$h = h^* = \text{const} \quad \text{или} \quad \bar{h} = 1. \quad (5.14.48)$$

Подстановка (5.14.48) в формулу (5.14.39) дает выражение для свободной энергии *двумерного зародыша* (отмечаемого индексом 2),

$$\Delta G_2(\bar{r}) = \Delta G^*(2 - \bar{r})\bar{r}, \quad (5.14.49)$$

в точности совпадающее с аналогичным выражением (5.14.20), полученным ранее для монослойного роста.

Согласно определению H в форме (5.14.38), условие (5.14.47) реализуется при

$$\sigma_{\perp} \gg \Delta\sigma. \quad (5.14.50)$$

Отсюда следует, что энергетически невыгодно вертикальное (по нормали к подложке) разрастание зародыша с увеличением его боковой поверхности, обладающей бóльшим поверхностным натяжением. Действительно, согласно (5.14.40) и (5.14.41), $\Delta G_{s\perp} = \sigma_{\perp} A_{\perp} \propto \bar{r}$ (при $\bar{h} = 1$) и $\Delta G_{s\parallel} = \Delta\sigma A_{\parallel} \propto \bar{r}^2$. Между поверхностными энергиями латерально разрастающегося зародыша сохраняется следующее соотношение (ср. формулу (5.14.45)):

$$2\Delta G_{s\parallel} = \Delta G_{s\perp}. \quad (5.14.51)$$

Следовательно, быстрее разрастаются параллельные подложке грани зародыша с площадью $A_{\parallel} = \alpha r^2$, в силу (5.14.50) дающие меньший вклад в увеличение свободной энергии системы.

Одномерное (вертикальное) разрастание (рис. 5.55 в) имеет место, когда размеры критического зародыша таковы, что

$$r^* \ll h^* \quad \text{или} \quad H \gg 1, \quad (5.14.52)$$

и по мере разрастания зародыша *сохраняется* изначальный критический поперечный размер, то есть (ср. формулу (5.14.48))

$$r = r^* = \text{const} \quad \text{или} \quad \bar{r} = 1. \quad (5.14.53)$$

Условие (5.14.52), в противоположность предыдущему случаю ($\bar{h} = 1$), реализуется при (ср. формулу (5.14.50))

$$\Delta\sigma \gg \sigma_{\perp}. \quad (5.14.54)$$

Следовательно, в этом случае энергетически невыгодно латеральное разрастание по той же ранее высказанной причине. Это обеспечивает выполнение равенства (5.14.53), которое приводит к следующему соотношению между поверхностными энергиями вертикально разрастающегося зародыша (ср. формулу (5.14.51)):

$$\Delta G_{s\perp} = 2\Delta G_{s\parallel} \bar{h}. \quad (5.14.55)$$

Подстановка равенства (5.14.53) в формулу (5.14.39) дает свободную энергию *одномерно* разрастающегося вертикального зародыша (отмечаемого индексом 1):

$$\Delta G_1(\bar{h}) = \Delta G^* = \text{const}. \quad (5.14.56)$$

Следовательно, при $r = r^*$ увеличение высоты зародыша $\bar{h} > 1$ не приводит к изменению свободной энергии, оставляя ее на вершине барьера зародышеобразования ΔG^* , равного (5.14.34).

С целью сравнения трех рассмотренных выше вариантов разрастания зародыша для них на рис. 5.56 приведен изобарный потенциал зародышеобразования в нормированном виде $\Delta G/\Delta G^*$ как функция характерного нормированного размера зародыша. Кривые построены по формулам (5.14.46), (5.14.49) и (5.14.56), соответственно, для объемного (трехмерного) роста при $\bar{h} = \bar{r}$, латерального (двумерного) роста при $\bar{h} = 1$ и вертикального (одномерного) роста при $\bar{r} = 1$. Штрих-пунктирная вертикальная линия соответствует критическому зародышу, слева от которой располагаются докритические зародыши ($\bar{r} < 1$ и $\bar{h} < 1$), а справа — закритические зародыши ($\bar{r} > 1$ и $\bar{h} > 1$).

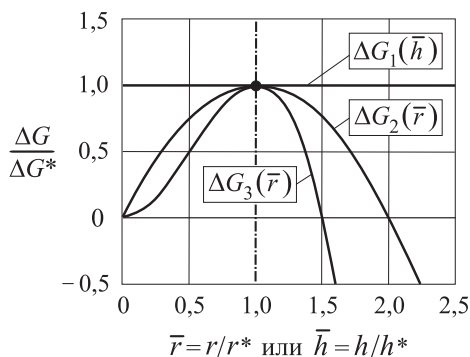


Рис. 5.56. Зависимость нормированной свободной энергии зародышеобразования от характерного нормированного размера зародыша для трех вариантов разрастания: одномерное разрастание ($\Delta G_1(\bar{h})$ при $\bar{r} = 1$); двумерное разрастание ($\Delta G_2(\bar{r})$ при $\bar{h} = 1$); трехмерное разрастание ($\Delta G_3(\bar{r})$ при $\bar{h} = \bar{r}$)

Трехмерное зародышеобразование возможно лишь при строгом соотношении между поверхностными натяжениями $\Delta\sigma$ и $\sigma_\perp$ в виде (5.14.38), что и определяет форму зародыша ($h/r = H$). Трехмерное и двумерное зародышеобразование происходит путем преодоления термоактивационного барьера высотой ΔG^* , в результате чего начинается и развивается спонтанный рост закритического зародыша.

В отличие от этого, для вертикального (одномерного) роста, как показывает прямая линия $\Delta G_1(\bar{h})$ на рис. 5.56, построенная по уравнению (5.14.56), нет необходимости в преодолении потенциального барьера ΔG^* : необходимо лишь удерживать зародыш на вершине этого барьера. Такая ситуация может реализоваться, во-первых, по механизму полимолекулярной адсорбции БЭТ (см. п. 5.11) при низкой поверхностной диффузии, когда падающие из газовой фазы молекулы напрямую питают вертикально растущую грань кристалла. Во-вторых, близкая ситуация возникает при росте так называемых *вискеров* (от англ. *whisker*), механизм формирования которых описан в следующем параграфе.

Если при вертикальном росте диффузия по поверхности подложки должна быть пренебрежимо малой, то латеральное разрастание зародыша, наоборот, эффективно развивается за счет диффузионного присоединения атомов к движущейся ступени.

Рассмотренные варианты разрастания критического зародыша представляют собой частные модельные случаи, работающие при $\bar{h} = \bar{r}$ (трехмерный рост), $\bar{h} = 1$ (двумерный рост) и $\bar{r} = 1$ (одномерный рост). Понятно, что в реальных условиях

должны наблюдаться отклонения от этих предельных моделей. В первую очередь, это касается модели одномерного разрастания, происходящего при $r = r^* = \text{const}$ и $\Delta G_1(\bar{h}) = \Delta G^* = \text{const}$. Такое состояние является метастабильным. Действительно, достаточно латерального присоединения нескольких атомов к периферии вертикально растущего зародыша, чтобы обеспечить его скатывание с вершины потенциального барьера ΔG^* и переход к комбинированному (вертикально-латеральному) разрастанию.

В заключение этого раздела кратко рассмотрим кинетические соотношения, характеризующие процесс зародышеобразования.

4. Кинетика зародышеобразования. Скорость зародышеобразования определяют как скорость разрастания критических зародышей в результате двух возможных механизмов:

а) за счет падения частиц из первичной питающей фазы непосредственно на поверхность зародыша (вертикальный рост);

б) за счет присоединения к периферии зародыша адсорбированных атомов (адатомов), перемещающихся по поверхности подложки путем диффузионных скачков (латеральный рост).

На начальном этапе кристаллизации, когда критические зародыши покрывают малую часть поверхности, доминирующим является второй механизм, зависящий от коэффициента поверхностной диффузии адатомов. В этом случае скорость зародышеобразования определяется как

$$v_{\text{зар}} = Z_{\alpha} N_{\alpha}^* \omega^* [1/(\text{м}^2 \cdot \text{с})]. \quad (5.14.57)$$

Здесь N_{α}^* — концентрация флуктуационно возникших критических зародышей, содержащих α^* частиц, ω^* — частота присоединения адсорбированных атомов к критическому зародышу, $Z_{\alpha} = (\Delta G^*/3\pi k_B T \alpha^{*2})^{1/2}$ — фактор неравновесности Зельдовича, обычно имеющий значение от 10^{-2} до 10^{-1} . Число частиц в критическом зародыше находится как $\alpha^* = V^*/\Omega$, где V^* — объем зародыша, вычисляемый для трехмерного и двумерного зародышей по формулам (5.14.10) и (5.14.13), а $\Omega = v^k/N_{\alpha} \equiv M/\rho^k N_{\alpha}$ — объем конденсата, приходящийся на одну частицу.

Вероятность флуктуационного возникновения критических зародышей на адсорбционных центрах описывается больцмановской формулой (5.13.10), поэтому их концентрация равна

$$N_{\alpha}^* = N_{\alpha} \exp\left(-\frac{\Delta G_{\alpha}^*}{k_B T}\right), \quad (5.14.58)$$

где $N_{\alpha} = 1/a_0^2$ — плотность адсорбционных центров на подложке, расположенных на расстоянии a_0 друг от друга ($N_{\alpha} \approx 10^{15} \text{ см}^{-2}$).

В стационарном состоянии концентрация адсорбированных атомов не меняется во времени ($dN_a/dt = 0$). В соответствии с уравнением (5.10.17), стационарная концентрация адатомов определяется из равенства плотностей потока адсорбируемых частиц, $J_{\text{адс}} \equiv dN_a^+/dt$, и десорбируемых частиц, $J_{\text{дес}} \equiv dN_a^-/dt = = N_a/\tau_a$, где использовано выражение (5.10.21). Среднее время жизни атомов в адсорбированном состоянии дается формулой (5.10.20), которую записывают в более общей форме,

$$\tau_a = \tau_0 \exp\left(\frac{\Delta G_{\text{дес}}}{k_B T}\right), \quad (5.14.59)$$

путем введения $\Delta G_{\text{дес}}$, свободной энергии активации процесса десорбции. Отсюда получаем выражение для стационарной концентрации адсорбированных на подложке атомов:

$$N_a = J_{\text{адс}} \tau_a = J_{\text{адс}} \tau_0 \exp\left(\frac{\Delta G_{\text{дес}}}{k_B T}\right). \quad (5.14.60)$$

Кроме тепловых колебаний в направлении нормали к поверхности, рассмотренных в п. 5.9, адатомы совершают также диффузионные скачки по поверхности от одного адсорбционного центра к другому. Частота таких скачков может быть по аналогии с выражением (5.14.59) представлена в форме

$$\nu_{\text{диф}} \equiv \frac{1}{\tau_{\text{диф}}} = \frac{1}{\tau_0} \exp\left(-\frac{\Delta G_{\text{диф}}}{k_B T}\right), \quad (5.14.61)$$

где $\Delta G_{\text{диф}}$ — свободная энергия активации процесса диффузии. Именно эти диффузионные скачки с частотой (5.14.61) обеспечивают присоединение адатомов к зародышам.

Частота ω^* присоединения адатомов к критическому зародышу равна числу адатомов $N_a l a_0$, примыкающих к зародышу по его периферийной линии длиной l , умноженному на частоту диффузионных скачков $\nu_{\text{диф}}$:

$$\omega^* = N_a l a_0 \nu_{\text{диф}} = J_{\text{адс}} l a_0 \exp\left(\frac{\Delta G_{\text{дес}} - \Delta G_{\text{диф}}}{k_B T}\right), \quad (5.14.62)$$

где использованы формулы (5.14.60) и (5.14.61).

Подстановка (5.14.58) и (5.14.62) в формулу (5.14.57) дает искомое выражение для скорости зародышеобразования:

$$v_{\text{зар}} = Z_{\alpha} J_{\text{адс}} \frac{l}{a_0} \exp\left(\frac{\Delta G_{\text{дес}} - \Delta G_{\text{диф}} - \Delta G_{\alpha}^*}{k_B T}\right). \quad (5.14.63)$$

Здесь длина l периферийной линии критического зародыша равна $2\pi r_2^*$ для двумерного зародыша (рис. 5.52) и $2\pi r_3^* \sin \theta$ для трехмерного сферического зародыша (рис. 5.48).

Как видно из (5.14.63), температурная зависимость скорости роста зародышей определяется конкуренцией процессов десорбции, поверхностной диффузии и зародышеобразования. При достаточно высоких температурах диффузионное перемещение адатомов в виде дискретных скачков может сменяться их движением в форме двумерного газа. В этом случае возрастает частота ω^* присоединения адатомов к зародышу, что увеличивает скорость образования центров кристаллизации. С понижением температуры может нарушаться равновесие между паром и адсорбционным слоем из-за увеличения времени τ_a жизни атомов в адсорбированном состоянии, а также равновесие в самом адсорбционном слое из-за снижения поверхностной подвижности адатомов.

Рост зародышей за счет поверхностной диффузии возможен только при высоких температурах подложки. При низких температурах можно считать, что адатомы прочно связаны с адсорбционными центрами, поверхностная диффузия малозначительна, и рост зародышей происходит по механизму прямого осаждения атомов из первичной фазы.

В общем случае результирующая скорость роста лимитируется наиболее медленным из вышеупомянутых процессов. Установление равновесия между физически адсорбированными атомами и первичной фазой обычно происходит достаточно быстро (в течение микросекунд). Поэтому в реальных условиях рост пленок контролируется либо процессами массопереноса в первичной фазе (при кристаллизации из жидкой фазы и химическом осаждении из газовой фазы), либо поверхностной диффузией (при физическом осаждении из атомно-молекулярных пучков).

Приведенные выше рассуждения предполагали, что исходная первичная фаза является однокомпонентной. В реальных системах формирование новой фазы происходит в среде, содержащей достаточно большое число различных веществ в соизмеримых количествах, причем каждое из них может войти в состав зародыша. Однако в процессе зародышеобразования происходит молекулярный отбор, в результате которого зародышами становятся кластеры, содержащие только те компоненты, которые мало различаются по структуре. Аналогичный отбор реализуется и в процессе дальнейшего роста зародышей. Поэтому если первичная фаза оказывается пересыщенной по многим компонентам, то в ней образуется смесь разнородных кластеров. Поскольку кластеры зарождаются одновременно, то практически всегда

(исключая выращивание монокристаллов) в пересыщенных средах образуются коллективы частиц разного размера, а формирующаяся фаза, как правило, полидисперсна.

5.15. Механизмы формирования и роста пленок на подложках

Формирование тонких пленок на поверхности подложек наиболее часто происходит в две стадии:

а) *стадия образования зародышей*, рассмотренная в предыдущем параграфе, на которой возникают критические зародыши, способные к дальнейшему росту;

б) *стадия роста пленки*, на которой критические зародыши разрастаются с образованием сплошной пленки.

Именно таким образом (в два этапа, начиная с зародышей) происходит процесс роста по следующим трем механизмам гетерогенного формирования пленок на идеальных подложках:

- островковый механизм роста (*модель Фольмера–Вебера*),
- послойный механизм роста (*модель Франка–Ван-дер-Мерве*),
- послойно-островковый механизм роста (*модель Странского–Крастанова*).

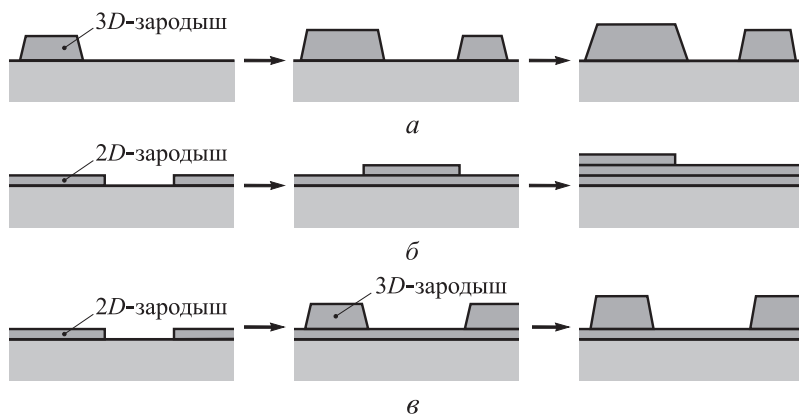


Рис. 5.57. Схематическое изображение трех основных механизмов роста пленок: а — островковый рост (модель Фольмера–Вебера), б — послойный рост (модель Франка–Ван-дер-Мерве), в — послойно-островковый рост (модель Странского–Крастанова). Стрелки показывают последовательность роста

Механизмы формирования и роста пленок на идеальных подложках, изображенные на рис. 5.57, будут рассмотрены ниже. Начнем рассмотрение с роста пленок на дефектных подложках.

1. Особенности роста пленок на несовершенных подложках. На шероховатой или дефектной поверхности могут работать специфические механизмы роста, не требующие начального образования зародышей. Как уже говорилось в предыдущем параграфе, моноатомная ступень (см. рис. 5.51) и, тем более, излом ступени являются местом безактивационного роста. Именно таким образом (без зародышеобразования) осуществляются *ступенчато-слоевой* и *спирально-дислокационный* механизмы роста на несовершенных подложках.

Ступенчато-слоевой рост реализуется при наличии на поверхности подложки ступеней, источником которых служит, в частности, естественная шероховатость граней с большими индексами Миллера. Эти грани проявляются в виде совокупности атомных ступеней, образованных участками плотноупакованных плоскостей с малыми индексами, как показано на рис. 5.58. При низких температурах фронт ступеней является атомно-гладким. Тепловые флуктуации, имеющие место при достаточно высоких температурах, приводят к возникновению изломов в ступенях.

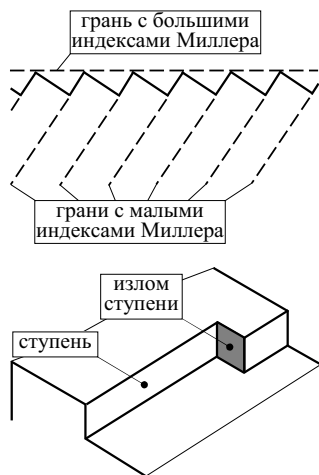


Рис. 5.58. Совокупность моноатомных ступеней на грани кристалла с большими индексами Миллера и излом моноатомной ступени

При ступенчато-слоевом механизме отсутствует необходимость в образовании зародышей, так что процесс роста пленки состоит из следующих последовательно протекающих стадий, схематически показанных на рис. 5.59 а:

- 1) *адсорбция частиц* первичной фазы на поверхности подложки в виде адатомов;
- 2) *поверхностная диффузия* адатомов по террасе по направлению к ступени с закреплением в ее угле;
- 3) *движение атомов* вдоль ступени с окончательным закреплением их в изломе ступени.

Поскольку ступень образует двугранный угол, а излом — трехгранный, то атом В (рис. 5.59 а), находящийся в изломе, сильнее связан с подложкой, чем атом Б, расположенный в угле ступени, а последний сильнее, чем атом А, адсорбированный гладкой поверхностью. В процессе осаждения каждая ступень на шероховатой поверхности подложки последовательно застраивается частицами, непосредственно поступающими из

первичной фазы, минуя стадию зародышеобразования. Результатом последовательного застраивания и перемещения ступеней является новый атомный слой. Перечисленные выше стадии показаны в виде стрелок ①, ②, ③ на рис. 5.59 а.

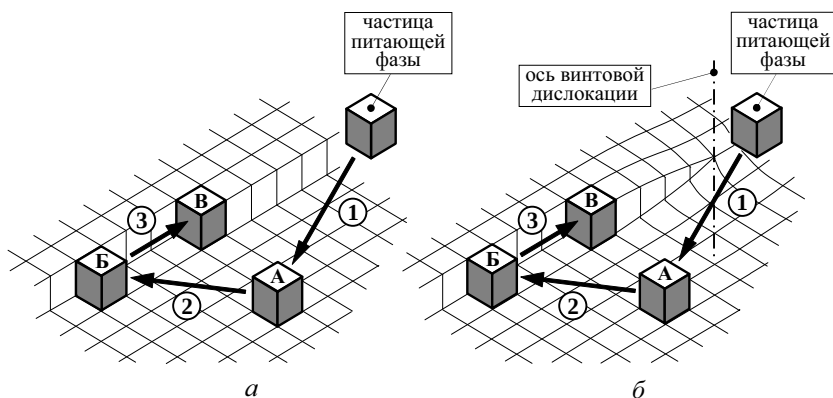


Рис. 5.59. Схема осаждения частицы первичной фазы на подложку с монокристаллической ступенью (*ступенчатый рост*) (а) и с винтовой дислокацией (*спиральный рост*) (б): ① — адсорбция, ② — поверхностная диффузия, ③ — движение вдоль ступени; А — атом на поверхности подложки, В — атом в двугранном угле ступени, В — атом в трехгранном угле излома ступени

Так как на гранях с большими индексами Миллера при любой температуре существуют ступени, показанные на рис. 5.58, то все они полностью не зарастают. Процесс роста этих граней идет непрерывно при любых, даже очень малых, пересыщениях. Более того, колебания внешних условий (температуры или концентрации осаждаемых частиц в питающей фазе) разрушают идеальные атомно-гладкие грани, придавая им атомно-шероховатую ступенчатую структуру. Это проявляется, в частности, при кристаллизации из жидкой фазы, когда даже рост идеальных граней с малыми индексами Миллера идет послойно, минуя стадию образования зародышей.

Несингулярная (диффузная) грань, схематически изображенная на рис. 5.14 а, всегда атомно-шероховатая, поскольку содержит большое число стохастически распределенных изломов. На несингулярной грани зародыши не возникают, а рост происходит путем присоединения атомов ко всем ступенькам и изломам одновременно. Такой механизм называют *нормальным ростом*, так как застраивание всей совокупности ступеней происходит по нормали к поверхности. Осаждающиеся атомы присоединяются

к атомам подложки практически в любой точке своего падения, находя там энергетически выгодное место для закрепления.

При нормальном росте зависимость скорости формирования слоев от переохлаждения носит линейный характер. Это является следствием статистической независимости актов присоединения и отрыва атомов на изломах. В процессе роста ступенчатая поверхность перемещается как целое, параллельно самой себе, вдоль нормали к исходной сингулярной грани. Это в принципе отличается от тангенциального движения ступени при послыном росте на идеальной грани по модели Франка–Ван-дер-Мерве, рассматриваемой ниже (см. рис. 5.57 б).

Спирально-дислокационный рост реализуется на подложках, имеющих на своей поверхности спиральные ступени с изломами в местах выхода винтовых дислокаций. Как показали эксперименты, в ряде случаев рост пленок из парогазовой фазы даже на гранях с малыми индексами Миллера становится уже заметным при весьма малых пересыщениях (порядка одного процента), когда зародышевый механизм практически исключен. Исследования подтвердили решающую роль винтовых дислокаций в подобных экспериментальных ситуациях.

Как показано на рис. 5.59 б (см. также рис. 3.15 б и относящийся к нему текст), выход винтовой дислокации создает на поверхности подложки ступень с изломом, исчезающую в процессе ее спирального застраивания. Наличие исчезающей ступени делает ненужным образование зародышей, так что рост слоев происходит при любых, даже очень малых, пересыщениях.

Последовательные стадии спирально-дислокационного роста совпадают с тремя рассмотренными выше стадиями ступенчатослоевого роста (рис. 5.59 а) и изображены на рис. 5.59 б такими же стрелками с цифрами ①, ②, ③. Однако, если при слоевом росте ступень перемещается параллельно самой себе, то при спиральном механизме она вращается вокруг оси винтовой дислокации. В результате чего образуется спиральная поверхность роста, сужение которой заканчивается так называемой пирамидой роста с вершиной на оси дислокации.

Рассмотрим специфические особенности эпитаксиальных процессов, которые будут необходимы для понимания механизмов роста пленок на идеальных подложках (см. рис. 5.57), в первую очередь, механизма Странского–Крастанова.

2. Особенности гетероэпитаксиального роста. Под эпитаксией понимают процесс ориентированного наращивания на подложке монокристаллического слоя, продолжающего в процессе

роста кристаллографическую ориентацию подложки (см. сноску на с. 29). В настоящее время принято различать *гомоэпитаксию* (называемую также *автоэпитаксией*) — ориентированное наращивание монокристаллической пленки на подложке из того же материала, и *гетероэпитаксию* — ориентированное наращивание пленки на чужеродной подложке. По мере наращивания пленки гетероэпитаксия сменяется гомоэпитаксией. Особенности гомоэпитаксии будут рассмотрены ниже при изложении послойного механизма Франка–Ван-дер-Мерве.

Термин «эпитаксия» впервые ввел Руайе и сформулировал правило, согласно которому эпитаксиальный рост возможен лишь в том случае, когда срастающиеся кристаллы имеют одинаковый тип химической связи и являются изоструктурными с расхождением периода решеток не более, чем на 15%. Последующие эксперименты показали, что даже при несоответствии периодов решетки вплоть до 90–100% возможно эпитаксиальное сращивание кристаллов различной кристаллографической симметрии, однако, с дефектами структуры в районе границы раздела.

Позднее Ван-дер-Мерве предложил модель образования псевдоморфных слоев, частично объясняющую результаты экспериментов. Под *псевдоморфизмом* понимают индуцированное подложкой изменение периода решетки в результате упругой деформации структуры пленки. Вынужденная деформация обеспечивает сопряжение кристаллической решетки пленки с решеткой подложки, что и называют псевдоморфизмом. Отклонение от собственной кристаллографической симметрии с изменением периода решетки, с одной стороны, снижает межфазную энергию границы пленка–подложка, а с другой стороны, увеличивает энергию механических напряжений, возникающих за счет упругой деформации пленки. Противоборство этих двух факторов заметным образом проявляется в пределах некоторого переходного слоя, названного *псевдоморфным слоем*.

Если несоответствие периодов решеток относительно невелико (в среднем не превышает 1–2%), то пленка упруго деформируется так, что кристаллы срастаются *когерентно*. В этом случае атомы, расположенные по разные стороны границы раздела, точно подходят один к одному, а в растущей пленке возникает упруго-напряженный псевдоморфный слой (ПС на рис. 5.60 а). В пределах этого слоя происходит плавный переход от периода решетки подложки к периоду решетки объемного материала пленки с соответствующим уменьшением упругих деформаций. По толщине ПС простирается на несколько периодов решетки, в отличие от показанной на рис. 5.60 толщины в два периода.

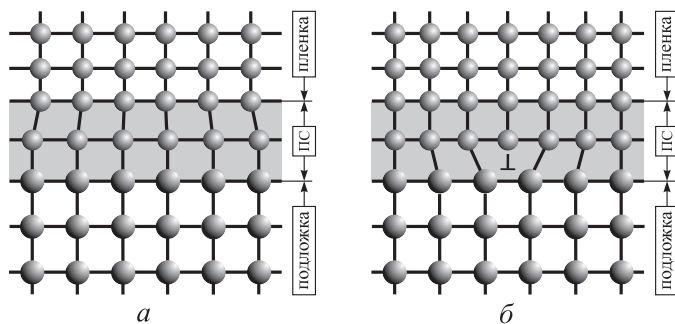


Рис. 5.60. Схема образования упруго-напряженного псевдоморфного слоя (ПС), выделенного серым цветом (а), и дислокации несоответствия $\perp$ (б) при гетероэпитаксиальном росте пленки на монокристаллической подложке

При большом несоответствии периодов двух решеток (в среднем более 4–5%) энергия упругих деформаций пленки настолько увеличивается, что когерентное срастание кристаллов становится энергетически невыгодным. В этом случае образуется *полукогерентная граница*, на которой зарождаются краевые дислокации (см. рис. 3.15 а), называемые **дислокациями несоответствия** (ДН), одна из которых показана на рис. 5.60 б и обозначена знаком $\perp$. Такие дислокации лежат в плоскости псевдоморфного слоя и, действуя совместно с остаточной упругой деформацией в этом слое, компенсируют различие периода решеток между растущей пленкой и подложкой.

Таким образом, путем генерации дислокаций несоответствия осуществляется «сброс» избыточной упругой энергии (накопленной в напряженном псевдоморфном слое), который называют *упругой релаксацией*. В результате этого упруго-деформированная структура с ПС превращается в дефектную структуру с ДН. Однако подобные дефектные структуры в большинстве случаев являются нежелательными для приборных применений.

В середине 80-х годов прошлого века с применением новейших методов исследования поверхности (в частности, СТМ и ДБЭ) было обнаружено, что в сильно рассогласованных гетероструктурах существует еще один механизм упругой релаксации (сброса упругих напряжений в ПС). Он реализуется путем роста по механизму Фольмера–Вебера (рис. 5.57 а) смачивающего (упруго-напряженного) слоя, который стимулирует и питает рост на своей поверхности трехмерных островков без генерации ДН. Такой механизм, носящий имя Странского–Крастанова и называемый *послойно-островковым* (рис. 5.57 в), будет описан ниже.

Переходим к рассмотрению сущности трех механизмов эпитаксиального роста пленок на совершенных подложках, схематически изображенных на рис. 5.57.

3. Особенности островкового роста пленок по механизму Фольмера–Вебера. В основе механизма островкового роста (см. рис. 5.57 а) лежит зарождение и разрастание трехмерных зародышей на атомно-гладких плотноупакованных гранях совершенного кристалла. Таковыми гранями являются сингулярные грани, имеющие малые индексы Миллера (см. рис. 5.13 а).

Необходимым условием реализации зародыше-островкового механизма роста является выполнение неравенства $\Delta\sigma > 0$, или $\sigma_{1п} < \sigma_{2п} + \sigma_{12}$ (см. рис. 5.50, формулу (5.14.17) и текст после нее). В этом случае атомы осаждаемого вещества связаны между собой сильнее, чем с подложкой, смачивание поверхности незначительно, и энергетически более выгодно формирование изолированных друг от друга трехмерных зародышей, разрастающихся далее в сплошную пленку. Зародыши возникают и растут при степени покрытия подложки веществом первичной фазы значительно меньшей, чем необходимо для образования монослоя. Сплошными пленки становятся лишь при определенной толщине и имеют, как правило, зернистую структуру.

Другим условием образования трехмерных зародышей является необходимость достаточно высокого пересыщения. Реально вероятность образования зародышей, а вместе с ней и скорость роста пленок, ничтожно малы вплоть до пересыщений, достигающих единиц и даже десятков процентов.

В основе образования и разрастания трехмерных зародышей в сплошную пленку лежат следующие элементарные процессы:

1) *массоперенос частиц* в первичной фазе (при помощи атомно-молекулярного пучка и диффузионного или конвективного потока), определяющий доставку осаждаемого вещества на поверхность подложки и растущего слоя;

2) *адсорбция частиц* первичной фазы (одиночных атомов и молекул), которые являются поверхностным источником, питающим процесс зародышеобразования;

3) *поверхностная диффузия атомов*, определяющая транспорт частиц к докритическим, критическим и закритическим зародышам с превращением их в центры кристаллизации.

Адсорбированные частицы, если их кинетическая энергия после столкновения с поверхностью остается достаточно большой, могут десорбироваться, то есть покинуть поверхность и возвратиться в первичную фазу. Однако чаще потеря энергии за счет

взаимодействия с молекулярным полем сил поверхностного притяжения оказывается большой. Тогда оставшейся энергии достаточно лишь для двумерного термически активированного движения частицы по поверхности в форме диффузионных скачков из одного адсорбционного положения в другое.

Результирующая скорость роста пленки лимитируется наиболее медленным из этих трех процессов. Установление равновесия между физически адсорбированными атомами и первичной фазой обычно происходит достаточно быстро (в течение микросекунд, см. пп. 5.9 и 5.10). Поэтому в реальных условиях рост пленок контролируется либо процессами массопереноса в первичной фазе (при кристаллизации из жидкой фазы и химическом осаждении из газовой фазы), либо поверхностной диффузией (при физическом осаждении из атомно-молекулярных пучков).

Рассмотрим два этапа эволюции гетероэпитаксиального роста пленок по островковому механизму Фольмера–Вебера.

Этап независимого роста зародышей и островков состоит в том, что адатомы, размещенные на поверхности с концентрацией N_a , взаимодействуют между собой, сталкиваясь в процессе диффузии, и образуют *кластеры* (скопления атомов), состоящие из α частиц. Метастабильные кластеры, находящиеся в равновесии с адсорбционным слоем, образуют *критические зародыши* с энергией образования ΔG_α^* , концентрация которых, согласно (5.14.58), равна $N_\alpha^* = N_\alpha \exp(-\Delta G_\alpha^*/k_B T)$. Критические зародыши разрастаются за счет адатомов, присоединяющихся к ним в результате диффузии. По мере разрастания зародыши превращаются в *островки*, форма и размеры которых зависят от свойств подложки, зародышевой фазы и условий роста.

На начальной стадии концентрация адатомов (N_a) заметно больше концентрации островков (N_o), при этом $N_a \propto \theta$ и $N_o \propto \theta^3$, где θ — степень покрытия поверхности. Следовательно, возникновение нового островка более вероятно, чем рост уже существующих за счет присоединения адатомов. Так как на этой стадии размеры островков намного меньше размеров свободных от них зон (называемых диффузионными зонами), то ростом первых за счет прямого попадания атомов из питающей фазы можно пренебречь. В ходе осаждения величина N_o увеличивается и приближается к N_a . При этом для адатома вероятности встретиться с другим адатомом или с краем островка становятся соизмеримыми. В итоге концентрация адатомов достигает максимума и начинает уменьшаться ($N_a \propto \theta^{-1/3}$), а концентрация островков продолжает расти, но с меньшей скоростью ($N_o \propto \theta^{1/3}$).

Когда N_0 возрастает до такой степени, что среднее расстояние между островками становится близким к средней диффузионной длине атомов, концентрация островков достигает своей максимально возможной величины, что обычно наблюдается при покрытии $\theta = 0,1-0,4$ монослоя (МС). Дальнейшее осаждение материала приводит исключительно к увеличению размера островков и к уменьшению N_0 в ходе их агрегации.

Этап укрупнения островков заключается в увеличении их среднего размера за счет уменьшения количества как результат следующих процессов, сопровождающих агрегацию островков:

- *миграция* кластеров по поверхности подложки,
- *коалесценция* (слияние) кластеров при соприкосновении,
- *остwaldовское созревание* (рост) больших кластеров за счет исчезновения малых.

Поверхностная миграция кластера небольших размеров происходит в виде перемещения его центра масс, что позволяет описывать ее таким же образом, как и поверхностную диффузию одиночных атомов. Установлено, что увеличение размера кластеров сопровождается уменьшением их подвижности и монотонным увеличением энергии активации поверхностной диффузии. В зависимости от конкретных условий поверхностная диффузия островков может происходить по различным механизмам, которые принято подразделять на две основные группы:

- а) индивидуальные механизмы, основанные на движении одиночных атомов, образующих островок (механизм последовательных перемещений, механизм испарения–конденсации и др.);
- б) коллективные механизмы, представляющие собой одновременное перемещение группы атомов (дислокационный механизм, механизмы сдвига, скольжения, переползания и др.).

Коалесценция островков возникает при непосредственном соприкосновении их границ в ходе миграции и одновременного разрастания. Форма образующихся после слияния островков определяется подвижностью краевых атомов, которая зависит от сил сцепления между частицами конденсата и подложки. Если взаимодействие между островками сильнее, чем с подложкой, и температура обеспечивает эффективное протекание диффузионных процессов, то под действием сил поверхностного натяжения происходит слияние островков.

В процессе коалесценции исчезают граничные поверхности по линии соприкосновения, уменьшаются суммарная площадь и размеры с одновременным ростом высоты островков. Они утрачивают свою первоначальную конфигурацию и образуют нано-

частицу с конфигурацией, отвечающей правилу Кюри–Вульфа (см. формулу (5.5.16)). Процесс реализуется путем растворения одних участков поверхности и роста других вплоть до установления равновесной конфигурации. На этапе коалесценции образуется перколяционная сетка зародышевой фазы на поверхности подложки с каналами большой протяженности, ограниченными кристаллографическими гранями.

Если краевая подвижность адатомов малая (при низких температурах или при затрудненном изменении формы островков из-за поверхностных дефектов и окисной пленки), то имеет место не слияние, а объединение островков, называемое *коагуляцией* (см. с. 614). В этом случае изменение исходной формы незначительно и происходит лишь в области соприкосновения, т. е. на участках с наибольшей кривизной поверхности. Две частицы срастаются друг с другом чисто механически, без изменения огранки, взаимного расположения и ориентации. В конечном итоге образующиеся пленочные покрытия часто имеют поликристаллическую структуру. Более того, при коагуляции или коалесценции соседних сильно разориентированных островков на их границах могут возникать дислокации несоответствия (см. рис. 5.60 б).

Оствальдовское созревание островков, названное по имени Оствальда (см. сноску на с. 298), наблюдается на последней стадии формирования пленки, когда адатомы как источник, питающий зародыши, практически полностью исчерпаны. В этой ситуации рост крупных островков происходит за счет более мелких в условиях так называемого «исчезновения без поедания», т. е. без процесса коалесценции.

В основе процесса лежит эффект Гиббса–Томсона, согласно которому частицы малых размеров менее устойчивы из-за большей кривизны поверхности (см. формулы (5.8.28), (5.8.30) и относящийся к ним текст). Поэтому малые островки исчезают, а крупные, потребляя их атомы, растут; при этом обмен атомами происходит с участием первичной (парогазовой и жидкой) фазы.

При кристаллизации из парогазовой фазы структурный порядок в пленках обеспечивается в основном подвижностью адатомов по подложке. Их высокая подвижность способствует преимущественному росту ориентированных зародышей. В зависимости от ориентирующих свойств подложки возникают либо монокристаллические эпитаксиальные, либо поликристаллические пленки. Когда подвижность адатомов уменьшается настолько, что атомы конденсируются непосредственно в точке их падения, практически не диффундируя, возникает большое число зароды-

шей, дающих поликристаллическую мелкозернистую или даже аморфную пленку. Такое происходит, во-первых, при низких температурах и, во-вторых, при наличии примесных атомов на подложке, стабилизирующих зародыши и снижающих подвижность адатомов. Отжиг пленки при высокой температуре превращает ее структуру в крупнозернистую.

Наряду с рассмотренным выше ПК-механизмом конденсации (пар–кристалл), в ряде случаев экспериментально наблюдается ПЖК-механизм конденсации (пар–жидкость–кристалл). Этот механизм обычно работает при достаточно высоких температурах: $T_{гр} < T < T_{пл}$. Здесь $T_{пл}$ — температура плавления объемного конденсируемого материала, $T_{гр}$ — граничная температура, соответствующая переходу от ПК-механизма конденсации к ПЖК-механизму ($T_{гр} \approx 0,68 T_{пл}$ для Bi, Sn и Pb). При этом чем выше температура подложки, тем до больших размеров сохраняется зародыш в виде жидкой фазы.

Это объясняется тем, что температура плавления малых частиц ниже температуры плавления объемного материала и приближается к ней по мере увеличения размеров (см. рис. 5.31). Поэтому на начальной стадии конденсации фазой, равновесной с паром, является жидкий расплав в форме островков. При увеличении размеров они существуют в переохлажденном состоянии в области температур $T_{гр} < T < T_{пл}$, где происходит формирование пленки по механизму пар–переохлажденная жидкость. После достижения островками величины, определяемой размерной зависимостью температуры плавления, они затвердевают, образуя поликристаллическую или даже аморфную пленку. Дальнейшая конденсация происходит в соответствии с рассмотренным выше ПК-механизмом.

В рамках термодинамической теории рядом авторов предложена диаграмма фазовых переходов жидкость–кристалл в координатах «температура–размер кластера» (рис. 5.61). Она позволяет проанализировать состояние частиц в островковых пленках как в процессе

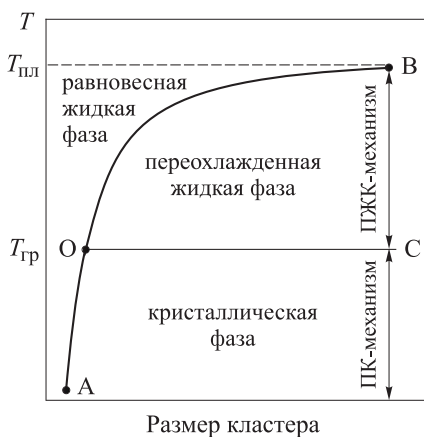


Рис. 5.61. Схематическая диаграмма фазовых переходов жидкость–кристалл для трехмерных зародышей

Она позволяет проанализировать состояние частиц в островковых пленках как в процессе

их затвердевания, так и при изменении температуры и размеров зародыша в уже закристаллизовавшихся пленках.

Кривая АОВ на рис. 5.31 соответствует размерной зависимости температуры плавления маленьких зародышей и ограничивает область существования термодинамически равновесной жидкой фазы. Область переохлажденной жидкости расположена между линиями ОС и ОВ, а область справа от линии АОС отвечает кристаллическому фазовому состоянию зародышей. Здесь $T_{гр}$ — граничная температура перехода от механизма пар-переохлажденная жидкость (при $T_{гр} < T < T_{пл}$) к ПК-механизму (при $T < T_{гр}$). Эта температура соответствует предельному переохлаждению при кристаллизации жидкой фазы на подложке и определяется ее природой.

Рост нановискеров является классическим примером реализации ПЖК-механизма. Нановискеры представляют собой нитевидные кристаллы, длина которых на порядок и более превосходит поперечные размеры (от 10 до 100 нм).

Процесс формирования вискера состоит из трех стадий. Первоначально на подготовленную поверхность подложки (например Si(111) или GaAs(111)B) наносят слой вещества-катализатора, уменьшающего поверхностную энергию (обычно слой золота толщиной около 1 нм). Затем поверхность разогревают до температуры выше эвтектической точки для смеси материала подложки с материалом катализатора (Au-Si или Au-Ga). На поверхности образуются капли жидкого раствора, находящиеся в термодинамическом равновесии с подложкой. На следующей стадии производится осаждение вещества (в рассматриваемом случае кремния или арсенида галлия), из которого и формируется далее вискер.

Насыщение раствора осаждаемым веществом происходит путем адсорбции частиц на поверхности жидкой капли непосредственно из парогазовой фазы. В результате капля жидкого раствора становится пересыщенной, и на ее границе с подложкой начинается кристаллизация и рост вискера в направлении по нормали к подложке с поперечным размером, примерно равным диаметру капли. Поскольку адсорбция продолжается, то капля жидкого раствора, расположенная на торцевой поверхности вискера, сохраняет свой размер и движется вверх (рис. 5.62). За счет

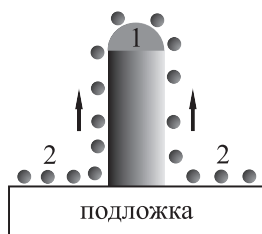


Рис. 5.62. Схема роста нановискера: 1 — капля жидкости, 2 — адатомы. Стрелками указано направление диффузии адатомов

активации поверхности рост под каплей происходит во много раз быстрее, чем на свободных участках. Приведенная на рисунке схема учитывает, во-первых, классический рост нановискеров по ПЖК-механизму за счет адсорбции атомов на полусферической поверхности капли и, во-вторых, диффузионный рост за счет поступления в каплю атомов, поднимающихся с подложки по боковой поверхности растущего вискера (что показано стрелками).

Вертикальный рост вискеров происходит в полном соответствии с моделью одномерного разрастания критического зародыша, разработанной в предыдущем параграфе (см. с. 708). При данном пересыщении Δg_v критический зародыш цилиндрической формы, образовавшийся с поперечным размером $r^* = 2\sigma_{\perp}/\Delta g_v$ (см. формулу (5.14.33) при $\beta/\alpha = 2$), примерно равным радиусу исходной жидкой капли, сохраняет его неизменным при условии (5.14.54). То есть требуется, чтобы поверхностное натяжение на боковой поверхности ($\sigma_{\perp}$) было существенно меньше величины $\Delta\sigma$, определяющей энергетический вклад торцевой грани вискера. В соответствии с (5.14.56), увеличение высоты вискера не требует затраты энергии, так как не приводит к изменению свободной энергии системы, оставляя ее на вершине барьера зародышеобразования ΔG^* , равного (5.14.34). Из вышеприведенной формулы для r^* видно, что возникновение вискера малого радиуса требует больших пересыщений Δg_v . Экспериментально установлено, что критический радиус вискера для разных материалов лежит в диапазоне от 5 до 50 нм.

Представляя собой одномерную кристаллическую систему, нановискеры, по сути, являются квантово-размерными объектами (см. п. 5.1) и могут быть практически использованы в устройствах наноэлектроники. Вискеры из двухкомпонентных полупроводниковых соединений $A^{III}B^V$ и $A^{II}B^{IV}$ (GaN, ZnO, InSb и др.) являются системами с уникальными оптическими свойствами, в которых проявляются различные квантовые эффекты. На их основе возможно создание сверхбыстрых интегральных схем с разрешением 10–50 нм, светоизлучающих устройств, устройств функциональной наноэлектроники и др. Кремниевые вискеры успешно используют в качестве ультраострого зонда кантилевров в атомно-силовой микроскопии (см. с. 58).

4. Особенности послойного роста пленок по механизму Франка – Ван-дер-Мерве. В основе механизма послойного роста (см. рис. 5.57 б) лежит зарождение и разрастание двумерных зародышей на атомно-гладких плотноупакованных гранях совершенного кристалла с малыми индексами Миллера. Модель

Франка–Ван-дер-Мерве чаще реализуется при гомоэпитаксиальном росте, когда пленка и подложка состоят из одного материала.

Необходимым условием реализации послойного механизма является выполнение неравенства $\Delta\sigma < 0$, или $\sigma_{1п} > \sigma_{2п} + \sigma_{12}$ (см. рис. 5.50, формулу (5.14.17) и текст после нее), то есть атомы осаждаемого вещества должны быть связаны с подложкой сильнее, чем друг с другом. В этом случае, когда адгезия является достаточно сильной, зародышеобразование на подложке происходит при очень малом пересыщении. При этом на идеальной плотноупакованной грани формируется двумерный зародыш критических размеров, который разрастается в монослой путем присоединения атомов к моноатомной ступени в результате поверхностной диффузии адатомов. В процессе образования монослоя пересыщение само по себе исчезает. Оно вновь становится необходимым, когда растущий атомный слой полностью заполнен и должен возникнуть новый плоский зародыш, который обычно зарождается у края грани и разрастается к ее середине.

Перемещение краев зародыша в форме движения ступени, определяющее скорость разрастания монослоя, происходит довольно быстро. Скорость роста пленки в целом лимитирована зарождением двумерного зародыша на поверхности совершенной грани, дающим начало новому монослою. Так обстоит ситуация при малых пересыщениях. Если пересыщение сильное, то рост новых плоскостей начинается еще до того, как завершится заполнение уже растущей грани, и чем больше пересыщение, тем вероятнее такой процесс. В этом случае *монослойный рост* сменяется *многослойным ростом*, а фигуры роста принимают вид лестницы со ступеньками, ведущими к центру граней.

Подобная смена видов роста зависит также и от температуры в силу конкуренции двух типов диффузии, обеспечивающих:

- *внутрислойный массоперенос* — диффузия адатомов по плоскости одной террасы;
- *межслойный массоперенос* — диффузия адатомов с террасы на террасу через ступень.

Для характеристики этих процессов на рис. 5.63 изображен потенциальный рельеф поверхности для адатома вблизи ступени. Адатом 1 на нижней террасе при столкновении со ступенью присоединяется к ней, так как при этом у него появляется большее число ближайших соседей и, следовательно, большая энергия связи (минимум А). Поэтому диффузией вверх с террасы на террасу можно пренебречь. Однако адатом 2 на верхней террасе, подойдя к краю ступени, сталкивается с дополнительным барьером

(максимум В), который превышает диффузионный барьер $\mathcal{E}_D$ на величину $\mathcal{E}_{ES}$, называемую *барьером Эрлиха–Швобеля*. Его появление вызвано тем, что при пересечении края ступени атом оказывается в положении с пониженным числом ближайших соседей.

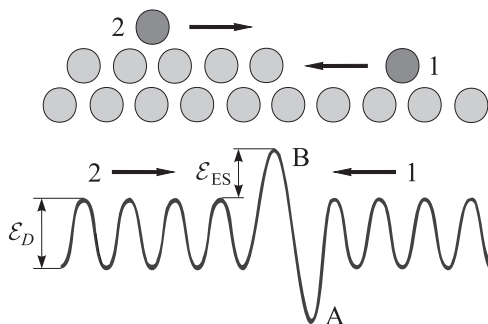


Рис. 5.63. Схематическое изображение потенциального рельефа поверхности вблизи ступени для адатомов 1 и 2, движущихся путем диффузионных скачков по нижней и верхней террасам в направлении, указанном стрелками

Вероятность перехода адатома с верхней на нижнюю террасу, то есть эффективность межслойного переноса, характеризуют коэффициентом прохождения $s = \exp(-\mathcal{E}_{ES}/k_B T)$. Если барьер Эрлиха–Швобеля очень мал или высока температура, то имеем $s \approx 1$, в противоположном случае — $s \approx 0$.

Понятно, что при достаточно высокой температуре и малом пересыщении двумерного газа адатомов их подвижность высокая, и все адатомы достигают края растущей ступени на «своей» террасе. При этом барьер Эрлиха–Швобеля не оказывает заметного влияния ($s = 1$), так что адатомы с верхней террасы «спрыгивают» на нижний зарастающий слой. Таким образом эффективно реализуется межслойный массоперенос, обеспечивая *двумерный монослойный рост* пленки (термодинамику монослойного зародышеобразования см. на с. 700–703).

Когда межслойный массоперенос подавлен ($s = 0$), что имеет место при высоком барьере Эрлиха–Швобеля и низкой температуре, адатомы не могут покинуть «своей» террасы (верхней плоскости того островка, на которой они оказались), и каждый из зародившихся островков начинает независимо разрастаться в латеральном и вертикальном направлениях, то есть имеет место *трехмерный многослойный рост* пленки (термодинамику многослойного зародышеобразования см. на с. 703–710). Индивидуальное зарождение островков при $s = 0$ происходит, если ширина террасы превышает среднюю диффузионную длину адатомов.

Для стимулирования двумерного роста в тех условиях, когда обычно имеет место трехмерный рост, нередко применяют специальные поверхностные примеси, наносимые в количестве одного или долей монослоя. Такие вещества называют *сурфактантами* (от англ. аббревиатуры *surf(ace) act(ive) a(ge)nt*, заимствованной из медицинской терминологии, что полностью эквивалентно термину «поверхностно-активные вещества»).

Атомы сурфактанта не должны растворяться в пленке, чтобы постоянно смачивать поверхность подложки, снижая поверхностное натяжение $\sigma_{2п}$ для растущего слоя. Сурфактант может оказывать двойное действие: с одной стороны, он повышает подвижность адатомов (что способствует росту монослоя за счет движения ступеней), а с другой стороны, его атомы, встраиваясь в край ступени, снижают барьер Эрлиха–Швобеля (что увеличивает межслойный массоперенос и способствует двумерному монослоистому росту).

Известным примером использования сурфактанта является эпитаксия пленки Ge на Si(111) в присутствии монослоя сурьмы. Без сурфактанта (в отсутствие Sb) на поверхности Si(111) 7×7 происходит формирование трехмерных островков Ge поверх псевдоморфного слоя Ge-Si толщиной 3 МС, то есть реализуется механизм Странского–Крастанова, рассматриваемый ниже.

5. Особенности послойно-островкового роста пленок по механизму Странского–Крастанова. Механизм роста пленок по Странскому–Крастанову (рис. 5.57 в) представляет комбинацию послойного механизма Франка–Ван-дер-Мерве (рис. 5.57 б) и островкового механизма Фольмера–Вебера (рис. 5.57 а). Этот механизм реализуется в гетероэпитаксиальных структурах при сильном рассогласовании параметров сопрягаемых решеток.

Суть механизма Странского–Крастанова состоит в спонтанном образовании трехмерных островков нанометровых размеров на поверхности предварительно сформированной тонкой однородной пленки (псевдоморфной и когерентной с подложкой), которую называют *смачивающим слоем* (СС). В результате роста трехмерных островков происходит частичное снятие упругих напряжений в *псевдоморфном* СС, возникшем *когерентно* (без дислокаций несоответствия) из-за рассогласования решеток слоя и подложки (см. рис. 5.60 и относящийся к нему текст).

Таким образом, возникновение трехмерных островков на поверхности СС является проявлением морфологической неустойчивости псевдоморфной упруго-напряженной пленки. Такая неустойчивость зависит от толщины слоя, поскольку с ее ростом

возрастает и упругая энергия, запасенная в этом слое. Толщину смачивающего слоя, соответствующую началу формирования когерентных трехмерных островков, называют *первой критической толщиной* h_1 . При $h < h_1$ имеет место двумерный рост СС, а в интервале толщин $h_1 < h < h_2$ формируются трехмерные островки. Значение h_2 , называемое *второй критической толщиной*, соответствует началу появления дислокаций несоответствия, обеспечивающих релаксацию упругих напряжений при $h > h_2$. Иными словами, спонтанное образование трехмерных островков в сильно рассогласованных структурах является первой стадией гетероэпитаксиального роста, предшествующей образованию дислокаций несоответствия (ДН).

На рис. 5.64 приведена кривая $h_2 = f(\Delta a/a)$, разделяющая две стадии роста — напряженный псевдоморфный рост (ниже кривой) и релаксированный дислокационный рост (выше кривой) для пленки $\text{Ge}_x\text{Si}_{1-x}$, выращенной на подложке Si(100) методом МЛЭ. Граница между этими стадиями задана второй критической толщиной h_2 , которая определяется долей x атомов Ge в растущей пленке $\text{Ge}_x\text{Si}_{1-x}$. Отсюда видно, что величина h_2 меняется на несколько порядков при изменении несоответствия решеток от долей процента до максимального значения $(\Delta a/a)_{\max} \approx 4,2\%$ при $x = 1$ ($a_{\text{Ge}} = 5,66 \text{ \AA}$ и $a_{\text{Si}} = 5,43 \text{ \AA}$).

Гетерогенная система германий на кремнии (Ge/Si) является наиболее исследованной экспериментально и теоретически в рамках модельных представлений Странского–Крастанова. Методами АСМ и СТМ (атомно-силовой и сканирующей туннельной микроскопии) в системе Ge/Si(001) обнаружены два типа 3D-островков, которые в литературе принято классифицировать как:

- **hut-кластеры**, то есть шалашеобразные островки (от англ. *hut* «шалаш») с прямоугольным основанием на плоскости (001) и боковой огранкой фасетками $\{105\}$ с наклоном под углом $11,3^\circ$;

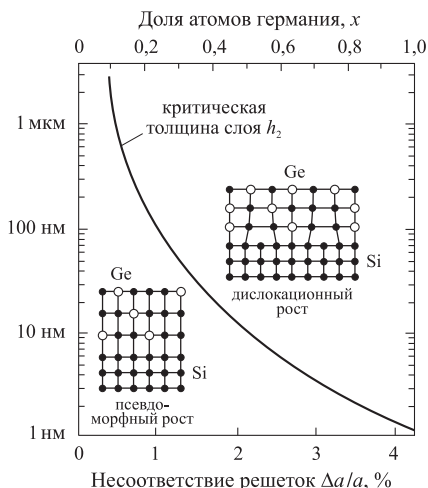


Рис. 5.64. Зависимость второй критической толщины h_2 смачивающего слоя от несоответствия решеток $\Delta a/a$ для пленки $\text{Ge}_x\text{Si}_{1-x}$ на подложке Si(100)

в частном случае квадратного основания *hut*-кластеры принимают форму *пирамиды* с равными боковыми гранями;

• **dome-кластеры**, то есть куполообразные островки (от англ. *dome* «купол») с восьмиугольным основанием на плоскости {001}, 8 боковыми гранями, образованными фасетками {102} и {113}, и 4 гранями при вершине из фасеток {105}; куполообразные островки больших размеров называют *superdome*-кластерами.

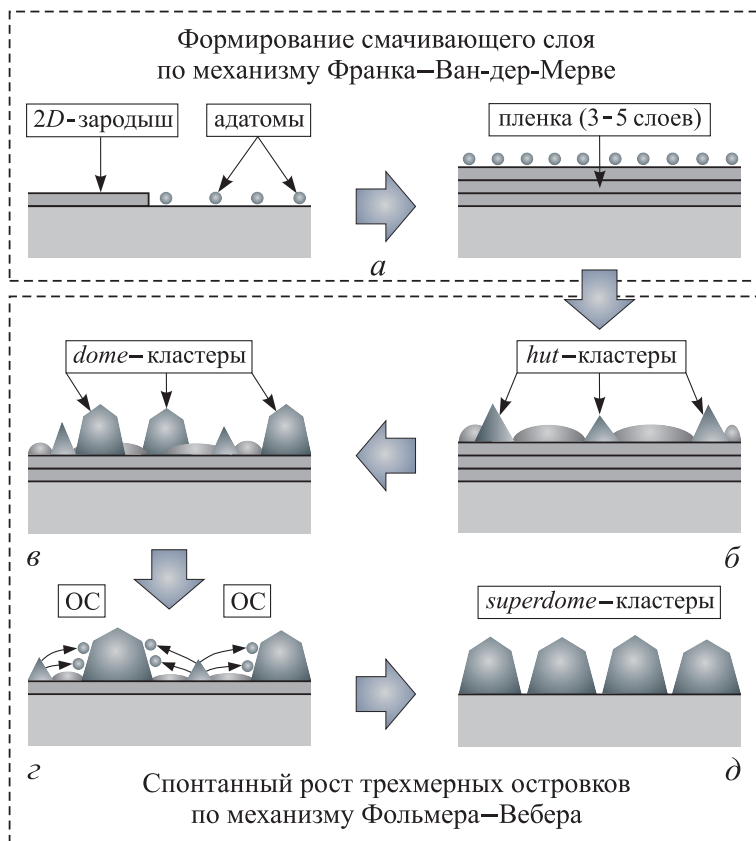


Рис. 5.65. Схема формирования когерентно напряженных трехмерных островков по механизму Странского-Крастанова в системе Ge/Si(100)

Экспериментально установлено, что эпитаксиальный рост Ge на подложке Si с ориентацией (100) происходит в несколько этапов, схематически изображенных на рис. 5.65. На первом этапе пленка Ge толщиной до 3-5 монослоев формируется путем разрастания 2D-зародышей по механизму Франка-Ван-дер-Мерве

(см. рис. 5.57 б и 5.65 а). Она находится в упруго-сжатом (из-за различия параметров a_{Ge} и a_{Si}) и когерентном (без ДН) состоянии. Именно эта пленка выполняет роль смачивающего слоя (СС), на поверхности которого имеется пересыщенный адсорбат, служащий первичным источником зарождения трехмерных *hut*-кластеров. В дальнейшем материал для роста *dome*-кластеров начинает поставлять также и СС. В результате возникает множество островков, форма, размер и количество которых зависят, прежде всего, от температуры и скорости осаждения материала.

В области низких температурах (от 350 до 500°C) образуются *hut*-кластеры с размером основания 15–20 нм и углом наклона боковых граней 11,3° (рис. 5.65 б), близкие по составу к чистому Ge. В них материал упруго релаксирован в среднем на 20%.

При более высоких температурах (550–700°C) возникают и начинают доминировать более крупные *dome*-кластеры, имеющие состав $x \approx 0,5$, с размерами основания 50–100 нм и отношением высоты к основанию около 1:4 (рис. 5.65 в). В таких островках из-за большого отношения высоты к основанию упругая релаксация достигает 50%, при этом механические напряжения снижаются от основания к вершине кластера. Несмотря на то, что высота *dome*-кластеров заметно превосходит вторую критическую толщину (как видно из рис. 5.64, $h_2 \approx 10$ нм при $x = 0,5$), они остаются когерентно сопряженными (без ДН) с подложкой.

Дальнейший рост *dome*-кластеров идет за счет поступления атомов, во-первых, от *hut*-кластеров по механизму ОС (оствадьдовского созревания, см. с. 722) и, во-вторых, от смачивающего слоя Ge, толщина которого уменьшается при переходе от рис. 5.65 б к рис. 5.65 г вплоть до полного его исчезновения на рис. 5.65 д, где показаны сверхбольшие *superdome*-кластеры. Их высота достигает 50 нм и более, что существенно превышает вторую критическую толщину h_2 . В результате этого возникают ДН, снимающие избыточную упругую энергию, накопленную в островках сверхбольших размеров.

Вышеописанный процесс гетероэпитаксии в системе Ge/Si, сопровождающийся утоньшением смачивающего слоя, характерен для роста 3D-островков лишь на поверхности (100). Иная картина наблюдается на подложке Si(111), где толщина СС остается неизменной, поскольку рост 3D-островков происходит за счет пересыщения поверхности смачивающего слоя адатомами (рис. 5.65 б), т. е. реализуется классический механизм роста.

Экспериментально было обнаружено, что энергия ступени на поверхности Si(111) в 5 раз больше, чем на поверхности (100). По-видимому, это является основной причиной, способствующей

большей атомарной гладкости и структурной устойчивости грани (111) по сравнению с поверхностью (100).

По той же причине 3D-островки Ge на поверхности Si(111) растут в форме трехгранной усеченной пирамиды с верхней гранью (111) и боковыми гранями {113}, наклоненными под углом $29,5^\circ$. При этом высота пирамиды составляет около 0,1 от поперечных размеров основания. В силу структурной устойчивости, такие кластеры по сравнению с *hut*-кластерами на поверхности Si(100), по-видимому, менее склонны к упругой релаксации. Поэтому уже на начальной стадии их образования наблюдаются дислокации несоответствия.

На сегодняшний день, кроме структуры Ge/Si(001), детально изучена также структура InAs/GaAs(001), в которой рассогласование периодов решетки составляет около 7%. Когерентно напряженные 3D-островки InAs выращивают методами МЛЭ и МОС-гидридной технологии. Смачивающий слой InAs(001) имеет первую критическую толщину $h_1 = 1,6-1,7$ МС, начиная с которой возникают 3D-островки, образующие при осаждении 4 МС плотный массив трехмерных островков. Каждый островок имеет квадратное основание со сторонами длиной 14 нм, ориентированными вдоль осей [100] и [010]. Боковые наклонные грани {101} образуют пирамиду высотой около 6 нм.

Таким образом, *движущей силой образования трехмерных островков* является **объемная упругая релаксация**, т. е. уменьшение упругой энергии в результате формирования 3D-островков по сравнению с упругой энергией когерентно напряженной (бездислокационной) эпитаксиальной пленки. Иными словами, при определенных условиях возникает морфологическая неустойчивость поверхности упруго-напряженного слоя. Это становится возможным, если выигрыш в свободной энергии за счет упругой релаксации механических напряжений в островках превышает добавку за счет увеличения площади поверхности и возникновения фасеточных граней с большой поверхностной энергией.

Не вдаваясь в детали, отметим, что термодинамическая модель процесса рассматривает систему, состоящую из подложки, эпитаксиальной пленки (смачивающего слоя) и ансамбля упруго-напряженных когерентных островков, с определенным количеством осажденного материала. При данной температуре исходят из условия баланса поверхностных энергий подложки, пленки и островков совместно с объемной энергией, включающей упругую энергию, запасенную в пленке и трехмерных островках.

С этой целью классическое уравнение (5.13.9), дающее изменение изобарного потенциала в процессе зародышеобразования,

дополняется членом, учитывающим упругую энергию:

$$\Delta G = \Delta G_v + \Delta G_s + \Delta G_{el}. \quad (5.15.1)$$

Дополнительное слагаемое ΔG_{el} (индекс *el* от англ. *elastic*) связано с упругой деформацией материала, что имеет место только в случае выращивания механически напряженных структур. При больших рассогласованиях решеток величина упругой энергии, $\Delta G_{el}(V_{\text{зар}}, h/r)$, зависит не только от объема зародыша $V_{\text{зар}}$, но и от его формы, т. е. от отношения высоты h к поперечному размеру r зародыша. Последнее является существенным при переходе от 2D-структур к 3D-структурам.

Модель рассматривает процесс формирования ансамбля трехмерных зародышей как фазовый переход первого рода. Смачивающий слой выступает в роли метастабильной конденсированной фазы, а островки представляют собой новую фазу. Мерой метастабильности является толщина смачивающего слоя: чем она больше, тем энергетически более выгодно образование трехмерных островков. Как и в ранее рассмотренных классических моделях считается, что процесс появления зародышей носит флуктуационный характер и обусловлен преодолением активационного барьера, который образуется за счет конкуренции энергетически выгодных и невыгодных вкладов в свободную энергию зародыша. Поэтому свободная энергия образования когерентного зародыша имеет максимум при определенном значении латерального размера r^* , который аналогичен критическому радиусу в моделях куполообразного сферического и плоского дискового зародыша.

Полагая, что в равновесном состоянии островки однородно распределены по размерам, ищут минимальное значение свободной энергии системы в зависимости от латерального размера: $G_{\min} = f(r)$. Затем при том же количестве осажденного материала величину $G_{\min}$ сравнивают с энергией системы при наличии лишь смачивающего слоя (G_{cc}). Если $G_{\min} < G_{\text{cc}}$, то в системе возникает устойчивый ансамбль когерентных трехмерных островков. При $G_{\min} > G_{\text{cc}}$ образование островков энергетически невыгодно и стабильным оказывается смачивающий слой.

Кроме термодинамических моделей, в литературе имеются и кинетические модели зарождения и эволюции когерентно напряженных трехмерных островков. Такие модели позволяют в условиях роста при известных энергетических параметрах системы рассчитать изменение во времени функции распределения островков по размерам и определить зависимость их структурных характеристик от управляющих параметров ростового процесса.

Спонтанное возникновение упорядоченных наноэлементов, каковыми являются массивы трехмерных когерентно напряженных островков в гетероструктурах Ge/Si и InAs/GaAs, служит ярким примером **самоорганизации** в конденсированных средах. Практическое использование принципов самоорганизации в наноэлектронной технологии создает альтернативный подход, пришедший на смену традиционной фотолитографии с присущими ей ограничениями по разрешению в наноразмерном диапазоне.

Синтез различных наноэлектронных структур путем создания технологических условий для самоорганизации формирования упорядоченных массивов квантовых точек является сегодня перспективным направлением развития оптоэлектроники. В гетеросистемах Ge/Si и InAs/GaAs получены квантовые точки в виде 3D-островков предельно малых размеров (10–100 нм) с высокой плотностью распределения по площади (10^{10} – 10^{11} см⁻²).

Существует возможность управления формированием упорядоченного массива путем создания мест преимущественного зарождения островков. Такими центрами зародышеобразования могут служить атомные ступени, если рост происходит на подложках, слегка отклоненных от кристаллографической плоскости. Этим методом удалось получить почти идеальную треугольную решетку островков Ge на висячей грани Si(100).

Упорядочение может происходить спонтанно из-за упругих сил отталкивания между островками. Эту тенденцию самопроизвольного упорядочения можно усилить, выращивая несколько слоев (один поверх другого) квантовых точек, разделенных тонкими слоями материала матрицы. Таким способом создают массивы вертикально связанных квантовых точек InAs в матрице GaAs путем поочередного осаждения 5,5 Å InAs и 15 Å GaAs.

В настоящее время, кроме трехмерных островков в рассогласованных гетероструктурах, имеются и другие типы спонтанно упорядоченных наноструктур, выращиваемых по принципу самоорганизации. Наибольший практический интерес в первую очередь представляют периодические фасетированные поверхности, позволяющие создавать упорядоченные массивы как квантовых точек, так и квантовых проволок.

6. Особенности роста наноструктур на фасетированных поверхностях. *Фасетирование поверхности* представляет собой процесс самопроизвольного образования на ней микроскопически упорядоченной структуры в виде «холмов и канавок», имеющей вполне определенный период повторения D . Существуют два типа фасетированных поверхностей:

- *вицинальные поверхности*, расположенные под небольшим углом к сингулярным граням с малыми индексами Миллера (см. рис. 5.13–5.14 и относящийся к ним текст);
- *высокоиндексные поверхности*, содержащие фасеточные грани с малыми индексами Миллера (см. рис. 5.66).

Вицинальные поверхности являются обычно атомно-гладкими и состоят из плоских террас, разделенных моноатомными ступенями. При отклонении от сингулярной грани в одном направлении ступени образуют периодическую структуру в перпендикулярном направлении (рис. 5.13 б). Ее двугранные углы служат местами образования $1D$ -нанообъектов (квантовых проволок). При отклонениях от сингулярной грани в двух направлениях ступени содержат регулярно повторяющиеся изломы (рис. 5.13 в), в трехгранных углах которых вырастают упорядоченные $0D$ -нанообъекты (квантовые точки).

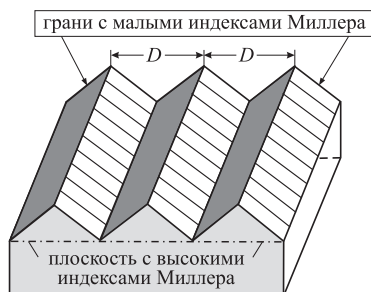


Рис. 5.66. Периодически-фасетированная поверхность с периодом D

Фасетирование высокоиндексных поверхностей экспериментально наблюдалось на $\text{AlAs}(311)$, $\text{GaAs}(311)$ и $\text{GaAs}(775)$, для вицинальных поверхностей использовали $\text{AlAs}(001)$ и $\text{GaAs}(001)$.

В основе спонтанного фасетирования плоской поверхности кристалла лежит ориентационная зависимость свободной поверхностной энергии. Если плоская граница кристалла имеет большую поверхностную энергию (что характерно для вицинальных и высокоиндексных граней), то спонтанная трансформация ее в структуру типа «холмов и канавок» (рис. 5.66) снижает общую свободную энергию, несмотря на увеличение результирующей поверхности. Морфология фасетированной структуры порождается требованием минимума поверхностной свободной энергии.

Кристаллографическая анизотропия определяет тензорную форму поверхностного натяжения, то есть создает различие между значениями σ на разных гранях кристалла. Поскольку смежные грани кристалла разделены ребром, то на каждом из них возникает скачок соответствующих компонентов тензора поверхностных натяжений. Этот скачок вносит вклад в энергию, запасенную в поле упругих деформаций, порожденных ребром как линейным источником механических напряжений в кристалле.

Качественная картина, поясняющая возникновение поля упругих сил $\mathbf{F}$, порожденных скачком поверхностного натяжения

на ребрах кристалла, разделяющих фасетки 1 и 2 с поверхностными натяжениями σ_1 и σ_2 , изображена на рис. 5.67.

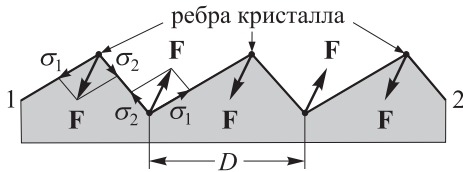


Рис. 5.67. Поле упругих сил F , порожденных скачком поверхностных натяжений σ_1 и σ_2 при переходе через ребра кристалла

Теория самопроизвольного образования периодически фасетированных поверхностей показывает, что свободная энергия поверхности содержит три вклада (в расчете на единицу площади изначально плоской поверхности),

$$\Delta G_s(D) = \Delta G_{\text{фас}}(D) + \Delta G_{\text{реб}}(D) + \Delta G_{\text{упр}}(D), \quad (5.15.2)$$

рассчитанные как функция периода D структуры. Теоретический анализ показывает, что входящие в (5.15.2) величины таковы:

- $\Delta G_{\text{фас}}(D) = \text{const}$ — поверхностная свободная энергия фасеточных граней (не зависит от периода D);
- $\Delta G_{\text{реб}}(D) = C_1/D$ — свободная энергия, запасенная ребрами;
- $\Delta G_{\text{упр}}(D) = -(C_2/D) \ln(D/a)$ — энергия упругой релаксации, порожденная ребрами как линейным источником deformаций, дающим логарифмическую функцию; a — постоянная решетки.

Поскольку энергия упругой релаксации отрицательна, то ее вклад уменьшает полную свободную энергию (5.15.2) фасетированной поверхности и делает ее минимальной при периоде структуры, равном $D_{\min} = a \exp(1 + C_1/C_2)$.

Периодически фасетированные поверхности служат подложкой для непосредственного выращивания упорядоченных массивов *квантовых проволок*, когда материал осаждается в «канавках» структуры, и *квантовых точек*, когда для этого используются изломы ступеней на вицинальных поверхностях с отклонениями от сингулярной грани в двух направлениях.

Анализ всевозможных режимов роста показал, что тип растущей структуры определяется степенью смачивания фасетированной поверхности осаждаемым материалом.

Если осаждаемый материал смачивает подложку (что имеет место в случае, когда $\Delta\sigma = (\sigma_{12} + \sigma_2) - \sigma_{1\text{п}} < 0$, см. рис. 5.50), то образуется однородное покрытие фасеточных граней, показанное на рис. 5.68 *a*. Именно таким образом осаждается AlAs на вицинальной поверхности GaAs(001), разориентированной на 3° в направлении $[110]$.

Если же осаждаемый материал не смачивает подложку, то изолированные кластеры осаждаемого материала формируются в «канавках» периодически фасетированной поверхности, образуя «тонкие кластеры» (рис. 5.68 б) или «толстые кластеры» (рис. 5.68 в). Такая ситуация экспериментально реализуется при осаждении GaAs на вицинальной поверхности AlAs(001), разориентированной на 3° в направлении $[1\bar{1}0]$, а также в случае структуры AlAs/GaAs(311). Дальнейшее наращивание приводит к образованию квантовых проволок.

В случае неоднородного покрытия периодическое фасетирование поверхности восстанавливается после осаждения нескольких монослоев. Тогда «холмы» на поверхности гетерофазной системы образуются над «канавками» подложки и наоборот (рис. 5.68 г). В этом случае возникает квантовая яма с модулированной толщиной. Например, для квантовой ямы GaAs в матрице AlAs был получен период модуляции $D = 12$ нм при амплитуде модуляции 1,2 нм.

Для террас, ограниченных со всех сторон фасетированной поверхности, возможно получение квантовых точек, поскольку адсорбированные на террасе атомы встраиваются преимущественно в трехгранные углы изломов ступени. Таким образом были получены квантовые точки InAs на вицинальной подложке GaAs(001), разориентированной в направлении $[010]$.

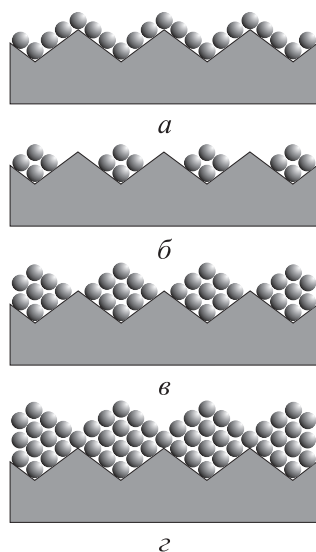


Рис. 5.68. Возможные гетерофазные структуры на фасетированной поверхности

5.16. Механизмы формирования герметичных соединений материалов

Необходимость в герметичном соединении материалов широко встречается в разнообразных задачах электронного приборостроения, в первую очередь, для создания вакуумно-плотных соединений в приборах вакуумной электроники.

Качество монолитного соединения материалов самым существенным образом зависит от физико-химических свойств соединяемых поверхностей. Ниже, не вдаваясь в детали и частности конкретных технологических процессов, кратко изложим общие

физические закономерности, лежащие в основе методов создания герметичных соединений.

Процесс создания монолитного герметичного соединения любых деталей состоит в образовании прочных химических связей между атомами соединяемых материалов. На первый взгляд кажется, что такой процесс термодинамически всегда возможен без подвода внешней энергии, так как при соединении поверхностей исчезает их поверхностная энергия, что понижает общую энергию системы. Однако в действительности это не так, и на практике требуется энергетическая активация процесса путем подвода энергии в форме либо тепла (*термическая активация*), либо упруго-пластической деформации материала (*механическая активация*), либо облучения высокоэнергетическими частицами (*радиационная активация*).

Физически процесс соединения материалов можно представить протекающим в две последовательные стадии:

1) *стадия физического контакта*, осуществляемая путем сближения соединяемых поверхностей до межатомных расстояний, при этом на границе возникает *активационный барьер*, порожденный близкодествующими силами взаимного отталкивания поверхностных атомов;

2) *стадия химического контакта*, осуществляемая путем энергетической активации за счет подвода внешней энергии, необходимой для перевода поверхностных атомов в возбужденное состояние, что делает их химически активными, то есть способными преодолеть активационный барьер с последующим образованием прочных химических связей.

Процесс герметичного соединения материалов может происходить как в жидкой фазе с нагревом, так и в твердой фазе под сильным давлением.

В первом случае физический контакт осуществляется силами *адгезии* при смачивании поверхности жидким материалом (припоем или расплавом). Термическая активация дополняется диффузией компонентов и их взаимной растворимостью в жидкой фазе. При затвердевании слабые адгезионные силы, обеспечивающие смачивание, заменяются прочными химическими связями, соответствующими природе соединяемых материалов.

Во втором случае физический контакт возникает в результате пластической деформации микронеровностей на соединяемых поверхностях при сжатии деталей. Механическая или термомеханическая активация, достигаемая давлением или давлением с нагревом, приводит к возбуждению поверхностных атомов преимущественно в местах выхода дислокаций. Эти возбужденные

атомы создают активные *центры схватывания* на соединяемых поверхностях. Между ними возникает химическое взаимодействие (ионного, ковалентного или координационно-ковалентного типа), приводящее к формированию устойчивых электронных конфигураций. Именно это и обеспечивает монолитность и герметичность соединения материалов.

Рассмотрим подробнее физическую сущность различных методов соединения материалов в жидкой и твердой фазах.

1. Соединение в жидкой фазе включает пайку припоями и различные виды сварки плавлением, среди которых выделяют газовую, аргоно-дуговую, плазменно-лучевую, электронно-лучевую и лазерную сварку.

Пайка припоями осуществляется с использованием мягких припоев ($T_{\text{пл}} < 450^\circ\text{C}$) или твердых припоев ($T_{\text{пл}} > 450^\circ\text{C}$) без расплавления соединяемых материалов.

Главным условием обеспечения герметичного соединения является *растекаемость* расплавленного припоя по обеим поверхностям. Смачивание поверхности жидкостью формирует каплю припоя в форме сферического купола, аналогичного трехмерному зародышу, показанному на рис. 5.48. Краевой угол смачивания θ определяется соотношением между поверхностными натяжениями на трех границах раздела: $\sigma_{1\text{п}} \equiv \sigma_{\text{гт}}$ (газ–твердое), $\sigma_{2\text{п}} \equiv \sigma_{\text{жт}}$ (жидкость–твердое) и $\sigma_{12} \equiv \sigma_{\text{гж}}$ (газ–жидкость), которые связаны формулой Юнга (5.14.1). Зависимость θ от $(\sigma_{\text{гт}} - \sigma_{\text{жт}})$ показана на рис. 5.50, где случаю растекаемости соответствует область полного смачивания при $\sigma_{\text{гж}} < (\sigma_{\text{гт}} - \sigma_{\text{жт}})$. Для выполнения этого неравенства необходимо так подобрать материал припоя, чтобы уменьшить сумму $(\sigma_{\text{гж}} + \sigma_{\text{жт}})$ по сравнению с величиной $\sigma_{\text{гт}}$.

Для повышения растекаемости спаиваемые детали часто покрывают гальваническим слоем меди, никеля, серебра или золота, которые, наряду с увеличением $\sigma_{\text{гт}}$, предупреждают межкристаллитную коррозию металлов. Напротив, окисные пленки на поверхности металлов снижают $\sigma_{\text{гт}}$ и тем самым препятствуют растекаемости припоя, приводя к его коагуляции (скатыванию в капли). По этой причине пайку чаще проводят не в вакууме, а в атмосфере водорода, восстанавливающего поверхностные окислы. Преимущество использования эвтектических припоев (типа медно-серебряного припоя ПСр72) для получения вакуумно-плотных спаев описано в п. 2.10.

Сварка плавлением осуществляется путем расплавления соединяемых материалов без применения припоя. Методы сварки

различаются способом подвода энергии к месту соединения с целью нагрева и активации процесса сварки.

Газовая сварка обеспечивает нагрев за счет теплоты сгорания водорода, природного или водяного газа в пламени газовой горелки. Она применяется как для сварки металлов, так и для создания вакуумно-плотных спаев металла со стеклом и керамикой, рассмотренных ниже.

Аргоно-дуговая сварка и плазменно-лучевая сварка осуществляются в атмосфере аргона. При аргоно-дуговой сварке нагрев происходит за счет тепла электрической дуги, горящей между неплавящимся вольфрамовым электродом и свариваемыми деталями. Плазменно-лучевая сварка обеспечивает нагрев плазменным лучом, вытекающим со сверхзвуковой скоростью из плазменной головки (называемой дуоплазмотроном), внутри которой сформирована высокотемпературная плазма.

Электронно-лучевая сварка и лазерная сварка используют ускоренные электроны и высокоэнергетические фотоны, энергия которых при торможении в поверхностных слоях металла передается свободным электронам, а от них — кристаллической решетке, вызывая нагрев глубинных слоев, в том числе под слоем тугоплавких окислов.

2. Соединение в твердой фазе включает разные виды сварки давлением, в том числе электроконтактную, термокомпрессионную, термодиффузионную, ультразвуковую и холодную сварку.

Электроконтактная и термокомпрессионная виды сварки занимают промежуточное положение между сваркой плавлением и сваркой давлением, так как сдавливание деталей сопровождается нагревом места соединения.

Электроконтактная сварка (точечная или роликовая) осуществляется сдавливанием металлических деталей двумя холодными медными электродами или вращающимися роликами с пропусканием импульса тока через область контакта материалов. На переходном сопротивлении контакта выделяется джоулево тепло, разогревающее эту область вплоть до ее расплавления, что приближает такой метод к сварке плавлением.

Термокомпрессионная сварка применяется для присоединения к металлизированной поверхности полупроводника алюминиевых или золотых проволок толщиной в десятки микрон. При этом надежный механический и электрический контакт между проволочным выводом и полупроводником формируется без расплавления соединяемых материалов. Разогрев места сварки до температуры $\sim 350^\circ\text{C}$ производится нагретой тонкой

иглой, часто в форме расщепленного электрода. Она прижимает вывод к контактной площадке полупроводника под давлением около 10 кг/мм^2 . Этого достаточно для пластической деформации проволоки и соединения материалов по механизму схватывания с образованием активных центров.

Термодиффузионная сварка проводится в вакууме для соединения металлических деталей (в том числе с керамикой) при температуре около $T_{\text{пл}}/2$, сжатых под давлением около 1 кг/мм^2 . Более высокие температуры и низкие давления по сравнению с термокомпрессией обеспечивают соединение материалов не по механизму схватывания через активные центры, а в результате взаимного диффузионного проникновения атомов контактирующих материалов. Относительно невысокое давление необходимо только для создания физического контакта на начальном этапе, когда принципиальную роль играет ползучесть материалов на микровыступах. Высокая температура активизирует взаимную диффузию материалов, приводящую к их монолитному и герметичному соединению.

Холодная сварка и ультразвуковая сварка выполняются под давлением без внешнего нагрева, хотя место соединения разогревается в процессе сварки до $300\text{--}500^\circ\text{C}$. Высокие давления при холодной сварке (до 100 кг/мм^2) разрушают окисные пленки, препятствующие пластическому течению материалов, и обеспечивают образование активных центров схватывания. Применение ультразвука также способствует разрушению окисных пленок, что позволяет снизить сжимающие усилия.

3. Спаи металлов со стеклом и керамикой имеют специфическую природу своего формирования. Как отмечалось в п. 4.6, основу строения стекол и керамики составляют окислы типа SiO_2 и Al_2O_3 , образующие силикатную, алюмосиликатную или алюмооксидную структуру — аморфную для стеклофазы и кристаллическую для кристаллофазы. Эти окислы взаимодействуют с окислами металлов по *донорно-акцепторному механизму*, когда один окисел (донор) отдает пару электронов, а другой (акцептор) ее принимает. Это приводит к образованию прочной координационно-ковалентной связи.

Спаи металлов со стеклом обычно требуют предварительного окисления поверхности металла в окислительном пламени газовой горелки. В результате этого обеспечивается не только стадия физического контакта из-за смачивания окисленной поверхности металла расплавленным стеклом, но и конечная стадия химического контакта. Взаимное растворение окислов, как

правило, сопровождается их донорно-акцепторным взаимодействием, при этом SiO_2 выполняет функцию акцептора, а окислы металлов — функцию донора.

Спаи металлов с керамикой осуществляют либо при помощи пайки по металлизации, либо с применением активной пайки. В обоих случаях соединение материалов происходит по донорно-акцепторному механизму, приводящему к образованию координационно-ковалентной связи.

Пайка по металлизации требует предварительного нанесения на поверхность керамики молибдено-марганцевой пасты с последующим ее вжиганием во влажном азоте. Эта операция создает окисел MnO_2 , играющий роль донора электронной пары, который взаимодействует с акцепторным окислом Al_2O_3 керамики, давая марганцевую шпинель $\text{MnO}_2 \cdot \text{Al}_2\text{O}_3$. При температуре вжигания $\sim 1400^\circ\text{C}$ частицы порошка молибдена спекаются друг с другом и с керамикой при помощи размягченной шпинели. В результате этого после охлаждения образуется прочный слой металлизации на поверхности керамики. Соединение металлов с металлизированной керамикой выполняют обычной пайкой твердым или мягким припоем. При этом для лучшей растекаемости припоя часто молибденовую металлизацию покрывают слоем гальванического никеля.

Активная пайка использует металлы переходной группы типа Ti, Ta, Zr, Nb и др. с незастроенной внутренней электронной оболочкой, которые могут выполнять роль акцептора по отношению к окислу Al_2O_3 керамики как донору электронной пары. Эти металлы либо применяют в виде конструкционного материала деталей, либо вводят их в состав активных припоев. В первом случае соединение осуществляется с помощью обычных припоев, в расплаве которых растворяется активный материал детали, обеспечивая донорно-акцепторное взаимодействие с окислом керамики. Во втором случае металл и керамику соединяют при помощи активного припоя на основе титана, вводимого в место соединения. В обоих случаях отпадает необходимость в предварительной металлизации поверхности керамики.

В заключение подчеркнем, что любое вакуумно-плотное монолитное соединение разнородных материалов можно осуществить лишь в том случае, если между компонентами этих материалов существует химическое или диффузионное взаимодействие. В результате такого взаимодействия возникает переходной слой, содержащий, наряду с исходными компонентами, также продукты их взаимного влияния.

5.17. Контрольные вопросы

1. В каких случаях становится важным учет поверхностных свойств физических объектов?
2. Опишите физические основания для размерной классификации изучаемых явлений и объектов различной природы.
3. В чем состоят квантовые размерные особенности нанообъектов пониженной размерности — квантовых ям, квантовых проволок, квантовых точек?
4. В чем состоят структурно-геометрические особенности нанообъектов атомной природы?
5. Опишите особенности икосаэдрического типа симметрии и квазикристаллического состояния материалов.
6. Каков смысл структурных и электронных магических чисел для наночастиц?
7. Какая индексация обычно используется для описания атомной структуры поверхности?
8. Дайте общее описание процессов релаксации и реконструкции поверхности.
9. В чем состоят специфические особенности реконструкции поверхности кремния?
10. Приведите характерные структурные дефекты поверхности реальных твердых тел.
11. Какие грани кристалла называются сингулярными, вичинальными и диффузными?
12. Какие новые физические понятия принято вводить с целью термодинамического описания поверхностных и межфазных явлений?
13. В чем заключается физический предположения Гиббса о введении разделяющей поверхности?
14. Каков физический смысл поверхностного натяжения и гиббсовской адсорбции?
15. Запишите адсорбционную формулу Гиббса и раскройте ее физическое содержание.
16. Сформулируйте принцип Гиббса–Кюри в применении к жидкостям и кристаллам.

17. В чем состоит смысл правила Кюри–Вульфа и его практической применимости?
18. Дайте общую характеристику поверхностно-активных и поверхностно-инактивных веществ; приведите примеры.
19. Какова роль ПАВ в механизмах удаления поверхностных загрязнений?
20. Опишите технологические особенности получения наноразмерных пленок и структур методом Ленгмюра–Блоджетт.
21. Дайте характеристику структурных типов пленок, получаемых по технологии Ленгмюра–Блоджетт.
22. Опишите технологические особенности золь-гель метода получения наноструктурированных материалов.
23. В чем заключается специфика термодинамики искривленных поверхностей?
24. Какова физическая причина появления капиллярного давления и осмотического давления?
25. Опишите сущность капиллярных явлений и областей их практической применимости.
26. Объясните физический смысл формулы Гиббса–Томсона в применении к испарению и растворимости частиц малых размеров.
27. Каким образом размер наночастицы влияет на температуру ее плавления?
28. Опишите природу физической и химической адсорбции на поверхности твердых тел.
29. Какой вид адсорбции является термоактивационным процессом и почему?
30. Дайте объяснение энергетической диаграммы для физической и химической адсорбции.
31. От чего зависит среднее время пребывания частицы в адсорбированном состоянии?
32. Качественно объясните температурную зависимость степени заполнения адсорбционных центров.
33. Что означают недиссоциативная/диссоциативная адсорбция и нелокализованная/локализованная адсорбция?

34. Изложите физическое содержание и способ описания модели адсорбции Ленгмюра.
35. Чем определяется кинетика недиссоциативной адсорбции и что характеризует уравнение изотермы Ленгмюра?
36. В чем заключается суть латерального взаимодействия при адсорбции и двумерной конденсации?
37. Изложите физическое содержание и способ описания модели адсорбции БЭТ.
38. Опишите физическую картину капиллярной конденсации в мезопористых адсорбентах.
39. Какова роль процессов адсорбции, растворения и диффузии в газопоглощении материалов?
40. Что является движущей силой процесса кристаллизации при физических и химических методах осаждения?
41. Какова физическая причина появления критического пересыщения и критических зародышей в первичной (жидкой или парогазовой) фазе?
42. В чем заключается специфика термодинамики гомогенного зародышеобразования?
43. Каковы термодинамические особенности гетерогенного зародышеобразования в рамках модели куполообразного сферического зародыша?
44. Каковы термодинамические особенности гетерогенного зародышеобразования в рамках модели монослойного зародыша?
45. Каковы термодинамические особенности гетерогенного зародышеобразования в рамках модели многослойного зародыша?
46. Чем определяется скорость гетерогенного зародышеобразования?
47. Всегда ли рост пленки на подложке происходит через стадию образования зародышей? Приведите примеры.
48. Опишите особенности выращивания пленок на несовершенных подложках при ступенчато-слоевом и спирально-дислокационном механизмах роста.
49. В чем состоит специфика эпитаксиального роста пленок на монокристаллических подложках?

50. Какова роль псевдоморфного слоя, дислокаций несоответствия и упругой релаксации в процессах гетероэпитаксиального роста пленок?
51. Опишите особенности островкового роста пленок по механизму Фольмера–Вебера.
52. Опишите особенности послойного роста пленок по механизму Франка–Ван-дер-Мерве.
53. Опишите особенности послойно-островкового роста пленок по механизму Странского–Крастанова.
54. Опишите особенности роста наноструктур на фасетированных поверхностях подложки.
55. Какова физическая сущность методов соединения материалов в жидкой и твердой фазах при формировании герметичных и вакуумно-плотных спаев и швов?

Приложение А

ТЕРМОДИНАМИЧЕСКИЕ РАСЧЕТЫ В ТЕХНОЛОГИЧЕСКИХ ЗАДАЧАХ ЭЛЕКТРОНИКИ

Цель данного приложения — дать читателю возможность закрепить полученные теоретические сведения из химической термодинамики и приобрести практические навыки термодинамического расчета на примере решения двух задач, имеющих непосредственное отношение к технологическим проблемам вакуумной электроники (приложение А.1) и полупроводниковой электроники (приложение А.2).

А.1. Состав газовой фазы и окисляемость металлов при термообработке оксидного катода

Общие сведения. Как уже отмечалось в п. 3.5, оксидные катоды, являющиеся одним из наиболее распространенных типов термокатода, представляют собой смесь оксидов бария, стронция и кальция, среди которых основным является ВаО. Будучи ди-электриком с шириной запрещенной зоны $\Delta E_g = 3,8$ эВ, оксид бария в чистом виде имеет работу выхода около 2,8 эВ. Согласно формуле Ричардсона–Дэшмана (3.5.5), такая большая работа выхода не позволяет обеспечить нужную для работы вакуумного прибора плотность тока термоэмиссии. Снижение работы выхода достигается в процессе активировки оксидного катода, проводимой путем его нагрева при откачке прибора.

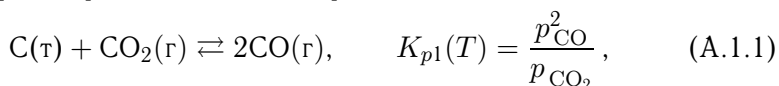
Процесс термообработки необходим для создания в оксидном покрытии атомов бария вследствие разложения ВаО, чтобы уменьшить количество катионных вакансий. В этом случае равновесие между вакансиями смещается в сторону преобладающего действия кислородных (анионных) вакансий, ведущих

себя как двухзарядные F -центры с донорными свойствами. Как результат этого, уровень Ферми поднимается вверх и работа выхода понижается до приемлемых значений 1,4–1,7 ГэВ. Кроме того, в никелевый kern реальных катодов вводятся специальные присадки-активаторы типа Mg, Si и W, способствующие химическому образованию свободного бария, например, по реакции $\text{BaO} + \text{Mg} \rightleftharpoons \text{MgO} + \text{Ba}$.

Как известно, оксиды щелочноземельных металлов Ba, Sr и Ca являются гигроскопичными и поэтому неустойчивы в обычных атмосферных условиях. Это не позволяет непосредственно формировать термоэмиссионное покрытие в виде окислов. Поэтому оксидная масса наносится на поверхность керна в виде смеси карбонатов BaCO_3 , SrCO_3 и CaCO_3 , приготовленной в форме суспензии на основе органического связывающего вещества. Последующий пиролиз карбонатов в процессе откачки дает необходимые оксиды BaO, SrO и CaO.

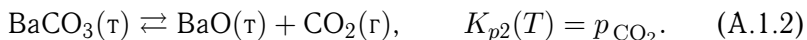
Вопросы термохимии обработки оксидного катода отчасти были затронуты в предыдущих параграфах, так что ниже собираем их воедино.

При термообработке катода в первую очередь происходит разложение органической связки, сопровождающееся бурным газовыделением с образованием углерода. Это вызывает нежелательное почернение оксидного покрытия, что повышает теплоизлучение с поверхности катода и снижает его рабочую температуру. Углерод удаляется из оксидного покрытия с помощью газотранспортной химической реакции (2.16.4):

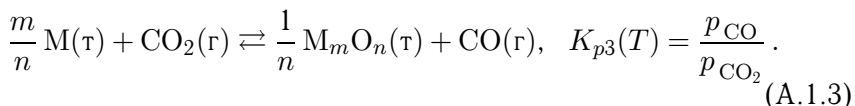


где справа записан закон действующих масс для этой реакции.

Углекислый газ CO_2 , входящий в реакцию (A.1.1), получается по реакции (2.12.17) разложением карбоната бария:



Таким образом, при термообработке оксидного катода формируется газовая среда, содержащая CO_2 и CO, наряду с другими газами. По отношению к металлам M такая среда обладает окислительно-восстановительными свойствами согласно реакции



Следовательно, возможность протекания реакции (А.1.3) в ту или другую сторону, управляемая уравнением изотермы Вант-Гоффа (2.13.5), связана с составом газовой фазы в вакуумном приборе. При этом давления p_{CO_2} и p_{CO} , формируемые реакциями (А.1.1) и (А.1.2), являются для реакции (А.1.3) неравновесными, так как навязываются ей со стороны, и в общем случае отличаются от равновесных для этой реакции давлений.

Таким образом, для оценки термодинамической возможности окисления металлических деталей вакуумного прибора необходимо сначала произвести расчет состава газовой среды, создаваемой реакциями (А.1.1) и (А.1.2), а затем на основании уравнения изотермы Вант-Гоффа (2.13.5) для реакции (А.1.3) определить направление ее протекания в этой газовой среде.

Расчет состава газовой фазы. Цель расчета — нахождение температурных зависимостей $p_{\text{CO}_2}(T_{\text{кат}})$ и $p_{\text{CO}}(T_{\text{кат}})$, где $T_{\text{кат}}$ — температура обработки катода. Обычно условия разложения карбонатов при термообработке таковы, что скорость газовыделения много больше скорости откачки, то есть практическая ситуация соответствует изохорному процессу, протекающему в объеме, ограниченном вакуумной оболочкой прибора. В этом случае для реакций (А.1.1) и (А.1.2) надо оперировать с законами действующих масс в «концентрационной» форме (2.12.6):

$$\begin{aligned} K_{c1}(T) &= \frac{c_{\text{CO}}^2}{c_{\text{CO}_2}}, \\ K_{c2}(T) &= c_{\text{CO}_2}. \end{aligned} \quad (\text{А.1.4})$$

При температуре обработки катода $T_{\text{кат}}$ из равенств (А.1.4) находим мольно-объемные концентрации газов:

$$\begin{aligned} c_{\text{CO}_2}(T_{\text{кат}}) &= K_{c2}(T_{\text{кат}}), \\ c_{\text{CO}}(T_{\text{кат}}) &= \sqrt{K_{c1}(T_{\text{кат}})K_{c2}(T_{\text{кат}})}. \end{aligned} \quad (\text{А.1.5})$$

Связь между константами равновесия $K_p(T)$ и $K_c(T)$ дается соотношением (2.12.7), подстановка которого в (А.1.5) дает

$$\begin{aligned} p_{\text{CO}_2}(T_{\text{кат}}) &= K_{p2}(T_{\text{кат}}), \\ p_{\text{CO}}(T_{\text{кат}}) &= \sqrt{K_{p1}(T_{\text{кат}})K_{p2}(T_{\text{кат}})}. \end{aligned} \quad (\text{А.1.6})$$

Таким образом, расчет состава газовой фазы сводится к нахождению температурной зависимости констант равновесия $K_{p1}(T)$ и $K_{p2}(T)$ для исходных реакций (А.1.1) и (А.1.2).

Расчет констант химического равновесия. В основе расчета константы равновесия $K_p(T)$ для любой реакции, записанной в обобщенной форме (2.12.1), лежит соотношение (2.13.3), с помощью которого константа равновесия K_p выражается через ΔG_T° :

$$\ln K_p(T) = -\frac{\Delta G_T^\circ}{RT}. \quad (\text{A.1.7})$$

Стандартный изобарный потенциал реакции ΔG_T° в этой формуле выражается через тепловой эффект ΔH_T° и энтропию ΔS_T° реакции в стандартных условиях с помощью известного соотношения (2.14.4):

$$\Delta G_T^\circ = \Delta H_T^\circ - T \Delta S_T^\circ. \quad (\text{A.1.8})$$

Таким образом, расчет зависимости $K_p(T)$ сводится к вычислению стандартных величин ΔH_T° и ΔS_T° для конкретной химической реакции при температуре T .

Исходными для термодинамических расчетов служат следующие выражения, полученные из формул (1.5.11), (2.14.5) и (2.14.6):

$$\begin{aligned} \Delta H_T^\circ &= \Delta H_{298}^\circ + \int_{298}^T \Delta C_P(T) dT, \\ \Delta S_T^\circ &= \Delta S_{298}^\circ + \int_{298}^T \frac{\Delta C_P(T)}{T} dT, \end{aligned} \quad (\text{A.1.9})$$

где ΔH_{298}° и ΔS_{298}° — стандартный тепловой эффект и стандартная энтропия реакции при $T = 298$ К, $\Delta C_P(T)$ — изменение изобарной теплоемкости за один пробег реакции слева направо при температуре T .

Величины ΔH_{298}° , ΔS_{298}° и $\Delta C_P(T)$ вычисляются на основе таблиц стандартных термодинамических величин (см. п. 1.13), короткие фрагменты которых приведены ниже в форме табл. А.1 и табл. А.2.

Для расчета ΔH_{298}° и ΔS_{298}° используют следующие выражения, аналогичные с формулами (2.14.5) и (2.14.6):

$$\Delta H_{298}^\circ = \sum_K \nu_K \Delta H_{298}^\circ(A_K) - \sum_H \nu_H \Delta H_{298}^\circ(A_H), \quad (\text{A.1.10})$$

$$\Delta S_{298}^\circ = \sum_K \nu_K S_{298}^\circ(A_K) - \sum_H \nu_H S_{298}^\circ(A_H). \quad (\text{A.1.11})$$

Величины $\Delta H_{298}^{\circ}(A_{к,н})$ и $S_{298}^{\circ}(A_{к,н})$, входящие в формулы (А.1.10) и (А.1.11), представляют собой стандартные молярные теплоты образования и стандартные энтропии. Они берутся из табл. А.1 и А.2 для конечных $A_{к}$ (стоящих в реакции справа) и начальных $A_{н}$ (стоящих в реакции слева) участников рассматриваемой реакции, обобщенная запись которой в форме (2.12.1) имеет стехиометрические коэффициенты $\nu_{к}$ и $\nu_{н}$.

Для расчета величины $\Delta C_P(T)$, входящей в (А.1.9), применяют выражение, аналогичное по форме с (А.1.10) и (А.1.11):

$$\Delta C_P(T) = \sum_{к} \nu_{к} C_P(A_{к}; T) - \sum_{н} \nu_{н} C_P(A_{н}; T). \quad (A.1.12)$$

Выражение (А.1.12) записано для участников реакции $A_{к}$ и $A_{н}$ как алгебраическая сумма их изобарных теплоемкостей, значения которых берутся из табл. А.1 и А.2. При этом температурная зависимость теплоемкостей $C_P(A_{к,н}; T)$ описывается как

$$C_P(T) = a + bT + cT^{-2}, \quad (A.1.13)$$

коэффициенты которой a , b и c для конкретного вещества даны в табл. А.1 и А.2. Здесь также приведены значения $C_P(A_{к,н}; 298)$ при $T = 298$ К.

Таким образом, расчет температурной зависимости константы равновесия $K_p(T)$, выражаемой через ΔG_T° в виде (А.1.7), следует выполнять по формулам (А.1.8)–(А.1.13).

Таблица А.1. Термодинамические данные для расчета состава газовой фазы

Вещество (состояние)	ΔH_{298}° $\frac{\text{ккал}}{\text{моль}}$	S_{298}° $\frac{\text{кал}}{\text{моль} \cdot \text{К}}$	$C_P(298)$ $\frac{\text{кал}}{\text{моль} \cdot \text{К}}$	$C_P(T) = a + bT + cT^{-2}$		
				a	$b \cdot 10^3$	$c \cdot 10^{-5}$
BaCO ₃ (т)	–287,3	26,80	20,40	20,77	11,70	–2,86
BaO(т)	–133,0	16,80	11,29	12,74	1,04	–1,98
SrCO ₃ (т)	–291,7	23,10	19,40	18,81	12,70	–2,20
SrO(т)	–141,7	13,11	10,71	11,50	2,11	–1,02
CaCO ₃ (т)	–285,4	21,80	19,86	16,31	15,73	–3,21
CaO(т)	–151,2	9,12	10,04	10,30	2,53	–1,41
C(т)	0	1,37	2,04	4,10	1,02	–2,10
CO(г)	–26,41	47,18	6,97	6,79	0,98	–0,11
CO ₂ (г)	–94,05	51,06	8,87	10,55	2,16	–2,04

Таблица А.2. Термодинамические данные
для расчета окисляемости металлов

Вещество (состояние)	ΔH_{298}° $\frac{\text{ккал}}{\text{моль}}$	S_{298}° $\frac{\text{кал}}{\text{моль} \cdot \text{К}}$	$C_P(298)$ $\frac{\text{кал}}{\text{моль} \cdot \text{К}}$	$C_P(T) = a + bT + cT^{-2}$		
				a	$b \cdot 10^3$	$c \cdot 10^{-5}$
Cu(т)	0	7,96	5,86	5,41	1,50	—
CuO(т)	-39,50	10,20	10,70	9,27	4,80	—
Cu ₂ O(т)	-40,0	22,44	16,70	14,90	5,70	—
Ni(т)	0	7,14	6,20	4,06	7,04	—
NiO(т)	-57,30	9,10	10,58	-4,99	37,58	3,89
Fe(т)	0	6,49	6,03	4,60	5,02	—
FeO(т)	-63,20	14,05	12,20	12,62	1,49	-0,76
Fe ₂ O ₃ (т)	-196,30	21,50	24,80	23,36	17,24	-3,08
Fe ₃ O ₄ (т)	-266,90	36,20	34,27	39,92	18,86	-10,01
Cr(т)	0	5,68	5,58	5,84	2,36	-0,88
CrO ₃ (т)	-142,10	17,20	—	—	—	—
Cr ₂ O ₃ (т)	-273,0	19,78	25,0	28,53	2,20	-3,74
Ag(т)	0	10,20	6,09	5,73	1,26	-0,06
Ag ₂ O(т)	-7,30	29,10	15,67	13,26	7,04	—
W(т)	0	7,80	5,93	5,74	0,76	—
WO ₂ (т)	-140,43	12,09	13,32	—	—	—
WO ₃ (т)	-200,93	19,50	19,04	—	—	—
Mo(т)	0	6,84	5,68	5,48	1,30	—
MoO ₂ (т)	-140,75	11,11	13,35	—	—	—
MoO ₃ (т)	-177,84	18,55	17,93	—	—	—
Ta(т)	0	9,92	6,04	—	—	—
Ta ₂ O ₅ (т)	-489,37	33,89	32,23	—	—	—
Ti(т)	0	7,33	5,98	5,28	2,40	—
TiO(т)	-129,40	8,31	9,53	—	—	—
Ti ₂ O ₃ (т)	-363,0	18,46	22,88	—	—	—

Продолжение таблицы А.2

Вещество (состояние)	ΔH_{298}° $\frac{\text{ккал}}{\text{моль}}$	S_{298}° $\frac{\text{кал}}{\text{моль} \cdot \text{К}}$	$C_P(298)$ $\frac{\text{кал}}{\text{моль} \cdot \text{К}}$	$C_P(T) = a + bT + cT^{-2}$		
				a	$b \cdot 10^3$	$c \cdot 10^{-5}$
TiO ₂ (т)	-225,50	12,00	13,49	17,14	0,98	-3,05
Ti ₃ O ₅ (т)	-586,84	30,14	36,02	—	—	—
Al(т)	0	6,77	5,82	4,94	2,96	—
Al ₂ O ₃ (т)	-400,30	12,18	18,88	27,38	3,08	-8,20
Zr(т)	0	9,30	6,01	6,83	1,12	-0,91
ZrO ₂ (т)	-261,47	12,03	13,40	16,64	1,80	-3,36
Nb(т)	0	8,74	5,87	—	—	—
NbO(т)	-406,0	11,98	9,86	—	—	—
NbO ₂ (т)	-795,40	13,01	13,72	—	—	—
Nb ₂ O ₅ (т)	-1898,0	32,74	31,53	—	—	—

Учет разной степени вклада теплоемкостей позволяет проводить термодинамический расчет в рамках трех последовательных приближений.

Первое приближение полностью пренебрегает вкладом теплоемкостей ($\Delta C_P = 0$), т. е. температурная зависимость термодинамических величин, даваемая формулами (А.1.9), опускается. Тогда выражение (А.1.8) принимает предельно упрощенный вид:

$$\Delta G_T^{\circ} = \Delta H_{298}^{\circ} - T \Delta S_{298}^{\circ}, \quad (\text{А.1.14})$$

где стандартные величины ΔH_{298}° и ΔS_{298}° вычисляются по формулам (А.1.10) и (А.1.11).

Второе приближение пренебрегает температурной зависимостью теплоемкостей, то есть подразумевается, что

$$\begin{aligned} \Delta C_P(T) &= \Delta C_P(298) \equiv \\ &\equiv \sum_k \nu_k C_P(A_k; 298) - \sum_n \nu_n C_P(A_n; 298) = \text{const.} \end{aligned} \quad (\text{А.1.15})$$

Тогда из формул (А.1.8) и (А.1.9) следует, что

$$\Delta G_T^{\circ} = \Delta H_{298}^{\circ} - T \Delta S_{298}^{\circ} - \Delta C_P(298) T f(T), \quad (\text{А.1.16})$$

где введена специальная функция

$$Tf(T) = T \left(\ln \frac{T}{298} + \frac{298}{T} - 1 \right), \quad (\text{А.1.17})$$

полученная вычислением интегралов в формулах (А.1.9) с учетом (А.1.15). Численные значения функции (А.1.17), называемой *функцией Улиха*, приведены в табл. А.3.

Таблица А.3. Функция Улиха и коэффициенты Темкина–Шварцмана

$T, \text{ К}$	$Tf(T), \text{ К}$	M_0	$M_1 \cdot 10^{-3}$	$M_{-2} \cdot 10^5$
500	56,65	0,1133	0,0407	0,0916
600	117,74	0,1962	0,0759	0,1423
700	195,58	0,2794	0,1153	0,1853
800	287,76	0,3597	0,1574	0,2213
900	392,46	0,4361	0,2012	0,2521
1000	508,30	0,5088	0,2463	0,2783
1100	634,16	0,5765	0,2922	0,2988
1200	769,12	0,6410	0,3389	0,3176
1300	912,42	0,7019	0,3860	0,3340
1400	1063,43	0,7595	0,4336	0,3484
1500	1221,58	0,8141	0,4814	0,3610
1600	1386,40	0,8665	0,5296	0,3723
1700	1557,48	0,9162	0,5780	0,3824
1800	1734,45	0,9635	0,6265	0,3915
1900	1916,97	1,0090	0,6752	0,3998
2000	2104,76	1,0525	0,7240	0,4072

Третье приближение учитывает точную зависимость $\Delta C_P(T)$, рассчитанную по формуле (А.1.12) на основе табличных данных. В этом приближении удобно воспользоваться *формулой Темкина–Шварцмана*

$$\Delta G_T^\circ = \Delta H_{298}^\circ - T \Delta S_{298}^\circ - T(M_0 \Delta a + M_1 \Delta b + M_{-2} \Delta c). \quad (\text{А.1.18})$$

Здесь Δa , Δb , Δc представляют собой алгебраические суммы коэффициентов, учитывающих температурный ход теплоемко-

стей в формуле (А.1.13), подсчитанные с учетом стехиометрических коэффициентов реакции (2.12.1) по общему правилу типа (А.1.10)–(А.1.12), например,

$$\Delta a = \sum_k \nu_k a(A_k) - \sum_n \nu_n a(A_n).$$

Коэффициенты Темкина–Шварцмана, входящие в (А.1.18), равны

$$M_0 = \ln \frac{T}{298} + \frac{298}{T} - 1,$$

$$M_n = \frac{T^n}{n(n+1)} + \frac{298^{n+1}}{(n+1)T} - \frac{298^n}{n} \quad \text{при } n \neq 0.$$

Численные значения коэффициентов M_0 , M_1 и M_{-2} приведены в табл. А.3.

Расчет окисляемости металлов. Для выяснения вопроса о возможности или невозможности окисления металла в газовой среде, сформированной при термообработке катода реакциями (А.1.1) и (А.1.2), необходимо прежде всего рассчитать по описанной выше методике температурный ход константы равновесия $K_{p3}(T)$ для окислительно-восстановительной реакции (А.1.3). Тогда изобарный потенциал, описываемый в реальных условиях уравнением изотермы Вант-Гоффа в виде (2.13.5), для реакции (А.1.3) вычисляется по формуле

$$\begin{aligned} \Delta G &= RT \left[\ln \frac{p_{\text{CO}}}{p_{\text{CO}_2}} - \ln K_{p3}(T) \right] = \\ &= RT \left[\ln \sqrt{\frac{K_{p1}(T_{\text{кат}})}{K_{p2}(T_{\text{кат}})}} - \ln K_{p3}(T) \right], \end{aligned} \quad (\text{А.1.19})$$

где T — температура рассматриваемого металла, отличная от температуры катода $T_{\text{кат}}$. Последнее равенство в (А.1.19) получено с использованием выражений (А.1.6) для p_{CO} и p_{CO_2} .

Таким образом, из соотношения (А.1.19) следует, что

а) при $\sqrt{K_{p1}(T_{\text{кат}})/K_{p2}(T_{\text{кат}})} > K_{p3}(T)$ имеем $\Delta G > 0$, т. е. реакция (А.1.3) сдвинута влево и рассматриваемый металл не окисляется при температуре T в данной газовой смеси;

б) при $\sqrt{K_{p1}(T_{\text{кат}})/K_{p2}(T_{\text{кат}})} < K_{p3}(T)$ имеем $\Delta G < 0$, т. е. реакция (А.1.3) сдвинута вправо и рассматриваемый металл окисляется при температуре T в данной газовой смеси.

Задание 1

1. Пользуясь термодинамическими данными, приведенными в табл. А.1, рассчитайте константы равновесия $K_{p1}(T_{\text{кат}})$ и $K_{p2}(T_{\text{кат}})$ по формулам (А.1.14), (А.1.16) и (А.1.18) для первого, второго и третьего приближений в интервале температур $T_{\text{кат}} = 1000 - 1500$ К.

2. Используя результаты из п. 1, рассчитайте по формулам (А.1.6) и постройте температурные зависимости $p_{\text{CO}}(T_{\text{кат}})$, $p_{\text{CO}_2}(T_{\text{кат}})$ и $p_{\text{CO}}/p_{\text{CO}_2} = f(T_{\text{кат}})$ для трех приближений.

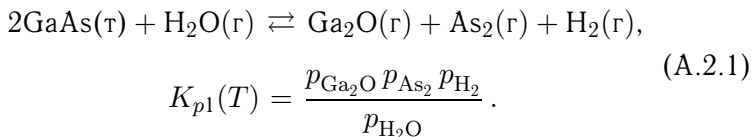
3. Пользуясь термодинамическими данными, приведенными в табл. А.2, для выбранной пары металл–окисел рассчитайте $K_{p3}(T)$ по формулам (А.1.14), (А.1.16) и (А.1.18) для первого, второго и третьего приближений в интервале температур $T = 500 - 1500$ К.

4. Используя результаты п. 2 и п. 3, постройте на одном графике температурные зависимости $p_{\text{CO}}/p_{\text{CO}_2} = f(T_{\text{кат}})$ и $K_{p3}(T)$ для трех приближений и по знаку ΔG в формуле (А.1.19) сделайте вывод о возможности или невозможности окисления выбранного металла.

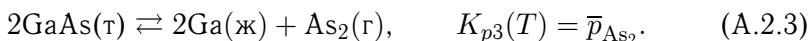
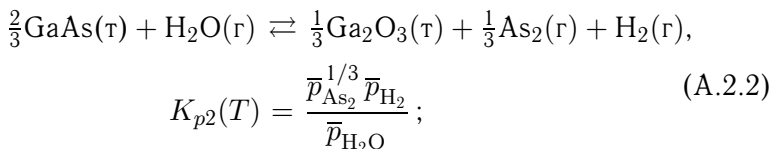
А.2. Химическое осаждение арсенида галлия из газовой фазы в системе GaAs–H₂O–H₂

Общие сведения. Газотранспортные химические реакции широко применяют в современной полупроводниковой технологии для эпитаксиального выращивания монокристаллических слоев с высокой степенью структурного и электрофизического совершенства. Для этой цели в газовый поток, создаваемый газом-носителем (каковым чаще является водород), вводится так называемый *транспортный агент*. Его основная функция — осуществлять травление полупроводникового материала с образованием газообразных продуктов реакции, переносимых потоком (конвекционным или диффузионным) из зоны травления в зону осаждения (см. п. 2.16).

Для арсенида галлия роль транспортного агента могут выполнять пары воды H₂O. Основной газотранспортной реакцией при этом является реакция с образованием летучего окисла Ga₂O:



Наряду с основной газотранспортной реакцией (А.2.1), в изучаемой системе возможно протекание конкурирующих реакций, препятствующих транспорту арсенида галлия в газовой фазе с помощью паров воды. Это происходит вследствие связывания галлия в форме новой конденсированной фазы. Такими фазами являются твердый окисел $\text{Ga}_2\text{O}_3(\text{т})$ и жидкий галлий $\text{Ga}(\text{ж})$, образующиеся по реакциям:



Возможность кристаллизации GaAs в соответствии с газотранспортной реакцией (А.2.1) при условии подавления конкурирующих реакций (А.2.2) и (А.2.3) оценивают по составу газовой фазы. Парциальные давления газообразных компонентов p_{H_2} , $p_{\text{H}_2\text{O}}$, $p_{\text{Ga}_2\text{O}}$ и p_{As_2} , формируемые реакцией (А.2.1), рассматриваются как неравновесные для конкурирующих реакций. Эти давления навязываются реакциям (А.2.2) и (А.2.3) основной реакцией (А.2.1) и в общем случае отличаются от давлений $\bar{p}_{\text{H}_2}$, $\bar{p}_{\text{H}_2\text{O}}$ и $\bar{p}_{\text{As}_2}$, равновесных для этих реакций. Подавление конкурирующих реакций имеет место при условии, если их изобарный потенциал, вычисленный по уравнению изотермы Вант-Гоффа (2.13.5), положителен ($\Delta G > 0$). В этом случае реакции (А.2.2) и (А.2.3) смещены влево, так что образования твердого окисла $\text{Ga}_2\text{O}_3(\text{т})$ и жидкого галлия $\text{Ga}(\text{ж})$ не происходит.

Очевидно, что первым этапом анализа является определение температурного хода констант равновесия $K_{p1}(T)$, $K_{p2}(T)$ и $K_{p3}(T)$ для реакций (А.2.1), (А.2.2) и (А.2.3) на основании таблиц стандартных термодинамических величин.

Расчет констант химического равновесия. Нередко в справочниках приведены термодинамические величины, связанные со значениями термодинамических функций G_T° и H_T° , полученными на основе спектроскопических данных или вычисленными квантово-статистическими методами. В частности, используется температурная зависимость новой функции

$$\Phi_T = -\frac{G_T^\circ - H_{298}^\circ}{T}, \quad (\text{А.2.4})$$

значения которой приведены в табл. А.4.

Таблица А.4. К расчету химического транспорта
арсенида галлия парами воды

Функция	T, К	Вещество (состояние)							
		H ₂ (г)	H ₂ O (г)	As ₂ (г)	As ₄ (г)	Ga ₂ O (г)	Ga ₂ O ₃ (г)	Ga (ж)	GaAs (г)
$\Phi_T, \frac{\text{ккал}}{\text{моль} \cdot \text{К}}$	700	33,15	47,41	59,60	80,34	73,03	25,63	19,90	16,27
	800	33,72	48,09	60,29	81,88	74,07	27,38	20,79	16,76
	900	34,25	48,75	60,95	83,34	75,08	29,07	21,57	17,20
	1000	34,76	49,38	61,58	84,76	76,04	30,71	22,27	17,65
	1100	35,24	49,99	62,17	86,11	76,96	32,30	22,90	18,10
	1200	35,69	50,58	62,74	87,35	77,83	33,82	23,48	18,43
	1300	36,13	51,14	63,28	88,56	78,66	35,24	24,01	18,88
$\Delta H_{298}^\circ, \frac{\text{ккал}}{\text{моль}}$	298	0	-57,8	48,0	34,5	-19,7	-240	1,335	-17,7

Нетрудно видеть, что в этом случае стандартный изобарный потенциал реакции ΔG_T° вместо выражения (А.1.8) принимает вид

$$\Delta G_T^\circ = \Delta H_{298}^\circ - T \Delta \Phi_T. \quad (\text{А.2.5})$$

Константа химического равновесия $K_p(T)$ связана с величиной ΔG_T° тем же соотношением (А.1.7), тогда с учетом (А.2.5) получаем

$$\ln K_p(T) = \frac{1}{R} \left(\Delta \Phi_T - \frac{\Delta H_{298}^\circ}{T} \right). \quad (\text{А.2.6})$$

Стандартную теплоту реакции ΔH_{298}° находим, как и ранее, по формуле (А.1.10), а величину $\Delta \Phi_T$ — по общему правилу типа (А.1.10)–(А.1.12):

$$\Delta \Phi_T = \sum_K \nu_K \Phi_T(A_K) - \sum_H \nu_H \Phi_T(A_H). \quad (\text{А.2.7})$$

При этом значения $\Phi_T(A_K)$ и $\Phi_T(A_H)$ берутся из табл. А.4 для конечных (A_K) и начальных (A_H) участников рассматриваемой реакции.

Расчет состава газовой фазы. Обычно процесс транспорта арсенида галлия парами воды проводится в открытой проточной системе при внешнем давлении 1 атм. Однако, для общности будем принимать суммарное давление равным P , тогда

$$P = p_{\text{H}_2} + p_{\text{H}_2\text{O}} + p_{\text{Ga}_2\text{O}} + p_{\text{As}_2}. \quad (\text{A.2.8})$$

На вход системы подается в качестве газа-носителя водород H_2 под давлением $p_{\text{H}_2}^0$, который насыщен парами воды с парциальным давлением $p_{\text{H}_2\text{O}}^0$ путем пропускания через специальный сосуд с водой, называемый барботером. Величина $p_{\text{H}_2\text{O}}^0$ регулируется при помощи температуры барботера и измеряется точкой росы увлажненного водорода. Относительная влажность водорода задается параметром

$$a = \frac{p_{\text{H}_2\text{O}}^0}{p_{\text{H}_2}^0} = \frac{n_{\text{H}_2\text{O}}^0}{n_{\text{H}_2}^0}, \quad (\text{A.2.9})$$

где $n_{\text{H}_2\text{O}}^0$ и $n_{\text{H}_2}^0$ — число молей паров воды и водорода, вводимых в систему.

Пусть по реакции (А.2.1) вступило во взаимодействие с арсенидом галлия $\xi n_{\text{H}_2\text{O}}^0$ молей паров воды, где ξ — степень превращения. Тогда в соответствии со стехиометрией реакции равновесное число молей в системе с учетом (А.2.9) таково:

$$\begin{aligned} n_{\text{H}_2} &= n_{\text{H}_2}^0 + \xi n_{\text{H}_2\text{O}}^0 = (1 + \xi a) n_{\text{H}_2}^0, \\ n_{\text{H}_2\text{O}} &= n_{\text{H}_2\text{O}}^0 - \xi n_{\text{H}_2\text{O}}^0 = (1 - \xi) a n_{\text{H}_2}^0, \\ n_{\text{Ga}_2\text{O}} &= n_{\text{As}_2} = \xi n_{\text{H}_2\text{O}}^0 = \xi a n_{\text{H}_2}^0. \end{aligned} \quad (\text{A.2.10})$$

Полное число газовых молей записываем на основе (А.2.8) в виде

$$n = n_{\text{H}_2} + n_{\text{H}_2\text{O}} + n_{\text{Ga}_2\text{O}} + n_{\text{As}_2} = [1 + (1 + 2\xi)a] n_{\text{H}_2}^0. \quad (\text{A.2.11})$$

Отсюда парциальное давление каждого компонента, рассчитанное с использованием (А.2.10) и (А.2.11) по формуле $p_i = (n_i/n)P$, равно

$$p_{\text{H}_2} = \frac{1 + \xi a}{1 + (1 + 2\xi)a} P \approx [1 - (1 + \xi)a] P \approx P, \quad (\text{A.2.12})$$

$$p_{\text{H}_2\text{O}} = \frac{(1 - \xi)a}{1 + (1 + 2\xi)a} P \approx (1 - \xi)a P, \quad (\text{A.2.13})$$

$$p_{\text{Ga}_2\text{O}} = p_{\text{As}_2} = \frac{\xi a}{1 + (1 + 2\xi)a} P \approx \xi a P, \quad (\text{A.2.14})$$

где приближенные равенства записаны с учетом малых значений величины a , лежащих в диапазоне от 10^{-4} до 10^{-2} .

Подставляя формулы (А.2.12)–(А.2.14) в закон действующих масс для реакции (А.2.1), получаем уравнение для нахождения

степени превращения ξ как функции трех управляемых параметров рассматриваемого технологического процесса P , T и a :

$$\xi^2 + \frac{K_{p1}(T)}{aP^2} \xi - \frac{K_{p1}(T)}{aP^2} = 0. \quad (\text{A.2.15})$$

Решение (A.2.15) дает искомую степень превращения:

$$\xi(T, P, a) = \frac{K_{p1}(T)}{2aP^2} \left(\sqrt{1 + \frac{4aP^2}{K_{p1}(T)}} - 1 \right). \quad (\text{A.2.16})$$

Найденная функциональная зависимость $\xi = f(T, a)$ при $P = \text{const}$ позволяет рассчитать по формулам (A.2.13) и (A.2.14) парциальные давления $p_{\text{H}_2\text{O}}$, $p_{\text{Ga}_2\text{O}}$ и p_{As_2} как функции температуры T и относительной влажности водорода a .

Область стехиометричности газовой фазы. Для термодинамической оценки возможности подавления конкурирующих реакций надо рассчитать их изобарный потенциал ΔG для давлений в системе, полученных из расчета состава газовой фазы, созданного основной реакцией (A.2.1). Такую газовую фазу принято называть *стехиометрической*. Уравнение изотермы Вант-Гоффа (2.13.5) для конкурирующих реакций имеет вид:

а) для реакции (A.2.2)

$$\Delta G_2(T, a) = RT \left[\ln \frac{p_{\text{As}_2}^{1/3}(T, a) p_{\text{H}_2}(T, a)}{p_{\text{H}_2\text{O}}(T, a)} - \ln K_{p2}(T) \right], \quad (\text{A.2.17})$$

б) для реакции (A.2.3)

$$\Delta G_3(T, a) = RT \left[\ln p_{\text{As}_2}(T, a) - \ln K_{p3}(T) \right]. \quad (\text{A.2.18})$$

Выражения (A.2.17) и (A.2.18) содержат парциальные давления p_{H_2} , $p_{\text{H}_2\text{O}}$ и p_{As_2} компонентов стехиометрической газовой смеси, ранее полученные в виде (A.2.12)–(A.2.14).

Решение уравнений $\Delta G_2(T, a) = 0$ и $\Delta G_3(T, a) = 0$, описывающих равновесие реакций (A.2.2) и (A.2.3), дает две линии 2 и 3 на плоскости T – a , показанные на рис. А.1. Эти линии ограничивают область ①, где протекает основная реакция (A.2.1) и подавлены конкурирующие реакции (A.2.2) и (A.2.3). За пределами этой области ниже линии 2, кроме основной реакции (A.2.1), протекает реакция (A.2.2) с образованием нелетучего окисла $\text{Ga}_2\text{O}_3(\text{т})$, а выше линии 3 — реакция диссоциации арсенида галлия (A.2.3) с образованием жидкого галлия $\text{Ga}(\text{ж})$. Одновременное подавление обеих конкурирующих реакций, гарантированное неравенствами $\Delta G_2 > 0$ и $\Delta G_3 > 0$, возможно лишь внутри

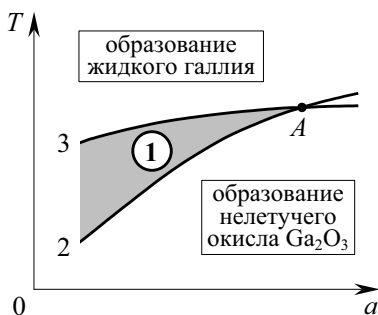


Рис. А.1. Область ① стехиометричности газовой фазы, сформированной по основной реакции (А.2.1), ограниченная линиями 2 и 3, соответствующими условиям равновесия $\Delta G_2(T, a) = 0$ и $\Delta G_3(T, a) = 0$ для конкурирующих реакций (А.2.2) и (А.2.3)

области ①, лежащей левее точки А, в которой пересекаются кривые 2 и 3 на рис. А.1. Эта область называется *областью стехиометричности* газовой фазы, так как для нее соотношение между парциальными давлениями газообразных компонентов удовлетворяет стехиометрии основной газотранспортной реакции (А.2.1). Внутренние точки области стехиометричности газовой фазы соответствуют оптимальному режиму эпитаксиального процесса при выращивании арсенида галлия в системе GaAs–H₂O–H₂.

Задание 2

1. Пользуясь термодинамическими данными, приведенными в табл. А.4, рассчитайте по формуле (А.2.6) константы равновесия для реакций (А.2.1), (А.2.2), (А.2.3) и постройте графические зависимости $K_{p1}(T)$, $K_{p2}(T)$, $K_{p3}(T)$ в интервале температур $T = 700 - 1300$ К.

2. Пользуясь давлениями насыщенного водяного пара, приведенными в табл. А.5, вычислите относительную влажность водорода $a = p_{\text{H}_2\text{O}}^0 / p_{\text{H}_2}^0$, соответствующую точкам росы $+20^\circ\text{C}$, $+10^\circ\text{C}$, $+5^\circ\text{C}$, 0°C , -5°C , -10°C , для выбранного давления $P = p_{\text{H}_2}^0 + p_{\text{H}_2\text{O}}^0$, в частности при $P = 1$ атм.

Таблица А.5. Давление насыщенного водяного пара

Точка росы	$+20^\circ\text{C}$	$+10^\circ\text{C}$	$+5^\circ\text{C}$	0°C	-5°C	-10°C
$p_{\text{H}_2\text{O}}^0$, Торр	17,53	9,20	6,54	4,58	3,02	1,95

3. Рассчитайте степень превращения $\xi(T, a)$ по формуле (А.2.16), задавая в качестве параметра вычисленные значения относительной влажности водорода $a = p_{\text{H}_2\text{O}}^0 / p_{\text{H}_2}^0$ для точек росы $+20^\circ\text{C}$, $+10^\circ\text{C}$, $+5^\circ\text{C}$, 0°C , -5°C , -10°C , и постройте температурные зависимости $\xi(T, a)$ в интервале температур 700–1300 К для выбранного давления P .

4. Рассчитайте по формулам (А.2.13) и (А.2.14) и постройте температурные зависимости $p_{\text{H}_2\text{O}}(T, a)$ и $p_{\text{Ga}_2\text{O}}(T, a) = p_{\text{As}_2}(T, a)$ для полученных ранее в п. 2 значений относительной влажности водорода a .

5. Рассчитайте по формулам (А.2.17) и (А.2.18) и постройте температурные зависимости $\Delta G_2(T, a)$ и $\Delta G_3(T, a)$ для вычисленных выше значений a .

6. Используя кривые, полученные в п. 5, вычислите графически корни уравнений $\Delta G_2(T, a) = 0$ и $\Delta G_3(T, a) = 0$ и по результатам графического решения постройте в координатах T – a границы области стехиометричности газовой фазы, качественно показанной на рис. А.1.

Приложение Б

СПРАВОЧНЫЙ МАТЕРИАЛ

Б.1. Важнейшие физические константы

Заряд электрона	$e = 1,602 \cdot 10^{-19}$ Кл
Масса электрона	$m_e = 9,109 \cdot 10^{-31}$ кг
Скорость света в вакууме	$c = 2,998 \cdot 10^8$ м/с
Электрическая постоянная	$\varepsilon_0 = 8,854 \cdot 10^{-12}$ Ф/м
Магнитная постоянная	$\mu_0 = 1,257 \cdot 10^{-6}$ Г/м
Постоянная Авогадро	$\mathcal{N}_A = 6,022 \cdot 10^{23}$ моль ⁻¹
Постоянная Больцмана	$k_B = 1,381 \cdot 10^{-23}$ Дж/К
Постоянная Планка	$\hbar = 1,055 \cdot 10^{-34}$ Дж·с
Постоянная Фарадея	$\mathcal{F} = e\mathcal{N}_A = 96485$ Кл/г-экв
Газовая постоянная	$R = k_B\mathcal{N}_A = 8,314$ Дж/(моль·К)

Б.2. Некоторые внесистемные единицы

$1 \text{ \AA} = 10^{-8}$ см = 10^{-4} мкм = 0,1 нм
$1 \text{ л} = 10^{-3}$ м ³ = 10^3 см ³ = 10^{15} мкм ³ = 10^{24} нм ³
$1 \text{ Торр (мм рт. ст.)} = 133,3$ Па (Н/м ²) = $1,316 \cdot 10^{-3}$ атм
$1 \text{ бар} = 10^5$ Па (Н/м ²) = 0,987 атм = 750 Торр (мм рт. ст.)
$1 \text{ атм} = 1,013 \cdot 10^5$ Па (Н/м ²) = 1,013 бар = 760 Торр (мм рт. ст.)
$1 \text{ ккал/моль} = 4,184$ кДж/моль = 0,0434 эВ/частица
$1 \text{ кГм} = 9,81$ Дж = 2,34 кал = 73,6 л·Торр = $6,12 \cdot 10^{19}$ эВ
$1 \text{ эВ} = 1,6 \cdot 10^{-19}$ Дж = $3,83 \cdot 10^{-20}$ кал = $1,2 \cdot 10^{-18}$ л·Торр
$1 \text{ а.е.м. (атомная единица массы)} = 1,66 \cdot 10^{-27}$ кг

Б.3. Основные обозначения

Приведенные ниже обозначения (с указанием размерности физических величин) соответствуют наиболее распространенным и общепринятым в литературе по физической, физико-химической и электрохимической тематике.

- a — период идентичности кристаллической решетки; относительная влажность водорода;
- a_i — активность i -го компонента в растворе;
- A — площадь поверхности (м^2);
- $\mathcal{A}$ — химическое сродство, $\mathcal{A} \equiv -\Delta G$ (Дж/моль, кал/моль);
- A — акцепторный атом;
- A^- — акцепторный ион;
- $[A]$ — концентрация акцепторных атомов (м^{-3});
- $[A^-]$ — концентрация акцепторных ионов (м^{-3});
- $A_{\text{н(к)}}$ — обобщенный химический символ начального (конечного) участника реакции, записанной в обобщенной форме (1.5.14);
- c_i — мольно-объемная концентрация i -го компонента в реальных условиях (моль/м^3 , моль/л);
- $\bar{c}_i$ — мольно-объемная концентрация i -го компонента в равновесных условиях (моль/м^3 , моль/л);
- C_P — изобарная теплоемкость (молярная) при постоянном давлении ($\text{кал}/(\text{моль} \cdot \text{К})$);
- C_V — изохорная теплоемкость (молярная) при постоянном объеме ($\text{кал}/(\text{моль} \cdot \text{К})$);
- ΔC_P — изменение изобарной теплоемкости системы, происходящее за один пробег слева направо обобщенной химической реакции (1.5.14);
- D — донорный атом;
- D^+ — донорный ион;
- $[D]$ — концентрация донорных атомов (м^{-3});
- $[D^+]$ — концентрация донорных ионов (м^{-3});
- D_i — коэффициент диффузии для i -го компонента ($\text{м}^2/\text{с}$);
- D_{ij} — бинарный коэффициент диффузии для i -го и j -го компонентов ($\text{м}^2/\text{с}$);

D_D	— коэффициент Дюфура ($\text{Вт} \cdot \text{м}^2$);
D_T	— коэффициент термодиффузии ($\text{м}^2/\text{с}$);
e^-	— подвижный электрон в зоне проводимости кристалла;
e^+	— подвижная дырка в валентной зоне кристалла;
$\mathcal{E}$	— энергия активации термоактивационного процесса (эВ);
E	— электродвижущая сила (В); электродный потенциал в электрохимическом элементе (В);
E°	— стандартный электродный потенциал (В);
ΔE	— обратимая ЭДС электрохимического элемента (В);
E_c	— энергия дна зоны проводимости кристалла (эВ);
E_v	— энергия потолка валентной зоны кристалла (эВ);
E_A	— энергетический уровень акцепторной примеси (эВ);
E_D	— энергетический уровень донорной примеси (эВ);
E_F	— энергия уровня Ферми (эВ);
ΔE_g	— ширина запрещенной зоны (эВ);
ΔE_A	— энергия ионизации акцепторной примеси (глубина залегания акцепторного уровня) (эВ);
ΔE_D	— энергия ионизации донорной примеси (глубина залегания донорного уровня) (эВ);
F	— свободная энергия Гельмгольца (изохорно-изотермический потенциал) (Дж, кал);
ΔF	— изменение свободной энергии в изохорно-изотермических условиях (Дж, кал); изохорный потенциал химической реакции (Дж/моль, кал/моль);
ΔF_T°	— стандартный изохорный потенциал при температуре T , в частности, ΔF_{298}° при 298 К (Дж/моль, кал/моль);
G	— свободная энергия Гиббса (изобарно-изотермический потенциал) (Дж, кал);
ΔG	— изменение свободной энергии в изобарно-изотермических условиях (Дж, кал); изобарный потенциал химической реакции (Дж/моль, кал/моль);
ΔG_T°	— стандартный изобарный потенциал при температуре T , в частности, ΔG_{298}° при 298 К (Дж/моль, кал/моль);
H	— энтальпия (Дж, кал);

- ΔH — изменение энтальпии (тепловой эффект Q_P изобарного процесса) (Дж, кал);
- ΔH_T° — стандартная энтальпия образования химических соединений из простых веществ¹⁾ или стандартный тепловой эффект в изобарных условиях при температуре T , в частности, ΔH_{298}° при 298 К (Дж/моль, кал/моль);
- i — плотность электрического тока (А/м²);
- I — электрический ток (А);
- J_α — термодинамические потоки;
- $\mathbf{J}_i$ — вектор плотности (диффузионного) потока частиц i -го сорта (м⁻²·с⁻¹, моль/(м²·с));
- $\mathbf{J}_q$ — вектор плотности теплового потока (Вт/м²);
- $\mathbf{J}_s$ — вектор плотности потока энтропии (Вт/(м²·К));
- $k_{1(2)}$ — константа скорости прямой (обратной) реакции;
- K_a — константа равновесия химической реакции, выраженная через активности a_i ;
- K_c — константа равновесия химической реакции, выраженная через мольно-объемные концентрации c_i ;
- K_p — константа равновесия химической реакции, выраженная через парциальные давления p_i ;
- K_x — константа равновесия химической реакции, выраженная через мольные доли x_i ;
- K_w — ионное произведение воды;
- K_i — константа электронно-дырочного равновесия в полупроводниках (индекс «i» от англ. «intrinsic»);
- K_A — константа равновесия квазихимической реакции ионизации акцепторной примеси;
- K_D — константа равновесия квазихимической реакции ионизации донорной примеси;
- K_{V_M} — константа равновесия квазихимической реакции ионизации вакансий в подрешетке М;
- K_{V_X} — константа равновесия квазихимической реакции ионизации вакансий в подрешетке Х;

¹⁾ Нередко в справочной литературе энтальпию образования ΔH_{298}° обозначают в виде ΔH_{f298}° с добавлением индекса f от англ. *formation*.

K_{Φ}	— константа равновесия процесса образования собственных дефектов по Френкелю;
$K_{Ш}$	— константа равновесия процесса образования собственных дефектов по Шоттки;
L	— произведение растворимости твердого электролита в воде;
$L_{\alpha\beta}$	— коэффициенты Онсагера;
M	— элементарный полупроводник или металлический элемент полупроводникового соединения MX или M_xX_y ;
M_I	— междоузельный атом M ;
M_I^+	— ионизованный междоузельный атом M ;
$[M_I]$	— концентрация междоузельных атомов M_I (м^{-3});
$[M_I^+]$	— концентрация междоузельных ионов M_I^+ (м^{-3});
m	— масса всей рассматриваемой системы (г, кг);
M_i	— молярная масса i -го компонента (г/моль, кг/кмоль);
n	— концентрация электронов в полупроводнике (м^{-3});
n_i	— собственная концентрация носителей заряда в полупроводниках (индекс « i » от англ. «intrinsic») (м^{-3});
n_i	— число молей i -го компонента в системе;
N_i	— число частиц i -го компонента в системе;
N_A	— растворимость (равновесная концентрация) акцепторной примеси в полупроводнике (м^{-3});
N_D	— растворимость (равновесная концентрация) донорной примеси в полупроводнике (м^{-3});
N_{Φ}	— поверхностная концентрация физадсорбционных центров (м^{-2});
N_x	— поверхностная концентрация химических центров (м^{-2});
p	— концентрация дырок в полупроводнике (м^{-3});
p_i	— парциальное давление i -го компонента (атм, Торр);
$\bar{p}_i$	— парциальное давление i -го компонента в равновесных условиях (атм, Торр);
p_i^0	— давление насыщенного пара чистого i -го компонента (атм, Торр);
pH	— водородный показатель водного раствора;

- P — суммарное давление в многокомпонентной газовой фазе (атм, Торр);
 $\mathcal{P}$ — производство энтропии (Вт/К, кал/(с · К));
 Q_V — тепловой эффект изохорного процесса, равный ΔU (Дж, кал);
 Q_P — тепловой эффект изобарного процесса, равный ΔH (Дж, кал);
 s_i — коэффициент растворимости i -го компонента в конденсированной (жидкой или твердой) фазе;
 S — энтропия (Дж/К, кал/К);
 S_T° — стандартная энтропия вещества при температуре T , в частном случае, S_{298}° при 298 К (Дж/(моль · К), кал/(моль · К));
 ΔS — изменение энтропии в ходе термодинамического процесса (Дж/К, кал/К);
 ΔS_T° — стандартная энтропия процесса при температуре T , в частном случае, ΔS_{298}° при 298 К (Дж/(моль · К), кал/(моль · К));
 t — время (с);
 T — абсолютная температура (К);
 U — внутренняя энергия системы (Дж, кал); падение напряжения (В);
 ΔU — изменение внутренней энергии (тепловой эффект Q_V изохорного процесса) (Дж, кал);
 ΔU_T° — стандартный тепловой эффект в изохорных условиях при температуре T , в частности, ΔU_{298}° при 298 К (Дж/моль, кал/моль);
 v_r — скорость r -ой химической реакции (моль/(м³ · с), моль/(л · с));
 $\mathbf{v}_i$ — вектор гидродинамической скорости движения i -го компонента (м/с);
 V — объем (м³, л);
 V_M — вакансии в подрешетке М;
 V_X — вакансии в подрешетке Х;
 V_M^- — заряженная вакансия в подрешетке М;
 V_X^+ — заряженная вакансия в подрешетке Х;

- $[V_M]$ — концентрация вакансий в подрешетке М (м^{-3});
 $[V_X]$ — концентрация вакансий в подрешетке Х (м^{-3});
 $[V_M^-]$ — концентрация заряженных вакансий в подрешетке М (м^{-3});
 $[V_X^+]$ — концентрация заряженных вакансий в подрешетке Х (м^{-3});
 w — вероятность элементарных процессов (с^{-1}); объемная плотность экстенсивного свойства W ;
 W — статистический вес, такой что энтропия $S = k_B \ln W$; работа (механическая $W_{\text{мех}}$, электрическая $W_{\text{эл}}$) (Дж);
 x_i — мольная доля i -го компонента;
 X_α — термодинамические силы;
 X — неметаллический элемент полупроводникового соединения MX или M_xX_y ;
 X_I — междоузельный атом Х;
 X_I^- — ионизованный междоузельный атом Х;
 $[X_I]$ — концентрация междоузельных атомов X_I (м^{-3});
 $[X_I^-]$ — концентрация междоузельных ионов X_I^- (м^{-3});
 z_i — зарядовое число частицы i -го сорта с электрическим зарядом $z_i e$;
 $Y_{\text{м(с)}}$ — коэффициент распыления металлов (химических соединений);
 z — координационное число;
 α — коэффициент прилипания при физической адсорбции; коэффициент электронного переноса;
 β_i — коэффициент диффузионного массопереноса i -го компонента (м/с);
 γ_i — коэффициент активности i -го компонента в растворе;
 Γ_i — гиббсовская адсорбция частиц i -го сорта (моль/м^2);
 δ_i — эффективная толщина диффузионного слоя для i -го компонента (м);
 η — электродное перенапряжение (В);
 θ — степень заполнения адсорбционных центров; краевой (контактный) угол;
 Θ — степень заполнения химических центров;

λ	— длина электромагнитной волны в вакууме; коэффициент теплопроводности (Вт/(м · К));
μ_i	— химический потенциал i -го компонента (Дж/моль, кал/моль);
μ_i°	— химический потенциал i -го компонента в стандартных условиях (Дж/моль, кал/моль);
$\tilde{\mu}_i$	— электрохимический потенциал заряженных частиц i -го компонента (Дж/моль, кал/моль);
$\nu_{\text{н(к)}}$	— стехиометрические коэффициенты для начальных (конечных) участников химической реакции, записанной в обобщенной форме (1.5.14);
ξ	— степень превращения вещества; число пробегов (степень протоноты) химической реакции;
π	— поверхностное давление (Н/м); осмотическое давление (атм, Торр);
ρ_i	— массовая плотность вещества i -го сорта (кг/м ³);
σ	— поверхностное натяжение (Дж/м ²);
σ_s	— плотность производства энтропии (Вт/(м ³ · К));
τ	— время релаксации (с);
τ_a	— среднее время жизни адсорбированных атомов (с);
φ	— электрический потенциал (В);
Φ	— работа выхода электронов (эВ);
χ	— электронное сродство (эВ);
г	— газообразное состояние;
п	— парообразное состояние;
ж	— жидкое состояние;
т	— твердое состояние;
к	— конденсированное (жидкое или твердое) состояние;
K	— общее число компонентов в системе;
C	— число термодинамических степеней свободы равновесной системы;
Φ	— число сосуществующих фаз в равновесной системе;
X	— число химических реакций в равновесной системе.

Литература, рекомендуемая для углубленного изучения

1. *Матвеев А. Н.* Молекулярная физика. 2-е изд. — М.: Высшая школа, 1987.
2. *Киреев В. А.* Краткий курс физической химии. 5-е изд. — М.: Химия, 1994.
3. *Стромберг А. Г., Семченко Д. П.* Физическая химия. 5-е изд., испр. — М.: Высшая школа, 2003.
4. *Карпетьянц М. Х.* Введение в теорию химических процессов. 5-е изд. — М.: Высшая школа, 1981.
5. *Ормонт Б. Ф.* Введение в физическую химию и кристаллохимию полупроводников. — М.: Высшая школа, 1982.
6. *Гурвич А. М.* Введение в физическую химию кристаллофосфоров. — М.: Высшая школа, 1982.
7. *Хенней Н.* Химия твердого тела. — М.: Мир, 1971.
8. *Антропов Л. И.* Теоретическая электрохимия. 3-е изд. — М.: Высшая школа, 1984.
9. *Гуров К. П.* Феноменологическая термодинамика необратимых процессов. — М.: Наука, 1978.
10. *Пригожин И., Дефэй Р.* Химическая термодинамика. — Новосибирск: Наука, 1966.
11. *Николис Г., Пригожин И.* Самоорганизация в неравновесных системах. — М.: Мир, 1979.
12. *Пригожин И.* Введение в термодинамику необратимых процессов. — Ижевск: НИЦ «Регулярная и хаотическая динамика», 2001.
13. *Пригожин И., Кондепуди Д.* Современная термодинамика. От тепловых двигателей до диссипативных структур. — М.: Мир, 2002.
14. *Эбелинг В.* Образование структур при необратимых процессах. Введение в теорию диссипативных структур. — Москва–Ижевск: Институт компьютерных исследований, 2003.
15. *Шаскольская М. П.* Кристаллография. — М.: Высшая школа, 1984.

16. Мейер К. Физико-химическая кристаллография. — М.: Металлургия, 1972.
17. Люпис К. Химическая термодинамика материалов. — М.: Металлургия, 1989.
18. Металлургия в электронике. /Сборник статей под ред. Рябцева Н.Г. и Кузнецова Ф.А. — М.: Металлургия, 1970.
19. Медведев С.А. Введение в технологию полупроводниковых материалов. — М.: Высшая школа, 1970.
20. Колобов Н.А. Основы технологии электронных приборов. — М.: Высшая школа, 1980.
21. Крапухин В.В., Соколов С.А., Кузнецов Г.Д.. Физико-химические основы технологии полупроводниковых материалов. — М.: Металлургия, 1982.
22. Пичугин И.Г., Таиров Ю.М. Технология полупроводниковых приборов. — М.: Высшая школа, 1985.
23. Курносое А.И., Юдин В.В. Технология производства полупроводниковых приборов и интегральных микросхем. 3-е изд., испр. — М.: Высшая школа, 1986.
24. Ефимов И.Е., Козырь И.Я., Горбунов Ю.И. Микроэлектроника: Физические и технологические основы, надежность. — М.: Высшая школа, 1986.
25. Черняев В.Н. Физико-химические процессы в технологии РЭА. — М.: Высшая школа, 1997.
26. Краткий справочник физико-химических величин. /Под ред. Мищенко К.П. и Равделя А.А. — Л.: Химия, 1972.
27. Рябин В.А., Остроумов М.А., Свит Т.Ф. Термодинамические свойства веществ (справочник). — Л.: Химия, 1977.
28. Вудраф Д., Делчар Т. Современные методы исследования поверхности. — М.: Мир, 1989.
29. Адамсон А. Физическая химия поверхностей. — М.: Мир, 1979.
30. Праттон М. Введение в физику поверхности. — Ижевск: НИЦ «Регулярная и хаотическая динамика», 2000.
31. Оура К., Лифшиц В.Г., Саранин А.А., Зотов А.В., Катаяма М. Введение в физику поверхности. — М.: Наука, 2005.
32. Ролдугин В.И. Физико-химия поверхности. — М.: Издательский дом «Интеллект», 2008.

Дополнительная литература, посвященная наноматериалам и нанотехнологиям, приведена в конце введения (см. с. 72).

Предметный указатель

Агрегатные превращения
206–210

Адиабатический процесс 89

Адсорбция 644–682

— гиббсовская 583, 585, 587,
590–592

— диссоциативная/недиссоциа-
тивная 652, 660

—, изотерма Генри 660

—, изотерма Ленгмюра 660, 663

— локализованная/нелокализо-
ванная 652, 654, 664, 670

—, модель БЭТ 672–676

—, модель двумерной конденса-
ции 654, 664–670

—, модель капиллярной conden-
сации 655, 676–682

—, модель Ленгмюра 655–659

— мономолекулярная (моно-
слойная) 654–671

— полимолекулярная (много-
слойная) 655, 671–676

—, роль в газопоглощении мате-
риалов 682–685

— физическая 647–650

— химическая 650–652

Активированный комплекс 437

Активировка катода 396, 747

Активность компонента 223

Акцептор 378–379, 386

Анион 293

Ассоциация молекул 217

Бертоллиды 263

Бифуркация 118, 129

Ближний порядок 215, 334

Буферные растворы 295

Вакансии 377, 379

Величины

— интенсивные 87, 123, 170

— локальные 170–174, 181–182

— парциальные мольные 124

— удельные 170, 172, 186,
197–198

— экстенсивные 87, 123, 170

Внутренняя энергия 79

— — идеального газа 128

Водородный показатель 295

Восстановление катодное 299,
310, 324

Время релаксации 80, 88, 101,
167–169, 398, 535, 662

Газопоглощение материалов
682–685

Газопроницаемость материалов
457–62

Гальванический элемент 299

Гидратация 217, 293

Гидрофильность/гидрофобность
597, 599

Грамм-атом 77

Грамм-ион 77

Грамм-молекула (моль) 77

Грамм-эквивалент 77

Граница

— отражающая 444

— связывающая 444

Давление

— диссоциации 274

— капиллярное 621–627

— насыщенного пара 221,
211–215, 230–237

— — — малых частиц 630–634

- осмотическое 627–630
- парциальное 128, 232
- стандартное 129, 132
- Дальний порядок 215, 334
- Дальтонида 263
- Дефекты структуры поверхности 577–581
- Дефекты решетки (точечные) 369–373, 377–398
 - —, ассоциация и компенсация 411–417
 - — примесные 370, 377, 406–410, 422–424
 - —, активаторы/соактиваторы 388–392
 - — — внедрения 370, 377, 379
 - — — замещения 370, 377, 378
 - —, сенсibilизаторы 388
 - — собственные 377, 418–428
 - — — антиструктурные 377
 - —, вакансии 370, 377, 379
 - —, междоузельные атомы 370, 377
 - — — по Френкелю 370, 421
 - — — по Шоттки 370, 420
 - — акцепторные 378–379
 - — донорные 378–379
 - — амфотерные 378, 428–431
- Диаграммы плавкости бинарных систем 241, 247
 - — без твердых растворов 249–252
 - — дистектические 265
 - — с неограниченным твердым раствором 252–255
 - — с ограниченными твердыми растворами 256–262
 - — с химическими соединениями в твердой фазе 262–268
- Дислокации и поверхностные дефекты 369, 373–376
 - винтовые 373–375, 716
 - , дефекты упаковки 376
 - несоответствия 718
 - краевые 373–375
- Диссипативные структуры 75, 107, 117, 119, 127, 129
- Диффузия 148–149
 - изотермическая/неизотермическая 192–196
 - конвективная 479
- Диффузионные задачи
 - — обезгаживания материалов 445–453
 - — легирования полупроводников 462–468
 - — — из поверхностного источника с отражающей границей 466–468
 - — — из полуограниченного источника 464–466
 - — — из постоянного источника 462–463
- Донор 378–379, 385
- Закон(ы)
 - Аррениуса 154
 - Генри 238
 - Гесса 93
 - Дальтона 128
 - действующих масс 269–274
 - Мура 40
 - разбавления Оствальда 298
 - Рауля, 232
 - Рауля–Генри 234
 - Сивертса 240
 - сохранения энергии 85–88
 - Фика 442–443
- Идеальный газ 127
 - —, внутренняя энергия 128
 - —, термодинамические функции 127–131
 - —, уравнение состояния 127, 600
 - —, химический потенциал 129
- Идеальный раствор 224–225
 - —, закон Генри 238
 - —, закон Рауля 232
 - —, закон Сивертса 240
- Ингибитор 438

- Ионное произведение воды 294, 318
Ионное равновесие 293–298
- Капиллярные явления 625–627
Катализатор 438
Катион 293
Квазихимический метод 399–401
Кинетика
— диффузионная 445–453, 462–468, 475–483, 483–487, 542–545
— зародышеобразования 710–712
— мономолекулярной адсорбции 660–662
— термического окисления 487–494
— химическая 435–438, 475–483
— —, основной постулат 152
— химического травления полупроводников 495–500
— химического транспорта веществ 514–540
— — — при реактивном распылении 527–540
— — — в проточных системах 514–518
— — — в сэндвич-системах 519–526
— электрохимическая 545–549
Коагуляция 614, 722, 739
Коалесценция 721–722
Конденсация 511
— двумерная 654, 664–670
— капиллярная 634, 655, 676–682
Коннода 248
Константа
— диссоциации 296
— скорости химической реакции 152, 435–437
— химического равновесия 190, 270–274, 436
Контакт
— материальный 85
— механический 85, 90
— тепловой 85, 89
— физический 738
— химический 738
Концентрация
— в мольных долях 123, 128
— мольно-объемная 123, 128
Коэффициент(ы)
— активности 223, 230
— газопроницаемости 458
— диффузии 149, 193, 194, 442
— диффузионного массопереноса 480, 490, 497
— Дюфура 194
— конденсации 657
— Онсагера 157
— прилипания 656
— распределения (сегрегации) примеси 245, 465, 485
— распыления 532
— растворимости 238–241
— Соре 195
— стехиометрические 94, 137, 151, 269
— Темкина–Шварцмана 754–755
— теплопроводности 150, 194
— термодиффузии 194
— электронного переноса 547
Краевой (контактный) угол 695
Кривая
— испарения 206–207
— плавления 206–207
— сублимации 206–207
Кристаллизация 685–737
— гомогенная 685, 693–694
— гетерогенная 685, 695–737
—, движущая сила процесса 685–691
— неравновесная 255
— равновесная 248, 249–251, 252–254, 259–261
Кристаллические решетки 342–349
— —, индексы Миллера 347–349
— —, классы и группы симметрии 342–343

- —, сингонии и ячейки Браве 344–347
- Кристаллическая структура 342, 349–369
 - — алмаза 361
 - — вольфрама 358
 - — вюрцита 363
 - — каменной соли 360
 - — магния 359
 - — меди 359
 - — сфалерита 362
 - — хлористого цезия 360
 - —, политипизм и полиморфизм 364–368
 - —, пространственная упаковка атомов 350–356
 - —, координационный многогранник 351, 352
 - —, координационное число 350
 - —, кристаллические пустоты 354–356
- Критический зародыш 689–691
 - —, работа образования 694, 697, 701, 705
 - —, размер 694, 697, 701, 705
- Критическое пересыщение 689
- Латеральное взаимодействие 656, 664, 666, 670, 676, 707
- Линия
 - ликвидуса 245, 246, 249, 253, 256–261
 - солидуса 245, 246, 253, 257–261
- Локальное квазиравновесие 80, 167–170, 174–176
- Люминесценция 387–392
 - внутрицентровая 388
 - рекомбинационная 388
- Магические числа 564–567
 - — структурные 567
 - — электронные 567
- Макросостояние системы 77
- Метод(ы)
 - Ленгмюра–Блджетт 598–605
 - золь-гель технологии 617–621
 - локальной микрообработки 29
 - модификации материала 29
 - нанодиагностики 65–71
 - осаждения материала 28
 - термической очистки 456
 - удаления материала 28
 - физического осаждения 686–688
 - химического осаждения 689
 - химической очистки 456
- Механизм(ы)
 - генерации собственных дефектов 418–422
 - диффузии атомов 439–442
 - кинетики химических реакций 435–439
 - молекулярного сита 220, 238, 460
 - роста пленок 713–737
 - — — на фасетированной поверхности 734–737
 - — — островковый (Фольмера–Вебера) 713, 719–725
 - — — послыйный (Франка–Ван-дер-Мерве) 713, 725–728
 - — — послыйно-островковый (Странского–Крастанова) 713, 728–734
 - — — спирально-дислокационный 716
 - — — ступенчато-слоевой 714–716
 - удаления загрязнений 592–598
 - формирования соединений 737–742
- Микросостояние системы 77
- Мицелла 595, 615
- Модель зародыша
 - куполообразного 695–699
 - многослойного 703–710
 - монослойного 700–703
- Моль 77
- Мольная доля 123, 128

- Нанодиагностика 65–71
— быстрыми электронами 70–71
—, зондовая микроскопия 71
— ионными и молекулярными пучками 67–69
— медленными электронами 66–67
—, полевая микроскопия 70–71
—, радиоспектроскопия 71
— электромагнитным излучением 69–70
Наноматериалы 42–55
— углеродные 43–55, 366
—, нанотрубки 51–55
—, фуллерены 44–48
—, фуллериды 49–50
—, фуллериты 48–49
—, эндоэдралы 50–51
— фрактальные 43
Нанотехнологии 55–64
— зондовые 57–63
—, молекулярно-лучевая эпитаксия 56, 511–514
—, МОС-гидридная эпитаксия 56, 506–507
—, реактивное распыление 56, 527–540
— фуллеренов и УНТ 63–64
—, химическая сборка 56, 507–511
Насыщенный раствор 221
Насыщенный пар 211, 221
—, —, парциальное давление 211
—, —, —, зависимость от критической температуры 630–633
—, —, —, зависимость от температуры 212–215
—, —, —, зависимость от полного давления 212–213
Начала термодинамики 73, 76, 82, 83, 87, 97–99
Область гомогенности 263
Область протекания процессов
— — — диффузионная 476, 494, 499, 526
— — — кинетическая 476, 493, 499, 526
Окисление анодное 299, 310, 324
Описание сплошной среды
— — — лагранжево 177
— — — эйлерино 177
Осаждение катодное 299, 310
Перенапряжение 300
— диффузионное 542
— водорода 327
— кислорода 327
— разряда 546
— фазовое 542
— электродное 327, 541
— электрохимическое 542
Перитектика 256–257
Перитектическая горизонталь 248, 257
Планарный процесс 28
Плотность
— потока внутренней энергии 182
— потока частиц 148, 181
— потока энтропии 182–186
— производства внутренней энергии 182
— производства частиц 181
— производства энтропии 182–186
— теплового потока 149, 186
— энергии, энтальпии, энтропии 170–173
Площадка невариантности 250, 260, 265
Поверхностно-активные вещества 590–591, 596–600
Поверхностное давление 600–601, 665–666, 670
Поверхностное натяжение 583–584
Поликонденсация 612
Полимеризация 612
Полупроводники
— компенсированные 381, 390
— примесные 379, 404–406

- — *n*-типа 381
- — *p*-типа 381
- сильно легированные 384
- собственные 381, 402–404
- Поляризация электрода
 - — концентрационная 300, 542
 - — химическая 300, 540
- Постоянная
 - Авогадро 77, 763
 - Больцмана 78, 763
 - универсальная газовая 127, 763
 - Фарадея 77, 763
- Потенциал
 - изобарно-изотермический (изобарный) 102
 - изобарный химической реакции 137, 276
 - изохорно-изотермический (изохорный) 102
 - изохорный химической реакции 279
 - ионизации 335
 - поверхностный 394
 - химический 87, 122–124, 132–133, 393
 - электрический 140–142, 308–309
 - электродный 306, 311–314
 - электрокинетический 615
 - электрохимический 140, 393
- Поток вещества
 - диффузионный 149, 180, 479, 489–491, 496
 - конвективный 180, 479
 - стефановский 480–482, 520
 - химический 477, 491, 497
- Правило
 - Кюри–Вульфа 588
 - рычага 248, 254–255
 - Трутона 215
 - фаз Гиббса 138–140
- Предельный ток диффузии 542
- Примесь
 - акцепторная 378
 - амфотерная 378
 - внедрения 377, 379
 - глубокая 384
 - донорная 378
 - замещения 377, 378
 - мелкая 379
- Принцип(ы)
 - вакуумного обезгаживания материалов 454–457
 - Гиббса–Кюри 588
 - детального равновесия 159
 - диффузионного легирования полупроводников 468–475
 - — — из парогазовой фазы 469
 - — — из поверхностных источников 469–471
 - компенсации заряда/валентности 389–391
 - Ле Шателье 165–166, 287
 - масштабирования 34
 - микроскопической обратимости 159
 - минимума производства энтропии 76, 103, 109, 160–166
 - симметрии Кюри 187–189
 - эволюции термодинамических систем 76, 199–201
 - экстремумов 76, 104
 - эпитаксии 501–514
 - — газофазной 504–511
 - — жидкофазной 501–504
 - — молекулярно-лучевой 511–514
- Проектная норма техпроцесса 35
- Произведение растворимости 297
- Производная
 - локальная 178
 - конвективная 178
 - субстанциональная 178
- Производство энтропии 99, 103, 109, 156, 161
 - — избыточное 113, 117
- Процесс
 - адиабатический 89
 - диффузионный 148–149, 439–442
 - изобарный 90–92

- изотермический 89, 101–106
- изохорный 90–92
- ионный 293–298, 317–322
- квазистатический 88, 95
- кинетический 150–153, 435–438
- необратимый 83, 89, 95–99
- неравновесный 83, 84
- обратимый 83, 89, 96–97
- равновесный 83, 84
- релаксационный 88
- самопроизвольный 95, 102–104
- термоактивационный 154
- Псевдоморфный слой 717, 728
- Работа 85–87
 - выхода электрона 392–394
 - механическая 85, 303
 - электрическая 303
- Равновесие
 - газов и конденсированных сред 211–215, 230–241
 - жидкой и твердой фаз 242–248
 - ионное 293–298
 - текущее 107
 - фазовое 135–136, 141
 - химическое 136–138
 - электронно-дырочное 406
 - электрохимическое 141–142, 306
- Раствор
 - атермальный 225–226
 - буферный 295
 - жидкий 217–218
 - коллоидный 613, 617
 - идеальный 224–225
 - насыщенный 221, 237
 - отрицательный 223, 228
 - положительный 223, 228
 - предельно-разбавленный 229
 - регулярный 226–230
 - твердый
 - — внедрения 216, 220
 - — вычитания 216
 - — замещения 216, 220
 - — неограниченный 216, 252
 - — ограниченный 216, 256
 - — упорядоченный 216
 - электролитов 293–298
- Растворение анодное 299, 310
- Растворимость 221
 - газов в конденсированных средах 220–221, 682–685
 - — атомарная 220, 239–241
 - — молекулярная 220, 237–239
 - примесей в полупроводниках 406–410
 - — —, ретроградный ход 410
- Реконструкция поверхности 397, 568, 570–571
- Релаксация поверхности 397, 568, 571–577
- Ряд напряжений 315, 317, 319
- Самоорганизация 12, 57, 61, 75, 116, 117, 513, 734
- Седиментация/пептизация 614–616
- Система
 - адиабатическая 86
 - гетерогенная 135
 - гомогенная 135
 - закрытая 86, 101–104
 - замкнутая 80, 82, 86, 100–101
 - квазизамкнутая 80, 167, 175
 - однокомпонентная 206–215
 - открытая 86, 38–40, 106–119
 - электрохимическая 298
- Скорость зародышеобразования 710–712
- Скорость химической реакции 151–153, 435, 477
 - — —, константа скорости 152, 436–437
- Сольватация 217, 293
- Состояние(я)
 - аморфное 334
 - жидкокристаллическое 335
 - квазикристаллическое 563
 - кристаллическое 334
 - переходное 437

- прекурсорное 652
- поликристаллическое 334
- стандартное 94, 132–134
- электронов
- — локализованные/нелокализованные 380
- — поверхностные 396–399
- — пониженной размерности 555–560
- — — —, квантовая нить 559
- — — —, квантовая точка 560
- — — —, квантовая яма 557
- Сродство
- к электрону 336
- химическое 138, 184, 190–192
- электронное 392, 394
- Статистический вес 78, 418
- Структурная упорядоченность 117–118
- Суспензия 231, 613, 748
- Температура
- кипения 214, 231, 233–234
- кристаллизации 207, 243–246
- перитектики 258
- плавления 207, 243–244
- — малых частиц 635–639
- — нанокластеров 639–644
- разложения 275
- , статистический смысл 83–84
- тройной точки 207
- фазового перехода 208
- эвтектики 249, 257
- Теорема
- Нернста 82, 134
- Остроградского–Гаусса 179, 183, 200, 443
- Пригожина 109–110, 160–164, 200
- Теплоемкость 91, 134, 751
- Теплота
- адсорбции 649, 652
- изобарного процесса 90–93, 122
- изохорного процесса 90–93, 121
- испарения 209, 219–220
- кристаллизации 245, 687
- некомпенсированная 96
- необратимых процессов 95–97
- образования соединения (стандартная) 94, 134, 750–753
- переноса 194, 196
- плавления 209, 219–220
- растворения 239, 240
- фазового перехода 93, 208
- химической реакции 94, 280–283
- Терминология и понятия
- коллоидной химии 612–617
- органической химии 605–612
- Термодинамические
- потоки 99, 156
- силы 99, 156, 201
- степени свободы 138
- Термостат 95
- Ток обмена 546
- Упругая релаксация 718, 729, 732, 736
- Уравнение(я)
- Аррениуса 437
- баланса 178–184
- — внутренней энергии 182
- — в общей форме 178–181
- — числа частиц 181
- — энтропии 182–184
- Гельмгольца 92, 121
- Гиббса–Гельмгольца 121
- Гиббса–Дюгема 124
- Гиббса–Томсона 632
- изобары Вант-Гоффа 280
- изотермы Вант-Гоффа 277
- изотермы Генри 660
- изотермы Ленгмюра 660, 663
- изохоры Вант-Гоффа 282
- Кертисса–Гиршфельдера 520
- Клапейрона–Клаузиуса 242
- Клапейрона–Менделеева 127
- непрерывности 443
- Нернста 305, 314

- Фика 443
- Шредера–Ван-Лаара 246
- Уровень вакуума 392
- Уровень Ферми 383, 392
- Условия
 - равновесия 95, 104
 - — фазового 136, 141
 - — химического 137–138, 141–142, 277
- самопроизвольности процессов 95, 104, 278
- Фаза 135
- Фазовые переходы
 - второго рода 210–211
 - первого рода 207, 210, 219
- Фигуративная точка 248
- Физадсорбция 647–650
- Физически бесконечно малая (фбм) частица 170, 177
- Физические свойства кристаллов
 - — люминесцентные 387–392
 - — термоэмиссионные 392–396
 - — электрические 379–387
- Флуктуации
 - термодинамические 100–101, 110–118, 143–145, 198
 - крупномасштабные 175
 - мелкомасштабные 175
- Флуктуационная устойчивость 110–116, 143–147, 197–199
- Формула
 - адсорбционная Гиббса 587
 - Гиббса–Томсона 632, 635
 - Лапласа 623
 - Тафеля 548
- Фугитивность 223
- Функция(и)
 - Гаусса 467
 - ошибок 450, 463
 - состояния 87, 119–121
 - Темкина–Шварцмана 754
 - термодинамические 119–120
 - Улиха 754
- Хемосорбция 650–654
- Химическая реакция
 - — в обобщенной форме 94, 137, 150, 269, 289, 476, 496
 - — газотранспортная 288–292
 - — гетерогенная 136, 142, 273–275
 - — гомогенная 136, 142, 270–273
 - — ионизации атома 309
 - — квазиравновесная 153
 - — квазихимическая 399–401
 - — окислительно-восстановительная 298, 310
 - — простая 153
 - — разряда иона 309
 - — экзотермическая 283
 - — эндотермическая 283
- Химическая связь 334–341
 - — водородная 341, 648
 - — донорно-акцепторная 339
 - — ионная 336–337
 - — ковалентная 337–340
 - — металлическая 340
 - — молекулярная 340–341, 647–648
- Химические соединения
 - — интерметаллические 263
 - — переменного состава 216, 263
 - — постоянного состава 262
 - — с инконгруэнтным плавлением 265
 - — с конгруэнтным плавлением 265
- Химический потенциал 87, 122–124, 132–134, 393
- Центры
 - люминесценции 388
 - прилипания/захвата 384, 391
 - рекомбинации 384
 - свечения 388, 391
 - окраски 372, 388
 - многозарядные 381
- Эвтектика
 - простая 249
 - сложная 257

- Эвтектическая горизонталь 248, 257
- Электрод
- водородный 313–314, 318–322
 - кислородный 318–322
 - обратимый относительно аниона 308, 311–312
 - обратимый относительно катиона 308, 311–312
 - редокс 308, 311
- Электродвижущая сила 299–300, 304–307
- Электродный потенциал 306, 311–314
- Электродные процессы 307–316
- Электролиз 293, 299
- Электролитическая ячейка 299
- Электролиты 293–298
- Электроника
- макро- 8, 11, 23 25–31
 - микро- 8, 11, 23, 32–41
 - нано- 8, 11, 23, 42–71
- Электроотрицательность 336
- Электрофорез 615–616
- Электрохимический элемент
- —, гальванический режим работы 301, 310, 322–324
 - —, электролитический режим работы 302, 310, 324–328
- Электрохимический потенциал 140, 393
- Эмульсия 231, 613
- Энергетические зоны кристалла 380, 391, 393, 402, 404
- Энергия
- активации 154, 437, 439, 651
 - внутренняя 79
 - ионизации 378
 - свободная Гельмгольца 102
 - свободная Гиббса 102
- Энтальпия 91
- Энтропия(и) 78–84, 95–100
- закон возрастания 82–83, 98
 - избыточная 113–116, 145, 158
 - информационная 79
 - колебательная 418
 - конфигурационная 418
 - необратимых процессов 95–97
 - обратимых процессов 95–97
 - приращение внешнее 98, 106–108
 - приращение внутреннее 98, 106–108
 - производство 99, 103, 109, 161
 - смешения 131, 218, 222, 225, 226, 418
 - стандартная 134, 750–753
- Эпитаксиальный рост 716–737
- , гетероэпитаксия 703, 716–718
 - , гомоэпитаксия 701, 726–728
- Эффект(ы)
- Дюфура (диффузионный термоэффект) 193–196
 - размерные квантовые 555–560
 - размерные структурные 560–567
 - Соре (термодиффузия) 164, 193–195