

Л.А.Грибов Н.И.Прокофьева

# Основы физики

учебник



Гармония

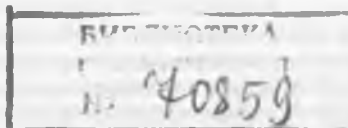
53  
11-82

Л.А.Грибов Н.И.Прокофьева

# Основы физики

издание третье

*Рекомендовано Министерством  
общего и профессионального  
образования Российской Федерации  
в качестве учебника для студентов  
естественно-научных направлений  
высших учебных заведений*



70854-70860  
7



Гардарики  
Москва, 1998

УДК 53(075.8)

ББК 22.3

Г82

Рецензенты:

Доктор физико-математических наук профессор *В.А. Дементьев*

Доктор физико-математических наук профессор *В.И. Баранов*

**Грибов Л.А., Прокофьева Н.И.**  
Г82 Основы физики: Учебник. — 3-е изд. — М.: Гардарика, 1998. — 564 с.: ил.

ISBN 5-7762-0045-8 (в пер.)

Изложены основы классической и релятивистской механики, термодинамики, включая элементы неравновесной, молекулярной физики, теории электромагнитных явлений, оптики и теории строения вещества на молекулярном уровне. Особое внимание уделяется вопросам перехода от существующего к качественно новому и самоорганизации в неживой и живой природе. Специальная глава посвящена демонстрации плодотворности применения открытых в физике общих закономерностей к теории познания окружающего мира, к проблемам управления и к пониманию общественных процессов.

Для студентов естественно-научных специальностей вузов и системы бакалавриата. Может быть использован в технических вузах с сокращенным курсом физики.

УДК 53(075.8)

ББК 22.3

ISBN 5-7762-0045-8

© «Гардарика», 1998

© Грибов Л.А., Прокофьева Н.И., 1998

## ПРЕДИСЛОВИЕ

Предлагаемое читателям второе (дополненное и исправленное) издание учебника «Основы физики» отличается от первого тем, что в него не только внесены мелкие исправления и устранены обнаруженные «ляпы», но сделан ряд существенных дополнений. Эти дополнения прежде всего имели своей целью усилить ту часть учебника, которая призвана показать роль физики в познании не статического, а развивающегося мира и обратить внимание на то, что некоторые фундаментальные выводы физики оказываются всеобщими и с успехом могут быть использованы в других областях человеческого знания, вплоть до истории. Примеры читатели найдут в заключительной главе учебника, которая так и названа: «Уроки физики».

Вообще, по мнению авторов, сейчас все более и более ощущается необходимость «наведения мостов» между физикой как таковой и другими науками и сторонами человеческой деятельности. Взгляд на физику только как на основу техники сейчас явно устарел. В наш век физика положила начало молекулярной биологии, на стыке математики и физики родилась кибернетика, физика открыла основной закон появления нового—самоорганизацию и разрастание флуктуаций.

Являясь всегда фундаментом естествознания, физика выработала специфическую форму мышления, которая может быть названа здоровым консерватизмом и которая как раз сейчас в быстро меняющемся и очень не простом мире крайне полезна.

Все это резко повышает общеобразовательное значение физики. Не случайно сейчас обсуждается вопрос о введении курса физики в гуманитарных высших учебных заведениях. Конечно, такая тенденция должна найти свое отражение в стиле и содержании учебников. Настоящее, второе, издание «Основ физики» предназначено прежде всего для естественно-научного бакалавриата. Именно поэтому в книге нет узкой специализированной направленности. Вместе с тем, авторы считали необходимым, как и в первом издании, уделить значительное внимание миру молекул, т. е. тем объектам, с которыми прежде всего имеют дело химики, биологи, специалисты сельского хозяйства. С целью сокращения объема книги в нее вообще не включены описательные разделы, достаточно полно излагаемые в школьном курсе физики. Это, прежде всего, вопросы строения



атомного ядра и теории элементарных частиц, радиоактивность и ряд других. Крупные исключения, по-видимому, вообще неизбежны в любом кратком учебнике, если только не превращать его в справочник.

Мы будем рады, если книга окажется полезной в учебном процессе. Мы будем благодарны за все замечания, которые читатели сочтут нужным нам сообщить.

Особую признательность мы выражаем профессорам В.А. Демьезу, А.Д. Гладуну, А.Д. Суханову, В.И. Баранову, Л.В. Тарасову, Л.Л. Дашкевичу, доц. В.А. Башлачеву, которые приняли на себя труд отрецензировать учебник или ознакомиться как с первым, так и с некоторыми разделами второго издания и высказали свои предложения.

Мы благодарим также всех наших сотрудников, которые в той или иной степени помогли нам при технической подготовке рукописи к печати.

## ВВЕДЕНИЕ

*«Все науки можно разделить на две группы—на физику и коллекционирование марок».*

*Э. Резерфорд*

### 1. Что такое физика?

В «Физическом энциклопедическом словаре» (т. 5, с. 311) на этот вопрос дан следующий ответ: «Слово физика означает науку о природе. Границы, отделяющие физику от других наук о природе, никогда не были четкими. Поэтому попытки дать строгое определение физике как науки путем ограничения класса изучаемых ею объектов оказываются неудачными. У любого объекта имеются такие общие свойства (механические, электрические и т. д.), которые служат предметом изучения физики. Ближе всего к истине определение современной физики как науки о формах материи (их иногда называют «первичными», «общими»), которые входят в состав любых сложных материальных систем, о взаимодействиях этих форм материи и их движениях».

На разных этапах исторического развития физики как науки в ее рамках было детально разработано учение об основных законах движения простых и сложных тел и совокупностей тел, обнаружены, исследованы и приведены в стройную систему электромагнитные явления, установлено строение атомов, молекул, твердых тел, получены фундаментальные результаты о свойствах жидкостей, газов, плазмы, развиты методы детального исследования хода химических реакций и т. д.

Показательно, что одно из величайших достижений науки нашего века—открытие строения двойной спирали ДНК, положившей начало триумфальному шествию молекулярной биологии, было сделано в чисто физической лаборатории<sup>1</sup>, той самой, в которой за несколько десятков лет до этого было открыто строение атома.

Вобрав в себя громадную область самых разнообразных знаний о природе окружающего нас материального мира, физика, еще на ранней стадии давшая объяснение строению Солнечной

<sup>1</sup> Кавендишская лаборатория, Кембридж, Великобритания.

системы, сейчас занимается разработкой тончайших вопросов, связанных с возникновением всей нашей Вселенной.

Накопление фундаментальных знаний по отдельным направлениям сопровождалось их дальнейшей детальной разработкой вплоть до появления важнейших прикладных областей.

Так, разработка основных законов механики привела к появлению теоретической механики, теории машин и механизмов и в конечном счете к созданию методов расчета и проектирования всех известных на сегодняшний день технических устройств. Экспериментальные и теоретические открытия в области электромагнитных явлений заложили основу всей современной электротехники, без которой уже трудно представить существование современного человечества. Развитие атомной физики привело к решению не только кардинальных проблем этой науки и перевороту в понимании строения окружающего нас материального мира, но и к величайшим политическим последствиям. В самом деле, вся послевоенная политика крупнейших держав, определившая в значительной степени жизнь человечества на нашей планете, строилась под знаком непрерывного соревнования в области совершенствования ядерного оружия.

Нуждаясь во все большей энергии и зная о том, что ее некоторые химические виды (нефть, уголь и др.) истощаются, человечество с надеждой обращает свой взор опять-таки к достижениям физики, рассчитывая на то, что в не столь уж отдаленном будущем заработают термоядерные реакторы, способные полностью обеспечить потребности человечества в энергии.

Достижения в области физики твердого тела позволили создать современную электронику, микроэлектронику, электронно-вычислительные машины, без которых уже невозможен прогресс в любых направлениях человеческой деятельности и роль которых непрерывно возрастает, ставя перед человечеством проблемы не только технического характера, но и социального и философского.

Старый раздел физики—оптика—в наше время в связи с открытием лазеров и развитием техники переживает невероятно бурный подъем. Появилась возможность не только воздействовать на вещество мощными потоками света, но и осуществлять связь на очень большие расстояния с большой емкостью информации, создавать объемные изображения, не отличающиеся от обычных видимых глазом изображений, решать целый ряд других важнейших задач.

В свое время великий русский ученый М.В. Ломоносов сказал: «Широко простирает химия руки свои в дела человеческие». С еще большим основанием можно сказать, что еще шире простирает руки свои в дела человеческие физика. Более того, теперь уже хорошо известно, что не только строение молекул, но и ход любой химической реакции может быть объяснен и даже во многих случаях рассчитан с достаточной степенью точности с помощью приемов, разработанных физи-

ками. Это дало основание некоторым физикам сказать, что химия есть часть физики, которая занимается вопросами строения и взаимодействий многоатомных молекул и их преобразованием при химических реакциях.

Не следует, конечно, думать, что химия как наука идет к концу. В самом деле, хотя мы сейчас уверены, что все химические реакции идут по законам, которые открыты в области физики, и что многие химические реакции действительно можно рассчитать, однако нет необходимости проводить такие расчеты во всех случаях.

Химические методы исследования, правила, накопленные в химии, не только не перестают быть нужными, но непрерывно обогащаются, развивая эту область человеческого знания как самостоятельную науку.

Особенностью физики является то, что она широко пользуется математикой. Более того, целый ряд разделов, начиная с основ дифференциального и интегрального исчисления, был создан физиками и под влиянием потребностей развития физики. Тесная взаимосвязь физики и математики была с самого зарождения этих наук их характерным свойством. Возможно, это объясняется тем, что физика, несмотря на всю ее сложность, все же изучает относительно простые соотношения между явлениями окружающего нас мира.

Великий физик Альберт Эйнштейн сказал, что строение атома—это детская игра по сравнению с детской игрой. Этим он хотел подчеркнуть, что несмотря на все величие физического знания, оно все-таки ограничивается значительно более простыми проблемами, чем, скажем, психология.

В иерархии естественных наук, таких, как физика, химия, биология, медицина, физика занимает первую ступень. В этом смысле ее надо считать простейшей. Но как нельзя построить здание, начиная с крыши, и необходимо начинать все-таки с фундамента, так и нельзя построить здание всех остальных наук без фундаментальных сведений, излагаемых в курсе физики.

Физика является основой учения о материальном мире и демонстрирует искусство решать проблемы этого мира. Базируясь на точном эксперименте, развитом математическом аппарате, всей мощи логического и интуитивного способа получения знаний, физика достигла своего совершенства именно потому, что полностью использовала все возможности, заключенные во всех этих отдельных сторонах познания.

Установить границы будущего развития физики, конечно, невозможно. Она будет продолжать развиваться в глубину, получая все более и более фундаментальные сведения об окружающем нас материальном мире, и вширь, создавая отдельные прикладные, но не менее важные для развития всего человечества и для его жизни направления. Одно можно сказать с достаточной определенностью. Предметом исследования физики всегда будут такие объекты и явления и соотношения между

отдельными объектами и явлениями, которые можно выразить либо на языке чисел, либо на языке четких понятий, в отношении которых можно сказать, истинны они или ложны.

В этом смысле такое важнейшее человеческое чувство, как любовь, являющееся безусловным фактом, никогда не будет предметом изучения физики. И не потому, что это чувство не столь уж важно, а просто потому, что оно по своему характеру не вписывается в те представления, в те способы описания, которые свойственны физике.

Изучение физики наряду с общеобразовательным, что, безусловно, само по себе принципиально важно, для специалистов в разных областях имеет и чисто прагматическое значение. Современное общество насыщено техникой. Для понимания работы этой техники необходимо знание вполне определенных разделов физики, прежде всего механики, учения об электромагнитных явлениях, оптики и др.

В сельском хозяйстве используется воздействие самых разнообразных полей на растения и на семенной материал. Никакая промышленность не может развиваться без метрологической службы, а она, в свою очередь, базируется на разнообразных специальных приборах, микропроцессорной технике и т. д.

Важнейшим процессом в природе является фотосинтез. Углубленное понимание этого процесса требует знания о взаимодействии света и вещества, строения вещества, тех преобразований, которые в веществе происходят при поглощении световой энергии. В основе селекции и выведения новых сортов растений лежит учение о механизме наследственности и о законах генетики. Без знания физики понять этот механизм и эти законы невозможно.

Может возникнуть вопрос: а не достаточно ли специалистам (не физикам) тех знаний, которые они получают уже в школьном курсе? Ведь не изучают же, например, в вузе сельскохозяйственного профиля такие предметы, как общая история, география, которые для жизни человека также весьма важны. Ответ на этот вопрос может быть только один: нет, недостаточно. В самом деле, при прохождении курса физики в школе приходится учитывать возраст учащихся. Хотя курс физики в школе является достаточно фундаментальным, но все-таки он рассчитан не на взрослых людей. Основная цель курса физики в вузах—за сравнительно небольшое число часов дать краткий, построенный на одном и при этом повышенном уровне очерк состояния этой области науки. Необходимость такого изучения курса физики ощущалась постоянно. Современная физика есть часть общечеловеческой культуры. Даже при чтении газет мы непрерывно сталкиваемся с обсуждением проблем энергетики, охраны окружающей среды и т. д. Все эти вопросы также невозможно понять, не зная физики в ее современном виде. Таким образом, изучение физики в нефи-

зических вузах есть не дань традиции, а абсолютная необходимость.

Несколько слов об отборе материала для учебника, о его концепции. Ограниченное время, отводимое на изучение физики в высших учебных заведениях химического, сельскохозяйственного и биологического профилей, а также в технических вузах с уменьшенным курсом физики, делает практически невозможным изложение, хотя бы и в сокращенном виде, всех традиционно составляющих курсы физики крупных разделов этой науки. Любая попытка «объять необъятное» с неизбежностью превратит учебник в подобие справочника, где понемногу говорится обо всем, но который ничему научить не может. Субъективизм в отборе материала поэтому неизбежен.

Исходя из этого, а также принимая во внимание, что студенты таких вузов должны иметь отчетливые знания достаточно обширного курса физики, авторы учебного пособия сознательно пошли на исключение целого ряда крупных разделов физики, таких, например, как физика твердого тела, физика ядра и элементарных частиц и др. Наоборот, усилен раздел о квантовой теории строения вещества (молекул), введена специальная глава о развивающихся процессах и т. д. Сделано это с целью дать студентам основу для понимания современной теоретической химии, биохимии, молекулярной биологии, процессов развития в живой природе. Довольно детально изучается механика, так как она не только составляет основу физики и имеет важнейшее прикладное значение, но и позволяет, опираясь на уравнение движения, продемонстрировать идеал любой науки—способность с высокой точностью предсказать развитие событий. Включение понятий релятивистской механики «в ткань» классической не только позволяет указать на противоречия последней, но и на сравнительно раннем этапе ввести соотношение  $E = mc^2$ , являющееся, как известно, основой для понимания не только ядерной энергетики, но и, что не менее важно, обычной.

В разделах об электромагнитных явлениях, оптики, термодинамики выделены наиболее фундаментальные положения, формирующие общий взгляд на явления, способы их описания и т. д., а также непосредственно используемые при изучении ряда специальных курсов. Где это уместно, авторы старались продемонстрировать «работу» физических законов на биологических и других примерах. Авторы стремились к тому, чтобы у студентов сложилось достаточно четкое представление о качественных соотношениях между физическими характеристиками. Цель научить студентов производить желаемые расчеты не ставилась. Поэтому лишь кратко затрагиваются вопросы систем единиц. Эту сторону дела предполагается раскрыть в специальном задачнике, который подготавливается к печати.

Сложность математического аппарата не выходит за рамки общематематической подготовки студентов соответствующих вузов. Авторы сознательно стремились к тому, чтобы студенты

получали необходимые формулы не в готовом виде, а могли как можно в большей степени сами провести необходимые выкладки. Это единственный прием, который может подкрепить простое запоминание созданием необходимых логических связей, а также дать определенную тренировку в строгом мышлении и овладении математическим аппаратом, в частности, элементами математического анализа и теории дифференциальных уравнений. В век компьютеризации без этого уже не обойдешься.

## 2. Вопросы методологии физики

Некоторые общие положения. В основе всех современных знаний в области физики лежат эксперименты и наблюдения над природными явлениями, когда эти природные явления по каким-либо причинам не могут быть воспроизведены в лаборатории. Однако физик всегда стремится к тому, чтобы то или иное изучаемое им явление было воспроизведено в лабораторных условиях. Это необходимо, так как только в лабораторных условиях можно выделить сам изучаемый процесс в его чистом виде, исключив массу побочных мешающих факторов, которые, как правило, сопутствуют соответствующему наблюдению в природных условиях.

Ум человеческий ограничен, только глупость бесконечна. Поэтому человек и не может охватить то или иное явление сразу во всей его сложности. Познание сложного явления всегда требует предварительного познания отдельных его составляющих (анализа). Лишь на последующих этапах происходит объединение отдельных сторон познаваемого явления в целостную картину (синтез). Такой анализ проще всего производить с помощью лабораторных экспериментов. Но после того как серия экспериментов проведена, начинается второй процесс, сплошь и рядом во времени развивающийся в известной степени параллельно первому. Это процесс поиска взаимосвязи между отдельными сторонами явления, выделение чего-то общего, обобщение. Такое обобщение, если оно охватывает весьма широкую область физических явлений, называется *физическим законом* или *принципом*. Как показывает история физики, подобное обобщение не всегда и даже весьма часто не является следствием чисто логических построений. Нередко окончательные выводы и обобщения не следуют формально логически из той совокупности фактов, которые к данному моменту накоплены наукой. Более того, иногда бывает чрезвычайно трудно поставить такие решающие эксперименты, обработка результатов которых с помощью методов формальной логики привела бы к окончательному выводу. Это относится даже к такому, казалось бы, простому закону, каким является всем хорошо известный второй закон Ньютона, характеризующий связь между силой, массой тела и ускорением, которое оно приобретает под действием этой силы.

Когда Максвелл записал свою знаменитую систему уравнений, он не располагал всеми необходимыми экспериментальными сведениями для построения этих уравнений. Отдельный дополнительный член в одном из этих уравнений был введен им из эстетических соображений красоты и симметрии этих уравнений. Лишь затем выяснилось, что введение этого нового члена позволило получить фундаментальный вывод о том, что свет представляет собой не что иное, как электромагнитную волну, и, более того, обосновать существование самой этой волны. Последующие эксперименты подтвердили этот вывод. Итак, в основе крупных обобщений, как правило, наряду с чисто логическими построениями и с экспериментальными основаниями значительную роль играет интуиция. Мы не знаем законов, по которым производится интуитивное построение в уме человека, однако оно существует и нередко оказывается правильным. При этом в своих рассуждениях физики постоянно пользуются методом, который в шутку можно назвать методом Микеланджело. Рассказывают, что когда этого великого живописца, скульптора и архитектора спросили, как ему удастся ваять столь прекрасные статуи, то он ответил, что это очень просто: надо взять кусок мрамора и отсечь все лишнее!

Приведем в этой связи лишь один простой пример.

Предположим, что мы проделываем эксперименты с целью выявить зависимость скорости движения тел от времени в разных условиях. В первой серии экспериментов первоначально разогнанное до скорости  $v_0$  тело в дальнейшем скользит по некоторой поверхности. Поскольку в этом варианте всегда присутствует трение скольжения, то скорость тела со временем будет уменьшаться вплоть до нуля (если время проведения опыта достаточно велико). Выбирая разные подложки, проведем эксперименты при разных коэффициентах трения, постепенно уменьшая его. Если построить графики зависимости скорости тела от времени, то получим серию прямых, изображенных на рис. В.1,а.

Пусть теперь тело движется по поверхности жидкости. Тогда оно будет преодолевать силы вязкого трения. Если менять жидкость, уменьшая ее вязкость, то будет наблюдаться картина, показанная на рис. В.1,б. Какие выводы можно сделать из серии подобных экспериментов? Прежде всего бросается в глаза то, что во всех случаях скорость движения убывает со временем. Значит, тело рано или поздно остановится. Именно такой вывод и сделал в свое время Аристотель, утверждавший, что все тела, однажды приведенные в движение, рано или поздно останавливаются.

Можно, однако, взглянуть на результаты всех экспериментов с другой точки зрения, обратив прежде всего внимание на то, что вне зависимости от фактора торможения (шероховатая поверхность или жидкость) при уменьшении трения графики зависимости скорости от времени приближаются к горизонтальной прямой, отвечающей начальной скорости. Выполняя достаточно большую серию экспериментов с постепенно уменьшаю-



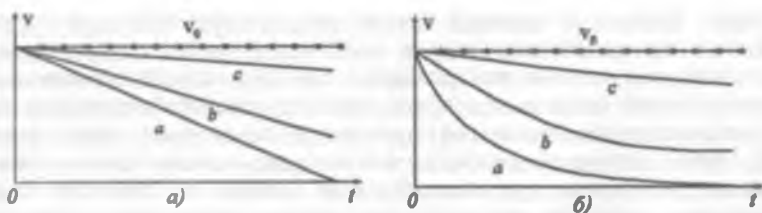


Рис. В.1. Зависимость скорости движения тела от времени при торможении за счет трения скольжения (а) и вязкого трения (б). Трение уменьшается в последовательности  $a \rightarrow b \rightarrow c$

щимся трением, увидим, что, во-первых, если время эксперимента не очень велико, то при малых значениях коэффициентов трения скольжения или вязкости жидкости график зависимости скорости движения тела от времени в любом случае будет представлять собой прямую, и, во-вторых, что угол между этой прямой и прямой, отвечающей начальной скорости, будет стремиться к нулю.

Отсюда можно заключить, что если отсечь мешающие факторы, то выделится нечто главное—способность тела сохранять свое равномерное и прямолинейное движение в отсутствии действующих на него сил. Этот вывод и сделал Г.Галилей, сформулировав знаменитый принцип инерции.

Можно поэтому сказать, что Аристотель лишние куски мрамора оставил, а Г. Галилей все лишнее убрал.

Именно этим качеством и отличаются все формулировки физических законов: они представляют собой концентрат лишь самого нужного без мешающих моментов.

В этом заключается сила законов физики. Однако имеется и слабость. Если мы захотим, опираясь на эти законы, предсказать результаты тех или иных экспериментов в сложных конкретных ситуациях, то, поскольку учесть точно все «мешающие факторы» практически невозможно, результат предсказаний совпадет с реальностью лишь с той или иной степенью точности. К счастью, эта точность во многих случаях оказывается высокой.

При этом в физике экспериментальным фактом считается только такой результат, который удастся получить независимо в разных лабораториях и разными исследователями. В противном случае каким бы интересным ни показалось то или иное сообщение в научном журнале о проведенном эксперименте, ему еще не придают значения факта до тех пор, пока в другом месте и другими людьми не будет получено то же самое. Неуклонное следование этому принципу избавляет от засорения науки разного рода спекуляциями, чего, к сожалению, в реальной жизни не так уж мало.

Важно заметить, что если обобщение охватывает широкий круг явлений и еще что-то, что в первоначальных экспериментах не наблюдалось, то это обобщение является генератором новых идей. В результате такого обобщения, как следствие, можно

получить указание на постановку новых экспериментов, которые, в свою очередь, расширяют наши сведения об окружающем мире. Ярким примером такого типа обобщения было построение уравнения Дирака, которое стало одним из величайших достижений квантовой механики. Прямым следствием этого уравнения было выявление наряду с электроном частицы, которая получила название позитрон. В момент написания Дираком уравнения эта частица не была известна, однако дальнейшие эксперименты подтвердили существование не только этой частицы, но и целого антимира. Итак, развитие физики как науки демонстрирует диалектическое правило спирали познания, которое известно в философии. Эту спираль можно изобразить следующим образом: эксперимент, логическая обработка, интуитивное обобщение, формулировка закона, изучение следствий из этого закона, указание на новые возможные эксперименты, получение в результате этих экспериментов нового знания, новая логическая обработка, снова интуитивное обобщение и т. д.

Конечно, это только грубая схема. В реальной жизни все происходит гораздо сложнее. В самом деле, ученые ведут исследования, подчиняясь «внутренней логике» науки. Этим термином обычно обозначается смутное внутреннее осознание ученых, что то или иное исследование надо вести в данном направлении. На этом пути может обнаружиться факт, совершенно неожиданный, который никак не укладывается в рамки уже накопленных знаний. Такая ситуация в физике встречалась нередко, и именно при попытках объяснения необычных результатов экспериментов возникал скачок в развитии науки. Не случайно среди физиков бытует шутливое высказывание, что чем дальше эксперимент от теории, тем ближе к Нобелевской премии. Известно, что Нобелевские премии даются за величайшие достижения в научных исследованиях, нередко действительно неожиданные, которые нельзя предсказать формальным логическим путем, опираясь на предыдущий ход развития науки.

Однако история развития науки научила физиков следующее: если в процессе экспериментального исследования появились результаты, объяснение которых, на первый взгляд, нельзя дать в рамках существующих представлений, не надо спешить утверждать, что здесь действительно появляется «таинственный незнакомец». Прежде всего нужно попытаться объяснить такое явление в рамках уже сложившихся представлений. Только тогда, когда эти возможности исчерпаны, нужно вводить что-то принципиально новое. Очень часто такой здоровый физический консерватизм не только избавляет от обилия совершенно ненужных ошибок и ненужной траты времени, но и приводит к весьма существенным открытиям.

В нашем веке в результате экспериментов по радиоактивному распаду было обнаружено явление, которое на первый взгляд противоречило закону сохранения энергии. Возникли даже утверждения, что этот закон имеет ограниченное хождение. Однако

впоследствии выяснилось, что закон сохранения энергии продолжает оставаться правильным, а наблюдаемое явление кажущегося отклонения от него объясняется тем, что при радиоактивном распаде появились новые частицы, получившие название нейтрино, которые в то время не были известны. Впоследствии свойства этих частиц были хорошо изучены и закон сохранения энергии оказался «спасенным».

Но история физики учит и другому: никогда не надо абсолютизировать те или иные физические законы. Мы можем быть уверены в том, что эти законы выполняются в пределах определенного круга явлений. Но если само новое явление не входит в этот круг, то закон может быть и неверным.

Тем не менее разумный консерватизм нужен. Еще до появления сомнений в отношении действия закона сохранения энергии при радиоактивном распаде Российская Академия наук еще в прошлом веке приняла постановление о том, что никакие проекты вечного двигателя не рассматриваются этой Академией и отправляются либо в архив, либо уничтожаются. Такое решение было вполне закономерным. Дело в том, что практически все проекты вечных двигателей основаны либо на обычных механических устройствах, либо на устройствах, которые используют электромагнитные, тепловые и некоторые другие явления.

Незыблемость в кругу этих явлений закона сохранения энергии доказана настолько хорошо и таким количеством экспериментов, что вероятность того, чтобы что-то осталось незамеченным, практически равна нулю. Однако если когда-нибудь обнаружатся явления, которые не сводятся к механическим, электромагнитным, оптическим и другим, хорошо изученным, то не исключена возможность того, что закон сохранения энергии окажется нарушенным.

История физики научила и другому: хорошо обоснованные законы физики, верные для большого круга изучаемых явлений, как правило, при дальнейшем развитии физики и при появлении новых фактов не «отменяются», они оказываются частным случаем некоторых более общих законов.

Все развитие физики показывает, что мы непрерывно движемся от понимания отдельных, частных ко все более общим законам природы. Появление механики больших скоростей (релятивистской механики) Эйнштейна отнюдь не «отменило» классическую механику Ньютона. Последняя оказалась следствием механики Эйнштейна при условии, что скорости движения являются малыми по сравнению со скоростью света. В свою очередь, законы макроскопической механики являются следствиями законов квантовой механики, управляющих микромиром. Можно показать, как квантовое уравнение Шредингера при определенном предельном переходе превращается в уравнение движения классической механики.

Громадное количество самых разнообразных взаимодействий в природе удалось свести к четырем:

1) сильные взаимодействия, проявляющиеся в ядрах атомов;

2) электромагнитные взаимодействия, которые еще в начале прошлого века рассматривались как отдельные электрические и магнитные взаимодействия;

3) гравитационные взаимодействия;

4) слабые взаимодействия, проявляющиеся в микромире при некоторых процессах взаимопревращений элементарных частиц.

В последние годы удалось объединить электрические и слабые взаимодействия, построив теорию электрослабых взаимодействий.

В свое время А.Эйнштейном была поставлена задача о сведении всех взаимодействий к одному какому-нибудь виду.

В этом смысле можно говорить о том, что с каждым новым шагом развития физики основные законы физики и теории упрощаются, а число их уменьшается. Все большее число известных раньше законов и положений становится следствием более общих. При этом старые утверждения как часть новых можно вывести, опираясь на законы формальной логики. Например, по мере того как развивается физика, число фундаментальных частиц и типов взаимодействий, как уже указывалось, становится все меньше и меньше. Почему так происходит? Мы этого не знаем, но это исторический факт. И этот факт интуитивно еще в XIV в. был осознан английским философом У. Оккамом и получил название «бритвы Оккама». Утверждение заключается в том, что чем ближе мы находимся к некоторой истине, тем проще оказываются основные законы, выражающие истину.

Значит ли это, что при изучении физики нужно всегда исходить из основных законов, чтобы уже дальше логически вывести всю невероятно большую гамму следствий? Конечно, нет. Сегодня мы знаем, что, опираясь на уравнение Шредингера, на законы квантовой механики, «управляющие» движением микрочастиц, при определенном предельном переходе можно получить уравнение движения Ньютона и, следовательно, закон, который «управляет» макроскопическими явлениями, составляющими сферу классической механики. Но какой смысл каждый раз изучать то или иное явление и получать тот или иной более частный вывод, начиная с самого начала? В этом нет никакого резона, не говоря уже о большой сложности. Развитие физики, появление все более и более фундаментальных законов ни в коем случае не делает менее нужными законы частного характера. Поэтому традиционное деление физики, возникшее исторически, на механику, термодинамику, учение об электромагнитных явлениях, атомную физику, квантовую физику и т. д. не только не потеряло своего значения, но продолжает существовать и сейчас, составляя основу практических приложений в любой сфере, которая вписывается в рамки соответствующих данному разделу явлений.

Тем не менее, как сказал известный ученый Поль Дюгем, физика не механизм, который можно разбирать и развинчивать,

физику надо воспринимать в целом. Отсюда, конечно, не следует, что нельзя при изучении этого механизма более или менее пристально рассматривать отдельные составные части.

**Модели в физике. Роль математики.** Важнейшим понятием в физике является понятие модели. Под моделью понимается некоторое упрощенное описание реального объекта, которое достаточно для того, чтобы понять то или иное интересующее нас явление или описать связь между теми или иными отдельными событиями.

В любом физическом явлении, как правило, можно выделить нечто главное, что интересует исследователя в данной конкретной ситуации. Все остальное является как бы мешающими побочными факторами. Вот это главное должно быть выделено в модели. Так, например, если надо ответить на вопрос, где находится самолет, летящий из Москвы в Санкт-Петербург, и описать его движение, то неважно, какую форму имеет этот физический объект, каковы размеры крыльев, фюзеляжа, сколько двигателей и т. д. Вполне можно представить этот самолет в виде маленькой точки и характеризовать положение самолета на пути от Москвы до Санкт-Петербурга расстоянием от этой точки до какого-то определенного пункта в Москве или Санкт-Петербурга. Однако если воспользоваться той же моделью для описания движения самолета в пределах аэродрома, это может привести к серьезным неприятностям.

При описании такого сложного объекта, каким является молекула, также можно применять в зависимости от ситуации различные модели. Если необходимо исследовать закономерности, связанные с давлением газа внутри замкнутого сосуда, то вполне достаточно представить молекулы, из которых состоит газ, в виде отдельных шариков. Но если исследуются более сложные явления, связанные, например, со спектральным проявлением молекул или с их реакциями, то такого представления недостаточно. Нужно не только указать, из каких атомов состоит молекула, но и представить ее в виде некоторой пространственной геометрической фигуры с определенными длинами связей и валентными углами, привести массы атомов, которые входят в состав молекулы, охарактеризовать распределение зарядов внутри молекулы и т. д. Степень сложности описания объекта должна отвечать степени сложности того явления, которое надо качественно или количественно охарактеризовать.

Понятие модельного представления верно не только по отношению к самому объекту, но и к условиям, в которых объект наблюдается. Так, например, при изучении движения тела, брошенного под углом к горизонту, можно принять или не принять во внимание сопротивление воздуха. При изучении оптических явлений в некоторых случаях можно не учитывать существование дифракции света и пользоваться представлениями геометрической оптики. При исследовании строения атома нас обычно не интересуют детали строения его ядра и можно

ограничиться представлением о ядре, как о некотором очень маленьком шарике, имеющем вполне определенные заряд и массу. При описании работы многих механизмов на первых этапах рассмотрения можно пренебречь эффектом трения и т. д. То, что мы вынуждены пользоваться при описании различных объектов вполне определенными моделями, связано не просто с невозможностью описания во многих случаях реальной ситуации во всей ее полноте, а главным образом с тем, что слишком подробное описание затеняет существо дела и делает невозможным сам процесс познания. Любая модель является некоторой абстракцией, некоторым упрощенным отражением реальной действительности. Именно с такими абстракциями всегда связана физика, особенно теоретическая, в задачу которой входит не просто описание данного объекта, но и предсказание его поведения с достаточно хорошей степенью точности на некоторый период времени вперед.

Выше мы уже сказали, что физика широко пользуется математикой и что это составляет не только ее особенность, но и главную силу. Однако при всей своей мощи математика в физике играет подчиненную роль. Математика используется для описания поведения вполне определенных объектов и явлений, которые описываются в физике с помощью моделей. Если эта модель выбрана неправильно, то как бы мы ни пытались исправить эту неправильность применением математического аппарата, все равно ничего не получится. Эта мысль в свое время была очень хорошо выражена Т. Гексли: «Математика подобно жерновам мельницы перемалывает то, что под нее засыпали». И как, засыпав плевелы, мы не получим доброкачественной муки, так и, заложив неправильную модель, мы не получим правильного ответа, какой бы математикой ни пользовались. Так, например, каким бы ни было теоретическое представление, с помощью которого мы хотели бы объяснить некоторые присущие только белковым молекулам специфические проявления, эта теория заведомо будет несостоятельной, если только в ней в самом начале при выборе соответствующей модели не заложена разница между простейшими полимерными молекулами типа полиэтилена и сложными белковыми образованиями.

Однако не следует сводить роль математики в физике только к производству вычислений. Д. Гиббс сказал: «Математика—это язык». Применительно к физике этот язык чрезвычайно емкий, четкий, образный. Именно на языке математики удается выразить наиболее простым и наглядным образом причинно-следственные соотношения между отдельными явлениями, что, в свою очередь, позволяет существенно расширить мощь познаний и предсказать ход событий.

Измерения в физике. Прямые и обратные задачи. Основания получения физических знаний—наблюдение природных явлений и лабораторный эксперимент, причем чаще всего именно

БИБЛИОТЕКА

№ 40853

последний. В свою очередь, в основе обработки эксперимента лежат измерения.

Измерения в физике принято делить на *прямые* и *косвенные*. Прямые—это те измерения, которые можно проделать либо непосредственно, полагаясь на имеющиеся у человека органы чувств, либо с помощью самых простейших приборов. Например, когда нужно измерить расстояние между некоторыми объектами, просто прикладывая измерительную линейку, это будет прямое измерение. Если мы взвешиваем какое-то тело на весах, это тоже прямое измерение. Однако если необходимо измерить расстояние между атомами в некотором кристалле, опираясь на дифракционную картину рассеяния рентгеновских лучей, это уже будет косвенным измерением. Такие косвенные измерения в большинстве случаев и делаются, составляя специфику современной науки, особенно в случае микромира. Для косвенных измерений характерно еще то, что мы измеряем одно, а хотим получить сведения о совершенно другом. В том же примере об измерении расстояния между атомами в кристалле непосредственно измеряемыми и наблюдаемыми являются не эти расстояния, а рефлексы или следы на фотографической пластинке, которые получаются после прохождения рентгеновского излучения через кристалл. Обработывая эти рефлексы, получают сведения о расстояниях между атомами, которые непосредственно не наблюдаются.

Другой пример: предположим, что необходимо собрать сведения об упругостях химических связей в многоатомной молекуле. Это можно сделать, если наблюдать спектр поглощения электромагнитного излучения в инфракрасной области. Такой спектр можно получить при прохождении излучения через слой газа или жидкости, состоящей из этих молекул. Это спектр поглощения на определенных длинах волн, а надо получить сведения о силах, которые отвечают химическим связям в многоатомных молекулах.

Использование косвенных измерений является неизбежным и отражает тот факт, что мы, к сожалению, обладаем хотя и развитыми, но все-таки весьма ограниченными по своим возможностям органами чувств. В самом деле, человеческий глаз способен в лучшем случае различать расстояние между двумя точками, которое не меньше нескольких десятых долей миллиметра. Рукой же нельзя ощутить разности в весе в несколько граммов. Мы не можем заглянуть внутрь атомного ядра и т. д. Поэтому в процессе развития цивилизации человек усилил свои органы чувств, используя разнообразные приборы, назначение которых, в сущности, состоит в том, чтобы перевести не наблюдаемую непосредственно человеком особенность какого-нибудь объекта в такие проявления, которые человек способен наблюдать с помощью своих органов чувств. Здесь возникает очень важный и интересный вопрос: как можно установить правильность той или иной модели нашего объекта или явления,

ее адекватность реальной окружающей нас природе, если в принципе мы не в состоянии непосредственно своими органами чувств наблюдать и измерять те параметры, которые характеризуют эту модель? Для этого, очевидно, необходимо располагать сведениями о связи между параметрами модели объекта или явления и теми их проявлениями в условиях конкретного эксперимента, результаты которых уже можно наблюдать с помощью органов чувств. Если такая связь установлена и она единственная, то тогда, предполагая возможность существования различных моделей, описывающих определенный круг явлений или объектов, можно предсказать, какие именно макроскопические, т. е. непосредственно наблюдаемые нашими органами чувств, явления мы должны наблюдать с помощью определенной совокупности экспериментальных приемов. Если некоторые из этих предсказаний совпадают с реальным наблюдением, то мы говорим, что соответствующая модель отвечает действительности. Непосредственная связь между параметрами объекта или явления и его наблюдаемыми проявлениями устанавливается, как правило, математическим путем. Эта связь может быть выражена в простейшем случае короткой формулой, состоящей всего из нескольких членов, или сложной формулой, системой уравнений, а в наиболее общем случае—алгоритмом или правилом, применение которого однозначно приводит к предсказанию экспериментально наблюдаемого результата. Такая алгоритмическая связь между параметрами объекта и его наблюдаемыми проявлениями устанавливается, как правило, на базе достаточно общих физических законов. И именно то обстоятельство, что эти физические законы позволяют давать правильные предсказания, подтверждает, в свою очередь, их правильность и общность.

В современной физике различают два класса задач: *прямые и обратные*.

Под прямыми задачами понимают расчет или вполне однозначное предсказание макроскопически наблюдаемых проявлений свойств того или иного объекта по заданным характеристикам модели объекта или явления. Под обратными задачами понимается, наоборот, восстановление характеристик данного объекта, явления или круга явлений по заданным экспериментальным наблюдениям.

В сущности, наиболее общим случаем решения таких обратных задач и является получение обобщений в форме физических законов на основании отдельных частных наблюдений. Ценность и общность таких обобщений (моделей) как раз и определяется количеством тех экспериментально наблюдаемых фактов, которые можно получить, если заданы конкретные характеристики модели.

Пример прямой задачи в физике—предсказание движения межпланетных космических кораблей на основании вполне определенной модели с указанием расположения в настоящем и ближайшем будущем планет и некоторых других тел солнечной



системы и законов движения тел в поле сил тяготения. Пример обратной задачи—восстановление орбиты движения малых небесных тел—комет и астероидов—по наблюдениям их положения в определенные моменты времени. Таким способом, например, в свое время была определена орбита кометы Галлея, приход которой в район Земли произошел в конце 1986 г. В процессе физического познания прямые и обратные задачи переплетаются. Как правило, обратные задачи решаются итерационным путем. Вначале мы имеем некоторые представления об объекте, которые характеризуем с помощью определенной совокупности качественных или количественных параметров. После этого, опираясь на развитые методы решения прямых задач, предсказываем результат того или иного эксперимента, который в дальнейшем и производится. Если совпадение предсказанных и реально измеренных результатов не очень хорошее и не удовлетворяет нас, то мы начинаем изменять параметры исходной модели, стремясь добиться такой ситуации, чтобы совпадение предсказанных и наблюдаемых эффектов было удовлетворительным. Если это удастся, то модель считается, при некоторых дополнительных условиях, подтвержденной.

Это совпадение может быть достигнуто на качественном уровне, т. е. когда указываем просто, что должно наблюдаться то или иное явление, и количественном уровнях, т. е. когда не только предсказываем, что то или иное явление должно наблюдаться, но и указываем соответствующую величину, например положение или яркость линии в спектре. Конечно, объем знаний в данной модели существенно возрастает, если удастся перейти на количественный уровень. В том-то и состоит значение современной физики, что она не только позволяет описать те или иные явления и установить между ними связь на качественном уровне, т. е. понять, что явление *A* является следствием явления *B*, но и предсказать это следствие на таком количественном уровне, который вполне проверяется экспериментом и удовлетворяет точности этого эксперимента. В упрощенном курсе, который предлагается читателям, мы, конечно, лишены возможности остановиться на этой сильной стороне физики, т. е. на возможности численно с хорошей точностью предсказывать те или иные конкретные события. Однако во всех случаях, когда это можно сделать, мы будем демонстрировать на конкретных примерах или указывать, каким способом можно произвести то или иное вычисление, опираясь на общие установленные физикой законы.

Заметим, что хотя этот курс и представляет некое достаточно единое изложение физики, начиная с самых основных понятий, однако мы ни на секунду не забываем, что его читатели уже должны быть знакомы с физикой в объеме средней школы. Это означает, что они уже располагают сведениями о большом числе фундаментальных экспериментов, которые привели к установлению тех или иных физических законов. Они знают о

существовании законов, знают во многих случаях их формулировки и простейшее математическое описание. Это дает возможность при изложении вузовского курса не придерживаться некоторой строгой последовательности и оперировать в разделе «Механика» понятиями электрического заряда и поля, считая, что эти понятия уже известны из школьного курса, хотя к вопросам существования электрических зарядов и разного рода закономерностей, связанных с проявлением взаимодействия между этими зарядами, мы подойдем только в разделе «Электромагнитные явления».

Аналогично будем поступать и в других случаях, где это необходимо по ходу изложения и может обратить внимание читателей на важные и интересные моменты.

### 3. О размерностях физических величин

Еще древнегреческие философы, заложившие основы науки об окружающем нас мире, утверждали, что все измеряется числом и мерой. Хотя мы знаем, что есть явления, числом и мерой не характеризуемые, однако и в настоящее время это утверждение сохраняется: все величины в физике характеризуются числами, которым приписывается определенный смысл, или размерность. Например, расстояние от одного предмета до другого составляет 100 м. Бессмысленно написать число 100, не указав, к чему оно относится. Сравнивать между собой можно только числа, имеющие один и тот же смысл, т.е. размерность. Например, длину с длиной, интервал времени с другим интервалом и т. д.

Большое число различных физических понятий и величин приводит к большому разнообразию размерностей.

Среди них принято различать *основные* и *производные*. В качестве основных выбирают те, которые связаны с наиболее фундаментальными понятиями.

Таковыми являются, в частности, пространство и время. Именно в пространстве и во времени развиваются любые события. Поэтому в качестве основных выбраны размерности длины  $[L]$  и времени  $[T]$ . После выбора размерности остается договориться, что конкретно принять в качестве единичной меры. Так, например, до недавнего времени в качестве единичной меры длины был принят определенный эталон, материально зафиксированный в виде металлического стержня, хранящегося в Международной палате мер и весов. Сейчас в качестве метра принята длина, отвечающая  $1650763,73$  длин волн излучения, которое возникает при переходах между определенными уровнями энергии атомов криптона-86. Основным требованием к любым подобным эталонам является возможность их максимально точно воспроизведения в разных странах и районах этих стран, а также в случае утраты в результате какой-нибудь катастрофы.

Эталоны и правила их воспроизведения устанавливаются международными договоренностями и меняются по мере развития науки и средств измерений.

Кроме размерностей длины и времени одной из основных является размерность массы, так как масса—фундаментальное свойство объектов. Кроме того, ее сравнительно легко калибровать. Поэтому, хотя согласно современным представлениям полная внутренняя энергия тела и его масса связаны между собой, удобнее положить в основу размерностей именно размерность массы  $[M]$ , а не энергии. Если размерности длины и времени заданы, то, например, производные размерности скорости и ускорения легко получить, воспользовавшись определением основных:

$$[v] = \frac{[L]}{[T]} = \frac{\text{длина}}{\text{время}},$$

$$[a] = \frac{[L]}{[T]^2} = \frac{\text{скорость}}{\text{время}} = \frac{\text{длина}}{\text{время} \cdot \text{время}}.$$

Размерность силы устанавливается на основании второго закона Ньютона:

$$[F] = [M][L]/[T]^2.$$

Размерность энергии

$$[E] = [M][v]^2 = [M][L]^2/[T]^2.$$

Таким образом, размерности скорости, ускорения, силы, энергии являются производными.

Общим правилом записи всех равенств в физике является одинаковость размерностей в правой и левой частях. Чтобы выполнить это требование, приходится иногда вводить коэффициенты размерности. Так, например, экспериментально устанавливается соотношение между силой притяжения двух планет, их массами и расстоянием между центрами планет (закон всемирного тяготения). Если запишем это соотношение в виде  $f = m_1 m_2 / r^2$  и воспользуемся введенными выше размерностями, то получим

$$[M][L]/[T]^2 \neq [M]^2/[L]^2.$$

Удовлетворить правилу размерностей можно, лишь введя справа дополнительный постоянный множитель, получивший название гравитационной постоянной  $G$  с размерностью  $[G] = [L]^3/[M][T]^2$ . Точно так же, чтобы правильно связать температуру (обозначение размерности  $[\Theta]$ ) газа и среднюю кинетическую энергию одной его молекулы, надо ввести некоторую постоянную (постоянную Больцмана) с размерностью  $[k] = [M][L]^2/[T]^2[\Theta]$ .

Рассмотренные примеры являются частными случаями нередко встречающейся в физике ситуации, когда требование уравнивания размерностей помогает не только проверять правильность записи тех или иных формул, но и корректировать их.

Приведенные размерности физических величин не зависят от конкретного выбора их эталонов и единиц, например метров или сантиметров для  $[L]$ , минут или секунд для  $[T]$  и т.д.

В некоторых случаях одну и ту же физическую величину можно ввести, опираясь на существенно разные эксперименты. Например, величину электрического заряда можно ввести, опираясь на закон Кулона, непосредственно измеряя силу взаимного притяжения двух одинаково заряженных металлических шариков, растягивающих калиброванную пружину.

В этом случае размерность заряда  $[Q] = [M]^{1/2}[L]^{3/2}/[T]$ .

С другой стороны, для определения величины заряда можно воспользоваться и измерением силы постоянного тока. Это можно сделать, либо калибруя источник тока и элементы электрической цепи, либо [как это и делается при введении соответствующей единицы в Системе Интернациональной (СИ)] измеряя силу взаимного магнитного притяжения двух очень длинных тонких параллельных проводов, входящих в состав одной цепи постоянного тока и расположенных на фиксированном расстоянии друг от друга.

Вспомнив, что такой ток связан с количеством заряда, переносимого через сечение проводника в единицу времени, и принимая в качестве основной (опорной) величины именно силу тока (размерность  $[I]$ ), нетрудно сообразить, что размерность заряда окажется следующей:  $[Q] = [I] \cdot [T]$ . Чтобы при таком выборе единицы заряда не возникла ошибка в записи закона Кулона, где в левой части должна стоять размерность силы, вводится постоянная (диэлектрическая постоянная, или диэлектрическая проницаемость вакуума), обратная величине  $\epsilon_0$  и имеющая размерность  $[\epsilon_0] = [I]^2[T]^4/[M][L]^3$ .

Подобные соображения приводят к существованию различных систем единиц. Основной сейчас считается СИ. В русской литературе чаще используется термин Международная система единиц. В этой системе в качестве основных принимаются единицы длины (метр), времени (секунда), массы (килограмм), силы электрического тока (ампер), термодинамической температуры (кельвин), количества вещества (моль) и силы света (кандела).

Разумеется, не составляет большого труда совершить при необходимости переход от одной системы единиц к другой. Этими замечаниями мы ограничимся, так как в дальнейшем в учебнике конкретные вычисления не производятся.

### Вопросы для самоконтроля

1. Роль и значение физики в современном естествознании. Влияние физики на технический прогресс и через него на политику крупнейших мировых держав. Какие примеры этого вам известны?
2. Какими особенностями должны отличаться объекты и явления, чтобы они могли быть предметом изучения физики в настоящем и будущем?
3. Соотношение анализа и синтеза и роль обобщений в процессе познания окружающего мира. Интуитивные и логические построения. Приведите примеры.
4. Перечислите иерархию ступеней познаний в области физики, а также других естественных наук. Почему математика играет такую большую роль в физике и почему ее можно называть языком современных точных наук?
5. При каких условиях то или иное явление можно назвать фактом с точки зрения современной физики?
6. Что такое законы в физике? Приведите примеры общих и частных законов.
7. Что такое модели в физике? Приведите примеры известных вам моделей из разделов физики, изучаемых в школьном курсе.
8. Корабль с космонавтами стартует с Земли к Луне. Когда космонавты могут рассматривать Луну как материальную точку и когда такое упрощенное представление является совершенно неверным?
9. Что такое прямые и косвенные измерения в физике? Приведите примеры. Что такое прямые и обратные задачи?
10. Роль размерностей физических величин. Какие соображения лежат в основе выбора тех или иных размерностей? Приведите примеры уточнения формул с помощью правила размерностей.



# ГЛАВА 1

## КИНЕМАТИКА

*«Движенья нет, сказал мудрец брадатый.  
Другой смалчал и стал перед ним ходить.  
Сильнее бы не мог он возразить;  
Хвалили все ответ замысловатый.  
Но, господа, забавный случай сей  
Другой пример на память мне приводит:  
Ведь каждый день пред нами Солнце ходит,  
Однако ж прав упрямый Галилей».*

А. Пушкин

### 1.1. Системы координат и степени свободы

Первым важнейшим разделом любого курса физики является кинематика. Кинематика—это раздел физики, который изучает характеристики движения тел и изменения их форм, не касаясь вопроса о причинах этих движений. Любое изменение состояния реального объекта в природе или любое изменение состояния модели, будь то простое перемещение или что-нибудь другое, называется *событием*. При изучении простых механических процессов под событием мы будем понимать только изменение либо положения тел, их совокупностей в пространстве, либо относительного расположения отдельных частей сложной системы. Все события происходят во времени и в пространстве. Привычные понятия времени и пространства, с которыми мы сталкиваемся почти сразу, как только начинаем ощущать окружающий нас мир, на самом деле являются сложными философскими категориями. В философском словаре этим понятиям даются следующие определения: «Время—атрибут, всеобщая форма бытия материи, выражающая длительность бытия и последовательность смены состояний всех материальных систем и процессов в мире. Время не существует само по себе, вне материальных изменений; точно так же невозможно существование материальных систем и процессов, не обладающих длительностью, не изменяющихся от прошлого к будущему» (Философский энциклопедический словарь. М., 1983. С. 94).

«Пространство есть форма бытия материи, характеризующая ее протяженность, структурность, сосуществование и взаимо-

действие элементов во всех материальных системах... Пространство и время неразрывно связаны между собой, их единство проявляется в движении и развитии материи» (Там же. С. 541).

Для того чтобы ввести основные характеристики движения, воспользуемся простейшей моделью—*материальной точкой*. Под материальной точкой понимается тело, которое имеет такие размеры, что ими можно пренебречь по сравнению с общим перемещением тела в пространстве. Для материальной точки имеют смысл лишь два вопроса: где эта точка находится в данный момент времени и как она движется. Для описания положения точки в пространстве используется, как известно, координатный способ.

Напомним понятие координаты. Предположим, что нужно охарактеризовать положение точки на некоторой прямой. Можно поступить следующим образом: задать на этой прямой некоторую точку, назвав ее началом координат. Положение любой другой точки на этой прямой определяется расстоянием от точки до начала координат. Это расстояние можно изобразить в виде числа (число мм или км), причем число, определяющее расстояние от начала координат до искомой точки, если эта точка лежит справа от начала координат, принимается положительным, а число, определяющее расстояние от искомой точки, если эта точка лежит слева от начала координат, принимается отрицательным. Такое число называют *координатой данной точки на выбранной оси*. Задавая это число, мы совершенно однозначным образом указываем положение выбранной точки на оси.

Для описания положения точки на плоскости можно ввести две координатные оси, т. е. две прямые, расположенные в простейшем случае перпендикулярно одна другой (рис. 1.1).

Для того чтобы выразить положение точки в этом случае через координаты, надо через эту точку провести прямую, параллельную оси  $X$ , и вторую прямую, параллельную оси  $Y$ . Величины отрезков, отсекаемые этими прямыми на осях ординат, называют координатами данной точки на плоскости. При этом, как и в случае одной прямой, числа, выражающие длины отрезков  $a_x$  и  $a_y$ , берут со знаком плюс, если отрезки лежат на оси  $X$  справа от начала координат и на оси  $Y$  на верхней полуоси, и со знаком минус, если отрезки лежат на отрицательных полуосях координат  $X$  и  $Y$  (см. рис. 1.1).

Если числа  $a_x$  и  $a_y$ , т. е. координаты точки, известны, то можно найти положение точки. Для этого надо поступить следующим образом: отложить по осям координат отрезки  $a_x$  и  $a_y$ , учитывая знаки координат, затем через точки, соответствующие концам этих отрезков, провести прямые, параллельные соответственно осям  $X$  и  $Y$ . Пересечение прямых укажет положение точки  $A$ . Если всегда придерживаться этого правила, то для однозначной характеристики положения точки на пло-

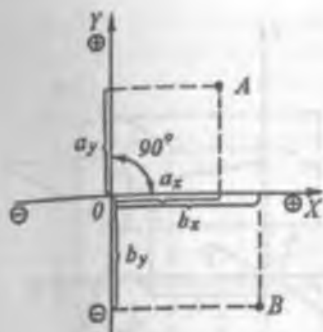


Рис. 1.1. Координаты точки  $A$  есть  $a_x > 0, a_y > 0$ . Координаты точки  $B$  есть  $b_x > 0, b_y < 0$

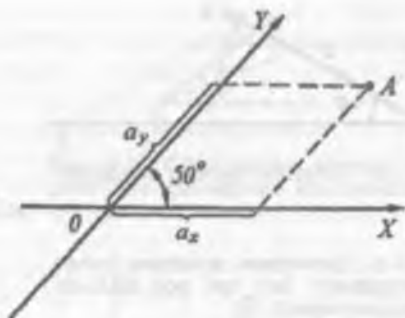


Рис. 1.2. Косоугольная система координат

скости достаточно указать числа  $a_x$  и  $a_y$ . Здесь уместно выделить одно обстоятельство. Для того чтобы охарактеризовать положение точки на плоскости, мы выбрали координатную систему, построенную на двух взаимно перпендикулярных осях  $X$  и  $Y$ . Это *декартова система координат*. Однако такое описание не единственно. Можно выбрать положение осей  $X$  и  $Y$  не под углом  $90^\circ$ , а под каким-то другим. Тогда получим *косоугольную систему координат* (рис. 1.2). В этом случае соответствующие координаты по осям  $X$  и  $Y$  представляют отрезки  $a_x$  и  $a_y$ , построенные так, как показано на рис. 1.2, т. е. снова отсекаемые прямыми, проведенными через точку  $A$  параллельно осям координат.

Еще один способ описания положения точки на плоскости состоит в следующем. Зададим одну только ось  $X$  и из положения, принятого за начало координат, проведем в точку  $A$  отрезок (рис. 1.3). В качестве координаты выбирается длина отрезка и угол между отрезком  $OA$  и прямой  $X$ . Это — *полярная система координат*.

Для характеристики положения точки в пространстве можно ввести три взаимно перпендикулярные координатные оси и задать для них положительное направление, приняв за начало координат точку пересечения осей (рис. 1.4). Чтобы найти координаты некоторой точки  $A$  в выбранной системе осей, надо поступить следующим образом: провести плоскость через точку  $A$  параллельно плоскости осей  $Y$  и  $Z$ , вторую плоскость провести через ту же точку параллельно плоскости осей  $X$  и  $Z$ , третью плоскость провести параллельно плоскости осей  $X$  и  $Y$ . Отрезки  $a_x, a_y, a_z$ , отсекаемые этими плоскостями на осях координат, и будут координатами данной точки в выбранной декартовой системе. При этом, как и в случае точки на плоскости, длины отрезков  $a_x, a_y, a_z$  берутся со знаками плюс, если они откладываются по положительным направлениям осей



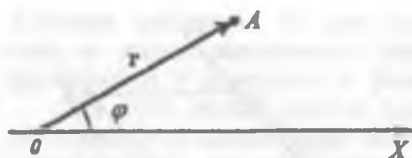
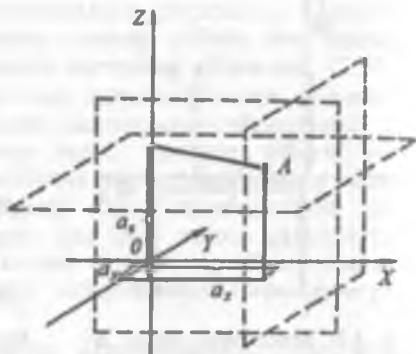


Рис. 1.3. Полярная система координат

Рис. 1.4. Трехмерная декартова система координат (все три оси взаимно перпендикулярны)



координат, и со знаками минус—в противоположном случае. Если заданы координаты точки  $A$ , то можно найти место ее положения, поступая следующим образом: надо отложить отрезок  $a_x$  на оси  $X$  и через точку, где отрезок кончается, провести плоскость, параллельную плоскости осей  $YZ$ . При этом отрезок откладывается справа от начала координат, если  $a_x > 0$ , и слева от него, если  $a_x < 0$ . Далее надо отложить по аналогичному правилу отрезок  $a_y$ , провести вторую плоскость, параллельную плоскости  $XZ$ , и так же поступить с третьей осью. В какой-то точке все плоскости пересекаются и место пересечения этих плоскостей дает положение точки  $A$ .

Так же как и в случае двух координат на плоскости, при описании положения точки в трехмерном пространстве можно поступить и иначе. Можно, например, ввести косоугольную систему (рис. 1.5) или сферическую систему координат (рис. 1.6), в которой уже задается угол  $\varphi$  поворота некоторого отрезка по отношению к выбранной оси, угол  $\theta$  между отрезком  $OA$ , соединяющим начало координат с точкой  $A$ , и плоскостью  $S$  и длина отрезка  $OA$ . Если договориться всегда поступать по определенным стандартным правилам, то, задавая координаты, можно однозначно указать положение любой точки в пространстве.

Чтобы охарактеризовать положение точки на прямой с помощью координаты, необходимо задать одно число. Чтобы охарактеризовать с помощью координат точку на плоскости, нужно задать два числа. Для описания положения точки в пространстве с помощью координатной системы задаются три числа. Принято называть число координат, с помощью которых однозначно задается положение точки или тела в пространстве, *числом степеней свободы*. Каждая координата при этом определяет возможное перемещение точки или тела и называется *степенью свободы*. Если точка находится на прямой линии, то она может двигаться только вдоль этой прямой. В этом случае она имеет одну степень свободы. Если точка лежит на плоскости,

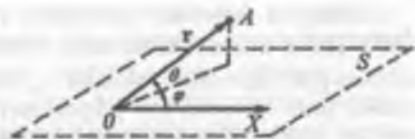


Рис. 1.6. Сферическая система координат для трехмерного пространства

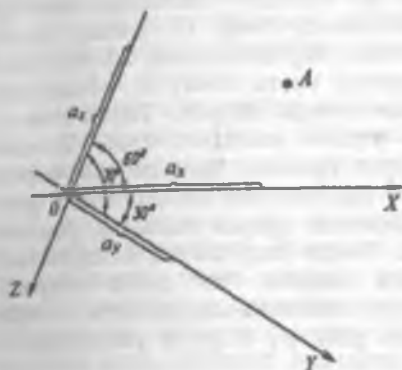


Рис. 1.5. Косоугольная трехмерная система координат

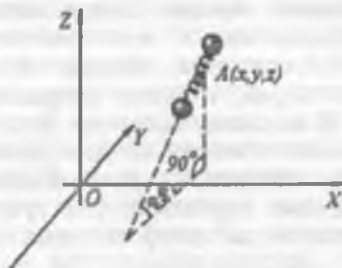


Рис. 1.7. Система с шестью степенями свободы

то она может перемещаться по этой плоскости в горизонтальном и вертикальном направлениях. В этом случае точка имеет две степени свободы. И наконец, точка, находящаяся в пространстве, имеет три степени свободы. Для материальной точки максимальное число пространственных степеней свободы равно трем. Поэтому мы говорим, что окружающий нас мир пространства трехмерен.

Важно также отметить, что для относительно небольших размеров пространства этот мир хорошо описывается *евклидовой геометрией*. Лишь при переходе к более обширным пространствам, связанным, например, с целой Вселенной, выясняется, что обычная геометрия Евклида уже не адекватна природе и надо пользоваться более сложной геометрией—*кривойной геометрией Римана*. В этом случае полагают, что мир искривлен.

Итак, для материальной точки возможны три пространственные степени свободы. Однако если перейти к более сложным телам, то могут возникнуть дополнительные вопросы. Например, если взять два маленьких шара (рис. 1.7), соединенных между собой пружиной, и искать способы описания состояния этой системы в пространстве, то кроме вопроса о том, где находится такая система, возникают еще два других:

1) как эта система повернута по отношению к осям координат, если в качестве осей координат выбрать декартову трехмерную систему?

2) на каком расстоянии находятся друг от друга эти шары?

Конечно, можно поступить так: приписать каждому из шаров (материальных точек) три координаты  $x_1, y_1, z_1$  и  $x_2, y_2, z_2$ , используя которые можно с помощью методов аналитической геометрии однозначно определить как расстояние между шарами, так и углы, образованные прямой, соединяющей шары, с осями координат. Такой способ, однако, не обладает наглядностью, и поэтому более естественно поступить по-другому. Например, если шары одинаковы, то в качестве точки, определяющей местоположение системы в пространстве, уместно взять точку  $A$ , лежащую по середине отрезка, соединяющего центры шаров. Такая точка характеризуется в пространстве тремя координатами. Затем можно ввести углы, которые образует прямая, проходящая через центры шаров, с осями координат (достаточно задать всего два угла). И наконец, взаимное расположение шаров опишем, указав расстояние  $r$  между их центрами. Тогда получается следующая система координат для описания модели из двух шаров, соединенных пружиной: три пространственных координаты точки  $A$ , принятой за центр системы, два угла  $\varphi_1$  и  $\varphi_2$  и расстояние  $r$  между центрами шаров.

Во всех рассмотренных выше примерах бросается в глаза одна закономерность. Она заключается в том, что число введенных координат независимо от того, какая координатная система была использована, является одним и тем же. Если имеются плоскость и материальная точка на ней, то независимо от того, какая берется система координат (прямоугольная, косоугольная или полярная), в любом случае для точного описания места положения точки необходимо вводить две координаты. Если материальная точка находится в пространстве, то опять-таки для ее описания во всех случаях достаточно трех координат, т. е. она обладает тремя степенями свободы. Система из двух точек, если они могут изменять расположение друг относительно друга, должна описываться шестью координатами, если, конечно, необходимо полностью охарактеризовать ее состояние. При этом неважно, применяется ли трехмерная декартова система координат или любая другая, например косоугольная или система координат, в которой задаются положение некоторой центральной точки, углы поворота прямой, соединяющей пары шаров, по отношению к осям и расстояние между шарами. Таким образом, можно прийти к важному выводу: *число степеней свободы является инвариантом и не зависит от выбора системы координат.*

Общее правило заключается в следующем. Если какое-либо тело представить состоящим из  $N$  частей (или  $N$  материальных точек, причем все эти материальные точки могут двигаться относительно друг друга), то для полного описания его состояния надо ввести либо  $3N$  декартовых координат (по три координаты на каждую материальную точку, составляющую тело), либо три координаты центра масс (см. § 3.1), три координаты, которые характеризуют поворот всей системы как целого по отношению

к некоторой системе координат, и, наконец,  $3N - 6$  координат, которые характеризуют относительное движение частей тела или его деформации. Число координат деформации системы тел может быть и меньше  $3N - 6$ . Это возможно тогда, когда относительное положение некоторых частей тела считается неизменным. Например, если деформации вообще нет (твердое тело), то для полного описания пространственного состояния объекта достаточно шести координат (положение и поворот по отношению к внешним осям).

В физике существует очень большое число самых разнообразных координат. Они получили название *обобщенных*. Выбор наиболее рациональной системы координат диктуется прежде всего соображениями удобства и простоты описания движения тела или системы тел и простотой получающихся при этом уравнений движения. Значение выбора наиболее выгодной системы координат очень хорошо может быть проиллюстрировано на примере описания движения планет вокруг Солнца. Мы знаем, что планеты движутся вокруг Солнца почти по окружностям и что при этом Солнце находится в центре окружностей. Такая простая картина движения получается только в том случае, если солнечная система рассматривается как бы со стороны, из мирового пространства. Можно сказать, что мы выбираем систему координат, начало которой расположено где-то в мировом пространстве. Если же рассматривать движение планет с точки зрения наблюдателя, находящегося на Земле, то тогда получается отнюдь не такое простое изображение. Планеты будут двигаться по весьма сложным кривым, не имеющим ничего общего с окружностями. Заслуга Н. Коперника, предложившего планетарную систему, как раз и заключается в очень нелегком процессе изменения точки зрения наблюдателя и мысленного перенесения его в мировое пространство.

Еще один наглядный пример введения рациональной системы обобщенных координат у всех перед глазами. Речь идет об описании возможных движений пальцев кисти руки. Легко видеть, что наибольшая простота достигается, если в качестве координат принять углы между направлениями соседних фаланг пальцев, углы между прилегающими к ладони фалангами пальцев и плоскостью ладони, углы между соседними пальцами и т. д. Бросается в глаза большое число степеней свободы (около 20), которые связаны с доступными движениями лишь одной кисти руки. Понятно, почему так трудно создать искусственную кисть руки и снабдить ею робот.

## 1.2. Траектория и мировая линия

Выше мы рассмотрели характеристики положения точки, тела или системы тел в пространстве. Чтобы охарактеризовать изменение этого положения в пространстве, или движение тела или группы тел, наряду с указанием соответствующих пространствен-

ных координат необходимо ввести еще одну дополнительную характеристику—время. Поэтому для описания движения даже простейшего тела—материальной точки—нужно воспользоваться четырьмя параметрами. Под четвертым параметром, или четвертым измерением, понимается время. При этом пространственные координаты (три в случае простого тела и более в случае сложного тела с изменяющимся расположением отдельных частей) могут иметь как положительный, так и отрицательный знаки, а четвертый параметр (время) может быть только положительным. Такую четырехмерную (если ограничиться случаем материальной точки) систему описания принято называть *пространственно-временной* системой отсчета. Время всегда изменяется только в одном направлении—от прошлого к будущему.

Рассмотрим снова материальную точку, находящуюся в данном месте пространства в некоторый момент времени  $t$ . Для наглядности сохраним только две пространственные переменные. Это позволит изобразить интересующую нас ситуацию на простом чертеже (рис. 1.8). Предположим, что точка перемещается в пространстве, т. е. переходит последовательно из положения 1 в положения 2, 3 и т. д. Такой переход изображается довольно сложной кривой. Эта кривая, называемая *мировой линией*, в пространственно-временной системе отсчета включает как пространственные координаты, так и время. В дальнейшем мы увидим, что по дополнительной оси надо откладывать не собственно время, а величину  $ict$ , где  $i$ —мнимая единица;  $c$ —скорость света и  $t$ —время.

Укажем некоторые важнейшие свойства мировой линии.

1. *Мировая линия всегда принимается непрерывной.* Это утверждение отражает тот экспериментальный факт, что тело, перемещаясь из одной точки в другую или сохраняя свое состояние покоя [последнее выражается мировой линией, перпендикулярной плоскости  $XU$  и параллельной оси времени (рис. 1.8)], всегда находится в каком-то состоянии и не исчезает неизвестно куда и не появляется вновь. Если тело переходит из начальной точки мировой линии в конечную, то оно проходит и все промежуточные точки.

2. *Мировая линия не может иметь резких изломов* (где математически теряет свой смысл понятие производной). Это связано с тем, что тело не может приобрести скорость от нуля сразу до какой-то заданной величины скачком, т. е. невозможно бесконечно большое ускорение.

3. *Мировая линия не может самопересекаться.* Это свойство мировой линии обусловлено тем, что все события развиваются лишь от прошлого к будущему. Как тут не вспомнить древний афоризм: «Нельзя дважды войти в одну и ту же реку!»

Наряду с представлением движения в пространстве четырех переменных (простейшего тела—материальной точки) распространено и представление о пространстве трех переменных  $x, y, z$ ,

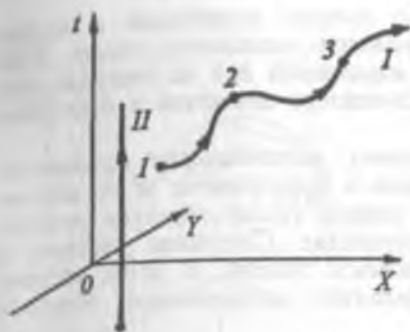


Рис. 1.8. Мировая линия двумерного пространства: I—переменное движение, II—состояние покоя

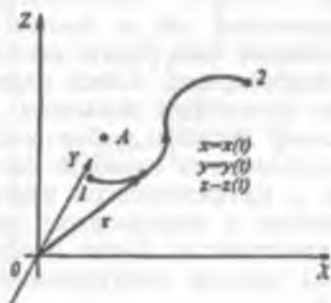


Рис. 1.9. Состояния материальной точки в пространственной системе координат. Неизменяющаяся точка A—состояние покоя; кривая (траектория) 1-2—состояние движения (материальная точка проходит всю кривую 1-2 в соответствии с течением времени);  $r = r(x, y, z, t)$ —радиус-вектор материальной точки

т. е. пространственных координат. В этом случае положение материальной точки математически изображается также некоторой точкой. В состоянии покоя точка неподвижна. Состояние движения описывается перемещающейся точкой из начального положения 1 в конечное положение 2 (рис. 1.9). Напомним, что при использовании пространственно-временной системы отсчета состояние покоя изображается точкой, движущейся в соответствии с течением времени вдоль мировой линии, расположенной параллельно оси времени.

Если тело переходит из начального положения в конечное, то оно последовательно проходит через все точки кривой, соединяющей начальное и конечное положения. Такая кривая определяет траекторию движения материальной точки. Эта траектория движения, так же как и мировая линия, непрерывна и на любом участке между двумя остановками имеет непрерывные первую и вторую производные по времени. Когда точка перемещается из состояния 1 в состояние 2, то мы полагаем, что координаты  $x$ ,  $y$  и  $z$  параметрически зависят от времени. Такое представление является обычным для ньютоновской, или классической, механики, в которой время принимается абсолютным и не меняется при переходе от одной системы координат к другой. В теории относительности пространственные и временные переменные взаимосвязаны и преобразование координат одновременно затрагивает их все.

В том случае, когда используется обычная трехмерная система декартовых координат, удобно применять для характеристики положения некоторой точки в пространстве понятие радиуса-вектора (см. рис. 1.9). Радиусом-вектором называется

вектор, который проводится из начала координат в точку пространства, где в данный момент находится тело. При перемещении тела радиус-вектор изменяется как по модулю, так и по направлению. Конец радиуса-вектора движется в пространстве по траектории движения.

Полная система любых координат, описывающих произвольные перемещения сложной системы в пространстве и ее деформации, с математической точки зрения также образует некоторый вектор в многомерном пространстве. Состояние системы в этом пространстве будет изображаться точкой. В этом состоит одна из причин популярности понятия материальной точки в физике.

Радиус-вектор в простейшем случае трехмерной евклидовой системы координат является частным случаем многомерного вектора. Как обычный вектор, так и многомерный при движении материальной точки или при изменении состояния системы соответствующим образом меняется.

Выше мы рассмотрели координатный способ описания состояния тел. В физике, как и в реальной жизни, бессмысленно задавать вопрос «где», если не указано «по отношению к чему». Поэтому как место начала координат, так и направления осей связываются с какими-нибудь телами или точками, которые однозначно на основании знания положения этих тел могут быть определены.

Такая совокупность тел, на которую «опирается» система координат, называется *системой отсчета*.

Введем теперь следующую важную кинематическую характеристику—скорость движения.

### 1.3. Скорость движения

Разобьем всю траекторию движения на ряд точек, соответствующих разным моментам времени (рис. 1.10). Соединим точки на траектории, соответствующие разным моментам времени, прямыми. Укажем направление движения точки вдоль траектории и каждого отрезка прямой. Каждый участок получающейся ломаной линии представляет вектор. Назовем его *вектором перемещения* точки на участке траектории за интервал времени  $[t_i, t_{i+1}]$ . Если воспользоваться понятием радиуса-вектора  $\mathbf{r} = \mathbf{r}(t)$ , то вектор перемещения равен  $\Delta \mathbf{r}_i = \mathbf{r}(t_{i+1}) - \mathbf{r}(t_i)$ . Если промежутки времени уменьшать, то ломаная линия будет все более и более точно соответствовать истинной траектории. Модули векторов, изображающих перемещение точки в данный отрезок времени, начнут уменьшаться. Направления этих векторов, которые совпадают с хордами, проведенными между двумя соседними точками на кривой (траектории), при уменьшении расстояния между точками стремятся к направлениям касательных. Поэтому в пределе при уменьшении интервала времени до

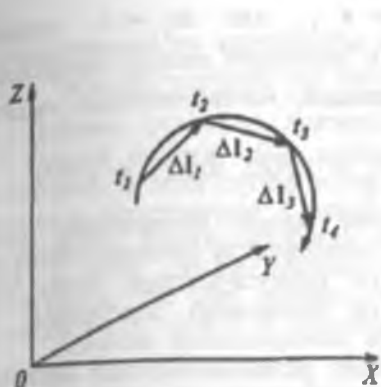


Рис. 1.10. Введение вектора перемещения

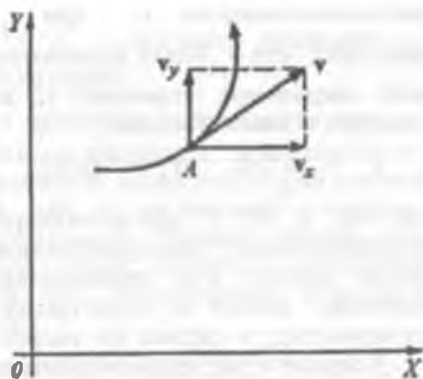


Рис. 1.11. Представление вектора скорости его компонентами (пример двумерного движения)

нуля направление вектора перемещения, который мы будем впредь обозначать символом  $\Delta l_i = \Delta r_i$ , совпадает с касательной к траектории движения в точке  $i$ .

Представим, что некоторая материальная точка за интервал времени  $\Delta t = t_2 - t_1$  изменила свое положение в пространстве так, что ее исходные координаты  $x_1, y_1, z_1$  (или радиус-вектор  $r_1$ ) приняли новые значения  $x_2, y_2, z_2$  (радиус-вектор  $r_2$ ). Тогда соответствующее перемещение может быть описано скоростью, в качестве которой выбирается вектор с компонентами

$$v_x = \frac{x_2 - x_1}{t_2 - t_1} = \frac{\Delta x}{\Delta t}, \quad v_y = \frac{y_2 - y_1}{t_2 - t_1} = \frac{\Delta y}{\Delta t}, \quad v_z = \frac{z_2 - z_1}{t_2 - t_1} = \frac{\Delta z}{\Delta t},$$

или в векторной записи

$$\mathbf{v} = \frac{\mathbf{r}_2 - \mathbf{r}_1}{t_2 - t_1} = \frac{\Delta \mathbf{r}}{\Delta t} = \frac{\Delta l}{\Delta t},$$

где  $\Delta \mathbf{r} = \Delta l$ .

Принимая как постулат классической физики, что координаты точки и время могут измеряться одновременно и независимо, и определяя положение точки в разные моменты времени внутри первоначального интервала  $\Delta t$ , получаем некоторую последовательность:

$$\mathbf{v}_i = \frac{\mathbf{r}_i - \mathbf{r}_1}{t_i - t_1} = \frac{\Delta \mathbf{r}_i}{\Delta t_i},$$

где  $t_1 \leq t_i \leq t_2$ . Направления векторов  $\mathbf{v}_i$  и их модули различны и зависят от интервалов  $\Delta t_i$ . Однако, исследуя поведение



последовательности  $v_i$  при  $t_i \rightarrow t_{i-1}$ , или  $\Delta t_i \rightarrow 0$ , можно заметить, что  $v_i$  будет стремиться к некоторой вполне определенной предельной величине  $v$ , имеющей математический смысл вектора с компонентами

$$v_x = dx/dt, \quad v_y = dy/dt, \quad v_z = dz/dt, \text{ или } v = dr/dt,$$

где  $dx$  и  $dt$  и другие—бесконечно малые приращения, или дифференциалы, пространственных координат и времени. Именно такой вектор, или производная от радиуса-вектора точки по времени, взятая в окрестности ее начального положения, и принимается в физике за *скорость материальной точки*.

Теоретически и экспериментально установлено, что модуль скорости любого тела при любых условиях не может быть больше скорости света в вакууме, т. е. величины  $c \approx 3 \cdot 10^8$  м/с, так что  $|v|$ —всегда конечная величина.

Так как вектор перемещения  $\Delta l$  точки при уменьшении отрезка времени стремится к  $dl$  и по своему направлению начинает совпадать с касательной к траектории, то и соответствующий вектор скорости, равный отношению вектора перемещения к промежутку времени, также будет стремиться к этой касательной. Поэтому *скорость движения точки изображается в любой момент времени вектором, касательным к траектории движения. Если траектория движения представляет прямую линию, то движение называется прямолинейным. Скорость этого движения все время имеет одно и то же направление в пространстве. Если движение происходит по кривой линии, то направление скорости в разных точках траектории различно. В этом случае говорят о криволинейном движении с переменной скоростью.*

Скорость движения, будучи вектором, представляется в виде суммы трех векторов, имеющих смысл компонент скорости по осям координат (рис. 1.11). Поэтому можно описать движение точки в пространстве, задавая скорость в каждой точке  $i$  траектории с помощью вектора  $v(i)$  или указывая каждый раз компоненты вектора скорости по осям координат, т. е. задавая в каждый момент значение трех чисел  $v_x(i)$ ,  $v_y(i)$ ,  $v_z(i)$ . Оба эти способа изображения в общем совершенно эквивалентны и отличаются один от другого только удобством и краткостью записи. Заметим, что запись скорости с помощью векторной величины  $v$  не содержит в себе указания на систему координат и в этом смысле не зависит от выбора системы координат.

Как уже отмечалось, при движении тела по криволинейной траектории скорость движения все время меняется. Построим график, показывающий изменение скорости. Этот график может быть построен аналогично графику, изображающему траекторию движения, с той разницей, что по осям координат будут

откладываться не координаты  $x, y, z$ , а значения проекций  $v_x, v_y, v_z$  вектора скорости на оси координат.

Принято различать некоторые простейшие виды движения.

Самый простой вид движения—равномерное прямолинейное движение. В этом случае точка движется вдоль некоторой прямой линии в пространстве с одной и той же постоянной скоростью. Траектория такого движения изображается прямой линией. Так как в этом случае скорость все время является постоянной величиной, то это значит, что она изображается некоторой неподвижной точкой в координатах  $v_x, v_y, v_z$ .

При равноускоренном прямолинейном движении точка движется вдоль прямой, но уже со скоростью, которая все время меняет свое значение пропорционально времени. В этом случае траектория движения по-прежнему выражается прямой линией, так же как и в случае равномерного прямолинейного движения, но график скорости в координатах  $v_x, v_y, v_z$  уже не представляется точкой, а имеет вид прямой линии. Это означает, что скорости по осям координат все время возрастают.

При равноускоренном движении точки по окружности ее скорость изменяется как по модулю, так и по направлению.

В общем случае любое движение точки может быть представлено в виде двух простейших: прямолинейного движения, происходящего вдоль какой-то прямой, и движения по окружности. Поэтому при изучении движения целесообразно изучать основные закономерности, которые отвечают этим двум простейшим видам движения.

В самом деле, радиус-вектор точки для любого момента времени  $t$  может быть представлен как  $\mathbf{r} = |\mathbf{r}|\mathbf{e}$ , где  $|\mathbf{r}|$ —модуль радиуса-вектора,  $\mathbf{e}$ —единичный направляющий вектор ( $|\mathbf{e}| = 1, \mathbf{e} \parallel \mathbf{r}$ ). Тогда малое приращение вектора  $\mathbf{r}$  равно  $\Delta \mathbf{r} = \Delta |\mathbf{r}|\mathbf{e} + |\mathbf{r}|\Delta \mathbf{e}$ . Видно, что  $\Delta \mathbf{r}$  представляется суммой двух векторов, один из которых направлен вдоль по вектору  $\mathbf{e}$ , и, стало быть, совпадает с исходным направлением радиуса-вектора  $\mathbf{r}$  (прямолинейное перемещение), а другой—перпендикулярно исходному направлению радиуса-вектора  $\mathbf{r}$  [по определению, единичный вектор  $\mathbf{e}$  может менять только свое направление, но не модуль], поэтому при малых перемещениях  $\Delta \mathbf{e} \perp \mathbf{e}$  (или  $\mathbf{r}$ ) (движение по окружности). Таким образом, общее перемещение получается как комбинация двух указанных движений.

Движение точки по окружности—это движение вокруг некоторой оси. Для простоты будем считать, что ось (вращения) всегда направлена по оси  $Z$ , а само движение происходит в плоскости  $XY$ . Траектория такого движения представляет окружность, проведенную с некоторым радиусом  $r$  вокруг начала координат (рис. 1.12).

Если движение происходит с одной и той же скоростью, т. е. с одним и тем же числом оборотов в единицу времени, то вектор

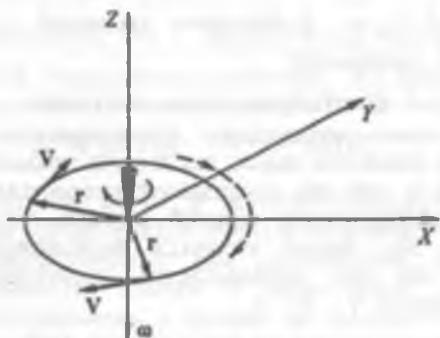


Рис. 1.12. Введение вектора  $\omega$

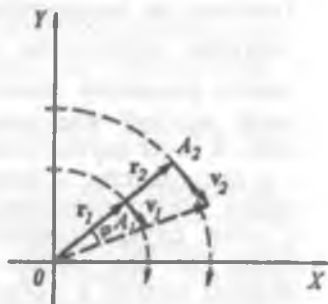


Рис. 1.13. К выводу соотношения  $|v|$  и  $|r|$

скорости вращения все время изменяет свое направление в пространстве, но сохраняет свой модуль.

Движение по окружности является как раз таким видом движения, которое удобно описывать, используя полярную систему координат. Тогда положение точки на окружности будет характеризоваться углом, отсчитываемым от некоторого начального радиуса-вектора.

Угловая скорость также представляется вектором. Этот вектор совпадает с осью вращения и направлен в ту сторону, в какую движется совмещенный с осью правосторонний винт, если его вращать в направлении движения точки. Если смотреть вдоль оси вращения, то при движении точки по часовой стрелке винт будет двигаться от наблюдателя, и направление угловой скорости в этом случае следует принять таким, как показано на рис. 1.12.

Если точка движется против часовой стрелки, то винт выкручивается из плоскости  $XY$ . Он будет в этом случае двигаться вверх, и направление угловой скорости следует принять вверх к наблюдателю (вверх по оси  $Z$ ).

Может возникнуть вопрос: а что изменится, если правосторонний винт заменить левосторонним? В механике ничего не изменится, так как все законы механики инвариантны относительно выбора типа винтов. Однако в более сложных явлениях, чем те, которые описываются механикой, дело обстоит иначе. Так, например, важнейшие биологически активные макромолекулы имеют не просто спиральную структуру, но почему-то (причина этого сейчас не ясна) чаще всего отвечающую именно правостороннему винту, т. е. правое и левое в природе оказывается неравноправным. Однако еще до того, как это стало известным, в физике и математике пользовались правосторонними винтами.

Модуль вектора выбирается пропорциональным числу оборотов вокруг оси или углу поворота в единицу времени.

Воспользуемся теперь векторным произведением. Тогда можно заключить, что вектор скорости  $v$ , направленный перпендикулярно вектору угловой скорости  $\omega$  и перпендикулярно радиусу-вектору, проведенному от оси вращения в данную точку траектории (см. рис. 1.12), по направлению совпадает с векторным произведением вектора угловой скорости и вектора  $r$ . Если мы рассмотрим несколько точек с одним и тем же числом оборотов в единицу времени, но расположенных на разном расстоянии от оси вращения (рис. 1.13), то за один и тот же промежуток времени точка, отстоящая от оси вращения дальше, будет проходить путь, больший как раз на величину отношения радиуса первой траектории к радиусу второй траектории. Следовательно, по модулю вектор скорости должен быть пропорционален числу оборотов в единицу времени и радиусу окружности. Окончательно заключаем, что вектор  $v$  можно записать как  $v = \omega \times r$ .

Напомним, что, по определению, модуль

$$|v| = |\omega| |r| \sin 90^\circ = |\omega| |r|.$$

С другой стороны,  $|\omega| = \frac{d\varphi}{dt}$ , где  $\varphi$ —угол поворота радиуса-вектора точки при ее движении вокруг оси по отношению к некоторому начальному положению.

Введение угловой скорости в форме вектора так, как описано выше, имеет то большое преимущество, что в этом случае мы одновременно характеризуем ось вращения, направление движения и скорость этого движения. Во все время движения угловая скорость, если это вращение равномерное, остается постоянной величиной, в то время как скорость движения  $v$  точки вдоль траектории движения (ее называют *линейной скоростью*) все время меняет свое направление.

Рассмотрим сначала простейший случай, когда точка движется вокруг оси с одной и той же угловой скоростью. При этом вектор скорости  $v$  постоянен по модулю, но все время меняет направление в пространстве. Этот вектор поворачивается, описывая, в свою очередь, окружность. График скорости движения в этом случае также представляет окружность, вычерчиваемую концом вектора, начало которого расположено в начале координат. Чем больше скорость вращения, тем больше радиус окружности, соответствующий графику скорости. Этот радиус пропорционален модулю вектора скорости (рис. 1.14).

Если совершается движение по окружности со все возрастающей угловой скоростью, то в этом случае вектор скорости  $v(t)$  все время растет по модулю. При каждом обороте скорость возрастает, значит, длина вектора  $v(t)$  тоже все время возрастает. По-прежнему график скорости может быть вычерчен как след конца вектора скорости, начало которого расположено в начале

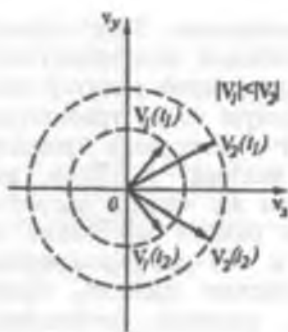


Рис. 1.14. График движения конца вектора  $\mathbf{v}$  в зависимости от времени  $t$

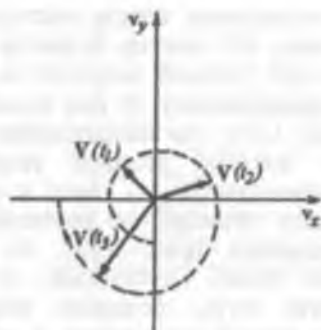


Рис. 1.15. График скорости неравномерного движения по окружности

координат. При вращении этого вектора получается спиральная кривая (рис. 1.15).

## 1.4. Ускорение движения

Ускорением называется производная скорости движения по времени. Понятие ускорения может быть введено совершенно так же, как понятие скорости движения, если разбить график скорости на отдельные участки, соответствующие интервалам времени, и заменить сплошную кривую, изображающую график скорости, ломаной линией. Каждый участок такой ломаной линии представляет величину, определяющую изменение вектора скорости  $\Delta \mathbf{v}_i$  на участке от момента времени  $t_i$  до момента времени  $t_{i+1}$ .

При переходе к пределу при  $\Delta t_i = t_{i+1} - t_i \rightarrow 0$  участки эти уменьшаются, в результате соответствующий вектор  $\Delta \mathbf{v}$  аналогично вектору перемещения будет стремиться к касательной к графику скорости. Таким образом, ускорение по отношению к скорости и графику скорости играет ту же самую роль, что и скорость по отношению к траектории перемещения. Можно поэтому сказать, что ускорение есть вектор, направленный по касательной к графику скорости в сторону, которая соответствует изменению скорости движения. По модулю вектор ускорения равен отношению модуля вектора изменения скорости  $\Delta \mathbf{v}$  к интервалу времени  $\Delta t$ , за который это изменение скорости произошло.

Для малых промежутков времени  $dt$  ускорение

$$\mathbf{a} = \frac{d\mathbf{v}}{dt} = \frac{d}{dt} \left( \frac{d\mathbf{r}}{dt} \right) = \frac{d^2 \mathbf{r}}{dt^2} = \frac{d^2 \mathbf{l}}{dt^2}$$

и математически представляет производную вектора  $\mathbf{v}(t)$  по времени или вторую производную радиуса-вектора точки по времени.

Выше мы отметили, что скорость (по модулю) перемещения любого объекта по отношению к другому объекту не может быть больше скорости света в вакууме. Для ускорения какие-либо большие ограничения неизвестны, однако поскольку любые ускорения появляются лишь при воздействии внешних сил, ясно, что бесконечно большие ускорения могут появиться лишь при наличии бесконечно больших взаимодействий. Все же известные нам в природе взаимодействия имеют конечные величины (при реальных возможных сближениях объектов).

Для того чтобы изучить основные закономерности ускорения, вернемся снова к двум простейшим движениям точки. При равномерном прямолинейном движении скорость точки остается все время постоянной. В этом случае, очевидно, ускорение равно нулю. При прямолинейном неравномерном движении скорость точки направлена вдоль прямой, соответствующей траектории движения. Ускорение также должно быть направлено вдоль прямой, соответствующей траектории движения, но векторы ускорения и скорости могут иметь разные направления. Если скорость возрастает, то ускорение направлено в том же направлении, что и вектор скорости. Если скорость убывает, то ускорение направлено в сторону, противоположную вектору скорости.

При криволинейном движении вектор скорости как вектор, который во всех точках должен быть касателен к траектории движения, меняет не только свой модуль, но и направление. В соответствии с этим и ускорение как величина, характеризующая изменение скорости, будет зависеть от двух факторов: от изменения направления вектора скорости движения и от изменения модуля этого вектора.

В рассмотренном выше случае прямолинейного движения изменяется только модуль вектора скорости движения, но не его направление: ускорение целиком связано только с этим фактором.

При движении точки вокруг некоторой оси с одной и той же угловой скоростью модуль вектора скорости  $v$  не меняется, а меняется только направление. В этом случае должно также наблюдаться ускорение, но здесь оно связано только с изменением направления движения, т. е. с изменением направления вектора скорости.

Посмотрим, каковы модуль и направление ускорения при движении точки по окружности.

Как уже указывалось, при равномерном движении точки по окружности график скорости представляет окружность. По определению, ускорение—это вектор, направленный во всех точках по касательной к графику скорости. Значит, ускорение при таком движении должно быть направлено так, как показано на рис. 1.16,а, т. е. перпендикулярно вектору скорости. Перпендикуляр к вектору скорости направлен, очевидно, по радиусу, соединяющему центр окружности и данную точку траектории.

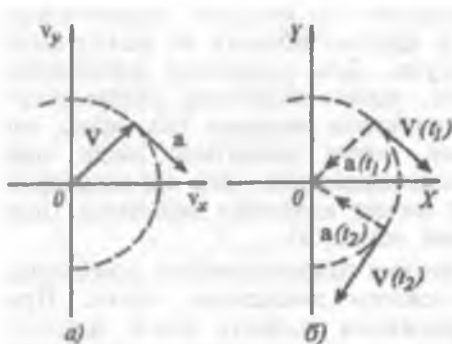


Рис. 1.16. Скорость и ускорение на графике скорости (а); скорость, ускорение и траектория равномерного движения по окружности (б)

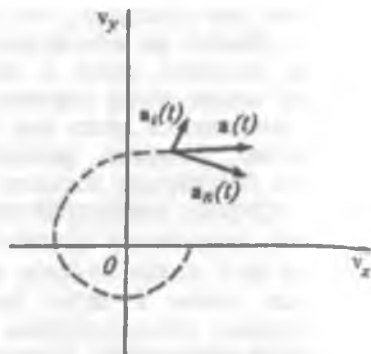


Рис. 1.17. Соотношение векторов ускорений при ускоренном движении по окружности

Направление вектора ускорения легко установить на основании простого рассуждения. Если происходит перемещение точки по окружности из точки 1 в точку 2, то вектор скорости меняет свое направление так, как показано на рис. 1.16, б. Его конец отклоняется в сторону, соответствующую направлению от траектории движения к центру. Следовательно, ускорение также направлено от траектории движения к центру. Такое ускорение называется *нормальным* и обозначается символом  $a_n$ .

Если снова вернуться к графику скорости, то увидим, что вектор ускорения, касательный к графику скорости, по аналогии с вектором скорости, касательным к траектории движения, может быть представлен в форме векторного произведения вектора угловой скорости  $\omega$  и вектора скорости  $v$ :

$$a_n = \omega \times v.$$

Тогда

$$|a_n| = |\omega| |v| \sin 90^\circ = |\omega| |v| = |\omega| |\omega| |r| = r |\omega|^2.$$

Предположим, что движение происходит вокруг оси с постоянно возрастающей скоростью, т. е. число оборотов точки вокруг оси в единицу времени постоянно возрастает. В этом случае график скорости представляет, как было указано выше, спираль, а ускорение—вектор, касательный к этой спирали. Из рис. 1.17 видно, что в этом случае вектор ускорения уже не перпендикулярен направлению скорости. Ускорение тогда можно представить в виде векторной суммы двух векторов. Один из них направлен перпендикулярно скорости и соответствует нормальной компоненте ускорения, а второй направлен вдоль вектора скорости движения. Вторая компонента называется *тангенциальным ускорением*  $a_t$ . Нормальное ускорение по-прежнему связано только с модулем вектора угловой скорости в данный момент времени.

Тангенциальное ускорение появится только в тот момент, когда эта угловая скорость будет меняться. Чем больше меняется скорость вращения, тем больше расстояние между витками спирали и тем значительнее тангенциальное ускорение. Тангенциальное ускорение, следовательно, должно быть тогда пропорционально, во-первых, радиусу окружности, по которой происходит движение точки, а во-вторых, угловому ускорению  $\alpha$ , т. е. величине, показывающей, насколько быстро изменяется скорость движения вокруг оси в единицу времени. Стало быть,  $|a_t| = r|\alpha|$ , где  $r$ —радиус окружности, по которой происходит движение.

Угловое ускорение  $\alpha$  также является вектором. В простейшем случае неизменной оси вращения  $\alpha \parallel \omega$  и направлено в ту же сторону, что и  $\omega$ , если угловая скорость увеличивается, и в противоположную, если угловая скорость уменьшается.

Пользуясь векторным произведением и учитывая направления векторов, запишем

$$a_t = \alpha \times r.$$

Вектор полного ускорения складывается из двух взаимно перпендикулярных векторов—тангенциального и нормального ускорений. Его модуль по теореме Пифагора равен

$$|a| = \sqrt{|a_n|^2 + |a_t|^2} = r\sqrt{|\omega|^4 + |\alpha|^2} = r\sqrt{\left(\frac{d\varphi}{dt}\right)^4 + \left(\frac{d^2\varphi}{dt^2}\right)^2}.$$

Если изобразить полное ускорение, траекторию на графике, то с учетом, что нормальное ускорение должно быть перпендикулярно скорости, а тангенциальное направлено вдоль вектора скорости, получаем картину, показанную на рис. 1.18.

При увеличении угловой скорости вектор полного ускорения тем больше отклоняется от радиуса в сторону касательной к траектории в направлении движения, чем больше ускорение вращения. Если, наоборот, происходит замедление угловой скорости, то вектор полного ускорения будет отклоняться от радиуса в сторону, противоположную движению, и тоже тем больше, чем больше замедление вращения в единицу времени. Эти два случая показаны на рис. 1.18.

Выражение для полного ускорения  $a$  при движении по окружности легко получить с помощью простых математических выкладок.



Рис. 1.18. Полное ускорение при ускоренном и замедленном движениях по окружности



Имеем  $\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$ . Отсюда по правилам дифференцирования получаем

$$\mathbf{a} = \frac{d\mathbf{v}}{dt} = \frac{d\boldsymbol{\omega}}{dt} \times \mathbf{r} + \boldsymbol{\omega} \times \frac{d\mathbf{r}}{dt} = \boldsymbol{\alpha} \times \mathbf{r} + \boldsymbol{\omega} \times \mathbf{v}.$$

Если ось вращения не меняется, то вектор углового ускорения совпадает с осью вращения. Тогда вектор  $\boldsymbol{\alpha} \times \mathbf{r}$  направлен вдоль по касательной к окружности. Это и есть вектор  $\mathbf{a}_t$ . Вектор  $\boldsymbol{\omega} \times \mathbf{v}$  направлен вдоль вектора  $\mathbf{r}$  в сторону оси вращения, т. е.  $\boldsymbol{\omega} \times \mathbf{v} = \mathbf{a}_n$ . При этом

$$|\mathbf{a}_t| = |\boldsymbol{\alpha}| |\mathbf{r}|, \quad |\mathbf{a}_n| = |\boldsymbol{\omega}| |\mathbf{v}| = |\boldsymbol{\omega}|^2 |\mathbf{r}|, \quad \mathbf{a} = \mathbf{a}_t + \mathbf{a}_n.$$

Отсюда и получается приведенное выше выражение для  $|\mathbf{a}|$ .

В заключение заметим, что для сложной системы с большим числом степеней свободы также можно ввести понятие траектории как следа движения изображающей систему точки в многомерном пространстве с координатами  $q_k$ . Скорость движения такой точки будет представляться вектором с компонентами  $dq_k/dt$ , имеющими смысл первых производных по времени от обобщенных координат.

Можно показать, что такой вектор также будет иметь смысл касательной к линии, отвечающей траектории в многомерном пространстве. Вектор с компонентами  $d^2q_k/dt^2$ , являющимися вторыми производными от обобщенных координат по времени, будет играть роль обобщенного ускорения.

Видно, что аналогия с описанием движения материальной точки в обычном пространстве весьма глубока.

## 1.5. Релятивистская кинематика

Интервалы времени и длины в разных системах отсчета. В предыдущих параграфах были введены основные кинематические характеристики, необходимые для описания состояния объектов и их движения,— скорость и ускорение тел, обсуждены вопросы применения различных координат. Было также отмечено, что максимально возможной скоростью перемещения любых тел, а также вообще распространения любых сигналов является скорость света. В данном параграфе мы остановимся на некоторых следствиях, вытекающих из существования максимально возможной скорости движения, равной скорости света, и вопросах взаимосвязи пространственных и временной координат. Ограничимся простейшим объектом—материальной точкой. Сразу же отметим, что при описании любого движения необходимо указать, относительно чего это движение совершается. Не будем пока останавливаться на том, существует или не существует точка отсчета, которую можно считать абсолютно покоящейся.

Попытаемся описать движение, как всегда привязывая его к каким-то конкретным окружающим объектам. Например, движение автомобиля будем рассматривать относительно поверхности земли или какого-то объекта (строения) на этой поверхности, которое и предполагается неподвижным; движение теплохода по реке—относительно берега. Ясно, что в таком описании движений содержится и некоторая условность. Она сразу же обнаруживается, если представить себе следующую картину. Пусть на двух параллельных железнодорожных путях стоят два поезда. Наблюдатель находится в одном из этих поездов и разглядывает в окно соседний поезд. Ничего другого, кроме соседнего поезда, соседнего вагона, он не видит. В такой ситуации, по-видимому, неоднократно бывали и читатели этой книги. Наверное, они неоднократно ловили себя на том, что совершали ошибку, приняв начало движения собственного поезда за движение соседнего. Человеку вообще свойственно считать себя центром мироздания, и если собственный поезд начал двигаться без толчка и с малым ускорением, то его пассажир по-прежнему будет считать себя неподвижным. Тем более трудно разобраться в ситуации, когда и соседний поезд одновременно пришел в движение относительно земли. Этот пример наглядно показывает, что, глядя в окно на соседний поезд, наблюдатель может судить только об относительном движении поездов, но не об их перемещениях по поверхности земли.

Именно то обстоятельство, что наблюдатель в целом ряде случаев сделает неверное заключение и будет считать свой поезд неподвижным, в то время как именно он пришел в движение по отношению к земле, и показывает, что понятие движения относительно.

Теперь представим, что наблюдатель, находящийся в первом поезде, через свое окно и окно второго поезда следит за перемещением каких-то объектов внутри второго поезда. Тогда второй поезд по отношению к нему будет движущимся, и некоторый предмет, перемещающийся внутри второго поезда, будет для первого наблюдателя участвовать как бы в двух движениях: с одной стороны, в движении всего второго поезда и, с другой, в движении по отношению к предметам, связанным со вторым поездом. Представим, что во втором поезде также находится наблюдатель, который рассматривает движение некоторого объекта внутри своего же собственного купе и, таким образом, регистрирует движение объекта лишь по отношению к нему. Наблюдателя, который находится во втором поезде и для которого предмет участвует только в одном движении, т. е. в движении относительно этого наблюдателя внутри вагона, мы будем называть подвижным. Разумеется, можно поместить первого наблюдателя во второй поезд и рассматривать движущийся внутри первого поезда предмет. Конечно, от такой перемены ничего не изменится. Введем две системы координат, одну из которых будем условно считать неподвижной, а вторую движу-

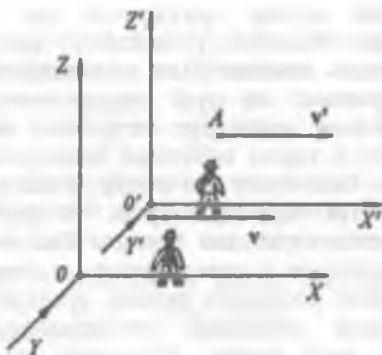


Рис. 1.19. Материальная точка  $A$  движется вдоль оси  $X$  со скоростью  $v'$  по отношению к наблюдателю, смещенному со штрихованной системой координат. Штрихованная система координат движется вдоль той же оси со скоростью  $v_0$  по отношению к нештрихованной системе координат (эту скорость регистрирует наблюдатель, «закрепленный» в нештрихованной системе координат и следящий за движением «штрихованного» наблюдателя)

щейся мимо первой. Пусть внутри второй системы координат имеется еще движущаяся материальная точка. Ограничимся простейшей ситуацией, т. е. будем считать, что движение одной системы координат по отношению к другой совершается только вдоль одной оси (вдоль оси  $X$ ), а также, что движение второй системы координат по отношению к первой системе координат равномерно и прямолинейно, т. е. совершается без ускорения. Обозначим систему координат, условно принятую за неподвижную,  $XYZ$ . Систему координат, которая принимается условно за движущуюся, обозначим  $X'Y'Z'$ . Будем также считать, что материальная точка, движение которой рассматривается по отношению как к первой, так и второй системам координат, перемещается также только вдоль оси  $X$  (рис. 1.19). Таким образом, вдоль осей  $Y$  и  $Z$  никакого движения не происходит. Тогда, как можно заключить из простых, наглядных соображений, скорость движения точки  $A$ , которую должен был бы регистрировать наблюдатель, находящийся в нештрихованной системе координат, должна была бы иметь вид  $v = v_0 + v'$ , где  $v'$ —скорость движения материальной точки по отношению к штрихованной, движущейся системе координат;  $v_0$ —скорость движения штрихованной системы координат по отношению к нештрихованной системе;  $v$ —та скорость, которая будет регистрироваться наблюдателем, находящимся в нештрихованной, условно принятой за неподвижную, системе координат. Такое сложение скоростей любой наблюдатель может легко проверить, если он представит себе скорость человека, идущего, например, по палубе парохода в направлении его движения, в то время как сам пароход движется мимо наблюдателя, находящегося на берегу. Так как, по нашему предположению, никакого движения вдоль осей координат  $Y$ ,  $Z$  не происходит, то из формул для скоростей движения, которые будут регистрироваться наблюдателями, находящимися либо в штрихованной, либо в нештрихованной системе координат (в штрихованной системе координат наблюдатель будет видеть, что материальная точка по отношению к нему движется только со скоростью  $v'$ ), можно

закл<sup>ю</sup>чить (учитывая, что  $v = v_0 + v'$  или  $\frac{dx}{dt} = v_0 + \frac{dx'}{dt}$ ), что между штрихованными и нештрихованными координатами должна быть следующая связь:  $dx = dx' + v_0 dt$  или  $x = v_0 t + x'$ . При  $t = 0$  точки 0 и 0' совмещены.

При этом предполагается, что, как это принято в классической механике Ньютона, переход от одной системы координат к другой никак не отражается на течении времени, т. е.  $t = t'$ .

Так как никакого движения вдоль этих осей  $Y$  и  $Z$  нет, то  $y = y'$ , а  $z = z'$ .

Полученные нами соотношения для координат с большой степенью точности могут быть проверены на многочисленных механических экспериментах, что и было сделано на разных этапах развития физики прошлого времени. Однако здесь следует сделать одну важнейшую оговорку. В самом деле, все подобные эксперименты не делались, да и не могли быть сделаны в прошлом при таких условиях, когда скорости движения как одной системы координат по отношению к другой, так и скорости перемещения материальной точки внутри штрихованной системы координат были бы сравнимы со скоростью света. Предполагая же, например, что скорость  $v_0 = (2/3)c$ , где  $c$ —скорость света, а скорость  $v' = (4/5)c$ , получаем, что наблюдатель, находящийся в нештрихованной системе координат, должен был бы увидеть материальную точку проносящейся мимо него со скоростью  $(22/15)c$ , т. е. большей, чем скорость света. Наш пример является чисто умозрительным, и исследователь прошлого ничего странного здесь бы не усмотрел. Более современный исследователь заметил бы противоречие с утверждением (базирующимся на конкретных новых экспериментах) о том, что скорость движения любого тела и вообще передачи любого сигнала не может быть больше, чем скорость света. Значит, записанные выше соотношения между координатами штрихованной и нештрихованной систем верны, но только в пределах тех экспериментов, где скорости не очень велики. Наша дальнейшая задача будет заключаться в том, чтобы выяснить, как надо преобразовать соотношения между пространственными координатами и временем, чтобы они оказались верными и в тех случаях, когда скорости движений очень велики и приближаются к скорости света.

Для этого примем в качестве исходной посылки утверждение, что скорость света является не только максимальной величиной для скоростей распространения любых сигналов и перемещения любых тел в пространстве, но и одинаковой как в системе нештрихованных координат, которая принята нами условно за неподвижную, так и в системе штрихованных координат, которая принята нами за движущуюся. Как уже указывалось, это утверждение и многочисленные его следствия проверены сейчас на большом количестве самых разнообразных экспериментов.

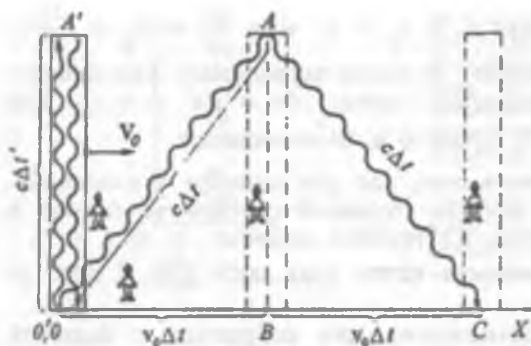


Рис. 1.20. Наблюдатель, находящийся на линейке (штрихованные координаты), регистрирует движение луча света вверх и вниз. Наблюдатель неподвижный (нештрихованные координаты) припишет лучу света движение по отрезкам  $OA$  и  $AC$ . В момент первой вспышки начала штрихованной и нештрихованной системы координат были совмещены

Представим следующий мысленный эксперимент: пусть мы имеем некоторую линейку, у нижнего конца которой располагается источник света, способный давать очень короткие импульсы. Такие источники действительно существуют и могут генерировать импульсы длительностью до  $10^{-12}$  с. Свет направляется вдоль линейки вверх и достигает ее конца, где расположено зеркало. Отражаясь от зеркала и возвращаясь вдоль линейки, свет достигает нижнего ее конца, где его приход регистрируется. Между вспышкой и ответным сигналом проходит некоторое время, которое принято называть *собственным временем* (и обозначать  $\tau$ ) в системе координат, связанной с этой линейкой.

Если вспышки света повторять через равные промежутки времени, избрав началом новой вспышки момент возвращения света в нижнюю часть линейки, то мы получим своеобразные часы, которые будут давать сигнал через интервалы времени  $\tau$ , равные времени прохождения света вверх и вниз по линейке.

Расположим линейку вдоль вертикальной оси. Предположим, что линейка движется вправо по отношению к некоторому наблюдателю, принятому за неподвижного, со скоростью  $v_0$ . Примем направление этого движения за ось  $X$  (рис. 1.20). Наблюдатель, находящийся в штрихованной системе координат и движущийся вместе с линейкой (для него линейка является неподвижной), заметит, что свет вдоль линейки проходит от момента вспышки до получения в том же месте отраженного сигнала путь, равный удвоенному расстоянию  $O'A'$ , т. е. путь  $O'A'$  и  $A'O'$ . Неподвижный же наблюдатель заметит, что линейка движется мимо него. Для него путь, проходимый светом, равен расстояниям  $OA$  и  $AC$ . Расстояние  $OC$  должно быть, очевидно, равным  $v_0 T$ , где  $T$ —период времени, который измерит неподвижный наблюдатель по своим часам, синхронизированным с такими интервалами, которые «отмеряет» свет, выходящий из начала линейки и снова достигающий после отражения этого же начала.

Наблюдатель, перемещающийся вместе с линейкой и измеривший по своим часам интервал, равный собственному времени,

зная скорость света  $c$ , должен заключить, что общий пройденный вдоль линейки путь равен  $ct$ . Наблюдатель, принятый за неподвижного, должен, в свою очередь, заключить, что полный пройденный путь от точки  $O$  до  $A$  и от  $A$  до  $C$  равен  $cT$ . Так как нет никаких оснований предполагать, что процессы движения импульса света вверх и вниз чем-нибудь отличаются (кроме направления движения), то можно принять интервалы времени  $\Delta t'$  и  $\Delta t$  до момента достижения светом зеркала на верхнем конце линейки равными  $\Delta t' = \tau/2$  и  $\Delta t = T/2$  (рис. 1.20). Так как  $O'A$ —диагональ прямоугольника  $O'(O)A'AB$ ,  $OA'$  и  $BA$ —стороны этого прямоугольника и, разумеется,  $O'A = BA < OA$ , то различаются либо скорости света в разных системах координат, либо отрезки времени  $\Delta t' = \tau/2$  и  $\Delta t = T/2$ .

Как указывалось, экспериментальным фактом является то, что скорость света как для неподвижного, так и для движущегося наблюдателя будет одной и той же. Стало быть,  $\tau$  и  $T$  неодинаковы, и сразу становится ясным, что наше утверждение о том, что  $t = t'$  (см. выше приведенные формулы для соотношения координат в движущейся и неподвижной системах), уже неверно. Точнее, мы вправе только утверждать, что при малых скоростях перемещения это будет правильным. По теореме Пифагора о прямоугольных треугольниках  $OA^2 = OB^2 + AB^2$ , откуда  $c^2 \Delta t^2 = v_0^2 \Delta t'^2 + c^2 \Delta t'^2$ , т. е.  $\Delta t' = \Delta t \sqrt{1 - v_0^2/c^2}$  или  $\tau\gamma = T$ . Символом  $\gamma$  в релятивистской механике принято обозначать величину,  $(1 - v_0^2/c^2)^{-1/2}$ .

В результате мы получим, что интервалы времени, которые измеряют движущийся и неподвижный наблюдатели отличаются друг от друга множителем  $\gamma$ , который при малых скоростях с очень хорошей степенью точности близок к единице и начинает заметно отличаться от единицы только при скорости перемещения одной системы координат по отношению к другой, сравнимой со скоростью света. Именно поэтому можно утверждать, что  $t = t'$  при сравнительно малых скоростях.

Мы получили весьма интересный результат: оказывается, что интервалы времени, которые будут наблюдаться для одной и той же последовательности между двумя событиями двумя наблюдателями, находящимися в подвижной и неподвижной системах координат, оказываются различными. Так как это есть свойство самого времени как понятия, но ни в коем случае не связано с каким-либо конкретным способом его измерения, то мы можем сказать, что в двух разных системах координат время будет течь как бы с разной скоростью. Так как  $\gamma > 1$ , то  $\tau < T$ . Значит, для любого наблюдателя, который рассматривает процессы таким образом, что для него объекты либо неподвижны, либо движутся с очень малой скоростью (а это значит, что он сам вместе с ними перемещается), время будет течь как бы быстрее. Измеря-

смый им интервал собственного времени будет наименьшим по сравнению с интервалами между теми же событиями, которые зарегистрируют наблюдатели, мимо которых объекты проносятся с большой скоростью. Этот эффект известен под названием *правила замедления хода часов в неподвижной системе координат*.

Можно ли проверить существование такого замедления времени с помощью каких-либо прямых наблюдений? В самом деле, когда говорят о замедлении времени, которое наблюдает неподвижный наблюдатель по отношению к подвижному наблюдателю, возникает такой вопрос: ведь мы же условно выбрали одного наблюдателя неподвижным, а другого движущимся? Помня их местами и тогда вся картина изменится на противоположную.

Нет ли здесь какого-то противоречия? Оказывается, что эффект замедления времени действительно обнаруживается, но при таких условиях, когда один наблюдатель по отношению к другому наблюдателю не будет уже полностью эквивалентным. Это достигается в тех случаях, когда один из наблюдателей хотя бы в какой-то момент времени движется не равномерно и прямолинейно по отношению к другому, а изменяет направление своего движения, т. е. получает некоторое ускорение. Обычно такая ситуация рассматривается на примере *парадокса близнецов*.

Парадокс этот заключается в следующем: допустим есть два близнеца; один из них остается на Земле, а второй отправляется в путешествие на космическом корабле (предполагаем, что полет совершается равномерно и прямолинейно). Затем корабль поворачивается (здесь уже возникает ускорение и появляется неравноправность одного близнеца по отношению к другому!) и начинает возвращаться на Землю. Через некоторое время близнецы встречаются. Если сравнить, насколько каждый из близнецов постарел, то мы обнаружим, что близнец на Земле постарел больше (его собственное время шло быстрее), несмотря даже на то, что с точки зрения космонавта часы на Земле шли медленнее.

Этот парадоксальный факт в 1971 г. в США был проверен с помощью эксперимента. Для этого были применены очень точные часы, в которых в качестве колеблющегося элемента использовались электроны атомов цезия. Такие часы как бы «тикают» около  $9 \cdot 10^9$  раз в секунду. Была изготовлена совершенно идентичная пара таких часов. Одни часы оставались на Земле, а другие помещались в самолет, облетавший вокруг Земли.

Самолет, летящий по круговой траектории вокруг Земли, движется с некоторым ускорением и, следовательно, находится не в такой же ситуации, в которой находятся часы, покоящиеся относительно Земли. В соответствии с законами релятивистской кинематики путешествующие в самолете часы должны были отстать от покоящихся на  $184 \pm 23$  нс. На самом деле наблюдаемое отставание составило  $203 \pm 10$  нс. Видно, что получа-

ющий результат согласуется с теоретическим в пределах погрешности измерения.

Было осуществлено еще большое число экспериментов, в которых проверялся эффект замедления времени, и все они подтвердили сделанный выше вывод. Таким образом, разное течение времени в системах координат, движущихся друг относительно друга равномерно и прямолинейно, которое будут регистрировать экспериментаторы, находящиеся соответственно в одной системе координат и в другой и рассматривающие явления как бы изнутри и со стороны, является непреложным экспериментальным фактом.

Обратимся теперь снова к опыту с движущейся линейкой. Введем обозначение  $\Delta l_{\perp}$  (и ниже  $\Delta l'_{\perp}$ ), чтобы различить случаи расположения линейки перпендикулярно и параллельно направлению движения. На основании свойств прямоугольного треугольника  $OAB$  можно записать  $c^2 \Delta t^2 = v_0^2 \Delta t'^2 + \Delta l_{\perp}^2$ , где  $\Delta l_{\perp}$  — длина линейки в вертикальном направлении, которую измерит неподвижный наблюдатель. Поэтому

$$\Delta t = \frac{\Delta l_{\perp}}{c \sqrt{1 - v_0^2/c^2}}.$$

Полное время прохождения света туда и обратно

$$2\Delta t = \frac{2\Delta l_{\perp}}{c \sqrt{1 - v_0^2/c^2}}.$$

Для подвижного наблюдателя, если линейка имеет длину  $\Delta l'_{\perp}$ , соответственно получим

$$2\Delta t' = \frac{2\Delta l'_{\perp}}{c}.$$

На основании этого и учитывая соотношение для отрезков времени в подвижной и неподвижной системах координат, находим, что  $\Delta l_{\perp} = \Delta l'_{\perp}$ . Другими словами, длина линейки для подвижного и неподвижного наблюдателей будет представляться совершенно одинаковой. Это значит, что соотношения  $y = y'$  и  $z = z'$  остаются верными и при больших скоростях движения, если оно происходит вдоль оси  $X$ .

Проведем теперь тот же самый опыт, но при этом линейку расположим горизонтально, т. е. в направлении движения. В этом случае свет, исходящий из левого конца линейки, пройдет путь  $\Delta l_1 + v_0 \Delta t_1 = c \Delta t_1$ , или  $\Delta t_1 = \frac{\Delta l_1}{c - v_0}$ .



В самом деле, пока свет будет распространяться по направлению линейки, правый конец линейки уже отойдет от своего прежнего положения на  $v_0 \Delta t_1$ . При обратном движении линейки отраженная световая волна встретит левый конец на более близком расстоянии, так как пока она будет двигаться налево, конец линейки движется навстречу ей. Стало быть,  $c \Delta t_2 =$

$$= \Delta l_1 - v_0 \Delta t_2, \text{ или } \Delta t_2 = \frac{\Delta l_1}{c + v_0}.$$

Общее затрачиваемое время в этом случае

$$\Delta t = \Delta t_1 + \Delta t_2 = \frac{2c \Delta l_1}{c^2 - v_0^2} = \frac{2 \Delta l_1}{c(1 - v_0^2/c^2)}.$$

Наблюдатель, движущийся вместе с линейкой, не заметит разницы между временем прохождения света до зеркала и от зеркала обратно к началу линейки. Для него полное время

прохождения сигнала  $\Delta t' = 2 \Delta l'_1$ . Так как  $\frac{\Delta t'}{\Delta t} = \sqrt{1 - \frac{v_0^2}{c^2}}$ ,

то  $\Delta l_1 = \Delta l'_1 \sqrt{1 - \frac{v_0^2}{c^2}}$ , или  $\Delta l_1 \gamma = \Delta l'_1$ .

Это значит, что неподвижному наблюдателю будет казаться, что линейка укоротилась по сравнению со своей прежней длиной

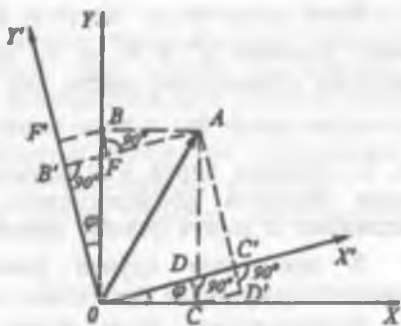
на величину  $\sqrt{1 - v_0^2/c^2}$ . Следовательно, не только время оказывается разным для разных систем координат, но и линейные размеры предметов.

**Преобразование Лоренца.** Выше мы получили соотношения между интервалами времени и длинами отрезков в системах координат, движущихся друг относительно друга равномерно и прямолинейно со скоростью  $v_0$ . Найдём теперь соотношения между координатами. Для этого вначале обратимся к некоторым математическим преобразованиям и их общим свойствам. Представим, что имеется двумерная (пока для простоты ограничимся этим случаем) декартова система координат  $XY$ , в которой задана с помощью радиуса-вектора  $OA$  точка  $A$ . Рассмотрим теперь ещё одну декартову систему координат  $X'Y'$  с тем же самым началом, но повернутую относительно первой на угол  $\varphi$  (рис. 1.21).

Так как отрезок  $OA$  есть диагональ квадратов  $OBAC$  и  $OB'A'C'$ , причём длины сторон квадратов равны координатам  $x, y$  и  $x', y'$  точки  $A$  в двух системах координат, то по теореме Пифагора

$$OA^2 = OB^2 + OC^2 = x^2 + y^2 = OB'^2 + OC'^2 = x'^2 + y'^2.$$

Рис. 1.21. К выводу формул для преобразования координат при повороте декартовых осей



Отсюда заключаем, что при повороте осей координат сумма квадратов координат любой точки в пространстве не меняется. Нетрудно видеть, что это правило выполняется и в трехмерном случае при произвольном повороте осей одной прямоугольной декартовой системы координат относительно другой, т. е.

$$x^2 + y^2 + z^2 = x'^2 + y'^2 + z'^2.$$

Найдем теперь, как преобразуются собственно координаты  $x$ ,  $y$  при повороте осей. Соответствующие соотношения можно получить на основании следующих простых рассуждений. Обратимся к рис. 1.21.

Координата  $x'$  точки  $A$  в повернутой системе численно равна длине отрезка  $OC'$ , причем  $OC' = OD + DC'$ . Точка  $D$  получается на пересечении перпендикуляра  $CD$ , проведенного к оси  $X'$  через конец отрезка  $OC$ , длина которого равна координате  $x$  в основной системе. Значит,  $OD = x \cos \varphi$ . По построению, фигура  $CDC'D'$  образует такой прямоугольник, что  $DC' = CD'$ , причем  $CD' = AC \sin \varphi = y \sin \varphi$ .

Окончательно

$$x' = OC' = x \cos \varphi + y \sin \varphi.$$

Аналогично длина отрезка

$$OB' = y' = OF' - F'B' = OF' - FB = y \cos \varphi - x \sin \varphi.$$

Итак,

$$x' = x \cos \varphi + y \sin \varphi, \quad y' = -x \sin \varphi + y \cos \varphi.$$

Полученные соотношения удовлетворяют требованию

$$x^2 + y^2 = x'^2 + y'^2.$$

В самом деле,

$$\begin{aligned} x'^2 + y'^2 &= x^2 \cos^2 \varphi + y^2 \sin^2 \varphi + 2xy \cos \varphi \sin \varphi + \\ &+ x^2 \sin^2 \varphi + y^2 \cos^2 \varphi - 2xy \cos \varphi \sin \varphi = x^2 + y^2. \end{aligned}$$

Если  $\cos \varphi = a$ ,  $\sin \varphi = b$ , то повороту координат отвечает условие  $a^2 + b^2 = 1$ , причем преобразования координат должны обязательно иметь вид

$$x' = ax + by, \quad y' = -bx + ay,$$

т. е. стоящие перед  $y$  в первом случае и перед  $x$  во втором в паре равенств коэффициенты должны быть одинаковыми по величине и иметь разные знаки.

В противном случае равенство  $x^2 + y^2 = x'^2 + y'^2$  не удовлетворяется ни при каких значениях коэффициентов  $a$  и  $b$ .

Воспользуемся этим общим свойством для вывода соотношения между переменными при переходе от неподвижной системы к движущейся.

Снова исходим из условия, что скорость света во всех системах координат, движущихся друг относительно друга равномерно и прямолинейно, одна и та же. Это означает, что если в начальный момент начала двух систем координат были совмещены друг с другом и в этот момент в соответствующей точке произошла вспышка света и начала распространяться во все стороны сферически симметричная световая волна, то в дальнейшем как движущийся, так и неподвижный наблюдатели будут регистрировать также сферические волны.

Известно, что в декартовой системе координат уравнение сферы имеет вид  $x^2 + y^2 + z^2 = R^2$ , где  $R$ —радиус сферы. Если от момента вспышки прошло время  $t$ , то  $R = ct$  и, значит,  $x^2 + y^2 + z^2 = c^2 t^2$ , где  $c$ —скорость света. Отсюда  $x^2 + y^2 + z^2 - c^2 t^2 = 0$ , причем это равенство должно быть справедливым для любой системы. Следовательно, для двух сравниваемых систем можно записать:  $x^2 + y^2 + z^2 - c^2 t^2 = x'^2 + y'^2 + z'^2 - c^2 t'^2$ .

Выведенные выше правила преобразования интервалов времени и длин отрезков при переходе от одной системы координат к другой, движущейся относительно первой равномерно и прямолинейно, показывают, что пространственные и временные координаты должны быть связаны друг с другом и уже нельзя считать течение времени везде одинаковым и, следовательно, нельзя отделять пространство от времени. Математическим следствием этого является то, что надо переходить от трехмерного к четырехмерному пространству, объединяя пространственные и временные переменные, и говорить уже о пространственно-временных системах отсчета. В евклидовом пространстве имеется только три действительные координатные оси. Поэтому при переходе к четырехмерному пространству приходится прибегать к комплексным числам и вводить комплексные параметры. Чтобы к тому же сохранить требование  $x^2 + y^2 + z^2 + u^2 = |S|^2$  (где  $|S|^2$ —квадрат модуля некоторого

вектора) и для четырехмерной системы координат (именно благодаря такому требованию и удастся удовлетворить общему принципу относительности и, если так можно выразиться, уравнивать в правах все равномерно движущиеся друг относительно друга системы отсчета и состояния покоя и равномерного прямолинейного движения) в качестве четвертой переменной примем величину  $u = ict$ , где  $i = \sqrt{-1}$  — мнимая единица.

Тогда для двух сравниваемых систем отсчета выполняется равенство

$$x^2 + y^2 + z^2 + (ict)^2 = x'^2 + y'^2 + z'^2 + (ict')^2,$$

по своей математической структуре аналогичное ранее рассмотренному для случая поворота одной системы координат по отношению к другой.

Теперь будем считать, что система отсчета движется относительно другой вдоль оси  $X$ . Из условия сохранения длин отрезков, расположенных перпендикулярно направлению движения, сразу можно заключить, что  $y = y'$  и  $z = z'$ . Значит, надо найти только соотношения между координатами  $x$  и  $x'$  и  $u$  и  $u'$ .

Мы уже знаем, что свойства сохранения сумм квадратов координат достигаются, если воспользоваться преобразованием их с коэффициентами  $a$  и  $b$  так, чтобы

$$x' = ax + bu, \quad u' = -bx + au,$$

причем  $a^2 + b^2 = 1$ . Так как  $u = ict$ , то

$$x' = ax + b(ict), \quad ict' = -bx + a(ict).$$

Координаты  $x$  и  $x'$  действительны, поэтому на основании первого равенства заключаем, что надо положить  $b = id$ . Тогда и во втором равенстве мнимая единица исчезнет и в результате получатся соотношения между действительными (а не мнимыми) величинами, характеризующими как пространственные, так и временные изменения.

Обратимся теперь к получающемуся после замены  $b$  на  $id$  первому равенству:  $x' = ax - dct$ .

Оно верно для произвольных координат  $x$  и  $x'$  в любой момент времени. Значит, для двух разных точек с координатами  $x_1, x_2$  и  $x'_1, x'_2$  в один и тот же момент времени получим

$$x'_2 - x'_1 = a(x_2 - x_1).$$

Разности  $x_2 - x_1$  и  $x'_2 - x'_1$  есть длины расположенных в направлении движения отрезков  $\Delta l$  и  $\Delta l'$ . Выше было показано, что для таких отрезков в разных системах координат должно выполняться соотношение

$$\Delta l' = (1 - v_0^2/c^2)^{-1/2} \Delta l.$$

Отсюда сразу находим  $a = (1 - v_0^2/c^2)^{-1/2}$ . Теперь учтем, что  $a^2 + b^2 = 1 = a^2 - d^2$ . Стало быть,  $d^2 = v_0^2/[c^2(1 - v_0^2/c^2)]$  и  $d = v_0/[c(1 - v_0^2/c^2)^{1/2}]$  (мы выбираем знак плюс при извлечении квадратного корня из тех соображений, чтобы найденное преобразование при  $v_0^2/c^2 \rightarrow 0$  совпадало с преобразованием координат Галилея, т. е. при малых скоростях).

Итак, окончательно искомые преобразования координат (преобразования Лоренца) имеют вид:  $x' = (1 - v_0^2/c^2)^{-1/2}(x - v_0 t)$ ,  $y' = y$ ,  $z' = z$ ,  $t' = (1 - v_0^2/c^2)^{-1/2}(t - v_0 x/c^2)$ .

Отметим, что найденное преобразование приобретает совсем близкое сходство с рассмотренным выше преобразованием простого поворота координат в двумерном евклидовом пространстве, если положить  $a = \text{ch } \varphi$  и  $d = \text{sh } \varphi$ , где  $\text{ch } \varphi$  и  $\text{sh } \varphi$  — гиперболические косинус и синус угла поворота  $\varphi$ . Для таких функций выполняется равенство  $\text{ch}^2 \varphi - \text{sh}^2 \varphi = 1$ .

В современной физике преобразования Лоренца играют очень большую роль. Эти преобразования показывают, что параметры пространства и времени связаны между собой и при переходе от одной системы к другой, если эти системы по отношению друг к другу движутся равномерно и прямолинейно, меняются одновременно. Значит, нет не только абсолютного пространства (покоя), но и времени. Это в принципе отличается от ньютоновской ситуации, где рассматриваются относительно малые скорости и где время является одним и тем же как в одной, так и в другой системах координат. Обратим внимание на то, что в выражении для переменной времени в преобразовании Лоренца второй член, содержащий пространственную координату, убывает значительно быстрее при малых  $v_0$ , чем первый член, поскольку он умножается на величину, обратную  $c^2$ . Поэтому время становится «классическим» быстрее, чем пространственная координата при переходе от больших скоростей к малым.

Итак, при малых скоростях можно считать, что временная координата не меняется при переходе от точки зрения одного наблюдателя к точке зрения другого наблюдателя, если и тот и другой движутся относительно друг друга равномерно и прямолинейно. Преобразуются только пространственные координаты.

Поэтому их в известной степени можно рассматривать независимо от времени и вести все рассуждения в трехмерном пространстве  $XYZ$ , считая, что каждая из координат  $x$ ,  $y$ ,  $z$  параметрически зависит от времени.

В релятивистском случае уже нельзя разделить пространственные координаты от времени и нужно рассматривать четырехмерное пространство в целом. В таком мире события изображаются мировой линией. Если точка покоится, то мировая линия

представляет прямую, параллельную оси времени. Если точка движется, то мировая линия представляется наклонной прямой. Без ограничения общности можно предполагать, что в начальный момент времени наблюдаемая точка находилась в начале координат. Если точка покоится, то мировая линия совпадает с направлением оси времени. Если точка движется, например, вдоль оси  $X$ , то это будет линия, лежащая в плоскости  $Xt$  и наклоненная под некоторым углом  $\varphi$ . Очевидно, что переход от одной ситуации, когда мировая линия совпадает с направлением оси времени, в другую ситуацию, когда мировая линия наклонена по отношению к этой оси, т. е. переход от состояния покоя к состоянию равномерного прямолинейного движения (или смена точек зрения наблюдателей), математически эквивалентен повороту в комплексном пространстве одной системы координат по отношению к другой на угол  $\varphi$ . В самом деле, этим поворотом ранее наклонная мировая линия может быть совмещена с линией, отвечающей оси времени.

Мы видели, что должно быть

$$x^2 + y^2 + z^2 - c^2 t^2 = x'^2 + y'^2 + z'^2 - c^2 t'^2.$$

Величина

$$S_{12} = \sqrt{c^2(t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2}$$

получила название *интервала между двумя событиями*. Величина квадрата интервала не зависит от того, какой из двух наблюдателей, находящихся в разных системах отсчета, измеряет его. Именно поэтому понятие интервала и играет такую роль в теории относительности, т. е. в общем учении о взаимосвязи пространства и времени и о преобразованиях пространства и времени при переходе от одной системы отсчета к другой.

Итак, переход от одной системы отсчета к другой системе (а это означает и переход от покоя к состоянию равномерного прямолинейного движения), если эти системы движутся относительно друг друга равномерно и прямолинейно, математически означает поворот этих координат около одного начала вокруг какой-то заданной оси. Это сразу указывает на то, что все такие системы следует рассматривать как равноправные. Мы, таким образом, не имеем оснований ожидать, что при переходе от одной системы отсчета к другой или от смены места наблюдателей будет меняться что-то физически, проявляться какие-то особенности экспериментов, которые в состоянии будут выделять одну систему отсчета из других. К этому вопросу, который непосредственно связан с принципом относительности, мы вернемся в дальнейшем после рассмотрения проблем, связанных с динамикой движения и его основными законами.

Сформулируем теперь правило релятивистского сложения скоростей. Оно выводится сразу же с помощью соотношений Лоренца, которые можно переписать в дифференциальной форме:

$$dx = \gamma dx' + \gamma v_0 dt', \quad dt = \gamma dt' + \gamma \frac{v_0}{c^2} dx'.$$

Разделив первое уравнение на второе, получим

$$\frac{dx}{dt} = \frac{dx' + v_0 dt'}{dt' + (v_0/c^2) dx'} = \frac{dx'/dt' + v_0}{1 + (v_0/c^2) dx'/dt'},$$

или

$$v = \frac{v' + v_0}{1 + (v_0/c^2)v'}.$$

Видно, что, даже если обе скорости  $v'$  и  $v_0$  будут достигать скорости света, все равно суммарная скорость будет равна только скорости света. Видно также, что при малых скоростях это соотношение переходит в привычную формулу сложения скоростей. Эта же формула показывает, что время и пространство при переходе от одной системы координат к другой преобразуются таким образом, что свет распространяется с одинаковой скоростью с точки зрения всех наблюдателей. Заметим, что при больших скоростях изменится и формула для ускорения. При равноускоренном движении скорость связана с ускорением следующим образом:  $v = at$ . Если время будет очень длительным, то формула приведет к бесконечному возрастанию скорости. Значит, она не верна. Как надо исправить эту формулу, мы увидим в дальнейшем (см. § 3.4).

Подведем итоги. Опираясь на эксперименты, мы постулировали, что скорость любых объектов не может превосходить скорость света. Скорость света есть максимальная скорость. Со скоростью, большей нее, никакие тела не могут двигаться и никакой сигнал, никакое действие не могут передаваться в пространстве. Время и пространство связаны между собой и при переходе от одной системы отсчета к другой, если эти системы движутся друг относительно друга равномерно и прямолинейно, временные и пространственные координаты преобразуются одновременно. Есть, однако, инвариант, который не меняется при переходе от одной системы координат к другой, и квадрат этого инварианта равен  $\Delta S^2 = c^2 \Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2$ . Одна система отсчета от другой отличается поворотом в пространстве, и в этом смысле все рассмотренные нами системы можно считать эквивалентными между собой и ничем не различающимися.

Если существует какая-нибудь система отсчета, в которой данное тело покоится, то существует и бесконечное множество (поскольку углы поворота можно выбирать как угодно) таких, в которых данное тело совершает равномерное и прямолинейное движение. Существование таких систем является прямым след-

ствиям первого закона Ньютона, в котором утверждается, что в тех случаях, когда на тела не действуют никакие внешние силы, они сохраняют свое состояние покоя или (подчеркиваем это) прямолинейного равномерного движения. Уже в этой формулировке, которая близка к исходной, основополагающей, оригинальной формулировке Ньютона, устанавливается эквивалентность покоя и равномерного прямолинейного движения.

#### Вопросы для самоконтроля

1. Дайте определения пространства и времени.
2. Какие системы координат вам известны? Из каких соображений должна выбираться система координат для описания того или иного круга явлений? Чего удалось в первую очередь достичь, рассматривая Солнечную систему в полярных координатах, центр которых совмещен с Солнцем?
3. Что такое степени свободы? Зависит ли число степеней свободы от выбранных координат?
4. Сколькими степенями свободы обладает указательный палец человека? Какими координатами лучше всего воспользоваться для того, чтобы наиболее наглядным образом охарактеризовать его возможные движения? Сколькими степенями свободы обладает кисть человека?
5. Какова взаимосвязь между пространственными координатами и временем в ньютоновской механике и в механике Эйнштейна?
6. Что такое траектория и мировая линия? Какими основными свойствами должны обладать эти кривые? Как изображается состояние покоя в ньютоновской механике и в четырехмерном пространстве—времени?
7. По каким правилам вводится вектор перемещения?
8. Что такое вектор скорости движения и как перейти от реально измеримой скорости движения к скорости движения, изображаемой производной от радиуса-вектора точки по времени?
9. Воспроизведите рассуждения, которые приводят к понятию вектора угловой скорости.
10. Как найти соотношение между векторами угловой и линейной скоростей?
11. Какие два простейших вида движения лежат в основе любых более сложных?
12. Что такое ускорение движения, угловое ускорение, нормальное, тангенциальное и полное ускорения при вращательном движении? Напишите формулу, связывающую полное ускорение при вращательном движении с нормальным и тангенциальным. Как меняется направление вектора полного ускорения по отношению к радиусу-вектору точки в данный момент времени при ускоренном и замедленном вращательном движении?
13. Какое противоречие с современными экспериментальными данными возникает при больших скоростях, если воспользоваться соотношением между координатами в двух движущихся друг относительно друга прямолинейно и равномерно системах координат по Ньютону? Приведите конкретный пример.
14. Как изменится интервал времени и отрезок длины при переходе от одной системы координат к другой, если системы движутся друг относительно друга равномерно и прямолинейно и если предположить, что скорость распространения света и в том и в другом случае будет одной и той же и является некоторой мировой константой?
15. Как преобразуются длины катетов и гипотенузы прямоугольного треугольника при переходе от одной системы координат к другой, движущейся относительно первой равномерно и прямолинейно вдоль направления, совпадающего с одним из катетов?
16. Как преобразуются двумерные прямоугольные декартовы координаты  $x$ ,  $y$  при повороте этих координат вокруг оси, перпендикулярной плоскостям  $OX$  и  $OY$ , на некоторый угол  $\alpha$ ? Какая характеристика останется при этом неизменной?
17. Воспроизведите рассуждения, которые приводят к тому, что в качестве координаты времени в четырехмерной системе пространство—время необходимо взять величину, равную  $ict$ , где  $i$ —мнимая единица.



18. Что такое интервал в релятивистской кинематике?

19. Запишите преобразование Лоренца для пространственных и временной координат при переходе от одной системы координат к другой, движущейся относительно первой равномерно и прямолинейно со скоростью  $v_0$  вдоль оси  $X$ .

20. Воспользовавшись формулой для релятивистского сложения скоростей, найдите суммарную скорость в случае, когда одна система по отношению к другой движется со скоростью  $\frac{1}{2}$  скорости света и объект внутри этой второй системы перемещается со скоростью соответственно  $\frac{1}{2}$  скорости света,  $\frac{1}{2}$  скорости света,  $\frac{1}{2}$  скорости света.



## ГЛАВА 2

### ДИНАМИКА

*«Сегодня в полдень пущена ракета.  
Она летит куда скорее света.  
И долетит до цели в семь утра  
Вчера».*

*Английская эпиграмма*

#### 2.1. Сила и масса, второй закон Ньютона

В гл. 1 мы рассматривали вопрос о том, как можно описать движение точки или тела. Были введены понятия мировой линии и траектории, числа степеней свободы, скорости тела и ускорения. Были также обсуждены специфические эффекты, возникающие при перемещении тела со скоростями, близкими к скорости света. Мы также выяснили, что понятия покоя тела или его движения по отношению к некоторой системе координат равномерно и прямолинейно зависят от позиции наблюдателя и, в сущности, являются эквивалентными.

Существенную особенность в движение вносит, однако, появление ускорения, которое нельзя свести к состоянию покоя или равномерного прямолинейного движения. В самом деле, в тех случаях, когда движение ускоренное, мировая линия уже не прямая, которая в простейшем случае исходит из начала координат. В общем случае мировая линия может иметь форму весьма сложной кривой. Какие бы повороты системы координат мы при этом ни производили, стремясь, как это делали раньше в случае равномерного прямолинейного движения, совместить мировую линию с осью времени (что и означало переход к такой системе координат, где тело покоится), нам не удастся это сделать. Действительно, кривую линию с прямой линией, представляющей собой ось времени, совместить, пользуясь только поворотами системы координат, никак нельзя. Здесь интересно отметить следующее. Математически можно допустить, что переход от одной системы координат к другой системе координат заключается не только в повороте осей, но и в изменении на различных участках масштаба элементарных длин, откладываемых вдоль осей координат. Нетрудно видеть, что всегда, даже

если мировая линия является кривой, используя различные интервалы единичной длины вдоль оси времени, можно как бы спрямить эту кривую линию. Тогда последующим поворотом системы координат можно было бы совместить мировую линию с осью времени.

Однако такая операция, вполне допустимая с математической точки зрения, физически означала бы, что мы допускаем неравномерное в разные моменты течение времени. Между тем именно принцип равномерного течения времени является одним из основных и главнейших принципов физики, тесно связанных с таким важнейшим законом природы, как закон сохранения энергии. Позже мы увидим, что допущение неравномерного течения времени автоматически приводит к допущению о невыполнении этого закона. Однако не имеется ни одного наблюдения или эксперимента, который показывал бы, что закон сохранения энергии может нарушаться. Стало быть, появление ускоренного движения есть появление некоторого нового качества, существенно отличающегося от состояния покоя или равномерного прямолинейного движения. В физике появление ускорения связывают с действием сил. Как и многое другое, понятие силы является расширением привычного бытового понятия.

В повседневной жизни мы часто пользуемся термином «сила», обозначая им мускульное действие, которое приходится затрачивать для того, чтобы привести какое-нибудь тело в движение или, наоборот, затормозить это движение до полной остановки. Во всех случаях меняется и скорость движения, т. е. появляется ускорение. В физике это бытовое определение силы расширяется. *Принято силой называть любую причину изменения скорости тела.* При этом скорость тела как вектор может менять и направление, и модуль.

Реальные физические силы не могут возникнуть из ничего. Появление силы связано с наличием взаимодействия между отдельными телами и телами и полями. Эти взаимодействия в конкретных условиях эксперимента могут проявляться самым разнообразным образом.

В результате многочисленных исследований в современной физике установлено, что *самые разнообразные взаимодействия могут быть сведены к некоторым основным фундаментальным типам.* Известно четыре типа таких взаимодействий. Это прежде всего *сильные взаимодействия* или *ядерные*. Они ответственны за существование ядер и скрепляют частицы, образующие атомные ядра, друг с другом в одну прочную систему. Следующий тип — *электромагнитные взаимодействия*. Они, в частности, проявляются в том, что если два тела имеют электрические заряды одного знака, то они отталкиваются одно от другого. Если же два тела имеют заряды противоположных знаков, то они притягиваются одно к другому. Следующим типом взаимодействия является также хорошо известное *гравитационное*

**взаимодействие.** Благодаря этому взаимодействию существуют, в частности, планетарные системы и та планетарная система, к которой принадлежит наша Земля. Наконец, последний тип взаимодействия—*слабые взаимодействия*, проявляющиеся в некоторых типах радиоактивного распада и связанные с некоторыми тонкими эффектами взаимодействия между собой элементарных частиц.

Глубокой мечтой физиков-теоретиков является попытка свести все эти взаимодействия к какому-нибудь новому, нам сегодня не известному, единственному типу фундаментальных взаимодействий. Удастся ли это сделать или не удастся, сказать трудно, однако необходимо отметить, что одним из очень крупных достижений теоретической физики последнего времени является построение теории *электрослабых взаимодействий*, которые позволили свести электрические (электромагнитные) и слабые взаимодействия к одному электрослабому взаимодействию, которое и проявляется в зависимости от условий конкретного эксперимента в виде либо электромагнитных, либо слабых взаимодействий, известных уже довольно давно. Таким образом, можно сказать, что сейчас мы уже имеем дело всего с тремя типами фундаментальных взаимодействий. Посмотрим, к чему приведет развитие теоретической и экспериментальной физики в дальнейшем.

Утверждение, что всякие силы есть результат взаимодействий между отдельными телами или телами и полями (заметим, что поля сами порождаются какими-нибудь источниками), аналогично в известной степени утверждению о том, что понятие движения имеет смысл тогда, когда указано, по отношению к чему тело движется. Как только мы наблюдаем какое-то ускорение, всегда должно задать себе вопрос: с чем произошло взаимодействие того тела, за ускорением которого следим?

Рассмотрим простые механические явления, которые происходят при относительно небольших скоростях и которые могут быть подвергнуты простой экспериментальной проверке.

Мы уже сказали выше, что силу мы будем определять как некоторую причину изменения скорости тела. Так как ускорение есть векторная величина, то естественно вводить силу так, чтобы математически она выражалась тоже вектором, совпадающим по направлению с вектором ускорения, а по модулю пропорциональным этому вектору. Соответствующий коэффициент пропорциональности, разумеется, должен быть скалярной величиной. Этот коэффициент пропорциональности принято называть *массой тела*. В дальнейшем будет показано, что на самом деле все не так просто и ввести массу таким способом можно только в том случае, если тело движется со скоростью, существенно меньшей скорости света. Соотношение  $f = ma$ , где  $f$ —вектор силы,  $a$ —вектор ускорения и  $m$ —масса материальной точки, было введено в физику Ньютоном и получило название *второго*

закона Ньютона. Этот закон по причинам, о которых мы поговорим ниже, имеет принципиальное значение.

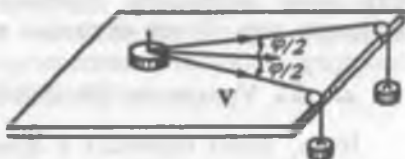
Обратим внимание на то, что в рассматриваемое соотношение помимо введенного в предыдущей главе ускорения, для определения и измерения которого достаточно располагать двумя кинематическими характеристиками—интервалами длины и времени,—входят еще две новых составляющих—вектор силы и масса. Направление вектора силы мы договорились принимать таким же, как и вектора ускорения. Остается, стало быть, выбрать его модуль. Сразу видно, что при попытке решить эту проблему мы столкнемся с тем фактом, что так как модуль силы  $|f|$  и масса входят в одно и то же равенство, то определить их независимым образом на основании каких бы то ни было экспериментов вообще невозможно. В самом деле, либо мы должны действовать разными силами на одну и ту же материальную точку с принятой за стандарт массой, что позволит сравнивать между собой модули сил и расположить их в определенном ряду, либо выбрать стандарт силы (калиброванный динамометр) и действовать этой силой на точки с разными массами, что позволит расположить в ряд эти последние. В любом случае вполне определенно можно найти лишь ряды относительных величин:  $m_1/m_2 = |a_2|/|a_1|$ ,  $m_1/m_3 = |a_3|/|a_1|$  и т. д. или  $|f_1|/|f_2| = |a_1|/|a_2|$ ,  $|f_1|/|f_3| = |a_1|/|a_3|$ .

Отсюда видно, что как  $m$ , так и  $|f|$  определяются лишь с точностью до констант.

Теперь вспомним, что всякое тело, в частности и материальная точка, обладает способностью не только приобретать под действием определенной силы соответствующее ускорение, но и сжимать или растягивать пружину находящихся на поверхности земли неподвижных относительно нее пружинных весов (простейшими весами такого рода является подвешенный на штатив всем известный из школьного курса физики динамометр).

Если сопоставить ряд чисел, передающих обратные относительные значения модулей ускорений, получаемых различными телами под действием одной и той же силы, и ряд чисел, передающих относительные значения отсчетов пружинных весов, когда на них воздействует та же последовательность тел, то выясняется, что эти ряды чисел будут с высокой степенью точности совпадать между собой. Такие эксперименты проводились неоднократно, еще И. Ньютоном. Но если И. Ньютону удалось установить совпадение обсуждаемых рядов чисел с точностью лишь до  $10^{-3}$ , то один из последних опытов, проведенных в 1974 г., показал, что соответствующие числа совпадают между собой с точностью до  $10^{-10}$ . Такое совпадение не может быть случайным и свидетельствует о том, что речь идет об одном и том же фундаментальном свойстве тел. Тем не менее в пределах экспериментов, основанных на классической механике, остается неопределенность, которая и послужила основанием для

Рис. 2.1. Под действием одного груза шайба будет двигаться вдоль направления тяги. Под действием одновременно двух одинаковых грузов шайба будет двигаться по биссектрисе угла между тягами



введения в физику двух понятий: инертной и гравитационной масс. Лишь позднейшие исследования А. Эйнштейна позволили уточнить эти понятия и показать, что во всех случаях мы имеем дело с одной и той же массой. С этими вопросами мы познакомимся ниже, в разделе о релятивистской динамике.

В заключение этого параграфа остановимся еще на одном вопросе. Поскольку ускорение есть вектор, его всегда можно представить векторной суммой двух векторов и более. Тогда и вектор силы, пользуясь введенным выше правилом, также можно представить векторной суммой некоторых частных сил. Физически такая ситуация реализуется, в частности, в простом эксперименте, показанном на рис. 2.1. В этом случае на лежащую на столе шайбу одновременно воздействуют две тяги. Шайба будет приобретать ускорение, направленное вдоль по вектору  $\vec{F} = \vec{f}_1 + \vec{f}_2$ , которым характеризуется одновременное действие на тело двух сил. Такой суммарный вектор называется *равнодействующей*. Существенно, что при построении равнодействующей принимается (это подтверждается многочисленными экспериментами), что действие каждой силы не зависит от действия другой, т. е. векторы частных сил складываются аддитивно. Этот факт—одно из проявлений довольно общего принципа суперпозиции.

Здесь необходимо сделать важное замечание. Математически сила изображается так называемым *скользящим вектором*. Скользящим называется вектор, для которого несущественно расположение начала вектора в пространстве, а имеют значение лишь величина (модуль) и направление. Скользящий вектор не меняется при переносе его вдоль прямой, совпадающей с собственным направлением вектора. Сама такая прямая называется *линией действия силы*. Возможность описания силы скользящим вектором в какой-то степени отражает тот экспериментальный факт, что результат будет одним и тем же, безразлично за длинную или короткую веревку человек тянет груженую вагонетку.

Известное из школьного курса геометрическое правило сложения векторов (правило параллелограмма) может быть применено к векторам сил только в том случае, когда линии действия всех векторов сил пересекаются в одной точке. Для сил, приложенных к материальной точке это условие, очевидно, всегда выполняется. В случае же твердого тела или еще более сложной системы это уже может быть не так (силы приложены к разным точкам тела). К каким последствиям это приводит и

как надо учитывать действие на тело одновременно многих внешних сил, будет показано в § 3.1.

## 2.2. Уравнение движения

Итак, снова вернемся к фундаментальному исходному соотношению:  $F = ma$ . Перепишем его, вспомнив определение ускорения как производной. Тогда  $m \frac{dv}{dt} = F$ , или, если перейти к декартовой системе координат и воспользоваться координатной формой,  $m \frac{dv_x}{dt} = F_x$ ,  $m \frac{dv_y}{dt} = F_y$ ,  $m \frac{dv_z}{dt} = F_z$ .

Будем считать массу постоянной величиной. Сила  $F$  может, конечно, зависеть, во-первых, от места положения точки в пространстве и, во-вторых, изменяться со временем. Поэтому более подробная запись должна иметь вид  $m \frac{dv_x}{dt} = F_x(x, y, z, t)$  и т. д. для других компонент.

Во всех случаях пока будем принимать тело за материальную точку, для которой существенно только, где она расположена, и которая описывается тремя степенями свободы. Так как скорость, в свою очередь, является производной от радиуса-вектора, то

$$m \frac{d^2 r}{dt^2} = F(r, t)$$

или в координатах

$$m \frac{d^2 x}{dt^2} = F_x(x, y, z, t), \quad m \frac{d^2 y}{dt^2} = F_y(x, y, z, t),$$

$$m \frac{d^2 z}{dt^2} = F_z(x, y, z, t).$$

С математической точки зрения эти соотношения представляют собой систему дифференциальных уравнений. Принято называть их *уравнениями движения*. Если каждая из компонент силы  $F_x$ ,  $F_y$ ,  $F_z$  зависит только от своей переменной ( $F_x$ —функция только  $x$ ,  $F_y$ —функция только координаты  $y$  и т. д.), то тогда получается три независимых дифференциальных уравнения, которые и можно решить по отдельности.

В простейших случаях так и бывает. Однако в общих случаях из-за того, что компонента силы  $F_x$  может зависеть от расположения данного тела вдоль оси  $Y$  или  $Z$  и наоборот, получается «зацепляющаяся» система уравнений, и тогда их надо разрешать в целом. Это представляет нередко довольно сложную задачу, но в общем-то это дело математики, и, опираясь на те или иные приемы, такую задачу всегда решить можно. Итак, посмотрим, что же получится, если удастся решить уравнения,

подобные описанным выше. Для простоты рассмотрим одномерное движение. Тогда соответствующие соотношения между определяющими движением величинами, являющиеся прямым следствием второго закона Ньютона, будут иметь вид

$$m \frac{d^2 x}{dt^2} = m \frac{dv_x}{dt} = F_x(x, t).$$

Как уже говорилось, основной задачей физики как науки является предсказание событий. Как же предсказать, где будет находиться материальная точка через некоторый промежуток времени и с какой скоростью она будет двигаться? Чтобы получить ответ на этот вопрос, рассмотрим ряд примеров.

Предположим, что на материальную точку вообще не действуют никакие силы. Тогда уравнение движения для этого тела имеет вид  $m \frac{dv_x}{dt} = 0$ . Сокращая массу, получаем  $\frac{dv_x}{dt} = 0$ .

Следовательно, движение тела характеризуется тем, что производная от скорости движения по оси  $X$  по времени равна нулю. Производная равна нулю только в том случае, если сама функция есть постоянная величина. Следовательно, решение дифференциального уравнения  $v_x = \text{const}$ . Обозначим эту константу  $c_1$ . Заметим, что никакого другого вывода, решая соответствующее дифференциальное уравнение, получить нельзя и константа остается неопределенной. Продолжим рассуждение. Скорость  $v_x$ , в свою очередь, определяется производной  $dx/dt$ .

Значит,  $\frac{dx}{dt} = c_1$ , т. е. производная от функции  $x$ , определяющей пространственное расположение тела, постоянная величина. Тогда функция  $x = c_1 t + c_2$ . Прибавление еще одной константы  $c_2$  вызвано тем, что производные всех функций, отличающихся друг от друга на постоянную величину, одинаковы. Следовательно, решая дифференциальное уравнение  $\frac{dx}{dt} = c_1$ , невозможно получить ответ лучше, чем с точностью до аддитивно прибавляемой постоянной. Итак, общий ответ, который мы получаем, решая уравнение движения для случая, когда на тело вообще не действуют никакие внешние силы, для скорости и пространственной координаты  $x$  имеет вид

$$v_x = c_1, \quad x = c_1 t + c_2.$$

Рассмотрим теперь еще один, тоже достаточно простой, пример. Предположим, что действующая на тело сила есть постоянная величина  $F_x = F_0$ . Уравнение движения тела

$m \frac{dv_x}{dt} = F_0$ . Тогда производная от функции  $v_x$  по времени,



равная  $F_0/m$ , есть постоянная величина. Следовательно, функция  $v_x = \frac{F_0}{m}t + c_1$ . Эта постоянная добавлена из тех же самых соображений, что и выше. Так как  $v_x = dx/dt$ , то можно записать новое дифференциальное уравнение:  $\frac{dx}{dt} = \frac{F_0}{m}t + c_1$ .

Решая его, находим

$$x = \frac{1}{2} \frac{F_0}{m} t^2 + c_1 t + c_2,$$

где  $c_2$ —новая постоянная.

Как видим, если в первом примере, где рассматривалось движение тела, когда на него не действуют никакие силы, мы получили, что  $x$  пропорционально времени, то во втором случае  $x$  пропорционально квадрату времени. По своему характеру первое движение отвечает равномерному прямолинейному движению, а второе—равномерно ускоренному. Но и в том и в другом случае, решая уравнения движения, мы только выяснили его характер, но не смогли ничего сказать о числовых значениях скорости или координаты  $x$  через определенный промежуток времени от начального до интересующего нас момента. Ответ получен неопределенный с точностью до некоторых постоянных  $c_1, c_2$ .

Опираясь только на уравнение движения или только на второй закон Ньютона, нельзя найти эти постоянные. Чего-то не хватает! Чтобы выяснить, чего именно, вспомним, что, когда говорят о движении тела, о его местоположении, всегда нужно указывать по отношению к чему. Точно так же, если надо предсказать, где будет тело через какой-то промежуток времени или с какой скоростью оно будет двигаться, необходимо указать, где оно находилось в начальный момент и с какой скоростью в этот начальный момент двигалось, т. е. откуда и каким образом началось движение. Такое указание есть задание начальных условий; оно является необходимым для получения определенного ответа. Будем считать, что при  $t = 0$  материальная точка имела координату  $x_0$  и скорость  $v_0$ ; тогда, подставляя  $t = 0$  в соотношение  $x = c_1 t + c_2$ , получаем  $x = x_0 = c_2$ . Скорость  $v_x$  для первого примера равна  $c_1$  и, стало быть, на основании начального условия ( $v_x = v_0$  при  $t = 0$ ) мы должны заключить, что  $c_1 = v_0$ . Итак, окончательно  $x = v_0 t + x_0$ . Это и есть хорошо известная нам из школьного курса формула для пути при равномерном прямолинейном движении. Подставляя во втором примере значения координаты и скорости в момент времени  $t_0$ , т. е.  $x = x_0$  и  $v_x = v_0$ , в выражения  $v_x =$

$= \frac{F_0}{m} t + c_1$ ,  $x = \frac{1}{2} \frac{F_0}{m} t^2 + c_1 t + c_2$ , находим  $c_2 = x_0$ ,  $v_x = \frac{F_0}{m} t + v_0$ . Таким образом, окончательные ответы получаются уже вполне определенными.

Рассмотрим третий, более сложный пример. Представим, что некоторое тело движется по-прежнему вдоль оси  $X$  в сопротивляющейся среде так, что сила сопротивления пропорциональна скорости этого движения. Будем считать, что никакие другие силы, кроме силы сопротивления среды, на тело не действуют. Такая ситуация реализуется, например, при движении корабля в море после выключения его двигателей. В этом случае скорость корабля постепенно уменьшается вплоть до нуля (до полной остановки). Изучим это движение. Запишем соответствующее уравнение движения корабля:

$$m \frac{dv_x}{dt} = F_x = -kv_x.$$

Знак минус выбран нами потому, что сила в данном конкретном случае препятствует движению и, стало быть, направлена в сторону, противоположную скорости;  $k$ —коэффициент пропорциональности (постоянная величина, связанная в конкретном случае со свойствами жидкости и с формой тела);  $v_x$ —скорость движения тела.

Это дифференциальное уравнение первого порядка. Решим его с помощью обычных математических приемов. Для этого перенесем дифференциал времени в правую часть уравнения:

$$\frac{dv_x}{v_x} = -\frac{k}{m} dt.$$

Проинтегрируем правую и левую части уравнения, т. е. выполним следующую операцию:

$$\int \frac{dv_x}{v_x} = -\frac{k}{m} \int dt.$$

После интегрирования получим  $\ln v_x = -k/m \cdot t + c_1$ . Появление постоянной  $c_1$  нам уже понятно. В самом деле, такое интегрирование можно выполнить только с точностью до постоянной. Дифференцированием легко проверить, что полученное решение удовлетворяет уравнению движения. Для того чтобы определить постоянную  $c_1$ , необходимо указать, как двигалось тело в начальный момент времени. При  $t = 0$  получаем, что  $c_1 = \ln v_0$ .

Итак,

$$\ln v_x - \ln v_0 = -(k/m)t, \text{ или } v_x/v_0 = e^{-(k/m)t}, \text{ или}$$

$$v_x = v_0 e^{-(k/m)t}.$$

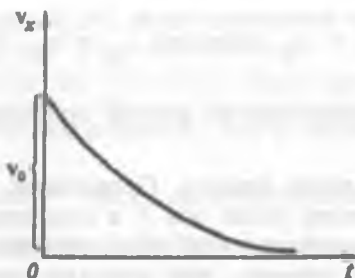


Рис 2.2. График зависимости скорости движения корабля в море при торможении его сопротивлением воды

Можно построить график зависимости  $v_x$  от времени (рис. 2.2). Он имеет вид спадающей экспоненты. Эта экспонента будет спадать тем быстрее, чем больше коэффициент сопротивления  $k$  и чем меньше масса, т. е. тела большей массы тормозится медленнее, что и подтверждается экспериментом.

Найдем теперь закон изменения пространственной координаты. Для этого воспользуемся законом изменения скорости тела:

$$v_x = \frac{dx}{dt} = v_0 e^{-(k/m)t}.$$

Снова получаем дифференциальное уравнение первого порядка. Перепишем его в виде

$$dx = v_0 e^{-(k/m)t} dt$$

и проинтегрируем правую и левую стороны:

$$x = v_0 (-m/k) e^{-(k/m)t} + c_2.$$

Пусть при  $t = 0$  материальная точка имела координату  $x_0$ . Подставляя это значение в полученную формулу, находим, что постоянная  $c_2 = x_0 + (m/k)v_0$ . Окончательно, в исследуемом нами движении координата

$$x = x_0 + \frac{m}{k} v_0 (1 - e^{-(k/m)t}).$$

Таким образом, пройденный путь тем больше, чем больше начальная скорость, масса тела и чем меньше коэффициент сопротивления  $k$  среды.

В качестве последнего примера рассмотрим явление невесомости в космосе. Его неоднократно все читатели, по-видимому, наблюдали во время телевизионной передачи, производимой с борта космического корабля.

Космонавты специально демонстрируют это явление, например, следующим образом. Космонавт берет в руки какой-нибудь достаточно тяжелый, по земным понятиям, предмет, например киноаппарат, поднимает его на уровень глаз и отпускает. На Земле этот аппарат немедленно должен упасть, а в космосе он

продолжает держаться около лица космонавта так, как если бы он потерял свой вес.

На первый взгляд, возникает некий парадокс. В самом деле, мы знаем, что силы всемирного тяготения продолжают действовать. Ведь если бы их не было, то спутник не совершал бы кругового движения вокруг Земли, а улетел бы куда-то в мировое пространство. В свою очередь, сила тяжести—это такая характеристика, которая и возникает потому, что имеется всемирное тяготение. И вот получается, что силы всемирного тяготения продолжают существовать, а вес якобы исчез. В чем же дело? Прежде всего обратим внимание на то, что экспериментальным фактом является именно движение космонавта и аппарата по одинаковой траектории уже после того, как непосредственный контакт между космонавтом и аппаратом исчез. Это и выражается в том наблюдаемом с Земли эффекте, что относительное расположение космонавта и аппарата не меняется. Попробуем понять это, опираясь на уравнение движения и на некоторые закономерности, отсюда следующие.

Рассмотрим неподвижное шарообразное тело массой  $M$ . Пусть на достаточно большом расстоянии от поверхности этого тела, таком, что можно в дальнейшем считать, что вся масса тела сосредоточена в его центре, движутся два малых тела массами  $m_1$  и  $m_2$ . Допустим, что никакие силы, кроме сил притяжения по закону всемирного тяготения, со стороны большого тела на эти малые тела не действуют. Тогда уравнение движения тела массой  $m_1$

$$m_1 \frac{d^2 r_1}{dt^2} = -(G M m_1 / |r_1|^3) \cdot r_1,$$

где  $r_1$ —радиус-вектор, соединяющий центр тела массой  $M$ , которое мы и принимаем в качестве неподвижного, с малым телом. Знак минус в этом уравнении ставится потому, что радиус-вектор  $r_1$  направлен от большого тела к малому, а сила тяготения, приложенная к малому телу, направлена от малого тела к большому, т. е. противоположно  $r_1$ . Так как

$$\left| G \frac{M m_1}{|r_1|^3} r_1 \right| = G \frac{M m_1}{|r_1|^2},$$

то принятая запись соответствует закону

всемирного тяготения Ньютона. Аналогичное уравнение может быть записано, очевидно, и для второго тела. И в первом, и во втором уравнениях можно сократить левые и правые части на массу  $m_1$  (или  $m_2$ ). В результате оба уравнения будут иметь одинаковый вид. Если к тому же считать, что в начальный момент времени два тела с массами  $m_1$  и  $m_2$  располагались друг относительно друга на столь малых расстояниях, что их

радиусы-векторы  $r_1$  и  $r_2$  также приблизительно одинаковы, то мы получим, что уравнения движения масс  $m_1$  и  $m_2$  по своему характеру не только похожи, но просто тождественны. Их решения, а стало быть, и движение объектов с массами  $m_1$  и  $m_2$  будут абсолютно одинаковыми, если совпадают исходные начальные условия. А как раз именно это и имеется внутри космического корабля, когда вначале космонавт держал в своих руках киноаппарат и, следовательно, находился с ним в каком-то контакте. Ясно, что в этом положении они движутся в пространстве совершенно одинаково. Когда контакт исчез, т. е. космонавт выпустил аппарат из рук, то оба тела (космонавт и аппарат) обязаны двигаться так же, как они двигались в тот момент, когда были в состоянии контакта. Именно это и наблюдает зритель, разглядывая всю эту картину с помощью телевизионной камеры, находящейся внутри космического корабля.

Можно показать, что такое явление невесомости будет наблюдаться во всех случаях, когда имеется движение каких-нибудь близко расположенных объектов в поле сил тяготения любого количества космических тел. Необходимо только, чтобы никакие другие силы, кроме тех, которые возникают в результате наличия всемирного тяготения, на рассматриваемые тела не действовали. Здесь уместно остановиться на еще одном интересном вопросе.

До недавнего времени все механические эксперименты проводились только на Земле. Обобщением таких экспериментов и явился второй закон Ньютона. Однако во всех этих экспериментах в той или иной степени присутствуют мешающие факторы. Таковыми являются и вращение Земли вокруг оси, и наличие притяжения всех тел к земной поверхности. Поэтому если подобные механические эксперименты производились бы с очень высокой степенью точности, то даже в тех случаях, когда скорости движения объектов будут существенно меньшими, чем скорость света, мы все равно получили бы некоторое отклонение от этого закона. Возникает вопрос, а нельзя ли указать такую постановку эксперимента, при которой законы Ньютона выполнялись бы с абсолютной точностью? Системы координат, в которых второй закон Ньютона выполняется с абсолютной точностью, называются *инерциальными*. Так вот, идеальными инерциальными системами координат являются такие, которые, если так можно сказать, находятся в состоянии свободного падения под действием всемирного тяготения. Стало быть, эксперименты по проверке точности могли бы быть поставлены именно в космических кораблях, находящихся в свободном полете, т. е. с выключенными двигателями, в мировом пространстве. Опыты, производящиеся в земных условиях, строго говоря, не корректны. Таким образом, в этом смысле второй закон Ньютона является определенной абстракцией. К счастью, поп-

равки на неидеальность земных условий оказываются небольшими.

Продemonстрируем значение решения уравнения движения и анализа этого решения, рассмотрев поведение так называемых следящих систем. Это чрезвычайно важный класс физических объектов, изучение которых позволяет понять, не только как действует, например, радиолокатор, но и почему не падает прямоходящий человек и даже как в простейшем случае функционируют государственные учреждения (см. последнюю главу учебника—«Уроки физики»).

Как и всегда, общие принципы постараемся понять, оперируя простейшей моделью.

Пусть имеются две материальные точки, которые могут двигаться только вдоль оси  $X$ . Их координаты по отношению к некоторому началу обозначим  $x_1$  и  $x_2$ . Будем считать, что на точку 2 действует также вдоль  $X$  некоторая сила  $f$ , которая зависит от разности  $x_1 - x_2$  и обращается в нуль, если  $x_1 = x_2$ . Если в начальный момент сила, действующая на точку  $x_2$ , станет отличной от нуля, точка  $x_2$  также придет в движение (будет «следить» за движением первой точки).

Примем, далее, как это и бывает в реальных следящих системах, что сигнал  $f = f(x_1 - x_2)$  воздействует на точку 2 не непосредственно, а через некоторое управляющее устройство, например усилитель. Сделаем еще один шаг в сторону реальных систем и учтем, что ни одно устройство управления не может работать совершенно неинерциально и его выходной сигнал всегда будет запаздывать по отношению к мгновенному изменению входного сигнала  $x_1 - x_2$ . Другими словами, если разность  $x_1 - x_2$  изменится скачком (точка  $x_1$  внезапно сдвинется на некоторое расстояние и далее остается неподвижной, а точка  $x_2$  не успела за ней последовать), то величина силы будет меняться во времени, как правило, так, как показано на рис. 2.3, т. е. приближаться к максимальному значению сигнала по закону экспоненты.

Такая реакция усилителя на входной сигнал описывается дифференциальным уравнением первого порядка:

$$df/dt + af = b(x_1 - x_2) = \text{const.}$$

В данном случае символом  $a$  ( $a > 0$ ) обозначен некоторый коэффициент, характеризующий управляющий элемент системы, а символом  $b$ —коэффициент размерности.

Пользуясь математическим правилом решения таких уравнений, не трудно найти, что

$$f = \frac{b(x_1 - x_2)}{a} [1 - \exp(-at)].$$

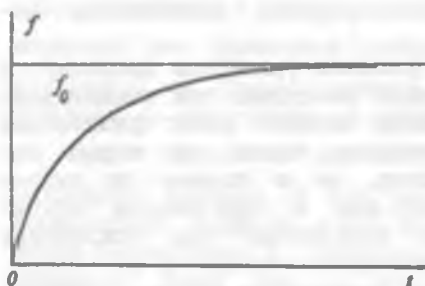


Рис. 2.3. Реакция усилителя на входной сигнал.  $f_0$ —максимальное значение выходного сигнала усилителя, отвечающего разности координат первой и второй точек. Считаем, в данном случае, что эта разность, возникнув в момент времени  $t_0$ , потом уже не меняется. Это возможно, если точка  $x_2$  еще не успела заметно сдвинуться

Видно, что поведение величины  $f$  вполне согласуется с рис. 2.3.

Принимая во внимание эти соображения, будем считать, что при непрерывном изменении положения как первой точки ( $x_1$ ), так и «следящей» за ней второй точки ( $x_2$ ) значение силы подчиняется уравнению

$$df/dt + af = b(x_1(t) - x_2(t)).$$

Разность  $\epsilon(t) = x_1(t) - x_2(t)$  называется управляющим сигналом, или сигналом рассогласования.

Обратимся ко второй точке. Каким должно быть уравнение движения этой точки?

Во-первых, на нее действует сила  $f$ . Оказывается, однако, что для хорошего функционирования следящей системы этого недостаточно. В самом деле, представим себе, что первая точка сдвинулась рывком на некоторое расстояние и затем остановилась. Вторая точка последовала за ней и догнала ее (случай  $x_1 = x_2$ ). В этот момент действующая на вторую точку сила обратится в нуль. Это не значит, однако, что точка остановится: она по инерции проскочит положение первой точки. Конечно, после этого снова возникнет не равная нулю разность координат двух точек и, следовательно, новая сила  $f$ , которая будет стремиться вернуть вторую точку к положению первой. При таком возвратном движении вторая точка снова проскочит положение первой и т. д. Не трудно сообразить, что в результате возникнут колебания второй точки около положения первой. Работа же идеальной системы должна быть такой, чтобы вторая точка, пусть и с некоторым запозданием, достигла положения первой, а потом не двигалась (если первая неподвижна). Стало быть, надо предусмотреть возможность быстрого уменьшения размахов таких колебаний, вплоть до их полного уничтожения. Это можно сделать, если ввести дополнительную тормозящую силу, действующую на вторую точку противоположно направлению ее движения и пропорциональную скорости

движения, т. е.  $f_{тр} = -k \frac{dx_2}{dt}$  ( $k > 0$ ). С такой силой (силой

вязкого трения) мы встречались выше, когда рассматривали задачу о замедлении скорости движения корабля в море после выключения двигателей.

Если все это учесть, то уравнение движения второй точки следует записать так ( $m$ —масса этой точки):

$$m \frac{d^2 x_2}{dt^2} + k \frac{dx_2}{dt} = f.$$

Введенная выше пара уравнений (одно для силы  $f$  и второе для координаты второй точки) может быть сведена к одному уравнению третьего порядка, которое решается численными методами при любом поведении первой точки, задаваемом функцией  $x_1(t)$ . Мы этого делать, однако, не будем, а приведем ряд результатов для некоторых простейших ситуаций.

На рис. 2.4—2.7 показана реакция следящей системы на входной скачкообразный сигнал для разных параметров системы: массы второй точки, инерции усилителя (коэффициент  $a$ ) и коэффициента вязкого трения ( $k$ ). Первый из этих рисунков показывает, как действует идеальная система: она быстро достигает нового состояния и почти сразу замирает около него. Во втором случае затухание очень велико и система с большим запаздыванием идет к желаемому новому положению, долго не достигая его. Если уменьшить коэффициент вязкого трения (затухание), то система начнет совершать затухающие колебания около нового положения. В зависимости от значения  $k$ ,  $m$  и  $a$  размахи этих колебаний и общая скорость достижения нового стабильного состояния будут меняться. Примеры на рис. 2.6—2.7 отвечают разным отношениям коэффициентов  $k$  и  $m$ : во втором случае отношение  $k/m$  больше, чем в первом.

Если первая точка движется равномерно и прямолинейно (верхняя прямая на рис. 2.8), то при значительном торможении (коэффициенте  $k$ ) вторая точка будет сильно отставать от первой. При уменьшении коэффициента  $k$  (затухания) отставание будет значительно меньшим, но движение второй точки (на довольно длительном этапе) уже равномерным не будет, а волнообразным (см. рис. 2.9).

На некоторых других интересных случаях реакции следящей системы на сигнал  $\epsilon(t)$  мы остановимся в последней главе учебника («Уроки физики»). Сейчас же ограничимся лишь этими простыми примерами и рассмотрим конкретные системы.

Начнем с радиолокатора. Для наблюдения за летящим самолетом радиолокатор должен «подсвечивать» самолет своим радиолучом. Направление этого луча указывает направление на цель и задает угловые координаты в сферической системе, в центре которой находится локатор, а время запаздывания отраженного сигнала по отношению к импульсу излучения позволяет определить расстояние. Ясно, что для непрерывного слежения за целью нужно все время поворачивать антенну так,



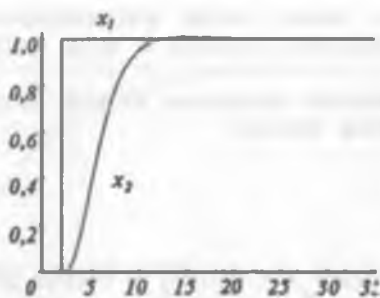


Рис. 2.4. Идеальная реакция ( $x_2$ ) следящей системы на входной ( $x_1$ ) скачкообразный сигнал

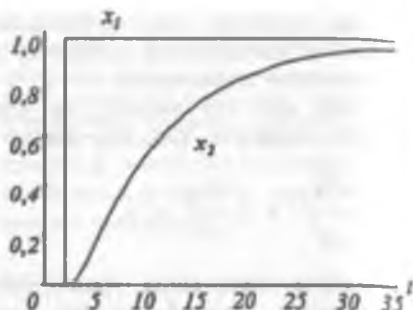


Рис. 2.5. Реакция следящей системы на входной сигнал при сильном торможении (большом коэффициенте вязкого трения)

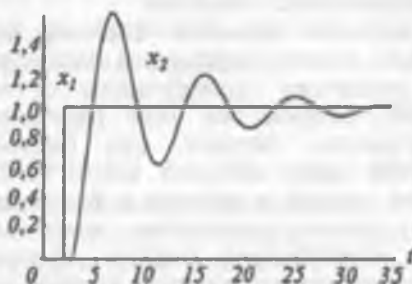


Рис. 2.6. Реакция следящей системы на входной скачкообразный сигнал при малом вязком трении и значительной массе второй точки

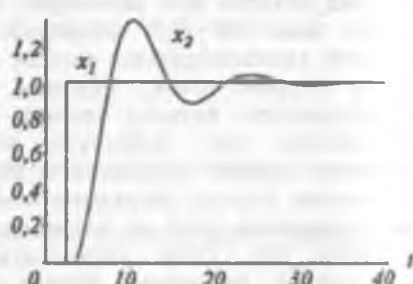


Рис. 2.7. Реакция следящей системы на входной скачкообразный сигнал при увеличении вязкого трения по сравнению с условиями рис. 2.6

чтобы ее ось проходила через цель. Если локаторный луч «сползет» с цели, то сила отраженного луча уменьшится по сравнению с максимальной. Соответствующая разность сигналов через усилитель воздействует на исполняющие двигатели антенны, которые срабатывают так, чтобы при повороте антенны отраженный сигнал снова возрос. Таким образом и происходит постоянное движение антенны вслед за целью.

Примерно так же действует подсолнух, поворачивая свою головку вслед за Солнцем. Разумеется, «двигатели» там другие.

А теперь мы обратимся к примеру, имеющему самое непосредственное отношение к любому из нас.

Каждый из читателей этой книги в детстве имел дело с куклами. Представительницы прекрасного пола постоянно с ними играли. Мальчики тоже знают, что это такое, и неоднократно держали их в руках. Всем известно поэтому, как трудно поставить куклу вертикально так, чтобы она не упала, особенно, если опора не строго горизонтальна.

Кукла представляет собой достаточно точное изображение реального человека. Почему же мы-то с Вами не падаем?

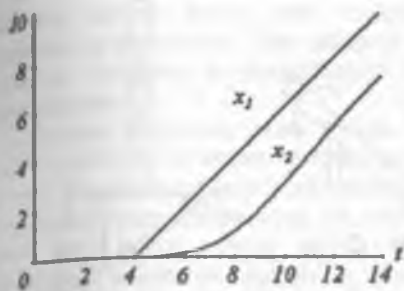


Рис. 2.8. Движение точки 2 вслед за точкой 1 при значительной вязкости

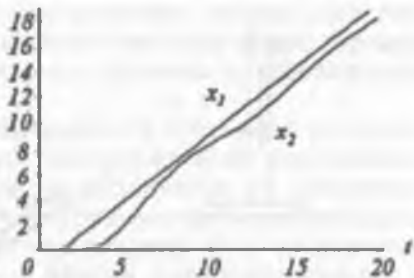


Рис. 2.9. Движение точки 2 вслед за точкой 1 при малом вязком трении и незначительной массе второй точки

Конечно, сразу же можно отметить, что человек—конструкция гибкая, подвижная, имеющая множество степеней свободы. Но представим себе солдата, стоящего на платформе по стойке «смирно». Он очень будет напоминать куклу. Однако если немного наклонить платформу, то кукла непременно упадет, а солдат—нет! Он все равно будет сохранять вертикальное положение, изменив автоматически угол между стопами и вертикалью (направлением действия силы тяжести).

Попробуем промоделировать ситуацию на примере очень простой системы, показанной на рис. 2.10.

Пусть на подставке с четырьмя ножками вертикально стоит длинный четырехгранный брусок с квадратным основанием. Подставка, в свою очередь, стоит на платформе, вначале расположенной горизонтально. Пусть теперь платформа наклонилась. Вместе с ней наклоняется и подставка. Отклонится от вертикального положения и брусок. Если он ничем к подставке не прикреплен, то ясно, что при достаточном наклоне брусок упадет. Не занимаясь здесь выкладками, укажем, что предельный угол наклона бруска по отношению к вертикали, при превышении которого он уже наверняка упадет, определяется отношением длины бруска к стороне квадрата его торца (основания). При отношении 10:1 этот угол будет составлять примерно  $6^\circ$ . Поэтому кукла и падает при небольших наклонах.

Усложним теперь конструкцию. Прикрепим к вертикально стоящему бруску отвес, то есть грузик на нити. При наклонах бруска между направлением отвеса (это направление всегда будет совпадать с направлением действия силы тяжести) и гранью бруска появится не равный нулю угол  $\varphi$  (см. рис. 2.10). Возникнет, таким образом, некоторый сигнал, который можно использовать в качестве управляющего. Представим себе, далее, что этот сигнал после надлежащего усиления и преобразования (например, в электрическое напряжение, подаваемое на клеммы электромотора) передается на исполнительный механизм. Пусть это действие сводится к тому, что соответствующие ножки подставки, на которой стоит брусок, удлиняются (ножки уже



Рис. 2.10. Иллюстрация действия простейшей следящей системы, обеспечивающей вертикальное положение свободно стоящего на подставке бруска

также не жесткие, а управляемые телескопические). За счет этого удлинения плоскость подставки начнет возвращаться к горизонтальному положению. Угол между направлением отвеса и гранью бруска будет уменьшаться, и брусок снова окажется в положении наибольшей устойчивости. Если дальше угол наклона платформы не меняется, то вся система перейдет из начального состояния (платформа горизонтальна, все ножки подставки имеют одинаковую длину, брусок стоит вертикально) к конечному (платформа наклонена, ножки подставки имеют разную длину, плоскость подставки снова горизонтальна и брусок сохраняет вертикальное положение).

Если предположить, далее, что платформа под подставкой все время то наклоняется, то выпрямляется, подобно палубе корабля на волнении в неспокойном море, то для удержания бруска потребуются непрерывная «работа» ножек подставки, согласованная с меняющимся управляющим сигналом. В целом вся конструкция представляет собой простой случай следящей системы.

Теперь вернемся к стоящему «смирно» солдату. Если приглядеться к нему, то легко заметить, что корпус солдата все время совершает небольшие движения около вертикального положения. Стоящая вертикально кукла таких движений не совершает: стоит себе, и все! Надо учесть, однако, что человек — не жесткая система. Вертикальное положение его корпуса удерживается напряжением соответствующих мышц. Какая-то из них может устать, что приведет к отклонению корпуса от положения вертикали. Так как само по себе вертикальное положение корпуса человека не устойчиво, то человек начал бы падать (как падает кукла, если ее немного наклонить). Однако падения человека не происходит именно потому, что немедленно приходит в действие некоторая управляющая система. Она называется вестибулярным аппаратом. Действие ее основано на преобразовании различных сигналов от определенных органов, реагирующих на отклонение корпуса от вертикального положения. Здесь мы не будем заниматься вопросами физиологии человека. Нам

важно только знать, что такая управляющая система действительно существует. Она выдает сигналы тем или иным группам мышц, которые и возвращают корпус человека в вертикальное положение.

Таким образом, если кукла находится в состоянии статического (хотя и не очень устойчивого) равновесия и на поддержание этого равновесия никакой энергии не расходуется, то равновесие человека поддерживается за счет постоянного функционирования тех или иных групп мышц. Даже, казалось бы, неподвижно стоящий «смирно» солдат в конце концов устанет и упадет, а кукла может стоять сколько угодно.

Теперь обратим внимание (см. рис. 2.4) на то, что следящая система не может сразу же перейти в новое состояние: всегда имеет место запаздывание, которое зависит от параметров системы. Значит, если внешняя ситуация изменится быстро, то система может не успеть сработать. Именно поэтому поскользнувшийся на льду человек падает, совершая при этом ряд судорожных движений: следящая система пытается отреагировать на изменение точки опоры, но не успевает!

Из приведенных рисунков также видно, что если параметры системы не оптимальны, то система переходит в новое состояние через стадию затухания колебаний иногда значительной величины. Отсюда следует, что при любом нарушении нормального действия вестибулярного аппарата человеку будет гораздо труднее сохранять вертикальное положение. Это проявляется в виде усилившихся колебаний человека по отношению к вертикали. Интересно отметить, что этот факт используется в медицинской диагностике состояния нервной системы. Пациенту предлагают встать вертикально, соединить подошвы и вытянуть вперед руки. Затем закрыть глаза. Здоровый человек сохранит и при этих условиях свое положение относительно долго, хотя покачивания его около вертикали увеличатся. Больной же почти сразу начнет падать. Причина этого в том, что в качестве управляющих сигналов для поддержания равновесия организм человека использует как внутренние ощущения, так и внешние—зрительные. Отключение части соответствующих анализаторов приводит к разбалансировке следящей системы, и устойчивость человека ухудшается. Разумеется, не вполне здоровая нервная система реагирует на это гораздо сильнее. Это и позволяет обнаружить патологию.

Итак, даже сохранение вертикального положения корпуса человека или вообще любой неизменной позы при опоре на две ноги, не говоря уже о движениях, требует непрерывного сокращения и расслабления практически всех мышц человека. Аналогичное кукле статическое состояние человека, при котором большая часть мышечной системы бездействует, вообще невозможно (если только человек не лежит плашмя). В результате за свое прямохождение человек расплачивается очень большими

энергозатратами, составляющими значительную часть всех вообще необходимых для его жизнедеятельности.

Вот как далеко завело нас обсуждение действия очень простой механической модели и решение несложного уравнения движения! Мы еще вернемся к следящим системам в главе «Уроки физики».

## 2.3. Принцип детерминизма

Выше было показано, что в тех случаях, когда заданы уравнение движения и начальные условия, решив это уравнение при заданных начальных условиях, можно получить точную характеристику движения в любой будущей момент времени. В таких случаях говорят, что движение полностью детерминировано. Будущее событие оказывается предопределенным заданием начальных условий. Эта предопределенность и составляет содержание принципа детерминизма.

Представим на минуту, что все события вокруг нас совершаются, подчиняясь лишь чисто механическим законам и определенным силовым взаимодействиям. Если бы это было так, то все события будущего были бы определены некоторым исходным, начальным толчком и совершались бы с фатальной неизбежностью. В свое время подобные рассуждения и приписывание принципу детерминизма некоторого абсолютного смысла привело к появлению определенной философской концепции, которая утверждала, что все в мире уже фатально предопределено, нет никаких случайных явлений и дело только в том, что мы не всегда в состоянии разобраться в сложности получающейся картины из-за трудностей ее познания. Но при ближайшем рассмотрении сразу видно, что такие рассуждения страдают известной непоследовательностью. В самом деле, для того чтобы охарактеризовать какое-то движение и предсказать события в некоторый наперед заданный момент, необходимо задать начальные условия. Даже если считать, что уравнение движения записано правильно, все равно необходимо учитывать то важнейшее обстоятельство, что задание начальных условий, которое подразумевает производство соответствующих измерений, не может быть выполнено с абсолютной точностью. Всегда при любом измерении будет допущена определенная неконтролируемая ошибка. Даже если эта ошибка и невелика, а интервал времени, на котором мы рассматриваем движение, становится очень большим, предсказание может быть выполнено со значительной погрешностью. Заметим, что поведение объектов в нашем мире в конечном счете определяется законами не классической, а более сложной (квантовой) механики. Эти законы описывают движение уже не макроскопических, а микроскопических объектов. Согласно же основным законам квантовой механики, можно указать положение и скорость движения макрочастицы только с определенной вероятностью.

Так как макроскопический мир состоит из микроскопических объектов, то такое поведение в микромире приводит в конечном счете к нарушению принципа полной определенности, или принципа детерминизма, во всем окружающем нас мире.

Это не означает, однако, что от принципа детерминизма нужно отказываться совсем. В самом деле, именно уверенность в правильности этого принципа в определенных условиях и позволяла утверждать, что направленная в сторону кометы Галлея земная астрономическая станция, движение которой рассчитывается в основном с помощью решения рассмотренных классических уравнений движения, действительно достигнет в ожидаемое время нужного района. Важный вопрос о научном прогнозе и его ограничении будет подробно рассмотрен в гл. 11.

## 2.4. Третий закон Ньютона, импульс

Перейдем теперь к обсуждению третьего закона Ньютона. Как мы знаем из школьного курса, этот закон утверждает, что *если имеются два тела, между которыми есть какое-то взаимодействие, то со стороны первого тела на второе будет действовать некоторая сила  $F_{12}$ , а со стороны второго тела на первое—сила  $F_{21}$ . Эти силы приложены соответственно ко второму и к первому телам, равны по модулю и противоположны по направлению*. Действие этого закона обнаруживается постоянно. Например, если попытаться привести в движение какой-нибудь тяжелый предмет, то нужно затрачивать определенную силу: этот предмет как бы отталкивает нас. Если мы при входе в здание открываем на себя тяжелую дверь, то явно чувствуем, как эта дверь начинает тянуть нас к себе.

Представим две шайбы, лежащие на гладком столе, между которыми находится сжатая пружина (рис. 2.11). Если пружине дать возможность распрямиться, то шайбы придут в движение со скоростями, направленными в противоположные стороны. Это, в сущности, и будет означать, что и к одной и к другой шайбе были приложены силы со стороны пружины. Если  $F_{12} = -F_{21}$  и массы шайб соответственно  $m_1$  и  $m_2$ , то

$$m_1 \frac{dv_1}{dt} = -m_2 \frac{dv_2}{dt}, \quad \text{или} \quad m_1 \frac{dv_1}{dt} + m_2 \frac{dv_2}{dt} = 0.$$

Введем теперь в рассмотрение величину  $P$ , равную произведению массы тела на его скорость и названную *импульсом*. Импульс также, очевидно, есть векторная величина, по направлению совпадающая с вектором скорости. Пользуясь понятием импульса, запишем

$$m_1 \frac{dv_1}{dt} + m_2 \frac{dv_2}{dt} = \frac{d}{dt}(mv_1) + \frac{d}{dt}(mv_2) = \frac{d}{dt}(p_1 + p_2) = 0,$$

где  $p_1 + p_2$ —полный импульс пары материальных точек.



Рис. 2.11. Шайбы лежат на столе и при освобождении пружины начинают двигаться без трения в разные стороны

Таким образом, производная полного импульса двух взаимодействующих материальных точек равна нулю. Следовательно, вектор полного импульса точек должен быть постоянным. Это утверждение названо *законом сохранения импульса*. Этот закон в данном случае получен как непосредственное следствие второго и третьего законов Ньютона.

Если рассмотреть не две точки, а большее их число и считать, что между всеми этими точками существуют силы взаимодействия, то для каждой пары из совокупности должен выполняться третий закон Ньютона. Распространяя рассуждения на всю совокупность точек, приходим к заключению, что полный импульс такой системы материальных точек, равный векторной сумме импульсов отдельных составляющих, также должен быть постоянным. Выбор этой постоянной величины диктуется опять-таки начальными условиями. Если в начальный момент все точки системы покоились, то в случае свободного движения точек под действием внутренних сил они будут двигаться так, что их полный импульс по-прежнему будет равен нулю.

Поскольку в простых механических явлениях всякая сила может проявиться только в результате взаимодействия между какими-нибудь телами, то закон о сохранении полного импульса системы, в сущности, означает, что невозможно создать движение одного тела, не вызвав одновременно противоположно направленного движения другого тела, с которым данное тело находится во взаимодействии. Пусть, например, из заряженной пушки, ствол которой имеет массу  $M$ , а заряд массу  $m$ , произведен выстрел. Так как в начальный момент снаряд и ствол покоились, то после выстрела импульс системы будет равен  $MV - mv = 0$ , где  $V$ —скорость ствола и  $v$ —скорость снаряда (минус ставится из-за того, что снаряд и ствол орудия движутся в противоположных направлениях). Следовательно, не только возникнут движения двух объектов, но их относительные скорости будут определяться отношением масс.

Третий закон Ньютона объясняет появление любого ускоренного движения в природе. Если какое-то тело изменило состояние покоя или равномерного движения, то это значит, что оно обязательно взаимодействует с каким-то другим телом. При этом и другое тело должно изменить свое состояние. Иногда это изменение состояния движения другого тела столь незначительно, что им можно пренебречь, но из этого не следует, что его в принципе нет. Например, когда человек идет по земле, то он отталкивается от земли ногами. При этом Земля должна вращаться в сторону, противоположную движению человека, но

так как масса Земли неизмеримо больше, чем масса человека, то смещение ее будет ничтожно и им практически можно пренебречь.

Очень важный пример проявления взаимодействия тел представляет реактивное движение. В этом случае происходит выбрасывание из корпуса, например ракеты, струи газа, обладающей определенной массой. При этом ракета начинает двигаться в противоположном направлении. Отношение скоростей ракеты и газа должно быть величиной, пропорциональной обратному отношению масс. Поэтому если ракете нужно развить большую скорость, выбрасывая в единицу времени сравнительно небольшую массу газа, то необходимо выбрасывать эту массу с очень большой скоростью. Скорость истечения газа есть величина, связанная с его температурой. Чем выше температура газа, тем выше скорость его истечения. Для того чтобы получить при том же расходе топлива большую скорость ракеты, необходимо повышать температуру отброшенного газа. Ограничения здесь ставят чисто технические трудности, связанные с тем, что применяемые в ракетостроении материалы не могут выдерживать слишком высокой температуры.

Можно привести еще большое число примеров, демонстрирующих выполнимость третьего закона Ньютона, но мы обратимся к другому чрезвычайно важному моменту, а именно, введем понятие поля.

## 2.5. Поле

При изучении взаимодействий тел прежде всего видно, что эти взаимодействия можно разделить на две группы.

К первой группе относятся взаимодействия, которые происходят при непосредственном контакте тел. Такими взаимодействиями являются, например, те, которые возникают в результате столкновения двух шаров друг с другом, а также в результате непосредственного приложения, например, мускульной силы человека к тем или иным предметам и т. д. Ко второй группе взаимодействий отнесем такие, которые возникают в тех случаях, когда непосредственного контакта двух тел нет. К ним относятся гравитационные взаимодействия, выражающиеся в притяжении двух тел, обладающих определенными массами. Действие на расстоянии возникает и в случае электромагнитных взаимодействий, в частности, в простейших взаимодействиях, описываемых законом Кулона. Для первого типа взаимодействий третий закон Ньютона выполняется, как показывают эксперименты, с очень высокой степенью точности, и никаких дополнительных вопросов не возникает. Далеко не так просто обстоит дело, если тела взаимодействуют на расстоянии (например, кулоновские, или электростатические взаимодействия).

Представим следующий эксперимент. Пусть два совершенно одинаковых металлических шарика, расположенных на рассто-



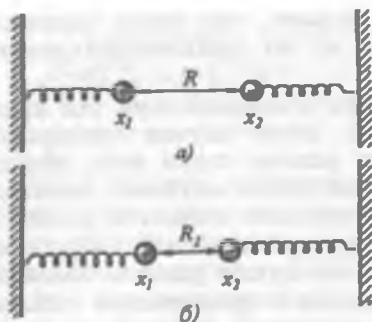


Рис. 2.12. К введению понятия поля

янии  $R$  друг относительно друга, прикреплены пружинами к неподвижным стенкам (рис. 2.12). Предположим, что одновременно эти шарики очень медленно начинают заряжать одним и тем же, но разноименным по знаку количеством электричества, присоединяя к полюсам электрофорной машины. Как только шарики начнут приобретать хотя бы очень малые электрические заряды, то по закону Кулона они начнут притягиваться друг к другу, растягивая удерживающие их пружины. Предположим, что шарики зарядили до величин, условно принятых за единицу, после чего дальнейшее зарядение прекратится. Шарики сблизятся на расстояние  $R_1$ , меньшее начального  $R$ , после чего вся

система будет находиться в состоянии равновесия. При этом пружины растянутся, если они совершенно одинаковы, на одинаковую величину. Значит, появившееся электростатическое взаимодействие привело к появлению силы, действующей со стороны шарика 1 на шарик 2, а со стороны шарика 2 на шарик 1, причем эти силы оказываются равными по модулю и противоположными по знаку, как это и предписывается третьим законом Ньютона. Представим теперь, что мы, не меняя заряда правого шарика, увеличиваем вдвое также очень медленно заряд левого шарика, сохраняя знак этого заряда. Сила кулоновского взаимодействия при этом должна возрасти, что приведет к дальнейшему сближению двух шариков и соответствующему растяжению пружины. После того как зарядение левого шарика прекратится, картина снова станет статической. При этом опять-таки растяжение пружин станет одинаковым, хотя заряды шариков различны. Снова мы получаем результат, полностью соответствующий третьему закону Ньютона. Видоизменим теперь эксперимент и предположим, что меняется заряд левого шарика вдвое скачком, причем очень резким скачком, за весьма малый промежуток времени. Согласно закону Кулона в тот момент, когда левый шарик приобрел удвоенный заряд, кулоновская сила взаимодействия между зарядами должна скачком возрасти вдвое. Это равносильно внезапному приложению сил к левому и правому шарикам. В результате система будет выведена из положения равновесия и оба шарика начнут совершать колебательные движения. Если в системе присутствует трение, то эти

колебания рано или поздно затухнут и система перейдет в состояние равновесия. Появление колебаний обоих шариков, когда меняется заряд только одного из них, опять-таки указывает на существование такого взаимодействия, когда первый шарик действует на второй, а второй шарик на первый. В результате они и приходят оба в состояния колебательных движений, причем если шарики одинаковые и пружины тоже одинаковые, эти колебательные движения будут подобными. Но остается открытым вопрос: а будут ли оба шарика приходить в колебательное движение абсолютно в одно и то же время или же в несколько разные моменты? Этот вопрос равносильен другому: «ощутит» ли правый шарик мгновенное изменение заряда левого шарика сразу же, как это изменение произошло, или спустя некоторое время? Можно задать еще один вопрос. Представим, что правого шарика вообще нет. Если теперь зарядить левый шарик удвоенным зарядом, то его состояние, конечно, не изменится. Стало быть, сам по себе наведенный заряд ни к какому движению не приведет. Появление колебаний всей системы и, в частности, колебания левого шарика связано с тем, что имеется и правый шарик. Таким образом, происходит как бы следующее: в левом шарике изменился заряд, это изменение заряда «почувствовал» правый шарик, на него начала действовать большая сила, правый шарик, естественно, должен прийти в движение. Однако отсюда еще никак не следует, что одновременно в движение должен приходить и левый шарик. Его движение, в свою очередь, можно объяснить только тем, что он каким-то образом «почувствовал», что начал двигаться правый шарик. Происходит как бы передача сигнала от левого шарика к правому и от правого шарика к левому. В разделе «Кинематика» мы отметили, что никакие влияния в природе одних тел на другие, выражающиеся либо в непосредственном переносе каких-то масс, либо в какой-то другой форме, не могут происходить мгновенно, а совершаются максимально со скоростью света.

Таким образом, информация об изменении состояния заряда первого шарика и информация о том, что другой шарик пришел в движение, будет передаваться от одного шарика к другому шарiku не мгновенно, а с некоторым, пусть малым, но все-таки запаздыванием. Если бы эксперимент провести с очень высокой степенью точности, то мы бы обнаружили, что колебания одного шарика происходят так, что они как бы сдвинуты, т. е. отличаются по фазе от колебаний другого шарика (рис. 2.13). В формулировке третьего закона Ньютона это очень важное обстоятельство никак не отражено. В тех случаях, когда расстояния между объектами не очень велики, таким запаздыванием можно пренебречь и считать, что третий закон Ньютона всегда верный. Есть, однако, такие явления, в которых даже при наличии сравнительно малых расстояний эффекты запаздывания оказываются уже заметными. Это относится, в частности, к электромагнитным явлениям (см. гл. 7).

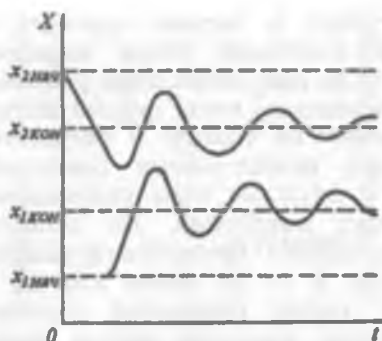


Рис. 2.13.  $x_{1нач}$ ,  $x_{2нач}$  и  $x_{1кон}$ ,  $x_{2кон}$  — начальные и конечные расположения шариков вдоль оси  $X$ . Кривые характеризуют затухающие колебательные движения шариков

Рассмотрим еще один эксперимент, на этот раз механический. Пусть имеются три тела массами  $M_1$ ,  $M_2$  и  $m$ . При этом  $M_1$  и  $M_2$  существенно больше  $m$ . Представим, что в начальный момент массы  $M_1$  и  $m$  соединены сжатой пружиной так, как показано на рис. 2.14. Пусть пружине дали возможность резко распрямиться. В этом случае масса  $M_1$  приобретает скорость  $V_1$ , равную  $V_1 = \frac{m}{M_1} v$ , где  $v$  — скорость меньшего тела; скорости

направлены в противоположные стороны. После того как действие пружины прекратится, каждое из этих тел начнет двигаться равномерно и прямолинейно.

Предположим, что опыт поставлен таким образом, что в своем дальнейшем движении тело массой  $m$  встречает тело массой  $M_2$  и при этом тела слипаются между собой так, что получается один объект. Воспользовавшись законом сохранения импульса, найдем, что такое объединенное тело приобретает некоторую скорость, направленную вправо (рис. 2.15), причем модуль этой скорости равен  $V_2 = mv/(m + M_2)$ . В результате мы обнаружим после всех этих событий, что тело  $M_1$  продолжает двигаться влево со скоростью  $V_1$ , а тело с объединенной массой  $M_2 + m$  движется в противоположную сторону со скоростью  $V_2$ . Отношение этих скоростей, т. е. скорости  $V_1$  к скорости  $V_2$ , равно отношению объединенной массы  $M_2 + m$  к массе  $M_1$ . Предположив, что масса  $m$  пренебрежимо мала по сравнению с массой  $M_2$ , получим, что отношения скоростей, которые в результате всего этого эксперимента приобретут тела с массами  $M_1$  и  $M_2$ , будут обратно пропорциональны отношению их масс. Результат будет таким, какого следует ожидать при мгновенном взаимодействии между телами  $M_1$  и  $M_2$  в соответствии с требованием закона сохранения импульса при взаимодействии пары точек. Масса  $m$  в этом эксперименте играет роль

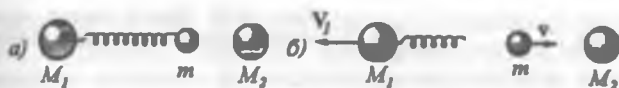


Рис. 2.14. Начальное (а) и промежуточное (б) состояния системы

переносчика взаимодействия и, как видно из всего рассмотрения, переносчика импульса.

Заметим, что в тот момент, когда левая масса пришла в состояние равномерного прямолинейного движения, правая масса еще не начала двигаться и находится в покое. Когда начинает двигаться масса  $M_2$ , масса  $M_1$  движется равномерно и прямолинейно и, стало быть, на нее не действуют никакие силы. Итак, если попытаться применить, забыв о переносчике импульса, третий закон Ньютона сразу к первому и второму телам, то надо еще предположить, что эти силы действуют на тела в разное время, что третьим законом Ньютона не предусматривается.

Аналогично можно объяснить и эффект взаимодействия между телами, находящимися на некотором расстоянии друг относительно друга, если ввести в рассмотрение некоторую материальную среду, которая окружает данные тела и которая, воспринимая импульс от первого тела, передает его затем второму телу. Такая *материальная среда* получила название *поля*.

Используя понятие поля, мы рассматриваем электромагнитные взаимодействия между удаленными телами следующим образом. При движении первого тела или изменении его заряда возникает некоторый импульс, который передается локально непосредственно полю, расположенному около первого тела. Воспринимают этот импульс кванты поля—частицы, которые получили название *фотонов*. Распространяясь в пространстве, фотоны достигают второго тела и передают ему некоторый импульс. В результате и наблюдается сдвиг второго тела. Но второе тело не только воспринимает импульс, но, в свою очередь, изменяя свое состояние, порождает изменение поля в пространстве, непосредственно к нему примыкающему. Возникает как бы обратное движение фотонов, обмен фотонами между двумя телами; в результате такого обмена и происходит наблюдаемое макроскопическое взаимодействие.

В случае ядерных сил в качестве таких носителей импульса выступают уже не фотоны, а  $\pi$ -мезоны.

Такое представление о взаимодействии тел названо *принципом близкого действия*: мы исключаем возможность взаимодействия удаленных тел через ничто или мгновенные взаимодействия на расстоянии. Взаимодействия происходят всегда с некоторым полем, которое прежде всего начинает меняться в области, непосредственно прилежащей к первому телу, а затем уже это изменение передается полем дальше и достигает других тел.

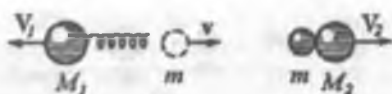


Рис. 2.15. Малый шарик долетает до правого большого и «прилипает» к нему



Поскольку в результате таких взаимодействий должен переноситься импульс, правильный закон сохранения импульса должен записываться в следующей форме:  $\sum_k p_k + p_{\text{поле}} = \text{const}$ ,

или *векторная сумма импульсов всех тел и импульса поля есть величина постоянная*. Такое утверждение представляет собой современную формулировку закона сохранения импульса. При этом следует помнить, что закон сохранения импульса относится к замкнутым системам, т. е. к таким системам, которые не взаимодействуют ни с какими окружающими телами или полями. В такой формулировке этот закон уже нельзя получить как следствие второго и третьего законов Ньютона, и он выступает как некое новое утверждение. В свою очередь, третий закон Ньютона может быть получен как следствие общей формулировки закона сохранения импульса в предположении существования непосредственного контакта тел, малой роли поля и практически мгновенного распространения действия одного тела на другое.

Таким образом, закон сохранения импульса оказывается значительно более общим и не сводимым к третьему закону Ньютона. Действительно, современная физика выделяет закон сохранения импульса как один из основных законов природы. Он выполняется как в случае больших скоростей, так и в случае движения микрочастиц, которые вообще в своем движении не подчиняются классическим законам, а подчиняются квантовым законам: в частности, в микромире понятие импульса, хотя и в несколько видоизмененной форме, по-прежнему имеет место. Так, от легко наблюдаемого и проверяемого в простых механических экспериментах третьего закона Ньютона мы пришли к формулировке нового закона, предположив дополнительно, что всякое действие в природе, всякое влияние распространяются с конечной скоростью. Как мы увидели, для этого нам надо было ввести новую физическую характеристику, называемую полем, и приписать полям способность переносить импульс, который и воспринимают кванты или частицы поля. Конечно, в зависимости от поля или от типа взаимодействий физический смысл таких квантов оказывается различным.

Здесь интересно отметить следующее. В рамках нашего рассмотрения, вообще говоря, нет существенной разницы между электрическими (или электромагнитными) взаимодействиями и гравитационными. И если некоторая логика приводит нас к необходимости введения понятия фотонов, то та же самая логика должна была бы привести нас к понятию квантов гравитационного поля, которые получили название *гравитонов*. Подобно электромагнитным волнам должны были бы существовать и

гравитационные волны. Поиск этих гравитационных волн ведется физиками в разных странах, однако, несмотря на постоянно возрастающую точность физического эксперимента, пока существование гравитационных волн экспериментально не подтверждено.

Введение понятия близкодействия и поля как передатчика близкодействующего влияния на далекие расстояния, фотонов или других частиц, образующих это поле и переносящих импульс, формулировка в общем виде закона сохранения импульса как одного из основных законов природы приводят к необходимости переформулировать и второй закон Ньютона, и уравнение движения.

Вместо выражения  $m \frac{dv}{dt} = f$  будем писать  $\frac{dp}{dt} = f$ . В такой формулировке естественно называть силой скорость изменения импульса тела. Импульс тела передается полю, и поле приобретает импульс со скоростью, численно равной силе. В частном случае для тела при  $m = \text{const}$  новое выражение переходит в прежнее.

Сила, таким образом, приобретает некоторый обобщенный смысл. Ведь, в самом деле, можно говорить о том, что тело передает свой импульс полю, в частности фотону электромагнитного поля, но нельзя говорить о том, что к фотону приложена некоторая сила в ее обычном ньютоновском понимании.

## 2.6. Кинетическая энергия материальной точки

Введем понятие кинетической энергии материальной точки. Кинетической энергией точки, как известно из школьного курса, называется скалярная величина, равная

$$T = \frac{1}{2} m |v|^2 = \frac{1}{2m} |p|^2.$$

Если воспользоваться разложением скорости по осям координат, то формула для кинетической энергии примет вид

$$T = \frac{1}{2} m (v_x^2 + v_y^2 + v_z^2).$$

Найдем частную производную от  $T$  по одной из компонент скорости, например  $v_x$ . Напомним, что частная производная от функции многих переменных (в данном случае  $T$  является функцией трех аргументов— $v_x$ ,  $v_y$ ,  $v_z$ ) находится так же, как и обычная производная, но те аргументы, по которым дифференцирование не производится, считают постоянными.

В соответствии с этим частная производная  $\partial T / \partial v_x = m v_x$ . Но это есть не что иное, как  $x$ -составляющая импульса точки.

Следовательно, должны иметь место такие важные соотношения:

$$\frac{\partial T}{\partial v_x} = p_x, \quad \frac{\partial T}{\partial v_y} = p_y, \quad \frac{\partial T}{\partial v_z} = p_z.$$

Подобные соотношения между производными от кинетической энергии по составляющим скорости справедливы и в других системах координат, а не только в декартовой. Это позволяет пользоваться понятием импульса (его тогда называют *обобщенным импульсом*) для любых известных в физике систем координат.

## 2.7. Движение точки по окружности

Выше в § 1.3 было показано, что при движении точки по окружности удобно ввести угловую скорость как вектор, направленный вдоль оси вращения, модуль которого пропорционален числу оборотов точки в единицу времени. Выбор направления вектора угловой скорости позволяет указать, по часовой стрелке или против нее совершается движение точки по окружности. При равномерном движении точки по окружности вектор угловой скорости все время остается постоянным.

Покажем, как можно записать уравнение движения точки по окружности, пользуясь не линейной, а угловой скоростью и полярной системой координат, чтобы математическая форма такого уравнения была аналогична ранее рассмотренной. Для этого предположим, что точка описывает окружность радиуса  $r$ . Если ввести угловую скорость  $\omega$ , то модуль линейной скорости  $|v| = r|\omega|$ . Подставляя это выражение в формулу для кинетической энергии, получаем

$$T = \frac{1}{2} m |v|^2 = \frac{1}{2} m r^2 |\omega|^2.$$

Величина  $mr^2$  называется *моментом инерции точки относительно оси вращения*. Обозначим эту величину символом  $I$ . Тогда кинетическая энергия точки

$$T = \frac{1}{2} I |\omega|^2 = \frac{1}{2} I (\omega_x^2 + \omega_y^2 + \omega_z^2),$$

где  $\omega_x$ ,  $\omega_y$  и  $\omega_z$  — проекции вектора угловой скорости  $\omega$  на декартовы оси координат.

Компоненты обобщенного импульса (момента импульса), соответствующего скорости  $\omega$ , введем аналогично компонентам импульса:

$$p_{\omega_x} = \frac{\partial T}{\partial \omega_x}, \quad p_{\omega_y} = \frac{\partial T}{\partial \omega_y}, \quad p_{\omega_z} = \frac{\partial T}{\partial \omega_z}.$$

Отсюда следует, что в качестве *вектора момента импульса* надо принять вектор  $P = I\omega$  при вращении материальной точки

вокруг оси. Математическая формула для момента импульса получается аналогичной формуле импульса с заменой массы на момент инерции и линейной скорости на угловую.

Найдем теперь форму записи уравнения движения точки по окружности. В разделе «Кинематика» было показано, что при движении точки по окружности вокруг фиксированной оси возникают две составляющие полного вектора линейного ускорения: нормальное ускорение и тангенциальное. Нормальное ускорение отвечает силе, удерживающей точку на окружности, а тангенциальное—силе, которая разгоняет точку вдоль окружности.

Для нормальной составляющей полной силы, действующей на точку, получаем  $|f_n| = m|a_n| = mr|\omega|^2$ . Видно, что ее модуль резко возрастает с увеличением угловой скорости и радиуса вращения.

Обратимся теперь к тангенциальной составляющей полной силы, действующей на точку, движущуюся ускоренно по окружности радиусом  $r$ . Согласно второму закону Ньютона  $f_t = ma_t$ . Наша задача заключается в том, чтобы получить аналог уравнению движения Ньютона, но так, чтобы в него входило не обычное, а угловое ускорение. Для этого надо записать уравнение Ньютона в форме  $K\alpha = N$ , где  $K$ —пока не известный скалярный коэффициент;  $\alpha$ —вектор углового ускорения ( $\alpha = d\omega/dt$ );  $N$ —вектор, параллельный  $\alpha$  и играющий роль силы при описании вращения точки по окружности.

Для решения этой задачи заметим (см. § 1.4), что вектор  $a_t$  направлен по касательной к описываемой точкой окружности и равен  $a_t = \alpha \times r$ . Если направление вектора угловой скорости (оси вращения) не меняется, а меняется лишь число оборотов точки в единицу времени, то модуль тангенциального ускорения равен  $|a_t| = |\alpha| |r|$ . Если вектор  $a_t$  векторно умножить на радиус-вектор точки, т. е. образовать вектор  $r \times a_t$ , то такой вектор будет направлен, очевидно, так же, как и вектор углового ускорения  $\alpha$ . По модулю это векторное произведение равно  $|r \times a_t| = |r| |a_t| = |r|^2 |\alpha|$ .

Значит, можно, возвращаясь к векторной форме, записать

$$r \times a_t = |r|^2 \alpha.$$

Учитывая этот результат, умножим равенство  $f_t = ma_t$  векторно на  $r$ :

$$r \times f_t = m |r|^2 \alpha.$$

Вектор  $r \times f_t$  при движении точки по окружности вокруг неизменной оси должен совпадать по направлению с вектором



углового ускорения  $\alpha$  (ведь  $f, \parallel a_t$ ). Такой вектор называется *моментом силы* и обозначается  $N$ . Введение понятия момента силы тесно связано с механикой рычагов. В частности, для равновесия рычага необходимо, чтобы векторная сумма моментов сил, действующих на плечи рычага, была равна нулю. Выражение, связывающее векторы углового ускорения и момента силы, позволяет определить и скаляр  $K$ , равный моменту инерции точки относительно оси вращения. Таким образом, окончательно получаем искомое уравнение движения материальной точки вокруг фиксированной оси в форме

$$I \frac{d\omega}{dt} = N \quad \text{или} \quad \frac{dP_\omega}{dt} = N.$$

Эти рассуждения позволяют заключить, что при описании движения точки по окружности существует аналогия с прямолинейным движением, если произвести в соответствующем уравнении движения следующие замены: вместо массы ввести момент инерции, вместо ускорения—угловое ускорение и вместо силы—момент силы относительно оси вращения.

Для того чтобы выяснить, как будет вращаться точка через некоторый промежуток времени, необходимо решить это уравнение с учетом начальных условий.

Мы увидим, в частности, что если  $N = 0$ , то движение по окружности будет совершаться с постоянной угловой скоростью. При постоянном моменте силы получим равноускоренное или равнозамедленное движение точки по окружности.

## 2.8. Работа

Понятие работы в физике является обобщением повседневного опыта человека.

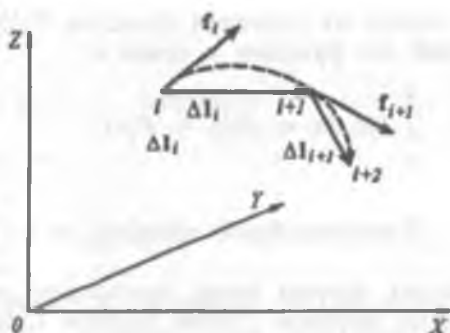
Под работой некоторой постоянной силы на прямолинейном участке пути понимается скалярная величина, равная произведению модуля силы на длину пути и на косинус угла между направлением движения и направлением вектора силы. Если воспользоваться правилом построения скалярного произведения двух векторов, то

$$\Delta A = |f| |\Delta l| \cos \varphi = f \Delta l,$$

где  $f$ —вектор действующей силы;  $\Delta l$ —вектор перемещения (пути);  $\varphi$ —угол между векторами  $f$  и  $\Delta l$ .

Рассмотрим выражение для работы переменной силы на участке пути произвольной формы. Любую кривую всегда с достаточной хорошей степенью точности можно заменить ломаной. Будем строить эту ломаную таким образом, чтобы, во-первых, ее прямолинейные участки были малы настолько, чтобы путь по ломаной линии практически не отличался от пути по кривой, и, во-вторых, на каждом таком прямолинейном участке ломаной

Рис. 2.16. К выводу выражения для работы



действующую на этом участке пути силу можно было бы считать неизменной.

Работа переменной силы на всем пути от начальной точки  $a$  до конечной точки  $b$ , очевидно, равна сумме элементарных работ силы на каждом участке ломаной. Обозначим  $\Delta l_i$  элементарный вектор ломаной линии, проведенный от точки  $i$  кривой в точку  $i + 1$  (рис. 2.16), а  $f_i$ —действующую на этом участке силу. Тогда работа

$$A = \sum_i f_i \Delta l_i$$

Если уменьшать и далее участки ломаной, то  $\Delta l_i$  можно заменить дифференциалом  $dl_i$ , а операцию суммирования—операцией вычисления определенного интеграла. Тогда

$$A = \int_a^b f dl.$$

Такой интеграл называется *криволинейным интегралом* вдоль определенной линии в пространстве (траектории движения) от начальной точки  $a$  до конечной  $b$ .

Воспользовавшись представлением скалярного произведения двух векторов через их координаты, можно записать

$$A = \int_a^b (f_x dx + f_y dy + f_z dz).$$

В простейшем случае

$$f_x = f_x(x), \quad f_y = f_y(y), \quad f_z = f_z(z).$$

Заметим, что для вычисления определенного интеграла от функции  $f(x)$ , т. е. величины  $\int_a^b f(x) dx$ , надо найти первообразную  $F(x)$  функции  $f(x)$ , т. е. найти неопределенный интеграл,

а затем из значения функции  $F(b)$  в точке  $b$  вычесть значение этой же функции в точке  $a$ :

$$\int_a^b f(x) dx = F(b) - F(a).$$

Криволинейный интеграл  $A = \int_a^b f dl$  можно заменить интегралом другого вида, вычисление которого оказывается гораздо более простым. Чтобы перейти к такой новой форме выражения для работы, заметим, что если тело покоится, то, очевидно,  $A = 0$ . Значит,  $A \neq 0$  только в том случае, когда тело перемещается. Конечно, надо исключить и случай свободного ( $f = 0$ ), прямолинейного, равномерного движения. Если тело движется, то  $dl = v dt = \frac{dr}{dt} dt$ , где  $v = \frac{dr}{dt}$  — скорость движения. Поэтому можно записать

$$A = \int_a^b f dl = \int_1^2 f v dt = \int_1^2 \left( f_x \frac{dx}{dt} + f_y \frac{dy}{dt} + f_z \frac{dz}{dt} \right) dt.$$

Здесь 1 и 2 — моменты времени, отвечающие началу и концу движения. Интегрирование ведется уже только по одной переменной (времени). При интегрировании надо принять во внимание, что сила  $f$  является функцией координат  $x, y, z$  точки, а они, в свою очередь, при перемещении точки зависят от времени.

Работа силы может быть положительной и отрицательной. Если действующая сила направлена в сторону движения, то угол  $\varphi$  меньше  $90^\circ$  и, следовательно,  $\cos \varphi > 0$ . Работа в этом случае положительна. Если же сила направлена в сторону, противоположную движению, то угол  $\varphi$  больше  $90^\circ$  ( $\cos \varphi < 0$ ) и работа будет отрицательной.

Физический смысл знака работы очень простой. Если сила составляет острый угол с перемещением, то она способствует движению. В этом случае сила совершает положительную работу. Если сила составляет с перемещением тупой угол (больше  $90^\circ$ ), то она тормозит движение и работа будет отрицательной. Если первое условие выполняется на всех участках криволинейного пути, то полная работа силы будет положительной. Работа будет отрицательной, если на всем пути сила тормозила движение. Такова, например, работа сил трения. Если часть пути сила тормозила движение, а на другой части способствовала ему, то в зависимости от ситуации полная работа может оказаться как положительной, так и отрицательной и равной нулю.

Предположим, что сила  $F$  представляется в виде суммы нескольких сил:  $F = f_1 + f_2 + f_3$ . По-прежнему работа силы  $F$  будет определяться выражением

$$A = \sum_i F_i \Delta l_i$$

Подставляя сюда  $F = f_1 + f_2 + f_3$ , получаем

$$\begin{aligned} A &= \sum_i (f_{1i} + f_{2i} + f_{3i}) \Delta l_i = \\ &= \sum_i f_{1i} \Delta l_i + \sum_i f_{2i} \Delta l_i + \sum_i f_{3i} \Delta l_i = A_1 + A_2 + A_3. \end{aligned}$$

Обратим внимание на различие в сложении сил и работ. Сила есть векторная величина, поэтому если надо найти действие совокупности сил, то складывают силы векторно. Работа—скалярная величина, поэтому если существует несколько действующих сил, то общая работа равна алгебраической сумме работ.

Здесь необходимо сделать пояснение. Пусть  $F = f_1 + f_2 = 0$ . Это значит, что  $f_1 = -f_2$  и  $A_1 = -A_2$ . В этом случае имеет смысл обсуждать либо работу силы  $f_1$  ( $A_1$ ), либо силы  $f_2$  ( $A_2$ ). Например, если тело под действием некоторой силы  $f$  движется равномерно и прямолинейно, преодолевая противоположное действие постоянной силы трения, то обычно исследуют только работу силы  $f$  против сил трения.

В первом примере рассмотрим работу по подъему массы  $m$  на высоту  $h$  с постоянной скоростью. В этом случае действующая сила должна быть по модулю равна силе тяжести  $mg = \text{const}$ . Считаем, что движение совершается вдоль оси  $Z$ . Тогда

$$A = \int_0^h mg dz = mgh.$$

Во втором примере вычислим работу, которую нужно совершить, чтобы сжать или растянуть пружину на величину  $x$  по сравнению с ее исходной длиной. Действующая сила должна преодолевать сопротивление пружины, которое по закону Гука должно быть пропорциональным отклонению от свободного состояния (когда пружина не деформирована). Значит,  $f = kx$ , где  $k$ —некоторый постоянный положительный коэффициент (коэффициент упругости). Подсчитаем, какая работа в этом случае совершается. Предположим, что пружина растягивается от нуля до какой-то величины  $l$ . Тогда работа

$$A = \int_0^l f dx = k \int_0^l x dx = \frac{1}{2} kl^2.$$

В общем случае работа связана с тем путем, по которому тело двигалось под действием силы. Это имеет место, например, при работе против сил трения. Существуют, однако, такие силы,

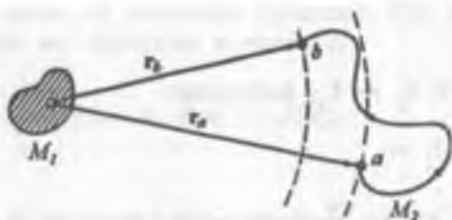


Рис. 2.17. Работа по перемещению тела  $M_2$  в поле сил тяготения тела  $M_1$  зависит только от изменения длин радиус-векторов  $r_a$  и  $r_b$ , но не от формы траектории

работа которых не зависит от пути. Их называют *потенциальными*. Ими, например, являются силы всемирного тяготения. Напомним, что всегда сила тяготения будет направлена по прямой, соединяющей центры двух тел. Если одно тело совершает движение по окружности постоянного радиуса вокруг другого, то всегда сила тяготения будет перпендикулярна касательной к окружности. Это означает, что работа сил тяготения при перемещении по такой траектории равна нулю. В результате получаем интересное свойство. Представим, что тело 2 перемещается из точки  $a$  в точку  $b$  по произвольной кривой (рис. 2.17). Перемещение тела 2 на любом участке траектории  $ab$  может быть представлено суммой перемещений по окружности и по радиусу. Перемещение по окружности не связано с работой силы тяготения. Она возникает при движении по радиусу. Таким образом, общая работа на всем участке пути зависит только от изменения расстояния тела 2 от тела 1 при переходе второго тела из точки  $a$  в точку  $b$ , но не от формы траектории, по которой движение происходило.

Такого типа силы играют в физике большую роль. Как уже говорилось, их называют потенциальными, а силовое поле, в котором они действуют,—потенциальным. Все известные фундаментальные силы или взаимодействия, о чем мы говорили в начале раздела «Динамика», относятся к категории потенциальных сил.

Установим связь между работой и кинетической энергией. Представим, что на некотором прямолинейном пути от точки 1 к точке 2 на тело (мы пока рассматриваем простейшее тело—материальную точку) действует некоторая сила  $f$ . Тогда работа, производимая на этом участке, равна  $A = \int_1^2 f dl$ .

Согласно второму закону Ньютона,  $f = m \frac{dv}{dt}$ . В свою очередь,  $dl = v dt$ . Тогда

$$A = \int_1^2 m \frac{dv}{dt} v dt, \quad \text{или} \quad A = \int_1^2 mv dv.$$

Интегрируя, получаем

$$A = \left( \frac{mv^2}{2} \right)_2 - \left( \frac{mv^2}{2} \right)_1.$$

Но первая точка отвечает началу движения, а вторая его концу. Поэтому работа равна приращению кинетической энергии тела на рассматриваемом участке пути.

## 2.9. Потенциальная энергия.

### Потенциальные кривые и поверхности

В § 2.8 было введено понятие работы и показано, как можно вычислить работу некоторой силы, действующей на материальную точку, при ее движении по любому пути. На основании приведенных формул легко заключить, что в произвольном случае работа зависит не только от действующей силы, но и от пути, по которому точка переходит из начального состояния в конечное.

Как уже отмечалось, существуют силы, работа которых не зависит от этого пути. К таким силам, которые получили название потенциальных или консервативных, относятся, в частности, все силы, связанные с гравитационными, кулоновскими, ядерными взаимодействиями. С математической точки зрения требование независимости определенного интеграла

$$A = \int_a^b (f_x dx + f_y dy + f_z dz) \quad \text{от пути, по которому точка пере-}$$

ходит из положения  $a$  в положение  $b$ , равносильно требованию, чтобы подынтегральные выражения  $f_x$ ,  $f_y$ ,  $f_z$  представляли собой частные производные соответственно по  $x$ ,  $y$  и  $z$  от некоторой функции, а элемент работы  $dA$  был полным дифференциалом этой функции. Если это так, то в результате интегрирования получим разность значений соответствующей функции для координат, отвечающих конечному и начальному положениям точки.

Рассмотрим теперь один простой пример, связанный с действием потенциальных сил. Пусть на расстоянии  $h$  над поверхностью некоторого массивного тела находится материальная точка. Если дать возможность этой точке свободно двигаться, то под действием силы притяжения к массивному телу она начнет падать на него, приобретая кинетическую энергию. На основании закона сохранения энергии (см. § 2.10) можно утверждать, что появление одного вида энергии неизбежно приводит к потере другого ее вида. В § 2.8 мы видели, что приобретенная телом кинетическая энергия  $\Delta T_{\text{кин}}$  равна работе  $\Delta A$  сил притяжения. Значит, именно на такую величину должна уменьшиться какая-то другая энергия. Эту энергию принято называть *потенциальной*. Если запас потенциальной энергии

охарактеризовать функцией  $U(h)$ , то следует принять  $\Delta A = -\Delta U$ .

Сохраним такое соотношение между порциями работы и потенциальной энергии для любых случаев. Тогда

$$dA = f_x dx + f_y dy + f_z dz = -\frac{\partial U}{\partial x} dx - \frac{\partial U}{\partial y} dy - \frac{\partial U}{\partial z} dz.$$

Отвечающая этому соотношению функция  $U(x, y, z)$  получила название *потенциальной функции*. Возрастание этой функции означает увеличение потенциальной энергии, а убывание—уменьшение. Полное изменение потенциальной энергии на участке 1—2

$$\Delta U = \int_1^2 \left( \frac{\partial U}{\partial x} dx + \frac{\partial U}{\partial y} dy + \frac{\partial U}{\partial z} dz \right) = -\Delta A = U(2) - U(1).$$

Следовательно, изменение потенциальной энергии материальной точки численно равно взятой с обратным знаком работе, которую надо затратить для того, чтобы за счет действия потенциальных сил переместить точку из начального состояния в конечное.

Потенциальная энергия есть функция взаимного расположения частей системы, и ее зависимость от координат, определяющих это взаимное расположение (например, расстояние между двумя притягивающимися по закону Кулона материальными точками, в которые помещены противоположные по знаку электрические заряды), может быть изображена графически. Если относительное расположение частей системы характеризуется одной переменной, то график зависимости потенциальная энергия—координата называется *кривой потенциальной энергии* или *потенциальной кривой*. Если для описания расположения тел друг по отношению к другу применяются две координаты, то графическое изображение потенциальной функции будет иметь вид обычной поверхности в трехмерном пространстве, характеризуемом тремя осями координат, по которым откладываются значения энергии (обычно по вертикальной оси) и пространственные параметры системы. Такая *поверхность* называется *потенциальной*. Если для описания относительного расположения частей системы используют три или более координат, то говорят о потенциальной поверхности в пространстве нескольких переменных. Графически в этом случае можно изобразить лишь сечения такой поверхности, когда все координаты, кроме одной, фиксируются, либо срезы, когда фиксируются все координаты, кроме двух.

Примеры потенциальных кривых для простейших случаев потенциальной функции тела, поднятого над землей на высоту  $h$ , и функции сжатой или растянутой пружины показаны на рис. 2.18, 2.19.

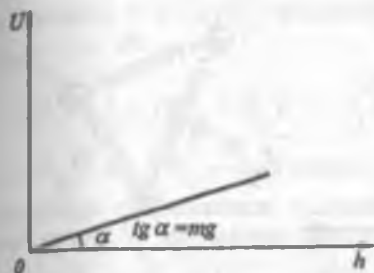


Рис. 2.18. График потенциальной функции подняемого над поверхностью земли на малую высоту тела

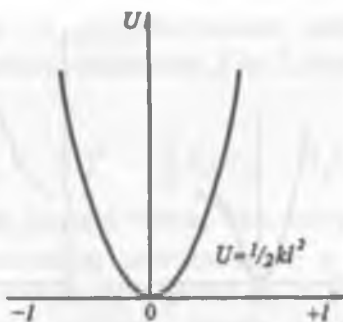


Рис. 2.19. График потенциальной функции сжимаемой или растягиваемой пружины

В первом случае график потенциальной функции представляет прямую, проходящую через начало координат с углом наклона, тангенс которого равен  $mg$ . Во втором случае потенциальная кривая есть парабола с минимумом в начале координат и кривизной, определяемой коэффициентом  $k$ .

Выше было показано, что

$$f_x = -\frac{\partial U}{\partial x}, \quad f_y = -\frac{\partial U}{\partial y}, \quad f_z = -\frac{\partial U}{\partial z},$$

т. е. компоненты силы, возникающей при взаимодействии тел в системе с потенциальной функцией  $U$ , зависящей от координат  $x, y, z$ , равны взятым с обратным знаком производным от  $U$  по  $x, y, z$ . Минус берется потому, что если потенциальная энергия возрастает, то приходится совершать работу, сила будет препятствовать движению. Если потенциальная энергия убывает, то возникают силы, которые помогают движению. Если взаимное расположение частей системы характеризуется одной координатой, то потенциальная энергия будет функцией одной координаты, и тогда для определения силы, возникающей при изменении этой координаты, надо вместо частной производной вычислять обычную производную от потенциальной функции по этой координате.

Рассмотренные графики (потенциальные кривые) играют в физике большую роль. В частности, по ним можно судить, будет ли система находиться в равновесии или нет.

Для того чтобы система могла находиться в устойчивом равновесии, необходимо, чтобы кривая или поверхность, изображающая зависимость потенциальной энергии от координат, характеризующих взаимное расположение частиц системы, обладала хотя бы одним минимумом. Чтобы пояснить сказанное, рассмотрим два графика (рис. 2.20, а, б).

В обоих случаях в точке  $x_0$  касательная к графику функции параллельна оси  $X$ , т. е.  $dU/dx = 0$ . Следовательно, когда система находится в состоянии, характеризуемом координатой  $x = x_0$ , никаких сил нет, т. е. система находится в состоянии



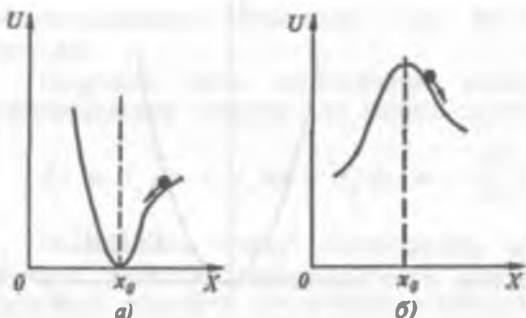


Рис. 2.20. Устойчивое (а) и неустойчивое (б) положения равновесия

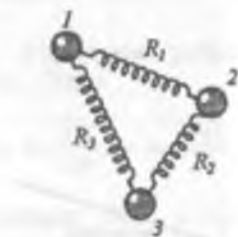


Рис. 2.21. Связанные упругими силами шары

равновесия. Если сместить систему из положения равновесия, то возникает действующая на тело сила, пропорциональная значению производной от  $U$  по  $x$  для новых точек. Однако в случае *а* эти силы будут стремиться вернуть систему в исходное положение, а в случае *б* еще больше отклонить ее от исходного положения. Можно наглядно это представить шариком, движущимся по кривой. В случае *а* шарик стремится сместиться к точке  $x_0$ , а в случае *б*, наоборот, упасть с холма, т. е. отойти от точки  $x_0$ . В случае *а* система обладает состоянием устойчивого равновесия, а в случае *б* нет. Говорят также, что в случае *а* имеется потенциальная яма.

Если относительное расположение тел в системе описывается двумя координатами, например координатами  $x$  и  $y$ , то математическое изображение потенциальной функции получается в виде потенциальной поверхности. Наглядным образом двумерной потенциальной поверхности является поверхность Земли с ее возвышенностями, оврагами и лощинами.

Вначале, когда мы вводили понятия координат и говорили о способах описания состояния различных систем, отмечалось, что можно использовать, в принципе, самые разнообразные координаты, лишь бы они имели наглядный геометрический смысл. Поскольку наличие потенциальной энергии всегда связано с взаимодействиями между отдельными составляющими рассматриваемой системы, то естественно описывать соответствующую потенциальную поверхность, используя координаты, отражающие взаимное расположение отдельных тел в пространстве. Так, например, если имеется несколько шариков (материальных точек), соединенных между собой пружинами (рис. 2.21), то потенциальная энергия такой системы проще всего выражается в координатах, которыми являются расстояния между отдельными парами шариков в системе. Потенциальная функция, зависящая в данном случае от трех переменных  $R_1, R_2, R_3$ , где  $R_1$  —

расстояние между шариками 1 и 2,  $R_2$ —расстояние между шариками 1 и 3,  $R_3$ —расстояние между шариками 2 и 3, имеет вид

$$U = \frac{1}{2} k_1 (R_{01} - R_1)^2 + \frac{1}{2} k_2 (R_{02} - R_2)^2 + \frac{1}{2} k_3 (R_{03} - R_3)^2,$$

где  $R_{01}$ ,  $R_{02}$ ,  $R_{03}$ —расстояния между парами точек, при которых пружины оказываются в своих нормальных состояниях, т. е. не растянуты и не сжаты. Нетрудно видеть, что, когда  $R_1 = R_{01}$ ,  $R_2 = R_{02}$ ,  $R_3 = R_{03}$ , потенциальная энергия, равная в этом случае нулю, достигает своего минимума; при любых других значениях  $R_i$  энергия  $U$  всегда больше  $U_0$ . Состояние  $U_0$  отвечает положению равновесия. При любой деформации системы, выводящей ее из положения равновесия, возникают силы (в данном случае силы упругости пружин), стремящиеся вернуть систему в положение равновесия.

С математической точки зрения, независимо от того, с помощью каких координат мы описываем относительные расположения тел в системе, состояние минимума всегда отвечает условию, при котором частные производные от потенциальной функции по выбранным координатам равны нулю. Это позволяет обобщить понятие силы (которое раньше вводилось как частная производная от потенциальной функции материальной точки соответственно по координатам  $x$ ,  $y$ ,  $z$ ) на случай любых координат и называть *компонентой обобщенной силы*, отвечающей изменению произвольной координаты  $q_i$ , производную, взятую с обратным знаком от потенциальной функции по этой координате.

Практический поиск положения равновесия сложной системы, если задана потенциальная функция как величина, зависящая от обобщенных координат, заключается в том, что с помощью вариации значений этих координат находятся такие состояния, когда потенциальная функция достигает минимума. Этих минимумов, вообще говоря, может быть много, и это означает, что система может обладать не одним, а многими положениями равновесия. Такой поиск положений равновесий с помощью вариаций координат называют в математике *вариационной процедурой*.

Так как система, находящаяся в состоянии равновесия, может быть выведена из этого состояния только за счет некоторого внешнего воздействия, то указанное состояние и будет наиболее вероятным. При разного рода внешних воздействиях система стремиться перейти из менее устойчивого в более устойчивое состояние равновесия.

Термин «менее или более устойчивое состояние» равновесия следует понимать следующим образом. При любом выходе системы из состояния равновесия появляются возвращающие

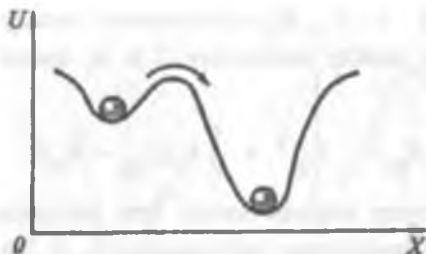


Рис. 2.22. Потенциальная кривая с двумя минимумами

силы, действующие только в пределах определенной деформации. Если система слишком далеко отошла от состояния равновесия, то она может перейти уже в состояние неравновесное и при дальнейшем изменении относительных координат системы может достичь нового более эффективного состояния равновесия. Для того чтобы вывести систему из этого нового состояния, надо еще больше изменить координаты. Наглядным примером тому является состояние шарика на некоторой кривой, которая имеет два минимума (рис. 2.22). Если шарик находится в первом (неглубоком) минимуме, то достаточно немного изменить его координату вдоль оси  $X$ , как он достигнет некоторого перевала, после чего скатится в следующий минимум. Простые опыты убеждают: для того чтобы вывести его из этого минимума, нужно затратить значительно больше энергии, приложить значительно большее усилие. Этот результат можно обобщить в виде следующего утверждения: чем более глубокой является потенциальная яма, тем более устойчиво положение системы в геометрическом состоянии, отвечающем этой яме, и тем большую энергию надо затратить, чтобы вывести систему из данного положения равновесия.

При внешних воздействиях система стремится перейти в это наиболее глубокое равновесное состояние. Такое правило, действительно наблюдаемое в многочисленных экспериментах, используется для установления наиболее вероятного состояния той или иной системы и называется *принципом минимума потенциальной энергии системы*.

Остановимся на одном важном следствии этого принципа, с помощью которого удастся объяснить и предсказать пространственное расположение атомов и молекул в кристаллах. Как известно, кристаллы—это тела, в которых атомы расположены в определенном порядке (рис. 2.23). Принято разделять кристаллы на два класса: *атомные и молекулярные*.

В атомных, как и в молекулярных, кристаллах существует вполне определенный порядок расположения атомов (рис. 2.24) и в этом смысле эти системы не отличаются друг от друга. Но в молекулярных кристаллах дополнительно можно выделить еще достаточно прочные образования, молекулы, которые и формируют упорядоченную пространственную структуру. При формировании такой структуры молекулы можно рассматривать как геометрические фигуры, обладающие вполне определенными

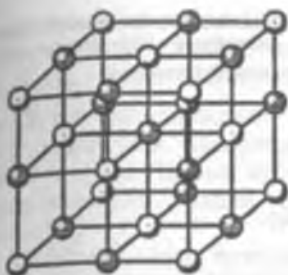
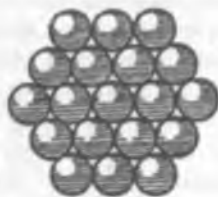


Рис. 2.23. Схематическое изображение кристалла NaCl. Атомы Na и Cl чередуются



а)



б)

Рис. 2.24. Слой из плотноупакованных шаров с квадратной симметрией (а): вокруг каждого шара четыре пустоты, вокруг каждой пустоты четыре шара. Слой из плотноупакованных шаров с шестиугольной симметрией (б): вокруг каждого шара шесть пустот, вокруг каждой пустоты три шара

длинами связей и валентными углами. И атомные, и молекулярные кристаллы существуют в целом как твердые образования только потому, что между отдельными атомами в кристаллах или между отдельными атомами разных молекул в молекулярных кристаллах имеются характерные взаимодействия. Они описываются потенциальными функциями.

Потенциальную функцию взаимодействия любой пары атомов для двух разных молекул можно представить в виде кривой (рис. 2.25). Аналитически такая кривая может быть описана функцией

$$U(R) = \frac{A}{R^6} - B \exp(-CR),$$

где  $R$ —расстояние между парами атомов двух разных молекул.

Физический смысл такого потенциала заключается в том, что если атомы расположены очень далеко друг от друга, то они практически не взаимодействуют. Математически это выражается в том, что «хвост» потенциальной кривой асимптотически устремляется к оси абсцисс. Когда атомы сближаются, то появляется определенное взаимодействие—притяжение между ними. Кривая устремляется вниз. При очень малых расстояниях между атомами они начинают наталкиваться один на другой, что приводит к появлению значительных сил отталкивания и росту потенциальной энергии взаимодействия. Где-то в промежуточном состоянии между совсем близким межатомным расстоянием и очень далеким существует минимум, который для двух атомов и будет определять положение их равновесия, т. е. наиболее устойчивое состояние. Потенциальная функция взаимодействия двух молекул

$$U = \sum_{i,j} [A_{ij}/R_{ij}^6 - B_{ij} \exp(-C_{ij} R_{ij})].$$

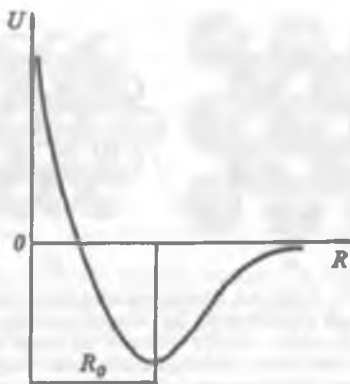


Рис. 2.25. Потенциальная кривая двух-атомной молекулы ( $R_0$ —равновесное расстояние между атомами)

Суммирование производится по всем парам атомов (по индексам  $i, j$ ), относящимся к разным молекулам. Коэффициенты  $A_{ij}$ ,  $B_{ij}$ ,  $C_{ij}$ —некоторые константы, которые свойственны данному сочетанию пар атомов. При изменении взаимного расположения двух молекул меняются и расстояния между атомами и может быть обнаружено такое состояние, которое отвечает минимуму потенциальной энергии или наиболее выгодному расположению одной молекулы по отношению к другой. Эти рассуждения можно распространить и на большое число молекул, составляющих молекулярный кристалл. Даже не производя никаких расчетов, можно заключить, что наиболее выгодными состояниями будут такие, когда в среднем все атомы расположены друг по отношению к другу на расстояниях, приблизительно отвечающих (для атомов одной молекулы, непосредственно

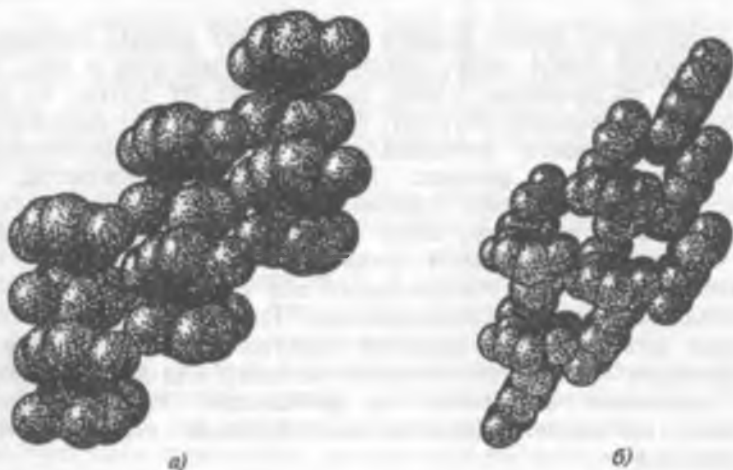


Рис. 2.26. «Выгодное» (а) и «невыгодное» (б) расположения молекул в кристалле  $C_6H_4Cl_{12}$

прилегающих к атомам другой молекулы) минимуму потенциальной кривой парного взаимодействия. Например, для молекул  $C_6H_4Cl_2$  (рис. 2.26) выгодным энергетическим расположением является расположение а, невыгодным—такое состояние б, когда эти молекулы расположены перпендикулярно друг к другу. В этом случае много атомов находится далеко друг от друга и, следовательно, вдали от положения минимума потенциальной энергии.

Условие, при котором кристалл образуется в случаях, когда удастся разместить все атомы или молекулы так, чтобы они располагались друг от друга на достаточно малых расстояниях, приблизительно отвечающих расстояниям минимумов потенциальных кривых парных взаимодействий, называют *принципом плотнейшей упаковки*. Этот принцип является одним из основных в *структурной кристаллохимии*—науке, которая занимается исследованием пространственных форм различных кристаллов. Примеры таких упаковок показаны на рис. 2.24.

В заключение параграфа обратимся к еще одному важному вопросу. Выше было показано, что для материальной точки в декартовой системе координат компоненты импульса связаны с кинетической энергией соотношениями

$$p_x = \frac{\partial T}{\partial v_x}, \quad p_y = \frac{\partial T}{\partial v_y}, \quad p_z = \frac{\partial T}{\partial v_z}.$$

В свою очередь, компоненты силы связаны с потенциальной энергией материальной точки соотношениями

$$f_x = -\frac{\partial U}{\partial x}, \quad f_y = -\frac{\partial U}{\partial y}, \quad f_z = -\frac{\partial U}{\partial z}.$$

Это позволяет по-иному записать уравнение движения материальной точки в поле потенциальных сил. В самом деле, если воспользоваться в общем случае движения в трехмерном пространстве уравнениями движения в форме

$$\frac{dp_x}{dt} = f_x, \quad \frac{dp_y}{dt} = f_y, \quad \frac{dp_z}{dt} = f_z,$$

то легко видеть, что их можно переписать так:

$$\frac{d}{dt} \left( \frac{\partial T}{\partial v_x} \right) + \frac{\partial U}{\partial x} = 0$$

(аналогично для  $y$  и  $z$ ).

Если ввести новую функцию  $L = T - U$ , получившую название функции Лагранжа и играющую важную роль в физике, то окончательно систему уравнений движения можно записать в виде

$$\frac{d}{dt} \left( \frac{\partial L}{\partial v_x} \right) - \frac{\partial L}{\partial x} = 0 \text{ и т. д.}$$

Выше неоднократно упоминалось, что состояние любой сложной системы может быть описано точкой в многомерном пространстве с помощью обобщенной системы координат и что в этом смысле имеется глубокая аналогия со случаем простой материальной точки.

Достаточно очевидно также, что если в системе действуют лишь внутренние потенциальные силы и она обладает потенциальной энергией, зависящей только от взаимного расположения тел, составляющих эту систему, то величина этой энергии никак не должна быть связана с тем конкретным набором координат, который используется для описания объекта. Любое изменение относительного расположения составляющих сложной системы будет сопровождаться либо совершением этой системой работы (когда потенциальная энергия уменьшается), либо работой над системой с увеличением ее потенциальной энергии. Разумеется, при этом будут действовать силы, отвечающие тем или иным деформациям системы, описываемым соответствующими координатами.

Опираясь на аналогию со случаем компонент силы, действующей на материальную точку в обычном трехмерном евклидовом пространстве, можно и в многомерном случае ввести вектор силы как вектор с компонентами  $f_k^{(q)} = -\frac{\partial U}{\partial q_k}$ , где  $q_k$ —одна из обобщенных координат. В более общем случае принимается, что  $f_k^{(q)} = \frac{\partial L}{\partial q_k}$ , где  $L$ —упомянутая выше функция Лагранжа.

Теперь вспомним, что не только компоненты простого импульса выражались через производные от кинетической энергии по компонентам вектора скорости ( $p_x = \frac{\partial T}{\partial v_x}$ ), но и компоненты вектора момента импульса ( $p_{\omega_x} = \frac{\partial T}{\partial \omega_x}$ ). Сохраняя это правило и в произвольном случае, определим обобщенный импульс как многомерный вектор с компонентами  $p_k^{(q)} = \frac{\partial T}{\partial \dot{q}_k}$ , и, далее,  $p_k^{(q)} = \frac{\partial L}{\partial \dot{q}_k}$ .

В результате состояние сложной системы в самом общем случае можно охарактеризовать совокупностью уравнений Лагранжа:

$$\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_k} \right) - \frac{\partial L}{\partial q_k} = 0 \quad (k = 1, 2, \dots, n).$$

Символом  $\dot{q}_k$  обозначена скорость изменения обобщенной координаты  $q_k$ .

Помимо того, что такая форма удобна для конкретных расчетов и пригодна для описания любых систем и для любого набора координат, важно еще и то, что, опираясь на функцию Лагранжа, можно сформулировать один из весьма общих физических принципов—*принцип наименьшего действия*. Он утверждает: любая система будет совершать такие движения, при которых

интеграл  $\int_1^2 L dt$ , взятый от начальной точки до конечной,

оказывается минимальным.

С помощью этого принципа можно получить не только обычные уравнения движения в форме Ньютона, но и описать значительно более широкий круг явлений, чем те, которые входят в область простых механических.

## 2.10. Сохранение энергии в механической системе

Выше были введены понятия кинетической и потенциальной энергий. Если тело движется равномерно и прямолинейно, то в этом случае его кинетическая энергия постоянна.

Изменение кинетической энергии тела на некотором участке траектории от точки  $a$  до точки  $b$  равно работе сил, которые замедляют или ускоряют его движение. Если кинетическая энергия возрастает, то совершается работа. Если кинетическая энергия уменьшается, то тело само совершает работу.

Сумму кинетической и потенциальной энергий называют *полной механической энергией* и обозначают  $H$ . Таким образом,  $H = T + U$ . Если в данной системе действуют только внутренние силы, то полная механическая энергия тел, составляющих систему, не изменяется. Это закон *сохранения механической энергии*.

Под внутренними силами понимают силы, которые возникают в результате воздействия одной части рассматриваемой системы на другую ее часть.

Кинетическая и потенциальная энергии могут переходить друг в друга, но при этом их сумма не меняется. Для примера рассмотрим шарик, висящий на нити. Если отклонить шарик в сторону, то система приобретает дополнительный запас потенциальной энергии. Отпустим шарик, и он придет в движение. В момент, когда шарик проходит положение равновесия, его потенциальная энергия будет минимальной, но зато он приобретает запас кинетической энергии. Это позволит ему, продолжая движение, подняться на некоторую высоту, в результате чего кинетическая энергия шарика перейдет снова в потенциальную. Затем все будет повторяться в обратном направлении. Здесь все время потенциальная энергия будет переходить в кинетическую, и наоборот, но сумма их сохраняется.

Рассмотрим теперь шарик, падающий с высоты  $h$  (рис. 2.27). Покажем, что его полная энергия в любой промежуточной точке



будет одной и той же. Считаем, что шарик в начальный момент покоился. Его кинетическая энергия в точке  $z$  равна  $T = \frac{1}{2}mv(z)^2$ , а потенциальная  $U = mgz$ . Полная энергия  $H = \frac{1}{2}mv(z)^2 + mgz$ . Скорость шарика в момент времени  $t$  равна  $v(z) = gt(z)$ . Таким образом,

$$T = \frac{1}{2}m[gt(z)]^2 = mg\left(\frac{1}{2}gt^2\right).$$

Но  $\frac{1}{2}gt^2$ —путь, проходимый шариком с ускорением  $g$ . Значит,  $T = mg(h - z)$ , откуда

$$H = T + U = mg(h - z) + mg(z) = mgh = \text{const.}$$

В заключение заметим, что механическая энергия системы может переходить в другие виды, например в тепловую. При этом, если движение частей системы совершается под действием только внутренних сил, оно может существовать до тех пор, пока не исчерпан запас полной энергии всей системы. Это правило устанавливает предел движения в механической системе.

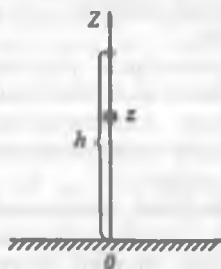


Рис. 2.27. К демонстрации закона сохранения энергии

Закон сохранения энергии более детально будем изучать в разделе «Термодинамика», поэтому здесь ограничимся теми двумя примерами, которые были приведены выше. Уместно, однако, обратить внимание на следующее.

Выше мы рассмотрели задачу о движении шарика, падающего с некоторой высоты, и показали, что механическая энергия шарика все время будет постоянной величиной. Если никаких потерь нет, а шарик падает на абсолютно упругую поверхность, то он через равные промежутки времени должен подскакивать на ту же высоту.

Так как абсолютно упругих поверхностей реально не существует, то в эксперименте заметим, что шарик будет подскакивать каждый раз на высоту, меньшую предыдущей, что легко трактовать как результат потери части энергии за счет неидеальности упругого удара, сопротивления воздуха и др. Что было бы, если бы течение времени было не равномерным, а например, убыстрялось при втором падении по сравнению с первым? Ясно, что это отразилось бы на значении кинетической энергии и описанная выше закономерность в подскакивании шарика уже не наблюдалась.

Таким образом, закон сохранения энергии тесно связан с принципом равномерного течения времени, и именно этим

объясняется широкое распространение этого закона и выполнимость его во всех тех случаях, которые сегодня доступны эксперименту.

### Вопросы для самоконтроля

1. Каким образом, наблюдая движение материальной точки в некоторой системе координат, можно сказать, действует ли на данную точку какая-нибудь сила или не действует?

2. Какие серии экспериментов приводят к понятиям инертной и гравитационной масс? Почему в рамках классической механики и при обычных ньютоновских представлениях о пространстве и времени нельзя доказать утверждение, что обе эти массы тождественны одна другой?

3. Что такое уравнение движения материальной точки? Следствием какого закона является уравнение движения? Какую роль играют начальные условия при попытках предсказать характер движения тела и его местонахождение через заданный промежуток времени от начала?

4. Как объяснить наблюдаемый по телевизору факт свободного полета космонавтов внутри космических кораблей? Можно ли воспроизвести такой эксперимент внутри движущегося самолета и что нужно для этого сделать?

5. Что такое принцип детерминизма и какие факторы ограничивают возможную точность предсказания местоположения тела и его движения через заданный промежуток времени, если известны уравнение движения тела и начальные условия?

6. Приведите примеры из бытовой практики, в которых обнаруживается действие третьего закона Ньютона.

7. Что нужно сделать, чтобы в процессе полета изменить направление движения ракеты?

8. В романе Жюль Верна «Из пушки на Луну» описывается, что в определенный момент, когда снаряд находился в некоторой промежуточной точке между Землей и Луной, летящие внутри снаряда путешественники обнаружили явление невесомости. В чем состояла ошибка великого романиста и в какой момент на самом деле наблюдалось бы явление невесомости, если бы подобный полет мог быть реально осуществлен?

9. Что такое следящая система? Как можно объяснить с помощью этого понятия прихождение человека? Приведите пример приема медицинской диагностики состояния нервной системы человека, основанного на свойствах следящей системы.

10. Что такое близкое действие и дальнее действие в физике? Почему приходится вводить понятие «поле»? Какую роль при введении этого понятия играет утверждение, что любые взаимодействия в природе распространяются не быстрее, чем скорость света?

11. Как должен быть записан закон сохранения импульса при наличии взаимодействий между телами через поля?

12. Как, зная кинетическую энергию тела, можно найти компоненты соответствующего импульса?

13. Как можно имитировать силу тяжести в спутниках?

14. Как в произвольном случае подсчитать работу, совершаемую некоторой силой, при перемещении материальной точки по криволинейному пути? Какой физический смысл придется положительной и отрицательной по знаку работе?

15. Некоторая материальная точка переместилась по поверхности стола, преодолевая все время на своем пути силу трения. Можно ли сказать, что в этом случае она приобрела некоторую потенциальную энергию?

16. Потенциальная функция для некоторой материальной точки имеет вид  $U = ax^2 + bx + c$ . Найти положение устойчивого равновесия точки на оси  $X$ .

17. Потенциальная функция некоторой материальной точки при ее движении вдоль оси  $X$  имеет вид  $U = ax^3 + bx^2 + cx + d$ . Найти местоположение устойчивого равновесия точки на оси  $X$ . Всегда ли такое равновесие будет существовать? Какое должно выполняться условие для коэффициентов  $a, b, c$ , чтобы существовало условие устойчивого равновесия?

18. Нарисовать график зависимости потенциальной энергии параллелепипеда, перекидываемого через ребро несколько раз вдоль определенного направления. Проследить, в каких случаях положение минимума потенциальной энергии оказывается более низким и в каких более высоким по отношению к тому, находится ли параллелепипед на малом по площади основании или на большом.

19. Что такое принцип плотнейшей упаковки? Как с его помощью можно объяснить и предсказать пространственное расположение атомов в кристаллах?

20. Выполнялся бы закон сохранения энергии, если бы течение времени не было равномерным?



## ГЛАВА 3

### ЭЛЕМЕНТЫ ДИНАМИКИ ТЕЛА И БОЛЬШИХ СКОРОСТЕЙ

*«Смотри в корни!»*

*К. Прутков*

#### 3.1. Движение твердого тела

При изучении законов движения тел мы в качестве модели использовали тело, размеры которого бесконечно малы по сравнению с областью, в которой тело перемещается, и для которого существен только вопрос, где тело находится в данный момент времени, с какой скоростью и с каким ускорением движется. Правда, мы уже упоминали о том, что могут быть и совокупности отдельных тел, например тело, состоящее из двух материальных точек, для которого уже можно поставить и другие вопросы: каково расстояние между материальными точками, как эта совокупность точек повернута по отношению к осям координат и т. д.? Для описания таких более сложных систем

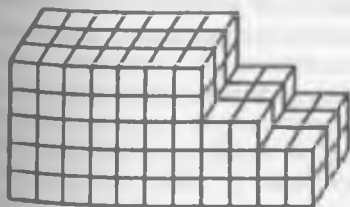


Рис. 3.1. Представление твердого тела в виде совокупности маленьких кубиков

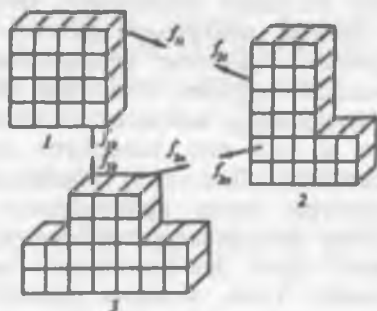


Рис. 3.2. Взаимодействия трех тел могут быть описаны совокупностью подчиняющихся третьему закону Ньютона взаимодействий друг с другом всех «кирпичиков» (материальных точек), образующих тела

нужно вводить большее число координат. При этом можно пользоваться как декартовыми координатами, так и обобщенными.

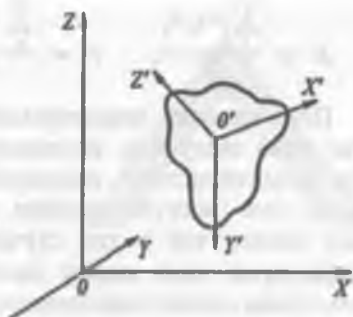
Прежде чем перейти к изучению темы, которая составляет содержание данного параграфа, обратим внимание на следующее. Очевидно, любой сложный объект, например тело произвольной формы, можно представить как объект, состоящий из большого числа материальных точек. Это можно сделать, если мысленно разрезать объект на большое число кубиков (рис. 3.1) и сложить их между собой так, чтобы они составили в целом то тело, которое имелось. Ясно, что точно так же можно поступить и с совокупностью тел (рис. 3.2). Каждое из тел, составляющих эту систему (в данном случае три тела), также разделим на совокупность маленьких частей (материальных точек). Взаимодействие между телами сведется к совокупности взаимодействий между отдельными материальными точками.

В дальнейшем, если необходимо изучить движение такой группы тел, можно составить систему уравнений движения для всех материальных точек, образующих эту группу, введя подходящим образом совокупность координат, которые описывают относительные расположения всех точек в пространстве и местоположение этих точек по отношению к системе координат, выбранной в качестве неподвижной. В общем случае для этого нужно задать  $3N$  степеней свободы, где  $N$ —число материальных точек, на которые разделены тела.

Здесь, однако, можно выделить один частный, но распространенный случай, а именно случай твердого тела. Под *твердым телом* мы будем понимать совокупность точек, расстояния между которыми в процессе движения не меняются. Если это так, то нет, очевидно, смысла следить за движением каждой из материальных точек, составляющих данное твердое тело. Можно выбрать другой прием, который позволит упростить описание и более наглядно представить движение всего тела.

Введем декартову систему координат (рис. 3.3), которую примем в качестве неподвижной и в которой будем задавать движение твердого тела. При описании тела необходимо ответить на следующие вопросы: в каком месте пространства находится тело и как оно повернуто по отношению к внешним осям координат. Для того чтобы охарактеризовать поворот тела, совместим (пока произвольно) дополнительную прямоугольную систему координат с твердым телом. При повороте тела одновременно будет поворачиваться и связанная с ним система координат. Углы, которые образует каждая из связанных с телом осей координат  $X'$ ,  $Y'$ ,  $Z'$  с неподвижными осями координат  $X$ ,  $Y$ ,  $Z$ , примем в качестве координат, характеризующих поворот твердого тела по отношению к лабораторной, или неподвижной, системе координат. Таких углов будет девять, по три для каждой из осей  $X'$ ,  $Y'$  и  $Z'$ .

Рис. 3.3. Оси координат  $X, Y, Z$  неподвижны. Оси координат  $X', Y', Z'$  смещаются в пространстве и поворачиваются вместе с телом



Оказывается, что независимыми среди этих девяти углов будут только три. Таким образом, мы получим три *вращательные степени свободы*.

Теперь выясним, где наиболее выгодно разместить точку, отвечающую началу координат  $X', Y', Z'$ , т.е. точку  $O'$ . Определим также, каким образом наиболее рационально ввести координаты, которые будут описывать перемещение твердого тела в пространстве, т.е. отвечать на вопрос, где находится тело. Для этого снова обратимся к модели твердого тела как совокупности материальных точек. При движении этого тела каким бы оно ни было, каждой материальной точке массой  $m_i$  можно приписать некоторую скорость  $v_i$  и импульс  $p_i$ . Полный импульс тела равен

$$P = \sum_i m_i v_i$$

Определим теперь полный импульс тела иным способом, т.е. как произведение полной массы  $M = \sum_i m_i$  тела на некоторую скорость  $V$ :  $P = MV$ . Это можно сделать, поскольку вектор  $P$  представляет сумму векторов  $p_i$ . Это позволяет говорить о векторе  $P$  как об импульсе некоторой точки, в которой сосредоточена вся масса твердого тела. Наша дальнейшая задача будет заключаться в определении координат этой точки. Это нетрудно сделать, пользуясь соотношением  $MV = \sum_i m_i v_i$ , которое можно записать в форме разложения по осям координат:

$$M \frac{dX}{dt} = \sum_i m_i \frac{dx_i}{dt}, \quad M \frac{dY}{dt} = \sum_i m_i \frac{dy_i}{dt}, \quad M \frac{dZ}{dt} = \sum_i m_i \frac{dz_i}{dt}.$$

Интегрируя эти выражения по времени, находим  $MX = \sum_i m_i x_i$  и т.д. Отсюда для координат точки, которой мы приписываем полный импульс тела, получаем

$$X = \frac{\sum_i m_i x_i}{M}, \quad Y = \frac{\sum_i m_i y_i}{M}, \quad Z = \frac{\sum_i m_i z_i}{M}.$$

Перемещение определенной таким образом точки в пространстве (она получила название *центра масс*) будем использовать для характеристики положения тела в пространстве и, в частности, его поступательного движения. Из определения центра масс видно, что в тех случаях, когда полный импульс системы равен нулю (это может быть либо тогда, когда тело покоится, либо тогда, когда оно вращается так, что ось вращения проходит через точку, отвечающую центру масс), не меняются и координаты центра. Таким образом, отличие от нуля полного импульса системы может быть принято в качестве признака наличия поступательного движения твердого тела и вообще системы взаимодействующих материальных точек.

Прежде чем перейти к выяснению закона движения центра масс, укажем, что центр масс обладает еще одним интересным свойством. Пользуясь правилом равновесия рычагов, нетрудно убедиться, что если подвесить твердое тело в точке, отвечающей центру масс, то оно будет находиться в состоянии безразличного равновесия. Если прикрепить этот подвес в любой другой точке, то тело примет такое расположение, при котором точка подвеса и точка, отвечающая центру масс, будут лежать на одной прямой, перпендикулярной поверхности земли (причем центр масс должен быть ниже точки подвеса).

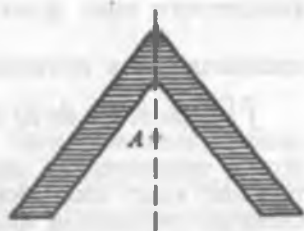
Вернемся к исходному выражению  $MV = \sum_i m_i v_i$  или  $P = \sum_i p_i$ . Продифференцируем его по времени:  $\frac{dP}{dt} = \sum_i \frac{dp_i}{dt}$ . По

определению силы можно записать:  $\frac{dp_i}{dt} = f_i$  и  $F = \sum_i f_i$ ;  $F$  будет иметь смысл силы, приложенной к центру масс и заставляющей этот центр масс приходить в движение;  $f_i$ —силы, приложенные к отдельным материальным точкам, составляющим тело.

На материальные точки, составляющие тело, действуют две группы сил—внутренние силы, которые скрепляют систему и превращают ее в единое целое, и внешние силы, возникающие за счет взаимодействия тела с каким-то другим. Таким образом,

$$F = \sum_i f_{i \text{ внеш}} + f_{i \text{ внутр}}.$$

рис. 3.4. Центр масс расположен в точке A



Согласно третьему закону Ньютона сумма всех внутренних сил должна быть равна нулю. Поэтому окончательно получаем, что

$$\mathbf{F} = \sum_i \mathbf{f}_{i, \text{внеш}}$$

Это позволяет записать общее уравнение движения для центра масс твердого тела в форме  $M \frac{d\mathbf{V}}{dt} = \mathbf{F}$ . Силу  $\mathbf{F}$  принято называть *равнодействующей всех внешних сил*. Как и для одной материальной точки, более общей формой записи является следующая:  $\frac{d\mathbf{P}}{dt} = \mathbf{F}$ .

Так как положение центра масс также определяется тремя координатами, то можно окончательно заключить, что движение твердого тела в общем случае описывается шестью координатами или, что то же самое, твердое тело имеет шесть степеней свободы.

В заключение отметим, что центр масс тела может не находиться внутри тела. Например, центр масс кольца находится в его геометрическом центре, где материала кольца нет. Центр масс угловой фигуры в тех случаях, когда распределение вещества вдоль по сторонам этого угла равномерно, находится в точке на биссектрисе угла (рис. 3.4).

В любом случае определение положения центра масс не составляет труда, если опираться на формулу координат центра масс, т.е. на формулу  $\mathbf{R} = \sum_i m_i \mathbf{r}_i / M$ . При непрерывном распределении масс в теле можно заменить соответствующее суммирование интегрированием. Для этого разобьем все тело на бесконечно малые объемы и припишем каждому объему плотность  $\rho(x, y, z)$ . При переходе от одной точки пространства к другой точке масса соответствующего кубика будет меняться. Полная масса тела определяется интегралом:

$$M = \iiint \rho(x, y, z) dV = \iiint \rho(x, y, z) dx dy dz.$$

Это тройной определенный интеграл, который может быть вычислен математическими средствами, вне зависимости от



конкретного вида функции  $\rho(x, y, z)$ . Аналогично сумма  $\sum_i \rho_i x_i$  превращается в интеграл по всему объему:

$$\iiint \rho(x, y, z) x \, dx \, dy \, dz.$$

Итак, окончательно при непрерывном распределении массы в твердом теле координаты центра масс могут быть вычислены согласно следующей формуле:

$$x = \iiint \rho(x, y, z) x \, dx \, dy \, dz / \iiint \rho(x, y, z) \, dx \, dy \, dz.$$

Аналогично для компонент  $y$  и  $z$ . Нетрудно убедиться с помощью, например, конкретных вычислений для простых примеров, что расположение центра масс по отношению к твердому телу является инвариантным к выбору внешней системы координат  $X, Y, Z$ , т. е. не зависит от выбора этой системы. Таким образом, центр масс является действительно характеристикой, присущей данному телу и не связанной с выбором внешней и внутренней системы координат.

Рассмотрим теперь вращение тела вокруг некоторой оси. Кинетическую энергию тела всегда можно записать в виде

$$T = \frac{1}{2} \sum_i m_i |v_i|^2.$$

Суммирование происходит по всем точкам, составляющим это тело. Если воспользоваться угловой скоростью вращения и учесть, что разные точки расположены от оси вращения на разном расстоянии  $r_i$ , то получим  $|v_i| = |r_i| |\Omega|$  и

$$T = \frac{1}{2} \sum_i m_i r_i^2 |\Omega|^2 = \frac{1}{2} \left( \sum_i m_i r_i^2 \right) |\Omega|^2,$$

где  $\Omega$ —вектор угловой скорости тела.

Сумма  $\sum_i m_i r_i^2 = \sum_i I_i$  называется *моментом инерции тела относительно выбранной оси вращения*. Из формулы видно, что момент инерции тела равен сумме моментов инерции составляющих это тело точек относительно оси вращения.

Как и в случае одной точки, дифференцируя выражение для кинетической энергии вращающегося вокруг фиксированной оси тела по компонентам угловой скорости, получаем составляющие по осям координат момента импульса тела. В векторной форме момент импульса имеет вид

$$\mathbf{p}_\Omega = I \Omega.$$

В свою очередь, дифференцирование момента импульса тела по времени приводит к выражению

$$\frac{d}{dt} p_{\Omega} = I \frac{d\Omega}{dt} = \sum_i \left( I_i \frac{d\Omega}{dt} \right).$$

Произведение  $I_i \frac{d\Omega}{dt}$  должно быть равно моменту силы, действующей на точку, относительно оси вращения. Сумма всех таких частных моментов сил называется *равнодействующим моментом сил относительно оси вращения* или *полным моментом сил*. Именно этот момент и определяет вращение тела как целого. Окончательно имеем

$$\frac{d}{dt} p_{\Omega} = I \frac{d\Omega}{dt} = \sum_i N_i = N.$$

Это выражение и есть *уравнение вращения тела вокруг некоторой оси*.

Если на тело действуют внешние силы, такие, что  $F$  и  $N$  не равны нулю, то оно будет перемещаться в пространстве, одновременно вращаясь.

В случае, когда направление оси вращения не меняется, его полная кинетическая энергия равна сумме энергий поступательного и вращательного движений:  $T = \frac{1}{2} M |v|^2 + \frac{1}{2} I |\Omega|^2$ .

Материал этого параграфа требует некоторых дополнительных пояснений. Это относится к правилам сложения приложенных к телу внешних сил и моментов. В самом деле, когда рассматривалось движение материальной точки и правило сложения действующих на нее сил, то все обстояло просто: линии действия всех сил пересекались в одной точке, и поэтому равнодействующая всех сил могла быть найдена последовательным применением к каждому двум правилу параллелограмма. В случае твердого тела ситуация усложняется: линии действия приложенных к телу внешних сил могут и не пересекаться друг с другом вообще. Чтобы обойти возникающие при этом трудности, поступают следующим образом.

Во-первых, можно с самого начала выделить из всей совокупности внешних сил такие, линии действия которых проходят через центр масс.

Достаточно очевидно, что под действием таких сил тело может двигаться только поступательно. Равнодействующая таких сил находится по обычным правилам.

Чтобы выяснить, как надо поступать с силами, линия действия которых не проходит через центр масс, выберем одну из них ( $F_A$ ), приложенную к точке  $A$ , расположенной на некотором расстоянии от центра масс (точка  $M$ ) (рис. 3.5).

Заметим, что поскольку речь идет о твердом теле, все элементарные участки которого жестко связаны между собой, то любая внешняя сила, где бы она ни была приложена, всегда

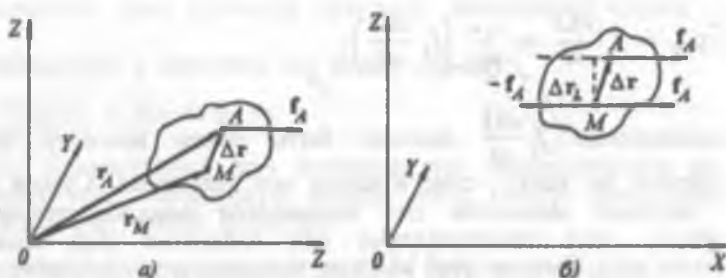


Рис. 3.5. Иллюстрация введения пары сил

будет воздействовать на все тело и приводить его в движение в целом.

Теперь представим себе эксперимент, который любой читатель при желании может легко проделать. Пусть имеется некоторое колесо, свободно вращающееся вокруг оси. Его можно привести во вращение, если приложить силу по касательной к ободу (толкнуть его рукой), но при этом, если удерживать ось вращения другой рукой, можно ощутить и силу, действующую на ось в том же направлении, что и сила, приложенная к ободу. Если бы ось была свободной, то мы обнаружили бы, что не только повернулось колесо, но и изменилось положение оси.

Вспомним, что любые движения твердого тела можно разделить на два вида: движение центра масс и поворот вокруг оси, проходящей через центр масс.

Математически такое разделение можно произвести, вводя связанную с центром масс систему декартовых координат  $X_M, Y_M, Z_M$  такую, что ее оси во все время движения остаются параллельными внешним осям  $X, Y, Z$ , и связанную с телом систему  $X', Y', Z'$ , которая может только вместе с телом поворачиваться вокруг центра масс, но не может от этого центра «оторваться».

Выше было отмечено, что приложенная к ободу колеса сила будет не только поворачивать колесо, но и тянуть его ось. Чтобы исключить в системе  $X', Y', Z'$  движение центра масс, необходимо, очевидно, при переходе к ней как-то преобразовать и силы. О принципе преобразования можно догадаться, вспомнив, что любое колесо легко повернуть так, чтобы ничем не удерживаемая его ось осталась неизменной, если одновременно поворачивать его, взявшись двумя руками за обод на противоположных концах диаметра колеса и прикладывая к ободу две одинаковые противоположно направленные силы. Такая комбинация получила название *пары сил*. Чтобы понять, как такая пара сил вводится математически, обратимся к рис. 3.5, б. Приложим к центру масс силу  $f_A$ , равную по величине силе, действующей на точку А. Скомпенсируем воздействие этой новой силы на центр масс равной ей по величине, но противоположно

направленной силой  $-\mathbf{f}_A$ . Ничего, конечно, в характере движения твердого тела при такой формальной операции не изменится.

Рассмотрим суммарный момент сил  $\mathbf{f}_A$  для точки  $A$  и  $-\mathbf{f}_A$  для центра масс относительно начала ( $O$ ) неподвижной системы координат  $X, Y, Z$ .

Получим

$$\mathbf{N} = \mathbf{r}_A \times \mathbf{f}_A - \mathbf{r}_M \times \mathbf{f}_A = \Delta \mathbf{r} \times \mathbf{f}_A.$$

Здесь  $\mathbf{r}_A$  и  $\mathbf{r}_M$  — радиус-векторы точек  $A$  и  $M$  в неподвижной системе координат и  $\Delta \mathbf{r} = \mathbf{r}_A - \mathbf{r}_M$  — радиус-вектор точки  $A$  относительно центра масс, т. е. в системе  $X_M, Y_M, Z_M$ . Сразу же бросается в глаза, что момент  $\mathbf{N}$  (момент пары сил) не меняется при поступательном движении тела (инвариантен относительно поступательного движения) и вообще при параллельном переносе вектора  $\Delta \mathbf{r}$  в пространстве. Видно также, что если векторы  $\Delta \mathbf{r}$  и  $\mathbf{f}_A$  не являются взаимно перпендикулярными, то возникнут две составляющие  $\mathbf{f}'_A$  и  $\mathbf{f}''_A$ , равные по величине и противоположные по знаку, «растягивающие» отрезок  $AM$ . Разумеется, ни к какому движению твердого тела эти две составляющие привести не могут. Таким образом, при построении момента пары сил надо учесть только составляющие сил  $\mathbf{f}_A$  и  $-\mathbf{f}_A$ , перпендикулярные вектору  $\Delta \mathbf{r}$ . Легко показать, что величина момента  $\mathbf{N}$  пары сил не изменится и тогда, когда вместо вектора  $\Delta \mathbf{r}$  берется вектор  $\Delta \mathbf{r}_\perp$ , образуемый перпендикуляром, проведенным из центра масс к линии действия силы  $\mathbf{f}_A$ . Поскольку, кроме того, действие на тело момента пары сил инвариантно относительно трансляций векторов  $\Delta \mathbf{r}$  или  $\Delta \mathbf{r}_\perp$ , то всегда для определенности можно совместить центр пары сил с центром масс.

В результате всех этих соображений получается следующий рецепт учета действия всех сил, приложенных в любых точках к твердому телу. Вначале все силы путем параллельного переноса надо привести к центру масс. Это позволит в дальнейшем найти соответствующую равнодействующую силу и решить задачу о поступательном движении тела полностью аналогично задаче о движении материальной точки. На основе тех сил, линии действия которых не проходят через центр масс, надо построить соответствующие моменты пар сил, совместив векторы  $\mathbf{N}$  для всех пар также с центром масс. Построенная после этого по обычным правилам равнодействующая всех моментов позволит найти направление оси вращения и решить соответствующее уравнение для вращения твердого тела.

### 3.2. Момент инерции

Сделаем ряд важных замечаний о моменте инерции твердого тела. Момент инерции зависит от формы тела, его массы и оси вращения. Если ось вращения изменится, то момент инерции тоже изменится.

Согласно теореме Штейнера при параллельном переносе оси вращения на расстояние  $l$  момент инерции меняется на  $Ml^2$ , если первоначальная ось проходила через центр масс тела. Другими словами, если  $I_0$ —момент инерции относительно оси, проходящей через центр масс, то момент инерции  $I$  относительно другой оси, параллельной исходной и отстоящей от нее на расстоянии  $l$ , равен

$$I = I_0 + Ml^2.$$

Здесь важно обратить внимание на следующее. Момент инерции всегда величина положительная. Это видно из определения. Кроме того,  $Ml^2 > 0$ . Отсюда заключаем, что  $I_0$  меньше, чем  $I$ . Мы получили важный результат: из всех моментов инерции относительно ряда параллельных осей наименьшим будет тот, для которого ось проходит через центр масс.

Пока мы говорили об одной оси, проведенной через центр масс. Ясно, что через центр масс можно провести множество осей. Однако для всякого тела существуют три выделенных момента инерции для трех определенным образом расположенных взаимно перпендикулярных осей. Эти моменты инерции для осей называют *главными моментами инерции тела*. Через главные моменты можно выразить любые другие моменты инерции тела при любых заданных осях.

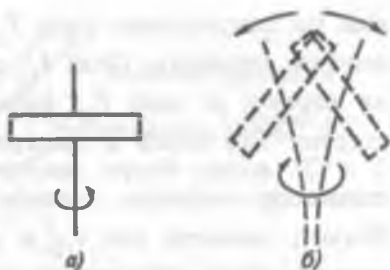
Направление осей главных моментов инерции обладают тем важным свойством, что свободные вращения тел, совершаемые вокруг осей с наибольшими и наименьшими значениями таких моментов, будут устойчивыми. Если вращение не совпадает ни с одной из таких осей, то оно будет неустойчивым. Эти случаи показаны на рис. 3.6.

Приведем некоторые значения моментов инерции. Наиболее простое тело—шар. Его центр масс лежит в центре шара. Здесь для всех трех взаимно перпендикулярных осей, проходящих через центр масс, получаются одни и те же моменты инерции. Момент инерции шара относительно оси, проходящей через его центр,

$$I = \frac{2}{5} MR^2,$$

где  $M$ —масса шара и  $R$ —радиус шара.

Рис. 3.6. Устойчивое вращение однородного стержня вокруг оси, проходящей через центр масс и совпадающей с направлением одной из главных осей момента инерции (а); неустойчивое вращение, когда ось вращения не совпадает с осью главного момента инерции (б)



Момент инерции однородного цилиндра радиуса  $R$  относительно его оси

$$I = \frac{1}{2} MR^2.$$

Момент инерции балки (т. е. параллелепипеда со сторонами  $a$  и  $b$ ) относительно оси, проходящей через центр этой балки,

$$I = \frac{1}{12} M(a^2 + b^2).$$

### 3.3. Закон сохранения момента импульса

При изучении динамики материальной точки указывалось, что сила действия одного тела на второе равна силе действия второго тела на первое; эти силы противоположно направлены. Роль силы в случае вращения будет играть момент силы, роль скорости—угловая скорость, роль массы—момент инерции. Если одно тело в результате взаимодействия с другим телом приходит во вращение и возникает момент силы, то на другое тело будет действовать момент, равный по числовому значению и противоположный по знаку. Таким образом, появится соотношение между моментами, аналогичное соотношению между силами.

Следствием этого является закон сохранения момента импульса, аналогичный закону сохранения импульса и утверждающий, что полный момент импульса в замкнутой системе тел есть величина постоянная. Если на тела системы действуют только внутренние силы и внутренние моменты силы, то момент импульса системы не меняется. Это значит, что если внутри системы возникло вращение, появился момент импульса, то обязательно внутри той же системы должно возникнуть вращение других ее частей в противоположном направлении так, чтобы общий момент импульса оставался равным нулю.

Покажем это на примере взаимодействия двух материальных точек 1 и 2, положения которых описываются во внешней системе координат (такую систему часто называют еще лабораторной) радиус-векторами  $\mathbf{r}_1$  и  $\mathbf{r}_2$ .

Если между точками 1 и 2 имеется механическое взаимодействие, то возникнут две равные по величине и противополож-

ные по направлению силы  $f_{12}$  и  $f_{21}$ . Линии действия этих сил должны совпадать. Сила  $f_{12}$  действует со стороны второй точки на первую, а сила  $f_{21}$  действует со стороны первой силы на вторую. Три точки в пространстве (начало координат и точки 1 и 2) всегда будут лежать в одной плоскости, и в той же плоскости, очевидно, должны лежать и векторы сил  $f_{12}$  и  $f_{21}$ . Значит, моменты сил  $f_{12}$  и  $f_{21}$  относительно начала координат должны быть направлены вдоль одной прямой—нормали к плоскости, но иметь разные знаки.

Моменты сил будут передаваться векторами соответственно  $N_1 = r_1 \times f_{12}$  и  $N_2 = r_2 \times f_{21}$ . Сумма этих моментов

$$N = N_1 + N_2 = r_1 \times f_{12} + r_2 \times f_{21} = r_1 \times f_{12} - r_2 \times f_{12} = (r_1 - r_2) \times f_{12}.$$

Вектор  $(r_1 - r_2) = \Delta r$  направлен вдоль прямой, соединяющей материальные точки 1 и 2, т. е. вдоль линии действия силы  $f_{12}$ . Поэтому векторное произведение  $\Delta r \times f_{12} = N_1 + N_2 = 0$ .

Таким образом, суммарный момент сил при взаимодействии двух материальных точек согласно третьему закону Ньютона оказывается равным нулю. Существенно при этом, что результат не зависит от выбора начала координат или от оси вращения. Если материальные точки закреплены на штангах, свободно вращающихся вокруг оси, то для них можно записать соответствующие уравнения движения (вращения). Получим

$$I_1 \frac{d\omega_1}{dt} = N_1 \quad \text{и} \quad I_2 \frac{d\omega_2}{dt} = N_2.$$

Так как  $N_1 + N_2 = 0$ , то

$$I_1 \frac{d\omega_1}{dt} + I_2 \frac{d\omega_2}{dt} = \frac{d}{dt} (I_1 \omega_1 + I_2 \omega_2) = 0.$$

Следовательно,  $I_1 \omega_1 + I_2 \omega_2 = P_{\omega_1} + P_{\omega_2} = \text{const.}$

Вывод этот не трудно распространить на все пары точек сложной системы, что и приведет к закону сохранения момента импульса системы точек:  $\sum_k P_{\omega_k} = 0$ .

Этот закон можно распространить и на систему, состоящую не из точек, а из тел.

Рассмотрим с этой точки зрения некоторые особенности конструкции вертолета. Вертолет имеет один несущий винт. При вращении винта под действием внутренних сил (двигателя) в силу требования постоянства момента импульса фюзеляж вертолета должен вращаться в другую сторону. Для того чтобы

Рис. 3.7. Компенсация вращения вертолета торможением хвостового винта



фюзеляж не крутился, его надо затормозить. Это можно сделать за счет внешних сил, которые создаются следующим образом (рис. 3.7): фюзеляж вертолета делается длинным, устанавливается дополнительный винт (с осью, перпендикулярной оси главного винта), в результате чего возникает сила, момент которой должен скомпенсировать вращение фюзеляжа. Так как момент силы пропорционален произведению силы на расстояние до оси вращения, то выгодно делать хвост вертолета длиннее, а мощность тормозящего дополнительного винта меньше. Таким образом, дополнительный боковой винт нужен для того, чтобы прекратить вращение кабины. В некоторых случаях этого винта нет, тогда делают не один несущий винт, а два с таким расчетом, чтобы они вращались в противоположных направлениях и компенсировали моменты импульсов друг друга.

Для твердого тела, вращающегося вокруг постоянной оси, уравнение движения имеет вид  $I(d\Omega / dt) = N$ . Математически оно аналогично уравнению поступательного движения. Однако надо иметь в виду, что в отличие от массы момент инерции может меняться, если в процессе вращения изменяется форма тела. Это изменение формы тела может происходить за счет внутренних сил. Например, когда фигурист на коньках, вращаясь вокруг оси, хочет остановиться, то он разводит руки.

Чтобы учесть общий случай, надо, полагая, что ось вращения неизменна, пользоваться записью уравнения для вращения тела в форме

$$\frac{d}{dt}(I\Omega) = \frac{dp_{\Omega}}{dt} = N.$$

Если  $N = 0$ , то  $p_{\Omega} = I\Omega = \text{const}$ .

Итак, если в процессе вращения под действием внутренних сил изменятся форма тела и его момент инерции, то обязательно изменятся и угловая скорость вращения. Фигурист, разводя руки в стороны, увеличивает момент инерции и снижает скорость вращения.

Так как момент импульса есть вектор, то при  $N = 0$  он должен сохранять не только свою величину, но и направление в пространстве. Этим объясняется способность волчков (в технике их называют гироскопами) сохранять при вращении вертикальное положение. Чем больше момент импульса волчка (его масса и угловая скорость вращения), тем труднее его опрокинуть. Этот



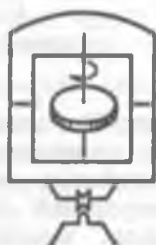


Рис. 3.8. Расположение гироскопа в вагоне поезда монорельсовой дороги



Рис. 3.9. Направление оси гироскопа при повороте ракеты сохраняется

же эффект позволяет понять, почему легко сохранять равновесие на быстро едущем двухколесном велосипеде и очень трудно в тех случаях, когда велосипед не движется. Способность гироскопа при вращении сохранять свою ось вращения используется в некоторых конструкциях монорельсовых дорог, где двухколесный экипаж находится над рельсом. Для того чтобы этот экипаж не опрокидывался во время остановок, в нем гироскоп располагается вертикально (рис. 3.8). Опрокидывание такого экипажа обязательно должно привести к изменению направления оси вращения, что требует значительных усилий. Сохранение гироскопом оси вращения используется также в системах ориентации ракет. Гироскоп располагается внутри ракеты и приводится в быстрое вращение. При движении ракеты (рис. 3.9) расположение оси гироскопа все время сохраняется и можно ориентировать эту ракету по отношению к этой оси.

### 3.4. Динамика больших скоростей

Мы в основном рассматривали тела, движущиеся с относительно малыми скоростями. Однако в тех случаях, когда скорости возрастают, проявляется связь между пространством и временем, так что пространственные и временные переменные уже нельзя рассматривать по отдельности и надо говорить о существовании четырехмерного мира. Поскольку никакое воздействие в природе не может передаваться со скоростью большей, чем скорость света, возникают определенные затруднения при толковании третьего закона Ньютона. Это вынуждает ввести понятия близкодействия и поля и перейти от третьего закона Ньютона к более общему закону—закону сохранения импульса.

Сейчас мы изучим еще ряд важных вопросов, связанных с динамикой больших скоростей.

Рассмотрим сначала уравнение движения в его первоначальной ньютоновской форме. Для простоты будем считать, что движение совершается вдоль одной оси  $X$ . Это позволяет воспользоваться уравнением движения не в векторной, а в скалярной форме. В механике Ньютона уравнение движения можно записать так:

$$m \frac{dv}{dt} = F \quad (v = v_x, \quad F = F_x) \quad \text{или} \quad \frac{dp}{dt} = F \quad (p = p_x).$$

Обе формы эквивалентны, так как по определению ньютоновский импульс  $p = mv$ , где масса играет роль постоянного коэффициента пропорциональности между векторами импульса и скорости. Такой коэффициент можно вносить под знак дифференциала или выносить из-под него, ничего не меняя в уравнении.

Решая уравнение при начальных условиях отсутствия движения, находим, что  $v = Ft/m$ . Стало быть, скорость должна возрастать прямо пропорционально времени. Значит, если это время велико (будем говорить, бесконечно велико) и действующая сила все время постоянна, то мы получим, что  $v$  может достигать любой величины, в том числе и большей скорости света. Это противоре-

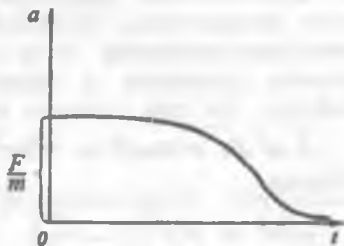


Рис. 3.10. Зависимость ускорения от времени

чит утверждению, что ни одно тело не может достигнуть скорости, равной скорости света, не говоря уже о том, чтобы иметь скорость больше скорости света. Так как при малых скоростях механика Ньютона верна, то остается только предположить, что при больших скоростях эта механика перестает быть верной. Если при малых скоростях  $v = Ft/m$ , а при больших скоростях  $v \rightarrow c$ , то график ускорения должен иметь вид, показанный на рис. 3.10: вначале ускорение постоянно, как и должно быть согласно ньютоновской механике, затем оно падает. Мы пока не знаем, по какому закону происходит это падение, и можно только утверждать, что так должно быть.

Посмотрим, что надо сделать, чтобы указанное противоречие разрешить.

Мы отмечали, что уравнение  $\frac{dp}{dt} = F$  является более общим, чем уравнение  $m \frac{dv}{dt} = F$ , в том смысле, что оно, в частности, позволяет говорить об изменении импульса поля, которое массой не обладает. Значит импульс выступает как некоторая самостоятельная характеристика, не обязательно определяемая произведением  $mv$ . Такое определение импульса должно, однако, непременно сохраняться для тел (именно эти физические объекты имеют свойство обладать массой), движущихся относительно

медленно. При больших скоростях определение импульса целесообразно уточнить. В частности, можно принять, сохраняя для непрерывности перехода классическое соотношение для вектора импульса  $p \sim v$  (р. IV), что скалярный коэффициент пропорциональности в соотношении между векторами  $p$  и  $v$  уже не равен массе, а зависит еще и от дополнительного множителя, который при малых скоростях должен обращаться в единицу.

Мы установили, что интервалы времени и длины преобразуются при переходе от одной системы отсчета к другой с помощью некоторого универсального множителя  $\gamma = (1 - v^2/c^2)^{-1/2}$ . Так как речь идет о преобразовании фундаментальных величин, то естественно предположить, что переход от классического импульса к релятивистскому также связан с этим множителем. Опираясь на эту догадку, примем в качестве импульса величину  $p = mv\gamma$ . Снова воспользовавшись уравнением движения в форме  $dp/dt = F$  и считая  $F = \text{const}$ , найдем, что при нулевых начальных условиях

$$p = Ft = mv(1 - v^2/c^2)^{-1/2}.$$

Отсюда

$$m^2 v^2 = F^2 t^2 \left(1 - \frac{v^2}{c^2}\right) \quad \text{и} \quad v^2 = \frac{F^2 t^2 c^2}{c^2 m^2 + F^2 t^2}.$$

Так как  $c^2$  и  $m^2$  — постоянные величины, то  $\lim_{t \rightarrow \infty} v^2$  оказывается

равным  $c^2$ . Таким образом, с помощью постоянной бесконечно долго действующей силы тело можно разогнать только до  $v = c$ , т. е. скорости света, но не больше. Значит, предположение, что  $p = mv\gamma$ , оказывается верным.

Перепишем выражение для  $v^2$  в другом виде:

$$v^2 = \frac{F^2}{m^2} t^2 \left(1 + \frac{F^2 t^2}{m^2 c^2}\right)^{-1}.$$

Но  $F^2/m^2 = a^2$ , где  $a$  — классическое ускорение тела (если сохранить опять-таки из требования предельного перехода к малым скоростям определение ускорения  $a$  таким же, как в ньютоновской механике). Тогда

$$v = at / \sqrt{1 + a^2 t^2 / c^2}.$$

Эта формула передает связь между скоростью и ускорением в релятивистском равноускоренном движении. Если ускорение невелико и время ограничено (это соответствует ньютоновской механике), то  $v = at$ .

Можно убедиться также, что релятивистский импульс (если применить формулу для ряда материальных точек) приводит к сохранению полного импульса при переходе от одной системы отсчета к другой. Таким образом, закон сохранения импульса справедлив и для тел, движущихся с большими скоростями.

Получим теперь еще одно важнейшее в физике соотношение. Для этого воспользуемся инвариантом

$$S^2 = c^2 t^2 - x^2 - y^2 - z^2.$$

Так как координаты  $x$ ,  $y$  и  $z$ , а также  $t$  имеют смысл отрезков, отложенных от нуля по соответствующим осям, то и для дифференциалов должно быть

$$(dS)^2 = c^2(dt)^2 - (dx)^2 - (dy)^2 - (dz)^2.$$

Выберем две системы отсчета, в одной из которых тело покоится, а в другой движется равномерно и прямолинейно. Тогда в первой системе отсчета  $dx' = dy' = dz' = 0$  и меняется только время.

Значит, опираясь на инвариантность  $(dS)^2$ , можно записать

$$c^2(dt')^2 - (dx)^2 - (dy)^2 - (dz)^2 = c^2(dt')^2.$$

Разделив обе части равенства на  $(dt')^2$  и учитывая, что  $(dt')^2 = (dt)^2(1 - v^2/c^2)$ , находим

$$c^2\gamma^2 - \left(\frac{dx}{dt}\right)^2\gamma^2 - \left(\frac{dy}{dt}\right)^2\gamma^2 - \left(\frac{dz}{dt}\right)^2\gamma^2 = c^2 \left(\gamma^{-1} = \sqrt{1 - v^2/c^2}\right).$$

Воспользовавшись определением релятивистского импульса, перепишем это равенство в форме

$$c^2\gamma^2 - (p_x^2 + p_y^2 + p_z^2)/m^2 = c^2$$

или

$$m^2 c^2 \gamma^2 - p^2 = m^2 c^2.$$

Это соотношение легко превратить в соотношение для квадратов энергии, если, сохраняя для возможности предельного перехода и в релятивистской механике размерность энергии как произведения размерности массы на размерность квадрата скорости, умножить левую и правую его части на инвариантную величину, имеющую размерность квадрата скорости. Такой инвариантной величиной является квадрат скорости света. Следовательно,

$$m^2 c^4 \gamma^2 - p^2 c^2 = m^2 c^4$$

или

$$E^2 - p^2 c^2 = E_0^2,$$

где  $E = mc^2\gamma = mc^2(1 - v^2/c^2)^{-1/2}$ ,  $E_0 = mc^2$ .

Обсудим физический смысл величин  $E$  и  $E_0$ . Величина  $E_0$  имеет размерность энергии и соответствует системе отсчета, в которой тело покоится. Естественно поэтому назвать эту энергию *энергией покоя*. Она соответствует некоторому внутреннему запасу энергии тела и не связана с его общим движением. В той системе отсчета, где тело движется, его полная энергия должна включать и кинетическую. По определению

$$T_{\text{кин}} = E - E_0 = mc^2(\gamma - 1).$$

Легко видеть, что это выражение при  $v \ll c$  переходит в привычное ньютоновское. Действительно, при  $v \ll c$  имеем

$$(1 - v^2/c^2)^{-1/2} \approx 1 + v^2/(2c^2),$$

и тогда

$$T_{\text{кин}} = E - E_0 = mc^2[v^2/(2c^2)] = mv^2/2.$$

Соотношение

$$T_{\text{кин}} = E - E_0, \text{ или } E_0 = E - T_{\text{кин}}, \text{ или } E = E_0 + T_{\text{кин}}$$

позволяет разрешить одно противоречие классической механики. В самом деле, хотя в классической физике тоже можно говорить о наличии внутренней энергии тела (например, если тело нагрето или имеется запас потенциальной энергии, когда тело сложное и состоит из отдельных связанных сжатыми пружинами частей), однако при переходе от системы отсчета, где это тело покоится, к системе, где оно движется поступательно равномерно и прямолинейно, внутренняя энергия принимается неизменной. Переход к равномерному движению означает признание появления у тела дополнительной, кинетической, энергии, которой неподвижное тело не обладает. Эта дополнительная энергия берется неизвестно откуда: с математической точки зрения преобразование систем отсчета—это просто эффект «поворота» координатных осей! Стало быть, налицо нарушение закона сохранения энергии!

Существование энергии покоя тела  $E_0 = mc^2$ , полной энергии  $E = mc^2\gamma$  и кинетической энергии  $T_{\text{кин}} = mc^2(\gamma - 1)$  «спасает» закон сохранения энергии.

Вывод о том, что всякое покоящееся тело или совокупность тел, центр масс которых не движется (и отсутствует вращение), обладает тем не менее энергией и энергией громадной, так как в выражение  $E_0 = mc^2$  входит квадрат скорости света, является одним из важнейших в физике.

Он лежит в основе объяснения и расчета энергии, выделяющейся при ядерных реакциях синтеза, когда сумма масс двух ядер атомов дейтерия больше массы ядра атома He, образующегося в результате слияния двух дейтронов. Выделяющаяся при таком слиянии энергия равна  $\Delta E = \Delta mc^2$ , где  $\Delta m = 2m_D - m_{\text{He}}$ —дефект массы.

Обычно, однако, не обращается внимание на то, что при более привычных чисто химических превращениях выделение энергии также подчиняется той же формуле. Так, и при образовании атома водорода из протона и электрона (или при образовании молекул из атомов) выделяется энергия  $\Delta E = (m_p + m_e - m_H)c^2$ , где  $m_p$ ,  $m_e$  и  $m_H$ —массы свободных протона, электрона и атома водорода. Энергия химических реакций мала по сравнению с энергией ядерных реакций, и

поэтому при химических превращениях дефекты масс также очень малы.

Меняются также, хотя и незначительно, массы тел при их нагревании и охлаждении и т. д.

«Уравнивание» массы и энергии, устанавливаемое формулой  $E = mc^2$ , позволяет дать новое определение массы как полной меры всех видов энергии, существующих в покоящемся теле. Это кинетическая энергия относительного движения составляющих тело частиц, потенциальная энергия их связи, энергия внутренних полей и т. д.

Становится понятным, почему громадный выигрыш в энергии, возникающей при слиянии двух дейтронов в He и связанный с тем, что ядерные силы велики и во много раз превосходят связывающие электрон и ядро кулоновские силы в атоме водорода, приводит к достаточно большому дефекту массы.

Масса системы, состоящей из  $n$  движущихся друг относительно друга равномерно и прямолинейно тел, равна

$$M = \left[ \left( \sum_{i=1}^n \frac{E_i}{c^2} \right)^2 - \left( \sum_{i=1}^n \frac{p_i}{c} \right)^2 \right]^{1/2}.$$

где  $\sum_{i=1}^n E_i$  — сумма энергий тел, а  $\sum_{i=1}^n p_i$  — их импульсов.

Для системы тел в релятивистской механике суммарная масса (ее надо связывать с центром масс и с системой отсчета, где

центр масс покоится)  $M \neq \sum_{i=1}^n m_i$ , однако, как и в классической

механике, выполняются соотношения

$$E_{\text{полн}} = \sum_{i=1}^n E_i, \quad p_{\text{полн}} = \sum_{i=1}^n p_i$$

т. е. энергия и импульс аддитивны.

Из формулы для полной энергии движущегося тела  $E = mc^2\gamma$  следует, что тела, для которых  $m \neq 0$ , не могут иметь скорость  $v = c$ , так как это приводит к бесконечно большой энергии, что невозможно. Если же  $m = 0$ , то энергия не обращается в бесконечность. В самом деле, при  $v \rightarrow c$  и  $m \rightarrow 0$  получается неопределенность типа  $0/0$ , что дает конечную величину. Это означает, что могут существовать объекты (фотоны, нейтрино), которые не в состоянии находиться в покое, а в любой системе отсчета всегда движутся со скоростью света.

Введение понятия импульса в общем виде, полной энергии и энергии покоя, обнаружение инвариантов относительно преобразования одной инерциальной системы к другой, запись уравнения

движения в форме  $\frac{dp}{dt} = F$  и введение силы как скорости изменения импульса и установление закона сохранения импульса, который верен для любой системы отсчета (как для той, где центр масс покоится, так и для той, где центр масс движется равномерно и прямолинейно), приводит к принципиальному заключению. Если механические явления подчиняются введенным выше законам, это означает, что проведенные в двух разных системах эксперименты должны дать одни и те же результаты. Значит, отличить одну систему от другой, производя внутри этих систем один и тот же эксперимент, нельзя. Это утверждение названо *принципом относительности*, который звучит следующим образом: *не существует такого эксперимента (причем не только механического), производя который можно отличить одну систему отсчета от другой, если они смещаются друг относительно друга равномерно и прямолинейно.*

### 3.5. Принцип эквивалентности

До сих пор мы исходили из того, что одна система отсчета движется по отношению к другой системе с постоянной скоростью. В этом случае уравнение движения будет иметь при переходе от одной системы отсчета к другой один и тот же вид. Представим теперь более общий случай, когда одна система отсчета движется по отношению к другой системе с переменной скоростью. В этом случае мировая линия будет искривленной и никаким поворотом осей не удастся свести ускоренное движение к равномерному или состоянию покоя. Чтобы разобраться в этом вопросе, вернемся к ньютоновской механике. Рассмотрим снова две системы координат. Одну условно примем за неподвижную и будем считать, что это нештрихованная система координат  $X, Y, Z$ ; вторую—за подвижную и обозначим соответствующие оси  $X', Y', Z'$ . Тогда для скоростей материальной точки, которые будут измерять наблюдатели в той и другой системах координат, получим  $v_x = v'_x + v_0$ , где  $v_x$ —скорость, измеряемая наблюдателем, условно принятым за неподвижного;  $v'_x$ —скорость, измеряемая исследователем, движущимся вместе с подвижной системой координат;  $v_0$ —скорость, отвечающая смещению вдоль оси  $X$  одной системы координат по отношению к другой. Если скорость  $v_0$  постоянна, то

$$a'_x = a_x = \frac{dv_x}{dt} = \frac{dv'_x}{dt}.$$

В ньютоновской механике силы одинаковы для двух наблюдателей (движущегося и неподвижного), поэтому мы получим одинаковые уравнения движения. Значит, результаты экспериментов, которые будут проводить наблюдатели, совпадут. Теперь представим, что

не постоянна (движение ускоренное), тогда  $a'_x \neq a_x$  и получаются два различных уравнения движения:

$$m \frac{dv_x}{dt} = F, \quad m \frac{dv'_x}{dt} + m \frac{dv_0}{dt} = F.$$

Таким образом, если для одного наблюдателя ничего не изменится, то другому наблюдателю будет казаться, что появилась некоторая дополнительная сила, пропорциональная массе. То, что такие силы при ускоренном движении действительно имеются, каждый из нас неоднократно ощущал. В самом деле, когда автомобиль резко трогается с места, то пассажиров прижимает к спинке сиденья; при торможении автомобиля, наоборот, тело подается вперед; при повороте автомобиля вправо возникает сила, которая действует на тело влево, и т. д. Все это убеждает в том, что наблюдатель, движущийся с ускорением, должен ощущать появление некоторых дополнительных сил, причем, что принципиально важно, сил, пропорциональных массе. А. Эйнштейн обратил внимание на то, что появляющиеся в ускоренных системах координат дополнительные силы, действующие на все тела пропорционально их массе и создающие как бы дополнительное поле, эквивалентны по своему проявлению действиям сил притяжения предметов, например к поверхности Земли.

На этом основании А. Эйнштейн сделал очень смелое предположение о том, что наш мир не прямолинеен, что его геометрия не есть простая прямоугольная геометрия Евклида, что в области, где расположены массивные тела, пространство искривляется и движение происходит уже не по прямым, а по некоторым кривым (геодезическим) линиям. Но в таком случае мы переходим к системам координат, не являющимся инерциальными, системам координат, которые движутся с ускорением. Согласно *принципу эквивалентности*, который ввел Эйнштейн, появление результирующих дополнительных сил в этом случае невозможно отделить с помощью какого-либо эксперимента от сил тяготения. Эйнштейн свел теорию тяготения к некоторой теории движения в искривленном пространстве. Эта идея лежит в основе *общей теории относительности*. Сопоставление кривизны траектории движения с массой источника гравитации и сведение сил тяготения к движению в пространстве более сложном, чем Евклидово, позволило преодолеть целый ряд принципиальных трудностей.

Например, если принять, что тяготение связано с искривлением пространства, то очевидно, что инертная и гравитационная массы должны совпадать и, более того, никакой гравитационной массы не существует, а есть только одна масса. И действительно, эксперименты показывают, что инертная и гравитационная массы совпадают между собой с очень большой точностью. Одним из следствий общей теории относительности, которое называется *гравитационным красным смещением* и которое действительно подтверждается экспериментом, является то, что свет, излучае-



мый массивными телами, должен иметь большую длину волны, чем свет, излучаемый менее массивными телами. Еще один эффект, предсказываемый общей теорией относительности и также проверенный экспериментально, заключается в искривлении светового луча, проходящего вблизи массивного тела.

#### Вопросы для самоконтроля

1. Сколько степеней свободы имеет твердое тело произвольной формы? Сколькими степенями свободы обладает бесконечно тонкая спица по отношению к внешним осям координат?

2. С помощью какого приема можно, зная законы движения материальной точки, получить на их основе законы движения твердого тела? Что такое парасила и для чего вводится это понятие? Какую роль играет понятие центра масс?

3. Чем отличается момент инерции точки, вращающейся вокруг некоторой оси, от момента инерции тела?

4. Как связана возможность устойчивого вращения твердого тела с расположением по отношению к этому телу главных осей инерции?

5. Почему, когда вращающийся на льду фигурист на коньках хочет остановиться, то разводит в стороны руки?

6. Как, располагая определенным количеством металла, создать маховик, который при заданном диаметре обладал бы максимальным моментом инерции?

7. Как поведет себя кабина вертолета, находящегося в полете, если по каким-либо причинам хвостовой винт перестанет работать?

8. Почему очень трудно удержаться на стоящем двухколесном велосипеде и совсем нетрудно сделать это, когда велосипед движется?

9. Как можно автоматически управлять движением ракеты в полете, поворачивая ее на вполне определенный заданный угол по отношению к внешним осям координат?

10. Как преобразуется уравнение движения в форме Ньютона, если мы переходим к описанию движения при очень больших скоростях?

11. Какова физическая причина того, что, даже если на материальную точку будет действовать длительное время постоянная сила, скорость движения материальной точки тем не менее не возрастет до бесконечности, а достигает предельной величины, равной скорости света?

12. Найти отношение релятивистского импульса движущейся материальной точки к классическому импульсу в тех случаях, когда скорость движения материальной точки меняется от 0,1 до 0,9с с шагом  $\frac{1}{10}$ . Построить соответствующий график.

13. Что такое дефект массы и какую роль он играет в ядерных реакциях? Могут ли существовать частицы с массой, равной нулю? Могут ли такие частицы находиться в покое? Если нет, то с какой скоростью они должны двигаться?

14. Что такое принцип эквивалентности и какую роль он играет в теории всемирного тяготения, развитой Эйнштейном?



## Г Л А В А 4

### КОЛЕБАНИЯ И ВОЛНЫ

*«Бросая в воду камешки, смотри на круги,  
ими образуемые; иначе такое бросание бу-  
дет пустою забавою».*

*К. Прутков*

#### 4.1. Общие положения

Колебательное движение является хотя и частным, но чрезвычайно важным и распространенным видом движения как в природе, так и в технике. Под *колебательным движением* мы будем понимать такое движение материальной точки или сложных систем, когда через определенные промежутки времени они возвращаются снова в свои исходные состояния или близкие к ним, так что процесс все время повторяется. Для того чтобы выполнялось это условие, т. е. чтобы система периодически возвращалась в свое первоначальное положение, необходимо существование сил, стремящихся вернуть ее в это состояние. Это условие, в свою очередь, выполняется, если потенциальная поверхность или функция имеет минимум и движение совершается в окрестностях этого минимума или в окрестности устойчивого равновесия. Примеры колебательных движений весьма многочисленны. Помимо хорошо всем известных разного рода маятников примерами колебательных движений являются, например, ходьба, биение сердца, миграция электромагнитной энергии в колебательном контуре и т. д. Не так давно русскими учеными В. Белоусовым и А. Жаботинским были обнаружены периодические химические реакции. Полагают, что подобного рода периодические химические реакции определяют существование в живых организмах биологических ритмов (биологических часов). Определенные ритмические колебательные процессы обнаружены в жизнедеятельности клеток.

Как и всегда, мы будем вначале пользоваться простейшими моделями. Пусть имеется материальная точка массой  $m$ , совершающая движение вдоль одной оси (одномерное движение) в окрестности некоторого равновесного состояния. Примем за начало отсчета положение минимума потенциальной кривой,

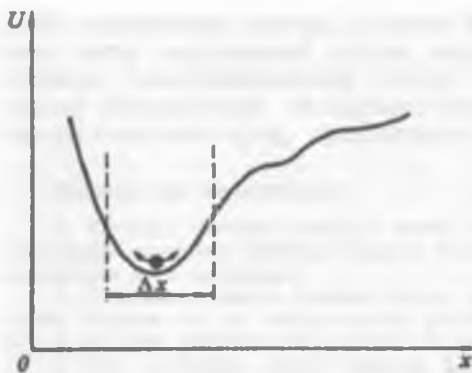


Рис. 4.1. Область колебательных движений  $\Delta x$

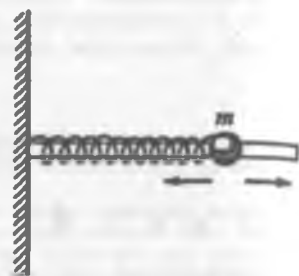


Рис. 4.2. Простой пружинный маятник

которая и определяет состояние равновесия. Точка будет совершать движение в пределах потенциальной ямы (рис. 4.1). Потенциальную яму любой формы без ограничения общности можно представить в виде

$$U = ax^2 + bx^3 + cx^4 + \dots,$$

где  $x$ —отклонение точки от положения равновесия (от дна потенциальной ямы). Тогда соответствующее уравнение движения материальной точки в области потенциальной ямы будет следующим:

$$m \frac{d^2 x}{dt^2} = -2ax - 3bx^2 - 4cx^3 - \dots$$

(см. формулу для связи между действующей силой и потенциальной функцией системы). Если отклонения от положения равновесия небольшие, то  $x^2 \gg x^3 \gg x^4$ . Тогда уравнение движения можно записать в упрощенном виде:

$$m \frac{d^2 x}{dt^2} = -2ax.$$

В качестве конкретного примера колеблющейся системы выберем шарик массой  $m$ , насаженный без трения на стержень и связанный пружиной со стенкой (рис. 4.2). Если шарик вывести из положения равновесия, то пружина растянется при движении вправо либо сожмется при движении влево. И в том и в другом случае возникает упругая сила, стремящаяся вернуть шарик в положение, соответствующее растянутой пружине.

Потенциальная функция пружины, подчиняющейся закону Гука, представляет параболу  $U = \frac{1}{2} kx^2$ . Такие параболические потенциалы получили название гармонических потенциалов, а соответствующие движения—гармонических колебаний. Гар-

Механические колебания возникают всегда, когда они совершаются с малыми отклонениями от положения равновесия, отвечающего минимуму потенциальных ям и поверхностей. Принципиальным является то, что любое другое более сложное колебательное движение может быть представлено как суперпозиция простых колебательных гармонических. Именно в этом заключается один из моментов, выделяющих колебательное гармоническое движение из всех остальных и делающих его наиболее важным. Для выяснения вопроса о характере гармонического движения рассмотрим, например, шарик на пружине (см. рис. 4.2) и запишем для него уравнение движения:

$$m \frac{d^2 x}{dt^2} = -kx.$$

В этом уравнении  $m$  — масса шарика,  $k$  — коэффициент упругости пружины,  $x$  — отклонение шарика от положения равновесия. Пружина считается невесомой. Так как возвращающая сила  $f = -dU/dx$  всегда направлена в сторону, противоположную смещению шарика из положения равновесия, то перед  $k$  получается минус. Перепишем это выражение в виде

$$\frac{d^2 x}{dt^2} = -\frac{k}{m} x, \quad \text{или} \quad \frac{d^2 x}{dt^2} = -\omega^2 x,$$

где  $\omega = \sqrt{k/m}$ .

В данном случае мы имеем уравнение, которое связывает вторую производную от смещения шарика из положения равновесия и само смещение. Это означает, что функция  $x = x(t)$  обладает тем свойством, что ее вторая производная по времени пропорциональна самой функции.

Таким свойством обладают функции  $y = \sin ax$  и  $z = \cos ax$ . Первые и вторые производные от этих функций имеют вид:

$$\frac{dy}{dx} = a \cos ax \quad \text{и} \quad \frac{dz}{dx} = -a \sin ax,$$

$$\frac{d^2 y}{dx^2} = -\frac{d}{dx} \left( \frac{dy}{dx} \right) = -a^2 \sin ax, \quad \frac{d^2 z}{dx^2} = -\frac{d}{dx} \left( \frac{dz}{dx} \right) = -a^2 \cos ax.$$

Таким образом, в результате двойного дифференцирования получаются снова исходные выражения с некоторыми коэффициентами. Эта особенность тригонометрических функций позволяет искать решение уравнения в форме

$$x = C_1 \sin \omega t + C_2 \cos \omega t.$$

Действительно, подставляя это выражение в исходное уравнение и учитывая, что

$$\frac{dx}{dt} = C_1 \omega \cos \omega t - C_2 \omega \sin \omega t,$$

$$\frac{d^2x}{dt^2} = \frac{d}{dt} \left( \frac{dx}{dt} \right) = -C_1 \omega^2 \sin \omega t - C_2 \omega^2 \cos \omega t,$$

получаем

$$\begin{aligned} \frac{d^2x}{dt^2} &= -C_1 \omega^2 \sin \omega t - C_2 \omega^2 \cos \omega t = \\ &= -\omega^2 x = -\omega^2 C_1 \sin \omega t - \omega^2 C_2 \cos \omega t. \end{aligned}$$

Таким образом, функция  $x = C_1 \sin \omega t + C_2 \cos \omega t$  есть решение уравнения.

Постоянные  $C_1$  и  $C_2$  находят из начальных условий. В качестве таких условий выбирают начальную скорость и начальное смещение шарика из положения равновесия. Если, например, начальное смещение шарика равно нулю, а начальная скорость шарика равна  $v_0$  (шарик подтолкнули), то должны выполняться равенства

$$x(t=0) = 0 = C_1 \sin(\omega \cdot 0) + C_2 \cos(\omega \cdot 0),$$

$$\begin{aligned} v(t=0) = v_0 &= \frac{dx}{dt}(t=0) = \\ &= C_1 \omega \cos(\omega \cdot 0) - C_2 \sin(\omega \cdot 0) = C_1 \omega, \end{aligned}$$

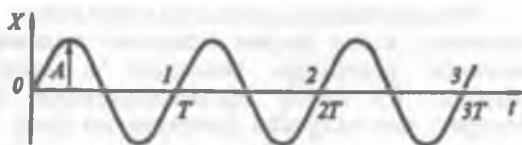
откуда  $C_2 = 0$  и  $C_1 = v_0/\omega$ .

В этом случае шарик совершает движения, описываемые функцией

$$x = C_1 \sin \omega t = \frac{v_0}{\omega} \sin \omega t.$$

Выясним смысл величины  $\omega$ . Для этого нарисуем график зависимости смещения шарика из положения равновесия во времени. Получим картину, изображенную на рис. 4.3. Мы видим, что шарик будет проходить положение равновесия через равные промежутки времени  $T$ , которые связаны с  $\omega$  соотношением  $\omega T = 2\pi$  или  $\omega = 2\pi/T$ . Промежуток времени между двумя совершенно одинаковыми состояниями шарика (точки 0—1, 1—2) называют *периодом колебаний*. Из графика видно, что шарик все время отклоняется до некоторого крайнего положения, затем возвращается к исходному положению, двигаясь в противоположную сторону на то же расстояние, снова возвращается к исходному положению и т. д. Максимальный отброс шарика из положения равновесия, т. е. высота пика на графике движения, равная  $C_1 = A$ , называется *амплитудой*

Рис. 4.3. График зависимости смещения колеблющейся точки из положения равновесия от времени



колебательного движения, а  $\omega = 2\pi/T$ —круговой частотой колебаний. Если, например, шарик переходит из положения 1 в 2 или из 2 в 3 за 1 с, то период колебаний равен 1 с, а  $\omega = 2\pi$  рад/с. При  $T = 0,5$  с шарик совершает два колебания в секунду и т. д.

В случае колебательного движения начало движения не обязательно совпадает с положением равновесия. Чтобы учесть это обстоятельство, можно ввести под знаком тригонометрической функции дополнительное слагаемое  $\varphi$ , называемое *начальной фазой колебания*. Выражение для смещения шарика из положения равновесия как функции времени в этом случае примет следующий вид:

$$x = A \sin(\omega t + \varphi).$$

Для того чтобы определить амплитуду  $A$  и коэффициент  $\varphi$ , необходимо задать начальные условия, например начальное смещение шарика из положения равновесия и начальную скорость движения шарика:

$$x(t=0) = x_0 = A \sin \varphi, \quad v(t=0) = v_0 = \frac{dx}{dt}(t=0) = A\omega \cos \varphi.$$

Тогда

$$\operatorname{tg} \varphi = \frac{\omega x_0}{v_0}, \quad A = \frac{x_0}{\sin \operatorname{arctg} \varphi}.$$

В случае простого гармонического движения легко найти максимальные значения скорости колебательного движения и ускорения. С помощью дифференцирования найдем

$$v_x = A\omega \cos(\omega t + \varphi), \quad a_x = -A\omega^2 \sin(\omega t + \varphi).$$

Таким образом, максимальная скорость, т. е. амплитуда скорости колебательного движения, равна произведению амплитуды колебательного движения на его частоту. Амплитуда ускорения колебательного движения равна произведению амплитуды колебательного движения на квадрат частоты.

Рассмотрение формул для смещения шарика из положения равновесия, скорости и ускорения позволяет сделать следующее заключение. Прежде всего мы видим, что если смещение шарика из положения равновесия пропорционально  $\sin \omega t$ , то скорость шарика пропорциональна  $\cos \omega t$ , а ускорение снова пропорционально  $\sin \omega t$ , но с обратным знаком.

Это означает, что максимальная скорость соответствует моменту, когда шарик проходит положение равновесия. Максимальное ускорение совпадает с максимальным отклонением шарика, но имеет противоположный знак. В таких случаях говорят, что скорость сдвинута по фазе на  $\pi/2$  по сравнению со смещением шарика, а ускорение сдвинуту по фазе на  $\pi$  по сравнению со смещением шарика.

## 4.2. Физический и математический маятники

Рассмотрим физический и математический маятники. *Физический маятник*—это любое твердое тело, подвешенное на оси таким образом, чтобы центр масс его был расположен ниже точки подвеса на  $l$  (рис. 4.4). Если вывести это тело из положения равновесия, т. е. отклонить в сторону, то возникнет за счет постоянно направленной вниз и приложенной к центру масс силы тяжести сила, стремящаяся вернуть тело в положение равновесия. Если после отклонения от положения равновесия предоставить телу возможность свободно двигаться, то оно под действием возвращающей силы будет совершать колебательные движения. Уравнение движения такого тела

$$I \frac{d\omega}{dt} = I \frac{d^2\beta}{dt^2} = N,$$

где  $\beta$ —угол между направлением вектора силы тяжести и прямой, проходящей через точку подвеса и центр масс;  $\omega$ —угловая скорость вращения;  $I$ —момент инерции тела относительно оси вращения;  $N$ —момент силы относительно той же оси. Так как векторы угловой скорости и момента силы имеют направления, совпадающие с осью вращения, то можно ограничиться скалярной записью уравнения движения.

Согласно общему правилу модуль момента силы равен произведению длины плеча, модуля действующей силы и синуса угла между направлением от центра подвеса к точке приложения силы и вектором силы. В данном случае действует лишь сила тяжести, которая всегда направлена вертикально вниз и приложена к центру масс. Тогда момент силы равен

$$N = Mgl \sin \beta.$$

Момент силы стремится вернуть тело в положение равновесия, т. е. действует в сторону, противоположную отклонению от него. Поэтому уравнение движения следует принять в форме

$$I \frac{d^2\beta}{dt^2} = - I M g \sin \beta.$$

Допустим, что маятник совершает малые колебания около положения равновесия. Известно, что в этом случае значение

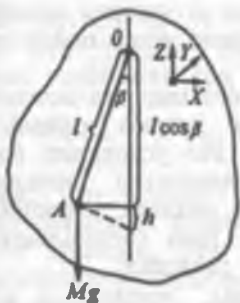


Рис. 4.4. Физический маятник (O—точка подвеса; A—центр масс)

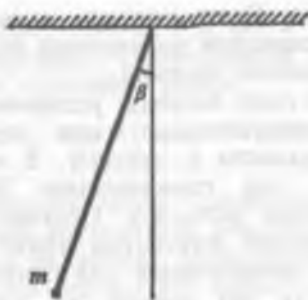


Рис. 4.5. Математический маятник

синуса с хорошей степенью точности равно значению угла. Учитывая это, перепишем уравнение движения в форме

$$I \frac{d^2\beta}{dt^2} = -lMg\beta \text{ или}$$

$$\frac{d^2\beta}{dt^2} = -\omega^2\beta,$$

где  $\omega = \sqrt{lMg/I}$ . Как видим, по форме это уравнение аналогично уравнению

$$\frac{d^2x}{dt^2} = -\omega^2x.$$

Следовательно, при тех же начальных условиях  $\beta = A \sin \omega t$ , причем  $\omega$  должна быть равна частоте гармонических колебаний системы. В этом случае частота гармонических колебаний оказывается зависящей от положения точки подвеса, расстояния от точки до центра масс, от массы тела, от ускорения свободного падения и от момента инерции тела.

Рассмотрим теперь простейший *математический маятник*. Он представляет собой материальную точку массой  $m$ , подвешенную на нити длиной  $l$  (рис. 4.5).

Для определения частоты колебаний математического маятника можно воспользоваться формулой для физического маятника, так как математический маятник является частным случаем физического маятника. Для такой системы момент инерции относительно оси подвеса равен  $I = ml^2$ .

Подставляя это выражение в формулу для частот колебаний физического маятника, находим

$$\omega = \sqrt{\frac{lmg}{I}} = \sqrt{\frac{lmg}{ml^2}} = \sqrt{\frac{g}{l}}.$$

Таким образом, частота колебаний определяется только ускорением свободного падения и длиной математического ма-



ятника. Если известны длина маятника и частота колебаний, то на основании полученной формулы можно легко найти ускорение свободного падения.

В свою очередь, ускорение свободного падения меняется, хотя и незначительно, при переходе от одного района земной поверхности к другому. В частности, это ускорение возрастает, если под поверхностью Земли расположено скопление руд тяжелых металлов. Поэтому принцип анализа вариаций частот колебаний маятников лежит в основе ряда специальных приборов, применяемых для геофизической разведки.

Как мы видели на всех примерах, частота  $\omega$  колебаний зависит только от свойств колеблющейся системы и не зависит ни от первоначальной энергии, которая была передана этой системе (при растягивании пружины, отклонении маятника от положения равновесия, за счет первоначального толчка), ни от начальных условий. Эти частоты, таким образом, являются характеристиками, присущими колебательной системе. Такие частоты получили название *собственных частот колебаний*. Независимость частоты или периода колебания любого маятника от того, каким образом этот маятник приведен в движение, дает основу для удобного отсчета времени. Тот факт, что период или частота есть величина постоянная, приводит к тому, что шкала времени может быть разбита на отдельные участки, отвечающие периоду колебания того или иного маятника, который принят за стандарт. В настоящий момент в качестве стандартного интервала принят интервал, отвечающий периоду колебаний электронов в атоме цезия-133. Этот интервал весьма мал, и в одной секунде укладывается  $9,19263177 \cdot 10^9$  этих интервалов. Выбор именно атомных процессов в качестве своеобразных часов для установления стандартных интервалов времени объясняется тем, что в этом случае удастся добиться максимальной стабильности хода таких часов.

### 4.3. Энергия гармонических колебаний

Определим энергию гармонических колебаний. Полная механическая энергия  $H$ , как известно, состоит из кинетической и потенциальной энергий. В данном конкретном случае потенциальная энергия равна энергии сжатой пружины или энергии математического или физического маятника, соответствующей поднятию центра массы тела на высоту  $h$ .

Для шарика на пружине энергия

$$H = T + U = \frac{1}{2} m v^2 + \frac{1}{2} k x^2$$

будет определяться равенством

$$H = \frac{1}{2} m A^2 \omega^2 \cos^2(\omega t + \varphi) + \frac{1}{2} k A^2 \sin^2(\omega t + \varphi).$$

Учитывая, что  $\cos^2(\omega t + \varphi) + \sin^2(\omega t + \varphi) = 1$  и  $\omega^2 = k/m$ , находим, что полная энергия системы, совершающей гармонические колебания под действием пружины, равна

$$H = \frac{1}{2} A^2 k = \frac{1}{2} A^2 m \omega^2.$$

Мы видим, что полная энергия гармонических колебаний не зависит от смещения шарика в каждый конкретный момент и не зависит от времени. Это постоянная величина, сохраняющаяся при гармоническом колебании. Энергия только все время переходит из одного вида в другой. В момент, когда шарик проходит положение равновесия, его потенциальная энергия равна нулю, во зато кинетическая энергия достигает максимума. Наоборот, в точке, в которой смещение шарика достигает амплитудных значений, максимальна потенциальная энергия шарика при нулевом значении кинетической энергии.

Аналогично выводится формула для полной энергии физического маятника:

$$\begin{aligned} H &= \frac{1}{2} I \left( \frac{d\beta}{dt} \right)^2 + Mgh = \\ &= \frac{A^2}{2} I \omega^2 \cos^2(\omega t + \varphi) + \frac{A^2}{2} Mgl \sin^2(\omega t + \varphi) = \frac{A^2}{2} Mgl. \end{aligned}$$

Здесь учтено, что  $h = l - l \cos \beta$  (см. рис. 4.5) и  $\cos \beta$  при малых  $\beta$  приближенно равен  $1 - \beta^2/2$ . Кроме того,  $\omega^2 = Mgl/I$ . Снова полная энергия для гармонического колебательного движения получается не зависящей от времени и от мгновенного положения физического маятника.

#### 4.4. Вынужденные колебания

Гармонические колебания, которые мы рассматривали до сих пор, совершаются под действием только внутренней силы—силы сжатия пружины или тяжести. В последнем случае, правда, необходимо рассматривать силу тяжести как внутреннюю возвращающую силу в системе маятник—Земля. Такое гармоническое движение называют *собственными колебаниями*. Получающаяся при этом частота колебаний называется *частотой собственных колебаний системы*. Частота собственных колебаний системы, как уже говорилось, тесно связана с ее физической природой и ею целиком определяется. Однако чрезвычайно важным и часто встречающимся в природе является колебательное движение, совершающееся под действием внешней периодической силы. Такое колебательное движение называется *вынужденным*.

Рассмотрим снова пример с шариком и пружинкой, но теперь будем считать, что дополнительно к системе приложена внешняя сила, периодически меняющаяся со временем:

$$F = F_0 \sin \omega t.$$

Уравнение движения в этом случае имеет вид

$$m \frac{d^2 x}{dt^2} = -kx + F_0 \sin \omega t,$$

или

$$\frac{d^2 x}{dt^2} = -\frac{k}{m} x + \frac{F_0}{m} \sin \omega t,$$

где  $\sqrt{k/m}$  — частота колебаний пружинного маятника в отсутствие внешней силы. Будем ее впредь обозначать символом  $\omega_0$ . Отношение  $F_0/m$  для сокращения записи обозначим  $f$ . Тогда уравнение колебаний маятника под действием внешней периодической силы имеет вид

$$\frac{d^2 x}{dt^2} = -\omega_0^2 x + f \sin \omega t.$$

В математике показывается, что решением такого уравнения является функция  $x = x(t)$ , равная сумме двух функций, а именно, решения уравнения  $d^2 x/dt^2 = -\omega_0^2 x$  и функции, пропорциональной  $\sin \omega t$ . (Уравнение  $d^2 x/dt^2 = -\omega_0^2 x + f \sin \omega t$  относится к типу неоднородных. Его решение есть сумма общего решения однородного уравнения  $d^2 x/dt^2 = -\omega_0^2 x$  и любого частного решения неоднородного. Таким частным решением, как легко убедиться, сделав необходимую подстановку, и будет  $x = C \sin \omega t$ .)

Следовательно, решение уравнения для колебания маятника под действием внешней силы имеет вид

$$x = C_1 \sin \omega_0 t + C_2 \cos \omega_0 t + C_3 \sin \omega t.$$

Подставим эту функцию в уравнение движения. Учитывая, что

$$\frac{d^2 x}{dt^2} = -\omega_0^2 C_1 \sin \omega_0 t - \omega_0^2 C_2 \cos \omega_0 t - \omega^2 C_3 \sin \omega t,$$

находим

$$\begin{aligned} \frac{d^2 x}{dt^2} &= -\omega_0^2 C_1 \sin \omega_0 t - \omega_0^2 C_2 \cos \omega_0 t - \omega^2 C_3 \sin \omega t = \\ &= -\omega_0^2 C_1 \sin \omega_0 t - \omega_0^2 C_2 \cos \omega_0 t - \omega_0^2 C_3 \sin \omega t + f \sin \omega t = \\ &= -\omega_0^2 x + f \sin \omega t. \end{aligned}$$

Сравнивая средние части равенства, заключаем, что

$$-\omega^2 C_3 \sin \omega t = -\omega_0^2 C_3 \sin \omega t + f \sin \omega t,$$

$$\text{или } C_3 = f/(\omega_0^2 - \omega^2).$$

Постоянные  $C_1$  и  $C_2$ , как и ранее, находятся по заданным начальным условиям. Если в начальный момент шарик находился в положении равновесия  $x = 0$  и скорость его также была равна нулю ( $v_0 = 0$ ) при  $t = 0$ , то

$$x_0 = 0 = C_1 \sin(\omega_0 \cdot 0) + C_2 \cos(\omega_0 \cdot 0) + C_3 \sin(\omega \cdot 0) = C_2,$$

$$v_0 = \frac{dx}{dt}(t = 0) = \omega_0 C_1 \cos(\omega_0 \cdot 0) + \\ + \omega_0 C_2 \sin(\omega_0 \cdot 0) + \omega C_3 \cos(\omega \cdot 0) = \omega_0 C_1 + \omega C_3.$$

Отсюда  $C_2 = 0$  и  $C_1 = -(\omega/\omega_0)C_3$ . Но  $C_3 = f/(\omega_0^2 - \omega^2)$ . Стало быть,  $C_1 = -\omega f/[\omega_0(\omega_0^2 - \omega^2)]$  и решением уравнения при заданных начальных условиях является функция

$$x = -\frac{\omega f}{\omega_0(\omega_0^2 - \omega^2)} \sin \omega_0 t + \frac{f}{\omega_0^2 - \omega^2} \sin \omega t.$$

Первый член в этом выражении описывает колебательные движения с частотой  $\omega_0$ . Они называются *свободными сопровождающими колебаниями*. Второй член описывает колебания с частотой  $\omega$  внешней силы. Они называются *вынужденными колебаниями*.

Таким образом, общее движение складывается из двух гармонических движений: сопровождающих свободных колебаний и вынужденных колебаний.

В реальных колеблющихся системах всегда присутствует трение, которое не учтено в записанном выше уравнении движения. Это трение приводит к тому, что постепенно сопровождающие свободные колебания затухают и остаются только вынужденные колебания с частотой  $\omega$  внешней силы. Такой режим движения называется *установившимся*. Амплитуда колебаний при этом оказывается равной  $A = f [(\omega_0^2 - \omega^2)^2 + 4\delta^2 \omega^2]^{-1/2}$ , где  $\delta$ —коэффициент трения. При приближении частоты  $\omega$  возбуждающей силы к частоте  $\omega_0$  собственных колебаний системы эта амплитуда будет возрастать, достигая предельной величины  $A = f/(2\delta\omega)$ . Если трение отсутствует, то в области  $\omega \approx \omega_0$  амплитуда будет линейно возрастать со временем, а само движение будет иметь вид, показанный на рис. 4.6. Такое явление, заключающееся в резком возрастании амплитуды вынужденного колебания при приближении частоты

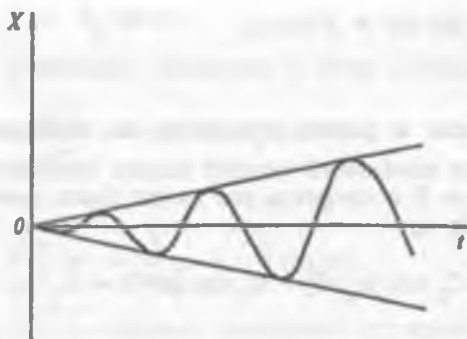


Рис. 4.6. График вынужденных колебаний точки при  $\omega = \omega_0$

вынуждающей силы к собственной частоте колебаний маятника, называется *резонансом*.

Выведем формулу для вынужденных колебаний маятника в области резонанса, т. е. в области  $\omega \approx \omega_0$ . Представим  $\omega = \omega_0 \pm \Delta\omega$ , где  $\Delta\omega$ —малая величина. Тогда общее решение уравнения движения маятника под действием внешней периодической силы без трения имеет вид

$$\begin{aligned} x &= -\frac{\omega}{\omega_0} \frac{f}{\omega_0^2 - \omega^2} \sin \omega_0 t + \frac{f}{\omega_0^2 - \omega^2} \sin (\omega_0 \pm \Delta\omega)t = \\ &= -\frac{\omega}{\omega_0} \frac{f}{\omega_0^2 - \omega^2} \sin \omega_0 t + \frac{f}{\omega_0^2 - \omega^2} \sin \omega_0 t \cos \Delta\omega t \pm \\ &\quad \pm \frac{f}{\omega_0^2 - \omega^2} \sin \Delta\omega t \cos \omega_0 t \end{aligned}$$

(использована формула синуса суммы или разности двух углов).

Пусть теперь  $\Delta\omega \rightarrow 0$  (или  $\omega \rightarrow \omega_0$ ). Тогда  $\lim_{\Delta\omega \rightarrow 0} (\omega/\omega_0) = 1$ .

Кроме того,  $\lim_{\Delta\omega \rightarrow 0} \cos (\Delta\omega t) = 1$  при  $t$  конечном. В результате при  $\Delta\omega \rightarrow 0$  получаем, учитывая, что при малых  $\Delta\omega \sin \Delta\omega t \approx \Delta\omega t$ :

$$x = \pm \frac{f}{\omega_0^2 - \omega^2} \Delta\omega t \cos \omega_0 t.$$

Вспомним, что  $\omega = \omega_0 \pm \Delta\omega$ , и тогда  $\omega_0^2 - \omega^2 = \mp 2\omega_0 \Delta\omega - \Delta\omega^2$ . Поэтому в области  $\Delta\omega \approx 0$   $\Delta\omega^2 \ll \Delta\omega$  и с хорошей точностью можно записать  $\omega_0^2 - \omega^2 = \mp 2\omega_0 \Delta\omega$ . Подставляя это выражение в формулу для  $x$ , найдем окончательно, что при  $\omega = \omega_0$   $x = -ft \cos \omega_0 t / (2\omega_0)$ , т. е. ту зависимость  $x = x(t)$ , которая показана на рис. 4.6.

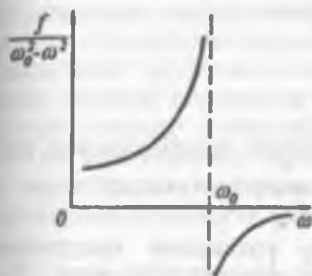


Рис. 4.7. Зависимость множителя  $f/(\omega_0^2 - \omega^2)$  от  $\omega$

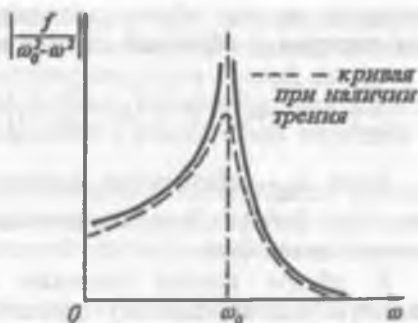


Рис. 4.8. Резонансные кривые

При наличии трения также наблюдается возрастание амплитуды вынужденных колебаний (если трение все же не очень велико), однако в этом случае подобное возрастание связано только с разностью квадратов частот  $(\omega_0^2 - \omega^2)$  и амплитуда при установившихся колебаниях не будет зависеть от времени.

График значения множителя  $f/(\omega_0^2 - \omega^2)$  показан на рис. 4.7. Видно, что он меняет свой знак справа и слева от частоты  $\omega$ , равной частоте  $\omega_0$  собственных колебаний системы. Это означает, что если в одном случае внешняя сила раскачивает систему, то в другом случае она начинает ее тормозить. Если нарисовать график зависимости амплитуды [модуля величины  $f/(\omega_0^2 - \omega^2)$ ] от  $\omega$ , то получится кривая (рис. 4.8), показанная сплошной линией.

Наряду с резонансом интенсивное возбуждение колебаний возможно и в автоколебательных процессах. Такие колебания при определенных условиях возникают в системах, которые могут воздействовать сами на себя через посредство специальной связи, которая называется *обратной*. Типичной системой такого рода являются часы с анкерным спуском. Колебания маятника в этих часах воздействуют периодически на механизм, обеспечивающий поступление в колеблющуюся систему дополнительной энергии от некоторого источника (поднятой гири или взведенной пружины). Назначение обратной связи заключается в том, чтобы, восприняв от колеблющейся системы сигнал, пропорциональный, например, ее смещению из положения равновесия, скорости или ускорению, преобразовать его к нужному виду и передать снова в систему в форме дополнительной силы, являющейся заданной функцией исходного смещения или его производных. Разумеется, для создания такой системы требуется постоянный источник энергии. Обратная связь как раз и управляет подачей в колеблющуюся систему порций энергии по заданному закону.

Опираясь на эти общие сведения, запишем уравнение движения для системы с обратной связью в форме

$$\frac{d^2x}{dt^2} = -\omega_0^2 x + F\left(x, \frac{dx}{dt}, \dots\right).$$

Здесь  $\omega_0$ —собственная частота свободно колеблющейся системы,  $F(x, dx/dt, \dots)$ —дополнительная, воздействующая через обратную связь сила.

В общем случае решение этого уравнения представляет сложную задачу, поэтому ограничимся частным примером. Представим пружинный маятник, на конце которого находится полосоной постоянный магнит, который при колебаниях маятника может вдвигаться в соленоид и выдвигаться из него. Из школьного курса физики известно (вспомним опыты Фарадея), что при движении магнита в соленоиде будет индуцироваться электрический ток, сила которого зависит от скорости движения магнита. Этот ток может быть усилен специальным усилителем и подан на исполняющее устройство, которое и будет подталкивать магнит с силой, пропорциональной скорости его движения.

Уравнение движения для такой автоколебательной системы примет вид

$$\frac{d^2x}{dt^2} = -\omega_0^2 x + k \frac{dx}{dt}, \quad \text{или} \quad \frac{d^2x}{dt^2} - k \frac{dx}{dt} + \omega_0^2 x = 0.$$

Это линейное дифференциальное уравнение второго порядка с постоянными коэффициентами. Его решения будут иметь разный вид в зависимости от соотношения постоянных  $\omega_0^2$  и  $k$ . В частности, если  $4\omega_0^2 > k$  и  $k > 0$  (положительная обратная связь), то

$$x = e^{kt/2}(C_1 \sin \omega t + C_2 \cos \omega t),$$

где  $\omega = \frac{1}{2}(4\omega_0^2 - k^2)^{1/2}$ . Видно, что при  $k = 0$  имеем уже привычное решение для свободно колеблющейся системы, совершающей малые гармонические колебания.

Если  $x = 0$  при  $t = 0$ , но  $v = v_0$ , то  $C_2 = 0$  и  $C_1 = v_0/\omega$ . Тогда

$$x = (v_0/\omega)e^{kt/2} \sin \omega t.$$

Значит, в системе после начального толчка возникнут колебания с частотой  $\omega$  и с возрастающей во времени по закону экспоненты амплитудой  $A = e^{kt/2}$ . Получим явление, напоминающее резонанс.

Значительное возрастание амплитуды при автоколебаниях и при резонансе возможно только при постоянном подводе внешней энергии. Амплитуда перестает возрастать, если потери энергии на преодоление сопротивления колебаниям (например, при преодолении трения) становятся сравнимыми с получаемой энергией от дополнительного источника.

Как резонансные, так и автоколебательные процессы являются в природе и технике основными причинами появления устойчивых периодических движений достаточно большой амплитуды.

В ряде случаев эти явления могут быть полезными, в ряде случаев вредными. Так, например, если случайно какой-нибудь механизм, содержащий вращающиеся части, плохо отцентрирован, то при каждом повороте вращательной части возникает удар по оси вращения. Следовательно, при вращении возникает некоторая периодическая возбуждающая сила, передаваемая всему механизму. В тех случаях, когда частота собственных колебаний механизма и частота вращения движущейся части близки между собой, возникает сильное раскачивание системы. Это в свое время вызвало большое количество затруднений при построении вертолетов. При вращении винтов вертолета создавалась сильная вибрация его корпуса, которая приводила к разрушению всей конструкции.

Явление автоколебаний лежит в основе флаттера. Явление это заключается в следующем. В тех случаях, когда скорость самолета приближается к скорости звука в воздухе и превосходит ее, возникает резкое возрастание сопротивления воздуха движению крыла самолета и крыло начинает двигаться как бы в жидкости. В этом случае могут возникнуть колебания крыла самолета. Такое явление легко наблюдать, если взять доску, вставить ее в поток воды под некоторым углом к потоку и двигать эту доску навстречу течению. При некоторых углах поворота этой доски относительно направления движения нетрудно ощутить возникновение колебательных движений. Если собственные колебания крыла самолета совпадут с возбуждаемыми встречным потоком воздуха колебаниями крыла, то амплитуда этих колебаний начнет быстро возрастать, что может привести к разрушению всей системы и катастрофе самолета.

Известно, что когда воинская часть проходит через не очень прочный мост, то всегда дается команда идти не в ногу. Это связано с тем, что при прохождении воинской части под действием шагов всех солдат, совершаемых в такт, возникает определенная, периодически действующая на мост сила. Может оказаться, что частота ударов ног солдат совпадет с частотой собственных колебаний моста. Мост начнет раскачиваться, что может привести к его разрушению.

Вместе с тем резонанс и автоколебания можно использовать и в полезных целях, а именно во всех тех случаях, когда задачей является раскачивание системы, приведение ее в состо-



яние колебаний. Уже указывалось, что автоколебательные раскачивания совершаются в часовых механизмах, имеющих маятники. Резонансные движения совершаются в простейших случаях катания на качелях. Хорошо известно, что для раскачивания качелей необходимо выбрать определенный такт движений. Этот такт движений должен совпадать с собственной частотой колебаний качелей вместе с человеком.

Явление резонанса наблюдается не только в технике, но и в живых организмах. Например, известно, что каждому человеку свойственна своя индивидуальная частота шага. При этом, как правило, мужчины шагают реже, чем женщины. Это явление легко понять, если обратить внимание на то, что ходьба представляет колебательное движение ног человека около подвеса, которым является место соединения бедра со скелетом. Чем более массивной и длинной является нога, тем больше она имеет момент инерции относительно точки подвеса. Следовательно, в согласии с формулой для физического маятника тем меньше должна быть собственная, или естественная частота колебаний ног.

Мы выше сказали, что при резонансе затрачивается наименьшее усилие для приведения системы в колебательное движение. Именно потому природа и отбирает такой такт ходьбы, который отвечает резонансным колебаниям ног движущегося животного.

Например, в спокойном состоянии крупные собаки шагают значительно реже мелких.

Эта же закономерность проявляется и в работе сердца, что тоже является примером колебательной системы. У крупных животных частота биения сердца всегда меньше, чем у мелких животных. Это также связано с тем, что колеблющаяся масса сердца крупного животного больше, чем у мелких животных, что и приводит к разности резонансных частот.

Выше мы уже касались важного вопроса о работе внешних сил при вынужденных колебаниях и, следовательно, вопроса об энергии, затрачиваемой внешним источником для поддержания таких колебаний.

Можно показать, что при наличии трения, когда система находится в состоянии установившихся вынужденных колебаний, работа внешней силы целиком связана только с преодолением сил трения. Для того чтобы такие колебания поддерживались длительное время, надо все время в систему подводить энергию, компенсирующую потери ее на трение, которые растут пропорционально амплитуде колебаний. При длительных колебаниях общий расход энергии внешнего источника может стать очень большим. Если запас энергии у внешнего источника ограничен, то это приведет рано или поздно к уменьшению амплитуды внешней силы и, следовательно, к уменьшению амплитуды вынужденных колебаний.

Как уже указывалось, в системах без трения или с малым трением в области резонанса ( $\omega_0 = \omega$ ) колебания будут быстро

возрастать по амплитуде. Для обеспечения этого возрастания и используется энергия внешнего источника. Так как в любой колеблющейся системе всегда имеются потери энергии на трение, то реальная резонансная кривая будет иметь вид, показанный на рис. 4.8 пунктиром. Эта кривая уже не достигает бесконечно больших значений (разрыва) в области резонанса.

Перейдем теперь к изучению чрезвычайно важного случая колебаний двух и более связанных маятников и волновых процессов.

## 4.5. Связанные маятники. Волны

Рассмотрим два математических маятника одинаковой длины, подвешенных на гибкой нити. Отклоним один из маятников из положения равновесия и дадим ему возможность свободно раскачиваться. Будем наблюдать следующее явление. Через несколько качаний первого маятника второй маятник, который по условию имеет ту же длину, что и первый, тоже начнет раскачиваться с той же частотой. Постепенно амплитуда его колебаний будет увеличиваться и одновременно начнет уменьшаться амплитуда колебаний первого маятника. В конце концов амплитуда колебаний второго маятника достигает максимальной величины и при этом первый маятник почти остановится. Затем весь процесс начнет развиваться в обратном направлении, т. е. начнет возрастать амплитуда колебаний первого маятника и уменьшаться амплитуда колебаний второго маятника.

Следовательно, колебательная энергия, первоначально находившаяся в первом маятнике, начнет переходить во второй маятник и обратно. В самом деле, при колебаниях первого маятника за счет связи его через гибкую нить с основанием второго маятника колебательное движение передается второму маятнику. Так как первый и второй маятники совершенно одинаковы, то они имеют и одинаковые собственные частоты. Поэтому колебательное воздействие первого маятника приводит к резонансному раскачиванию второго маятника. Однако такое резонансное раскачивание требует постоянного подведения внешней энергии. Эта энергия берется от первого маятника, в результате, очевидно, должна уменьшаться энергия первого маятника, т. е. уменьшаться амплитуда его колебаний. После того, как вся энергия перейдет ко второму маятнику, он начинает играть роль возбуждающей системы и раскачивает первый маятник. Снова возникает резонансный процесс, и энергия от второго маятника возвращается в первый маятник. Такая передача энергии от одного маятника к другому будет происходить с наибольшей легкостью именно при одинаковых маятниках. Если маятники отличаются друг от друга, то передача энергии будет сильно затруднена и хотя второй маятник и начнет раскачиваться, но с очень небольшой амплитудой.

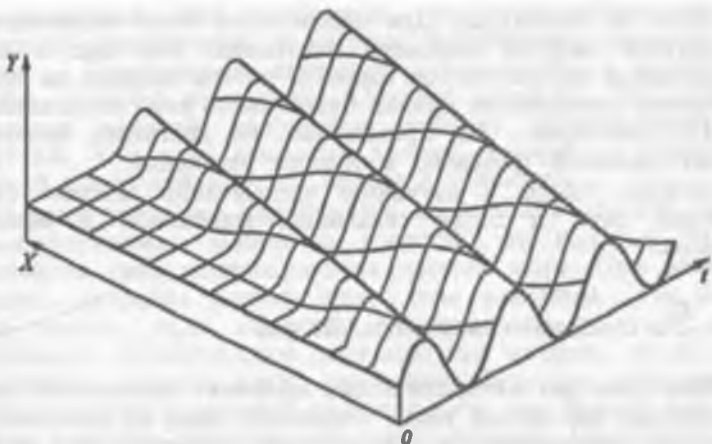


Рис. 4.9. Трехмерное изображение функций, отвечающей бегущей волне

Систему можно усложнить и взять вместо пары маятников ряд маятников, одинаково связанных между собой. В этом случае, если раскачивать первый маятник, то постепенно его энергия приведет в движение второй маятник. Колебательное движение второго маятника, в свою очередь, передается третьему маятнику и т. д. Начнется процесс перекачивания энергии вдоль по цепи, состоящей из целого ряда маятников. Такой процесс называется *волновым процессом передачи энергии*. Каждый из маятников при этом остается на своем месте. Передача энергии выражается в том, что первоначально неподвижный маятник постепенно переходит в колебательное движение. Изобразим смещение одного из маятников из положения равновесия символически  $u$ . Координатой  $x$  будем обозначать положение маятника на оси абсцисс. Если первый маятник начнет колебаться согласно закону  $u = A \sin \omega t$ , то постепенно другие маятники придут в колебательные движения с той же частотой (все маятники по условиям одинаковы), однако эти колебательные движения возникнут не сразу, а с известным запаздыванием по отношению к первому маятнику. Это запаздывание характеризуется определенной фазой. В свою очередь, фаза (отставание) зависит, с одной стороны, от расстояния данного маятника от первого (пропорциональна этому расстоянию или координате  $x$ ), а во-вторых, будет тем больше, чем меньше скорость распространения энергии вдоль по цепи связанных маятников. Изобразить в этом случае колебания всех связанных маятников можно с помощью формулы

$$u = A \sin \omega(t - x/v).$$

Наглядное трехмерное изображение этой функции показано на рис. 4.9. Отношение  $x/v$  имеет смысл *времени запаздывания*, которое характеризует отставание колебаний данного маятника,

расположенного в точке  $x$ , по отношению к первому маятнику, расположенному в начале координат ( $x = 0$ ). Эта формула представляет аналитическую запись волны. В самом деле, если рассматривать систему одинаковых маятников, то при передаче энергии от первого маятника вдоль по цепи все маятники будут совершать колебательное движение, напоминающее распространение волны вдоль поверхности воды.

Смысл отдельных членов в этой формуле следующий. Как уже указывалось,  $y$ —это смещение каждого из маятников из положения равновесия;  $\omega$ —частота колебаний маятника;  $t$ —текущая координата;  $x$ —координата, определяющая положение данного маятника по отношению к первому;  $v$ —скорость распространения энергии волны вдоль системы маятников;  $A$ —амплитуда колебаний. Предполагается, что  $A = \text{const}$ .

Простейшие волновые процессы, описываемые выражением

$$y = A \sin \omega(t - x/v),$$

обычно возникают при сильной связи между маятниками, т. е. когда каждый последующий маятник вовлекается в колебательное движение уже при первом качании предыдущего маятника.

Так же как и в случае двух маятников, происходит резонансная передача энергии и выполняется закон сохранения энергии. Энергия распространяется вдоль цепи маятников, но не исчезает и не возникает ни в одной из частей системы. Если рассмотреть такой частный случай, когда первый маятник все время приводится в колебательное движение внешней возбуждающей силой, то будет наблюдаться и определенный постоянный поток энергии вдоль цепи из совокупности маятников. Так как энергия колебания каждого маятника пропорциональна квадрату амплитуды его колебания, то энергия волны также должна быть пропорциональна квадрату ее амплитуды.

Для того чтобы такое распространение энергии было максимально легким, необходимо, очевидно, чтобы все маятники были одинаковыми. Если встретятся на пути волны отличающиеся по своим частотам маятники, то это затруднит прохождение волны дальше и приведет к появлению отраженной волны, движущейся в противоположную сторону.

Рассмотренная совокупность маятников является наглядной моделью, которая позволяет понять распространение энергии в твердых телах, в воздухе, в жидкости. Передача энергии с помощью такого колебательного движения вдоль системы связанных между собой упругими силами частиц получила название *бегущей волны*.

Для того чтобы поддерживать бегущую волну, необходимо все время вводить энергию в систему. Если первоначально были возбуждены колебания в какой-то части среды и затем энергию извне перестали подводить, то постепенно колебательная энергия начнет переводить в колебательное движение все большее и большее количество частиц и колебательное возмущение будет

распространяться во все стороны от места возмущения. Так как запас первоначальной энергии конечен, то постепенно с ростом числа маятников, участвующих в колебании, будет падать амплитуда этих колебаний и волны начнут затухать.

В волновых процессах, которые характерны прежде всего для сплошных сред (твердые тела, жидкости), мы встречаемся с новым видом движения, а именно с движением энергии без одновременного переноса массы. Но имеется и нечто общее. В самом деле, движущееся тело также переносит энергию как в виде кинетической энергии, так и в форме полной энергии, заключенной в его массе. Движущуюся массу тела с присущей ей кинетической энергией можно трактовать как перемещающийся в пространстве сгусток энергии. В свою очередь, и волновой процесс передачи энергии может происходить таким образом, что будет распространяться не синусоидальная волна, а «горб». Такой горб можно трактовать как сгусток колебательной энергии. Получится некая аналогия с перемещающимся телом. Поэтому такой сгусток энергии или волновой пакет получил название *квазичастицы*. Представление о квазичастицах играет в современной физике очень большую роль.

## 4.6. Типы волн

Ограничимся пока исследованием волновых движений в непрерывных средах. Такими непрерывными средами являются твердые тела или жидкости. Все эти объекты состоят из частиц, которые расположены близко друг к другу и для которых возможны небольшие смещения из положений их равновесия или мгновенных состояний. Причем взаимодействие частиц приводит к тому, что смещение части из них затрагивает соседние частицы. За счет этого и происходит передача энергии. Каждую такую частицу можно рассматривать как некий маленький маятник (осциллятор), а весь объект как совокупность осцилляторов. Если первоначально возбуждаются колебания небольшой группы таких осцилляторов, то в результате начинается распространение энергии в твердом теле или в жидкости, т. е. совершается волновое движение.

Различают следующие типы волн:

*одномерные волны*; примером таких волн являются волны, создаваемые в струне или натянутом жгуте;

*поверхностные (двумерные) волны*—это волны, возникающие, например, на поверхности жидкости;

*объемные (трехмерные) волны*—это волны, которые возникают при распространении звука в жидкости или в твердых телах.

Принято различать также волны продольные и поперечные. *Продольными* называются волны, при которых колебательные движения отдельных осцилляторов, составляющих среду, совершаются в том же направлении, в котором распространяется энергия. *Поперечными* называются волны, при которых коле-

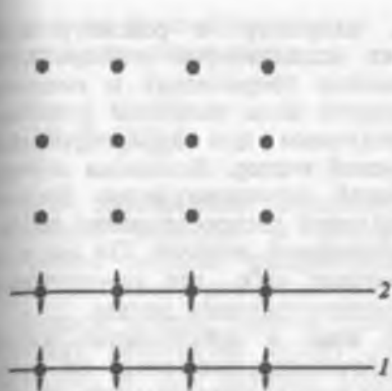


Рис. 4.10. Схема образования плоского фронта волны

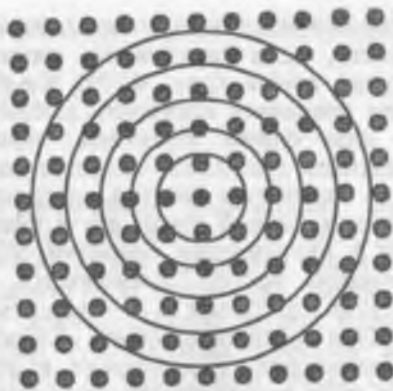


Рис. 4.11. Схема образования сферического фронта волны

бания осцилляторов происходят перпендикулярно направлению, вдоль которого передается энергия. Примером поперечных волн может служить волна, возбуждаемая в струне, когда движение частиц, составляющих струну, совершается вверх-вниз, а движение энергии—вдоль струны. Звуковая волна, распространяющаяся в жидкости,—пример продольных волн. Эти два типа волн являются крайними случаями. Все другие колебания можно представить суммой продольного и поперечного колебаний.

С понятием продольных и поперечных волн связано и понятие поляризации. Выделяют два типа поляризации. Мы говорим о *поперечной поляризации* в тех случаях, когда колебательное движение осцилляторов совершается перпендикулярно направлению движения волн, и о *продольной поляризации*, когда движение осцилляторов совершается в направлении распространения энергии.

Важным является понятие *фронта волны*. Чтобы ввести это понятие, представим среду, состоящую из отдельных молекул или атомов, имитирующих набор осцилляторов (рис. 4.10). Допустим, что вначале были возбуждены колебания осцилляторов, соединенных между собой некоторой линией 1. Через некоторое время колебания этих осцилляторов передаются следующей группе осцилляторов, соединенных прямой 2, затем придут в движение осцилляторы следующего слоя и т. д. Назовем поверхность, проведенную через совокупность колеблющихся с одинаковыми фазой и амплитудой осцилляторов, *фронтом волны*. В общем случае фронт волны может иметь весьма сложный вид, однако принято различать два крайних случая: *плоские* и *сферические* фронты. *Плоскими волнами* называются такие волны в пространстве, у которых фронт волны представляет плоскость. *Сферические волны* имеют фронт в виде сферической поверхности.

Такие волны могут возникнуть, например, в рассмотренной модели, если раскачивать один из осцилляторов, находящихся внутри среды (рис. 4.11). Комбинация сферических и плоских волн позволяет построить любые другие виды волновых фронтов.

Еще одним понятием, используемым для характеристики волнового движения, является волновой вектор. Волновым вектором называется вектор, проведенный перпендикулярно фронту волны в данной точке его в направлении распространения волны или в направлении передачи колебательной энергии. По модулю волновой вектор принимается равным  $\omega/v$ , где  $\omega$ —круговая частота колебаний и  $v$ —скорость распространения волны.

При распространении плоских волн в пространстве число осцилляторов, вовлекаемых в колебательное движение, на каждом шаге распространения волны все время остается одним и тем же, поэтому и энергия, необходимая для раскачивания новой совокупности осцилляторов, будет сохраняться одной и той же во все время распространения волны. Поскольку энергия колебания связана с их амплитудой, то при распространении плоских волн амплитуда волны, если только исключить разного рода постоянные потери, например на трение, будет сохраняться все время постоянной. В случае сферической волны число частиц, вовлекаемых при распространении волны в колебательное движение, растет пропорционально площади сферы радиуса  $r$ . На колебание каждой частицы среды с амплитудой  $A$  затрачивается энергия, пропорциональная  $A^2$ . Так как в отсутствие потерь на трение общая энергия должна быть постоянной, то выполняется соотношение  $r^2 A^2 = \text{const}$ . Стало быть, амплитуда колебания частиц сферической волны падает пропорционально  $1/r$ . Тогда для сферической волны, если радиус-вектор текущей частицы проводить из центра волны (из источника колебательного движения), можно записать

$$y = \frac{1}{r} \sin \omega \left( t - \frac{|r|}{v} \right).$$

Вернемся к выражению  $y = A \sin \omega(t - x/v)$ , которое может быть интерпретировано как аналитическая запись плоской волны, распространяющейся вдоль оси  $X$ . Величина  $y$  зависит от двух переменных:  $t$ —времени и  $x$ —расстояния от источника волны до точки, в которой производится наблюдение. Найдем частные производные от  $y$  по  $x$  и по  $t$  до второго порядка:

$$\frac{\partial y}{\partial x} = -A \frac{\omega}{v} \cos \omega \left( t - \frac{x}{v} \right), \quad \frac{\partial^2 y}{\partial x^2} = -A \frac{\omega^2}{v^2} \sin \omega \left( t - \frac{x}{v} \right),$$

$$\frac{\partial y}{\partial t} = A \omega \cos \omega \left( t - \frac{x}{v} \right), \quad \frac{\partial^2 y}{\partial t^2} = -A \omega^2 \sin \omega \left( t - \frac{x}{v} \right).$$

Сравнивая частные производные второго порядка от  $y$ , т. е. от смещения осциллятора из положения равновесия в бегущей волне,

по  $x$  и по  $t$ , из приведенных выше равенств заключаем, что имеет место соотношение

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 y}{\partial t^2}.$$

Это соотношение называется *волновым уравнением*. Это уравнение позволит в дальнейшем показать, что электромагнитная энергия распространяется в пространстве также в форме волнового движения электрических и магнитных полей.

Введем теперь понятие длины волны. *Длиной волны* будем называть расстояние между такими ближайшими точками на волне ( $x_1$  и  $x_2$ ), которые колеблются в одинаковой фазе. Это означает, что углы, стоящие под знаком  $\sin$ , различаются друг относительно друга для этих точек только на  $2\pi$ , поэтому  $\omega(t - x_1/v) = \omega(t - x_2/v) + 2\pi$ . Отсюда сразу можно найти, что

$$\lambda = x_1 - x_2 = 2\pi v/\omega.$$

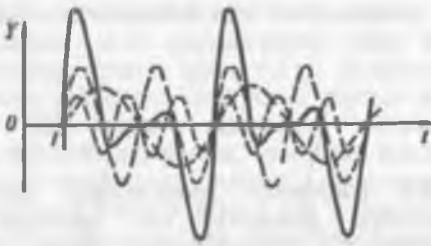
Таким образом, наименьшее расстояние между двумя одинаково колеблющимися осцилляторами в бегущей волне, или длина волны, пропорционально скорости распространения волны и обратно пропорционально частоте колебаний.

Дадим теперь некоторое наглядное толкование скорости распространения волны. Вспомним волны, возникающие на поверхности воды, если в нее бросить камень. Они будут иметь вид расходящихся от места падения камня концентрических окружностей. Нетрудно видеть, что гребень волны (максимум волны) или провал между двумя гребнями (минимум) будет распространяться по поверхности воды с определенной скоростью. Опираясь на аналитическую запись волны  $y = A \sin \omega(t - x/v)$ , найдем, что между текущей координатой  $x$  и временем для той ситуации, когда  $y = 0$ , т. е. для минимума волны, должно выполняться соотношение  $\omega(t - x/v) = 0$ . Отсюда получаем  $x = vt$ , т. е. координата, отвечающая минимуму колебаний в волне, перемещается в пространстве со скоростью  $v$  и перемещается равномерно. Очевидно, что с такой же скоростью будет перемещаться и гребень волны, т. е. перенос энергии волной совершается со скоростью  $v$ . Такая скорость получила также название *фазовой*. Это название связано с тем, что мы говорим о перемещении некоторой точки в волне, все время находящейся в одной и той же фазе колебаний.

Когда волна достигает какого-то участка пространства среды, осцилляторы этой среды начинают приходить в колебательное движение. Для того чтобы привести эти осцилляторы в колебания, необходимо сообщить им определенную энергию. Ранее мы установили, что для одного осциллятора энергия колебаний



Рис. 4.12. Результат сложения волн



определяется выражением  $H = \frac{1}{2} A^2 \rho \omega^2$ . Именно такую энергию и переносит плоская волна при своем распространении. Этой энергии можно сопоставить определенный поток энергии в единицу времени через некоторую поверхность единичного размера, проведенную перпендикулярно волновому вектору в данной точке пространства.

Выше мы рассмотрели простейшие волны (синусо- или косинусоидальные), которые возникают в тех случаях, когда колебания осцилляторов, составляющих среду, являются гармоническими. Реальные волны могут иметь более сложную форму. Как показывается в математике, любые сложные по форме периодические процессы могут быть представлены в виде суммы отдельных синусо- и косинусоидальных гармонических процессов. Значит, любая сложная волна может быть представлена в виде суммы простых гармонических волн. Так, например, показанная на рис. 4.12 сложная периодическая волна может быть представлена суммой гармонических синусоидальных волн. Более того, даже непериодический процесс, некоторый горб, может быть также представлен в виде суммы, правда, в этом случае очень большого числа, отдельных синусоидальных волн. Таким образом, если наблюдаем распространение в среде некоторого сгустка колебательной энергии, т. е. некоторой области, где имеется повышенное значение амплитуд колебаний по отношению к другим областям пространства, то такой сгусток колебательной энергии (волновой пакет) также может быть представлен как результат наложения большого числа синусоидальных волн.

На рис. 4.13 показана концентрация колебательной энергии в определенном месте пространства. Такой волновой пакет или сгусток колебательной энергии будет перемещаться в пространстве, не изменяя своего вида, если только все составляющие гармонические волны имеют одну и ту же скорость распространения. Если это не так, то со временем вид волнового пакета меняется и он будет размываться.

На рис. 4.14 продемонстрирован процесс размывания волнового пакета (его изображения даны через равные промежутки времени) в предположении, что различные волны, составляющие

Рис. 4.13. Распространение волнового пакета

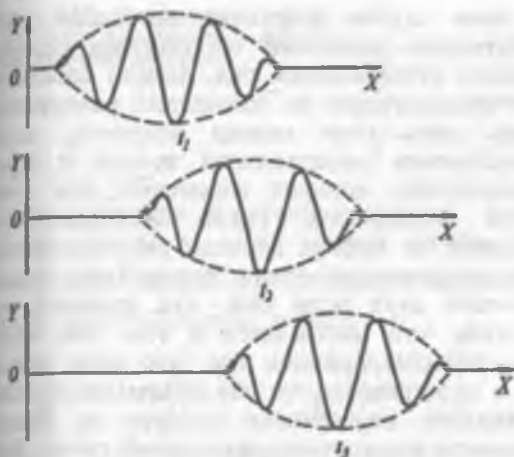
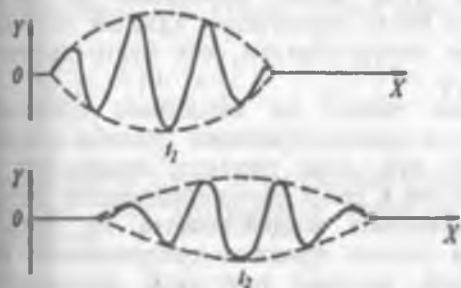


Рис. 4.14. Разыгрывание волнового пакета



первоначально этот волновой пакет, имеют отличающиеся друг от друга на постоянную величину  $\Delta v$  скорости распространения.

В дальнейшем мы увидим, что характерными для волн являются явления дифракции и интерференции, и так как волновые пакеты или квазичастицы представляют совокупность отдельных более простых волн, то для таких образований также должны быть свойственны явления дифракции и интерференции. Оказывается, что микрочастицы также обладают свойством давать дифракционные или интерференционные эффекты. Именно это и натолкнуло исследователей на мысль предложить для описания движения микрочастиц волновую, или квантовую, механику.

## 4.7. Сложение волн. Интерференция

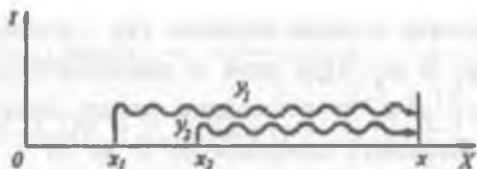
Пока мы рассматривали случаи, когда в среде распространялась лишь одна волна. Однако если имеется два или больше источников волн или возникают отражения, то в каком-то месте пространства все эти волны могут пересекаться. В результате их встречи наблюдается уже сложное колебательное движение. На

каждый осциллятор в этом случае действует несколько сил, приводящих их в колебательное движение (со стороны разных волн). Опираясь на принцип суперпозиции сил, можно ожидать, что суммарное смещение осцилляторов из положения равновесия при действии, например, двух волн можно получить, если каждый осциллятор представлять материальной точкой и просуммировать векторно колебания, которые возникают при действии каждой волны по отдельности. Такое предположение действительно подтверждается во многих случаях экспериментом и называется *принципом суперпозиции волн*. Эксперимент показывает, что при пересечении двух волн они, как правило, не воздействуют друг на друга, что выражается в том, что если после пересечения волны распространяются так, что одна волна и другая уже разделены в пространстве, то обе волны ведут себя так, как будто бы никакого пересечения вообще не было. Наглядным примером является пересечение двух лучей света. Мы знаем из школьного курса, что свет—это электромагнитные волны. Если скрестить два луча света, исходящие из двух фонарей, то легко видеть, что после пересечения каждый из этих лучей будет распространяться таким образом, как будто другого вообще не было.

Сказанное, однако, верно только до известного предела. Обычно, говоря о волнах и распространении колебательной энергии, подразумевают, что при этом никаких изменений в самой среде не происходит. Если в среде распространяются очень мощные волны, то они начинают воздействовать на среду, менять ее свойства. В этом случае принцип суперпозиции не выполняется. Группы явлений, которые при этом возникают, называют *нелинейными эффектами*.

Мы ограничимся пока только линейными эффектами, для которых принцип суперпозиции верен и суммирование волн сводится к суммированию колебаний, возникающих у осцилляторов среды при действии одновременно двух волн или больше. Это суммирование в общем случае должно быть векторным. Только в том случае, если поляризации двух волн совпадают, суммирование может производиться скалярно. Посмотрим детально, что будет происходить в месте встречи двух волн. Для этого выберем некоторую точку в пространстве и будем наблюдать результат наложения колебательных движений, которые переносятся в эту точку от некоторого источника волнами  $u_1$  и  $u_2$ . В качестве модели примем линейные волны, распространяющиеся вдоль оси  $X$  от двух источников  $x_1$  и  $x_2$  (рис. 4.15). Будем также считать, что обе эти волны поляризованы перпендикулярно оси  $X$  вдоль направления  $Y$ . Чтобы выяснить физику явления и максимально упростить вычисления, предположим, что частоты каждой из волн одинаковы и одинаковы амплитуды колебаний и скорости распространения волн, т. е.  $A_1 = A_2 = A$ ,  $\omega_1 = \omega_2 = \omega$ ,  $v_1 = v_2 = v$ . Будем также считать, что в начальный момент времени для

рис. 4.15. Результат сложения  
двух волн наблюдается в некото-  
рой текущей точке  $x$



точки, которая отвечает источнику первой волны, фаза колебаний равна  $\varphi_1$ . Для второй волны примем начальную фазу колебания равной  $\varphi_2$ . Тогда для первой волны получим следующую аналитическую запись:

$$y_1 = A \sin \left[ \omega \left( t - \frac{x - x_1}{v} \right) + \varphi_1 \right],$$

где  $x - x_1$  — расстояние от источника первой волны до точки  $x$ , в которой проводится наблюдение результата сложения двух волн. Аналогично для второй волны получим

$$y_2 = A \sin \left[ \omega \left( t - \frac{x - x_2}{v} \right) + \varphi_2 \right].$$

Пусть  $\varphi_1 \neq \varphi_2$ . Согласно принципу суперпозиции, результирующее движение в точке  $x$  (смещение осцилляторов и для первой и для второй волн совершается вдоль оси  $Y$ ) можно определить по формуле

$$\begin{aligned} y &= y_1 + y_2 = \\ &= A \sin \left[ \omega \left( t - \frac{x - x_1}{v} \right) + \varphi_1 \right] + A \sin \left[ \omega \left( t - \frac{x - x_2}{v} \right) + \varphi_2 \right] \end{aligned}$$

или

$$y = 2A \cos \left( \frac{x_1 - x_2}{2v} \omega + \frac{\varphi_1 - \varphi_2}{2} \right) \sin \left[ \omega \left( t - \frac{x}{v} + \frac{x_1 + x_2}{2v} \right) + \frac{\varphi_1 + \varphi_2}{2} \right],$$

где  $A' = 2A \cos \left( \frac{x_1 - x_2}{2v} \omega + \frac{\varphi_1 - \varphi_2}{2} \right) = 2A \cos \alpha$  явно от времени не зависит и должна быть отнесена к амплитуде суммарного колебания или к результату сложения двух волн. Второй член, включающий синус, явно зависит от времени и представляет бегущую волну. Аргумент под знаком  $\cos$  остается постоянным во времени при двух условиях: если относительное расстояние между источниками волн не меняется и не меняются значения их относительных фаз. Такие источники волн получили названия когерентных. При указанных условиях в результате будем наблюдать в точке  $x$  бегущую волну, которая, правда, уже исходит не из точки  $x_1$  или точки  $x_2$ , а из некоторой промежуточной

точки, которая задается как значениями  $x_1$  и  $x_2$ , так и фазами  $\varphi_1$  и  $\varphi_2$ . При этом в зависимости от разности фаз источников  $\varphi_1$  и  $\varphi_2$  и расстояния между этими источниками  $\cos \alpha$  может принимать значения от 0 до 1, т. е. будет происходить либо гашение волны волной, либо, наоборот, усиление одной волны другой волной. Такое усиление волны волной или гашение волны волной, которое происходит при сложении волн двух когерентных источников и поэтому сохраняется во времени, получило название *интерференции волн*. Найдем условия появления интерференции, предполагая для простоты, что  $\varphi_1 = \varphi_2$ . Для того чтобы происходило гашение одной волны другой волной, необходимо, очевидно, чтобы аргумент, стоящий под знаком  $\cos$  в выражении для амплитуды суммарной волны, был кратен нечетному числу  $\pi/2$ . Поэтому 
$$\frac{\omega(x_1 - x_2)}{2v} = \frac{\omega \Delta x}{2v} = (2n + 1) \frac{\pi}{2}.$$
 Здесь  $n$  — натуральные числа, принимающие значения  $n = 0, 1, 2, \dots$

Учитывая, что длина волны  $\lambda = 2\pi v/\omega$ , напомним  $\Delta x = (2n + 1)\lambda/2$ . Величина  $\Delta x$  получила название *разности хода*. Таким образом, гашение одной волны другой волной наблюдается, когда разность хода стабильна и кратна нечетному числу  $\lambda/2$ . Максимум интерференционной картины или усиление волны волной будет наблюдаться тогда, когда  $\cos \alpha$  обращается в единицу и, стало быть, аргумент  $\alpha$  равен кратному числу  $\pi$ . Следовательно,  $\omega \Delta x / (2v) = n\pi$ ,  $\Delta x = \lambda n$ , где  $n = 1, 2, \dots$ . Таким образом, интерференционная картина возникает не всегда, а только при определенных условиях.

Полученный результат можно распространить и на более сложный, нежели рассмотренный выше, достаточно искусственный случай. Для этого представим, что в некоторых точках пространства с радиус-векторами соответственно  $\mathbf{r}_1$  и  $\mathbf{r}_2$  находятся два когерентных источника волн. Выберем в том же пространстве произвольную точку, положение которой определим радиусом-вектором  $\mathbf{r}$ . В этой точке и будем наблюдать результат одновременного действия двух волн. Предположим, что обе волны, исходящие из точек 1 и 2, поляризованы перпендикулярно плоскости, проведенной через точки 1 и 2 и точку наблюдения  $\mathbf{r}$ , и имеют одинаковую частоту  $\omega$ .

Такой случай соответствует сложению двух волн, распространяющихся от точечных источников по поверхности жидкости, а также сложению двух, правда, при некоторых дополнительных условиях, сферических световых волн. В обоих рассматриваемых примерах результат действия двух волн можно найти, пользуясь правилами алгебраического (а не векторного) сложения.

Складывая возникающие в точке наблюдения колебания, возбуждаемые каждой из волн, получаем

$$y = y_1 + y_2 = A_1(|r - r_1|) \sin \omega \left( t - \frac{|r - r_1|}{v} \right) + \\ + A_2(|r - r_2|) \sin \omega \left( t - \frac{|r - r_2|}{v} \right),$$

где  $|r - r_1|$  и  $|r - r_2|$  — соответственно расстояния от источников волн до точки наблюдения;  $A_1(|r - r_1|)$  и  $A_2(|r - r_2|)$  — амплитуды волн в той же точке. Мы учли, что амплитуды должны быть уже не постоянными, а, как это бывает в круговых и сферических волнах, уменьшаться с расстоянием от точки наблюдения до источников.

Так как значения синуса любого угла колеблются между +1 и -1, то максимальным значением суммарной амплитуды может быть только величина  $y = A_1 + A_2$ , а минимальным  $y = A_1 - A_2$ . Первое условие выполняется тогда, когда аргументы синусов в выражении для двух волн отличаются друг от друга на угол, кратный  $2\pi$ .

Таким образом, условию максимума должно удовлетворять соотношение

$$\omega \left( t - \frac{|r - r_1|}{v} \right) + 2n\pi = \omega \left( t - \frac{|r - r_2|}{v} \right).$$

Отсюда

$$|r - r_1| - |r - r_2| = n \frac{2\pi v}{\omega} = n\lambda \quad (n = 1, 2, 3, \dots),$$

где  $|r - r_1| - |r - r_2|$  — разность хода волн от источников до точки наблюдения.

Условию минимума должно отвечать соотношение

$$\omega \left( t - \frac{|r - r_1|}{v} \right) + (2n + 1)\pi = \omega \left( t - \frac{|r - r_2|}{v} \right) \\ (n = 0, 1, 2, \dots).$$

Разность хода

$$|r - r_1| - |r - r_2| = (2n + 1) \frac{\pi v}{\omega} = (2n + 1) \frac{\lambda}{2}.$$

Следовательно, мы получаем те же правила для появления максимумов и минимумов интерференционной картины, что в первом простейшем примере.

Эффект взаимного гашения волн очень хорошо можно наблюдать при сложении волн, набегающих на берег моря под некоторым углом к нему, и отраженных волн. Этот факт используется на практике для построения волноломов, препятствующих разрушению берега волнами.



Рис. 4.16. Схематическое изображение опыта Юнга на поверхности воды

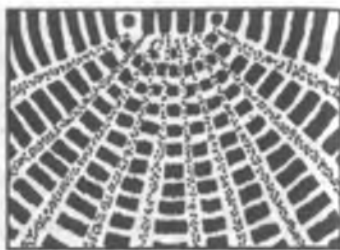


Рис. 4.17. Схематическое изображение расположения максимумов и минимумов интерференции волн на поверхности воды в опыте Юнга

Классическим опытом по наблюдению интерференции является опыт Юнга. Представим, что имеются два острия, расположенных на небольшом расстоянии друг относительно друга. Эти острия помещаются на одну и ту же горизонтальную планку (рис. 4.16) и одновременно приводятся в колебательное движение таким образом, чтобы они касались поверхности воды (точки  $B$  и  $C$ ). Мы получим два когерентных источника. Каждый из этих источников на поверхности воды будет порождать систему двумерных круговых волн. То, что эти источники действительно когерентны, понятно из конструкции прибора. Так как волны, исходящие из источников  $B$  и  $C$ , поляризованы одинаковым образом, а именно перпендикулярно поверхности воды, то в точке  $A$  будет происходить алгебраическое сложение двух волн. При этом, как было показано выше, если разность хода до точки  $A$  от точек  $B$  и  $C$ , или разность двух отрезков  $AB - AC$ , равна нечетному числу  $\lambda/2$ , то наблюдается гашение одной волны другой волной, т. е. участок гладкой поверхности воды. Если же  $AB - AC$  равно  $n\lambda$ , то в этом случае наблюдается стабильное увеличение амплитуды. Вся картина будет устойчива во времени, и ее легко наблюдать. На рис. 4.17 схематически показано расположение на поверхности воды максимумов и минимумов интерференционной картины.

Как известно, геометрическим местом точек, разность расстояния от которых до двух фиксированных точек (полюсов) постоянна, является гипербола. Поэтому как максимумы интерференционной картины, так и ее минимумы будут располагаться вдоль семейства гипербол.

Как уже указывалось, для когерентных источников эта картина будет стабильной. Если источники некогерентные, то картина получается размазанной и интерференция не наблюдается.

Здесь необходимо обратить внимание на один чрезвычайно важный момент. Мы видим, что в интерференционной картине на поверхности воды происходит чередование участков, где волна

усиливает волну. В тех местах, где имеется увеличение амплитуды, происходит накопление колебательной энергии. Таким образом, колебательная энергия оказывается локализованной в различных вполне определенных участках поверхности воды.

Такой результат «спасает» закон сохранения энергии. В самом деле, если одна волна гасит другую волну, то колебательное движение в этих местах исчезает и колебательная энергия убывает до нуля. Тогда возникает вопрос: а куда же она делась? Энергия на самом деле не исчезает, а перераспределяется в пространстве. При равномерном распределении колебательного состояния по всей поверхности воды перенос энергии пропорционален  $A^2$ . При наличии интерференции в некоторых местах энергия обращается в нуль, в некоторых местах она достигает  $4A^2$ . Суммарная энергия, если произвести интегрирование по всей поверхности воды, окажется постоянной величиной, как этого и требует закон сохранения энергии. Таким образом, в результате интерференции происходит только перераспределение энергии в пространстве, но не исчезновение ее.

## 4.8. Принцип Гюйгенса и принцип Ферма

Выше мы рассмотрели результат интерференции от двух когерентных источников. Если этих источников больше, то, очевидно, нужно произвести геометрическое или алгебраическое (последнее в том случае, если волны поляризованы одинаковым образом) сложение волн, исходящих от разных источников и достигающих данную точку. Оказывается, что результат такого сложения можно использовать при построении фронта волны в разные моменты времени. Такой способ построения волны привел голландского физика Х. Гюйгенса к формулировке принципа, который назван его именем. Чтобы сформулировать этот принцип, обратимся к следующему эксперименту.

Представим себе, что мы периодически касаемся одним острием поверхности воды. Вокруг этого острия во все стороны распространяются по поверхности воды круговые волны. Окружим такой источник волн круговой стенкой так, чтобы этот источник оказался в центре окружности. Проведем в кольцевой стенке на небольших равных расстояниях друг от друга серию щелей, которые можно открывать или закрывать по желанию исследователя. Если открыта только одна щель, то дальше по поверхности воды будет распространяться также круговая волна (рис. 4.18). Такая же картина должна, очевидно, наблюдаться для всех щелей, если последовательно открывать одну из них, а остальные считать закрытыми. Стало быть, можно сделать вывод о том, что каждая из этих щелей является источником круговых волн. Если теперь одновременно открыть, например, две щели, расположенные на противоположных концах диаметра, и наблюдать возникающие от каждой из этих щелей круговые волны, то нетрудно убедиться, что картина возникновения этих волн и их





Рис. 4.18. Подходящая к отверстию в кольцевом экране волна возбуждает за этим отверстием систему кольцевых волн

распространения будет абсолютно симметричная. Так как колебательная энергия к щелям доносится от первоначального источника и так как само окружающее кольцо с отверстиями отвечает форме круговых волн, возникающих от первоначального источника, результаты этих экспериментов можно трактовать следующим образом: каждая точка фронта волны является источником кольцевых волн, причем эти источники оказываются когерентными. Если одновременно открыть все щели и если их число будет достаточно велико, то мы увидим на некотором расстоянии от кольца с прорезями, что у нас снова стали распространяться кольцевые волны на поверхности воды так, как будто этого кольца вообще не было. Отвечающий этому и другим подобным экспериментам принцип Гюйгенса заключается в следующем. Пусть в некотором пространстве распространяется волна, которой можно приписать определенный перемещающийся во времени фронт. Гюйгенс предложил считать каждую точку фронта источником сферических волн. Такие расположенные на фронте волны источники волн все по построению и определению являются когерентными. В результате интерференции волн от всех этих источников образуется новый фронт волны, который может быть построен как касательная поверхность к совокупности сфер, описанных около расположенных на некотором расстоянии друг от друга точек первоначального фронта волны так, что радиусы этих сфер равны скоростям распространения волны в той части среды, которая прилегает к одной из точек, умноженным на один и тот же малый интервал времени. Такое построение показано на рис. 4.19.

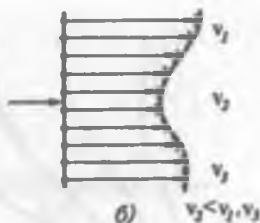
Очевидно, что если во всех областях среды скорость распространения волн будет одна и та же, то в результате такого построения мы получим фронт волны по форме такой же, как и исходный фронт. Если скорость распространения волн в разных участках среды будет различной, то форма фронта волны исказится по сравнению с первоначальным. Эти два случая показаны на рис. 4.20. Принцип Гюйгенса, который может быть обоснован с помощью точных вычислений, основанных на сложении волн, идущих от большого числа разных источников, расположенных определенным образом вдоль по фронту, помогает не только построить новый фронт волны через небольшой промежуток времени, но и, поступая так много раз, вообще



рис. 4.19. Построение фронта волны по принципу Гюйгенса



а)



б)

Рис. 4.20. Искажение фронта волны при распространении волны в неоднородной среде

описать, как будет вести себя фронт волны за достаточно длительный промежуток времени.

Можно также прийти к следующему важному заключению, что в однородных средах волны сохраняют свою форму, т. е. если они были плоскими, то и будут оставаться плоскими; если сферическими, то продолжат оставаться сферическими. В *неоднородных средах*, т. е. в таких средах, где в разных участках скорость распространения волны оказывается различной, фронт волны может исказиться. В частности, создавая определенную неоднородность пространства, можно превратить плоскую волну в сферическую, и наоборот. В сущности, в этом и заключается всем хорошо известное действие разного рода линз.

Для характеристики направления распространения энергии в волне мы раньше ввели волновой вектор. Если построить такие волновые векторы для близких участков фронта волны (небольших последовательно повторяющихся временных интервалов, рис. 4.21) и провести касательные ко всем этим волновым векторам, то мы получим путь распространения волны.

Вспомнив принцип Гюйгенса и сказанное выше о распространении волн в однородных и неоднородных средах, можно сразу заключить, что пути распространения волн представляют собой в общем случае неоднородных сред кривые линии. Лишь в однородных средах волна распространяется вдоль по пути, представляющему прямую линию.

Пути распространения колебательной энергии в принципе всегда можно найти, если разбить все пространство на малые интервалы, а время на малые отрезки и, задавая скорости распространения волны в каждом участке среды, построить эти фронты и затем соответствующие волновые векторы и касательные к ним, непосредственно опираясь на принцип Гюйгенса. Такое построение, однако, является довольно громоздким, и его можно заменить расчетами, опирающимися на *принцип Ферма*.

Этот принцип гласит: путь волны от точки А к точке В имеет форму кривой, которая отвечает наименьшему времени распространения в среде колебательного возмущения. Заметим, что речь идет не о кратчайшем расстоянии, а о наименьшем

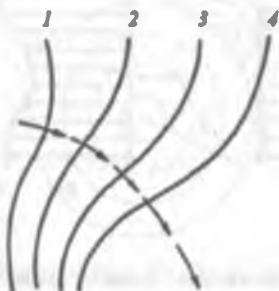


Рис. 4.21. Цифрами 1, 2, 3 и 4 обозначены фронты волны через равные промежутки времени. Стрелками показан один из путей волны

затрачиваемом для прохождения волны времени. Прямая получается только в том простейшем случае, когда среда, заполняющая пространство между точками  $A$  и  $B$ , однородна. Этот принцип является одним из примеров вариационных принципов, которые в физике весьма распространены.

Применим принцип Ферма сначала к эффекту отражения. Не касаясь вопроса, откуда берутся отраженные или преломленные волны, проведем следующий эксперимент. Создадим на поверхности воды некоторую волну, фронт которой будет представлять прямую линию. Такую волну легко получить, если взять достаточно длинную линейку и заставить ее одновременно всем своим краем периодически касаться поверхности воды. Если стенка ванны, в которой распространяется волна, находится на сравнительно небольшом расстоянии от линейки под некоторым углом к ней, то мы будем наблюдать систему плоских волн, доходящих до этой стенки и отраженных под некоторым углом (рис. 4.22). Этот эксперимент сразу указывает на то, что если взять две точки, отстоящие от отражающей поверхности на одинаковом расстоянии  $h$  (точки  $A$  и  $B$  на рис. 4.23), то волна, идущая от точки  $A$ , должна достигнуть поверхности в некоторой точке  $C$  (пока еще нами не определенной) и затем отразиться под каким-то углом и попасть в точку  $B$ .

Выясним, какому условию должна удовлетворять точка  $C$ , предполагая, что распространение волны от точки  $A$  к точке  $C$  и далее в точку  $B$  происходит в однородной среде. Пусть падающая волна касается отражающей поверхности в точке  $C$ , расположенной на расстоянии  $x$  от места пересечения с этой поверхностью перпендикуляра, проведенного из точки  $A$ . Общее время движения волны вдоль  $ACB$ , очевидно, равно  $t = AC/v + CB/v$ . Будем считать, что скорость  $v$  распространения волны в пространстве не меняется и среда однородна. Пользуясь правилами прямоугольных треугольников, запишем

$$t = \frac{\sqrt{h^2 + x^2}}{v} + \frac{\sqrt{h^2 + (AB - x)^2}}{v}.$$

Если варьировать положение точки  $C$  или, что то же самое, расстояние  $x$ , то время будет меняться: увеличиваться или уменьшаться. Минимум времени распространения будет тогда,

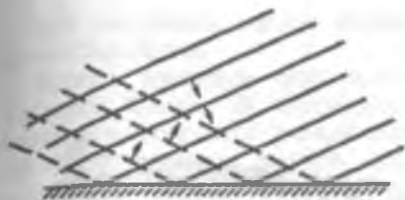


Рис. 4.22. Вид набегающих и отраженных от стенки волн на поверхности воды

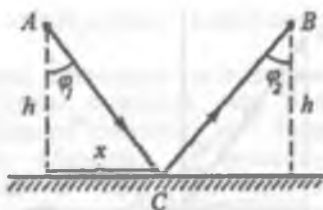


Рис. 4.23. К выводу закона отражения волны от плоскости

когда производная от времени по координате  $x$  окажется равной нулю. Это отвечает общему математическому утверждению, согласно которому любая функция от одного аргумента имеет максимум или минимум в точке, где производная от функции по своему аргументу обращается в нуль. Получим

$$\frac{dt}{dx} = \frac{2x}{2v\sqrt{h^2 + x^2}} - \frac{2(AB - x)}{2v\sqrt{h^2 + (AB - x)^2}} = 0.$$

Отношение  $x/\sqrt{h^2 + x^2} = \sin \varphi_1$  (угол  $\varphi_1$  указан на чертеже). Для угла  $\varphi_2$  имеем аналогично  $\sin \varphi_2 = (AB - x)/\sqrt{h^2 + (AB - x)^2}$ . Так как  $dt/dx = 0$  для минимума времени прохождения луча, то заключаем, что  $\sin \varphi_1 = \sin \varphi_2$  или  $\varphi_1 = \varphi_2$ . Таким образом, при отражении угол падения волны должен быть равен углу отражения.

Обратимся теперь к явлению преломления волны. При переходе волны из среды менее плотной в среду более плотную скорость распространения волны уменьшается. В результате угол преломления будет меньше угла падения: волновой вектор приближается к перпендикуляру, проведенному к поверхности раздела двух сред (рис. 4.24).

Будем считать, что волна исходит из точки  $A$ , преломляется в точке  $B$  и попадает к наблюдателю в точке  $C$ . Пусть расстояние от точки  $A$  до границы раздела сред будет  $h_1$ , а такое же расстояние для точки  $C$  —  $h_2$ . Расстояние между точками  $A'$  и  $C'$  вдоль границы раздела сред обозначим  $l$ . Расстояние от проекции точки  $A$  на границу раздела сред до точки преломления волны (точки  $B$ ) примем равным пока не определенной величине  $x$ . Тогда аналогичное расстояние до точки  $B$  от проекции точки  $C$  на границу раздела сред равно  $(l - x)$ . Время прохождения луча от точки  $A$  до  $B$  равно  $t_1 = AB/v_1$ . Время прохождения волны от  $B$  до  $C$  равно  $t_2 = BC/v_2$ , где  $v_1$  и

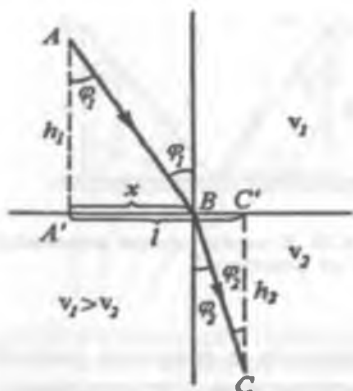


Рис. 4.24. К выводу закона преломления волн



Рис. 4.25. Явление атмосферной рефракции

$v_2$ —соответственно скорости прохождения волн в первой и второй средах.

Тогда общее время прохождения луча от точки  $A$  до точки  $C$

$$t = t_1 + t_2 = AB/v_1 + BC/v_2.$$

Пользуясь свойствами прямоугольных треугольников  $AA'B$  и  $CC'B$ , получаем

$$t = \frac{\sqrt{h_1^2 + x^2}}{v_1} + \frac{\sqrt{h_2^2 + (l-x)^2}}{v_2}.$$

Видим, что общее время распространения волны от источника к приемнику через границу раздела сред является функцией переменной  $x$ , конкретное значение которой нам и предстоит получить.

Найдем производную  $dt/dx$  и приравняем ее нулю:

$$\frac{dt}{dx} = \frac{2x}{2v_1\sqrt{h_1^2 + x^2}} - \frac{2(l-x)}{2v_2\sqrt{h_2^2 + (l-x)^2}} = 0.$$

В треугольниках  $AA'B$  и  $CC'B$  отношения  $x/\sqrt{h_1^2 + x^2}$  и  $(l-x)/\sqrt{h_2^2 + (l-x)^2}$  равны, очевидно,  $\sin \varphi_1$  и  $\sin \varphi_2$ , где  $\varphi_1$  и  $\varphi_2$ —соответственно углы падения и преломления. Тогда

$$\frac{\sin \varphi_1}{v_1} - \frac{\sin \varphi_2}{v_2} = 0 \quad \text{или} \quad n = \frac{v_1}{v_2} = \frac{\sin \varphi_1}{\sin \varphi_2}.$$

Отношение  $n = v_1/v_2$  называется коэффициентом преломления волны.

Мы получили, таким образом, соотношение, известное как закон Снеллиуса.

Можно привести еще целый ряд примеров использования принципа Ферма. В частности, нетрудно показать, что движущийся невысоко над поверхностью Земли световой луч будет в известной степени огибать ее, что приводит к атмосферной рефракции. Такое явление связано с тем, что плотность атмосферы с высотой уменьшается. Из-за атмосферной рефракции восход Солнца, например, будет виден чуть раньше, чем это должно было бы быть, если бы свет распространялся по прямой линии (рис. 4.25).

Разный прогрев атмосферы и связанная с этим разная плотность воздуха над разными участками земной поверхности приводят к необходимости вводить поправки при использовании лазерных геодезических инструментов и т. д.

## 4.9. Волны в замкнутой среде. Стоячие волны

Чрезвычайно важен случай волн в ограниченной среде. Простейшим примером таких сред является стержень конечной длины. Если приводить в колебание один из концов стержня, то колебательное движение начнет распространяться по нему и дойдет до его другого конца. В этом месте, поскольку энергия уже не может дальше распространяться в пустоту, поток энергии изменит свой знак по закону сохранения энергии. В результате по стержню начнет распространяться встречная волна, которая будет складываться с прямой волной. Если все время поддерживать колебания одного конца стержня, то возникнут две постоянные бегущие навстречу друг другу волны.

Рассмотрим, что при этом получится, на модели двух расположенных в точках  $x_1$  и  $x_2$  источников одномерных волн.

Пусть из точек  $x_1$  и  $x_2$  навстречу друг другу идут две одинаковые волны:

$$y_1 = A \sin \omega \left( t - \frac{x - x_1}{v} \right), \quad y_2 = A \sin \omega \left( t - \frac{x_2 - x}{v} \right).$$

В уравнении волны второй член в скобке всегда должен быть положительным, так как он отражает запаздывание по времени по отношению к начальному моменту, и поэтому в первом случае берем  $x - x_1 > 0$ , а во втором  $x_2 - x > 0$ , так как эта величина также больше нуля. Наблюдение сложения волн производится в точке  $x$ , лежащей между точками  $x_1$  и  $x_2$  (рис. 4.26).

При суммировании действия двух волн в точке  $x$  найдем

$$y = y_1 + y_2 = 2A \cos \left[ -\frac{\omega x}{v} + \frac{\omega}{2v} (x_1 + x_2) \right] \sin \omega \left( t - \frac{x_2 - x_1}{2v} \right).$$

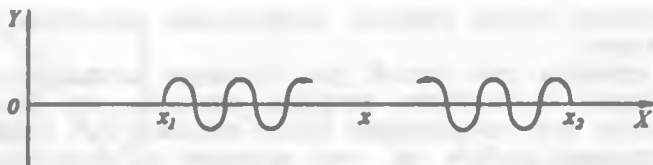


Рис. 4.26. Встречные волны, приводящие к появлению стоячей волны на отрезке от  $x_1$  до  $x_2$

В зависящем от времени множителе исчезает текущая координата  $x$ . Стало быть, все точки при колебаниях одновременно проходят положения равновесия, т. е. нет запаздывания по времени, характерного для бегущей волны. Нет, следовательно, и движения энергии в направлении волны. Поэтому волна называется *стоячей*.

Амплитуда волны равна

$$2A \cos \left[ -\frac{\omega x}{v} + \frac{\omega}{2v} (x_1 + x_2) \right],$$

т. е. является периодической функцией координаты  $x$ . График функции  $\cos \left( -\frac{\omega x}{v} + \frac{\omega}{2v} (x_1 + x_2) \right)$  отличается от графика функции  $\cos \left( -\frac{\omega x}{v} \right)$  лишь сдвигом по оси абсцисс. Поэтому можно рассмотреть лишь график функции  $\cos \frac{\omega x}{v} = \cos \left( -\frac{\omega x}{v} \right)$  (рис. 4.27).

Для двух точек на этом графике ( $x$  и  $x'$ ) имеем  $\cos \frac{\omega x}{v} = \cos \frac{\omega x'}{v}$ , откуда  $\frac{\omega x}{v} + 2\pi = \frac{\omega x'}{v}$ , т. е.  $x' - x = \lambda = 2\pi v/\omega$ , где  $\lambda$ —длина стоячей волны.

На концах отрезка в точках  $x_1$  и  $x_2$  для амплитуды имеем:

$$\begin{aligned} \cos \left[ -\frac{\omega x_1}{v} + \frac{\omega}{2v} (x_1 + x_2) \right] &= \cos \frac{\omega}{2v} (x_2 - x_1) = \\ &= \cos \left[ -\frac{\omega x_2}{v} + \frac{\omega}{2v} (x_1 + x_2) \right] = \cos \frac{\omega}{2v} (x_1 - x_2) = \cos \frac{\omega}{2v} (x_2 - x_1). \end{aligned}$$

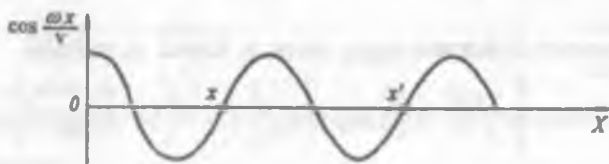
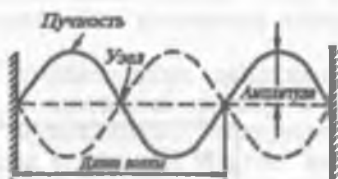


Рис. 4.27. График амплитуды стоячей волны

Рис. 4.28. Стоячая волна и ее характеристики



Стало быть, амплитуды стоячей волны на концах отрезка одинаковы. Положим их равными нулю. Тогда  $\cos \frac{\omega}{2v} (x_2 - x_1) = 0$  или

$$\frac{\omega}{2v} (x_2 - x_1) = (2k + 1) \frac{\pi}{2} \quad (k = 0, 1, 2, \dots),$$

$$x_2 - x_1 = (2k + 1) \frac{v\pi}{\omega} = (2k + 1) \frac{\lambda}{2}.$$

Следовательно, между длинами стоячих волн и величиной отрезка  $x_2 - x_1$  (размерами среды) имеется вполне определенное соотношение. Это важнейшее свойство стоячих волн.

Места, соответствующие максимальной амплитуде колебаний, называются *пучностями стоячей волны*. Точки, которые соответствуют минимальным амплитудам, называются *узлами стоячих волн* (рис. 4.28).

Как уже указывалось, важнейшим отличием стоячей волны от бегущей является то, что в установившейся стоячей волне энергия уже не передается от одной колеблющейся точки к другой. Энергия локализуется в областях около пучностей волн. Местоположение пучностей и узлов не меняется во все время существования стоячей волны.

В реальных ситуациях распространения волн в замкнутых средах встречные волны образуются в результате многократного отражения от границ среды (рис. 4.29). Стоячие волны образуются как результат сложения ряда таких встречных бегущих волн. Процесс установления волны (его называют процессом *релаксации*) совершается, как правило, за малый промежуток времени. Этот промежуток времени (*время релаксации*) определяется размерами среды и скоростью распространения в ней волн. Процесс релаксации является типичным *нестационарным процессом*. Такие процессы имеют особенное значение в статистических системах в связи с общей проблемой направленности развития событий.

Каждая стоячая волна обладает вполне определенной длиной волны или частотой. Так как в среде может существовать только дискретный набор стоячих волн, то отсюда легко получить, что среда заданных размеров может резонансно воспринимать колебательную энергию не при всех частотах внешнего воздействия, а лишь из заданного набора, т. е. дискретно. Здесь мы впервые



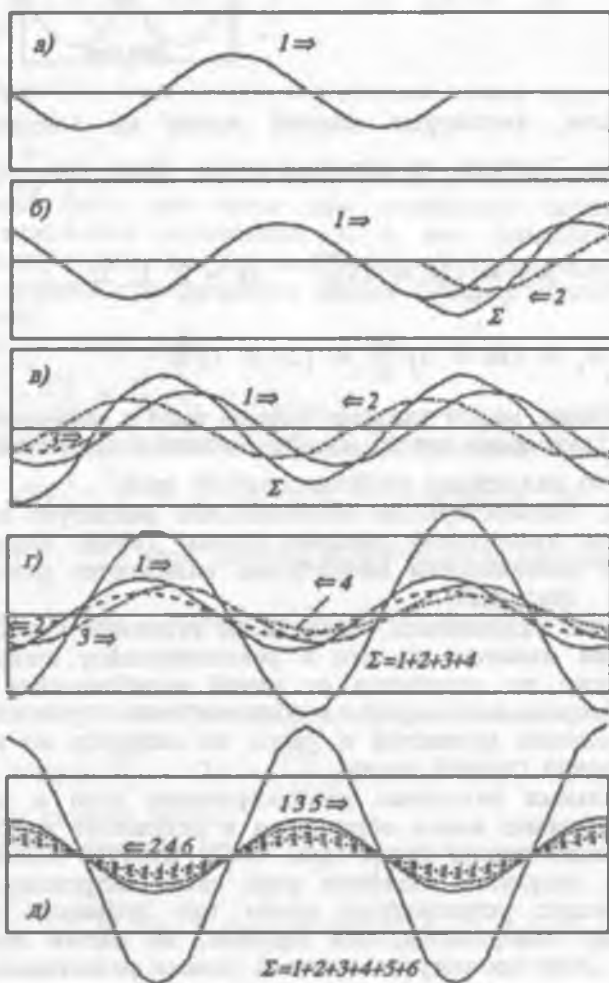


Рис. 4.29. Процесс установления стоячей волны в замкнутом пространстве с отражающими стенками с учетом частичной потери энергии в результате проникновения волн вправо в окружающую среду (расчет выполнен В. Дементьевым). Стрелками с цифрами на графиках показаны направления распространения бегущих волн при многократных отражениях. Знаком суммы с соответствующими цифрами обозначен результат сложения встречных бегущих волн. Последний рисунок отвечает почти установившемуся режиму стоячей волны

встречаемся со своеобразным эффектом квантования, что, как известно, является особенно характерным в поведении микро-частиц.

Стоячие волны как некоторые колебательные движения в замкнутой среде позволяют понять, в частности, работу музыкальных инструментов. Струнные музыкальные инструменты производят звук благодаря колебаниям струн. На струне могут возникать только вполне определенные стоячие волны, связанные с ее длиной, другими словами—вполне определенные частоты колебаний (длина волны связана с ее частотой). Следовательно, меняя длину струны, можно менять тон музыкального инструмента. Это и делает музыкант, прижимая пальцем струну к грифу инструмента. В музыкальных духовых инструментах происходят колебания продольного столба воздуха, заключенного в этом инструменте. Опять-таки, меняя длину этого столба, открывая те или иные клапаны или выдвигая шток, как это делается, например, в тромбоне, можно менять длину волны стоячих волн, образуемых в этом инструменте. Возбуждение стоячих волн имеет резонансный характер, поэтому если существует какая-то переменная внешняя сила, частота которой совпадает или близка к одной из резонансных частот, а длина волны, следовательно, совпадает с одной из возможных стоячих волн в данном музыкальном инструменте, то в результате можно раскачать столб воздуха или струну в музыкальном инструменте и достичь достаточной громкости звука.

К вопросам о волнах, об их свойствах мы еще раз вернемся в разделе «Оптика», так как свет является не чем иным, как электромагнитной волной. Там мы еще раз коснемся интерференции волн, чтобы оценить целый ряд экспериментов, связанных с оптическими явлениями, и дифракции, что позволит, в частности, понять, в каких случаях можно пользоваться представлениями геометрической оптики, а в каких уже надо учитывать волновую природу света.

#### Вопросы для самоконтроля

1. Приведите примеры колебательных движений в природе, растениях и животных и в технике. Какими общими признаками обладают такие движения?
2. Будет ли колебание с возвращающей силой, имеющей вид  $f = ax^2 + bx^3$ , гармоническим? Почему малые колебания можно считать гармоническими?
3. Будет ли математический маятник совершать колебания с одной и той же частотой в разных точках земной поверхности? Как должна измениться частота колебаний математического маятника в некоторой области земной поверхности, если под этой поверхностью находится какое-то массивное тело, например крупное месторождение металлических руд?
4. От каких основных факторов зависит энергия гармонического колебания?
5. Представим, что мы раскачиваем маятник с помощью периодических толчков. В каком соотношении интервал времени между двумя толчками должен находиться с собственным периодом колебаний маятника?
6. Приведите примеры резонансных и автоколебательных движений в растительных и животных организмах.
7. Как, пользуясь общими закономерностями возникновения резонансных явлений, объяснить, что частота шага, естественная для высоких людей, является меньше, чем частота шага для людей низкого роста? Как известно,

это создаст неудобства при движении солдатского строя. Как, пользуясь теми же закономерностями, объяснить, что частота биения сердца у мелких животных выше, чем у крупных?

8. Почему большие птицы машут крыльями реже в состоянии оптимального полета, чем маленькие?

9. Где легче может распространяться волна: в системах, состоящих из одинаковых осцилляторов, или в системах, где имеются области с резко различающимися частотами собственных колебаний осцилляторов?

10. Известно, что подводные лодки обнаруживают присутствие других подводных лодок в толще воды с помощью звуколокации, т. е. по отражению звуковой волны от лодки противника. Можно ли при таком обнаружении считать, что звуковой луч в толще воды всегда распространяется по прямой линии или есть какие-то мешающие факторы, приводящие к искривлению луча и затрудняющие звуколокацию? Что это за факторы?

11. Как изменится длина волны при переходе из среды менее плотной в более плотную?

12. Какое физическое явление, связанное с волновым движением, используется для уменьшения воздействия волн на берег с помощью волноломов? Почему вытянутые к морю сравнительно узкие молы могут играть роль таких волноломов?

13. Пользуясь принципом Гюйгенса, показать, как меняется фронт волны, если эта волна распространяется в среде, в которой скорость распространения волны равномерно меняется вдоль вертикальной оси или вдоль фронта волны.

14. Виолончелист, двигая в момент исполнения пьесы пальцы по грифу, прижимает струну, делая тем самым ее звучащую часть более длинной или более короткой. Как меняется высота звука, если струны укорачиваются?

15. Как различаются длины стоячих волн, которые могут образоваться в стержнях длиной  $l_1$  и  $l_2$ .

16. Какой принцип используется в музыкальных инструментах разного типа для того, чтобы изменить высоту издаваемых ими звуков?



## Г Л А В А 5

### ЭЛЕМЕНТЫ СТАТИСТИЧЕСКОЙ ФИЗИКИ

*«...когда исследователь достигает стадии, на которой он перестает видеть за деревьями лес, он слишком охотно склоняется к разрешению этой трудности путем перехода к изучению отдельных листьев»*

*Ланцет*

#### 5.1. Кинетическая теория газов, элементы статистической физики

До сих пор мы изучали поведение единичных объектов: материальных точек, тел или систем, состоящих из сравнительно небольшого числа частей. Лишь в случае распространения волн мы говорим об объектах, которые состоят из многих частиц, образуя непрерывную среду, но и в этом случае мы рассматривали специфическое движение, а именно согласованное движение большого числа частиц, возникающее при волновых процессах. Теперь перейдем к объектам, которые состоят из очень большого числа отдельных элементов, причем каждый отдельный элемент этого большого *статистического ансамбля* ведет себя в известной степени независимо от всех остальных. Примерами таких объектов являются прежде всего газы и жидкости. Мы знаем, что газы и жидкости состоят из отдельных частиц—атомов или молекул,—находящихся в постоянном движении. В таких объектах можно выделить два типа явлений. Явления первого типа, которые получили название *микроскопических*, связаны с поведением отдельных или малого числа частиц. Явления второго типа, которые принято называть *макроскопическими*, связаны с одновременным действием либо всех, либо большей части отдельных элементов сложной статистической системы так, что воздействие на измерительный прибор одной частицы уже вообще не ощущается. В соответствии с этим выделяют описания таких объектов двух типов: *микроскопическое* и *макроскопическое*. Так, например, если ставится цель проследить за движением одной конкретной частицы газа, то такое описание будет микроскопическим. Движение частицы

будет подчиняться законам механики, и средства описания ее движения являются обычными, уже ранее изученными для материальных точек или несколько более сложных тел. Но если надо описать действие массы газа, находящегося в некотором объеме, на стенку сосуда, выделяя на этой стенке квадрат размером, например,  $1 \text{ мм}^2$ , то, хотя в принципе и можно оценить это действие, разделяя прослеживая движения многих частиц, которые могут в единицу времени достигнуть выделенной площади на поверхности сосуда, ограничивающей область, где находится газ, однако, как сразу становится очевидным, такой подход абсолютно нерационален. Именно поэтому при переходе к таким множественным объектам или статистическим ансамблям в физике меняют и средства описания. Такое изменение средств описания является одним из примеров общего диалектического закона о переходе количества в качество.

Не следует думать, что понятия макроскопическое и микроскопическое относятся только к объектам очень малого размера, например к атомам или молекулам. В самом деле, когда изучается общее гравитационное поле, которое создается какой-нибудь удаленной галактикой или туманностью, то можно описать это поле, предполагая некоторое непрерывное распределение массы во всем объекте. Если рассматривать это явление на больших расстояниях от галактики, то действие каждой отдельной входящей в эту галактику звезды, хотя эта звезда может иметь громадные размеры, несущественно. Существенно только их суммарное действие. Такой подход будет макроскопическим. Если следить только за одной звездой из этого скопления, то можно говорить о микроскопическом подходе.

Некоторые основные черты поведения статистических ансамблей и способы их описания лучше всего рассмотреть на модели газа, как простейшей статистической системы. Газ, как мы уже говорили, является для этой цели вполне подходящим объектом. Это позволяет обнаружить некоторые качества, которые присущи именно системам, состоящим из многих частиц, и которые отсутствуют в системах, состоящих из небольшого числа частиц. Речь прежде всего идет о таком важнейшем моменте, как появление направленных процессов, существование которых и дает возможность ввести понятие времени.

По аналогии с введением простейшего тела—материальной точки—в механике, при изучении движения которой и удастся наиболее просто и наглядно сформулировать основные закономерности движения, при рассмотрении статистического объекта удобнее пользоваться моделью идеального газа.

*Идеальный газ*—это такой газ, между частицами которого нет никаких взаимодействий, кроме тех, которые происходят в момент упругого столкновения.

Модель идеального газа является в общем не такой уж идеализированной. Многие газы при не очень больших давлениях и не очень низких температурах ведут себя как идеальный газ.

Например, воздух при обычных условиях подчиняется законам идеального газа. Мы в дальнейшем выясним, когда можно пользоваться этой моделью и когда нельзя.

Для характеристики поведения множественных объектов, в частности такого объекта, как идеальный газ, необходимо изменить по сравнению с механикой средства описания. Посмотрим, как это можно сделать. Прежде всего укажем, что вопрос о том, где находится объект, имеет смысл как в случае, когда этот объект одиночный, так и в случае, когда под объектом понимается совокупность частиц. Конечно, бессмысленно указывать, где в каждый момент времени находится каждая из частиц, составляющих этот объект. Можно задать только вопрос: где находится весь ансамбль? Ответ на этот вопрос получается, если характеризовать местоположение объекта тем объемом пространства, в котором он находится. Например, по отношению к газу—тем объемом, который газ занимает. Для того чтобы ограничить задачу и сделать ее конечной, в дальнейшем будем подразумевать, что газ находится в некотором сосуде, т. е. его местоположение ограничено замкнутой поверхностью. Очевидно, что никаких специфических эффектов нельзя ожидать в газе, если такую поверхность как целое перемещать из одной области пространства в другую. Таким образом, вопрос о том, как движется (как целое) объем газа в пространстве, не имеет для нас смысла.

Хорошо установленным экспериментальным фактом является то, что все частицы газа находятся в постоянном хаотическом движении. Каждый газ, находящийся в замкнутом объеме, обладает некоторым запасом внутренней энергии, которая прежде всего проявляется в форме кинетической энергии всей совокупности частиц. Отсюда следует, что в качестве одной из характеристик состояния газа целесообразно ввести характеристику, которая определяет запас внутренней энергии системы. Такой характеристикой является температура (см. ниже).

Теперь уместно задать еще один вопрос: с помощью какого физического эксперимента можно обнаружить, есть ли газ внутри замкнутой поверхности? Простейшим будет эксперимент воздействия (давления) газа на стенку сосуда. Это воздействие (давление) легко обнаружить, если представить, что часть поверхности, ограничивающей объем газа (сосуда), представляет собой упругую резиновую мембрану. Если газа в сосуде нет, то мембрана будет плоской. При наличии газа (имеется в виду, что сам сосуд находится в вакууме) мембрана выгибается наружу. Мы получаем три характеристики, с помощью которых можно характеризовать явления, возникающие в газе. Это объем, температура и давление. Заметим, что характеристики имеют смысл только для статистического ансамбля и когда мы изучаем явления, возникающие при одновременном действии большого числа элементов этого ансамбля. Если в замкнутом сосуде имеется, например, десять молекул или атомов, то говорить о

температуре этих частиц или о давлении, которое они оказывают на стенку сосуда, бессмысленно.

Как видим, при переходе к статистическому ансамблю приходится менять способ его описания. Вместо радиуса-вектора тела вводится объем, который занимает статистический ансамбль. Вместо кинетической энергии частицы говорим об энергии всей совокупности частиц. Вместо силы будем пользоваться понятием давления газа на стенку сосуда.

Эти три величины можно рассматривать как степени свободы, которые характеризуют газ. Можно ввести систему координат, где состояние газа будет определяться некоторой точкой. Такая система в физике называется *фазовым пространством*. При изменении состояния газа точка движется по некоторой кривой, называемой *траекторией* точки в фазовом пространстве. В качестве четвертого измерения можно ввести время. В этом случае мы получим некоторый способ описания, математически сходный со способом описания материальной точки в трех- или четырехмерном пространстве.

Обычно состояние газа принято характеризовать на координатной плоскости объем (ось абсцисс)—давление (ось ординат). Это связано с тем, что при постоянной массе газа в объеме (например, отвечающем грамм-молекуле) из трех величин (объем, температура, давление) лишь две являются независимыми.

Здесь необходимо сделать важные пояснения, касающиеся равновесных и неравновесных состояний в газах и обратимых процессов.

*Стационарными* состояниями в газах называются такие, при которых в течение достаточно длительных промежутков времени не меняется форма ограничивающей объем с газом поверхности и помещенные в разных участках поверхности датчики температуры и давления дают неизменные показания.

Такие состояния возникают всегда, если газ находится в сосуде достаточно длительное время и если не меняется внешнее поле, действующее на все частицы газа.

Если внешнее поле отсутствует или его действием можно пренебречь (это можно сделать при экспериментах около поверхности Земли, когда сосуд относительно мал по размерам, хотя поле сил тяжести и действует на все частицы газа), то естественным состоянием является *равновесное*. Признаком такого состояния будет одинаковость показаний детекторов температуры и давления независимо от места их расположения на ограничивающей объем газа поверхности.

Если отмеченные выше условия не выполняются, то газ находится в *неравновесном* состоянии. Предоставленный сам себе газ всегда самопроизвольно переходит из неравновесного в равновесное состояние. Этот процесс также называется *релаксацией*, а потребное для установления равновесия время—*временем релаксации*.

При быстром неоднородном изменении внешних условий (резко двигается поршень в цилиндр с газом, газ быстро нагревается в одной части сосуда и др.) равновесное состояние газа нарушается. По прошествии времени газ снова перейдет в равновесное состояние, но уже при других условиях (при других объеме, давлении, температуре).

Перевести газ из одного состояния в другое можно либо с помощью ряда неравновесных процессов, либо путем бесконечно малых медленных шагов, при каждом из которых выполняются условия равновесия. Такие переходы называются *квазиравновесными* и их можно изображать на диаграмме состояния (диаграмме  $p, V$ ) непрерывными линиями. Переходы через серию неравновесных состояний на подобной диаграмме изобразить нельзя, так как при ее использовании подразумевается, что такие параметры, как давление и температура, одинаковы как для всего объема газа в целом, так и для любой его части.

В дальнейшем мы всегда, если нет особых оговорок, будем считать, что мы имеем дело либо с равновесным состоянием газа, либо с квазиравновесным процессом.

Неравновесные и квазиравновесные процессы обладают еще одним важным свойством: первые относятся к типу так называемых *необратимых*, а вторые — *обратимых*.

Рассмотрим эти процессы на примерах. Пусть имеется газ в цилиндре с невесомым поршнем. Пусть масса дополнительного груза на нем равняется  $M_0$ . Пусть в начальный момент давление газа в точности компенсировалось силой тяжести  $M_0g$ , т. е. поршень находился в состоянии равновесия. Снимем часть груза ( $M_1$ ). Равновесие нарушится и поршень будет двигаться до тех пор, пока в результате падения давления не установится новое равновесие. Если поршень поднимается на высоту  $H$ , то газ совершит работу  $(M_0 - M_1)gH$ .

Вернем поршень в прежнее положение. Ясно, что для этого надо, чтобы масса груза на нем стала никак не меньше  $M_0$ . На самом деле она должна быть даже больше  $M_0$ , так как при движении поршня вниз внешняя сила будет совершать над газом работу (не меньшую  $M_0gH$ ), в результате чего внутренняя энергия должна возрасти (подобно энергии сжимаемой пружины), что выразится в повышении температуры. Газ же при более высокой температуре будет оказывать и большее сопротивление движению поршня из-за давления. Чтобы газ точно вернулся к исходному состоянию, когда при нижней отметке равновесие наступает при нагрузке на поршень массой  $M_0$ , нужно для снижения его температуры часть тепла отвести во внешнюю среду.



Рассмотренный выше процесс называется *необратимым*. При таком процессе газ либо не вернется к исходному состоянию, либо что-то изменится во внешней среде.

Предположим теперь, что подъем поршня совершается не в один прием, а путем большого числа очень малых шагов. Предположим, что при каждом таком шаге снимается одинаковая малая масса  $m$ . Работа газа тогда будет равняться  $A_1 =$

$$= \sum_k g(M_0 - km)h_k, \text{ где } h_k \text{—расстояние, проходимое поршнем}$$

при каждом шаге. Если такие шаги будут очень малыми и каждый раз в газе будет устанавливаться равновесие, то мы получим типичный квазиравновесный процесс. Поскольку в каждой точке газ находится в положении равновесия, то особых явлений при малых изменениях этого положения в ту или иную сторону наблюдаться не будет.

Если подобным же образом с помощью малых шагов, последовательно нагружая поршень шаг за шагом массами  $m$ , вернуть поршень в исходное состояние, то сделанная над газом работа будет равна, очевидно,  $A_2 = \sum_k g[M_0 - (k - 1)m]h_k$ .

Разность модулей работ при подъеме поршня и при его опускании будет равна, как не трудно видеть,  $\Delta A = gmH$ .

Если  $m \rightarrow 0$ , то  $\Delta A \rightarrow 0$ . В результате такого процесса газ вернется в исходное состояние, и при этом можно обойтись без связи с внешней средой (если же связь есть, то ничего в ней не произойдет). Такой процесс называется *обратимым*.

Обратимые процессы лежат в основе так называемой *равновесной термодинамики*, а необратимые—в основе *неравновесной линейной и нелинейной термодинамик*. В этом и следующем параграфе мы будем иметь дело с равновесными состояниями и квазиравновесными (обратимыми) процессами.

Итак, при определении состояния газа используются три переменных:  $p$ —давление,  $V$ —объем,  $T$ —температура. *Шкала Цельсия*, в которой за нулевую точку отсчета принимается температура, отвечающая температуре таяния льда при нормальных условиях, а за сто градусов—температура кипения воды на поверхности Земли, по ряду причин используется в физике не вполне удобной. В физике поэтому пользуются температурой, измеряемой по *шкале Кельвина*. В этой шкале за начальную точку отсчета принимается температура, при которой прекращается поступательное движение молекул в газе и при которой давление газа обращается в нуль.

Эксперименты показывают, что если изобразить зависимость давления газа от температуры по шкале Цельсия, то для разных газов в широком диапазоне температур эти зависимости будут, во-первых, линейными, а во-вторых, продолжение соответствующих прямых в область отрицательных температур по шкале Цельсия будет давать одну и ту же точку пересечения (нуль

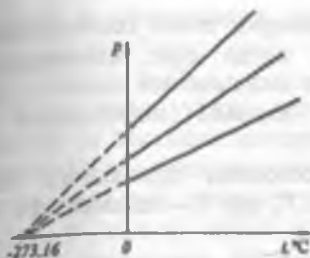


Рис. 5.1. Демонстрация определения температуры 0 К

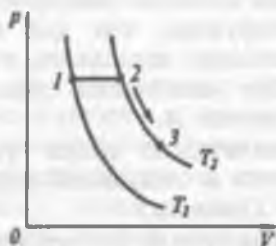


Рис. 5.2. К выводу обобщенного уравнения состояния газов

давления) с осью абсцисс (рис. 5.1). Такая точка приблизительно отвечает температуре  $-273,16^\circ\text{C}$ . На шкале Кельвина эта точка принимается за начало отсчета температуры, которая, таким образом, может иметь только положительные значения. Точка 0 К имеет, очевидно, гораздо более глубокий физический смысл, чем  $0^\circ$  по шкале Цельсия.

Как известно из школьного курса, существует ряд экспериментальных газовых законов.

Первый гласит, что  $pV = \text{const}$  при  $T = \text{const}$ . Это закон Бойля—Мариотта. Согласно второму  $p/T = \text{const}$  при  $V = \text{const}$ . Третий закон заключается в утверждении, что  $V/T = \text{const}$  при  $p = \text{const}$ . Это законы Гей-Люссака. Эти законы легко проверяются экспериментально. Их можно объединить, получив уравнение состояния идеального газа. Заметим, что для простоты мы всегда будем предполагать, что в сосуде находится моль газа. Это удобно, потому что моль газа содержит одно и то же число молекул вне зависимости от природы газа.

Выберем некоторое начальное состояние газа, которое характеризуется параметрами  $V_1, p_1, T_1$  (рис. 5.2). Совершим переход из точки 1 в точку 2 так, чтобы давление не менялось. Тогда согласно закону Гей-Люссака  $V_1/T_1 = V_2/T_2$ . Так как в точке

2 и в точке 1 давление одинаково, то  $\frac{V_1 p_1}{T_1} = \frac{V_2 p_2}{T_2}$ . Теперь из

точки 2 перейдем в некоторую точку 3 так, чтобы температура не менялась. Тогда  $\frac{p_2 V_2}{T_2} = \frac{p_3 V_3}{T_2}$ . Так как температура в точке

3 должна быть точно такой же, как в точке 2, то  $\frac{p_1 V_1}{T_1} =$

$\frac{p_2 V_2}{T_2} = \frac{p_3 V_3}{T_3}$ . Так как точки 1 и 3 выбирались произвольно,

то можно заключить, что для всех состояний газа выполняется соотношение  $\frac{pV}{T} = \text{const}$ .

Если провести эксперименты по проверке этого соотношения, то обнаружим, что для всех газов, когда они подчиняются приближению идеального, эта константа оказывается одной и той же. Она называется *универсальной газовой постоянной* и обозначается  $R$ .

Окончательно общее уравнение состояния идеального газа запишется в следующей форме:  $pV = RT$ , где  $T$ —температура по шкале Кельвина.

Мы получили это уравнение, опираясь на экспериментальные газовые законы. Выведем теперь это же уравнение, принимая во внимание микроскопическую структуру и пользуясь законами механики. Итак, будем считать, что в некотором объеме находится газ, содержащий частицы одного и того же сорта, причем примем эти частицы за материальные точки, т. е. будем интересоваться только их поступательным движением. Пусть масса каждой точки равна  $m$ .

Экспериментальным фактом является то, что частицы газа при любой температуре, если только эта температура такова, что газ остается газом, находятся в постоянном хаотическом движении.

Выделим на стенке сосуда участок единичной площадью. На этот участок все время налетают частицы газа, ударяются и отлетают обратно внутрь сосуда. Будем считать, что эти удары абсолютно упругие.

В эксперименте регистрируются удары не отдельных частиц, а некоторое совокупное действие ударов многих частиц за определенный достаточно большой по сравнению с временем удара частицы о стенку промежуток времени. Можно поэтому считать, что на стенку действует в течение всего времени некоторая усредненная сила, которая и называется, если ее отнести к единице площади, *давлением*.

Ясно, что чем больше частиц в единице объема газа, или чем больше плотность газа, то тем большее число частиц может в единицу времени долететь до стенки сосуда и удариться о нее. Воспользуемся экспериментальным фактом, а именно: давление во всех частях сосуда, в котором находится газ, или на все участки площади поверхности, внутри которой заключен газ, одинаково. Это может быть только в том случае, если внутри газа нет никакого направленного движения частиц, так что если выделить любой малый объем, то частицы, находящиеся в нем, в равном приблизительно количестве в любой промежуток времени будут двигаться как в правую сторону, так и в левую, как вверх, так и вниз, как на наблюдателя, так и от него. Значит, если рассматривать некоторый небольшой объем газа недалеко от стенки, где производится измерение давления, то только  $1/6$  частиц, заключенных в этом объеме (обозначим  $n_0$ —число частиц в объеме), может долететь до стенки и принять

участие в образовании постоянно действующей силы. Следовательно,  $p \sim \frac{1}{6} n_0$ .

Далее надо учесть, что до стенки сосуда в единицу времени долетят те частицы, которые не только движутся в сторону этой стенки, но и находятся не слишком далеко от нее. Будем считать также, что все частицы движутся от соударения со стенками сосуда и между собой до соударения с одной и той же по модулю скоростью. Назовем эту скорость *средней скоростью* и обозначим ее  $\bar{v}$ .

На самом деле частицы, соударяясь друг с другом и со стенками сосуда, приобретают различные скорости. Распределение числа частиц по скоростям передается некоторой функцией, названной *распределением Максвелла* (см. ниже). Распределение частиц по скоростям нужно принимать во внимание, например, при объяснении охлаждения объема жидкости при ее испарении, самого эффекта испарения жидкости даже при низких температурах и др. Однако при описании давления это обстоятельство можно игнорировать и считать, что все частицы имеют одну и ту же скорость.

На этом этапе мы встречаемся с некоторым новым моментом, которого не было при рассмотрении механики простых объектов, состоящих из небольшого числа тел. Раньше каждому телу или каждой точке приписывалась своя индивидуальная скорость. При описании ансамбля большого числа частиц вводится некоторая средняя скорость. Эту скорость, строго говоря, нельзя приписать ни одной реальной частице газа. Зато такое понятие оказывается чрезвычайно удобным в модельном представлении при описании действия большого числа частиц. Если ввести понятие средней скорости частицы, то надо считать, что за единицу времени к стенке долетят только те частицы, которые находятся на расстоянии от этой стенки, не большем величины, численно равной средней скорости. Таким образом,  $p \sim \frac{1}{6} n_0 \bar{v}$ .

Каждая долетевшая до стенки частица при абсолютно упругом ударе о стенку передаст ей некоторый импульс, который и будет восприниматься при большом числе ударов в виде постоянно действующей силы. Эта сила численно равна скорости изменения импульса. Летящая в сторону стенки частица имеет импульс  $m\bar{v}$ , отлетающая после удара частица движется в противоположном направлении, но сохраняет значение импульса. Это и есть условие *абсолютно упругого удара о неподвижную стенку*. Таким образом, полное изменение импульса равно  $2m\bar{v}$ . С учетом этого обстоятельства окончательно можно записать, что

$$p = \frac{1}{6} n_0 \bar{v} (2m\bar{v}) = \frac{1}{3} n_0 m \bar{v}^2.$$

Если считать, что внутри замкнутой поверхности находится моль вещества, то  $n_0 = N/V$ , где  $N$ —постоянная Авогадро;  $V$ —объем, заключенный внутри поверхности. Тогда

$$pV = \frac{1}{3} N m \bar{v}^2 \quad \text{или} \quad pV = \frac{2}{3} N \left( \frac{1}{2} m \bar{v}^2 \right),$$

где  $\frac{1}{2} m \bar{v}^2$ —средняя кинетическая энергия одной частицы. Если эту энергию умножить на число частиц, то получим полную кинетическую энергию всех частиц газа. Тогда последняя формула может быть переписана в форме

$$pV = \frac{2}{3} T_{\text{кин}}.$$

Если сравнить полученное соотношение с уравнением идеального газа ( $pV = RT$ ), то нетрудно заключить, что  $T_{\text{кин}} = \frac{3}{2} RT$ .

Итак, температура и полная кинетическая энергия отличаются друг от друга только постоянной величиной. Значит, можно утверждать, что температура есть мера кинетической энергии объема газа. Поскольку частицы идеального газа взаимодействуют между собой только при непосредственном ударе, а никаких других взаимодействий у них нет, то можно считать, что кинетическая энергия является полной энергией системы. Обозначим эту полную энергию буквой  $E$ . Тогда получим, что

$$E = \frac{3}{2} RT.$$

В результате проведенного рассмотрения становится понятным смысл температуры 0 К. Это такая температура, при которой прекращается всякое поступательное движение частиц газа.

Нельзя, однако, сказать, что в этом случае прекращается вообще всякое движение. Например, колебательное движение как внутри отдельных молекул, так и в твердом теле при 0 К сохраняется.

Разделив выражение  $\frac{3}{2} RT$  на число частиц, получим энергию, приходящуюся в среднем на одну частицу. Эта энергия равна  $\frac{3}{2} kT$ , где  $k = R/N$ —постоянная Больцмана. Таким образом, на одну частицу идеального газа в среднем приходится энергия  $\frac{3}{2} kT$ . Заметим, что коэффициент три не случаен. Он явился следствием предположения, что частицы идеального газа участвуют только в поступательном движении. Такое движение характеризуется тремя степенями свободы. Можно сделать вывод: в идеальном газе на каждую степень свободы приходится энергия, равная  $\frac{1}{2} kT$ .

Это правило можно обобщить и на случай газов, состоящих из молекул достаточно сложной формы. Эти молекулы могут не только двигаться поступательно, но и вращаться. Вращательные степени свободы тоже надо учитывать, так как часть энергии при столкновении может переходить из энергии поступательного движения в энергию вращения. Кроме того, реальные молекулы не являются абсолютно жесткими: их атомы могут совершать колебательные движения друг относительно друга. При столкновениях могут возбуждаться и эти колебательные движения (колебательные степени свободы).

Доказанное впервые Л. Больцманом и известное как *теорема о равномерном распределении энергии по степеням свободы* утверждение гласит, что вне зависимости от того, о какой степени свободы мы говорим, в любом случае, будь она поступательной или вращательной, в среднем на одну степень свободы приходится энергия, равная  $\frac{1}{2} kT$ . Частица с несколькими степенями свободы (пусть число этих степеней равно  $i$ ) обладает энергией  $\frac{i}{2} kT$ . Разумеется, все это относится к средней энергии. В каждый момент времени каждая частица может обладать и другой энергией.

Нетрудно получить соотношение между средней скоростью частицы и температурой газа. Выше было сказано, что для идеального газа средняя кинетическая энергия, приходящаяся на долю одной частицы, равна  $\frac{3}{2} kT$ . С другой стороны, средняя энергия равна  $\frac{1}{2} m \bar{v}^2$ , откуда  $\bar{v} = \sqrt{3kT/m}$ . Следовательно, при одной и той же температуре тяжелые частицы движутся с меньшей скоростью.

## 5.2. Распределения Максвелла и Больцмана

Использованные выше макрохарактеристики—объем, давление, температура—не полностью определяют состояние газа как статистической системы. Необходима дополнительная детализация—переход к описанию *статистических распределений*.

При рассмотрении давления и выводе формулы для давления на основании представления о газе как о совокупности движущихся хаотическим образом частиц мы ввели понятие средней скорости и предположили, что все частицы газа имеют одинаковую скорость. Это предположение верно для упрощенной модели, пригодной лишь для вывода уравнения состояния. Реально из-за соударения частиц друг с другом и со стенкой всегда появляются частицы с относительно малыми скоростями и частицы с повышенными скоростями.

Один из важнейших вопросов заключается в том, каким именно будет распределение частиц по скоростям, т. е. какое

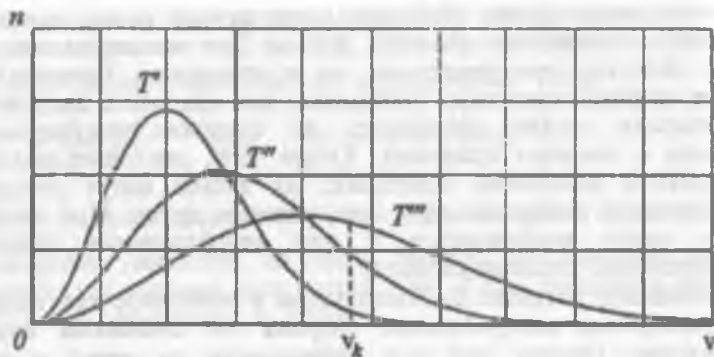


Рис. 5.3. Плотность состояний числа частиц газа в зависимости от скорости (распределение Максвелла) для разных значений температуры. Температура возрастает от  $T$  к  $T''$ . Отмечено значение критической скорости, при которой может быть преодолена энергия взаимного притяжения частиц

число частиц будет обладать скоростями, лежащими в определенном интервале значений модулей скоростей  $\Delta v$  или их компонент по осям координат.

Этот вопрос был изучен английским физиком Дж. Максвеллом и найденная им зависимость получила название *распределения Максвелла* (рис. 5.3). Среди частиц, движущихся с разными скоростями, условно можно выделить медленные частицы, движущиеся с наиболее вероятной скоростью, как можно показать, довольно близкой к введенной нами выше средней скорости, и, наконец, совокупность быстро движущихся частиц. Важной отличительной чертой кривой распределения Максвелла является то, что она не симметрична. Число медленно движущихся частиц оказывается меньшим, чем число быстро движущихся.

Вывод аналитического вида кривой распределения Максвелла довольно громоздок. Поэтому мы подробности опустим, а ограничимся лишь изложением некоторых основных физических соотношений и окончательным результатом.

Прежде всего уточним, что, собственно, мы хотим получить. Для этого вообразим, что имеется некоторый микроприбор-счетчик (конструкция его сейчас несущественна), способный регистрировать на частицы, скорости которых попадают в некоторый малый интервал от заданного значения  $v_1$  до  $v_1 + dv$ . Частицы, скорости которых не лежат внутри такого интервала (больше  $v_1 + dv$  или меньше  $v_1$ ) прибор не регистрирует. Пусть при каждом попадании частицы внутрь прибора раздается щелчок и на счетчике добавляется одна единица. Будем считать, что приемник прибора мал и его можно размещать в разных участках объема газа. Поместим прибор в некоторую область, занятую газом. Если открыть приемник прибора на бесконечно

малый промежуток времени, то может получиться и так, что ни одна частица в прибор залететь в данный момент времени не успеет. Однако, если такое время взять достаточно большим, то частицы с требуемыми скоростями рано или поздно в него залетят. Более того, при достаточно большом фиксированном интервале времени, в течение которого работает прибор, число зарегистрированных в нем частиц окажется в среднем постоянным вне зависимости от выбранного для опыта момента времени. Оно не будет зависеть и от того, в какую часть объема газа прибор помещен. Это следует из условия, что газ находится в равновесии. Значит, число частиц, обладающих скоростями в интервале от  $v_1$  до  $v_1 + dv$  и отмеченных прибором за единицу времени есть в среднем величина постоянная. Обозначим это число как  $n_0(v_1)$ .

Так как все микрообъемы газа эквивалентны, то легко найти общее число частиц в газе с выделенными значениями скоростей. Обозначим это число как  $N_v(v_1)$ .

Перенастроим теперь микроприбор так, чтобы он регистрировал частицы со скоростями в интервале от  $v_2$  до  $v_2 + dv$ . Это даст новое значение  $N_v(v_2)$ . Перебирая все возможные значения скоростей частиц от нуля до очень большой величины и производя измерения при большом числе заданных значений скорости  $v$ , расположенных равномерно вдоль оси  $v$  с шагом  $dv$ , получим ряд чисел:

$$N_v(v_1), N_v(v_2), \dots, N_v(v_k), \dots$$

Каждое такое число можно изобразить столбиком высотой  $n_v(v_k)$  и основанием  $dv_k$ , так что  $N_v(v_k) = n_v(v_k)dv_k$ . Получающаяся при этом картина называется *гистограммой*.

Располагая эти столбики вдоль оси  $v$ , уменьшая интервал  $dv_k$  и соединяя вершины столбиков непрерывной линией (см. рис. 5.4), получим некоторую функцию  $n(v)$ . Такая функция называется *функцией плотности состояний*.

Вспоминая геометрический смысл определенного интеграла, легко сообразить, что число частиц, скорости которых лежат в

широком интервале от  $v_1$  до  $v_2$ , будет равно  $\Delta N = \int_{v_1}^{v_2} n(v)dv$ .

Если интегрирование проводить от 0 до  $\infty$ , то должно быть, очевидно,

$$N = \int_0^{\infty} n(v)dv,$$

где  $N$ —полное число частиц в объеме.



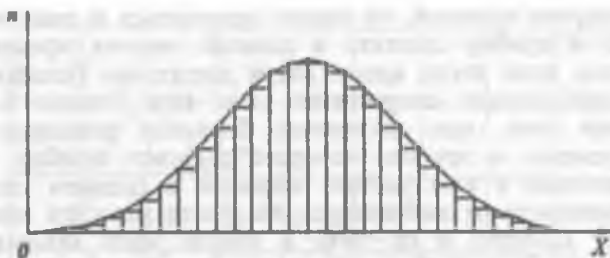


Рис. 5.4. Принцип построения гистограммы и функции плотности состояний для гауссова распределения значений величины  $X$ . Высота каждого столбика отвечает числу значений величины  $X$  в соответствующем интервале  $\Delta X$ , полученному в результате серии измерений

Таким образом, знание функции  $n(v)$  полностью разрешает задачу определения числа частиц газа, имеющих любые скорости в узком или широком интервале возможных их значений.

Именно эту функцию и предстоит нам определить.

Будем исходить из следующего. Как уже указывалось, давление газа на стенки сосуда объясняется столкновением частиц со стенками. Значит, газ оказывает на весь сосуд некоторое постоянное воздействие. Воздействие это, однако, не приводит к перемещению сосуда с газом в пространстве. Это значит, что газ не передает сосуду какого-то макроскопического направленного импульса. Это может быть, только если полный импульс всех находящихся в хаотическом движении частиц газа в любой момент времени равен нулю.

Если массы всех частиц одинаковы (ограничимся только этим случаем) и равны  $m$ , а мгновенные скорости каждой из частиц (нумеруем их индексом  $k$ ) равны  $v_k$ , то тогда для полного

импульса  $N$  частиц получим:  $p = \sum_k^N p_k = \sum_k^N m v_k = m \sum_k^N v_k =$   
 $= 0 = \sum_k^N v_k$ . Отсюда следует, что  $\sum_k^N v_{kx} = 0$ ,  $\sum_k^N v_{ky} = 0$  и

$\sum_k^N v_{kz} = 0$ , где  $v_{ki}$  есть компоненты скоростей частиц по осям координат. Если теперь разделить эти последние равенства на число частиц  $N$ , то будем иметь  $\frac{1}{N} \sum_k^N v_{kx} = 0$  (аналогично для  $Y$  и  $Z$ ).

Эти выражения по своей математической форме полностью совпадают с выражениями для среднего арифметического значения компонент некоторой скорости, в результате  $k$  измерений которой и получены частные значения  $v_1, v_2, \dots, v_N$ . Видно, что

среднее арифметическое этой величины при очень большом числе измерений ( $N \sim 10^{23}$ ) равно нулю. При этом предполагается, что сама скорость может принимать любые значения в достаточно широком интервале.

Реально следить за скоростью каждой молекулы нет не только никакой возможности, но и надобности, так как в подавляющем числе экспериментов (если только речь не идет об очень разреженных газах) всегда будет проявляться коллективное действие пусть не очень уж большой ( $10^2 - 10^3$  частиц), но все же совокупности молекул. Тогда все молекулы можно разделить на  $N_j$  групп по признаку близкого расположения их в данный момент времени в одном микрообъеме. Все молекулы одного такого микрообъема будут обладать некоторым суммарным импульсом  $p_j$ , который, конечно, может быть отличным от нуля. Вместе с тем ясно, что по-прежнему должно быть

$$\sum_j p_j = \sum_k p_k = 0.$$

Если теперь во всех микросовокупностях молекул всегда выбирать одинаковое (или почти одинаковое) число молекул, то и суммарные массы микрогрупп будут одинаковыми. Это позволяет записать, что  $\sum_j v_j = 0$  и  $\sum_j v_{jx} = 0$ ,  $\sum_j v_{jy} = 0$  и

$\sum_j v_{jz} = 0$ . Далее снова можно перейти к средним арифмети-

ческим:  $\frac{1}{N_j} \sum_j v_{jx} = 0$  и др.

Это, как и выше, позволяет ввести, например, для  $x$ -компоненты некоторую величину  $v_x$ , измерения которой с учетом случайного разброса дают значения  $v_{jx}$ . Среднее арифметическое значение этой величины при большой выборке также должно быть равно нулю. Это выполняется, как следует из теории случайных явлений, только в том случае, если возможные частные значения  $v_x$  распределены симметрично относительно нуля.

Теперь попробуем сообразить, каким еще свойством должно обладать распределение обсуждаемой величины  $v_x$ . Для этого зададим себе вопрос, что является более вероятным для суммарной скорости микрообъема (скорости его центра масс): иметь эту скорость близкой к нулю или существенно отличной от нуля? Нетрудно сообразить, что в силу хаотичности движения молекул в каждом микрообъеме существенно отличные от нуля суммарные

скорости будут преимущественно наблюдаться только при очень малом числе молекул (порядка 10). С ростом же числа молекул наиболее вероятным результатом сложения всех их скоростей в микрообъеме будет величина, близкая к нулю. Это означает, что при проведении измерений  $v_x$  или при последовательных многих измерениях величины  $v_x$  наибольшее количество значений будет лежать в области, близкой к нулю.

Введем теперь так называемую функцию распределения плотностей вероятностей значения любой величины  $x$ . Будем откладывать по горизонтальной оси все возможные значения этой величины, а по вертикали—вероятность того, что при проведении измерения будет обнаружено некоторое число  $x_k$ .

Тогда получится некоторая кривая  $f(x)$  или *кривая вероятности*.

Поясним это простым примером. Вспомним игральную кость. На ее гранях нанесены цифры от 1 до 6. Если мы будем бросать кость, например, 600 раз, то каждое из этих чисел выпадет примерно 100 раз (любой читатель может проверить это!). Значит, вероятность наблюдения каждого из значений от одного до шести одинакова. Гистограмма состояний будет представлять собой шесть одинаковых по высоте столбиков. Теперь возьмем кость, на гранях которой нанесены цифры 1, 2, 2, 3, 4, 5, т. е. цифра 2 повторяется дважды. Тогда при повторении эксперимента она появится уже не 100, а 200 раз. Значит, в области значения измеряемой величины, равного 2, на гистограмме будет существовать максимум. Если теперь заменить обычную игральную кость многогранником с очень большим числом граней и разместить на каждой из них по одной цифре в интервале от 1 до 6 с очень небольшим шагом (например, 1; 1,01; 1,02 и т. д.), то можно будет построить «плотную» гистограмму и перейти затем к функции плотности состояний. Она, очевидно, будет представлять прямую линию. Если же несколько раз на разных гранях нанести одинаковые цифры около значения равного 2, то функция плотности состояний в этом месте увеличится пропорционально числу таких повторений.

Пользуясь опытом этого примера и возвращаясь к анализу распределения скоростей молекул в газе, можно заключить, что функция вероятности  $f(v_x)$  должна представлять собой симметричную колоколообразную кривую с максимумом при  $v_x = 0$ . В теории случайных процессов доказывается, что разброс случайной величины  $x$  подчиняется так называемой *гауссовой кривой вероятности*, имеющей вид  $f(x) \sim \exp[-a(x - x_0)^2]$ , где  $a = \text{const}$  и  $x_0$ —наиболее вероятное значение случайной величины  $x$  (см. рис. 5.4). Именно это значение и будет найдено при большом числе измерений в результате вычисления среднего арифметического. Стало быть,  $x_0 = \frac{1}{N} \sum_k x_k$ .

Гауссова кривая отвечает так называемому нормальному закону распределения случайной величины и является универсальной для всех физических явлений, удовлетворяющих критерию случайности. Именно поэтому эта функция и играет в физике очень большую роль.

Смысл этой кривой можно пояснить, воспользовавшись наглядным примером стрельбы из закрепленного ружья по мишени. Если цель находится на достаточно большом расстоянии, то, даже если положение ружья неизменно, при большом числе выстрелов мы никогда не получим, чтобы пули все время попадали в одну и ту же точку. Всегда будет наблюдаться некоторый разброс попаданий, хотя наибольшее количество пуль ложится около центра мишени. По мере отхода от центра мишени к краям число попаданий  $y$ , в области все более и более удаляющейся от центра мишени, уменьшается и передается

функцией  $y = e^{-ax^2}$ , где  $x$  — отклонение от выбранного центра. Разброс происходит потому, что невозможно создать совершенно одинаковые условия для двух выстрелов. Чуть-чуть отличаются массы патронов, качество пороха, может в результате небольшого ветра измениться плотность воздуха на пути от ружья к мишени и т. д. Если все эти факторы действуют случайно, неконтролируемо, то наблюдается именно колоколообразное гауссово распределение вероятностей попаданий в зависимости от отклонения от центра мишени. Очевидно, случайными будут и значения компонент скорости  $v_{jx}$  для микрогрупп. Поэтому следует ожидать, что кривая распределения вероятностей величины  $v_x$  будет иметь вид  $f(v_x) \sim \exp(-av_x^2)$ . Напомним, что наиболее вероятным значением  $v_x$  является нуль.

Поскольку в газе нет никаких особенностей, отличающих направления  $X$ ,  $Y$  и  $Z$ , то аналогичные распределения получаются для  $f(v_y)$  и  $f(v_z)$ . Из независимости и равноправности событий, характеризующих величинами  $v_x$ ,  $v_y$ ,  $v_z$ , следует, что полная вероятность, отвечающая одновременному наблюдению трех величин  $v_x, v_y, v_z$ , должна быть связана с функцией

$$F(v) = f(v_x)f(v_y)f(v_z) = \exp[-a(v_x^2 + v_y^2 + v_z^2)] = \exp(-av^2).$$

Это утверждение основано на теореме математической теории вероятностей, согласно которой вероятности независимых событий перемножаются. Видно, что в показателе экспоненты в функции  $F(v)$  стоит уже просто квадрат модуля скорости группы частиц.

Теперь учтем еще следующий важный момент. Величина  $v^2 = v_x^2 + v_y^2 + v_z^2$  зависит лишь от длины вектора скорости (от его модуля) и не меняется при любых сочетаниях  $v_x$ ,  $v_y$  и  $v_z$ , составляющих квадрат скорости неизменным. В таких случаях говорят, что имеет место вырождение соответствующей величины

относительно, в данном случае, ее компонент. Если теперь по осям координат отложить компоненты  $v_x$ ,  $v_y$  и  $v_z$ , то одно и то же значение  $v$  будет получаться для всех точек сферы с радиусом  $v = \sqrt{v_x^2 + v_y^2 + v_z^2}$ .

Подобно тому, как повторение одинакового символа на гранях игральной кости приводит к увеличению вероятности выпадения этого символа при многократном бросании кости, так и отмеченное выше свойство должно иметь своим следствием соответствующее возрастание вероятности появления частиц со скоростями в области заданного значения  $v$  от  $v$  до  $v + \Delta v$ . Степень этого вырождения должна быть равна площади соответствующей сферы, т. е.  $S = 4\pi v^2$ .

Тогда для функции распределения вероятности наблюдения скорости со значением модуля, близким к заданному  $v$ , получим

$$F(v) \sim 4\pi v^2 \exp(-av^2).$$

Число частиц с такой скоростью будет определяться просто произведением общего числа частиц на эту вероятность и на величину интервала  $dv$  около значения  $v$ , т. е.

$$N_v = n(v) dv = NA \cdot 4\pi v^2 \exp(-av^2) dv.$$

Функция  $n(v) \sim v^2 \exp(-av^2)$  есть искомая функция распределения плотности вероятности числа частиц по скоростям. Постоянную  $a$  и коэффициент пропорциональности  $A$  можно найти из условий, что сумма частиц, обладающих всеми возможными скоростями, должна быть равна полному числу частиц и что определенная по правилам вычисления математического среднего некоторого распределения усредненная скорость должна быть равна

$\bar{v} = \sqrt{\frac{3kT}{m}}$ . В результате всех этих действий окончательно найдем

$$n(v) = N \left( \frac{m}{2\pi kT} \right)^{3/2} 4\pi v^2 \exp \left( -\frac{mv^2}{2kT} \right).$$

Это и есть распределение Максвелла. Графики функции  $n(v)$  при разных значениях температуры  $T$  показаны на рис. 5.3. Получается, что основная масса частиц должна иметь скорость в районе «горба». Видно, далее, что, во-первых, кривая распределения Максвелла является асимметричной. Во-вторых, что эта асимметрия резко возрастает с ростом температуры. Это означает, что очень быстро (нелинейно) растет число быстрых частиц, в частности, превосходящих некоторую критическую скорость  $v_{\text{крит}}$ .

Это позволяет понять некоторые наиболее важные закономерности испарения жидкостей при их нагревании и охлаждения при испарении без доступа тепла.

Хотя жидкость и нельзя представить как идеальный газ, так как в модели идеального газа исключаются любые взаимодействия между частицами, кроме тех, которые возникают при упругом столкновении, однако в одном имеется принципиальное сходство: и в том и в другом случаях частицы в объеме статистического ансамбля находятся в состоянии хаотического движения. Так как именно такой характер движения и приводит к специфическому характеру распределения числа частиц по интервалам модулей скоростей (распределению Максвелла), то асимметрия в числе быстрых и медленных частиц должна наблюдаться и в жидкости.

Процесс испарения заключается в том, что некоторые частицы, приобретая в результате многократных столкновений повышенный запас кинетической энергии, оказываются способными прорваться через границу раздела газ—жидкость и перейти в газовую фазу. Для такого перехода необходимо преодолеть силы взаимного притяжения частиц жидкости. Очевидно, что такими частицами могут являться лишь быстрые частицы. Если бы и частицы, обладающие наиболее вероятной скоростью, тоже могли преодолеть поверхностный барьер, то жидкость очень быстро испарилась бы. Испарение будет возрастать с ростом температуры, так как все большее и большее число частиц будет иметь скорость выше некоторой  $v_{\text{крит}}$ , достаточной для прорыва поверхностной пленки. Кривая Максвелла правильно передает тот экспериментальный факт, что интенсивность испарения растет гораздо быстрее, чем температура. Разумеется, для повышения температуры требуется постоянный подвод тепла. На первый взгляд может показаться, что член  $v^2$  (связанный с вырождением по скоростям) при обсуждении процессов испарения не должен приниматься во внимание, так как не все частицы со скоростями выше  $v$  заданной будут прорываться через поверхностную пленку, а только те, которые движутся перпендикулярно к ней. Однако это не так. Надо учесть, что пленка эта не является идеально ровной. Она имеет сложный микроскопический рельеф, и поэтому даже частицы, движущиеся под большими углами к макроскопической поверхности пленки, могут преодолеть ее действие.

Теперь представим себе, что к слою жидкости внешнее тепло не подводится. Несмотря на это, в жидкости все равно может иметься заметное число частиц, обладающих скоростью выше  $v_{\text{крит}}$ . Вспомним, что на каждую частицу в газе (и жидкости) в среднем приходится  $\frac{3}{2} kT$  энергии (если частицу представить в виде шарика).

Для быстрых частиц, очевидно, такая энергия будет больше. Поэтому вылетающие из жидкости быстрые частицы уносят в среднем больше энергии, чем унесли бы частицы, обладающие наиболее вероятной скоростью. Поэтому общая энергия жидкости

должна уменьшаться, а это уменьшение отражается в понижении ее температуры. Чем быстрее идет испарение, если мы, конечно, не подводим внешнюю теплоту к жидкости, тем, естественно, быстрее происходит охлаждение этой жидкости.

Это охлаждение будет происходить не до нулевой температуры. В конце концов даже быстрые частицы будут не в состоянии преодолеть силы взаимного сцепления частиц жидкости, которые приводят к образованию поверхностной пленки. Тогда испарение прекратится. Если температура все-таки еще достаточно велика, то жидкость будет продолжать испаряться, постепенно охлаждаясь.

Такой процесс понижения температуры при испарении лежит в основе терморегуляции всех живых существ. Например, для нормального функционирования организма человека температура его тела должна поддерживаться в определенных пределах.

Мы все хорошо знаем, что когда организм начинает перегреваться, человек потеет. На поверхности тела образуется пленка воды, которая начинает испаряться и за счет этого понижать свою температуру. Понижение температуры пленки приводит к понижению температуры всего тела. Аналогичный механизм терморегуляции имеют и растения, которые, как известно, также испаряют влагу.

В заключение остановимся еще на одном очень важном соотношении. Вспомним, что величины скоростей частиц и их энергии (в идеальном газе в отсутствии внешних полей имеется лишь кинетическая энергия) связаны между собой. Поэтому, пользуясь формулой для распределения частиц по скоростям, можно получить выражение для распределения частиц по энергиям. Для этого надо от переменной  $v$  перейти к переменной

$$\epsilon = \frac{1}{2}mv^2.$$

Выше было показано, что число частиц  $N(v)$ , обладающих скоростью в интервале от  $v$  до  $v + dv$ ,

$$N(v) \sim v^2 \exp(-mv^2/2kT) dv.$$

Откуда сразу получаем, учитывая, что  $v \sim \sqrt{\epsilon}$  и  $d\epsilon = mv dv \sim v dv$ ,

$$N(\epsilon) \sim \sqrt{\epsilon} \exp(-\epsilon/kT) d\epsilon.$$

Функция  $N(\epsilon)$  характеризует число частиц, имеющих кинетическую энергию в интервале от  $\epsilon$  до  $\epsilon + d\epsilon$ . Выше мы рассматривали газ, на частицы которого не действуют никакие внешние силы. Силы эти, однако, могут появиться, если объем с газом находится в поле сил тяжести (гравитационном) или, например, электрическом поле. В отсутствие поля частицы газа распределяются равномерно по всему объему и различие в состоянии частиц проявляется только в разных значениях их скоростей. Если же на частицы газа будет действовать внешнее поле, то это приведет уже к неравномерному распределению

числа частиц в пространстве, где газ находится. В самом деле, двигаясь в новых условиях внутри объема, частицы будут приобретать или терять потенциальную энергию. Возникнет конкуренция двух процессов. За счет хаотического теплового движения частицы газа будут по-прежнему стремиться распределиться равномерно по всему объему. Силы же, обусловленные действием на все частицы газа внешнего поля, будут «загонять» частицы в области с наименьшей потенциальной энергией. В результате возникнет новое усредненное состояние газа, которое следует уже характеризовать не только распределением частиц по скоростям, но и числом частиц в разных областях пространства. Функция плотности состояний в этом случае должна уже зависеть не только от компонент скоростей  $v_x$ ,  $v_y$  и  $v_z$ , но и от пространственных координат  $x$ ,  $y$ ,  $z$ . Число частиц (обозначим его  $N_{v_x}$ ), которые будут обладать скоростью с компонентами от  $v_x$  до  $v_x + \Delta v_x$  и др. и которые при этом будут находиться в области пространства (в кубике) с координатами вершин  $x$ ,  $y$ ,  $z$ ;  $x + \Delta x$ ,  $y$ ,  $z$  и др., определяется формулой:

$$N_{v_x} = n(v_x, v_y, v_z, x, y, z) \Delta v_x \Delta v_y \Delta v_z \Delta x \Delta y \Delta z.$$

Поскольку наложение внешнего поля не может изменить хаотического характера теплового движения и два отмеченных выше фактора (тепловое движение и поле) действуют независимо друг от друга, естественно считать, что  $n(v_x, \dots, x, \dots) = n_v(v_x, \dots) n_x(x, \dots)$ . При этом, как и ранее,

$$n_v(v_x, v_y, v_z) \sim \exp(-\alpha v_x^2) \exp(-\alpha v_y^2) \exp(-\alpha v_z^2) = \exp(-\epsilon/kT).$$

Поскольку в данном случае нас интересуют лишь состояния, при которых компоненты скоростей частиц находятся в районе заданных значений  $v_x$ ,  $v_y$  и  $v_z$ , а не состояния отвечающие одинаковому модулю вектора скорости, то дополнительный множитель  $v^2$  перед экспонентой учитывать не надо.

Обратимся теперь к функции  $n_x(x, y, z)$ , характеризующей неравномерное по пространству распределение числа частиц в малом объеме, вызванное действием поля. Как уже говорилось, если поля нет, то должно быть  $n_x = \text{const}$ . При наличии поля  $n_x \neq \text{const}$ . Действие поля приведет к тому, что частицы будут «вытесняться» из областей с большими значениями соответствующей потенциальной функции. «Тепловой фактор» будет, наоборот, «заталкивать» частицы в эти области. В результате установится равновесие. Пользуясь некоторой аналогией с силами, условие этого равновесия вдоль оси  $X$  можно записать в виде



$$kT \frac{\partial n_0}{\partial x} = -n_0 \frac{\partial U}{\partial x}.$$

Здесь выражение  $kT \frac{\partial n_0}{\partial x}$  отражает действие «теплового фактора» (или кинетического). Действие этого фактора, очевидно, пропорционально средней кинетической энергии частиц газа, т. е. величине  $kT$ , и производной  $\frac{\partial n_0}{\partial x}$ , которая характеризует перепад плотности частиц от одной малой области пространства к соседней в направлении оси  $X$ . Ясно, что чем большим будет этот перепад, тем с большей вероятностью частицы из одного микрообъема будут «перетекать» в соседний (см. также § 5.3). Поскольку газ в среднем находится в стационарном состоянии, то его температура всюду одинакова. В противном случае наблюдались бы тепловые потоки. «Силовой» (потенциальный), действующий в противоположном направлении (отсюда знак минус) фактор должен зависеть от скорости изменения потенциальной функции в выбранной области пространства также вдоль оси  $X$  (вспомним, что  $x$ -компонента силы, действующей на материальную точку, находящуюся в поле с потенциальной функцией  $U = U(x, y, z)$ , есть  $f_x = -\frac{\partial U}{\partial x}$ ) и  $n_0$ , так как поле действует сразу на все частицы в единице объема.

Переходя от выражения  $kT \frac{\partial n_0}{\partial x} = -n_0 \frac{\partial U}{\partial x}$  к дифференциалам  $kT dn_0 = -n_0 dU$  и интегрируя последнее с учетом  $kT = \text{const}$ , получим

$$\ln n_0 = -\frac{U}{kT}, \quad \text{или} \quad n_0 = n_0(U=0) \exp\left(-\frac{U}{kT}\right).$$

Здесь  $n_0(U=0)$ —число частиц в единице объема при минимальном значении потенциальной энергии и  $U$ —потенциальная энергия частиц в силовом поле. Такое распределение частиц называется *распределением Больцмана*. Его частным случаем является так называемая барометрическая формула, описывающая изменение давления воздуха в зависимости от высоты  $h$  над поверхностью Земли. Барометрическая формула имеет вид

$$p = p_0 \exp(-mgh/kT).$$

Учитывая теперь, что  $n_v \sim \exp(-\epsilon/kT)$  и  $n_x \sim \exp(-U/kT)$ , найдем окончательно

$$n(v, x) \sim \exp\left(-\frac{\epsilon}{kT} - \frac{U}{kT}\right) = \exp\left(-\frac{H}{kT}\right).$$

Здесь  $H = \epsilon + U$  — полная энергия молекул газа в данной области пространства.

Распределение с функцией  $n(v, x) \sim \exp(-H/kT)$  получило название *распределения Максвелла—Больцмана*. Оно полностью характеризует состояние газа как по их скоростям, так и по пространственному распределению частиц. Распределение с функцией  $n \sim \exp(-H/kT)$  играет очень важную роль в физике, так как сохраняется и при переходе к квантовым явлениям и дискретным значениям энергии.

В последнем случае, если возможны значения энергии  $H_i$  (в квантовой механике принято обозначение  $E_i$ ), число частиц с этой энергией равно

$$n_i = A \exp(-H_i/kT) = A \exp(-E_i/kT).$$

Здесь  $A$  — нормировочный множитель, также учитывающий степень вырождения. Экспонента называется *множителем Больцмана*.

Видно, что как распределение Максвелла, так и барометрическая формула тесно связаны с распределением Больцмана. Поскольку такое распределение устанавливается в системе самопроизвольно в результате хаотического движения частиц, то оно получило также название *нормального*.

### 5.3. Нестационарные процессы

Пока мы рассматривали состояния газа, которые во времени не меняются. Но наряду с такими простыми ситуациями встречаются и другие, более сложные. Так, например, может случиться, что в одной части объема концентрация частиц оказывается выше, чем в другой, имеется разная температура, наряду с хаотическим движением существует и направленное движение частиц газа, т. е. ветер.

Если такое неравномерное распределение параметров газа создано в начальный момент, а затем газ был предоставлен самому себе, то мы обнаружим, что по прошествии некоторого времени во всем объеме установится одно и то же состояние. Таким образом, это неравновесное состояние газа будет всегда по прошествии времени переходить в равновесное состояние, которое отвечает равномерному распределению свойств газа по всему объему, где он находится. Считаем, что внешнее поле отсутствует.

Здесь мы встречаемся с новым явлением, которое не наблюдали раньше, когда изучали механику небольшого числа объектов. Возникает некоторый направленный процесс, который всегда идет от неравновесного состояния к равновесному. Получается ли такая ситуация из-за того, что вступают в действие какие-то новые еще не известные нам (не описываемые в механике) законы взаимодействия частиц, или же здесь проявляется эффект коллективного движения большого числа объектов?

Представим, что с помощью ЭВМ на основе законов классической механики вычислены положения и скорости движения большого числа (до 1 000, например) частиц в сосуде.

Результат расчета положений хаотически движущихся и сталкивающихся частиц можно через равные промежутки времени выводить на дисплей и анализировать. Мы увидим, что исходная неравновесная картина, когда основная масса частиц собрана в левой части объема, а в правой части имеется область, где частиц вообще нет (вакуум), очень быстро исчезнет и в дальнейшем будет наблюдаться стабильная картина, когда в правой и левой частях объема находится приблизительно одинаковое количество частиц.

Вообще говоря, такая ситуация, когда в результате случайных столкновений и хаотического движения частиц снова в одной части сосуда образуется вакуум, не исключена полностью, однако вероятность этого для  $n$  частиц равна  $2^{-n}$ . Это означает, что если бы имелось всего сорок частиц и если бы ЭВМ выдавала по одной картине в секунду, то картинку с опустевшей правой частью сосуда получили бы приблизительно через сто тысяч лет. Ясно, что такую ничтожную вероятность вполне можно не принимать во внимание.

Итак, получаем следующий очень важный вывод. С одной стороны, мы не ввели в рассмотрение никаких новых физических законов и новых физических взаимодействий, которые бы отличали взаимодействие небольшого числа частиц, например двух-трех, от взаимодействия и столкновений большого числа частиц. С другой стороны, получается вполне определенный воспроизводящийся результат, заключающийся в том, что если первоначально все частицы собрались в одной части сосуда, а в другой части сосуда их нет (или их значительно меньшее число), то всегда процесс будет развиваться таким образом, что через некоторое время установится равновесие между различными частями сосуда.

Подобные эксперименты можно провести и в другом варианте, когда, например, частицам в одной части сосуда изначально приписываются повышенные начальные скорости по сравнению с частицами в другой части сосуда. Это означает, что газ в некоторой части сосуда обладает большей температурой. Снова через некоторое время мы обнаружим, что средняя кинетическая энергия частиц во всем объеме оказывается одинаковой, т. е. частицы с большей энергией передали излишек ее, а частицы с меньшим запасом кинетической энергии приобрели дополнительную энергию, заимствовав ее у более быстрых частиц. Опять-таки, хотя и нельзя полностью исключить возможность обратного процесса, вероятность его оказывается чрезвычайно малой. Результаты машинного эксперимента находятся в полном соответствии с результатами обычного физического эксперимента. Здесь, однако, надо сказать, что для правильной постановки машинного эксперимента начальные скорости частиц должны

быть выбраны случайными (см. гл. 10). Необходимости во введении каких-либо новых правил для характеристики взаимодействия или движения отдельных частиц такого множественного объекта по сравнению с описанием одиночных объектов нет. Но тем не менее такая сложная множественная система приобретает новое качество: способность создавать направленные процессы.

Это понятие направленного процесса относится уже не к каждой частице по отдельности, а ко всей их совокупности. Это снова проявление диалектического закона перехода количества в качество.

Так как такие направленные процессы характерны для множественных объектов, или статистических ансамблей, то целесообразно ввести для них свои вполне определенные законы, которые не сводятся к микроскопическим законам, определяющим состояние каждого отдельного объекта из этих ансамблей, но представляют нечто новое, свойственное именно коллективу.

Обратимся к детальному изучению некоторых неравновесных процессов, исходя уже из того, что эти неравновесные процессы развиваются во вполне определенных направлениях.

## 5.4. Явления переноса

Рассмотрим явления диффузии, теплопроводности и внутреннего трения. Все эти процессы относятся к категории неравновесных и возникают тогда, когда в двух достаточно близких объемах жидкостей или газов какие-то параметры системы оказываются неодинаковыми. К таким параметрам могут относиться концентрации частиц, средние энергии этих частиц или температуры. В разных соприкасающихся областях могут быть и различные направленные скорости движения частиц. Переход к стационарному, равновесному состоянию будет происходить за счет переноса определенного свойства. Например, в явлении диффузии переносятся частицы из той области, где их больше, в ту область, где их меньше, вплоть до выравнивания. При эффекте теплопроводности переносится кинетическая энергия частиц, так что постепенно их кинетическая энергия в данном объеме и кинетическая энергия в соседнем объеме также выравниваются и др.

Во всех последующих рассуждениях для упрощения будем считать, что имеем один и тот же газ, а не смесь газов. Полагаем, кроме того, что в каждом отдельном случае различным является только один какой-нибудь параметр.

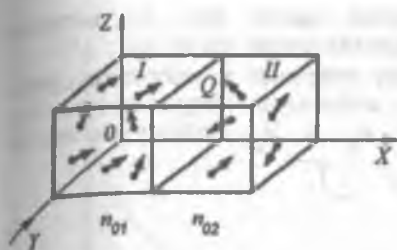
Введем понятие свободного пробега частицы в газе. Если имеется одна частица в замкнутом объеме, то она будет летать от стенки к стенке, и путь, который она пройдет от одного соударения до другого соударения, равен расстоянию между стенками. Если же имеется много частиц, то между ними тоже будут происходить соударения. Частица может много раз столкнуться с другими частицами, прежде чем она дойдет до стенки,

а может в принципе случиться и так, что она до стенки вообще не дойдет. Если частица движется со средней скоростью  $\bar{v}$  и при этом ни с чем не сталкивается, то за единицу времени она пройдет расстояние, численно равное средней скорости. На самом деле частица сталкивается с другими частицами и пройдет меньшее расстояние. Это расстояние тем меньше, чем больше вероятность того, что частица встретит на своем пути другую частицу или несколько частиц за заданный интервал времени. Опять-таки нельзя проследить за тем, как поведет себя каждая отдельная частица. Но можно ввести некоторую вероятностную величину и говорить, например, о среднем числе столкновений в единицу времени, среднем расстоянии  $\bar{s}$ , которое проходит частица от одного удара до другого удара. Ясно, что чем больше концентрация  $n_0$  (чем больше плотность газа), тем больше вероятность столкновения одной частицы с другой. Если число соударений в единицу времени обозначить  $\bar{n}$ , то можно написать, что  $\bar{n} \sim n_0$ . Чем больше средняя скорость частиц, тем с большим числом партнеров они могут столкнуться в единицу времени. Это означает, что  $\bar{n} \sim n_0 \bar{v}$ . До сих пор мы рассматривали молекулы как частицы, размеры которых бесконечно малы. Если это так, то частицы не могут столкнуться или их вероятность столкновения равна бесконечно малой величине. В действительности надо принять во внимание размеры частиц. Вообще говоря, молекулы любого вещества могут иметь весьма сложную геометрическую форму. Особенно это относится к молекулам, содержащим большое число атомов. Однако в достаточно грубом приближении, которым мы в наших дальнейших рассуждениях и ограничимся, в качестве приемлемой модели можно принять модель молекулы в виде сферы. Диаметр сферы принимается равным расстоянию между максимально удаленными друг от друга атомами молекулы. Несложные наглядные рассуждения показывают, что число соударений молекул-шариков между собой тем больше, чем больше площадь поперечного сечения сферы, окружающей молекулу. Поэтому радиус такой сферы называется *эффективным радиусом соударения*. В результате получим, что число соударений  $\bar{n}$  в единицу времени пропорционально  $r^2$ , где  $r$ —эффективный радиус соударения. *Эффективный радиус*—это величина, присущая каждой молекуле и зависящая от ее конкретных свойств.

Мы рассмотрели все факторы, влияющие на  $\bar{n}$ , и поэтому окончательно можно записать  $\bar{n} = n_0 r^2 \bar{v}$ .

Среднюю величину свободного пробега  $\bar{s}$  молекулы получим, разделив путь, проходимый ею за единицу времени без столкновений (т. е.  $\bar{v}$ ), на число столкновений за то же время. Таким образом, длина свободного пробега равна  $\bar{s} = \bar{v} / \bar{n} = 1 / n_0 r^2$ . Видно, что в замкнутом объеме длина среднего пробега частиц

Рис. 5.5. К выводу формул для коэффициентов явлений переноса



не зависит от температуры и определяется только плотностью газа и эффективными размерами частиц, составляющих этот газ.

Перейдем теперь к исследованию явления диффузии, которое заключается в переносе частиц газа из одного объема в другой, если числа частиц в них оказываются в начальный момент времени неодинаковыми. Выделим в объеме, занятом газом, два близко расположенных единичных объема I и II (рис. 5.5). Будем считать, что число частиц в объеме I равно  $n_{01}$ .

Соответствующее число частиц в объеме II обозначим  $n_{02}$ . Считаем, что  $n_{01}$  больше, чем  $n_{02}$ . Проведем некоторую

граничную плоскость Q, разделяющую эти объемы. Будем, как уже указывалось выше, считать, что средняя кинетическая энергия частиц в левом и правом объемах одна и та же, т. е.  $T_1 = T_2 = T$ . Одинаков также и тип частиц, находящихся в

левом и правом объемах, т. е.  $m_1 = m_2 = m$ . В результате

хаотического теплового движения частицы левого объема будут проходить через границу раздела в правый объем, и наоборот. Обозначим  $\Delta n_{12}$ —число частиц, переходящих в единицу времени

из объема I в объем II, а  $\Delta n_{21}$ —число частиц, переходящих за

то же время из объема II в объем I. Очевидно, что число

частиц, которое переносится слева направо или справа налево,

при хаотическом движении пропорционально средней скорости

хаотического движения частиц и среднему числу частиц в

единице объема, т. е.  $\Delta n_{12} \sim \bar{v} n_{01}$  и  $\Delta n_{21} \sim \bar{v} n_{02}$  (напомним, что

так как  $T_1 = T_2$  и  $m_1 = m_2$ , то  $\bar{v}_1 = \bar{v}_2 = \bar{v}$ ).

Так как  $n_{01} \neq n_{02}$  и, более того,  $n_{01} > n_{02}$ , число частиц,

переходящих слева направо, больше, чем число частиц, перехо-

дящих справа налево, на величину  $\Delta n \sim \bar{v}(n_{01} - n_{02})$ . Таким

образом, имеется направленное движение частиц, причем можно

считать, что это движение совершается вдоль оси X.

Концентрация частиц будет, очевидно, меняться при движе-

нии вдоль оси X слева направо не мгновенно, не скачком, а

непрерывно, т. е. представлять некоторую функцию. Поведение

этой функции вдоль оси X можно охарактеризовать, введя

«скорость» изменения числа частиц вдоль оси  $X$ , которую математически мы запишем как производную  $dn(x)/dx$ .

В математике такие производные отвечают компоненте градиента некоторой функции вдоль заданного направления. Для двух достаточно близких точек  $x_1$  и  $x_2$ , пользуясь такой производной, можно записать

$$f(x_1) - f(x_2) = \frac{df}{dx} \Delta x,$$

где  $\Delta x = x_1 - x_2$ . Следовательно,

$$\Delta n \sim \bar{v} \frac{dn}{dx} \Delta x,$$

где  $\Delta x$  принимаем равной удвоенной длине свободного пробега молекул. В самом деле, именно с расстояния длины свободного пробега от каждой стороны плоскости раздела находящиеся в объеме частицы ( $1/6$  их, см. § 5.1) могут с максимальной вероятностью без соударений пересечь плоскость раздела двух объемов газа. Так как мы вводим величину  $\Delta x$  не в точности равную  $\bar{s}$ , то безразлично, является ли эта величина одинаковой для молекул в первом и во втором объемах газа или различается для них. Поэтому можно считать, что величина  $\bar{s}$  взята усредненной. В результате получим, что

$$\Delta n \sim \frac{1}{6} \bar{v} \cdot 2\bar{s} \frac{dn}{dx}, \quad \text{или} \quad \Delta n \sim \frac{1}{3} \bar{v} \bar{s} \frac{dn}{dx} \sim \frac{1}{3} \sqrt{\frac{T}{m}} \frac{1}{n_0 r^2} \frac{dn}{dx}.$$

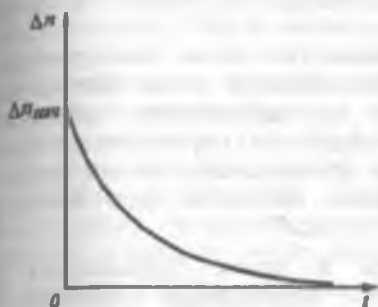
При этом  $n_0$  также следует рассматривать как некоторую среднюю плотность частиц в единице объема.

Конечно, такие средние величины, в свою очередь, можно ввести тогда, когда мы рассматриваем ситуацию, при которой число частиц в левом объеме в начальный момент незначительно отличается от числа частиц в правом объеме, что и соответствует обычной экспериментальной практике.

На основании проведенных выкладок можно заключить, что диффузия происходит тем быстрее, чем выше температура газа. Тяжелые частицы диффундируют медленнее, чем легкие. Скорость диффузии будет больше, если частицы обладают меньшими размерами. Наконец, диффузия тем меньше, чем больше плотность среды.

При прочих равных условиях диффузия частиц вдоль выбранного направления  $X$  выше в тех случаях, когда градиент числа частиц в направлении  $X$  выше. Значит, если в начальный момент этот градиент был достаточно велик, то диффузия будет происходить интенсивнее. Это приведет к постепенному выравниванию числа частиц в единице объема в правой и левой частях от границы раздела и к уменьшению, в свою очередь, градиента  $dn/dx$ , снижению скорости диффузии. График, иллю-

Рис. 5.6. Зависимость числа частиц, переносимых при диффузии через границу направленным образом, от времени



стрирующий зависимость скорости диффузии от времени, показан на рис. 5.6, где по горизонтальной оси отложено время, а по вертикальной оси—число частиц  $\Delta n$ , которое переносится в единицу времени из левой части объема газа в правую часть.

Коэффициент  $D = \frac{1}{3} \bar{v} \bar{s} = \sqrt{\frac{T}{m}} \frac{1}{n_0 r^2}$  называется коэффициентом диффузии.

Мы рассмотрели случай, когда справа и слева от границы раздела находится один и тот же газ. Не составляет особенного труда рассмотреть более сложную ситуацию, когда справа и слева от границы раздела находятся частицы газа разного сорта, имеющие, например, разную массу и поперечные размеры. Основные выводы при этом останутся теми же. Мы обнаружим, что если первоначально, например, в левой части находился газ с более легкими молекулами, а справа газ с более тяжелыми молекулами, то левые частицы быстрее проникают в правый объем, чем правые частицы в левый объем. Также более медленным окажется проникновение крупных частиц в среду мелких частиц.

Поскольку, как мы уже указывали, кинетические представления применимы в значительной степени и для жидкости, а жидкость—это весьма плотная среда по сравнению с газом, можно сразу же понять, почему диффузия в жидкости происходит значительно медленнее, чем в газе, если даже отношение температуры к массе, т. е. величина, пропорциональная квадрату средней скорости, и в газе и в жидкости одинакова.

Пока мы рассматривали случай, когда граница раздела имеет не физический, а чисто математический, условный смысл. Через такую границу частицы проникают слева направо и справа налево, не встречая никакого сопротивления. Представим теперь такой эксперимент, когда правая и левая части разделены реальной границей, причем такой, что через эту границу вправо частицы проходят легче, а влево сложнее. Если бы вообще частицы могли проходить только в одну сторону, то постепенно все они собрались бы, например, в правом объеме. В реальных



экспериментах всегда имеется определенная вероятность и обратного прохождения частиц.

Когда, например, левые частицы в таком эксперименте начинают проникать в правую часть объема и число частиц там начинает возрастать, то за счет их хаотического движения возрастает давление этих частиц справа на перегораживающую стенку. Это повышает вероятность прохождения частиц справа налево и создаст новое равновесное состояние, при котором  $n_{10} \neq n_{20}$ .

Свойством пропускать частицы в одном направлении и задерживать их движение в другом направлении обладают, в частности, стенки растительных клеток (биологические мембраны). Мембраны могут как бы отсасывать вещество из окружающей среды и собирать его в определенной области пространства. В результате в этой области пространства не только накапливается определенное вещество, но и возникает повышенное давление, которое получило название *осмотического*, а само явление—явление осмоса. Это явление оказывается чрезвычайно важным именно в растительном мире. Оно объясняет не только способность растительных организмов накапливать нужные вещества, но и возникновение вполне определенной устойчивости растений, например, в вертикальном положении и, более того, способность растений двигаться при изменении, например, условий освещенности. Хорошо известным фактом является то, что головка подсолнечника поворачивается вслед за движением Солнца таким образом, чтобы все время получать максимальное количество солнечной энергии. Подобная способность присуща целому ряду других растений. Такое движение растений вызывается возникновением в определенных клетках синхронно изменяющегося с освещенностью осмотического внутреннего давления.

Рассмотрим теперь следующее явление переноса—*теплопроводность*. Это явление возникает тогда, когда в двух соприкасающихся объемах жидкости или газа частицы находятся при различных температурах. Будем, как и ранее, рассматривать два близко расположенных, практически соприкасающихся, небольших объема газа. Для простоты также считаем, что газ справа и слева состоит из одинаковых частиц, т. е.  $m_1 = m_2 = m$ . Концентрации частиц в правой и левой частях объема также полагаем одинаковыми, т. е.  $n_{01} = n_{02} = n_0$ . Температуры же или средние скорости частиц в двух объемах различны:  $\bar{v}_1 \neq \bar{v}_2$  ( $\bar{v}_1 > \bar{v}_2$ ). За счет хаотического движения частицы из одного объема будут переходить в другой, и наоборот. Более быстрые частицы, т. е. частицы, обладающие большей кинетической энергией, сталкиваясь с более медленными в правом объеме, станут отдавать часть своей кинетической энергии. При этом температура в правой части объема возрастет или, что то же самое, возрастает средняя кинетическая энергия

движения частиц в правом объеме. Температура в левой части объема падает. Через некоторое время эти температуры сравняются и тогда, хотя частицы по-прежнему будут перелетать из одного объема в другой, макроскопическая передача тепловой энергии из одного объема в другой прекратится.

При изучении явления теплопроводности, как и ранее при изучении диффузии, исходим из того, что в единицу времени в результате хаотического движения слева направо и справа налево через границу переходит некоторое число частиц. Каждая частица в среднем имеет энергию, пропорциональную  $kT$ . Если обозначить  $\Delta T_{\text{кин}}^{(12)}$  — среднюю энергию, которая переносится частицами, движущимися из левого объема в правый, то можно написать:

$$\Delta T_{\text{кин}}^{(12)} \sim \frac{1}{6} \bar{v}_1 n_0 k T_1 \sim \frac{1}{6} \sqrt{\frac{T_1}{m}} n_0 k T_1 = \frac{1}{6} n_0 k m^{-1/2} T_1^{3/2}.$$

Появление множителя  $n_0 \bar{v}_1$  понятно, поскольку это число частиц, которые в среднем за единицу времени переходят из левого объема в правый объем через границу раздела. Суммарная энергия, которая переносится из правого объема в левый при соответствующем встречном движении частиц, очевидно, равна

$$\Delta T_{\text{кин}}^{(21)} \sim \frac{1}{6} n_0 k m^{-1/2} T_2^{3/2}.$$

Так как  $T_1 > T_2$  и  $\Delta T_{\text{кин}}^{(12)} > \Delta T_{\text{кин}}^{(21)}$ , то перенос энергии слева направо происходит в больших количествах, чем перенос энергии справа налево. В результате получаем

$$\Delta T_{\text{кин}} = \Delta T_{\text{кин}}^{(12)} - \Delta T_{\text{кин}}^{(21)} \sim \frac{1}{6} n_0 m^{-1/2} (T_1^{3/2} - T_2^{3/2}).$$

Будем считать, что изменение температуры (градиент температуры) наблюдается в направлении оси  $X$ . Полагая, что скачкообразных перепадов температуры нет и она достаточно непрерывно спадает при движении от левой части объема через границу раздела вдоль оси  $X$  и, следовательно, представляет некоторую функцию от  $x$ , запишем

$$\begin{aligned} T_1^{3/2} - T_2^{3/2} &= f_1(x) - f_2(x) = \frac{df}{dx} \Delta x = \frac{d(T^{3/2})}{dT} \frac{dT}{dx} \Delta x = \\ &= \frac{3}{2} T^{1/2} \frac{dT}{dx} \Delta x. \end{aligned}$$

Как и ранее, примем величину  $\Delta x$  по порядку равной  $\bar{L}_s$ , т. е. удвоенной средней длине свободного пробега. Тогда

$$\frac{\Delta T}{\text{кин}} \sim \frac{1}{2} n_0 m^{-1/2} T^{1/2} \frac{1}{n_0 r^2} \frac{dT}{dx} = \frac{1}{2} \sqrt{\frac{T}{m}} \frac{1}{r^2} \frac{dT}{dx},$$

где  $\eta \sim \frac{1}{r^2} \sqrt{\frac{T}{m}}$  — коэффициент теплопроводности.

Таким образом, перенос энергии из одной части объема газа в другую происходит тем быстрее, чем больше градиент температуры вдоль направления оси  $X$ . Этот перенос также больше у более нагретых газов. Тяжелые и крупные частицы медленнее переносят теплоту.

Проведенное рассмотрение позволяет понять один довольно интересный факт. Известно, что в космических аппаратах применяются для дыхания человека два рода газовых смесей. Одна газовая смесь состоит в основном из кислорода и азота и представляет смесь, близкую к тому составу воздуха, которым мы дышим. Вторая смесь представляет смесь кислорода и гелия. Гелий, так же как и азот, не участвует непосредственно в процессе дыхания, но позволяет поддерживать нужное давление газа в кабине. Так как он значительно легче воздуха, то можно взять большее количество дыхательной смеси в космический корабль. Если температура окружающего воздуха, как это обычно и бывает в комфортной для человека обстановке, меньше температуры человеческого тела, то будет наблюдаться постоянный отток тепла человеческого тела в атмосферу. Процесс этот подчиняется тем закономерностям, которые мы выше рассмотрели. Следовательно, атмосфера, состоящая из гелия и кислорода, поскольку гелий значительно легче азота, обладает повышенной теплопроводностью, повышенной способностью передавать тепло от слоя, более нагретого и непосредственно примыкающего к телу человека, к другим слоям. Человек в этой атмосфере мерзнет быстрее. Это надо компенсировать специальной одеждой.

Последний процесс переноса, который мы рассмотрим, относится к внутреннему трению или к переносу импульса. Представим, что частицы газа в левом объеме 1 наряду с хаотическим движением участвуют и в некотором направленном вертикально вверх движении, которое характеризуется скоростью  $V_1$ . Частицы в правом объеме 2 также участвуют в направленном движении вверх, но с иной скоростью  $V_2$ . При этом полагаем, что  $V_1 > V_2$ . Будем считать, что все условия, кроме того, в объемах 1 и 2 газа одинаковы, т. е.  $m_1 = m_2 = m$ ,  $T_1 = T_2 = T$ . Такая ситуация возникает в тех случаях, когда в толще газа с помощью какого-то источника создается поток — ветер. Тогда скорости частиц газа приобретают направленную составляющую и соответствующий компонент импульса. За счет того, что частицы при хаотическом движении переходят из левого объема в правый и сталкиваются с частицами в правом объеме, они передают этот компонент последним. В результате частицы в

правом объеме также приобретут большую скорость в вертикальном направлении. Если неравенство направленных скоростей создано только в начальный момент и не поддерживается длительное время внешним источником, то через некоторый промежуток времени эти скорости в левом и правом объемах уравниваются.

Таким образом, вертикальный компонент импульса будет переноситься из области, где он был больше, в область, где он был меньше. Снова возникает направленный процесс.

Общий импульс, который передается от левых частиц к правым, очевидно, пропорционален числу частиц, переходящих из левого объема в правый, и среднему импульсу направленного движения, т. е.  $\Delta p_{12} \sim \frac{1}{6} n_0 \bar{v} m V_1$ . Соответственно из правого объема в левый переносится в среднем за единицу времени импульс  $\Delta p_{21} \sim \frac{1}{6} n_0 \bar{v} m V_2$ . Так как  $\Delta p_{12} \neq \Delta p_{21}$ , то в результате возникает некоторый «поток» импульса слева направо, который в единицу времени может быть охарактеризован величиной

$$\Delta p = \Delta p_{12} - \Delta p_{21} \sim \frac{1}{6} n_0 m \sqrt{\frac{T}{m}} (V_1 - V_2) = \frac{1}{6} n_0 \sqrt{Tm} (V_1 - V_2).$$

Воспользовавшись, как и ранее, представлением о градиенте скорости и приняв  $\Delta x = 2\bar{s}$ , можно окончательно записать

$$\Delta p \sim \frac{1}{3} n_0 \sqrt{Tm} \frac{1}{n_0 r^3} \frac{dV}{dx} = \frac{1}{3} \frac{1}{r^2} \sqrt{Tm} \frac{dV}{dx},$$

где  $\xi \sim \frac{1}{r^2} \sqrt{Tm}$  — коэффициент внутреннего трения. Он пропорционален  $T^{1/2}$ ,  $m^{1/2}$ ,  $r^{-2}$ .

Таким образом, способность одного слоя газа увлечь за собой другой слой газа, если первый слой приобретает некоторые направленные движения, тем больше, чем больше температура газа, масса частиц газа и чем меньше геометрические размеры частиц.

Изученные нами примеры показывают, как, базируясь на микроскопических представлениях о том, что газ состоит из хаотически движущихся частиц, переходя к некоторым средним величинам, можно предсказать и изучить макроскопические эффекты переноса частиц (или переноса массы, что, очевидно, одно и то же, так как каждая частица обладает массой), переноса энергии, переноса направленного импульса, т. е. получить макроскопические результаты, которые могут быть непосредственно проверены макроскопическими экспериментами, где движение каждой отдельной частицы уже не играет роли.

В качестве последнего примера, на первый взгляд далекого от обсужденных выше, но, как увидим, имеющего с ними важные общие черты, рассмотрим некоторые особенности течения несжи-

маемой жидкости в трубе переменного сечения. Трением потока жидкости о стенки сосуда и внутренним трением слоев жидкости пренебрежем. Последнее можно сделать, если считать, что все микрообъемы жидкости в области любого сечения трубы движутся с одной и той же поступательной скоростью. При относительно малых скоростях можно, кроме того, считать течение ламинарным. Это означает, что траектории движения вдоль по трубе любых микрообъемов жидкости не пересекаются между собой и не образуют петель.

Упростим задачу еще и условием, что на частицу жидкости не действует никакое постоянное внешнее поле. Иногда это условие называют приближением невесомой жидкости. С довольно высокой степенью точности такую модель жидкости можно принять в том случае, когда труба расположена горизонтально по отношению к поверхности Земли.

Ясно, что если труба везде имеет одинаковое сечение, то и скорость потока жидкости будет постоянной. Однако, если сечение трубы на разных ее участках различно, то скорость, очевидно, также будет меняться. Если жидкость плотностью  $\rho$  несжимаема, то общее количество переносимой через сечение  $S$  в единицу времени жидкости везде одинаково. Это количество, или величина потока,  $I = \rho Sv$ , где  $v$  есть скорость течения жидкости. Должно быть  $I = \text{const}$ , или  $Sv = \text{const}$ . Это соотношение называют правилом неразрывности струи.

Выберем теперь некоторое сечение трубы  $S_1$ . Пусть на слой («срез») жидкости около сечения  $S_1$  действует слева давление  $P_1$ . В результате этот слой будет двигаться вправо, выталкивая последующие слои и заставляя их также приходить в движение.

Теперь на некотором расстоянии от первого сечения выберем сечение  $S_2$ , такое, что  $S_1 \neq S_2$ . Если течение жидкости совершается слева направо, то на сечение  $S_2$  будет действовать встречное давление  $P_2$  со стороны последующих слоев жидкости. Чтобы движение вообще стало возможным, надо, очевидно, чтобы величины давлений  $P_1$  и  $P_2$  были различными. Поскольку  $S_1 \neq S_2$ , то и  $v_1 \neq v_2$  и появится прирост кинетической энергии, связанный с различным характером движения слоев жидкости в районе сечений  $S_1$  и  $S_2$ . Этот прирост будет равен

$$\Delta T_{\text{кин}} = (m_1 v_1^2)/2 - (m_2 v_2^2)/2,$$

где  $m_1$  и  $m_2$  — массы слоев жидкости. Они, в свою очередь, равны  $m_1 = \rho \Delta V_1$  и  $m_2 = \rho \Delta V_2$ . Здесь  $\Delta V_1$  и  $\Delta V_2$  — объемы жидкости, пронесимые через сечения  $S_1$  и  $S_2$  за единицу времени.

Из уравнения неразрывности струи  $Sv = \text{const}$  следует, что  $S_1 \Delta x_1 = \Delta V_1 = S \Delta x_2 = \Delta V_2 = \Delta V$ ,

где  $\Delta x_1$  и  $\Delta x_2$  — расстояния, проходимые бесконечно тонкими «срезами» жидкости за единицу времени.

Таким образом,  $\Delta T_{\text{кин}} = (\rho/2)\Delta V(v_1^2 - v_2^2)$ .

Прирост кинетической энергии должен быть равен работе сил давления  $P_1$  и  $P_2$ . Учитывая, что они действуют в разном направлении, найдем

$$\Delta A = F_1 dx_1 - F_2 dx_2 = P_1 S_1 \Delta x_1 - P_2 S_2 \Delta x_2 = (P_1 - P_2) \Delta V.$$

На основании равенства  $\Delta T_{\text{кин}} = -\Delta A$  получим:

$$(\rho/2)(v_1^2 - v_2^2) = -(P_1 - P_2).$$

Минус берется потому, что работа производится над жидкостью за счет энергии внешнего источника. Переходя теперь к близко расположенным на расстоянии  $\Delta x$  вдоль оси трубы сечениям и дифференциалам, найдем:

$$\rho v (dv/dx) \Delta x = -(dP/dx) \Delta x, \quad \text{или} \quad \rho v (dv/dx) = -(dP/dx).$$

Учитывая, что  $v = \text{const}/S$ , можно, далее, записать:

$$dv/dx = -(S/\rho \text{const})(dP/dx).$$

Значит, если  $(dv/dx) > 0$  (скорость вдоль направления трубы возрастает), то  $(dP/dx) < 0$ . Стало быть, давление вдоль того же направления падает. Принимая во внимание, что  $v$  пропорционально  $S^{-1}$ , заключим, что давление будет большим на более широких участках трубы.

Соотношение  $\rho v (dv/dx) = -dP/dx$  можно переписать в другом виде, который окажется полезным при дальнейшем обсуждении. Для этого продифференцируем уравнение неразрывности струи по координате  $x$ . Получим:

$$d(Sv)/dx = 0 = v dS/dx + S dv/dx.$$

Отсюда

$$dv/dx = -(v/S)(dS/dx) = -(\text{const}/S^2)(dS/dx)$$

и, далее,

$$\rho v = (S^2/\text{const})(dS/dx)^{-1}(dP/dx).$$

Величины  $dS/dx$  и  $S$  — характеристики трубы в выбранной ее части и от условий протекания жидкости не зависят. Поэтому можно записать:  $\rho v = a(dP/dx)$ , где  $a$  — некоторый постоянный коэффициент.

Величина  $\rho v$  есть поток жидкости через единицу поверхности, расположенной перпендикулярно к направлению течения струи жидкости.

Теперь вспомним другие соотношения этого параграфа:

$$\Delta n \sim D dn/dx; \quad \Delta T_{\text{кин}} \sim \eta dT/dx; \quad \Delta p \sim \xi dV/dx.$$

Присоединим к этому и закон Ома в дифференциальной форме, т. е. соотношение между потоком заряженных частиц  $j_x$  в направлении  $X$  и изменением потенциала электрического поля (см. § 7.3):

$$j_x = -\rho^{-1} \partial \rho / \partial x.$$

Несмотря на разнородность всех этих явлений, которые объединяет лишь то, что все они характеризуют неравновесные процессы, мы получили один и тот же результат: поток некоторой величины  $J$  вдоль направления  $X$  с точностью до постоянной (при фиксированных условиях проведения эксперимента) равен компоненте вдоль того же направления градиента соответствующей функции.

По аналогии с обобщенной силой в механике ( $j_x = \partial L / \partial x$ , где  $L$  есть функция Лагранжа) такие компоненты называются обобщенными термодинамическими силами, или просто термодинамическими силами. Обычно их обозначают символами  $X$ , а коэффициенты пропорциональности символами  $L$ .

Тогда все вышеприведенные соотношения можно записать в одном виде:

$$J = LX.$$

В общем случае, когда поток зависит от градиента нескольких величин (одновременный перепад температуры, плотность газа и т. д.) и анализируется несколько процессов вблизи от равновесия, они описываются системой равенств:

$$J_1 = L_{11}X_1 + L_{12}X_2 + \dots + L_{1n}X_n,$$

$$J_2 = L_{21}X_1 + L_{22}X_2 + \dots + L_{2n}X_n,$$

.....

$$J_n = L_{n1}X_1 + L_{n2}X_2 + \dots + L_{nn}X_n.$$

Постоянные коэффициенты  $L_{nk}$  называются *кинетическими коэффициентами*.

В общей теории неравновесных процессов показывается, что  $L_{nk} = L_{kn}$  ( $k \neq n$ ), если отсутствуют внешние воздействия (теорема Онсагера).

Коэффициенты  $L_{kk}$  (одинаковые индексы!) называются также *собственными* и всегда положительны.

Коэффициенты  $L_{nk}$  при  $k \neq n$  (они называются *коэффициентами взаимности*) могут быть как положительными, так и отрицательными. Неравновесные процессы, описываемые подобными уравнениями, получили название *линейных*. В более сложных (нелинейных) процессах коэффициенты  $L_{nk}$  уже перестают быть постоянными.

## 5.5. Реальные газы

До сих пор мы пользовались моделью идеального газа, в котором свойства частиц близки к свойствам материальных точек и для которого взаимодействия между частицами наступают только тогда, когда эти частицы по закону абсолютно упругого удара сталкиваются между собой. Мы также предполагали, что размеры частиц бесконечно малы, так что уменьшать объем данной массы газа можно было практически до нуля.

Правда, при рассмотрении явлений переноса нам пришлось отступить от этой идеальной модели и считать, что частицы имеют хоть и малые, но конечные размеры. В противном случае столкновения между частицами просто не возникали бы. Учет размеров частиц есть первый шаг при переходе от идеального газа к реальным, так как эффективные размеры частиц различны для различных газов.

Посмотрим, какие изменения надо внести в уравнение состояния газов, если учитывать свойства реальных частиц. Прежде всего надо, очевидно, принять во внимание то, что если частицы имеют конечные размеры, то считать газ так, чтобы объем области, в котором он находился, уменьшился до нуля, уже нельзя. Если представить молекулы в виде шариков конечных размеров и рассмотреть газ в цилиндре под поршнем, то легко понять, что двигать поршень в сторону дна цилиндра можно только до тех пор, пока под ним не окажется совокупность плотно уложенных шариков-молекул. Для того чтобы еще дальше уменьшать свободный объем под поршнем, надо начать сжимать сами эти шарики, что, конечно, потребует несравненно больших усилий. Все это указывает на то, что в уравнение состояния газа надо ввести поправку, которая учитывала бы тот совокупный объем, который занимают плотно уложенные друг относительно друга молекулы, меньше которого нельзя сократить область пространства, занимаемую газом. Нетрудно видеть, что все рассуждения, с помощью которых мы получили уравнение состояния идеального газа  $pV = RT$ , можно сохранить и в том случае, когда предполагается, что частицы обладают определенным общим объемом. Для этого достаточно учесть, что реальное изменение объема под поршнем в цилиндре может проводиться не до нуля, а только до некоторой величины  $b$ . Уравнение состояния тогда переписывается в виде

$$p(V - b) = RT.$$

Это уравнение уже отражает специфику каждого газа, ведь постоянная  $b$  характеризует свойства вещества и зависит от размеров и форм частиц; для разных газов она будет различной.

Другим отличием реальных газов от идеального является наличие межмолекулярных взаимодействий, среди которых основную роль играют электростатические взаимодействия. Действительно, как известно из школьного курса, молекулы хотя и электронейтральны, но могут обладать дипольным моментом. Он



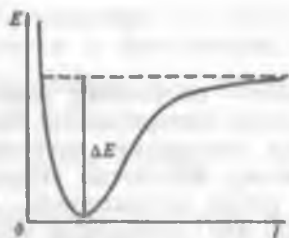


Рис. 5.7. Потенциальная энергия для двух взаимодействующих молекул как функция расстояния между ними ( $\Delta E$ —глубина потенциальной ямы)

возникает в том случае, когда имеется неравномерность распределения зарядов в самой молекуле (распределения заряда, создаваемого электронами и ядрами атомов). За счет этого вокруг молекул возникает электрическое поле, напряженность которого по модулю пропорциональна  $l^{-3}$ , где  $l$ —расстояние от молекулы до данной точки пространства. Хотя это поле убывает значительно быстрее, чем поле заряженной частицы, но тем не менее оно существует, проявляется на значительно меньших расстояниях, чем поле точечного заряда. Более подробное рассмотрение показывает, что даже если молекулы не обладают дипольным моментом, так как все равно в любой молекуле имеется некоторая неравномерность распределения положительных и отрицательных зарядов, то на близких расстояниях от молекулы наблюдается некоторое электростатическое поле, которое спадает с ростом расстояния от молекулы обратно пропорционально четвертой степени этого расстояния.

Поэтому на малых расстояниях молекулы испытывают электростатические взаимодействия, в частности притяжение, даже если они являются электрически нейтральными и бездипольными.

Конечно, характер таких взаимодействий целиком определяется свойствами самих молекул и для каждого сорта газа индивидуален. Если обозначить  $l$ —среднее расстояние между двумя молекулами,  $E = E(l)$ —потенциальную энергию взаимодействия, то потенциальную функцию межмолекулярных взаимодействий можно изобразить в виде кривой, показанной на рис. 5.7. Минимум этой функции, как и всегда, отвечает наиболее устойчивому положению равновесия, правый хвост отвечает области притяжения, а левая резко возрастающая часть кривой определяет область взаимного отталкивания молекул. Часто, пользуясь современными методами квантовой химии, можно вычислить это взаимодействие для конкретных случаев двух взаимодействующих молекул.

Уже очень давно такое взаимодействие было известно физикам, в частности на основании следующего простого эксперимента. Представим, что в некоторой области пространства за перегородкой находится плотно сжатый газ. Затем внезапно эта перегородка вынимается, и газ свободно распространяется во всем объеме сосуда. Оказывается, что в результате не только газ распределится по всему объему, но и понизится его температура. Если бы между частицами газа существовали только такие

взаимодействия, которые удовлетворяют условиям упругого удара, а эти условия таковы, что при ударе должны сохраняться полная энергия соударяющихся частиц и их полный импульс, то никакого понижения температуры при таком распространении газа по всему объему не должно было бы наблюдаться. Ситуация меняется, если рассматривать частицы, которые притягиваются друг к другу на расстоянии, значительно большем, чем область непосредственного контакта. Тогда если одна частица отлетает от другой, то она преодолевает силы притяжения и, следовательно, должна потерять часть своей кинетической энергии. В результате и наблюдается понижение температуры при расширении первоначально сжатого газа, так как уже все частицы начинают при таком расширении терять свою кинетическую энергию, преодолевая слабые силы сцепления.

Посмотрим, что нужно сделать, чтобы принять во внимание возникающие в реальных газах силы межмолекулярных взаимодействий. Существуют несколько способов, в известной степени эмпирических, основанных не на строгих рассуждениях, а на догадке, учета таких взаимодействий. Они приводят к разным формам уравнений состояния для реальных газов. Наибольшее распространение получило уравнение состояния в форме Ван-дер-Ваальса (по имени физика, его предложившего).

Для того чтобы получить некоторые указания на то, что именно нужно изменить в уравнении состояния идеального газа, чтобы учесть наличие межмолекулярных сил, представим крайний случай, когда силы взаимного притяжения между частицами значительны. Это означает, что глубина потенциальной ямы на кривой взаимодействия пары молекул, т. е. величина  $\Delta E$  (см. рис. 5.7), достаточно велика:  $\Delta E \gg kT$ . Почему берется именно такое соотношение? Вспомним, что  $kT$ —это мера средней кинетической энергии молекулы при заданной температуре. Если выполняется соотношение  $\Delta E \gg kT$ , то это означает, что частица, попавшая в область потенциальной ямы, не сможет из нее выскочить, так как ее кинетическая энергия для этого слишком мала. Значит, состояние, когда частицы удерживаются одна около другой в течение длительного периода времени, зависит от двух величин: от  $\Delta E$ —глубины потенциальной ямы, которая связана с сортом взаимодействующих молекул, и от температуры, которую имеет в данных условиях эксперимента статистическая система. Если условие  $\Delta E \gg kT$  хорошо выполняется, то при сближении двух частиц в результате хаотического движения они образуют устойчивую пару или, как говорят, *димер*. К этой паре может подлететь в результате хаотического движения третья частица, которая будет захвачена этой парой, и образуется *тример*. Эти рассуждения можно продолжать, и в конце концов мы придем к заключению, что в результате хаотического движения все частицы, постепенно сталкиваясь друг с другом, образуют некоторый устойчивый конгломерат. Такая ситуация действительно реализуется, но не в газах, а в

жидкостях, для которых как раз и характерно выполнение условия  $\Delta E \gg kT$ . В результате жидкость, предоставленная самой себе, например в условиях эксперимента, производимого в космическом полете, принимает форму шара.

За счет взаимодействия между отдельными частицами, составляющими жидкость, происходит как бы сдавливание шара и образование некоторого поверхностного давления. Если повысить температуру частиц, находящихся в этом шаровом жидком образовании, то может наступить такая ситуация, когда  $\Delta E$  окажется приблизительно равной  $kT$  или даже меньше  $kT$ . В этом случае одна частица легко отрывается от другой частицы и жидкое состояние становится невозможным. Получается свободно или почти свободно разлетающаяся совокупность частиц, что характерно для газа. Таким образом, переход системы из газообразного в жидкое состояние связан с величиной  $\Delta E$  и температурой всего статистического ансамбля. При высоких температурах один и тот же сорт частиц находится в газообразном состоянии, при достаточно низких температурах этот конгломерат частиц превращается в жидкость. Появление в жидком состоянии за счет сил сцепления сил поверхностного натяжения или поверхностного давления наводит на мысль о том, что и в газе дополнительное притяжение между отдельными частицами должно приводить как бы к повышению давления, оказываемого со стороны стенок сосуда на газ. Стало быть, уравнение состояния для реальных газов должно быть еще подправлено и его надо записать в следующей форме:

$$(p + \Delta p)(V - b) = RT.$$

Будем рассуждать дальше. Для того чтобы силы сцепления между отдельными частицами обнаружились и оказали существенное, заметное действие, необходимо, чтобы частицы в среднем могли находиться на достаточно малых расстояниях друг от друга, что, очевидно, возможно только тогда, когда  $n_0$  достаточно велико. Напомним, что  $n_0$  — концентрация частиц. Следовательно,  $\Delta p$  должно возрастать с увеличением  $n_0$ . Если общее число частиц в системе равно постоянной Авогадро (мы всегда имеем дело с молекул вещества), то  $n_0 = N/V$ , где  $N$  — постоянная Авогадро;  $V$  — объем, занимаемый системой. Итак,  $\Delta p$  должна быть функцией объема, причем с увеличением объема  $\Delta p$  стремится к нулю. В общем случае  $\Delta p$  можно представить в виде ряда по обратным степеням объема:

$$\Delta p = \frac{c}{V} + \frac{a}{V^2} + \frac{d}{V^3} + \dots$$

Посмотрим, какое из этих слагаемых должно быть наиболее значимым. Рассмотрим для этого некоторую частицу, находящуюся на периферии статистического ансамбля, например на поверхности жидкости, находящейся в космическом корабле и имеющей форму

шара. Эта частица испытывает силу притяжения со стороны всех близко расположенных к ней частиц, находящихся в более глубоких слоях жидкости или газа. Такая сила притяжения пропорциональна числу частиц, непосредственно находящихся около выбранной, т. е. пропорциональна  $n_0$ . Давление есть сила, приходящаяся на единицу поверхности. Это значит, что нужно рассматривать не силу, действующую на одну единственную частицу, а силу общего притяжения некоторого слоя газа или жидкости, находящегося у поверхности, к внутренним слоям. Так как число поверхностных частиц также равно  $n_0$ , то дополнительное давление  $\Delta p$  пропорционально  $n_0^2$  или  $1/V^2$ . Значит, в разложении  $\Delta p$  по обратным степеням объема прежде всего следует сохранить член  $a/V^2$ . Таким образом, уравнение состояния газа может быть переписано в форме

$$\left(p + \frac{a}{V^2}\right)(V - b) = RT,$$

где  $a$  и  $b$ —некоторые постоянные коэффициенты, зависящие от сорта газа ( $b$ —коэффициент, пропорциональный общему объему всех частиц газа и, таким образом, связанный с эффективными размерами частиц газа;  $a$ —коэффициент, связанный с межмолекулярными взаимодействиями между частицами газа).

Уравнение  $(p + a/V^2)(V - b) = RT$  и представляет уравнение состояния реального газа в форме Ван-дер-Ваальса. При больших  $V$  это уравнение переходит в уравнение идеального газа. Поэтому можно сказать, что все газы, если они достаточно разреженные, должны вести себя как идеальные, что и подтверждается экспериментом. Таким образом, различие между реальными и идеальными газами будет наблюдаться в плотных средах при достаточно больших внешних давлениях, когда средние расстояния между частицами уже сравнительно небольшие и суммарный объем, занимаемый частицами газа, уже нельзя считать бесконечно малой величиной по сравнению с общим объемом, в котором находится статистическая система.

Изобразим на диаграмме  $p, V$  изотермы состояния газа, подчиняющегося уравнению Ван-дер-Ваальса при разных температурах  $T_1 < T_2 < T_3$  (рис. 5.8). Напомним, что в случае идеального газа семейство изотерм представляется семейством гипербол. Для реальных газов такие зависимости  $p$  от  $V$  наблюдаются лишь при достаточно высоких температурах. При снижении температур на графике появляются небольшие изгибы. Начиная с некоторой температуры, названной критической, этот изгиб будет достаточно заметным, так что образуется небольшая область, где изотерма окажется параллельной оси абсцисс. При еще более низких температурах появляется волнообразный участок, который при последующем снижении температуры становится все более и более ярко выраженным. Таким образом,

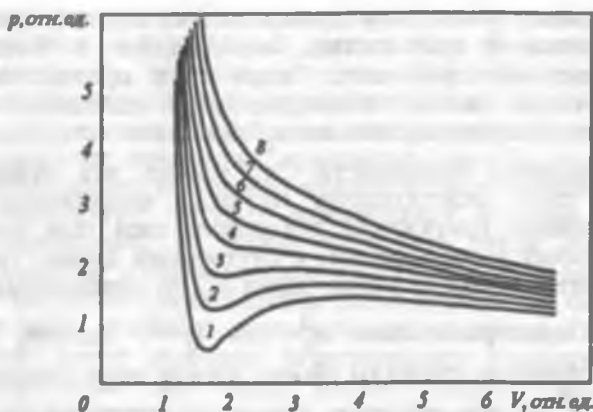


Рис. 5.8. Изотермы Ван-дер-Ваальса:  $T_1 < \dots < T_8$

изотермы реального газа при температурах ниже критической уже принципиально по своему виду отличаются от изотерм идеального газа.

Обратимся теперь к эксперименту и проверим ход изотерм Ван-дер-Ваальса, сжимая газ в цилиндре с поршнем в условиях, когда температура поддерживается постоянной. Мы увидим, что вплоть до критической температуры поведение изотерм будет таким, как предсказывает уравнение Ван-дер-Ваальса.

Однако в тех местах, где по уравнению Ван-дер-Ваальса следует ожидать появления волнообразных участков, на самом деле наблюдается существенно иная картина: вместо волнообразных участков появятся отрезки прямых (рис. 5.9). При достижении этих участков мы обнаружим также, что газ принципиально меняет свое состояние: появляются капли жидкости. Именно поэтому давление на соответствующем участке изотермы не меняется: газ постепенно переходит в жидкость. Когда весь газ перейдет в жидкое состояние, дальнейший ход цилиндра будет сопровождаться сжатием жидкости, что сделать очень трудно. Поэтому изотермы круто пойдут вверх.

То, что газ при снижении температуры и увеличении давления (т. е. увеличении числа частиц в единице объема и, следовательно, уменьшении среднего расстояния между частицами газа) должен превращаться в жидкость, следует из приведен-

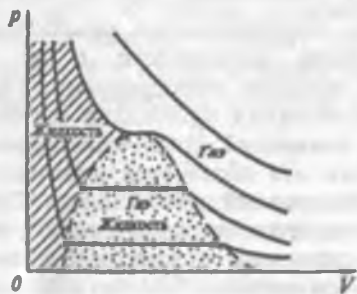


Рис. 5.9. Изотермы реального газа, системы газ—жидкость и жидкости

ного выше рассуждения о характере межмолекулярных взаимодействий. Неожиданным является то, что, хотя уравнение Ван-дер-Ваальса и не содержит в себе информации о фазовом переходе газ—жидкость, однако правильно предсказывает область  $p, V, T$ , где такой переход должен наблюдаться.

Везде, где есть участки горизонтальных линий на изотермах, присутствуют два компонента—газ и жидкость, которые находятся в равновесии. При уменьшении объема возрастает количество жидкости, а при увеличении его—газа. Газ, соответствующий таким участкам, называется насыщенным паром.

Анализируя уравнение Ван-дер-Ваальса, можно найти значения такой точки, где появление плоского участка только начинается. Эта точка называется критической, а соответствующие значения  $p, V, T$ —критическими значениями. Эти критические значения связаны с константами  $a$  и  $b$  следующим образом:

$$T_{кр} = \frac{8a}{27bR}, \quad V_{кр} = 3b, \quad p_{кр} = \frac{a}{27b^2}.$$

Видно, что при подходящих условиях (снижение температуры и сжатие) любой газ можно перевести в жидкое состояние.

Жидкий газ, даже помещенный в нормальные условия, способен долго сохранять свою температуру, подобно тому как сохраняет свою температуру вода, доведенная до кипения, до тех пор, пока вся не перейдет в пар. Следовательно, используя жидкие газы, можно в определенных объемах сохранять длительное время низкие температуры.

Как уже говорилось, для ожижения газов надо понизить их температуру ниже критической. Это, в свою очередь, можно сделать, если заставить газ совершать работу без доступа внешней энергии (см. раздел «Термодинамика»). Сравнительно легко получить жидкий азот, который при кипении сохраняет температуру около  $-170^\circ\text{C}$ . Такой жидкий азот широко применяется в научных исследованиях и технике. В металлургии для создания кислородного дутья используется жидкий кислород. Ожиженный пропан и другие газы служат в качестве источников энергии на транспорте и т. д.

#### Вопросы для самоконтроля

1. Что такое микроскопический и макроскопический подходы в физике?
2. При каких условиях водяной пар можно рассматривать как идеальный газ и при каких этого делать нельзя? Каково различие между обратимыми и необратимыми процессами в газах? Можно ли изобразить необратимый процесс на диаграмме объем—давление?
3. Почему при выводе уравнения состояния идеального газа с помощью микроскопического рассмотрения целесообразно и возможно вводить понятие средней скорости движения и оперировать именно ею?
4. Какой экспериментальный факт лежит в основе утверждения о том, что при поступательных движениях молекул энергия, приходящаяся на каждую степень свободы, является одной и той же?
5. Для объяснения какого явления необходимо принимать во внимание тот факт, что скорости движения молекул в реальном газе являются различными и меняются от нуля до некоторой достаточно большой величины?

6. Объяснить принцип регулирования температуры, который заложен в живых организмах.

7. Известно, что в Средней Азии, в областях с жарким климатом принято устраивать дворники, крыши которых состоят из переплетающихся ветвей дикого винограда. Внутри такого дворника поддерживается не очень высокая температура. Почему?

8. Какой принципиально новый класс явлений наблюдается в объектах, состоящих из многих частиц?

9. Какие явления переноса вам известны и как зависят коэффициенты, характеризующие эти явления, от температуры среды, массы молекул газа и плотности среды или количества частиц в единице объема? Что такое обобщенные термодинамические силы и как с их помощью можно единым образом описать все явления переноса?

10. Известно, что головка подсолнечника поворачивается в течение дня вслед за движением Солнца. Каким физическим явлением это можно объяснить?

11. Известно, что в российских космических кораблях для поддержания жизнедеятельности космонавтов применяется газовая смесь, близкая по составу к земному воздуху. В американских космических кораблях для этой цели применяют кислородно-гелиевую смесь. В каких условиях тепловая защита космонавтов должна быть более сильной, чтобы не происходило переохлаждение человеческого тела?

12. Какие особенности частиц реальных газов приводят к тому, что уравнение состояния реальных газов приходится изменить довольно существенно по сравнению с уравнением состояния идеальных газов?

13. Чем отличаются изотермы Ван-дер-Ваальса от изотерм, реально наблюдаемых при экспериментах с газами?

14. Чем объяснить тот факт, что кипящая жидкость сохраняет свою температуру?



## Г Л А В А 6

### ТЕРМОДИНАМИКА

*«Я развожу таких животных, которые находятся в термодинамическом равновесии с окружающей средой. Разводить теплокровных—это значит обогревать на свои деньги мировое пространство»*

В. Нернст

#### 6.1. Первое начало термодинамики

*Термодинамика*—раздел физики, изучающий общие закономерности обмена тепловой энергией между системами, системой и внешней средой и производства работы за счет этой энергии. Важнейшие выводы термодинамики широко используются в других науках, в частности в химии.

Мы будем заниматься лишь некоторыми, наиболее важными для физики, приложениями.

Понятия энергии и работы были введены в курсе механики, в частности, когда рассматривалась задача о работе, совершаемой при сжатии пружины. Эта работа приводит к возрастанию потенциальной энергии системы. Наоборот, если система совершает работу, то при этом ее потенциальная энергия уменьшается, причем расход потенциальной энергии равен производимой работе.

Изучая механику, мы говорили о двух формах энергии: кинетической и потенциальной. При переходе к изучению состояния газов мы ввели понятие температуры, рассматривая ее как меру энергии хаотического движения молекул. Энергию хаотического движения молекул мы называли *внутренней энергией газа*.

При изучении термодинамики для достижения максимальной наглядности будем широко пользоваться модельными представлениями. Самая общая термодинамическая модель должна включать три части: внешний источник тепловой энергии, рабочее тело и внешний приемник тепловой энергии, в который можно отводить энергию, накопленную в рабочем теле. Мы не будем конкретизировать ни тип источника или приемника энергии, ни



характер связи между ними и рабочим телом. Модель рабочего тела конкретизируем и примем в качестве него цилиндр с подвижным поршнем, внутри которого находится идеальный газ. Именно анализируя состояние идеального газа, легко ввести целый ряд весьма общих понятий.

Если с помощью внешнего источника тепловой энергии нагреть цилиндр, удерживая поршень неподвижным, то никакая макроскопическая работа совершаться не будет, но частицы газа будут двигаться быстрее, т. е. возрастает запас внутренней энергии рабочего тела. Энергия, переданная системе при ее нагревании, израсходовалась на повышение внутренней энергии газа. Одновременно с увеличением внутренней энергии, т. е. с повышением температуры газа, возрастет и его давление. Если не сдерживать поршень, то в результате он будет подниматься. Таким образом, в этом случае все подведенное количество теплоты расходуется двояко: во-первых, нагревается система, увеличивается ее внутренняя энергия и, во-вторых, перемещается поршень, совершается механическая работа.

Вообще, отвлекаясь от конкретной характеристики внутренней энергии системы, можно в общем случае сформулировать правило: количество теплоты  $\Delta Q$ , подведенное к системе, расходуется на повышение внутренней энергии  $\Delta E$  системы и на работу  $\Delta A$ , которая производится этой системой. Символически это правило записывается следующим образом:

$$\Delta Q = \Delta E + \Delta A$$

и называется *первым началом или первым законом термодинамики*.

Первое начало термодинамики является частным случаем закона сохранения энергии. В каком бы виде ни подводилась энергия к системе, всегда эта энергия будет израсходована на два процесса: 1) на повышение внутренней энергии системы; 2) на совершение системой работы. С проявлением закона сохранения энергии мы встречались в механике, когда говорили о сохранении полной энергии изолированной механической системы.

Важнейшими процессами, изучаемыми в термодинамике, являются изотермический и адиабатический. При изотермическом сжатии или расширении идеального газа его температура не меняется, т. е. не меняется внутренняя энергия. Вообще, вне зависимости от вида системы мы будем называть *изотермическим* процесс, при котором температура системы не меняется.

Для поддержания внутренней энергии неизменной при изменении состояния газа необходимо, чтобы он находился в контакте с внешней средой. Из внешней среды количество теплоты может поступать в систему или удаляться из нее. При изотермическом процессе  $\Delta E = 0$ , т. е.  $\Delta Q = \Delta A$ .

Рассмотрим другой случай. Представим, что контакт цилиндра с внешней средой исключен, например цилиндр обложен асбестом, который является хорошим изолятором. В этом случае  $\Delta Q = 0$ , т. е. теплота в систему не подводится и не отнимается.

Тогда  $\Delta E = -\Delta A$ . Такой процесс называется *адиабатическим*. При адиабатическом процессе любое изменение внутренней энергии связано с совершением данной системой работы.

Связь между  $\Delta A$  и  $\Delta E$  хорошо иллюстрируется опытом с нагреванием и охлаждением газа в цилиндре с поршнем при резком сжатии или расширении его.

Если газ внезапно расширяется, выталкивая поршень, то производимая им работа  $\Delta A$  должна быть равна изменению внутренней энергии  $\Delta E$ . Поскольку  $E$  зависит от температуры, должно наблюдаться уменьшение температуры газа.

Изменению температуры газа при быстром сжатии или расширении можно дать кинетическое объяснение. Рассматривая удары молекул о стенку сосуда, мы считали стенку неподвижной. При упругом ударе о стенку скорость молекулы по модулю не менялась. Представим теперь, что стенка движется, например, навстречу молекуле. Если в этом случае молекула столкнется с движущейся стенкой, то отскочит от нее со скоростью большей, чем до удара. Движение стенки будет передано молекуле. Следовательно, в результате столкновений молекул со стенкой скорости их будут возрастать, а это и означает повышение температуры. Если стенка будет отступать, то после удара частицы о стенку скорость частицы уменьшится. Значит, в этом случае температура понижается.

## 6.2. Теплоемкость

Теплоемкость характеризует количество теплоты, которое необходимо затратить, чтобы нагреть систему на  $1^\circ\text{C} = 1\text{ K}$ , т. е.  $c = \Delta Q / \Delta T$ . Как измерить эту теплоемкость? Количество теплоты  $\Delta Q$  в общем случае расходуется на повышение внутренней энергии системы и на работу этой системы. Поэтому

$$c = \frac{\Delta Q}{\Delta T} = \frac{\Delta E}{\Delta T} + \frac{\Delta A}{\Delta T}.$$

При изотермическом процессе ( $\Delta T = 0$ )  $c = \infty$ .

В других случаях теплоемкость зависит от того, производит ли система работу или нет. Поэтому для определенности надо выбрать стандартные условия. Первое заключается в том, что работа вообще не производится. Это значит, например, что осуществляется опыт с нагреванием газа в цилиндре с закрепленным поршнем. Все количество теплоты тогда расходуется только на повышение внутренней энергии. Если поршень в цилиндре не движется, то не изменяется и объем системы. Поэтому теплоемкость при  $\Delta A = 0$  ( $\Delta V = 0$ ) называется *теплоемкостью при постоянном объеме* и обозначается  $c_V$  (индекс  $V$  показывает, что нагревание происходит при постоянном объеме).

Легко видеть, что

$$c_V = \Delta E / \Delta T.$$

Обратимся к идеальному газу.

Напомним, что если принять во внимание только поступательные и вращательные степени свободы молекулы и учесть, что на каждую такую степень свободы в среднем приходится энергия, равная  $\frac{1}{2} kT$ , то полная внутренняя энергия моля идеального газа равна  $E = (i/2)RT$ . Следовательно,

$$\Delta E = (i/2)R\Delta T \quad \text{и} \quad c_V = (i/2)R.$$

Для одноатомного газа (например, гелия)  $i = 3$ , для двухатомного или состоящего из линейных молекул газа (например, ацетилена)  $i = 5$ , а для газа, состоящего из молекул сложной формы,  $i = 6$ . Теперь введем теплоемкость с учетом, что производится какая-то работа. Для определенности принято полагать, что при этом постоянным остается давление. Такая теплоемкость называется *теплоемкостью при постоянном давлении* и обозначается  $c_p$ . Чтобы найти эту величину, запишем первое начало термодинамики в форме

$$\Delta Q = \Delta E + \Delta A = c_V \Delta T + \Delta A.$$

Далее учтем, что работа в общем случае равна произведению силы на перемещение. В газе роль силы будет играть давление, а роль перемещения—изменение объема. Значит, вместо обычного выражения для работы можно записать  $\Delta A = p\Delta V$ .

Вспомним уравнение состояния газа:  $pV = RT$ . Если давление постоянно, то  $p\Delta V = R\Delta T$ . Отсюда  $\Delta Q = c_V \Delta T + R\Delta T$  (при постоянном давлении). Теплоемкость равна отношению  $\Delta Q/\Delta T$ . Разделив последнее равенство на  $\Delta T$ , получим

$$c_p = \Delta Q/\Delta T = c_V + R,$$

где  $R$ —молярная газовая постоянная, всегда положительная, поэтому  $c_p$  больше, чем  $c_V$ . Этот результат можно сразу предсказать на основании такого рассуждения. На что тратится теплота, когда система нагревается при постоянном объеме? Работа не совершается. Происходит только изменение внутренней энергии системы. Если мы рассматриваем теплоемкость при постоянном давлении, то внутренняя энергия тоже меняется, но, кроме того, дополнительно совершается работа. Но для того чтобы совершить работу, нужно затратить энергию за счет того количества теплоты, которое вводится в систему. Следовательно, необходимо ввести большее количество теплоты, чтобы изменить температуру системы на  $1^\circ\text{C}$ . Большей соответственно будет теплоемкость. Так как  $c_V = (i/2)R$ , то  $c_p = \frac{i}{2}R + R = R\left(\frac{i+2}{2}\right)$ . Как  $c_V$ , так и  $c_p$  зависят только от  $i$  и  $R$ .

Следовательно, все газы независимо от их природы будут иметь одну и ту же теплоемкость, если число степеней свободы

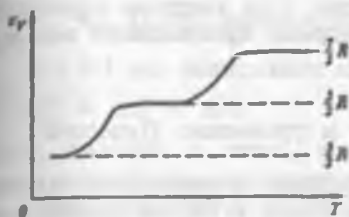


Рис. 6.1. Теоретическая (пунктир) и экспериментальная (сплошная линия) зависимости теплоемкости  $c_v$  от температуры.

Появление ступенек связано с постепенным возбуждением вращательных и колебательных движений ядер в молекуле

их молекул будет одинаковым. Здесь полезно сделать дополнительное замечание. При выводе формул для теплоемкости мы воспользовались теоремой о равномерном распределении средней энергии по степеням свободы, уравнив поступательные и вращательные степени свободы (см. § 5.4). Но мы намеренно ничего не сказали о колебательных степенях свободы, а они также существуют, если частицы сложные (молекулы) и у них могут меняться длины связей, валентные углы и т. д. На деформацию молекул также должна расходоваться подводимая к системе внешняя энергия (теплота). В результате хаотических столкновений при подаче энергии извне не только увеличится скорость их движения, но молекулы приобретут дополнительную потенциальную энергию, связанную с тем, что появится некоторая неизменная во времени постоянная средняя деформация всех частиц. Такой дополнительный расход внешней тепловой энергии и должен быть учтен в виде некоторой поправки к величине  $c_v$ .

В самом деле, более точные рассуждения, которые здесь не приведены, дают следующий результат:

$$c_v = (i_{\text{пост}} + i_{\text{вращ}} + 2i_{\text{кол}})R/2,$$

но по-прежнему  $c_p = c_v + R$ . Здесь  $i_{\text{пост}}$ ,  $i_{\text{вращ}}$ ,  $i_{\text{кол}}$  — соответственно числа поступательных, вращательных и колебательных степеней свободы.

Чем больше степеней свободы у молекулы, т. е. чем больше в ней атомов, тем больше теплоемкость. Поскольку теплоемкость связана с числом степеней свободы и постоянной  $R$ , то можно ожидать, что теплоемкость не зависит от температуры.

График любой теплоемкости, например  $c_v$ , как функции температуры, должен иметь вид прямой (рис. 6.1).

Оказывается, однако, что если провести эксперимент при высокой и низкой температурах, то зависимость изображается кривой, которая при более низкой температуре будет падать, а при высокой — возрастать. Лишь на «ступеньках» графика все будет так, как предсказывает классическая теория. Такое сложное поведение зависимости теплоемкости от температуры объясняется тем, что при низких температурах заметно проявляются квантовые эффекты, связанные с тем, что молекулы могут изменять свои энергетические вращательные и колебательные

состояния не непрерывно, а дискретно. При высоких температурах могут сказываться также и изомерные превращения молекул и их разрушение.

### 6.3. Адиабатический процесс и уравнение Пуассона

Обратимся к адиабатическому процессу и попробуем вывести уравнение, связывающее давление и объем в случае адиабатического расширения или сжатия газа.

Воспользуемся уравнением состояния идеального газа  $pV = RT$  и перейдем к дифференциалам:  $p dV + V dp = R dT$ . При адиабатическом процессе любая работа может совершаться только за счет изменения внутренней энергии.

Значит,  $p dV = -c_V dT$ , откуда  $p dV + V dp + \frac{R}{c_V} p dV = 0$  и

$$\frac{c_V + R}{c_V} p dV + V dp = \frac{c_p}{c_V} p dV + V dp = 0.$$

Отношение  $c_p/c_V$  принято обозначать  $\gamma$ . Тогда  $\gamma p dV + V dp = 0$ .

Умножив это равенство на  $V\gamma^{-1}$ , увидим, что оно превращается в дифференциал функции  $(pV^\gamma)$ , т. е. приобретает вид  $d(pV^\gamma) = 0$ . Отсюда окончательно получим уравнение адиабаты (уравнение Пуассона):

$$pV^\gamma = \text{const.}$$

Отношение  $c_p/c_V$  можно измерить, например, изучая распространение звука в воздухе. Когда распространяется звук, возникают периодические уплотнения воздуха. Они совершаются быстро, и теплота не успевает рассеиваться, т. е. происходит адиабатический процесс. В этом случае отношение между объемом и давлением определяется отношением  $c_p/c_V$ . Измерив скорость звука, можно вычислить это отношение. Скорость звука будет разной в зависимости от газов. Например, скорость в гелии одна, в парах воды—другая.

Из уравнения  $pV^{c_p/c_V} = pV^\gamma$  можно получить отношение между  $T$  и  $V$ :

$$pV^\gamma = pVV^{\gamma-1} = RTV^{\gamma-1} = \text{const.},$$

т. е.  $TV^{\gamma-1} = \text{const}$  для адиабатического процесса.

Адиабатический процесс играет большую роль в технике. При работе тепловых машин (паровые машины, двигатели внутреннего сгорания), где движение поршня совершается с большой скоростью, как правило, расширение и сжатие газа происходят адиабатически хотя бы на некоторых участках полного рабочего цикла.

## 6.4. Термодинамический потенциал—энтропия

В § 6.1 мы записывали первое начало термодинамики в виде  $\Delta Q = \Delta E + \Delta A$ .

Здесь имелось в виду, что знак  $\Delta$  характеризует некоторые приращения всех входящих в это выражение величин.

При вычислениях и общих рассуждениях удобно конкретизировать математический смысл таких приращений. Для этого надо перейти к строго определенным в математике дифференциалам. Напомним, что понимается под дифференциалом.

Если функция  $y$  зависит только от одной переменной  $x$ , то дифференциалом  $dy$  этой функции называется ее бесконечно малое приращение при бесконечно малом приращении аргумента  $dx$ .

Выполняется соотношение  $dy = y' dx$ , где  $y'$ —производная от функции  $y$  по аргументу  $x$ .

Если  $z$ —функция двух аргументов  $x$  и  $y$ , то для нее вводится понятие полного дифференциала  $dz$  так, что

$$dz = \frac{\partial z(x, y)}{\partial x} dx + \frac{\partial z(x, y)}{\partial y} dy,$$

где  $\frac{\partial z(x, y)}{\partial x}$  — частная производная функции  $z$  по аргументу  $x$ .

Для нас важно следующее: если существует некоторая функция  $z(x, y)$ , полный дифференциал которой  $dz$  есть подынтегральное выражение в определенном интеграле

$$\int_1^2 dz = \int_1^2 \left( \frac{\partial z}{\partial x} dx + \frac{\partial z}{\partial y} dy \right),$$

то значение этого интеграла не зависит от пути, соединяющего точки 1 и 2 на плоскости  $XY$ .

Функция  $z$  в этом случае называется *первообразной полного дифференциала*  $dz$ .

Именно это свойство мы и использовали ранее, когда вводили понятие потенциальной функции в механике. Напомним, что работа на участке пути определялась нами так:

$$\Delta A = \int_1^2 (f_x dx + f_y dy)$$

(если ограничиться плоскостью), где  $f_x$  и  $f_y$ —компоненты силы.

В свою очередь,

$$f_x = - \frac{\partial U(x, y)}{\partial x}, \quad f_y = - \frac{\partial U(x, y)}{\partial y},$$

где  $U(x, y)$ —потенциальная функция (если силы, разумеется, потенциальные). Значит, элемент работы  $dA$  есть также полный дифференциал.

Если, например, ввести потенциал  $\varphi(x, y)$  для электростатического поля, то элемент работы равен

$$dA = -\frac{\partial \varphi}{\partial x} dx - \frac{\partial \varphi}{\partial y} dy, \quad \Delta A = -\int_1^2 \left( \frac{\partial \varphi}{\partial x} dx - \frac{\partial \varphi}{\partial y} dy \right).$$

Теперь заметим, что такой определенный интеграл равен разности взятых в точках 2 и 1 первообразных подынтегральных выражений (или разности потенциалов!). Это значит, в свою очередь, что  $\Delta A$  уже не зависят от пути от точки 1 к точке 2, и если двигаться от точки 2 снова к точке 1 по другому пути (полное движение совершается по замкнутому контуру), то  $\Delta A = 0$ .

Это важнейшее свойство физических величин, малые изменения которых можно характеризовать полными дифференциалами подходящих функций, и представляет большой интерес не только с вычислительной точки зрения, но и для формулировки фундаментальных утверждений. Рассмотрим с этой точки зрения снова идеальный газ. Его состояние мы характеризуем тремя величинами:  $p$ ,  $V$  и  $T$ .

Приращения этих величин можно связать с соответствующими дифференциалами:  $dp$ ,  $dV$  и  $dT$ . Именно поэтому мы выше и могли записать уравнение состояния газа в дифференциальной форме. Но вот величины  $p dV$  или  $V dp$  уже не есть полные дифференциалы, так как, в частности, работа газа  $\Delta A = p dV$  зависит от пути, по которому состояние газа изменялось. Значит, не является полным дифференциалом и величина  $\Delta Q$ —переданное в систему количество теплоты.

Посмотрим, нельзя ли в термодинамике ввести такую функцию, приращение которой, с одной стороны, зависело бы от измеряемых параметров, а с другой, обладало бы свойством полного дифференциала.

Оказывается, это действительно можно сделать. Обратимся к выражению первого начала термодинамики, записав его для идеального газа:

$$\Delta Q = c_V \Delta T + p \Delta V = c_V dT + p dV = c_V dT + RT \frac{dV}{V}.$$

Если разделить это выражение на  $T$ , то нетрудно увидеть, что  $\Delta Q/T$ —полный дифференциал некоторой функции, которую принято обозначать  $S$  и называть *энтропией* или *приведенной теплотой* (последний термин сейчас практически не употребляется, хотя и представляется более ясным, чем термин «энтропия»).

В самом деле,  $\frac{\Delta Q}{T} = c_V \frac{dT}{T} + R \frac{dV}{V}$ , но  $\frac{1}{T} = \frac{d(\ln T)}{dT}$ .

Аналогично получаем выражение для  $1/V$ . В результате имеем

$$\frac{\Delta Q}{T} = \frac{d(\ln T^{c_V})}{dT} dT + \frac{d(\ln V^R)}{dV} dV.$$

Если учесть, что  $\ln(ab) = \ln a + \ln b$ , то

$$\frac{\Delta Q}{T} = \left[ \frac{\partial}{\partial T} \ln(T^{c_V} V^R) \right] dT + \left[ \frac{\partial}{\partial V} \ln(T^{c_V} V^R) \right] dV.$$

Если принять теперь  $S = \ln(T^{c_V} V^R) + \text{const}$ , то и получим

$$\frac{\Delta Q}{T} = dS = \frac{\partial S}{\partial T} dT + \frac{\partial S}{\partial V} dV.$$

Здесь  $dS$ —полный дифференциал, а это приводит к важнейшему свойству энтропии: изменение ее при изменении состояния системы, соответствующего замкнутому контуру на  $p, V$ -диаграмме, равно нулю.

Применительно к газу, это означает, что если газ произвольным образом менял свое состояние, но вернулся к исходному (совершается замкнутый цикл), то энтропия его не меняется. Стало быть, энтропию можно рассматривать как некоторый потенциал и разного рода процессы с передачей теплоты характеризовать с помощью этого потенциала. При этом энтропия есть функция состояния самой системы и ее изменение зависит лишь от начального и конечного состояния системы, но не от пути перехода между этими состояниями. Заметим, что такое выражение для энтропии применимо лишь для обратимых квазистационарных процессов, т. е. процессов, которые можно рассматривать как последовательность мало отличающихся воспроизводимых в любом направлении равновесных состояний.

В общем случае для участка обратимого процесса имеем:

$$\Delta S = \ln(T_2^{c_V} V_2^R) - \ln(T_1^{c_V} V_1^R) = c_V \ln \frac{T_2}{T_1} + R \ln \frac{V_2}{V_1}.$$

Для изотермического участка:  $\Delta S_{\text{изот}} = R \ln \frac{V_2}{V_1}.$

Для изохорического ( $V = \text{const}$ ):  $\Delta S_{\text{изох}} = c_V \ln \frac{T_2}{T_1}.$

Для адиабатического:  $\Delta S_{\text{адиаб}} = 0.$

В самом деле, для адиабатического процесса имеем  $TV^{\gamma-1} = \text{const}$ . Значит, если взять  $\ln$  для двух состояний газа (1 и 2), то получим:



$$\ln T_1 + (\gamma - 1) \ln V_1 = \ln (\text{const}),$$

$$\ln T_2 + (\gamma - 1) \ln V_2 = \ln (\text{const}).$$

Вычитая эти равенства одно из другого и принимая во внимание, что  $\gamma - 1 = (c_p - c_v)/c_v = R/c_v$ , находим

$$c_v \ln \frac{T_1}{T_2} + R \ln \frac{V_2}{V_1} = \Delta S = 0.$$

Используя понятие энтропии, запишем первое начало термодинамики так:

$$T dS = c_v dT + p dV.$$

Значит, при изотермическом процессе работа газа  $p dV = T dS$ .

Рассмотрим теперь изотермическое расширение газа от объема  $V_1$  до объема  $V_2$ . Обсудим некоторые вероятностные соотношения.

Пусть мы с помощью некоторого прибора определяем наличие какой-то частицы газа в единичном объеме. Очевидно, отношение вероятностей найти частицу в единичном объеме перед началом расширения и после его окончания должно быть обратно пропорционально отношению  $V_1/V_2$ . Если обозначить такую вероятность  $\alpha$ , то  $\alpha_1/\alpha_2 = V_2/V_1$ . Соответствующее отношение вероятностей для большого числа  $N$  независимых частиц равно  $(\alpha_1/\alpha_2)^N = (V_2/V_1)^N$  (вероятности независимых событий перемножаются). Для изотермического процесса:

$$\Delta S = R \ln \frac{V_2}{V_1} = \frac{R}{N} N \ln \left( \frac{\alpha_1}{\alpha_2} \right) = k \ln \left( \frac{\alpha_1}{\alpha_2} \right)^N,$$

если принять  $N$  равным постоянной Авогадро (число молекул в моле). Это позволяет придать изменению энтропии  $\Delta S$  наглядное физическое толкование (см. гл. 10).

## 6.5. Тепловые машины

Этот параграф посвящен работе тепловых машин. Тема эта представляет большой практический интерес. Мы узнаем, в частности, почему нельзя сделать тепловую машину с коэффициентом полезного действия выше определенной величины.

Представим рабочую часть тепловой машины в виде цилиндра с поршнем, внутри цилиндра находится идеальный газ. Поршень с помощью штока соединен с механизмом, воспринимающим движение, например с маховым колесом. Будем считать, что масса газа во время работы машины остается одной и той же и газ не меняется. Такое условие является верным лишь в некоторых случаях, например для тепловой машины, действующей под давлением водяного пара.

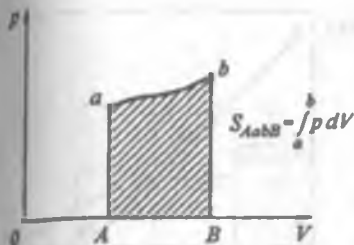


рис. 6.2. Работа газа на пути  $a \rightarrow b$  определяется площадью под кривой



Рис. 6.3. При обратимом процессе работа газа определяется площадью внутри замкнутой кривой между точками  $a$  и  $b$

В двигателе внутреннего сгорания процесс более сложный. Вначале имеется какая-то смесь, затем происходит химический процесс и вместо смеси получают продукты сгорания. Однако в определенном приближении и в этом случае можно считать, что газ в цилиндре не меняется.

Если газ переходит от состояния  $a$  в состояние  $b$  так, что изменение его состояний изображается кривой  $ab$ , то газ производит работу.

Напомним, что работа газа при малом изменении объема равна  $dA = p dV$ .

Полная работа газа на криволинейном участке диаграммы от  $a$  до  $b$  равна сумме работ на каждом отрезке кривой  $ab$ . Такая сумма приводится к определенному интегралу

$$\int_a^b p dV = A.$$

Заметим, что по абсолютной величине интеграл  $\int_a^b p dV$  равен площади под кривой  $ab$  (рис. 6.2) на  $p, V$ -диаграмме.

Посмотрим, как работает тепловая машина. Для того чтобы получилось непрерывное движение, необходимо, чтобы поршень много раз поднимался и опускался.

Зависимость  $p$  от  $V$  на диаграмме  $p, V$  изображается участками кривых. Для получения непрерывного движения необходимо, чтобы после каждого цикла система возвращалась к исходному состоянию. На графике можно изобразить это следующим образом (рис. 6.3). Газ при расширении совершает работу, сообщая движение внешнему устройству. При сжатии газа, наоборот, внешнее устройство совершает работу над газом. Работа газа над внешней системой по сравнению с работой, совершаемой внешней системой над газом, больше.

Полезная работа равна разности этих двух работ (площадь кривой, заключенной между нижней и верхней кривыми на участке от  $a$  до  $b$ ).



Рис. 6.4. Участок изотермы

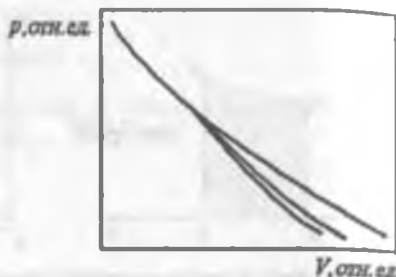


Рис. 6.5. График адиабат; адиабаты (две нижние кривые) идут круче изотермы и отвечают одноатомному газу и газу, состоящему из сложных молекул

Найдем работу газа при изменении его состояния для некоторых простейших случаев.

Пусть давление газа меняется от некоторого значения в точке  $a$  до значения в точке  $b$  по изотерме (рис. 6.4). Мы знаем, что при изотермическом процессе  $p dV = T dS$ , значит,

$$\Delta A = T \Delta S = TR \ln \frac{V_b}{V_a} \quad (T = \text{const}).$$

Итак, работа при изотермическом изменении объема газа равна произведению молярной газовой постоянной  $R$  на температуру  $T$ , соответствующую этой изотерме, и на логарифм отношения конечного объема к начальному объему газа.

Второй процесс, который мы рассмотрим, адиабатический (рис. 6.5).

При адиабатическом процессе отсутствует теплообмен с окружающей средой, следовательно,  $\Delta E = -p \Delta V$ , откуда  $p dV = -c_V dT$ , т. е. работа на малом участке адиабаты для газа равна теплоемкости  $c_V$ , умноженной на приращение температуры. Полная работа при адиабатическом процессе равна

$$A = \int_a^b p dV = - \int_a^b c_V dT = -c_V \int_a^b dT = -c_V (T_b - T_a).$$

Теперь можно перейти к изучению работы тепловой машины по циклу Карно. Все этапы этого цикла считаются обратимыми.

Значение этого цикла заключается, в частности, в том, что коэффициент полезного действия машины Карно является наибольшим из всех возможных при данных условиях.

Рассмотрим цилиндр с газом. Будем считать, что вначале происходило изотермическое расширение газа. Это значит, что к системе все время подводится какое-то количество теплоты от источника, который мы будем называть *нагревателем* (рис. 6.6).

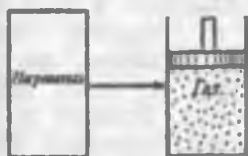


Рис. 6.6. Часть элементов тепловой машины

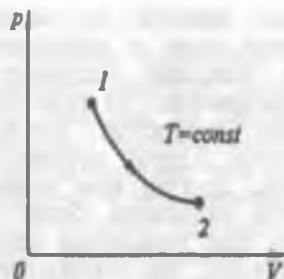


Рис. 6.7. Первый шаг цикла Карно

Обозначим исходное состояние газа на диаграмме  $p, V$  точкой 1 (рис. 6.7). Рассмотрим изменение состояния газа по изотерме до точки 2. Чтобы поддерживать температуру постоянной, нагреватель должен иметь температуру не ниже или равную температуре  $T_1$ . После того как газ достигнет при изотермическом расширении точки 2, совершив резкое адиабатическое расширение газа до точки 3 (рис. 6.8). Адиабата идет круче изотермы. При адиабатическом процессе теплота в систему не поступает, значит, можно считать, что, когда совершается адиабатический процесс, разъединяется ключ, который соединяет нагреватель и цилиндр. В результате адиабатического расширения газ охладится и достигнет температуры  $T_2$ , причем  $T_2 < T_1$ . В результате этих двух процессов поршень поднимется до предела. Для того чтобы осуществить непрерывное движение, поршень нужно опустить вниз, т. е. вернуться к исходному состоянию и снова начать сжимать газ. На это сжатие затрачивается энергия извне, т. е. должна быть совершена внешняя работа. Сжатие газа совершается по изотерме до точки 4 (рис. 6.9). Но при сжатии газ должен был бы нагреваться. Для того чтобы он не нагревался, необходимо от газа все время отводить теплоту. Как это можно сделать? Здесь мы встречаемся с проявлением второго начала термодинамики, о котором будем говорить подробнее позднее. Сейчас же сформулируем это важнейшее начало следующим

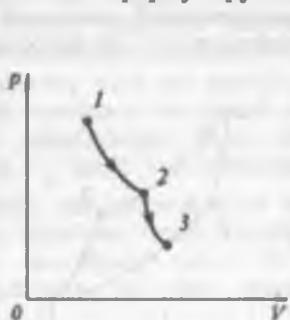


Рис. 6.8. Продолжение шагов цикла Карно

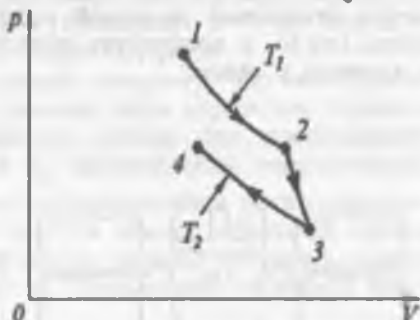


Рис. 6.9. Три первых шага цикла Карно

образом: теплота не может переходить самопроизвольно от холодного тела к нагретому телу. Например, если взять два бруска металла, нагретых неодинаково, и привести их в соприкосновение, то тело более горячее охладится, а тело более холодное нагреется. Теплота перейдет от горячего тела к холодному. Другой пример. Если слить две жидкости, одну при большей, другую при меньшей температуре, то более горячая жидкость охладится, более холодная нагреется. Вообще, во всех случаях при появлении контакта двух систем, имеющих разные температуры, их температуры будут выравниваться с течением времени, если только при этом за счет внешнего источника энергии не совершается какая-нибудь работа. Это приводит к важному следствию. Для того чтобы на участке от точки 3 до точки 4 осуществить изотермическое сжатие газа, необходимо из цилиндра отводить теплоту. Эта теплота не может снова перейти в нагреватель, так как  $T_2 < T_1$ . Значит, для того чтобы отвести эту теплоту, надо еще какое-то дополнительное устройство. Это устройство называется *холодильником*.

Итак, для циклической работы любой тепловой машины по второму началу термодинамики необходимо иметь нагреватель и холодильник, а между ними — цилиндр с поршнем (рис. 6.10). И вот именно эта необходимость иметь холодильник приводит к тому, что коэффициент полезного действия машин не может быть равен единице, а всегда меньше ее.

Чтобы вернуться в исходное состояние 1 (рис. 6.11), совершим адиабатическое сжатие газа, отсоединив при этом цилиндр от холодильника.

При адиабатическом сжатии газа произойдет повышение его температуры от  $T_2$  до  $T_1$  (см. рис. 6.11). В результате этих процессов получаем замкнутый цикл: от первой точки до второй по изотерме, от второй к третьей по адиабате, от третьей к четвертой по изотерме и от четвертой к первой снова по адиабате. Этот цикл и называется *циклом Карно* или *идеальным циклом*.

*Коэффициентом полезного действия* тепловой машины называется отношение полезной работы, совершаемой машиной при одном цикле, к количеству теплоты, переданному из нагревателя в цилиндр с газом:

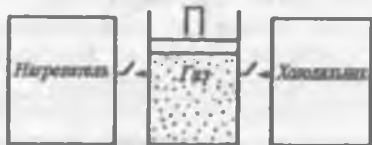


Рис. 6.10. Схема тепловой машины

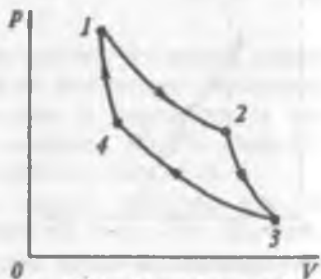


Рис. 6.11. Полный цикл Карно

$$\eta = \Delta A / \Delta Q.$$

Таким образом, если количество теплоты полностью переходит в работу, то коэффициент полезного действия двигателя равен единице. Такой двигатель был бы идеален, но создать его нельзя и не только потому, что происходят потери энергии на трение. Поэтому количество теплоты всегда больше, чем работа, которую совершает газ за счет этой теплоты, а это означает, что коэффициент полезного действия всегда меньше единицы. Для того чтобы подсчитать КПД, найдем работу цикла Карно, которая производится, когда совершается весь цикл от точки 1 к точке 2, от 3 к 4 и снова от точки 4 к точке 1. Однако прежде чем это сделать, выясним, почему именно цикл Карно является при заданных ограничениях на возможные изменения объема и температуры самым выгодным.

Будем рассуждать так. Примем, что в любом случае объем может изменяться только в пределах от  $V_1$  до  $V_3$ , а температура от  $T_1$  до  $T_2$ .

Предположим, что после движения по изотерме до точки 2 переход к точке 3 идет не по адиабате, а каким-то другим образом. Однако любой путь, кроме адиабаты, требует дополнительного количества теплоты от внешнего источника (возрастет знаменатель в выражении  $\eta = \Delta A / \Delta Q$  и уменьшится КПД цикла). Допустим, что и движение от точки 4 к точке 1 совершается не по адиабате, а иным образом. Но в этом случае надо передавать в систему не только энергию, достаточную для повышения температуры, но еще и совершить работу в соответствии с требованиями первого начала термодинамики. Но эта работа совершается над системой и, следовательно, уменьшает полезную работу машины. Поясним сказанное примером. Пусть после достижения точки 4 система при неподвижном поршне нагревается до температуры  $T_1$ . Для этого надо от нагревателя передать энергию  $\Delta Q_{41} = c_V(T_1 - T_2)$ . Очевидно, что  $\Delta Q_{41}$  будет равна работе внешних сил при адиабатическом сжатии  $4 \rightarrow 1$ . Но надо еще дополнительно сдвинуться от точки пересечения изохоры (вертикали) с изотермой 1—2 до точки 1, а это потребует дополнительных затрат энергии.

Значит, путь по адиабатам самый выгодный. Таким образом, любой наиболее выгодный цикл должен быть ограничен справа и слева адиабатами. Если потребовать, чтобы при произвольном цикле  $T_1$  была максимальной, а  $T_2$  минимальной температурой, то ясно, что всякий путь вправо, отличающийся от изотермического, пройдет ниже изотермы 1—2, а всякий обратный путь — выше изотермы 3—4. В результате общая площадь цикла уменьшится, а это приведет к уменьшению полезной работы  $\Delta A$ .

Цикл Карно действительно является наивыгоднейшим, и поэтому его КПД есть максимальная при данном перепаде

температур величина. Все другие циклы в этом смысле будут только хуже.

Вычислим теперь этот коэффициент. Это очень просто сделать, если воспользоваться понятием энтропии и учесть, что для изотермы работа равна  $\Delta A = T\Delta S$ , а для адиабаты  $\Delta A = -c_V(T_2 - T_1)$ . Тогда, поскольку участки адиабат 2—3 и 4—1 соединяют одни и те же изотермы, работа (полезная) на участке 2—3 полностью компенсируется работой над системой на участке 4—1. Полная работа определяется только движениями по изотермам:

$$\Delta A = T_1 \Delta S_{12} + T_2 \Delta S_{34} = (T_1 - T_2) \Delta S_{12},$$

так как полное изменение энтропии при движении по обратимому замкнутому циклу  $\Delta S = \Delta S_{12} + \Delta S_{34} = 0$ . Переданное в систему количество теплоты на изотермическом участке 1—2 равно  $\Delta Q = T_1 \Delta S_{12}$ . Следовательно,

$$\eta = \frac{\Delta A}{\Delta Q} = \frac{T_1 - T_2}{T_1} < 1.$$

Чтобы увеличить коэффициент полезного действия тепловой машины, надо либо увеличивать  $T_1$ , либо уменьшать  $T_2$ . Однако получить  $T_2 = 0$  К практически невозможно. Таким образом, КПД любой тепловой машины меньше единицы. В обычных условиях сделать температуру холодильника ниже температуры окружающего воздуха—трудоемкая задача. Поэтому для тепловых машин температура холодильника в лучшем случае равна температуре окружающей среды. Следовательно, температура  $T_2$  довольно велика. Например, 20°C соответствуют приблизительно 300 К. Температура  $T_1$ —это температура нагревателя. При более высокой температуре нагревателя получается более высокий КПД. Если бы температура  $T_1$  была бесконечно большой, то КПД снова бы стремился к единице. Реально температура  $T_1$  не может быть очень большой. В лучших паровых машинах температура водяного пара достигает сотен градусов. Например, при  $T_1 = 800$  К  $T_2/T_1 = 3/8$ , т. е.  $\eta$  равно лишь 5/8. В двигателе внутреннего сгорания температура гораздо выше, чем в паровых машинах. Еще больший КПД можно получить в газовых турбинах. Коэффициент полезного действия зависит от формы рабочего цикла, причем любой другой цикл имеет  $\eta$  меньше, чем цикл Карно.

Так как изотермический процесс—процесс медленный, то все реальные машины работают по циклам, где нет изотермических участков. А это, в свою очередь, также приводит к понижению КПД.

Вся конструкторская, инженерная мысль направлена на поиск циклов, которые дают более высокий КПД.

Производимую машиной работу можно определить, как и количество теплоты, переданной от нагревателя в цилиндр. Значит, можно вычислить  $\eta$ , а это позволяет построить термодинамическую шкалу температур.

Предположим, что надо сравнить температуры двух тел. Примем одно из них в качестве нагревателя, а другое — холодильника. Присоединим к ним цилиндр с газом и осуществим цикл Карно, вычисляя при этом  $\eta$ . Тогда  $T_1/T_2 = (1 - \eta)$ .

Если  $\eta = 0$ , то  $T_2 = T_1$ . Если  $\eta = 1$ , то  $T_2 = 0$  К (разумеется, такой опыт можно провести лишь теоретически). Так как совершенно неважно, что является рабочим телом, включенным между холодильником и нагревателем, то измерение температур не должно зависеть от свойств конкретных тел и является абсолютным.

Практически в подавляющем числе случаев измерение температур производится на основании расширения тел при нагревании. Термодинамический способ измерения температуры используется лишь при очень низких температурах, близких к 0 К.

Тепловая машина, действующая по циклу Карно, как указывалось, — самая выгодная машина. КПД двигателя Карно является максимально возможным. При этом не учитываются всевозможные потери, связанные с трением, с несовершенством механизма и т. д.

Если мы сделаем другую машину, тоже совершенную в том смысле, что не будет потерь на трение и т.п., все равно другой цикл будет иметь КПД ниже, чем цикл Карно. Это приводит к принципиальным ограничениям, связанным с повышением КПД тепловых машин любого рода. Зная температуру холодильника и температуру нагревателя, можно указать максимально возможный в данном случае КПД тепловой машины.

Реальные тепловые двигатели имеют довольно низкий КПД. Например, для паровой машины хорошим КПД считается 15–20%.

КПД двигателя внутреннего сгорания в зависимости от конструкции варьируется и достигает более высокого процента — до 30%. Еще более высокий КПД имеют газовые турбины за счет того, что повышается температура нагревателя, которая может достичь 1500–2000°C. Но все равно КПД реальной тепловой машины не бывает, как правило, выше 50%.

Всегда при любых условиях получить КПД = 1 мешает второе начало термодинамики, которое запрещает переходить теплоте самопроизвольно от холодного тела к горячему.

Если бы можно было осуществить передачу теплоты от холодного тела к горячему самопроизвольно, то можно было бы построить вечный двигатель. Поэтому говорят, что второе начало



термодинамики запрещает построить вечный двигатель второго рода.

Теперь предположим, что вся система (холодильник, цилиндр, нагреватель) помещена в замкнутый сосуд. В этом случае, после того как машина совершила один цикл, часть теплоты будет передана холодильнику. В результате температура нагревателя понизится, температура холодильника повысится. После второго цикла еще больше понизится температура нагревателя и повысится температура холодильника. Это будет продолжаться до тех пор, пока температура  $T_1$  не станет равной  $T_2$  и двигатель перестанет работать.

Чтобы работа двигателя была длительной, нужно все время за счет внешнего источника поддерживать постоянную температуру нагревателя. В обычной паровой машине это можно сделать за счет энергии сжигаемого топлива.

Нельзя также оставить совершенно замкнутым и холодильник. Предположим, что осуществляется только один процесс: все время поддерживается температура нагревателя постоянной за счет ввода новой энергии. Однако по-прежнему при каждом цикле должна повышаться температура холодильника. Если холодильник все время нагревается, то температура  $T_2$  станет равной температуре  $T_1$  и машина снова остановится. Чтобы обеспечить длительную работу машины, нужно все время отводить теплоту от холодильника, отводить в такую систему, которая бы практически от этого не нагревалась. Такой системой является внешнее окружающее пространство.

Таким образом, рассматриваемая тепловая система должна быть открытой системой, т. е. к этой системе должна извне приходить энергия, которая расходуется на поддержание постоянной температуры нагревателя, и, кроме того, из этой системы все время должна отводиться теплота. Только тогда можно осуществить работу тепловой машины бесконечно долго. Но это, конечно, не вечный двигатель.

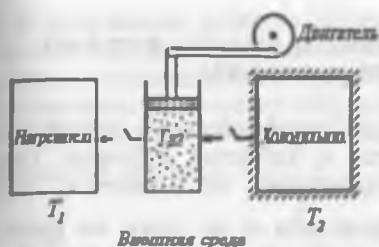
Посмотрим теперь, как работают промышленные или домашние холодильники, т. е. устройства, позволяющие получить низкие температуры в замкнутой камере.

*Бытовые холодильники делятся на два класса: тепловые и компрессорные.*

Рассмотрим, например, холодильник ЗИЛ, относящийся к компрессорному типу. В устройстве имеется двигатель, который сжимает или расширяет газ в цилиндре по заданному циклу.

Будем считать, что холодильник заключен в теплоизолирующую оболочку (рис. 6.12). Рассмотрим, как можно осуществить охлаждение камеры, которая теплоизолирована от внешней среды и играет роль холодильника.

Рис. 6.12. Схемы работы компрессорного холодильника



Предположим, что в начальный момент система имела состояние при температуре камеры  $T_2$ . Расширяя с помощью двигателя адиабатически газ, можно понизить температуру в цилиндре, сделав ее ниже  $T_2$ .

После этого цилиндр с газом связывается с камерой холодильника и совершается изотермическое расширение газа при температуре ниже температуры камеры холодильника. Теплота при этом начнет переходить в цилиндр, а сама холодильная камера, изолированная от внешней среды, охлаждаться. Прекратим изотермическое расширение газа в тот момент, когда температуры цилиндра и холодильной камеры станут почти равными. Отсоединив цилиндр от холодильной камеры, сожмем газ адиабатически. При этом газ нагреется. Будем продолжать сжатие газа до тех пор, пока его температура не станет больше температуры внешней среды. В этот момент присоединим цилиндр к внешней среде и продолжим сжатие, но уже изотермически. Так как температура цилиндра выше температуры внешней среды, то теплота будет переходить из него во внешнюю среду. Окончив этот процесс, снова адиабатически расширим газ, снизив его температуру ниже температуры холодильника. Затем повторим весь процесс снова. В результате теплота все время будет отниматься у холодильной камеры и выводиться во внешнюю среду. Таким образом, охлаждение камеры холодильника получается за счет работы двигателя, перемещающего поршень цилиндра с газом. Так как на каждом этапе теплота по-прежнему переходит от горячего тела к холодному, то никакого противоречия со вторым началом термодинамики не возникает.

Для того чтобы холодильник работал, необходимо выполнение двух условий.

1. Необходимо заключить холодильную камеру в теплоизолирующую оболочку. В этом случае запас тепловой (внутренней) энергии, содержащейся в холодильной камере, будет конечным и, отбирая энергию, можно охладить камеру.

2. Температура  $T_1$  должна поддерживаться одинаковой, и теплота из камеры должна извлекаться и передаваться в атмосферу.

## 6.6. Теорема Карно и обобщение понятия энтропии как термодинамического потенциала

Выше мы показали, что КПД тепловой машины, использующей цикл Карно и идеальный газ в качестве рабочего тела, равен  $\eta = 1 - T_2/T_1$ . Этот коэффициент не меняется, если машина будет работать в тех же условиях и по тому же циклу, но с другим рабочим телом.

Будем рассуждать следующим образом. Если машина работает по циклу Карно, то так как на адиабатических участках произведенная и затраченная работы всегда взаимно компенсируют друг друга (если только внутренняя энергия рабочего тела есть линейная функция температуры), а на изотермических участках, согласно первому началу термодинамики, равны количеству теплоты, переданной рабочему телу или отнятой от него, то можно записать, что

$$\eta = \frac{\Delta Q_1 - \Delta Q_2}{\Delta Q_1} \quad \text{или} \quad \Delta Q_2 = (1 - \eta)\Delta Q_1,$$

где  $\Delta Q_1$ —количество теплоты, переданной нагревателем рабочему телу;  $\Delta Q_2$ —количество теплоты, переданной рабочим телом холодильнику во время обратного хода поршня при изотермическом сжатии.

Представим, что имеются две машины Карно с разными рабочими телами (обозначим их индексами штрих и два штриха). Пусть машина с идеальным газом маркируется одним штрихом. Предположим, что обе эти машины могут последовательно присоединяться к одному и тому же нагревателю и холодильнику.

Пусть первая машина совершит  $n'$  рабочих циклов и в результате произведет работу  $A'$ , отняв у нагревателя количество теплоты  $\Delta Q'_1$  и передав холодильнику количество теплоты  $\Delta Q'_2$ .

Используем эту работу для того, чтобы заставить работать вторую машину в качестве холодильной, т. е. двигаться по циклу Карно против часовой стрелки. Тогда вторая машина отнимает у холодильника количество теплоты  $\Delta Q'_2$  и передаст нагревателю количество теплоты  $\Delta Q'_1$ . Выбирая число циклов достаточно большим, всегда можно добиться того, чтобы за счет работы  $A'$  вторая машина совершила сколь угодно близкое к целому количество полных циклов. Если никаких потерь на трение нигде нет, то положительная работа первой машины должна быть равна работе над второй машиной, т. е. полная работа двух машин должна быть равна нулю ( $A' = A''$ ), а так как обе машины обратимы, то  $\eta'\Delta Q'_1 = \eta''\Delta Q'_1$ .

В результате полного кругового процесса, при котором обе машины произвели одинаковую работу (но разного знака), вся система, состоящая из двух устройств с рабочими телами (холодильник и нагреватель), вернется в исходное состояние. При этом нагреватель потеряет количество теплоты  $\Delta Q_1'$  и получит количество теплоты  $\Delta Q_1''$ , т. е. общее его приращение равно

$$\Delta Q_1'' - \Delta Q_1' = \frac{\eta' - \eta''}{\eta''} \Delta Q_1'.$$

Это соотношение легко получить, вычитая из правой и левой частей равенства  $\eta' \Delta Q_1' = \eta'' \Delta Q_1''$  одну и ту же величину  $\eta'' \Delta Q_1'$  и деля затем результат на  $\eta''$ . Предполагая  $\eta' > \eta''$ , находим

$$\Delta Q_1'' - \Delta Q_1' > 0.$$

Холодильник в результате всего процесса отдаст количество теплоты

$$\Delta Q_2'' - \Delta Q_2' = (1 - \eta'') \Delta Q_1'' - (1 - \eta') \Delta Q_1' = \Delta Q_1'' - \Delta Q_1' > 0.$$

Таким образом, в результате всего процесса количество теплоты, равное  $\Delta Q_1'' - \Delta Q_1'$ , будет передано нагревателю, а так как полная работа по условию равна нулю, то получится, что количество теплоты от более холодного тела самопроизвольно, т. е. без совершения работы, перешло к более нагретому! Однако именно это и запрещается вторым началом термодинамики. В нашем рассуждении единственным уязвимым местом было то, что мы предположили  $\eta' > \eta''$ . Значит, это предположение неверно.

Если вторая машина будет совершать положительную работу, а первая работать в качестве холодильной, то условие  $\eta'' > \eta'$  опять будет неверно. Остается одно—считать, что  $\eta'' = \eta'$ .

Мы получили результат принципиальной важности (это и есть *теорема Карно*), согласно которому вне зависимости от того, каково рабочее устройство, КПД соответствующей тепловой машины будет равен  $\eta = 1 - T_2/T_1 = 1 - \Delta Q_2/\Delta Q_1$ , если только процесс совершается по циклу Карно. Ранее мы показали также, что это цикл с максимально большим КПД при данном перепаде температур нагревателя и холодильника. Из равенства, которое, таким образом, оказывается верным для любого вещества, используемого в качестве рабочего тела в тепловой машине ( $T_2/T_1 = \Delta Q_2/\Delta Q_1$ ), следует  $\Delta Q_1/T_1 = \Delta Q_2/T_2$ . Это приводит к условию, что сумма отношений  $\Delta Q/T$ , взятых для всех бесконечно малых участков обратимого цикла равна нулю.

В свою очередь, это означает, что величина  $dS = \Delta Q/T$  для любого малого участка обратимого процесса вне зависимости от вещества есть полный дифференциал, а изменение этой величи-

ны, названной энтропией, зависит только от начального и конечного состояний системы.

Получается, что энтропия—это некоторая универсальная функция, характеризующая фундаментальные процессы, происходящие в сложных системах, состоящих из большого числа объектов. К этим вопросам мы еще вернемся в гл. 10.

## 6.7. Энтропия в необратимых процессах

Мы ввели понятие энтропии как некоторого термодинамического потенциала для обратимых процессов. Если переход из одного равновесного состояния системы в другое необратим, то на участках внутри перехода рассматривать изменение энтропии как полный дифференциал уже нельзя. Можно, однако, по-прежнему найти полное изменение энтропии как разность ее значений в начале и конце.

Рассмотрим следующий простой пример. Имеется замкнутый сосуд, разделенный пополам перегородкой, не пропускающей газ, но хорошо проводящей тепло (это может быть тонкая медная фольга). Пусть в каждой половине сосуда находятся одинаковые количества одного и того же газа (идеального), но в начальный момент при разных температурах  $T_1$  и  $T_2$ , причём  $T_1 > T_2$ . Так как обе порции газа находятся в тепловом контакте, то через некоторое время температуры газа в обеих половинах сосуда выравниваются и станут равными  $T_3$ , причём  $T_1 > T_3 > T_2$ , т. е. горячий газ охладится, а холодный нагреется. Этот результат хорошо известен.

Посмотрим, как в этом случае изменится энтропия всей системы. Если число молекул в каждом из смежных объемов одно и то же ( $N$ ), а число степеней свободы каждой молекулы равно 3 (простейший случай), то, так как на каждую молекулу в среднем приходится энергия  $\frac{3}{2} kT$  ( $k$ —постоянная Больцмана),

общий запас энергии в начальный момент  $E = \frac{3}{2} Nk(T_1 + T_2)$ ,

а после установления теплового равновесия соответственно  $E = 3NkT_3$ .

Так как система замкнута, то по закону сохранения энергии

$$\frac{3}{2} Nk(T_1 + T_2) = 3NkT_3,$$

откуда

$$T_3 = (T_1 + T_2)/2.$$

Воспользуемся теперь формулой для изменения энтропии при изменении объема и температуры газа. Так как в данном случае объемы не меняются в процессе эксперимента, а меняется только

температура, то изменение энтропии в сумме для двух частей газа в разных половинах сосуда равно

$$\Delta S = c_V \ln \frac{T_3}{T_1} + c_V \ln \frac{T_3}{T_2} = c_V \ln \left( \frac{T_1 + T_2}{2T_1} \right) + c_V \ln \left( \frac{T_1 + T_2}{2T_2} \right) = \\ = c_V \ln \left( \frac{1}{2} + \frac{1}{2} \frac{T_2}{T_1} \right) + c_V \ln \left( \frac{1}{2} + \frac{1}{2} \frac{T_1}{T_2} \right).$$

Так как  $T_1 > T_2$ , то  $\frac{T_2}{T_1} < 1$ ,  $\frac{T_1}{T_2} > 1$  и, следовательно,

$$\left( \frac{1}{2} + \frac{1}{2} \frac{T_2}{T_1} \right) < 1, \quad \left( \frac{1}{2} + \frac{1}{2} \frac{T_1}{T_2} \right) > 1.$$

Значит,

$$\ln \left( \frac{1}{2} + \frac{1}{2} \frac{T_2}{T_1} \right) < 0, \quad \ln \left( \frac{1}{2} + \frac{1}{2} \frac{T_1}{T_2} \right) > 0.$$

Далее,  $\frac{T_2}{T_1} < \frac{T_1}{T_2}$  и  $\left| \ln \left( \frac{1}{2} + \frac{1}{2} \frac{T_2}{T_1} \right) \right| < \left| \ln \left( \frac{1}{2} + \frac{1}{2} \frac{T_1}{T_2} \right) \right|$ .

В результате получаем, что в целом  $\Delta S = \Delta S_1 + \Delta S_2 > 0$ . Таким образом, изменение энтропии в рассматриваемом необратимом процессе положительно.

Рассмотрим второй необратимый процесс: диффузию газа, первоначально находившегося в одной половине сосуда, в весь объем после снятия перегородки, разделяющей обе половины сосуда. Здесь можно воспользоваться соотношением  $\Delta S = R \ln (V_2/V_1) = k \ln (\alpha_1/\alpha_2)^N$ . Когда газ заполнит весь сосуд,  $\Delta S$  может только возрасти. Действительно, вероятность найти частицу в единичном объеме в первом случае ( $\alpha_1$ ), когда газ занимает меньший объем, выше, чем вероятность  $\alpha_2$  найти частицу во втором случае, когда частица летает в большем общем пространстве. Значит,  $\alpha_1/\alpha_2 > 1$  и  $\Delta S > 0$ .

В рассмотренных ситуациях процессы необратимы и  $\Delta S > 0$ . Данные результаты верны для всех необратимых процессов в замкнутых системах и обобщаются в виде весьма важного утверждения: *во всех необратимых процессах в замкнутых системах энтропия всегда возрастает!*

При этом возрастание энтропии сопровождается и выравниванием (в нашем случае) температур или плотностей газа двух соединяемых объемов. Если под *порядком* понимать сосредоточение частиц или энергии в определенном месте пространства, а под *беспорядком* — равномерное распределение их во всем объеме, то возрастание энтропии при совершающихся без внешних воздействий необратимых процессах отражает глубокое

природное стремление систем самопроизвольно, если это возможно, переходить от состояния более упорядоченного к состоянию менее упорядоченному.

Закон возрастания энтропии или закон о направлении развития любых спонтанных (самопроизвольных) процессов потому и имеет исключительно важное значение, что он указывает на существование в природе преимущественного развития сложных статистических систем.

В свою очередь, именно существование таких развивающихся лишь в определенных направлениях процессов и делает необходимым введение такого фундаментального понятия, как время, которое всегда меняется лишь от прошлого к будущему.

Подробному обсуждению этого принципиального вопроса и других смежных специально посвящена глава 10.

Сейчас же мы ограничимся лишь рассмотрением ряда дополнительных, но очень важных сведений о поведении энтропии в неравновесных процессах.

Вернемся к выражению для изменения энтропии идеального газа

$$\Delta S = c_v \ln(T_2/T_1) + R \ln(V_2/V_1).$$

Перепишем его в другом виде, учитывая, что  $n_{01} = N/V_1$  и  $n_{02} = N/V_2$  (число частиц в единице объема). Тогда

$$\ln(V_2/V_1) = -\ln(n_{02}/n_{01}).$$

В результате получим:

$$\Delta S = c_v \ln(T_2/T_1) - R \ln(n_{02}/n_{01}).$$

Отсюда

$$dS = c_v \frac{dT}{T} - R \frac{dn_0}{n_0} = \left( \frac{\partial S}{\partial T} \right) dT + \left( \frac{\partial S}{\partial n_0} \right) dn_0,$$

где  $S = \ln(T^{c_v} n_0^{-R})$ .

По своей математической форме это выражение совпадает с формой для элемента работы потенциальных сил:

$$dA = -dU = \left( -\frac{\partial U}{\partial x} \right) dx + \left( -\frac{\partial U}{\partial y} \right) dy$$

(ограничимся двумя переменными). Если воспользоваться обобщенными силами и функцией Лагранжа, то (см. § 2.9)

$$dA = dL = \left( \frac{\partial L}{\partial q_1} \right) dq_1 + \left( \frac{\partial L}{\partial q_2} \right) dq_2.$$

Пользуясь этой аналогией, можно производным  $\left( \frac{\partial S}{\partial T} \right)$  и  $\left( \frac{\partial S}{\partial n_0} \right)$  приписать также смысл некоторых обобщенных сил. Обозначим их соответственно как  $X_{ST}$  и  $X_{Sn_0}$ .

Найдем теперь производные от изменения энтропии по времени. Эта величина называется *производством энтропии*. Как было показано выше,  $dS/dt > 0$ . В случае равновесия  $dS/dt = 0$ . Разделим равенство  $dS = \left(\frac{\partial S}{\partial T}\right) dT + \left(\frac{\partial S}{\partial n_0}\right) dn_0$  на  $dt$ .

Получим:

$$\frac{dS}{dt} = \left(\frac{\partial S}{\partial T}\right) \frac{dT}{dt} + \left(\frac{\partial S}{\partial n_0}\right) \frac{dn_0}{dt} = X_{ST} \frac{dT}{dt} + X_{Sn_0} \frac{dn_0}{dt}.$$

Если процессы совершаются в замкнутой системе без контакта с внешней средой, то производные  $\frac{dT}{dt}$  и  $\frac{dn_0}{dt}$  могут быть отличными от нуля только тогда, когда система не находится в равновесии. Если эту систему разделить на достаточно большие участки, то для каждого из них можно говорить о своей температуре и о своей плотности частиц. Следовательно, в неравновесной ситуации внутри области, занятой газом, появляются, например, перепады температуры и числа частиц в единице объема.

Процесс перехода системы от неравновесного к равновесному состоянию, что и сопровождается возрастанием энтропии, заключается в постепенном выравнивании температуры и плотности газа во всех частях системы. Совершенно очевидно, что это обязательно будет связано с переносом энергии и вещества из одних областей полного объема газа в другие. В свою очередь, это означает существование соответствующих внутренних температурных (энергетических) и диффузионных потоков.

Закономерности возникновения таких потоков и средства их описания мы изучали выше. По самому определению поток любой характеристики есть величина этой характеристики, переносимая через некоторую поверхность в единицу времени.

Представим теперь, что имеется некоторая область с избыточным, по сравнению с равновесным, числом частиц в ней (или избыточной энергией). Тогда «лишние» частицы будут уходить во внешнее пространство. Окружим эту область сферой. Тогда можно говорить о потоке частиц через поверхность сферы. В силу закона сохранения полного потока числа частиц (или энергии) можно сразу заключить, что полный поток ушедших наружу частиц должен быть равен убыли их внутри сферы за единицу времени. Мощность потока, следовательно, можно характеризовать величиной  $dA/dt$ , где  $A$  есть мера данной характеристики (число частиц, запас энергии и т. д.). Опираясь на введенное выше (см. § 5.4) обобщенное описание потоков через кинетические коэффициенты и обобщенные силы, запишем для потока произвольной величины  $A$  соотношение

$$\frac{dA}{dt} = L_A X = L_A \frac{\partial A}{\partial x}.$$



Здесь  $X$ —обобщенная сила. Если воспользоваться этим соотношением, то нетрудно получить:

$$\frac{dS}{dt} = X_{ST} L_T \frac{\partial T}{\partial x} + X_{S n_0} L_{n_0} \frac{\partial n_0}{\partial x}.$$

Мы ограничились здесь лишь переменной  $x$ . Это возможно, если предположить, что все неоднородности в газе распределены лишь в направлении соответствующей оси, а вдоль осей  $Y$  и  $Z$  неоднородностей нет. Анализ полученного позволяет сделать следующие выводы.

Во-первых, величина производной  $dS/dt$ , или скорость производства энтропии, в неравновесной системе будет расти с ростом производных  $\partial T/\partial x$  и  $\partial n_0/\partial x$ , т. е. первоначальной степени неоднородности системы.

Во-вторых, так как  $X_{ST} = \frac{\partial S}{\partial T} = c_V/T$  и  $X_{S n_0} = \frac{\partial S}{\partial n_0} = -R/n_0$ ,

то энтропия при прочих равных условиях будет медленнее нарастать в более нагретых и менее плотных средах (обратите внимание на знак минус перед величиной  $R/n_0$ !). Другими словами, рост энтропии будет меньшим в системах, уже обладающих высокими значениями энтропии. Напомним, что  $S = \ln(T^{c_V} V^{R_0})$  и увеличивается с ростом температуры и объема (значит, падения  $n_0$ ) газа.

Далее, если газ находится в равновесном состоянии, то  $T$  и  $n_0$  постоянны и  $\partial T/\partial x$  и  $\partial n_0/\partial x$  равны нулю. Значит, в равновесном состоянии  $\frac{dS}{dt} = 0$ .

Очень важное утверждение состоит в том, что при переходе из одного равновесного состояния в другое через ряд квазиравновесных положений производство энтропии оказывается минимальным (теорема Пригожина).

Для доказательства этой теоремы учтем, во-первых, что если энтропия вообще производится, то  $dS \neq 0$  и, значит,  $X_{ST}$ ,  $X_{S n_0} \neq 0$ , т. е. производство энтропии всегда есть функция этих переменных, и, во-вторых, что минимум или максимум функции  $z = f(x, y)$  достигается в точке, где  $\partial z/\partial x$  и  $\partial z/\partial y$  равны 0. При изучении скорости производства энтропии можно говорить лишь о ее минимуме. Вычислим производную  $\frac{\partial}{\partial X_{ST}} \left( \frac{dS}{dt} \right)$  и приравняем ее к нулю. Тогда  $\frac{\partial}{\partial X_{ST}} \left( \frac{dS}{dt} \right) = L_T \frac{\partial T}{\partial x} = 0$ . Отсюда следует, что условие минимума производства энтропии и равенство нулю градиента  $\partial T/\partial x$  ( $L_T = \text{const}$ ) эквивалентны. Но именно отсут-

ствис градиента температуры и является условием протекания квазиравновесных процессов.

Аналогично можно рассмотреть и производную  $\frac{\partial}{\partial X_{S n_0}} \left( \frac{\partial S}{\partial t} \right)$ .

Тем самым теорема Пригожина доказана.

Если число участков в газе с достаточно заметными отклонениями значений  $T$  и  $n_0$  от среднего будет невелико, то ясно, что перенос энергии и частиц из этих участков в окружающее пространство в нем мало что изменит. Тогда можно считать, что силы  $X_{ST}$  и  $X_{S n_0}$  постоянны. На начальных стадиях процесса

можно считать постоянными и компоненты градиентов  $\partial T / \partial x$  и  $\partial n_0 / \partial x$ . Тогда  $\frac{\partial^2 S}{\partial t^2} = 0$ , и возрастание энтропии будет линейным во времени.

Наличие в среде макроскопических областей с заметным отличием локальных температур и плотностей газа от средних можно назвать частично упорядоченной структурой.

Рост энтропии связан, таким образом, как с нарастанием температуры всей среды или с увеличением ее объема (уменьшением средней плотности вещества), так и с разрушением (диссипацией) упорядоченной структуры. При этом роль этого второго фактора может оказаться значительно большей, чем первого.

Затронутые выше и другие вопросы составляют предмет исследования так называемой *термодинамики неравновесных процессов* или *неравновесной термодинамики*. Если начальное состояние не слишком сильно отличается от равновесного, то коэффициенты  $L_k$ , как уже упоминалось в § 5.4, можно считать постоянными. Тогда говорят о *линейной неравновесной термодинамике*.

В более сложном случае коэффициенты  $L_k$  могут уже не быть постоянными и возникнут так называемые нелинейные процессы. Они составляют предмет *нелинейной неравновесной термодинамики*. Мы не будем в этой книге касаться этой сложной, во многом еще неразработанной области науки.

#### Вопросы для самоконтроля

1. На каком принципе основан процесс охлаждения при получении жидких газов с помощью специальных компрессоров?
2. Какая физическая причина приводит к тому, что теплоемкость тел зависит от условий, при которых совершается данное измерение? Может ли теплоемкость быть бесконечно большой и при каком процессе такая ситуация может наблюдаться?
3. Можно ли, не производя никаких выкладок, показать, что адиабата в координатах давления и объема идет круче, чем изотерма?
4. Какие признаки энтропии как функции состояния идеального газа определяют ее сходство с электростатическим потенциалом и потенциалом

гравитационного поля, что и позволяет рассматривать энтропию как некоторый термодинамический потенциал?

5. Как связана энтропия с параметрами состояния идеального газа?

6. Чему равно изменение энтропии при изотермическом процессе?

7. Записать выражение для работы газа при адиабатическом, изотермическом, изохорическом, изобарическом процессах.

8. Почему именно цикл Карно играет такую большую роль при исследовании тепловых машин любого вида, хотя известно, что реально по циклу Карно не работает ни один из двигателей?

9. Какая физическая причина приводит к тому, что в качестве составной части любой тепловой машины должен быть холодильник?

10. Какая основная причина заставляет конструкторов при создании реально действующих двигателей внутреннего сгорания, турбин или паровых двигателей отказываться от цикла Карно, хотя и известно, что в этом случае коэффициент полезного действия был бы наибольшим? Что выигрывается и что проигрывается при таком отказе?

11. Объясните, как можно произвести сравнение температур двух тел с помощью тепловой машины и почему шкала Кельвина получила название абсолютной или термодинамической?

12. Какие условия должны выполняться, чтобы тепловая машина могла работать очень долгое время? Можно ли осуществить длительную работу тепловой машины, если холодильник будет замкнутым и неспособным сохранять свою низкую температуру?

13. Опишите, как работает обычный холодильник, построенный по компрессионному принципу. Можно ли нагревать внутренние помещения зданий за счет тепловой энергии, накопленной во внешней атмосфере, если температура окружающего воздуха значительно ниже температуры воздуха внутри жилых помещений?

14. Зависит ли коэффициент полезного действия тепловой машины, работающей по циклу Карно, от рабочего тела, т. е., например, от состава газа, находящегося под поршнем цилиндра?

15. Как меняется энтропия в том случае, если в двух объемах, которые могут обмениваться друг с другом теплотой, происходит выравнивание температуры? Как меняется энтропия, если происходит диффузия газа из одной половины сосуда в весь сосуд, если первоначально вторая половина сосуда не содержала газа?

16. Как связано изменение энтропии в необратимых процессах с направленным изменением времени (от прошлого к будущему)?

17. В чем заключается смысл теоремы Пригожина?

18. Какие процессы изучает неравновесная термодинамика?



## ГЛАВА 7

### ЭЛЕКТРОМАГНИТНЫЕ ЯВЛЕНИЯ

*«Физическая теория подобна костюму, сшитому для природы. Хорошая теория подобна хорошо сшитому костюму, а плохая — тришкину кафтану».*

Я. Френкель

#### 7.1. Заряды. Электрические и магнитные поля. Характеристики поля

Электричество и магнетизм — один из наиболее общих и важных разделов учения о строении мира. Характерным свойством электрических и магнитных сил является то, что их действие обнаруживается на расстоянии. Для объяснения такого действия необходимо вводить понятие поля. Поэтому учение об электромагнитных взаимодействиях есть прежде всего учение о создаваемых заряженными телами и электрическими токами полях и их свойствах.

В разделе «Механика» мы ознакомились с фундаментальным свойством частиц и тел — массой. Ею обладают все частицы и тела (напомним, однако, что элементарные частицы могут иметь нулевую массу).

Введем в рассмотрение еще одно важное свойство частиц — *электрический заряд*. Это свойство было обнаружено еще в глубокой древности с помощью опытов по электризации сил трением. Опыты эти хорошо известны из школьного курса физики. При трении друг о друга, например, кожи и стеклянной палочки происходит разделение зарядов, элементарными носителями которых являются электроны (отрицательные заряды) и ядра атомов (положительные заряды). В природе существует только эти два типа зарядов. При трении часть электронов с одного тела может перейти на другое. В результате второе тело получит отрицательный заряд, а тело, отдавшее электроны, приобретет точно такой же по величине, но положительный заряд.

Все окружающие нас тела состоят из атомов и молекул. В свою очередь, каждый атом состоит из электронов и ядер.

Каждое ядро содержит определенное число протонов. Заряды электрона и протона равны по величине и противоположны по знаку. Эти заряды являются наименьшими для частиц, которые могут существовать индивидуально. Поэтому такие заряды принимаются в качестве единичных. Есть, однако, объекты, получившие название кварков, которые имеют дробные заряды, но такие объекты не могут существовать индивидуально, а только в комбинациях, образуя более крупные частицы. Заметим, что иметь заряд, равный заряду электрона (или протона) или кратный ему, могут только такие материальные образования, которые обладают ненулевой массой.

С помощью простейших опытов по электризации тел трением можно установить, что тела, заряженные одноименными зарядами, отталкиваются друг от друга, а заряженные разноименно—притягиваются.

Фундаментальным законом природы является закон сохранения заряда, согласно которому заряды могут возникать и исчезать только парами (положительные и отрицательные заряды одной величины), так что сумма положительных и отрицательных зарядов в природе всегда остается постоянной. Поэтому Вселенная в целом электрически нейтральна.

В простейших опытах с помощью электрофорной машины можно создать на одном кондукторе избыточный положительный заряд, а на другом—отрицательный, но в целом машина будет электрически нейтральна.

В электрофорной машине заряды не возникают и не исчезают, а лишь разделяются в пространстве. Имеются, однако, эксперименты, когда такое возникновение или исчезновение зарядов действительно есть. Например, мощный сгусток электромагнитной энергии (гамма-квант) при определенных условиях может превратиться в пару частиц—электрон и позитрон (позитрон—частица, во всем сходная с электроном, но имеющая положительный заряд), каждая из которых уже несет заряд.

Наоборот, при столкновении электрона и позитрона происходит их аннигиляция и в результате появляется незаряженное образование—гамма-квант.

Электрические силы являются одними из наиболее важных в природе. Эти силы определяют существование атомов. В атомах отрицательные электроны притягиваются к положительно заряженному ядру, что и обеспечивает стабильное существование атома. В свою очередь, атомы могут объединяться между собой, образуя более крупные системы—молекулы. При этом никаких новых сил, кроме электрических, не возникает, только движение электронов становится более сложным. Из молекул состоят еще более сложные ассоциаты, например клетки и т. д.

Зная взаимодействия заряженных частиц, в принципе можно, хотя это и сложно, объяснить все разнообразие явлений из атомно-молекулярном уровне, встречающееся в окружающем нас мире.

Наша задача заключается в том, чтобы изучить наиболее фундаментальные свойства электромагнитных сил. Остановимся пока на характеристике сил, действующих на заряды. При этом вначале будем считать, что заряды локализованы на материальных точках, находящихся в вакууме.

Рассмотрим некоторое пространство, в котором находится несколько неподвижных или движущихся зарядов. Выделим из всех этих зарядов один, величину которого обозначим  $q$ . На этот заряд действует сила со стороны всех остальных зарядов. Эта сила зависит от величин остальных зарядов, от их расположения по отношению к выделенному и от того, находятся ли выделенный заряд и остальные в движении или нет. Существенно, что если в пространстве имеется заряд, то другой заряд будет испытывать его действие во всех точках окружающего первый заряд пространства. Следовательно, существует силовое поле, создаваемое первым зарядом.

Многочисленными экспериментами установлено, что выражение для обсуждаемой силы в общем случае можно записать в виде

$$\mathbf{F} = q\mathbf{E} + \frac{q}{c} [\mathbf{v} \times \mathbf{B}'] = q\mathbf{E} + q[\mathbf{v} \times \mathbf{B}] \quad \left( \mathbf{B} = \frac{\mathbf{B}'}{c} \right).$$

Это выражение называется *формулой Лоренца*, а сила  $\mathbf{F}$  — *силой Лоренца*. Здесь  $\mathbf{v}$  — скорость рассматриваемого заряда в данной точке пространства;  $c$  — скорость света. В этой формуле две части. Первая часть ( $q\mathbf{E}$ ) не зависит от  $\mathbf{v}$ , т. е. не зависит от скорости движения заряда, и определяет компоненту силы, которая будет действовать как на неподвижный, так и на движущийся заряд. Вектор  $\mathbf{E}$  называется *напряженностью электрического поля*. Ее можно определить как силу, действующую на неподвижный единичный заряд ( $q = 1$ ), помещенный в данную точку пространства, со стороны других зарядов, находящихся в том же пространстве.

Вторая часть формулы ( $q[\mathbf{v} \times \mathbf{B}]$ ) определяет компоненту силы, которая возникает только тогда, когда при прохождении данной точки пространства выделенный пробный заряд имеет скорость  $\mathbf{v} \neq 0$ . Эта сила исчезает, если заряд неподвижен. Вектор  $\mathbf{B}$  называется *индукцией магнитного поля*. Магнитное поле, в свою очередь, может быть создано только при наличии в данном пространстве движущихся зарядов. Индукцию  $\mathbf{B}$  магнитного поля уже нельзя, подобно напряженности  $\mathbf{E}$ , определить через силу, действующую в данной точке пространства на единичный пробный заряд, движущийся со скоростью  $\mathbf{v}$ . В самом деле, соответствующая сила, равная ( $q[\mathbf{v} \times \mathbf{B}]$ ), зависит не только от модулей векторов  $\mathbf{v}$  и  $\mathbf{B}$ , но и от их взаиморасположения. Если  $\mathbf{v} \parallel \mathbf{B}$ , то  $[\mathbf{v} \times \mathbf{B}] = 0$ , даже если  $\mathbf{B} \neq 0$ .

Связь между электрической и магнитной компонентами, создаваемыми зарядом, движущимся в пространстве равномерно и прямолинейно со скоростью  $\mathbf{v}$ , выражается формулой

$$\mathbf{B} = \frac{1}{c^2} [\mathbf{v} \times \mathbf{E}].$$

При этом необходимо заметить, что движущийся заряд создает электрическое поле, отличающееся от поля, создаваемого этим же, но неподвижным зарядом.

Появление второго члена в формуле для силы, действующей на заряд в пространстве со стороны других зарядов, связано с тем важнейшим обстоятельством, что любое влияние одного объекта на другой передается не мгновенно, а с конечной скоростью. С конечной, хотя и очень большой, скоростью (скоростью света) распространяется и создаваемое каждым зарядом электрическое поле. Если, например, в некоторой точке пространства возникает пара противоположно заряженных частиц, пара зарядов, то находящиеся на некотором расстоянии от этой пары другие заряды «почувствуют» присутствие новых зарядов не в момент появления пары, а через конечный, хотя и малый, промежуток времени, зависящий от скорости распространения действия этой пары зарядов, т. е. скорости распространения электрического поля, и расстояния зарядов от пары частиц.

Если какой-нибудь заряд переместился из одной точки в другую, то, очевидно, силы, действующие со стороны этого заряда на другие заряды, изменятся. При непрерывном движении заряда эти силы также должны меняться непрерывно; однако, если распространение действия заряда совершается с конечной скоростью, это изменение будет отставать от перемещения заряда, что приводит к значительным усложнениям теории действия электрических сил. Чтобы учесть специфические эффекты, возникающие при наличии движущихся зарядов, вводится дополнительная характеристика, которую мы и назвали индукцией магнитного поля.

Введение этой характеристики позволяет существенно упростить всю теорию электрических явлений и не задумываться о том, что электрическое воздействие распространяется в пространстве с конечной скоростью. Так как приходится пользоваться двумя понятиями: напряженностью  $\mathbf{E}$  электрического поля и индукцией  $\mathbf{B}$  магнитного поля,—то явления, которые мы будем в дальнейшем изучать, получили общее название *электромагнитных явлений*.

Таким образом, известные еще из школьного курса магнитные силы представляют не что иное, как проявление электрических действий, вызванных движущимися зарядами. В природе не существует никаких особых магнитных зарядов, а есть только электрические заряды двух типов, условно называемые *положительными* и *отрицательными*.

В заключение заметим, что в выражении для силы Лоренца не случайно стоит коэффициент  $c$ . Его значение соответствует скорости света в вакууме, а это как раз та самая максимальная

скорость, с которой может распространяться электрическое поле заряда, возникшего в данной точке пространства.

Мы уже пользовались (см. § 2.5) терминами «электрическое поле», «магнитное поле», «поле сил». Уточним смысл этих понятий. В математике и физике под *полем* некоторой величины, заданной в пространстве, понимается то, что значение этой величины вполне и единственным образом определено во всех точках этого пространства.

Например, во всех точках неравномерно нагретого тела существует температура, значение которой меняется непрерывно при переходе от одной точки к другой. В таких случаях говорят, что в данном теле существует поле температур. Так как температура—это скалярная величина, то поле, подобное температурному, называется *скалярным*.

Другой пример. Рассмотрим массивное тело. Поместим в окружающем это тело пространстве другое тело относительно малых размеров. Назовем это тело *пробным*. Тогда на пробное тело со стороны центрального массивного тела будет действовать сила тяготения, которая может быть представлена вектором, имеющим вполне определенные модуль и направление. Если сместить пробное тело, то на него все равно будет действовать сила тяготения к центральному телу, но имеющая в другой точке уже иное направление и модуль. Вообще, в какую бы точку окружающего массивное центральное тело пространства мы ни помещали пробное тело, на него будет действовать сила притяжения. Следовательно, центральное массивное тело создает в окружающем пространстве поле сил тяготения. Поскольку сила изображается вектором, подобное поле будет векторным.

Отвлекаясь от физической природы векторного поля (а в дальнейшем мы в основном будем рассматривать именно векторные поля), назовем *векторным полем* пространство, в каждой точке которого задан вполне определенный и единственный вектор. Векторное поле называется *однородным*, если при перемещении от одной точки к другой вектор не меняется ни по модулю, ни по направлению. Наиболее часто встречаются поля неоднородные, т. е. такие, в которых при переходе от одной точки пространства к другой вектор меняется как по модулю, так и по направлению.

Как охарактеризовать поле? Один из способов (и самый общий) заключается в том, чтобы указать модуль и направление вектора в каждой точке пространства. Например, для характеристики поля тяготения в пространстве вокруг массивного центрального тела (для простоты будем считать его сферическим) достаточно задать в каждой точке пространства векторы, которые направлены к центру по радиусам, проходящим через данную точку пространства и центр сферы, а по модулю обратно пропорциональны квадрату расстояния от данной точки до центра сферы и прямо пропорциональны массе центрального тела. Каждый такой вектор характеризует напряженность создаваемого



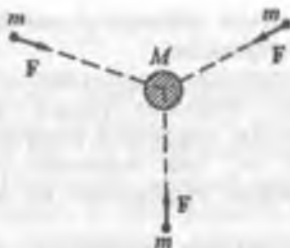


Рис. 7.1. Векторы поля сил тяготения



Рис. 7.2. Изображение поля сил тяготения

центральной телом поля тяготения в данной точке пространства (рис. 7.1).

Напряженность поля имеет конкретное физическое содержание. В данном случае это сила, которая действует на тело с единичной массой, помещенное в выбранную точку поля, со стороны центрального тела.

При описании поля аналитически задают функцию, которая характеризует вектор напряженности поля в любой точке рассматриваемого пространства. Однако часто для большей наглядности бывает удобно представить поле графически. Это можно сделать, задавая векторы напряженности поля в некоторых отдельных точках пространства, стараясь выбирать их так, чтобы общая графическая картина наглядно и достаточно полно помогала представить характер всего поля (рис. 7.2).

Можно сделать и так: начертить линии, которые в любой точке будут касательными к этим векторам. Для характеристики напряженности поля в этом случае можно проводить линии напряженности (*силовые линии*) гуще в тех местах, где напряженность поля больше, и реже там, где она меньше. Тогда получим картину силовых линий для поля тяготения, создаваемого некоторым массивным телом (рис. 7.3).

Однородное поле при таком способе будет изображаться совокупностью параллельных прямых (рис. 7.4), а неоднородное — системой сгущающихся и расходящихся кривых (рис. 7.5).

Та точка пространства, в которой силовые линии сходятся, называется *узлом* или *источником поля*.

Важным свойством линий напряженности является то, что они не пересекаются.

Обратимся к некоторым математическим характеристикам поля, которые позволят нам впоследствии производить вычисления характеристик электрического и магнитного полей для конкретных случаев и сформулировать общие законы электромагнетизма, являющиеся, подобно законам Ньютона в механике, обобщением всех известных в настоящее время экспериментальных фактов. Начнем с понятия циркуляции вектора по замкнутому контуру. Проведем в пространстве, в котором задано векторное поле  $F$ , произвольную замкнутую кривую  $L$  (рис. 7.6). Выберем на этой кривой определенное направление обхода,

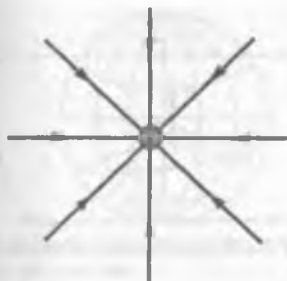


Рис. 7.3. Линии напряженности поля тяготения

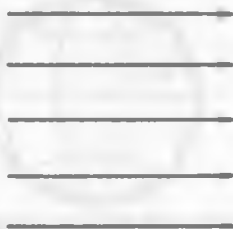


Рис. 7.4. Линии напряженности однородного поля

указанное стрелками. Для простоты будем считать, что все точки этой кривой лежат в одной плоскости. Заменим кривую ломаной линией, составленной из векторов  $\Delta l$ , малой длины (рис. 7.7).

Таким приемом мы уже пользовались ранее.

Длину вектора  $\Delta l$  выберем такой, чтобы на малом участке пространства, порядка  $|\Delta l|$ , поле можно было считать однородным. Необходимо также, чтобы ломаная линия с большой точностью совпадала с кривой  $L$ .

Составим скалярные произведения векторов поля  $F_i$  на каждом участке  $\Delta l_i$  на векторы  $\Delta l_i$ , т. е. произведения вида  $F_i \Delta l_i$ . Если вдоль кривой движется шарик, на который действует сила  $F_i$ , то произведение  $F_i \Delta l_i$  будет представлять элементарную работу, совершаемую силой  $F_i$  на участке  $\Delta l_i$ . Составим сумму  $\sum F_i \Delta l_i$  для всех участков кривой. Эта сумма равна полной работе, совершаемой переменной силой  $F$  при перемещении шарика вдоль всей кривой.

При стремлении  $\Delta l_i$  к нулю и числа участков ломаной к бесконечности вместо суммы получим интеграл  $\oint F dl$  ( $\oint$  указывает, что совершается полный обход кривой  $L$ ). Интеграл  $\oint F dl$  называется *циркуляцией вектора  $F$  вдоль замкнутой кривой*

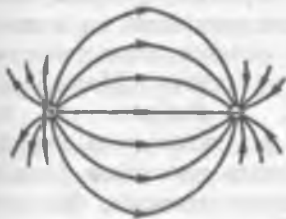


Рис. 7.5. Линии напряженности неоднородного поля (диполя)



Рис. 7.6. Замкнутая кривая в векторном поле



Рис. 7.7. Замена кривой ломаной линией ( $\Delta l_i \rightarrow dl_i$  при  $i \rightarrow \infty$ )

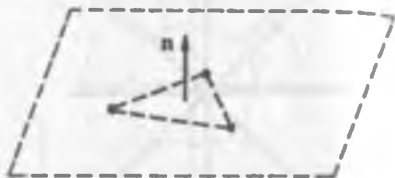


Рис. 7.8. К введению элемента площади

(контура)  $L$ . В частном случае циркуляция вектора равна работе некоторой силы по замкнутому пути. В общем случае, конечно, циркуляцию произвольного вектора  $F$  вдоль замкнутой кривой нельзя трактовать как работу. Эту величину следует рассматривать как математическую характеристику, весьма полезную при описании важнейших свойств векторного поля.

Второе понятие, с которым мы встретимся,—это понятие потока вектора через поверхность. Чтобы ввести это понятие, учтем следующее. Для того чтобы указать положение некоторой плоскости в пространстве, можно поступить разными способами. Например, можно указать, в каких точках данная плоскость пересекает декартовы оси координат.

Можно поступить иначе, а именно: провести вектор перпендикулярно плоскости (нормаль к плоскости) и задать его координаты в пространстве. Так как вектор, перпендикулярный плоскости, связан с положением этой плоскости вполне однозначно, то задание такой нормали также единственным и определенным образом характеризует положение в пространстве самой плоскости.

Если плоскость меняет свое положение, то одновременно поворачивается и нормаль. Такая характеристика плоскости чрезвычайно удобна. Поэтому она часто используется в математике (рис. 7.8).

Можно расширить такой способ описания и выбирать длину нормали, численно равной площади этой плоскости. Тогда, указывая нормаль, можно задавать положение плоскости в пространстве, а задавая длину этой нормали, указать площадь плоскости. Так как обычно пространство разделено на бесконечно малые элементы, то не будет никакой ошибки, если считать, что выбранная плоскость имеет форму квадрата. Можно задать вектор, перпендикулярный квадрату, и выбрать его длину, численно равной площади квадрата, что полностью определяет рассматриваемую плоскость. Если площадь плоскости бесконечно мала, то плоскость называется *элементом площади*. Обозначается этот элемент площади  $dS$ . Поскольку мы условились характеризовать площадь вектором, элемент площади тоже вектор.



Рис. 7.9. Разбиение сферической поверхности на участки



Рис. 7.10. Замена сферической поверхности «ломаной» сферой

Теперь рассмотрим в пространстве некоторую замкнутую область, например объем внутри сферической поверхности (рис. 7.9).

Подобно тому как мы вместо кривой строили ломаную, вместо сферической поверхности можно построить поверхность, которая состоит из участков плоскостей (рис. 7.10).

Если выбирать эти плоские участки достаточно маленькими, то точность такого представления поверхности сферы «ломаной» сферой может быть сколь угодно велика.

Предположим, что в пространстве задано векторное поле вектора  $F$ . Будем выбирать участки «ломаной» сферы столь малыми, чтобы модули векторов  $F_i$  во всех точках каждого из участков «ломаной» сферы можно было считать неизменными ( $i$ —индекс участка сферы).

Составим скалярные произведения векторов  $F_i dS_i$  для каждого плоского участка рассматриваемой поверхности, после чего просуммируем эти произведения по всем участкам поверхности. Эта сумма называется *поток вектора  $F$  через поверхность  $S$* .

Если участки сферы выбрать достаточно малыми, то суммы  $\sum F_i dS_i$  можно выразить интегралом. В дальнейшем будем различать замкнутые и незамкнутые поверхности. *Замкнутой поверхностью* будем называть такую поверхность, которая ограничивает со всех сторон какой-то объем пространства (например, сфера; внутри нее находится вполне определенный объем). Если взять только часть сферы, то поверхность называется *незамкнутой*. Она никакого объема не ограничивает. В соответствии с этим вводятся два обозначения для потока вектора сквозь поверхность.

Если мы пишем  $\oint_S F dS$ , то это значит, что мы рассматриваем поток вектора  $F$  сквозь поверхность  $S$ , являющуюся замкнутой. Замкнутость указывается кружком на символе интеграла. Физический смысл понятия потока вектора сквозь поверхность можно представить на следующем наглядном примере. Рассмотрим поток текущей жидкости. Скорость течения жидкости в каждой ее точке можно охарактеризовать вектором. Если поток



Рис. 7.11. Поток вектора  $F = \rho v$  сквозь поверхность  $dS$

жидкости установившийся, то этот вектор не будет меняться во времени. Тогда количество жидкости, которое пройдет в единицу времени через малую поверхность  $dS_i$ , расположенную в данной точке под углом  $\alpha_i$  к вектору скорости  $v_i$ , равно  $\rho |v_i| |dS_i| \cos \alpha_i$  (рис. 7.11). Здесь  $\rho$ —плотность жидкости.

Если из таких малых участков составить большую поверхность, то общее количество жидкости, которая пройдет в единицу времени через всю поверхность, равно  $\sum_i \rho v_i dS_i$ , или  $\int_S \rho v dS$  для незамкнутой и  $\oint_S \rho v dS$  для замкнутой поверхности.

Задавая поток вектора  $F$  через  $S$ , можно судить о модуле вектора  $F$ , что нам в дальнейшем и потребуется.

Введенные понятия потока и циркуляции вектора являются в учении об электричестве основными. С помощью этих общих понятий можно получить много конкретных выводов. Все законы, описывающие электромагнитные явления, могут быть сформулированы с использованием этих двух понятий.

К сказанному необходимо добавить еще одно важное замечание. Если имеется несколько источников поля (например, несколько электрических зарядов), то общая напряженность или индукция поля в любой точке пространства будет выражаться вектором, равным векторной сумме напряженностей в этой точке поля, создаваемых каждым зарядом в отдельности. Это утверждение называется *принципом суперпозиции полей* и выполняется как для электрического, так и для магнитного полей.

Принцип суперпозиции позволяет находить электрическое или магнитные поля в случае, когда имеется несколько зарядов или проводников с токами.

## 7.2. Электростатика

**Закон Кулона. Поле нескольких зарядов.** Электростатические поля молекул. *Электростатика*—раздел общего учения об электричестве и магнетизме, в котором изучаются свойства электрических полей, создаваемых совокупностью неподвижных электрических зарядов. Если заряды неподвижны и неизменны по величине, то в пространстве возникают только электрические поля, а магнитных полей нет.

Основным законом для статических электрических полей и взаимодействия зарядов является закон Кулона.

Этот закон формулируется следующим образом: два точечных заряда  $q_1$  и  $q_2$  взаимодействуют между собой с силой, направленной вдоль прямой, соединяющей эти заряды, и прямо пропорциональной произведению величин зарядов и обратно пропорциональной квадрату расстояния между ними. Этот закон непосредственно вытекает из прямых экспериментов, которые могут быть поставлены в разных вариантах.

В векторной записи этот закон имеет вид

$$F_{12} = \alpha \frac{q_1 q_2}{|r_{12}|^3} r_{12},$$

где  $F_{12}$ —сила, действующая на второй заряд со стороны первого;  $r_{12}$ —вектор, проведенный от первого заряда ко второму;  $\alpha$ —коэффициент пропорциональности, зависящий от размерности.

В Международной системе единиц (СИ)  $\alpha = (4\pi\epsilon_0)^{-1}$ , где  $\epsilon_0$ —электрическая постоянная. Поскольку мы не будем заниматься конкретными вычислениями величин разного рода действующих на заряды сил, чтобы не усложнять дальнейших формул, в ряде случаев примем коэффициент  $\alpha = 1$ .

Прямым следствием закона Кулона является следующее утверждение: *всякий точечный заряд создает вокруг себя статическое электрическое поле, напряженность которого представляет вектор, направленный вдоль вектора, соединяющего центр заряда и точку наблюдения, и по величине прямо пропорциональный заряду и обратно пропорциональный квадрату расстояния от заряда до точки измерения.*

В векторной форме это утверждение может быть записано так:

$$E = \frac{\alpha q}{|r_A|^3} r_A \quad \text{или} \quad E = \frac{q}{|r_A|^3} r_A,$$

если считать  $\alpha = 1$ . Здесь  $r_A$ —радиус-вектор точки А. Начало координат совмещено с зарядом.

В этой записи указывается и направление вектора напряженности  $E$  электрического поля, и его значение. Видно, что если заряд положителен, то  $E \parallel r_A$ , а если отрицателен, то векторы  $E$  и  $r_A$  имеют противоположные направления. Напомним, что под напряженностью электрического поля мы понимаем силу, действующую на точечный заряд единичной величины, причем этот заряд (его обычно называют *пробным*) всегда полагается положительным. Если центральный заряд положителен, то пробный заряд будет отталкиваться от него (действующая на пробный

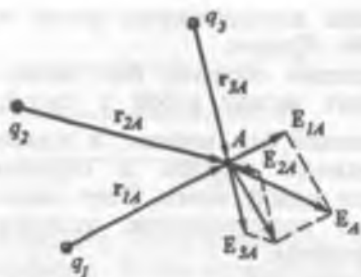


Рис. 7.12. Нахождение напряженности  $E$  в заданной точке поля, созданного тремя зарядами

заряд сила направлена по радиусу-вектору), а если отрицателен, то притягиваться (сила противоположна  $r_A$ ).

Для электрических полей, создаваемых многими источниками, выполняется принцип суперпозиции. Это означает, что напряженность поля в некоторой точке  $A$  в пространстве, где имеются заряды  $q_1$ ,  $q_2$  и т. д., может быть вычислена по правилам векторного сложения (рис. 7.12):

$$E = E_{1A} + E_{2A} + \dots = \sum_{i=1}^n E_{iA} = \sum_{i=1}^n \frac{q_i r_{iA}}{|r_{iA}|^3}.$$

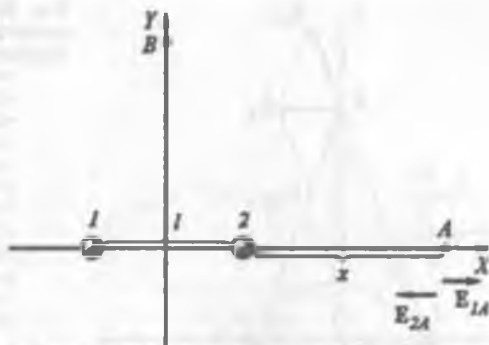
Для примера вычислим напряженность поля двух точечных зарядов (диполя).

**Диполем** называются два одинаковых по величине, но противоположных по знаку точечных заряда, расположенных друг по отношению к другу на расстоянии  $l$ . Рассмотрим напряженность электрического поля в точках  $A$  и  $B$ . Одна из этих точек расположена на оси диполя (рис. 7.13), а другая — перпендикулярно оси диполя, симметрично по отношению к этим зарядам. Напряженность поля, создаваемого в точке  $A$  на оси диполя первым зарядом, равна  $|E_{1A}| = q/(x + l)^2$  и направлена так, как показано на рис. 7.13. Напряженность от второго заряда равна по модулю  $|E_{2A}| = q/x^2$  и направлена в противоположную сторону. Полная напряженность равна вектору, направленному по оси диполя и по модулю равному

$$|E| = \left| |E_{1A}| - |E_{2A}| \right| = q \left| \left( \frac{1}{(x + l)^2} - \frac{1}{x^2} \right) \right|.$$

Если мы выберем точку наблюдения на достаточно большом расстоянии от диполя, то расстояние между зарядами можно рассматривать как малую величину по сравнению с расстоянием от точки наблюдения до диполя. В таком случае  $l$  может рассматриваться как малое приращение аргумента функции, зависящей от расстояния от точки наблюдения до крайнего правого заряда. Если воспользоваться правилом представления

Рис. 7.13. Напряженность поля, создаваемого в точке  $A$  на оси диполя



значения функции в точке, лежащей близко к данной, то можно записать:

$$\frac{1}{(x + l)^2} = \frac{1}{x^2} - \frac{2}{x^3}l$$

Воспользовавшись этим выражением, найдем, что в точке  $A$ , лежащей на оси диполя на большом расстоянии от него, напряженность поля равна

$$|E_A| = 2ql/x^3.$$

Таким образом, напряженность электрического поля в этом случае убывает обратно пропорционально кубу расстояния.

Найдем теперь напряженность электрического поля на прямой, проходящей через центр диполя, причем точку наблюдения  $B$  выберем на достаточно большом расстоянии  $y$  от оси диполя. Напряженность электрического поля, создаваемого в точке  $B$  первым зарядом,—это вектор (рис. 7.14), модуль которого

$$|E_{1B}| = q \frac{1}{|r_{1B}|^2} = \frac{q}{(y^2 + l^2/4)}.$$

Напряженность электрического поля от второго заряда диполя равна по модулю той же напряженности, но направленной так, как показано на рис. 7.14.

Полная напряженность равна

$$|E_B| = \frac{2q \sin \alpha}{(y^2 + l^2/4)} = \frac{ql}{(y^2 + l^2/4)^{3/2}}.$$

При достаточно большом расстоянии от точки наблюдения до диполя величина  $l$  становится малой по сравнению с  $y$  и тогда:

$$|E_B| \approx ql/y^3.$$

Таким образом, напряженность электрического поля и в этом случае обратно пропорциональна кубу расстояния.

И в том, и в другом случаях модуль вектора напряженности электростатического поля зависит от произведения  $p = ql$ ,



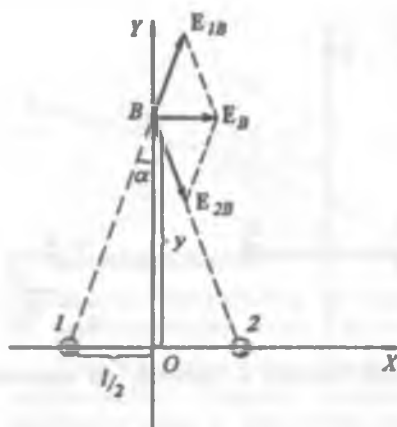


Рис. 7.14. Напряженность поля, создаваемого в точке  $B$  на прямой, перпендикулярной оси диполя

которое называется *дипольным моментом* и определяется лишь величиной одного из зарядов и расстоянием между ними. Если имеются два диполя, то, поскольку каждый из них создает в пространстве электростатическое поле, между ними возникнет взаимодействие. В частности, если диполи расположены так, как показано на рис. 7.15,а, они будут притягиваться друг к другу, а если так, как на рис. 7.15,б, то отталкиваться (стрелками указаны направления действующих на диполи сил). Реально, однако, притяжение всегда будет превалировать над отталкиванием, так как состояние б неустойчиво в том смысле, что любое малое отклонение диполей от оси  $X$  будет приводить к силе, стремящейся повернуть диполи к состоянию а.

Можно рассмотреть и еще ряд примеров, касающихся совокупности положительных и отрицательных зарядов. В таких случаях, как правило, напряженность электрического поля уменьшается значительно быстрее, чем напряженность электрического поля от одного точечного заряда. Именно поэтому, хотя все тела состоят из положительных и отрицательных зарядов (электроны и ядра атомов), но тем не менее они ведут себя как нейтральные тела. Это происходит потому, что напряженность электрического поля, создаваемого этими зарядами, спадает очень быстро.

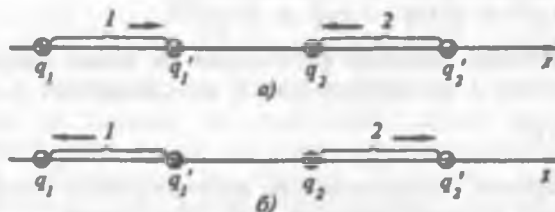
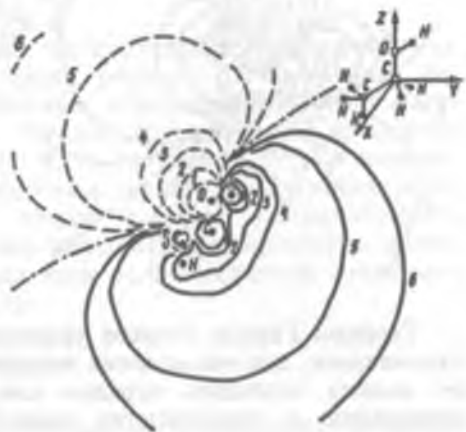


Рис. 7.15. Взаимодействие двух диполей

Рис. 7.16. Карта электростатического поля вокруг молекулы этилового спирта.

Изображены линии равного потенциала в плоскости симметрии молекулы. Области положительного потенциала показаны сплошными линиями, а отрицательного — пунктирными. Линия раздела областей положительного и отрицательного потенциалов показана прерывистой линией с точками. Жирными точками обозначены положения атомов на плоскости чертежа. Относительные величины потенциалов для линий, маркированных цифрами в порядке от 1 до 6, меняются следующим образом: 1; 0,5; 0,2; 0,1; 0,02; 0,01.

Видно, во-первых, что электростатическое поле вокруг молекулы имеет довольно сложную форму, а во-вторых, что уже на расстоянии, приблизительно равном размеру самой молекулы, потенциал поля убывает на два порядка по сравнению с областью, непосредственно прилегающей к атомам молекулы. Это последнее объясняет, в частности, почему межмолекулярные взаимодействия проявляются лишь на очень малых расстояниях между молекулами.



Весьма сложную форму имеют электростатические поля многоатомных молекул. Примеры таких полей, изображенных с помощью эквипотенциальных поверхностей<sup>\*)</sup> (или изопотенциальных линий на плоскости), приведены на рис. 7.16, 7.17. Наличие вокруг молекул областей повышенных положительных и отрицательных потенциалов существенно сказывается на ходе ион-молекулярных реакций, в частности реакций протонирования и окисления. Наличие полярных заместителей в молекулах, даже удаленных от реакционных центров, способно иногда радикально изменить вид потенциала около реакционного центра (рис. 7.17), а это, в свою очередь, приводит к изменению хода реакций. Так, например, именно изменением электростатического поля объясняются многие закономерности хода реакций (зависимости от сорта и расположения заместителя) нуклеофильного и электрофильного замещения в ароматических соединениях. Нейтральные молекулы на значительных расстояниях от них часто создают поля дипольного типа. Поэтому диполь-дипольные взаимодействия играют важную роль в межмолекулярных взаимодействиях в газах, жидкостях и молекулярных кристаллах.

<sup>\*)</sup> Определение понятия дано ниже.



Рис. 7.17. Карта потенциала электростатического поля вокруг молекулы хлорзамещенного этилового спирта.

Все обозначения и расположение молекулы по отношению к осям координат те же, что и на рис. 7.16. Бросается в глаза, что введение заместителя с большим зарядом (сильно полярного) приводит к существенному изменению вида поля вокруг молекулы. Такой эффект объясняет влияние полярных заместителей на реакционную способность органических соединений

**Теорема Гаусса (первое уравнение Максвелла).** Как уже указывалось, для вычисления напряженности электрического поля от многих точечных зарядов или от одного распределенного непрерывно в пространстве заряда необходимо провести суммирование напряженностей электрических полей от каждого заряда. В общем случае это суммирование может оказаться довольно сложным. Тогда можно поступить иначе.

Пусть точечный заряд окружен сферой с центром в этом заряде. Радиус сферы равен  $r$ .

Посмотрим, каким будет поток напряженности электрического поля сквозь поверхность сферы. Согласно закону Кулона напряженность поля на всей поверхности сферы представляет вектор, направленный перпендикулярно поверхности сферы во всех ее точках и равный по модулю  $|E| = \alpha q / |r|^2$ . По определению *поток вектора сквозь малый участок сферы*—скалярное произведение вектора напряженности электрического поля на вектор элемента площади. Напомним, что вектор элемента площади по модулю равен малому участку площади, выделяемой на сфере, а по направлению совпадает с перпендикуляром к этой сфере. Выберем направление такого вектора из сферы наружу, тогда скалярное произведение

$$(E \Delta S) = |E| |\Delta S| = q \frac{\alpha}{|r|^2} |\Delta S| = q \frac{\alpha}{r^2} \Delta S.$$

Полный поток напряженности электрического поля сквозь сферу радиуса  $r$  равен сумме всех потоков через каждый элемент площади сферы в отдельности. Так как модуль  $|E|$  на всех участках сферы один и тот же, то полный поток напряженности

$$\oint_S E dS = |E| \sum_i |\Delta S_i| = |E| \oint_S dS,$$

где  $\oint_S |dS|$  — полная поверхность сферы. Следовательно, полный поток оказывается равным

$$\oint_S E dS = |E| \cdot 4\pi r^2 = \frac{q\alpha}{r^2} \cdot 4\pi r^2 = 4\pi q.$$

Смысл этого выражения таков: полный поток напряженности электрического поля сквозь сферу радиуса  $r$  не зависит от радиуса этой сферы. Полученное утверждение может быть обобщено в виде теоремы Гаусса или первого уравнения Максвелла. Эта теорема гласит: если окружить любую совокупность точечных или распределенных зарядов любой замкнутой поверхностью, то полный поток вектора напряженности электрического поля сквозь эту замкнутую поверхность всегда равен с постоянным коэффициентом полному заряду, заключенному внутри этой поверхности:

$$\oint_S E dS = 4\pi \sum_i q_i = \frac{1}{\epsilon_0} \sum_i q_i$$

Теорема Гаусса для любой замкнутой поверхности и любой совокупности зарядов получается, если произвольная замкнутая поверхность заключена между двумя сферами — малой и большой (рис. 7.18). Так как поток вектора  $E$  сквозь сферы 1 и 2 должен быть одним и тем же, то он будет таким же и сквозь любую находящуюся между сферами поверхность  $S$ . Не играет никакой роли и месторасположение заряда внутри сферической или иной замкнутой поверхности. Значит, если внутри поверхности расположить несколько зарядов, то поток  $E$  станет равен их сумме. Это следствие закона Кулона и принципа суперпозиции. Теорема Гаусса позволяет, в частности, проводить вычисление напряженности электрического поля во многих конкретных случаях. Самое важное заключается в том, что теорема Гаусса сохраняется и для движущихся зарядов, когда закон Кулона уже не действует. Рассмотрим применение приведенного здесь общего закона для некоторых частных примеров.

Возьмем любой очень тонкий провод либо цилиндр достаточно длинный, чтобы его радиус был мал по сравнению с длиной самого цилиндра. Зарядим этот проводник равномерно по всей поверхности. Выясним направление напряженности электрического поля вокруг проводника. Для этого разобьем проводник на участки очень малой длины. Каждый такой участок можно считать точечным зарядом (точка  $A$ ).

Рассмотрим напряженность электрического поля в точке  $B$ , создаваемую участком проводника (рис. 7.19).

Вектор напряженности  $E_{AB}$  электрического поля в этом случае имеет направление, показанное на рис. 7.19. Так как проводник очень длинный, то всегда может быть найдена, очевидно, другая точка на проводнике — точка  $A'$ , которую мы



Рис. 7.18. К выводу теоремы Гаусса

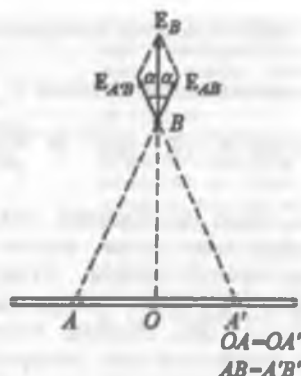


Рис. 7.19. Напряженность электрического поля в точке  $B$ , создаваемая участком проводника

назовем сопряженной точкой, отличающейся тем, что напряженность  $E_{A'B}$  электрического поля, создаваемая зарядом, сосредоточенным в точке  $A'$ , направлена так, как показано на рис. 7.19. Сопряженную точку будем выбирать следующим образом.

Проведем через точку наблюдения перпендикуляр к заряженному проводнику. Отложим на проводнике справа от точки пересечения перпендикуляра с проводником отрезок, равный отрезку от точки  $A$  до пересечения перпендикуляра к проводнику. Конец этого отрезка и будет играть роль сопряженной точки. Нетрудно видеть, что векторы напряженности электрических полей в точке  $B$  от каждой из пары таких сопряженных точек образуют один и тот же угол с перпендикуляром и одинаковы по модулю. Пользуясь правилом сложения векторов, определим, что полная напряженность электрического поля каждой пары таких сопряженных точек представляет вектор, направленный перпендикулярно направлению заряженного проводника. Так как проводник бесконечно длинный, то его поверхность может быть разбита на пары сопряженных точек. Такое рассмотрение показывает, что электрическое поле около длинного проводника должно описываться векторами, направленными перпендикулярно направлению проводника так, как показано на рис. 7.19, т. е. векторами, перпендикулярными проведенной вокруг заряженного линейного проводника цилиндрической поверхности, осевая линия которой совпадает с проводником. Вектор напряженности электрического поля во всех точках этой поверхности перпендикулярен ей и одинаков по модулю. Применим теорему Гаусса к этому проводнику. Длину проводника примем равной  $l$ . Замкнутая поверхность  $S$  образуется цилиндрической поверхностью и ее торцами. Будем считать, что на каждой единице длины проводника расположен заряд  $\sigma$ . Тогда общий заряд, распределенный на всем проводнике и, следовательно

но, заключенный внутри замкнутой поверхности  $S$ , равен  $\sigma l$ . Полный поток вектора напряженности электрического поля сквозь замкнутую поверхность, создаваемую боковой поверхностью цилиндра и его торцами, равен

$$\oint_S \mathbf{E} d\mathbf{S} = \frac{1}{\epsilon_0} \sigma l.$$

Если проводник бесконечно длинный и радиус окружающего цилиндра мал по сравнению с длиной, то поток сквозь торцевую поверхность можно считать пренебрежимо малым по сравнению с потоком напряженности электрического поля сквозь боковую поверхность. Следовательно, поток вектора напряженности электрического поля в этом случае равен потоку напряженности вектора электрического поля через боковую поверхность цилиндра. Поскольку вектор напряженности электрического поля по модулю на всех участках этой поверхности один и тот же, а по направлению на любом участке боковой поверхности проводника перпендикулярен поверхности, для потока получим следующее выражение:

$$\sum_i |\mathbf{E}_i| |\Delta S_i| = \oint_S \mathbf{E} d\mathbf{S} = |\mathbf{E}| \oint_S dS = |\mathbf{E}| \sum_i \Delta S_i,$$

где  $\oint_S dS_i = \sum_i \Delta S_i$  — площадь боковой поверхности цилиндра.

Если радиус цилиндрической поверхности принять равным  $r$  (расстояние от точки наблюдения до заряженного проводника), то полный поток определяется формулой

$$\oint_S \mathbf{E} d\mathbf{S} = \frac{1}{\epsilon_0} \sigma l = |\mathbf{E}| 2\pi r l.$$

Отсюда находим, что напряженность электрического поля на расстоянии  $r$  от цилиндра равна

$$|\mathbf{E}| = \frac{1}{2\pi\epsilon_0} \sigma / r.$$

Итак, напряженность электрического поля очень длинного проводника или цилиндра на расстояниях от него, малых по сравнению с длиной проводника (цилиндра), пропорциональна плотности заряда на единицу длины и обратно пропорциональна расстоянию от проводника.

Для точечного заряда мы получили зависимость напряженности электрического поля, обратно пропорциональную квадрату расстояния. В данном случае напряженность поля падает медленнее.

Рассмотрим еще один пример. Представим, что имеется равномерно заряженная плоскость больших размеров. Будем рассматривать поле на расстоянии значительно меньшем, чем размеры этой плоскости. Пусть на каждую единицу площади

приходится заряд  $\sigma$ , тогда общий заряд равен  $\sigma S$ , где  $S$ —площадь всей плоскости.

Определим направление напряженности электрического поля около такой плоскости, выбирая расстояние до плоскости малым по сравнению с размерами плоскости. Выберем некоторый малый участок на плоскости (точка  $A$  на рис. 7.20). Рассмотрим некоторую точку  $B$  вне плоскости. Напряженность поля, создаваемого участком около точки  $A$  на плоскости в точке  $B$ , показана на рис. 7.19.

Проведем перпендикуляр через точку  $B$  к плоскости и плоскость через перпендикуляр и точку  $A$ . Эта плоскость пересечет заряженную плоскость вдоль прямой  $AA'$ . Отложим на этой прямой точку  $A'$  на расстоянии от перпендикуляра, равном расстоянию от точки  $A$  до перпендикуляра. Назовем такую точку точкой, сопряженной для точки  $A$ . По условию в окрестности этой точки сосредоточен такой же заряд, что и в окрестности точки  $A$ . Напряженность поля от такого заряда, который можно рассматривать как точечный заряд, равна  $E_{A'B}$ . Простым геометрическим построением можно показать, что векторы  $E_{AB}$  и  $E_{A'B}$  образуют с перпендикуляром к плоскости одинаковые углы и по модулю тоже одинаковы. Следовательно, вектор полной напряженности поля, создаваемого в точке  $B$  двумя сопряженными точками на поверхности заряженной плоскости, перпендикулярен плоскости. Поскольку плоскость имеет бесконечные размеры, то она вся может быть разбита на пары таких сопряженных точек. Следовательно, в любой точке около плоскости вектор напряженности электрического поля перпендикулярен поверхности плоскости. Если плоскость заряжена положительно, то над плоскостью напряженность электрического поля будет, очевидно, представляться вектором, показанным на рис. 7.20. Под плоскостью напряженность электрического поля направлена в противоположную сторону.

Заклучим мысленно плоскость в прямоугольный ящик (рис. 7.21). Напряженность электрического поля на верхних и нижних границах ящика перпендикулярна его поверхности. Модуль вектора  $E$  будет одним и тем же на каждой такой поверхности, как и в случае прямого длинного провода. Влиянием боковых поверхностей ящика пренебрегаем, считая размеры этих поверхностей бесконечно малыми по сравнению с размерами верхней и нижней границ ящика. Учитывая это, для потока вектора напряженности электрического поля сквозь верхнюю и нижнюю границы ящика получим

$$\oint_S E dS = |E| \cdot 2S.$$

По условию этот поток должен быть равен полному заряду плоскости, точнее величине  $\sigma S/\epsilon_0$ . Отсюда заключим, что модуль

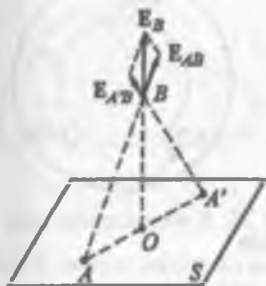


Рис. 7.20. К выводу формулы для напряженности поля равномерно заряженной плоскости

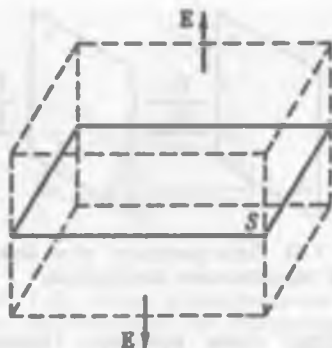


Рис. 7.21. Плоскость в "ящике"

вектора напряженности электрического поля такой заряженной плоскости равен

$$|E| = \sigma / (2\epsilon_0).$$

Таким образом, напряженность поля во всех точках около равномерно заряженной плоскости будет одна и та же. Такое поле мы ранее называли однородным. Следовательно, для того чтобы получить однородное электрическое поле, достаточно взять плоскость больших размеров, зарядить эту плоскость равномерно и проводить измерения на малом расстоянии по сравнению с размерами плоскости.

Рассмотрим теперь поле между двумя заряженными плоскостями. Возьмем две одинаковые плоскости, расположенные параллельно друг другу на малом по сравнению с размерами этих плоскостей расстоянии и зарядим одну плоскость зарядом с поверхностной плотностью  $+\sigma$ , а другую плоскость зарядом  $-\sigma$ . Картина расположения силовых линий поля около плоскостей в этом случае показана на рис. 7.22. Так как плоскости заряжены разноименно, то векторы напряженности поля одной плоскости направлены от плоскости, а другой—к плоскости.

Напряженность поля в пространстве согласно принципу суперпозиции равна векторной сумме напряженностей от каждой из плоскостей. Так как эти напряженности перпендикулярны плоскостям и обе по модулю равны  $\sigma / (2\epsilon_0)$ , то, как видно из рис. 7.22, для напряженности поля между плоскостями справедливы равенства

$$E = E_1 + E_2, \quad |E| = \sigma / \epsilon_0.$$

Вне плоскостей, так как векторы  $E_1$  и  $E_2$  направлены в разные стороны, полная напряженность равна нулю. Итак, напряженность поля между плоскостями удваивается, а вне плоскостей оказывается равной нулю.



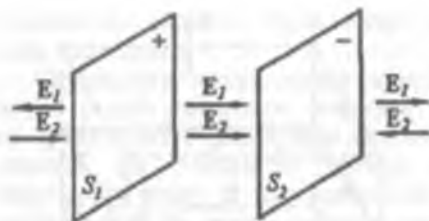


Рис. 7.22. Электрическое поле между двумя заряженными плоскостями

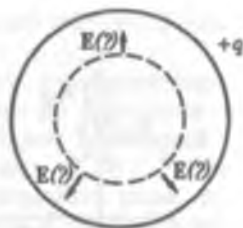


Рис. 7.23. К определению поля внутри шара

Еще один пример. Представим, что полый металлический тонкостенный шар заряжен. Оценим напряженность поля внутри шара. Проведем мысленно внутри шара сферу с центром в центре шара. Так как в этом случае никаких преимущественных направлений нет, то напряженность  $E$  поля должна была бы быть направлена по радиусу либо из центра к краям, либо, наоборот, от краев к центру, но по-прежнему  $|E|$  в любой точке проведенной сферы должен быть одним и тем же (рис. 7.23).

Если радиус сферы, проведенной внутри заряженного по поверхности шара, равен  $r$ , то  $\oint_S E dS = 4\pi r^2 |E|$ . Этот интеграл

должен быть пропорционален сумме зарядов, находящихся внутри сферы. Так как внутри сферы зарядов нет, то

$$4\pi r^2 |E| = 0 \quad \text{или} \quad |E| = 0.$$

Итак, электрического поля внутри заряженной сферы (будем говорить о равномерно заряженной сфере) нет. Этот эффект используют, когда надо защитить какой-нибудь прибор от воздействия сильных внешних электрических полей. В этом случае прибор заключают в металлическую оболочку. Оболочка может быть заряженной, но внутри поле равно нулю. Можно показать, что какой бы ни была форма поверхности, сделанной из металла, все равно внутреннее поле равно нулю.

Работа электрических сил. Второе уравнение Максвелла. Рассмотрим снова точечный заряд  $q_1$ . Вне этого заряда проведем замкнутую кривую. Отметим на этой кривой точки  $A$  и  $B$ . Пусть точка  $A$  соответствует минимальному расстоянию на замкнутой кривой до точечного заряда, а точка  $B$ —максимальному расстоянию до точечного заряда. Предположим, что точечный центральный заряд имеет положительный знак и что вдоль кривой от точки  $A$  к точке  $B$  и затем обратно перемещается тоже положительный заряд  $q_2$  (рис. 7.24). На участке от  $A$  до  $B$ , поскольку происходит отталкивание зарядов, направление действия электрической силы и направление перемещения совпадают. Следовательно, в этом случае работа будет положительной. На



Рис. 7.24. К выводу работы по переносу заряда  $q_2$  в электрическом поле заряда  $q_1$

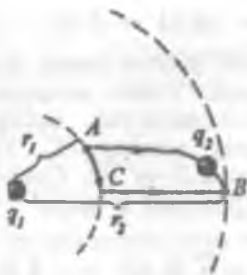


Рис. 7.25. Работа в электрическом поле при перемещении по поверхности и радиусу сферы

участке от  $B$  к  $A$  при движении по другой части кривой направление электрической силы и перемещения противоположны, поэтому на этом участке работа отрицательна.

Вспомним теперь картину линий напряженности электрического поля около точечного заряда. Окружим заряд сферой радиуса  $r$  и будем считать, что заряд перемещается вдоль поверхности этой сферы. В этом случае, так как напряженность электрического поля во всех точках сферы перпендикулярна поверхности сферы, работа, совершаемая под действием электрических сил при движении по сфере, равна нулю. В самом деле, как мы знаем из механики, работа представляет собой на каждом отдельном малом участке величину, равную скалярному произведению силы, действующей на этом участке, и вектора перемещения. Следовательно, на каком бы расстоянии от точечного заряда мы ни находились, если только движение совершается по поверхности сферы, окружающей данный точечный заряд, работа всегда равна нулю.

Рассмотрим произвольное перемещение заряда из одной точки пространства в другую. Это перемещение всегда может быть представлено как перемещение по радиусу сферы и затем по поверхности (рис. 7.25). Так как работа перемещения по поверхности сферы равна нулю, то полная работа электрических сил при переходе из одной точки в другую точку будет определяться только работой, соответствующей переходу со сферы одного радиуса на сферу другого радиуса. В случае замкнутой кривой (см. рис. 7.24) работа по перемещению точечного заряда из точки  $A$  в точку  $B$  определяется работой, соответствующей переходу заряда от сферы с радиусом, проведенным через точку  $A$ , до сферы, проведенной через точку  $B$ . На обратном пути от  $B$  к  $A$  работа должна быть равна той же величине, но взятой с обратным знаком. Отсюда мы заключим, что полная работа по перемещению любого точечного заряда в поле другого точечного заряда по замкнутой кривой всегда равна нулю в электростатическом поле.

Работа на каждом участке пути определяется следующей формулой:

$$\Delta A = qE \Delta l = F \Delta l.$$

Полная работа равна сумме работ по всем участкам замкнутой кривой. Такая величина, как было указано во введении к курсу электричества, называется циркуляцией вектора силы по замкнутому контуру  $L$ . Так как поле электростатическое, то вектор силы  $F = qE$ . Следовательно, формула для работы может быть переписана в виде

$$A = q \sum_i E_i \Delta l_i \rightarrow q \oint_L E dl = 0$$

(если  $L$ —замкнутый контур). На основании этой формулы можно заключить, что циркуляция вектора напряженности электрического поля по замкнутому контуру  $L$  в поле одного точечного заряда равна нулю.

Выше было сказано, что имеет место принцип суперпозиции полей. Это означает, что напряженность поля, создаваемого рядом точечных или протяженных источников, равна векторной сумме напряженностей полей от каждого точечного источника. Чтобы определить работу электрических сил в таком сложном электрическом поле, достаточно просуммировать, очевидно, работы в полях, создаваемых каждым точечным источником. В результате можно сделать общее заключение, что *циркуляция вектора напряженности электрического поля по замкнутому контуру в любом электростатическом поле равна нулю*. Это утверждение иногда называют вторым уравнением Максвелла для электростатики. Это уравнение записывается в форме

$$\oint_L E dl = 0.$$

**Энергия совокупности точечных зарядов.** Поскольку точечные заряды либо притягиваются друг к другу, либо отталкиваются и поскольку сила взаимодействия зависит лишь от относительного расположения зарядов, а работа по их перемещению зависит не от пути, по которому перемещались заряды при переходе от начального положения к конечному, а от характеристик этих состояний, то система точечных зарядов должна обладать потенциальной энергией.

Для двух точечных зарядов, если  $l_1$ —начальное расстояние между ними, а  $l_2$ —конечное, изменение потенциальной энергии равно взятой с обратным знаком работе кулоновских сил на участке 1—2:

$$\Delta U = -\Delta A = -\int_1^2 a \frac{q_1 q_2}{r^2} dl = a q_1 q_2 \left( \frac{1}{l_2} - \frac{1}{l_1} \right).$$

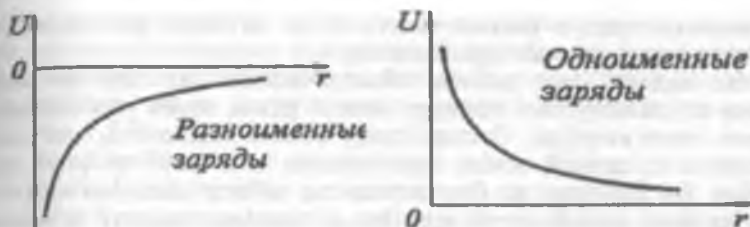


рис. 7.26. Зависимость потенциальной энергии от расстояния между зарядами

Если конечное расстояние считать очень большим и ввести радиус-вектор  $r_{12}$ , проведенный от первого заряда ко второму, то потенциальная энергия (потенциальная функция)

$$U = \alpha q_1 q_2 / |r_{12}|.$$

Кривая зависимости потенциальной энергии от расстояния между зарядами имеет вид гиперболы (рис. 7.26). Такая кривая не имеет минимума. Это означает, что система двух взаимодействующих зарядов ни при каком расстоянии между ними не может находиться в состоянии устойчивого равновесия.

Рассмотрим систему многих точечных зарядов. Каждый заряд будет испытывать притяжение или отталкивание со стороны других зарядов независимо от того, сколько имеется этих зарядов. Полная потенциальная энергия системы равна сумме потенциальных энергий этих пар:

$$U = \sum_{ij} \alpha q_i q_j \frac{1}{|r_{ij}|},$$

где  $|r_{ij}|$  — расстояние между каждой парой зарядов. Каждая составляющая этой суммы обратно пропорциональна расстоянию  $r_{ij}$ . Видно, что и в этом случае система точечных зарядов не имеет состояния равновесия. Это утверждение названо *теоремой Ирншоу* и имеет важное значение для понимания строения атома. В самом деле, атом состоит из разделенных в пространстве точечных положительных и отрицательных зарядов. Если бы все эти заряды были неподвижными, то по теореме Ирншоу можно было бы утверждать, что атом не может находиться в состоянии устойчивого равновесия, т. е. не может существовать. Отсюда следует вывод: для того чтобы атом существовал, сохраняя раздельные частицы в пространстве, эти частицы обязательно должны двигаться. Впоследствии мы увидим, что даже и в таком предположении, являющемся значительно более близким к истине, все равно не удастся объяснить в пределах классической физики существование стабильного атома.

**Потенциал.** *Потенциалом  $\varphi$*  называется взятая с обратным знаком работа электрических сил, которая совершается при перенесении пробного (точечного, единичного и положительного) заряда из данной точки пространства, где имеется электро-

статическое поле, в бесконечность, т. е. на такое расстояние, где действие поля уже не проявляется.

Мы видели, что работа электрических сил при движении заряда по замкнутому контуру всегда равна нулю независимо от формы этого контура. Отсюда сразу можно заключить, что каким бы путем из данной точки пространства точечный пробный заряд ни был бы выведен на бесконечность, работа электрических сил должна быть одной и той же. Это и позволяет ввести потенциал как скалярную величину, зависящую лишь от выбранной начальной точки и характеристик электростатического поля.

Для точечного заряда  $q$  потенциал точки, находящейся от него на расстоянии  $r$ , равен взятой с обратным знаком работе кулоновской силы при перемещении пробного заряда по радиусу на очень большое расстояние:

$$\varphi = - \int_r^{\infty} \alpha \frac{q}{r^2} dr = \frac{\alpha q}{r}.$$

Отметим следующие важнейшие свойства потенциала.

1. *Потенциал поля—скалярная величина.* Поэтому если имеется несколько зарядов, то результирующий потенциал некоторой точки поля по принципу суперпозиции полей равен алгебраической сумме всех потенциалов, которые создают в данной точке все заряды по отдельности. В самом деле, при наличии, например, двух зарядов кулоновская сила, действующая на пробный заряд, равна векторной сумме сил, действующих на такой заряд со стороны двух других.

Со скалярными величинами обращаться при различных вычислениях значительно проще, чем с векторными. Поэтому при расчетах характеристик электростатических полей обычно пользуются именно потенциалом.

2. *Потенциал  $\varphi = \varphi(x, y, z)$ —функция местоположения точки в пространстве по отношению к зарядам, которая может менять свою величину и знак при переходе от одной области пространства к другой.* Нетрудно видеть, что при переходе из одной точки в другую совершается работа, равная алгебраической разности потенциалов в этих двух точках.

В любом электростатическом поле можно выделить такие поверхности, при движении вдоль которых электростатические силы вообще не совершают работы. Например, для одиночного точечного заряда такой поверхностью будет любая сфера, окружающая заряд так, что он находится в ее центре. В случае равномерно заряженного длинного проводника аналогичной поверхностью будет боковая поверхность цилиндра.

В случае плоскости такой поверхностью является любая плоскость, проведенная параллельно заряженной плоскости на малом расстоянии от нее.

Обладающие таким свойством поверхности называются *эквипотенциальными* или *поверхностями равного потенциала*. С помощью таких поверхностей или линий, если в пространстве

выделяется какая-нибудь плоскость и проводятся линии пересечения эквипотенциальных поверхностей с выбранной плоскостью, удобно изображать (достигается большая наглядность) сложные электрические поля, окружающие, например, молекулы (см. рис. 7.16, 7.17).

3. Если необходимо в явном виде найти векторы напряженности электрического поля для разных точек пространства вокруг зарядов, то это легко сделать, пользуясь понятием потенциала, по аналогии с вычислением компонент силы по заданной потенциальной функции. Имеем:

$$E_x = -\frac{\partial \varphi}{\partial x}, \quad E_y = -\frac{\partial \varphi}{\partial y}, \quad E_z = -\frac{\partial \varphi}{\partial z}.$$

Заряды на проводниках. Емкость. Рассмотрим некоторый проводник. Под *проводником* (типичными проводниками являются металлы) мы понимаем такое тело, в котором имеются свободные заряды (в металлах—электроны), которые могут легко перемещаться под действием электрического поля.

Если на поверхность проводника нанести электрический заряд, то под действием кулоновских сил отталкивания эти заряды будут определенным образом распределяться и довольно быстро установится стационарная картина. При этом весь избыточный заряд сосредоточится в очень тонком слое на поверхности проводника. Утверждение, что через малый промежуток времени распределение заряда действительно будет стационарным, следует из того экспериментального факта, что в заряженном проводнике отсутствуют электрические (как поверхностные, так и объемные) токи, т. е. потоки движущихся зарядов. Как мы увидим ниже, такие движущиеся заряды создают магнитные поля. Отсутствие магнитных полей в пространстве около статически заряженных проводников и является признаком отсутствия токов. В свою очередь, это означает, что напряженность электрического поля внутри заряженных проводников всюду должна быть равна нулю, а напряженность электрического поля у поверхности заряженного проводника определяется векторами, нормальными к поверхности проводника в любых точках этой поверхности. В противном случае появились бы составляющие напряженности электрического поля, направленные вдоль поверхности проводника. Под действием этих составляющих заряды начали бы перемещаться вдоль поверхности проводника и возник бы поверхностный ток. Если напряженность электрического поля во всех точках проводника направлена перпендикулярно его поверхности, то перемещение любого заряда по поверхности проводника не связано с работой. Другими словами, разность потенциалов между двумя любыми точками проводника равна нулю. Следовательно, поверхность заряженного проводника является эквипотенциальной.

При последовательном переносе некоторого заряда из бесконечности на проводник затрачивается вполне определенная

работа. Полная работа определяется в конечном счете количеством заряда, распределенного на поверхности проводника, и потенциалом этой поверхности. Между потенциалом  $\varphi$  поверхности и количеством заряда  $q$ , нанесенным на эту поверхность, существует следующее соотношение:

$$q = C\varphi.$$

Коэффициент пропорциональности  $C$  называют *емкостью проводника*. Такое простое соотношение получается потому, что поверхность проводника является эквипотенциальной, и безразлично, в какую точку этой поверхности переносится дополнительная порция заряда.

Рассмотрим заряженную проводящую сферу. Заряд распределяется по поверхности ее равномерно. Потенциал в точке на расстоянии  $r$  от сферы равен

$$\varphi(r) = aq/r.$$

Потенциал поверхности сферы:

$$\varphi = aq/r_0,$$

где  $r_0$ —радиус сферы. С другой стороны, потенциал поверхности, если эта сфера—проводник, равен

$$\varphi = \frac{1}{C} q \quad \text{или} \quad C = a r_0.$$

Следовательно, емкость в данном случае с точностью до постоянной равна радиусу сферы. Вообще, емкость связана с геометрией проводника и является его функцией.

Здесь мы сделаем небольшое отступление. Интересно выяснить, зависит ли поверхностная плотность зарядов на проводнике и напряженность поля у поверхности от ее формы. Для решения этого вопроса будем рассуждать следующим образом. Выделим на поверхности проводника некоторую малую площадь (элемент площади). Количество заряда на этом элементе равно  $q = \sigma|\Delta S|$ , где  $\sigma$ —поверхностная плотность заряда. Заклучим этот элемент площади в ящик (рис. 7.27). Так как в любом месте поверхности напряженность поля направлена перпендикулярно поверхности и отсутствует внутри проводника, то поток вектора  $E$  сквозь все стороны ящика равен

$$\oint_S E dS = |E| |\Delta S| = \frac{1}{\epsilon_0} \sigma |\Delta S|.$$

Таким образом,  $|E| = \frac{1}{\epsilon_0} \sigma$ , т. е. модуль вектора напряжен-

ности электрического поля около поверхности проводника пропорционален плотности поверхностного заряда.

Теперь рассмотрим некоторый проводник произвольной формы, поверхность которого замкнута (рис. 7.28). Небольшой

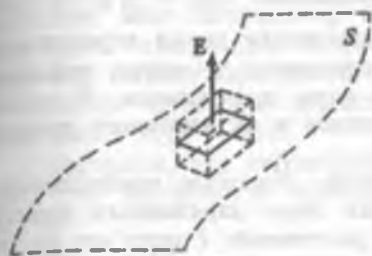


Рис. 7.27. К выводу формулы для напряженности поля у поверхности заряженного проводника



Рис. 7.28. Участки поверхности, описываемые сферой

выпуклый участок такой поверхности с хорошей точностью можно описать прилегающей сферой. Участок выберем так, чтобы в пределах его можно было считать, что  $\sigma = \text{const}$ . Тогда, очевидно, напряженность поля около этого участка по модулю и направлению будет совпадать с напряженностью поля, создаваемого в той же области пространства равномерно заряженной сферой радиуса  $R$  с той же плотностью заряда. Будут одинаковыми и потенциалы поверхности проводника и гипотетической сферы. Только что мы показали, что потенциал заряженной сферы равен  $\varphi = \alpha Q/R$ , где  $Q$ —полный заряд сферы. Для сферы плотность его всюду одинакова. Поэтому

$$\varphi = \alpha(1/R)4\pi\sigma R^2 = 4\pi\alpha\sigma R = \frac{\sigma R}{\epsilon_0}$$

( $S = 4\pi R^2$ —площадь поверхности сферы). В то же время потенциал всех точек поверхности любого заряженного проводника один и тот же. Значит, для выпуклых участков поверхности заряженного проводника выполняются соотношения  $R\sigma = \text{const}$  и  $R|E| = \text{const}$ .

Таким образом, плотность заряда и модуль вектора  $|E|$  напряженности электрического поля должны возрасти в местах сильных искривлений поверхности проводника, в частности в окрестности острых выступов. При зарядке проводников именно с острия и начинается пробой и образование искрового разряда в пространстве вокруг проводника.

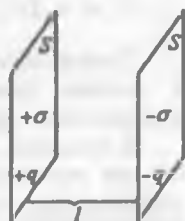


Рис. 7.29. Плоский конденсатор

Перейдем теперь к двум металлическим телам, образующим конденсатор, т. е. систему, у которой одно тело заряжено положительным зарядом  $+q$ , а другое таким же отрицательным (например, две металлические пластины; рис. 7.29). Между этими телами возникает разность потенциалов  $\Delta\varphi$ .



Имеет место следующее соотношение:  $\Delta\varphi = \frac{1}{C} q$ . В этом равенстве  $\Delta\varphi$ —разность потенциалов между двумя заряженными телами, образующими данный конденсатор;  $q$ —заряд, помещенный на одно из тел;  $C$ —емкость пары проводников. И в этом случае емкость зависит от проводников и их взаимного расположения.

Рассмотрим плоский конденсатор. *Плоским конденсатором* называется система, состоящая из двух заряженных пластин площадью  $S$ , расположенных на расстоянии  $l$  одна от другой. Заряды на пластинах равны по модулю и противоположны по знаку.

Найдем емкость  $C$  плоского конденсатора. Между двумя заряженными пластинами разность потенциалов равна

$$\Delta\varphi = \frac{1}{\epsilon_0} \sigma l,$$

где  $\sigma$ —плотность заряда;  $l$ —расстояние между двумя пластинами (толщина конденсатора). В самом деле, работа по перемещению единичного заряда от одной пластины к другой равна

$$A = \sum_l E_l dl = |E|l = \frac{1}{\epsilon_0} \sigma l.$$

Если заряд на пластине равен  $q$ , а площадь каждой пластины  $S$ , то

$$A = \Delta\varphi = \frac{1}{\epsilon_0} \sigma l = \frac{1}{\epsilon_0} lq/S,$$

где  $q/S = \sigma$ .

Сравнивая это выражение с выражением  $\Delta\varphi = q/C$ , находим

$$C = \epsilon_0 S/l.$$

Таким образом, емкость плоского конденсатора увеличивается с ростом площади пластины конденсатора и уменьшается с увеличением расстояния между пластинами.

В технике, особенно радиотехнике, конденсаторы используются для накопления зарядов. Повышать количество запасенного в конденсаторе заряда только за счет увеличения разности потенциалов нельзя, так как произойдет пробой конденсатора. Поэтому увеличение количества запасенного заряда обычно достигается за счет увеличения емкости  $C$  конденсатора. В частности, емкость плоского конденсатора  $C = \epsilon_0 S/l$  возрастает за счет увеличения площади  $S$  пластин. Это можно сделать, если взять, например, две длинные ленты металлической фольги, поместить между ними ленту провощенной бумаги и свернуть всю систему в трубку. Такой бумажный конденсатор часто используется в технике.

Найдем работу, которую надо произвести, чтобы зарядить проводник или конденсатор от нуля до потенциала  $\varphi$ . Будем рассуждать следующим образом. Пусть имеется проводник, например шарик, заряженный до потенциала  $\varphi$ . Этот потенциал  $\varphi$  является функцией заряда  $q$ , который уже находится на проводнике:  $\varphi(q) = q/C$ .

Представим, что небольшой заряд  $dq$  перенесли из бесконечности на поверхность проводника. В этом случае совершена работа

$$dA = \varphi(q)dq.$$

Далее,  $\varphi(q) = q/C$ , поэтому

$$dA = \varphi(q)dq = \frac{q}{C} dq.$$

Полная работа по зарядке проводника от 0 до  $Q$  равна

$$A = \int_0^Q \frac{q}{C} dq = \frac{1}{2} \frac{Q^2}{C}.$$

Умножим и разделим это выражение на  $C$ :

$$A = \frac{1}{2} \frac{Q^2 C}{C^2},$$

но  $Q/C$ —это конечный потенциал проводника. Обозначим его  $\Phi$ . Тогда

$$A = \frac{1}{2} \Phi^2 C.$$

Для конденсатора работа по его зарядке от нуля до заряда  $Q$  или разности потенциалов  $\Delta\Phi$  равна

$$A = \frac{1}{2} \Delta\Phi^2 C.$$

Для плоского конденсатора

$$\Delta\Phi = \frac{1}{\epsilon_0} \sigma l = El \quad \text{и} \quad C = \epsilon_0 S/l,$$

где  $E$ —напряженность поля между пластинами конденсатора. Поэтому

$$A = \frac{\epsilon_0}{2} E^2 l^2 \frac{S}{l} = \frac{\epsilon_0}{2} E^2 lS,$$

где  $lS$ —объем пространства между пластинами плоского конденсатора. Поэтому величину  $A/(lS) = E^2 \epsilon_0 / 2$  можно назвать удельной работой зарядки конденсатора или энергией электростатического поля напряженностью  $E$  в единице объема.

Обозначая эту величину  $W$ , запишем  $W = \frac{1}{2} \epsilon_0 E^2$  (плотность энергии).

Мы получили это выражение для плоского конденсатора. Однако можно показать, что оно справедливо и для конденсатора произвольной формы. Поэтому это выражение является общим для энергии электростатического поля.

Если внести в любую точку пространства между двумя обкладками конденсатора единичный заряд, то на него будет действовать сила. Это значит, что в этом месте пространства существует запас потенциальной энергии. Поэтому мы говорим, что электростатическое поле в любой малой области пространства, где оно существует, имеет вполне определенный запас энергии

$$W = \frac{\epsilon_0}{2} |E|^2.$$

Полная энергия поля в пространстве, ограниченном замкнутой поверхностью  $S$ , равна

$$W = \frac{\epsilon_0}{2} \iiint |E|^2(x, y, z) dx dy dz.$$

Интегрирование ведется по всему объему внутри некоторой замкнутой поверхности.

То, что электростатическое поле действительно обладает запасом энергии, легко обнаружить с помощью следующего опыта. Зарядим плоский конденсатор от электрофорной машины. Присоединим к пластинам конденсатора контакты электроскопа. При этом стрелка электроскопа отклонится на величину, пропорциональную разности потенциалов между обкладками конденсатора. Внесем между прокладками конденсатора лист диэлектрика, например стекла. В таком диэлектрике электроны не могут свободно мигрировать по всему объему, а лишь меняют свое положение около атомов, если на диэлектрик наложено внешнее электрическое поле.

Если в начальный момент была симметричная система в том смысле, что электроны равномерно окружали ядро так, что центр тяжести положительного и отрицательного зарядов находился в центре ядра, то после наложения электрического поля электроны начнут смещаться в одну сторону и в результате центры тяжести положительного и отрицательного зарядов не совпадут между собой. Система будет нейтральной в целом, но в этом случае она уже приобретает свойство диполя, т. е. системы, у которой заряды (один отрицательный, другой положительный) расположены на малом расстоянии друг от друга.

Чтобы разделить на небольшом расстоянии положительный и отрицательный заряды, нужно затратить энергию.

Вернемся теперь к нашему опыту с электроскопом. Если внести между обкладками конденсатора лист стекла (типичный

диэлектрик), то отклонение стрелки электроскопа уменьшится. Мы вправе трактовать этот результат как уменьшение разности потенциалов между обкладками конденсатора и, следовательно, как уменьшение средней напряженности поля ( $\Delta\varphi = |E|l$ ) внутри конденсатора и средней энергии поля, которая пропорциональна  $|E|$ .

Действительно, если часть энергии затрачивается на поляризацию, то напряженность электрического поля должна быть меньше при введении диэлектрика. При не очень большой напряженности поля можно считать, что напряженность поля в случае введения диэлектрика равна напряженности  $E$  поля в вакууме, деленной на величину  $\epsilon > 1$ , которая называется диэлектрической проницаемостью вещества:

$$|E| = \frac{1}{\epsilon} |E|_{\text{вак}}.$$

При небольшом поле можно считать, что  $\epsilon$  не зависит от напряженности поля и является характеристикой только данного тела.

Величину  $\epsilon$  можно связать с микроскопическими свойствами тела. Так, если тело состоит из одинаковых молекул, обладающих значительными дипольными моментами, то

$$\epsilon = 1 + \frac{4}{3} \pi n_0 \frac{p^2}{kT},$$

где  $n_0$ —концентрация молекул;  $p$ —дипольный момент молекулы;  $k$ —постоянная Больцмана;  $T$ —термодинамическая температура тела.

Таким образом, измеряя  $\epsilon$ , можно получить некоторые сведения о внутреннем строении вещества.

Для конденсатора  $\Delta\varphi = |E|l = q/C$ . Так как при введении в конденсатор диэлектрика величина заряда не может измениться, то  $C_{\text{вак}} |E|_{\text{вак}} l = q = C_{\text{вак}} \epsilon |E| l$ . Значит, емкость конденсатора с диэлектриком  $C = C_{\text{вак}} \epsilon$ .

Вспомним теперь (см. школьный курс физики), что частота электрического колебательного контура связана с емкостью конденсатора соотношением  $\omega \sim 1/\sqrt{C}$ . При малых изменениях  $C$  получим

$$\Delta\omega \sim -\frac{1}{C\sqrt{C}} \Delta C = -\frac{1}{C\sqrt{C}} C_{\text{вак}} \Delta\epsilon = -\frac{1}{C\sqrt{C}} C_{\text{вак}} \frac{4}{3} \pi \frac{p^2}{kT} \Delta n_0,$$

где  $\Delta n_0$ —изменение концентрации молекул.

Таким образом, наблюдая расстройку колебательного контура, можно решать задачи количественного анализа в аналитической химии.

Меняется  $\varepsilon$  и во время хода химических реакций, что можно заметить, если кювета с реагирующими веществами находится внутри конденсатора. В этом случае могут меняться как  $n_0$ , так и  $\rho$ . Такой прием используется в схемах высокочастотного титрования.

### 7.3. Постоянный ток

Постоянный ток и первое правило Кирхгофа. Рассмотрим ряд вопросов, связанных с постоянными токами.

Назовем *током проводимости* поток заряда, пронесимого через единицу поверхности точечными электрическими зарядами. Число таких зарядов, переносимых в единицу времени через нормально расположенную к направлению их движения единичную поверхность, равно  $n_0|v|$ , где  $n_0$ —концентрация зарядов;  $|v|$ —модуль их средней направленной скорости (скорость дрейфа). Если элемент поверхности характеризовать, как и ранее, вектором  $\Delta S$  и считать, что вектор скорости дрейфа  $v$  зарядов не обязательно направлен нормально поверхности, то для числа пересекающих поверхность частиц получим  $\Delta l = n_0 v \Delta S$ . Если  $q$ —заряд каждой частицы, то переносимый в единицу времени через элемент площади полный заряд или количество электричества (элементы силы тока) равен  $I = n_0 q v \Delta S$ . Произведение  $j = n_0 q v$  получило название *вектора плотности тока*.

*Постоянным током* мы будем называть такой поток заряженных частиц, при котором в любой момент времени через плоскость проходит с одной и той же скоростью одно и то же количество одинаковых по величине зарядов.

Потоком вектора  $j = n_0 q v$  через поверхность  $S$  называется сумма

$$I = \sum_i j_i \Delta S_i$$

где  $\Delta S_i$ —элементы площади поверхности.

При малых  $\Delta S_i$  эта сумма превращается в интеграл  $\int_S j dS = I$ .

Поток вектора сквозь поверхность характеризует общее количество заряда, пронесимого через эту поверхность в единицу времени. Интеграл  $\int_S j dS$  называют также *силой тока* через поверхность  $S$ .

Если имеется некоторый источник зарядов, то для потока их через замкнутую поверхность, окружающую источник, получим

$$\oint_S \mathbf{j} dS = -dQ/dt.$$

Здесь  $Q$ —полный заряд внутри поверхности. Это соотношение достаточно очевидно и выражает тот факт, что если какие-то заряды выносятся через поверхность  $S$  наружу, то при этом должен уменьшаться запас электричества внутри поверхности.

Если рассмотреть замкнутую поверхность и такое движение зарядов, при котором внутри замкнутой поверхности заряд не накапливается, то общее число зарядов, вошедших внутрь замкнутой поверхности, равно числу зарядов, выходящих через поверхность. Если условимся направление вектора  $\mathbf{j}$  выбирать таким, чтобы оно характеризовало движение зарядов, то в тех случаях, когда заряд выходит через поверхность, мы получим одно значение потока. В тех случаях, когда заряд входит через поверхность, мы получим другое, противоположное ему по знаку значение потока вектора через поверхность. Если число зарядов, входящих и выходящих через поверхность, одинаково, то одним участкам поверхности будет соответствовать положительное значение потока, а другим участкам поверхности—отрицательное значение потока. Полный поток вектора сквозь замкнутую поверхность тогда равен нулю:

$$\oint_S \mathbf{j} dS = 0.$$

Рассмотрим несколько проводников, соединенных в узел (рис. 7.30). Если окружить узел сферой, то полный поток вектора  $\mathbf{j}$  через поверхность этой сферы равен

$$\oint_S \mathbf{j} dS = \int_{S_1} \mathbf{j}_1 dS + \int_{S_2} \mathbf{j}_2 dS + \int_{S_3} \mathbf{j}_3 dS + \int_{S_4} \mathbf{j}_4 dS.$$

Здесь  $S_1, S_2, S_3, S_4$ —площади сечения проводников, а  $\mathbf{j}_1, \mathbf{j}_2, \mathbf{j}_3, \mathbf{j}_4$ —плотности токов в проводниках.

Так как, по определению,  $\int_{S_1} \mathbf{j}_1 dS = I_1$ —сила тока в проводнике, то, учитывая, что  $\oint_S \mathbf{j} dS = 0$ , получаем

$$I_1 + I_2 - I_3 - I_4 = 0.$$

Другими словами, алгебраическая сумма всех постоянных токов, протекающих через узел, равна нулю. Выведенное соотношение называется *первым правилом Кирхгофа для постоянных токов*. При вычислении этой суммы токи, текущие к узлу, надо брать с плюсом, а токи, текущие от узла, с минусом.



Рис. 7.30. К выводу первого правила Кирхгофа

Рассмотрим некоторые важные соотношения для плотности тока внутри проводника. Для того чтобы внутри проводника свободные электрические заряды пришли в движение и возник электрический ток, необходимо создать неравновесное распределение заряда вдоль поверхности проводника. В самом деле, если на поверхность проводника внести электрический заряд, например избыток электронов, то в первый момент времени все остальные электроны, находящиеся внутри проводника, начнут отталкиваться от этих излишних электронов. Возникает движение электронов, т. е. ток, которое будет продолжаться до тех пор, пока весь электрический заряд не распределится так, чтобы силы взаимного отталкивания электронов компенсировали друг друга. Излишний заряд в этом случае весь будет сосредоточен на поверхности проводника, напряженность поля внутри проводника равна нулю.

При наличии тока такая стационарная картина отсутствует: внутри проводника есть электрическое поле и наблюдаются движущиеся заряды. Плотность тока в этом случае зависит от напряженности электрического поля внутри проводника и обратно пропорциональна некоторому коэффициенту  $\rho$ , который зависит только от свойств проводника и называется его *удельным сопротивлением*:

$$j = \frac{1}{\rho} E_{\text{внутр}}.$$

Эту запись принято называть *дифференциальной записью закона Ома*. Выясним, при каких условиях закон Ома соблюдается и какой смысл должна иметь величина  $\rho$ . Перейдем для этого к рассмотрению микроскопической картины. Электроны в проводнике можно рассматривать как некий электронный газ (этим термином часто и пользуются). Согласно общим кинетическим представлениям частицы такого газа должны находиться в состоянии хаотических движений. Направленное движение, которое и воспринимается как электрический ток, возникает, если воздействовать на такую систему направленной силой. Этой силой и является напряженность  $E$  электрического поля. Под действием этой силы каждый из свободных зарядов приобретает ускорение, определяемое законом Ньютона:  $m \frac{dv}{dt} = qE$ . Если бы заряды двигались совсем свободно, то их скорость возрастала бы со временем и простого соотношения  $j = \frac{1}{\rho} E$  не могло бы быть.

На самом деле электроны не могут двигаться свободно: они будут сталкиваться между собой и с атомами кристаллической решетки. Если считать, что электрон начинает двигаться из состояния покоя и после удара об атом останавливается (довольно грубая, но в общем-то близкая к реальности картина), и принять, что до соударения прошло время  $\Delta t$ , то максимальная скорость, которую может приобрести электрон, будет равна

$v = \frac{q}{m} E \Delta\tau$ . После этого электрон остановится и будет затем разгоняться в направлении  $E$  снова с нуля. Таким образом, электрон набирает скорость не непрерывно, как должно было бы быть при отсутствии соударений, а порциями, стартуя каждый раз с нуля. Поэтому, несмотря на длительное действие силы  $F = qE$ , скорость дрейфа остается постоянной и равной  $v = (q/m)E\Delta\tau$ , где  $\Delta\tau$ —эффективное время между соударениями. В свою очередь,  $\Delta\tau \sim 1/(N_0 r_0^2)$ , где  $N_0$ —концентрация центров столкновения;  $r_0$ —эффективный радиус столкновений (сечение рассеяния электронов). Вектор плотности тока равен

$$j = n_0 q v = n_0 \frac{q^2}{m} \frac{1}{N_0 r_0^2} E,$$

или

$$\frac{1}{\rho} = n_0 \frac{q^2}{m N_0 r_0^2}.$$

Для электронов отношение  $q^2/m$  есть постоянная величина. Значение  $\rho$  должно меняться при переходе от одного вещества к другому в зависимости от  $n_0$ , т. е. от плотности подвижных носителей зарядов и произведения  $N_0 r_0^2$ . И то и другое является свойством данного проводника.

Электронная проводимость типична для металлов. Однако, например, в проводящих полимерах и жидких электролитах подвижными носителями зарядов могут являться и другие частицы (ионы). Для них отношение  $q^2/m$  может очень сильно отличаться от «электронного». Из-за малой величины  $\Delta\tau$  скорость дрейфа даже электронов в металлах при прохождении тока оказывается очень малой (порядка  $10^{-2}$  см/с).

Рассмотрим теперь некоторый участок проводника, причем будем полагать, что напряженность  $E_{\text{внутр}}$  поля на всем протяжении участка одна и та же, а сечение участка равно  $S$ . Считаем также, что вектор  $E_{\text{внутр}}$  направлен по оси проводника. Если  $\Delta l$ —длина участка проводника, а  $\Delta\varphi$ —разность потенциалов поля начала и конца участка, то  $\Delta\varphi = |E_{\text{внутр}}| \Delta l$  и, значит,

$$|j| = \frac{1}{\rho} |E_{\text{внутр}}| = \frac{1}{\rho} \frac{\Delta\varphi}{\Delta l}.$$



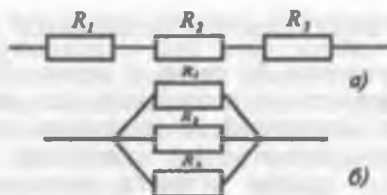


Рис. 7.3. 1. Последовательное (а) и параллельное (б) соединения проводников

Если для простоты считать, что плотность тока внутри проводника по всему сечению одна и та же и проводник однородный, то можно написать:

$$I = \int_S j dS = \int_S \Delta\varphi \frac{1}{\rho \Delta l} dS = \Delta\varphi \frac{1}{\rho} \frac{S}{\Delta l} = \Delta\varphi \frac{1}{R}.$$

Полученное соотношение называется *законом Ома для участка цепи*. Величина  $R = \rho \Delta l / S$  называется *сопротивлением участка проводника*. Оно растет при увеличении длины участка и падает с увеличением площади сечения. Разность потенциалов  $\Delta\varphi$  принято называть *разностью напряжений* или *падением напряжения* на участке проводника.

Формула  $R = \rho \Delta l / S$  позволяет легко получить известное соотношение для сопротивления последовательно и параллельно соединенных проводников (рис. 7.31).

Для первого случая имеем

$$R_{\text{общ}} = \rho_1 \frac{l_1}{S_1} + \rho_2 \frac{l_2}{S_2} + \rho_3 \frac{l_3}{S_3} = R_1 + R_2 + R_3.$$

Во втором случае все три проводника будут действовать так же, как скрученный втрое проводник. Найдем

$$\frac{1}{R_{\text{общ}}} = \left( \frac{S_1}{\rho_1 l_1} + \frac{S_2}{\rho_2 l_2} + \frac{S_3}{\rho_3 l_3} \right) = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}.$$

Почему по проводнику идет ток? Электродвижущие силы. Второе правило Кирхгофа. Чтобы получить ответ на этот далеко не тривиальный вопрос, будем исходить из ряда экспериментальных фактов, хорошо известных из школьного курса физики. Мы знаем, что в проводнике, концы которого подключены к гальваническому элементу, возникает и длительно поддерживается неизменным электрический ток. Если проводник металлический, то ток может быть связан только с движением электронов, так как положительные заряды (ядра атомов) «закреплены» на своих местах, образуя жесткую структуру металла.

Если проводник однородный, то нигде внутри проводника заряды не могут накапливаться. Некоторый лишний заряд может образоваться только на поверхности проводника.

Движение зарядов может совершаться как по поверхности проводника, так и внутри проводника. Легко убедиться, что если заменить, например, цилиндрический проводник очень тонкой металлической трубкой такого же диаметра, то при прочих равных условиях (если подключить и цилиндрический проводник, и трубку к одному и тому же источнику тока) сила тока в трубке уменьшится в случае очень тонких стенок трубки практически до нуля. Так как замена сплошного проводника трубкой не может отразиться на прохождении поверхностного тока, то не этот ток, а ток, связанный с дрейфом электронов внутри проводника, регистрируется, когда происходит передача электрической энергии вдоль по цепи, соединяющей полюсы источника тока.

Для того чтобы при этом условии существовал непрерывный перенос заряда по замкнутому контуру, внутри проводника должно быть некоторое поле сил, действующих на способные к перемещению заряды. Это поле формально может характеризоваться вектором напряженности, постоянным в любом сечении проводника (если проводник имеет цилиндрическую форму) и направленным вдоль его оси. Так как ток проходит и через гальванический элемент, то и на участке этого элемента можно говорить о наличии такого поля сил. Такие силы получили название *сторонних*.

Как уже указывалось, движение зарядов (электронов) по цепи не является свободным. В самом деле, прохождение тока по проводнику сопровождается его нагреванием, что связано с передачей части энергии направленного дрейфа электронов ядрам атомов вследствие столкновений. Следовательно, перенос заряда по контуру требует определенной работы. Это означает, что для сторонних электрических сил циркуляция вектора  $E_{\text{стор}}$  по замкнутому контуру, выражаемая интегралом  $\oint_L E_{\text{стор}} d\mathbf{l}$ , должна быть отлична от нуля:

$$\oint_L E_{\text{стор}} d\mathbf{l} \neq 0.$$

Напомним, что для вектора напряженности электрического поля в электростатике  $\oint_L E_{\text{кул}} d\mathbf{l} = 0$ .

Попробуем разобраться в процессах, которые приводят к появлению не равного нулю поля внутри проводника. Для этого рассмотрим прямой цилиндрический проводник радиуса  $r$ . Представим, что на левый конец проводника от постороннего источника перенесен отрицательный заряд в виде совокупности электронов. Так как электроны в проводнике обладают достаточной подвижностью и отталкиваются друг от друга, то они практически мгновенно равномерно распределятся по поверхности проводника. Пока картина не установится, будет наблюдаться,

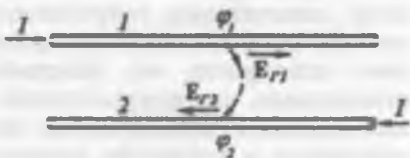
очевидно, поверхностный электрический ток. Время перехода к равновесному состоянию мало.

После установления равновесного состояния распределенные по поверхности статические заряды будут создавать вокруг проводника статическое электрическое поле, линии напряженности которого всюду перпендикулярны (нормальны) поверхности проводника. Внутри проводника поле будет отсутствовать. Такая картина электрического поля возникает как результат совокупного действия большого числа зарядов. В самом деле, любой малый участок поверхностного заряда можно рассматривать как точечный заряд. Такой заряд должен создавать электрическое поле как вне проводника, так и внутри него. Однако в сумме все поля таких точечных зарядов, на которые можно разбить общий поверхностный заряд, по принципу суперпозиции компенсируют друг у друга отдельные составляющие направленных внутрь проводника полей, приводя к общему нулевому внутреннему полю. Важно отметить, что необходимое для такой компенсации распределение электронной плотности на поверхности устанавливается для любого проводника автоматически в зависимости от его формы.

Если изменить отвечающее этим условиям распределение электронной плотности, то внутри проводника появится отличное от нуля электрическое поле. Это можно сделать, если с правого конца цилиндрического заряженного проводника удалить часть поверхностного отрицательного заряда. В первый момент распределение отрицательного заряда на поверхности проводника станет неравномерным: слева его больше, а справа меньше. Возникнет градиент поверхностного заряда и разность потенциалов вдоль поверхности. Если обеспечить постоянное поступление зарядов к одному концу проводника и в то же время обеспечить уход зарядов с другого конца, то картина неравномерного распределения заряда сохраняется во времени.

Такой, хотя уже и неравновесный, но постоянный по времени распределенный по поверхности заряд будет снова действовать как статический и создавать в пространстве постоянное электрическое поле, причем с отличной от нуля напряженностью внутри проводника. Появится и еще одна особенность: так как неравномерно распределенный заряд создается движущимися электронами (постоянная во времени картина распределения поверхностной электронной плотности получается потому, что уходящие с определенного места электроны сразу замещаются в том же количестве другими), то должен существовать и постоянный поверхностный ток. Это означает, что наряду с перпендикулярной (как в статике) к поверхности проводника составляющей вектора напряженности электрического поля существует и горизонтальная (параллельная поверхности) составляющая. Для цилиндрического проводника такая дополнительная составляющая направлена вдоль оси проводника. Составляющие поверхностной плотности от разных участков не компенсируют друг друга, что

Рис. 7.32. Форма линии напряженности электрического поля между двумя входящими в одну цепь параллельными проводами, по которым протекают противоположно направленные токи ( $\varphi_1$  и  $\varphi_2$ —потенциалы проводов,  $E_{r1}$  и  $E_{r2}$ —горизонтальные составляющие поля у поверхностей проводников)



и приводит к общему направленному вдоль оси проводника внутреннему электрическому полю.

Изменится поле и вне проводника: наличие горизонтальной составляющей скажется в том, что линии напряженности уже будут не перпендикулярны к поверхности проводника, а иметь некоторый наклон. Это подтверждается экспериментально. Например (рис. 7.32), линии напряженности электрического поля между двумя длинными параллельными входящими в одну и ту же цепь с током проводами вдали от концов этих проводов имеют вид изогнутых кривых. В такой форме линий напряженности легко убедиться, если учесть, что линия направлена от области с большим потенциалом к области с меньшим, т. е. от одного провода к другому, и что токи в проводах текут в разные стороны, а значит, в разные стороны направлены и горизонтальные составляющие вектора  $E$  на оси проводов.

Внутри прямых проводников отсутствует составляющая напряженности электрического поля, перпендикулярная оси проводника. Пусть проводник искривляется. Ток должен следовать за искривлением. Для этого необходимо, чтобы внутреннее поле приобрело компоненту, направленную по радиусу искривления. Это реализуется, если поверхностная плотность электронов оказывается разной на внешней и внутренней сторонах кольцевого участка провода.

Здесь надо сделать дополнительные пояснения. Несовпадающее с чисто статическим, однако не меняющееся со временем неравновесное и неоднородное по длине проводника распределение поверхностного заряда будет заметным лишь при коротких проводниках и в местах резких изгибов или неоднородностей. В длинном прямом однородном цилиндрическом проводнике неодинаковый по плотности вдоль поверхности в направлении оси проводника поверхностный заряд не возникает. Однако внутреннее поле все равно существует. Происходит это потому, что поверхностные заряды не неподвижны и имеется постоянный поверхностный ток. Ток этот поддерживается потому, что «наильно», против сил кулоновского отталкивания, нагнетаемые с одного конца электроны (поскольку соответствующие силы действуют против сил кулоновского отталкивания, то они кулоновскими быть не могут, почему и получили название сторонних) будут отталкивать близлежащие на проводнике отрицательные заряды, которые будут воздействовать на последующие и т. д. С другого конца провода электроны снимаются. В

результате появляются «свободные места». Они сразу же занимают электронами прилегающих участков провода, испытывающими отталкивание электронами предыдущих. Общее количество электронов на поверхности прямого провода остается неизменным, но поверхностные электроны в результате близкого действующего кулоновского взаимодействия все время будут находиться в состоянии направленного движения. Теперь учтем, что и поверхностные электроны испытывают сопротивление своему движению, хотя и меньшее, чем внутренние. Согласно законам механики их движение возможно только, если вдоль поверхности действует постоянная нескомпенсированная сила — напряженность электрического поля. Это поле не может затухнуть сразу за пределами поверхности и проникает внутрь проводника, порождая ток дрейфа. Таким образом, движущиеся поверхностные электроны как бы увлекают внутренние за собой. Вся картина, хотя и связанная с движением, во времени не меняется. Именно поэтому обсуждаемое электрическое поле является, подобно обычному кулоновскому, потенциальным.

Таким образом, в зависимости от формы проводника в нем для каждого случая создается неравновесное распределение поверхностных зарядов, которое, наряду с постоянным поверхностным током, и приводит к требуемой напряженности электрического поля внутри проводника.

Роль источника тока заключается в том, чтобы создать такие распределения, которые устанавливаются практически мгновенно, и поддерживать их неизменными, а также поддерживать постоянным поток зарядов во всей цепи. Этот поток связан с дрейфом электронов внутри проводника и ионов, возникающих в результате химической реакции, внутри гальванического элемента. Именно отвечающий дрейфу электронов ток и измеряется при подключении к внешнему участку цепи амперметра или по нагреванию проводника. Хотя скорость дрейфа электронов и мала, однако, так как дрейф возникает практически одновременно во всех участках проводника, это и воспринимается как мгновенное прохождение тока по проводнику.

Известный факт проводимости тонких полимерных пленок, зажатых между двумя пластинами плоского конденсатора, и отсутствие тока проводимости в нитях из того же материала наглядно показывают роль поверхностного распределения зарядов: хотя и в том, и в другом случаях в объеме полимера имеются свободные или почти свободные носители заряда, в первом эксперименте поле внутри полимера возникает за счет действия собственно конденсатора, а во втором — диэлектрическая поверхность полимера препятствует существованию поверхностного тока и образованию требуемого неравновесного зарядового распределения.

Рассмотрим замкнутую электрическую цепь (рис. 7.33) и выделим на ней участок 1—2, где действуют сторонние силы, возникающие в результате химической реакции в элементе, и



Рис. 7.33. Замкнутая электрическая цепь

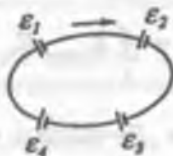


Рис. 7.34. Электрическая цепь с несколькими элементами

внешний участок 2—1, где проявляются только кулоновские силы, создаваемые неравновесным поверхностным зарядом. Заметим, что и на участке 1—2 действуют кулоновские силы, так как взаимодействия между зарядами по-прежнему существуют.

Для плотности тока в любых сечениях на этих участках можно записать:

$$j_{12} = \frac{1}{\rho_{12}} (E_{\text{стор}} + E_{12\text{Кул}}), \quad j_{21} = \frac{1}{\rho_{21}} E_{21\text{Кул}},$$

$$j_{12} \rho_{12} = E_{\text{стор}} + E_{12\text{Кул}}, \quad j_{21} \rho_{21} = E_{21\text{Кул}}.$$

Умножая эти равенства скалярно на  $d\mathbf{l}$  (векторы перемещения вдоль по цепи) и учитывая, что  $j_{21}, j_{12} \parallel d\mathbf{l}$ , так как ток течет вдоль провода и вдоль цепи направлены векторы  $E_{\text{Кул}}$  и  $E_{\text{стор}}$ , получаем

$$\rho_{12} |j_{12}| |d\mathbf{l}| = |E_{\text{стор}}| |d\mathbf{l}| + |E_{12\text{Кул}}| |d\mathbf{l}|,$$

$$\rho_{21} |j_{21}| |d\mathbf{l}| = |E_{21\text{Кул}}| |d\mathbf{l}|.$$

При равномерной плотности тока по всему сечению имеем, кроме того,  $|j| = I/S$ , где  $S$ —площадь сечения проводника. Так как заряд нигде не накапливается, то  $I = \text{const}$  ( $I$ —сила тока в цепи). Тогда

$$I \rho_{12} \frac{dl}{S_{12}} = E_{\text{стор}} dl + E_{12\text{Кул}} dl, \quad I \rho_{21} \frac{dl}{S_{21}} = E_{21\text{Кул}} dl.$$

Складывая эти равенства и интегрируя по всему замкнутому контуру (по всей цепи), записываем

$$I \oint_L \rho \frac{dl}{S} = \oint_L E_{\text{стор}} dl + \oint_L E_{\text{Кул}} dl.$$

Для кулоновского поля  $\oint_L E_{\text{Кул}} dl = 0$  (на участке 2—1 под действием поля электроны ускоряются, а на участке 1—2 поле тормозит их).

Окончательно

$$I \oint_L \rho \frac{dl}{S} = \oint_L E_{\text{стор}} dl,$$

где  $\oint_L \rho \frac{dl}{S} = R$ —полное сопротивление цепи (включая сопротивление источника тока). Поэтому можно записать  $R = R_{\text{внеш}} + R_{\text{внутр}}$ . Таким образом,

$$\oint_L E_{\text{стор}} dl = I(R_{\text{внеш}} + R_{\text{внутр}}),$$

где  $\oint_L E_{\text{стор}} dl = \mathcal{E}$ —работа сторонних электродвижущих сил. Следовательно,

$$\mathcal{E} = I(R_{\text{внеш}} + R_{\text{внутр}}).$$

Это соотношение называется *законом Ома для полной электрической цепи*;  $\mathcal{E}$ —работа, необходимая для перемещения единичного заряда по всему контуру, т. е. для поддержания постоянной силы тока в цепи. Принято называть величину  $\mathcal{E}$  *электродвижущей силой*.

Если электрическая цепь состоит из нескольких участков с батареями (рис. 7.34), то, рассуждая аналогично, получаем

$$\sum_i \mathcal{E}_i = I \left( \sum_i R_{\text{внеш}} + \sum_i R_{\text{внутр}} \right).$$

При этом  $\mathcal{E}_i$  берутся со знаком  $\leftrightarrow$  или  $\leftarrow$  в зависимости от расположения полюсов батарей при обходе контура в одном направлении. Так, например, в изображенном на рис. 7.34 контуре надо взять  $\mathcal{E}_1$  и  $\mathcal{E}_2$  с одним знаком, а  $\mathcal{E}_3$  и  $\mathcal{E}_4$ —с противоположным.

Последнее соотношение называется *вторым правилом Кирхгофа: алгебраическая сумма ЭДС в цепи пропорциональна полному сопротивлению цепи*. Коэффициент пропорциональности равен силе тока в цепи. При вычислении надо выбрать определенное направление обхода контура и в соответствии с этим выбирать знаки  $\mathcal{E}_i$ . Пользуясь правилами Кирхгофа, можно рассчитать сложную цепь, сводя задачу к решению системы уравнений.

## 7.4. Магнитостатика

Общие сведения. *Магнитостатикой* называется раздел учения об электромагнитных явлениях, изучающий свойства

постоянных магнитных полей, создаваемых постоянными токами, текущими по проводникам.

Основное утверждение магнитостатики состоит в том, что в пространстве вокруг постоянных токов создаются только постоянные магнитные поля.

Понятие магнитного поля возникло раньше, чем стала ясной связь между электрическими и магнитными явлениями (см § 7.1).

При изучении магнитостатических явлений можно пользоваться магнитной стрелкой. Известно, что магнитное поле отклоняет магнитную стрелку. Установившееся в магнитном поле направление стрелки принято за направление линии индукции В поля в данном месте пространства.

Многочисленные эксперименты по определению вида линий индукции показали, что это замкнутые линии без узлов. Напомним, что линии напряженности электростатических полей начинаются на зарядах одного знака и заканчиваются (либо уходят в бесконечность) на зарядах другого знака. В этом коренное отличие линий индукции магнитных полей от линий напряженности электростатических полей. Такое расположение линий индукции позволяет утверждать, что, во-первых,  $\oint_S \mathbf{B} d\mathbf{S} = 0$  и, во-вторых, что  $\oint_L \mathbf{B} d\mathbf{l} \neq 0$ . Первое равенство

интерпретируется как факт отсутствия магнитных зарядов (монополей), подобных электрическим. Второе утверждение следует из того, что всегда вдоль замкнутой линии индукции магнитные стрелки располагаются по правилу «голова к хвосту», т. е. всегда северный полюс одной стрелки «смотрит» в южный полюс следующей. Если считать, что вектор В поля направлен вдоль стрелки всегда от S к N, то  $\oint_L \mathbf{B} d\mathbf{l} \neq 0$  при обходе одной

линии индукции.

Экспериментально проверяется также и соотношение  $\oint_L \mathbf{B} d\mathbf{l} = \mu_0 I$ , если контур L—окружность и ток I проходит через центр окружности перпендикулярно ей. Здесь  $\mu_0$ —множитель размерности (магнитная проницаемость вакуума). При этом вектор В направлен по касательной к окружности. Обобщение результатов подобных экспериментов приводит к формуле

$$\oint_L \mathbf{B} d\mathbf{l} = \mu_0 \int_S \mathbf{j} d\mathbf{S},$$

где L—произвольный замкнутый контур; S—ограниченная им незамкнутая поверхность (закон Ампера).

Магнитное поле можно обнаружить и по его действию непосредственно на заряд. Посмотрим, как это можно в принципе сделать, ограничившись примером однородного поля. Представим, что в некоторую область пространства помещен единич-



ный пробный электрический заряд (заряженная материальная точка). Пусть этот заряд будет свободным в том смысле, что он не удерживается на месте никакими внешними механическими силами. В том случае, если заряд, первоначально неподвижный, продолжает покоиться, можно утверждать, что в рассматриваемой области пространства отсутствует электрическое поле. Однако из неподвижности заряда не следует, что отсутствует и магнитное поле. В самом деле, в § 7.1 мы отмечали, что магнитная компонента полной, действующей на заряд силы зависит от скорости движения заряда и поэтому не проявляется, если заряд не движется. Приведем заряд в движение. Измерим скорость и ускорение заряда, что, очевидно, всегда можно сделать независимым путем. Вернем заряд к исходному положению и вновь сообщим заряду импульс, но в другом направлении. Будем так поступать многократно, перебирая всевозможные направления импульсов. Если во всех случаях после начального воздействия заряд затем продолжает двигаться равномерно и прямолинейно, то мы вправе утверждать, что и магнитное поле в выбранной нами области пространства отсутствует. Однако может случиться, что заряд приобретет не только скорость, но и ускорение, зависящее от направления начальной скорости. Если возникнет ускорение, то это служит признаком появления направленной силы. Эту силу можно связать с существованием в пространстве, где производятся измерения, нового поля, которое не сводится к электростатическому. Это поле и получило название магнитного. Таким образом, если индикатором электрического поля может служить свободный первоначально покоящийся заряд, который под действием электрического поля приходит в движение, то индикатором наличия магнитного поля служит движущийся заряд, который под действием поля изменяет скорость и направление движения.

При многократном повторении эксперимента обнаружится, что имеется такое единственное направление, двигаясь вдоль которого пробный заряд не приобретает ускорения. Соответствующий вектор скорости обозначим  $v_0$ . Если в точку исходного положения заряда поместить центр магнитной стрелки, то она расположится вдоль такого выделенного направления движения. Естественно поэтому принять его в качестве направления вектора индукции В магнитного поля.

Если теперь проследить за направлением векторов ускорений, или, что то же самое, за направлениями действующей на заряд силы, то можно найти, что во всех случаях такие силы будут направлены перпендикулярно плоскостям, образуемым парами векторов: вектором скорости  $v$  и выбранного по указанному выше правилу вектором  $v_0$  или вектором индукции В магнитного поля. Выяснится также, что действующая на заряд сила пропорцио-

нальна синусу угла между этими двумя векторами и по модулю достигает максимума, если  $\mathbf{v} \perp \mathbf{v}_0$ .

На основании всех рассуждений можно определить индукцию магнитного поля в данной точке пространства:

$$\mathbf{B} = \frac{m}{qv_{\perp}^2} (\mathbf{a}_m \times \mathbf{v}_{\perp}).$$

Здесь  $m$ —масса материальной точки;  $q$ —ее заряд;  $\mathbf{a}_m$ —вектор, отвечающий максимальному ускорению при различных взаимных положениях  $\mathbf{v}$  и  $\mathbf{v}_0$ . Как уже указывалось, этот максимум достигается, если  $\mathbf{v} \perp \mathbf{v}_0$ . Может быть взят любой из таких векторов, так как ни направление вектора  $\mathbf{B}$ , ни его модуль от этого не изменятся. Это и приводит к выражению для магнитной компоненты силы, действующей на заряд  $q$ , в форме

$$\mathbf{F} = q(\mathbf{v} \times \mathbf{B}).$$

Рассмотрим примеры вычисления индукции  $\mathbf{B}$  различных магнитных полей с помощью основных уравнений.

Поле прямого тока. Пусть имеется проводник, по которому течет ток  $I$  так, как показано на рис. 7.35. Проведем вокруг него окружность радиусом  $r$ . Уравнение  $\oint_L \mathbf{B} d\mathbf{l} = \mu_0 I$  показывает,

что в данном случае составляющая  $B$ , касательная к окружности (см. рис. 7.35), должна быть отлична от нуля и во всех точках окружности по модулю должна быть одинакова (это непосредственно следует из отсутствия в силу симметрии картины каких-либо преимуществ точки  $a$  на окружности по сравнению с точкой  $b$  на той же окружности). Две другие возможные составляющие вектора  $\mathbf{B}$  (параллельная радиусу и параллельная току) равны нулю согласно уравнению  $\oint \mathbf{B} d\mathbf{S} = 0$ . В самом деле, если окружить провод цилиндрической поверхностью с осью, совпадающей с проводом с током, то первая составляющая дает поток вектора  $\mathbf{B}$  сквозь боковую поверхность этого цилиндра, а вторая—сквозь торцы цилиндра. Эти потоки равны нулю только в том случае, если соответствующие составляющие вектора  $\mathbf{B}$  равны нулю.

Итак, линии  $\mathbf{B}$  направлены в любой точке по касательной к окружности, проведенной через эту точку вокруг оси, совпадающей с проводником. Направление вектора  $\mathbf{B}$  определяется по правилу винта (рис. 7.36).

Для определения модуля  $B$  воспользуемся уравнением

$$\oint_L \mathbf{B} d\mathbf{l} = \mu_0 I,$$

выбрав в качестве  $L$  окружность радиуса  $r$ . Учтем, что

$$\oint_L \mathbf{B} d\mathbf{l} = \sum_i |\mathbf{B}_i| |d\mathbf{l}_i| \cos(\mathbf{B}_i, \hat{d\mathbf{l}}_i),$$

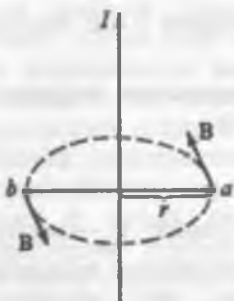


Рис. 7.35. Магнитное поле прямого тока

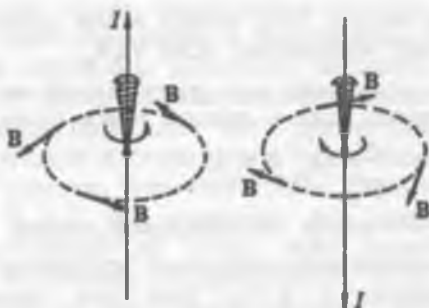


Рис. 7.36. Правило винта

если отрезки  $|dl_i|$  малы. Так как вектор  $B$  в любой точке окружности касателен к ней и во всех ее точках по модулю одинаков, то

$$\cos(\hat{B}_r dl_i) = 1,$$

$$\sum_i |B_i| |dl_i| \cos(\hat{B}_r dl_i) = |B| \sum_i |dl_i| = |B| 2\pi r,$$

где  $r$ —радиус окружности относительно оси, совпадающей с проводником с током. Следовательно,

$$|B| 2\pi r = \mu_0 I \quad \text{и} \quad |B| \sim \frac{I}{r}.$$

Таким образом, прямой ток создает вокруг себя поле, индукция  $B$  которого уменьшается обратно пропорционально расстоянию от проводника.

Поле двух параллельных проводников с током. Определим индукцию поля, создаваемого двумя параллельными проводами, расположенными на расстоянии  $l$  один от другого, по которым текут одинаковые токи, на средней линии между токами и на расстоянии  $x$  от правого проводника с током (рис. 7.37).

На рис. 7.38,а указаны направления векторов  $B$  в случае, когда токи текут в одну сторону, а на рис. 7.38,б—в случае, когда токи текут в противоположные стороны.

Предположим, что токи по обоим проводникам текут в одном направлении. Заметим, что для индукции магнитных полей, как и для напряженности электрических, справедлив принцип суперпозиции. Если есть несколько источников, которые создают электрические или магнитные поля, то суммарная напряженность электрического поля или индукция магнитного поля равна соответственно векторной сумме напряженностей электрических



Рис. 7.37. К нахождению магнитного поля двух параллельных токов

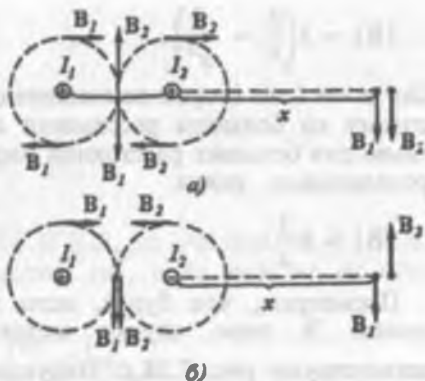


Рис. 7.38. Нахождение напряженности магнитного поля параллельных и антипараллельных токов

или индукций магнитных полей, создаваемых каждым из источников в отдельности:

$$B = \sum_k B_k.$$

Направления векторов  $B_1$  и  $B_2$  полей, создаваемых каждым из токов, показаны на рис. 7.38, а. Воспользуемся принципом суперпозиции и выражением для  $|B|$  прямого тока. Так как по условию по проводникам текут одинаковые токи, то на прямой между проводниками суммарная индукция равна

$$B = B_1 + B_2 = B_1 - B_1 = 0$$

(векторы  $B_1$  и  $B_2$  равны по модулю и направлены противоположно). В некоторой точке на расстоянии  $x$  от правого проводника с током индукция поля равна  $B = B_1 + B_2 \neq 0$ , так как  $B_1$  и  $B_2$  направлены в одну и ту же сторону. Так как силы токов одинаковы, то

$$|B| = |B_1| + |B_2| \sim I \left( \frac{1}{x+l} + \frac{1}{x} \right).$$

Если точка  $x$  находится на большом расстоянии от проводников (под большим расстоянием будем понимать расстояние, которое велико по сравнению с расстоянием  $l$  между двумя проводниками), то  $l$ —малое приращение и можно воспользоваться формулой дифференцирования:  $\frac{1}{x+l} = \frac{1}{x} - \frac{1}{x^2}l$ . Учитывая это, перепишем формулу для индукции следующим образом:

$$|B| \sim I \left( \frac{2}{x} - \frac{l}{x^2} \right).$$

Если  $x \gg l$ , то второй сомножитель убывает быстрее, чем первый, поэтому на большом расстоянии им можно пренебречь. Окончательно для больших расстояний индукция поля, создаваемая двумя проводниками, равна

$$|B| \sim I \frac{2}{x}.$$

Посмотрим, что будет, если токи текут в разных направлениях. В этом случае направления индукций  $B_1$  и  $B_2$  соответствуют рис. 7.38, б. Индукция поля на средней линии двух прямых токов равна удвоенной индукции поля одного тока.

В точке  $x$  вектор индукции  $B_1$  направлен вниз. Вектор индукции  $B_2$  поля в этой же точке от правого проводника направлен вверх. Следовательно, в этом случае векторы индукций магнитных полей должны вычитаться:

$$|B| = \left| |B_1| - |B_2| \right| \sim I \left( \frac{1}{x+l} - \frac{1}{x} \right).$$

Для больших расстояний справедливо соотношение

$$|B| \sim I \frac{l}{x^2}.$$

Таким образом, индукция поля вне пары проводников убывает обратно пропорционально квадрату расстояния между ними. Такое уменьшение индукции поля вне определенным образом расположенных проводников особенно ярко проявляется в соленоиде. *Соленоидом называется изолированный проводник с током, свернутый в спираль* (рис. 7.39).

Поле соленоида. На основе рассмотренного выше примера можно понять, что поле внутри спирали будет однородным, а вне спирали практически равно нулю. За счет большого количества витков можно внутри соленоида создать сильные магнитные поля. Именно поэтому соленоиды являются главной составной частью электромагнитов.

Найдем индукцию поля внутри (на оси) соленоида. Картина расположения векторов  $B$  показана на рис. 7.39. Для вычисления  $B$  воспользуемся равенством

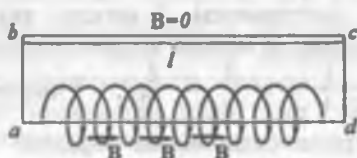


Рис. 7.39. Поле соленоида

$$\oint_L \mathbf{B} d\mathbf{l} = \mu_0 I.$$

Контур  $L$  выберем в форме прямоугольника  $abcd$  (см. рис. 7.39). Так как

$$\oint_L \mathbf{B} d\mathbf{l} = \sum_i |\mathbf{B}_i| |d\mathbf{l}_i| \cos(\mathbf{B}_i, \hat{d\mathbf{l}_i}),$$

то, учитывая, что на участке  $ad$   $\mathbf{B} \parallel d\mathbf{l}$ , на участке  $ab$   $\mathbf{B} \perp d\mathbf{l}$  или равна нулю (аналогично на участке  $cd$ ), а на участке  $bc$  всюду равна нулю, получаем

$$\oint_L \mathbf{B} d\mathbf{l} = |\mathbf{B}|_{ad} \sum_{ad} |d\mathbf{l}| = |\mathbf{B}|_{ad} l_{ad}$$

Символы  $|\mathbf{B}|_{ad}$  и  $\sum_{ad}$  означают, что учитываются только

составляющие индукции поля на оси соленоида на участке  $ad$  длиной  $l$ . Общая сила тока  $I$ , проходящего через контур  $abcd$ , равна произведению числа  $N$  витков, укладываемых на участке длиной  $l$ , и силы тока  $I$  в проводнике.

Таким образом, окончательно для индукции поля на оси соленоида имеем

$$|\mathbf{B}| l = \mu_0 N I \quad \text{или} \quad |\mathbf{B}| = \mu_0 \frac{N}{l} I = \mu_0 I n_0,$$

где  $n_0 = N/l$ —число витков на единице длины, т. е. плотность катушки. Следовательно, для того чтобы индукция поля была больше, можно поступить двумя способами: либо пропустить по соленоиду большой ток, либо небольшой, но плотность катушки соленоида в этом случае должна быть больше. В магнитах, которые используются в промышленности, ток и число витков достаточно велики.

Выше при выводе формулы индукции магнитного поля мы воспользовались интегральным уравнением, связывающим циркуляцию вектора магнитного поля по замкнутому контуру и поток вектора через поверхность, ограниченную этим контуром. При этом просматривается вполне определенная аналогия с выводом выражений для электрических полей с помощью теоремы Гаусса. В электростатике, однако, те же самые результаты можно получить и просто суммируя (интегрируя) по определенным правилам действия отдельных элементарных зарядов, размещенных вдоль заряженного линейного проводника или на плоскости.

Может возникнуть вопрос: а нельзя ли применить аналогичный прием в магнитостатике?

Это действительно можно сделать, если ввести так называемый элемент тока, т. е. вектор, равный  $I d\mathbf{l}$ , где  $I$ —сила тока и  $d\mathbf{l}$ —малый вектор (часть ломаной, с достаточной степенью точности передающей кривую линию), направленный вдоль участка проводника с током в сторону движения тока.

Если от середины такого участка провести радиус-вектор  $r$  в некоторую точку пространства, то вектор магнитной индукции в этой точке, создаваемой таким элементом тока, будет равен

$$dB = \frac{\mu_0}{4\pi} \frac{I[d\mathbf{l} \times \mathbf{r}]}{r^3}.$$

Это соотношение называют *законом Био—Савара—Лапласа*.

Здесь надо заметить следующее. Хотя это соотношение и напоминает закон Кулона для поля точечного заряда, однако самостоятельного физического смысла оно не имеет, а служит лишь математическим приемом для вычисления полного поля от проводника с током произвольной формы по правилу

$$\mathbf{B} = \frac{\mu_0}{4\pi} I \sum_k \frac{[d\mathbf{l}_k \times \mathbf{r}_k]}{r_k^3}.$$

В пределе при малых  $d\mathbf{l}_k$  получим

$$\mathbf{B} = \frac{\mu_0}{4\pi} I \int \frac{[d\mathbf{l} \times \mathbf{r}]}{r^3}.$$

Интегрирование ведется по всей длине проводника.

Магнитное поле как релятивистский эффект. До сих пор мы рассматривали магнитное поле как реальность, пользуясь для его обнаружения магнитной стрелкой. В § 7.1 говорилось, что движущиеся заряды взаимодействуют между собой не так, как неподвижные: сказывается запаздывание передачи воздействия одного из них на другой через посредство электрических полей. Однако подробно этот вопрос не рассматривался.

Постоянные магнитные поля создаются постоянными токами. Картина получается стационарной, и, казалось бы, никакого запаздывания учитывать не надо.

Чтобы разобраться в этом, вернемся снова к магнитному полю, создаваемому текущим по прямому проводнику постоянным током. Магнитное поле на неподвижный заряд не действует. Этот факт отражен в формуле Лоренца, где магнитная компонента полной действующей на заряд силы становится отличной от нуля, только если заряд начинает двигаться. Наибольшее воздействие магнитное поле прямого тока будет оказывать на пробный заряд, движущийся вдоль тока. Тогда в зависимости от относительного движения зарядов в проводнике и внешнего заряда пробный заряд будет либо приближаться к проводнику, либо отклоняться от него. Сила, действующая на пробный заряд, движущийся вдоль тока, направлена перпендикулярно ему и обратно пропорциональна расстоянию  $r$  от пробного заряда до проводника с током. Проводник с током неподвижен, а заряд движется. Но можно изменить точку зрения (перейти к другой системе координат) и считать пробный заряд неподвижным, а проводник с током движущимся. Физическая

картина, согласно общему принципу относительности, должна остаться прежней: если заряд притягивался к проводнику с током, то он и должен притягиваться, а если отталкивался, то отталкиваться.

Но если считать заряд неподвижным, то действовать на него магнитное поле не может, и для появления силы, направленной перпендикулярно проводнику, необходимо электрическое поле.

Здесь надо обратить внимание на то, что в проводнике с током имеются два вида зарядов: положительные—ядра атомов кристаллической решетки металла (считаем, что проводник металлический), которые движутся по отношению к неподвижному заряду с некоторой скоростью  $v_0$ , и отрицательные (электроны), которые по отношению к наблюдателю, совмещенному с пробным зарядом, движутся со скоростью  $v_0 \pm v_{\text{дрейфа}}$ . Знак

выбирается в зависимости от направления тока проводимости в проводнике. Поскольку и те и другие заряды равномерно заполняют все пространство проводника так, что в каждый момент времени в любом участке проводника плотность и тех и других зарядов одна и та же во все время наблюдения, то ситуация будет аналогична такой, как если бы мимо наблюдателя с разными скоростями проносились две разноименно заряженные и вложенные одна в другую трубки. Но если трубки движутся, то должно наблюдаться релятивистское сокращение их длин и, стало быть, если сами индивидуальные заряды электронов и ядер не меняются, увеличение линейной плотности зарядов (число зарядов на единицу длины). Так как положительные и отрицательные заряды движутся с разными скоростями по отношению к неподвижному наблюдателю, то линейные плотности зарядов должны быть различными для ядерной и электронной составляющих.

Если это так, то электрическое поле положительных зарядов уже не компенсирует поле отрицательных и возникает сила, направленная перпендикулярно проводнику, а по модулю обратно пропорциональная  $r$  (вспомним, чему равна напряженность электрического поля около заряженного прямого проводника!).

При переходе от движущейся системы координат к неподвижной величины зарядов элементарных носителей не меняются. В этом легко убедиться, считая, что если атомы металла неподвижны, то электроны даже в отсутствие внешнего поля хаотически движутся подобно частицам газа (мы уже пользовались термином «электронный газ»). Если бы при появлении движения заряды электронов менялись, то не существовали бы электрически нейтральные металлические тела и, более того, при нагревании металлов, что вызывает увеличение скоростей электронов, эти металлы должны были бы заряжаться. Но как раз это и не наблюдается.

Таким образом, появление магнитного поля токов есть чисто релятивистский эффект и никакой новой физической субстанции



(например, в виде магнитных зарядов) появляться не должно, что и подтверждается экспериментально.

Некоторые другие вопросы взаимосвязи электричества и магнетизма мы рассмотрим в § 7.5.

## 7.5. Электродинамика

Поле движущихся зарядов и действие полей на движущиеся заряды. *Электродинамика*—раздел физики, в котором изучаются свойства электрических и магнитных полей, создаваемых движущимися зарядами и меняющимися токами, и воздействия электрических и магнитных полей на движущиеся заряды.

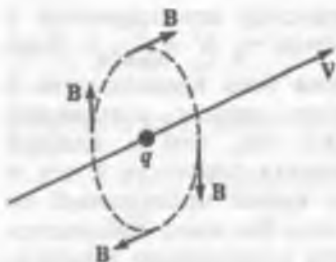


Рис. 7.40. Магнитное поле движущегося заряда

Рассмотрим вопрос о полях, создаваемых движущимися зарядами, и о воздействии этих полей на движущиеся частицы. Как уже упоминалось, всякий равномерно и прямолинейно движущийся заряд меняет свое электрическое поле и создает вокруг себя магнитное поле (рис. 7.40), индукция которого (см. § 7.1)

$$B \sim [v \times E].$$

На всякий электрический заряд действует сила

$$F = qE + q[v \times B].$$

Здесь  $E$ —напряженность электрического поля;  $q$ —заряд;  $v$ —скорость движения заряда;  $B$ —индукция магнитного поля.

Первое слагаемое в этой формуле определяет действие электрической составляющей поля на заряд и не зависит от того, движется или покоится заряд, а второе слагаемое связано с движением заряда и зависит от магнитной компоненты поля.

Вторую составляющую полной силы, действующей на заряд, т. е. величину  $q[v \times B]$ , также называют *силой Лоренца*.

Всякое возмущение в пространстве распространяется со скоростью не выше скорости света. В частности, электрическое поле при смещении точечного заряда не просто переносится вместе с зарядом, как в случае бесконечно большой скорости распространения поля, а меняется более сложным образом. Возникают эффекты, связанные с запаздыванием появления поля на больших расстояниях от заряда, которые могут быть описаны введением индукции магнитного поля.

В разделе «Механика» уже рассматривался вопрос о ряде особенностей, возникающих в случае, если скорости движения тел достаточно большие. Указывалось, что при переходе от неподвижной к движущейся системе координат являются переменными время, длины отрезков. Применим эти результаты к анализу тех изменений в электрическом поле, которые возника-

ют, если заряд начинает двигаться, и обоснуем высказанные выше утверждения.

Для этого предварительно рассмотрим неподвижную и движущуюся плоскости. Будем считать, что плоскость имеет большие размеры, а исследуемая точка пространства находится близко от плоскости. Если  $Q$ —заряд плоскости, а ее площадь  $S$ , то модуль вектора напряженности поля  $|E| = \frac{1}{2\epsilon_0} \frac{Q}{S}$ , причем напряженность направлена перпендикулярно плоскости.

Пусть теперь плоскость смещается равномерно параллельно самой себе. В случае движущихся систем соотношение

$\oint_S E dS = \frac{1}{\epsilon_0} \sum_i q_i$  выполняется и можно исследовать электрические поля совершенно так же, как и поля неподвижных зарядов, с той лишь разницей, что надо учитывать происходящие при этом изменения линейных размеров пространства и связанные с этим деформации поверхностей, внутри которых заключены заряды.

Здесь, однако, надо отметить, что хотя для движущихся зарядов теорема Гаусса сохраняется и не меняются величины зарядов, однако отсюда совсем не следует, что сохраняется и закон Кулона и поле точечного заряда продолжает оставаться сферически симметричным. Напротив, как мы увидим ниже, это поле изменяется.

Вернемся к движущейся плоскости. Она сожмется вдоль одной из сторон и в результате площадь  $S$  поверхности уменьшится в  $\sqrt{1 - v_0^2/c^2}$  раз. В соответствии с этим возрастет поверхностная плотность зарядов  $\sigma$ , равная  $Q/S$ , и возрастет составляющая поля  $E$ , направленная перпендикулярно плоскости.

Как в случае движущейся, так и в случае неподвижной плоскости перпендикулярная составляющая напряженности электрического поля около плоскости равна  $E_y = \frac{1}{2\epsilon_0} \frac{Q}{S}$ . При переходе от одной системы координат к другой меняются только линейные размеры плоскости вдоль оси  $X$ . Поэтому должно быть  $S = S' \sqrt{1 - v_0^2/c^2}$ . Отсюда, в свою очередь, следует, что  $E'_y = E_y \sqrt{1 - v_0^2/c^2}$ . Здесь, как и ранее, штрих относится к движущейся вместе с объектом системе координат.

Мы нашли, таким образом, что при движении плоскости создаваемое ею электрическое поле в направлении, перпендикулярном движению, должно возрасти.

Рассмотрим теперь другой пример—заряженную сферу. Если сфера неподвижна, то поле в пространстве вокруг нее сферически симметрично.

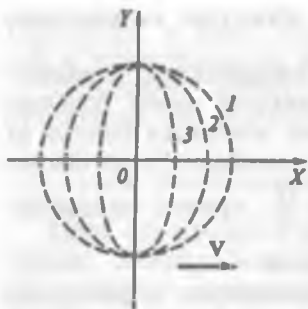


Рис. 7.41. Заряженная сфера: неподвижная (1), движущаяся (2, 3) с разными скоростями

В системе координат, в которой сфера движется равномерно в направлении оси  $X$ , заряженный объект примет вид сплюснутого в направлении оси  $X$  эллипсоида (рис. 7.41).

Сферическая симметрия нарушается. Принципиально изменится и характер электрического поля в окружающем пространстве: поле также перестанет быть сферически симметричным. Как нетрудно видеть, это, в свою очередь, приведет к тому, что циркуляция вектора напряженности электрического поля по замкнутому контуру уже не будет равной нулю (как это имеет место в электростатике).

Вычисления показывают, что напряженность поля  $E$  вокруг движущейся сферы должна увеличиваться в направлении  $Y$ , подобно полю плоскости, и уменьшаться в направлении  $X$ . Заметим, что именно поэтому полный поток вектора электрического поля сквозь поверхность, окружающую сферу, остается неизменным независимо от того, движется сфера или нет.

Сфера малых размеров имитирует точечный заряд, и поэтому напряженность электрического поля движущегося заряда возрастает с ростом скорости его движения в направлении, перпендикулярном направлению движения, и убывает в направлении движения. Этот вывод подтверждается многочисленными экспериментами.

Предположим, что заряд движется ускоренно. Так как напряженность его поля зависит от скорости движения, то при появлении переменной скорости, связанной с ускорением, напряженность электрического поля заряда также будет переменной и зависеть от времени. В этом случае в полном согласии с индукционными членами (см. ниже) в уравнениях Максвелла должно появиться и переменное магнитное поле. Другими словами, *в пространстве вокруг движущегося с ускорением точечного заряда возникает переменное электромагнитное поле, которое будет распространяться во все стороны (в вакууме) до бесконечности. Такое поле называется излучением.*

Следовательно, всякий ускоренно движущийся электрический заряд излучает электромагнитное поле.

Проведенное рассуждение позволяет понять еще один интересный факт. Предположим, что имеются два точечных заряда. Если заряды неподвижны, то между ними действует кулоновская сила притяжения или отталкивания (в зависимости от сочетания знаков зарядов). Допустим, что эти заряды приходят в движение с постоянной скоростью, перпендикулярной прямой, их соединяющей, и движутся параллельно. Тогда напряженность поля, создаваемого каждым зарядом в месте расположения второго, возрастает так, как было указано выше. Казалось бы, должно

измениться и взаимодействие между зарядами. Однако при этом одновременно каждый заряд создаст свое магнитное поле  $B \sim [v \times E]$ . Взаимодействие зарядов определяется формулой Лоренца

$$F = qE + q[v \times E],$$

причем магнитная составляющая силы Лоренца компенсирует изменение электрической составляющей, вызванное возрастанием компоненты электрического поля, перпендикулярной направлению движения заряда. В результате не возникнет эффектов, которые позволили бы, наблюдая поведение этих зарядов, отличить движение от неподвижности.

Таким образом, введение магнитного поля отвечает требованиям принципа относительности и тесно с ним связано. Электрические и магнитные взаимодействия составляют две части одного и того же явления—электромагнитных взаимодействий частиц.

**Взаимодействие токов.** Пусть два параллельных проводника с токами расположены на расстоянии  $l$  друг от друга (рис. 7.42, а). Считаем вначале, что токи текут в одном направлении. Тогда векторы индукции магнитных полей, создаваемых токами в местах, где расположены проводники, направлены так, как показано на рис. 7.42, б, и равны  $|B_1| \sim I_1/l$ ,  $|B_2| \sim I_2/l$ .

Если по каждому из проводников течет ток, то на них действует со стороны магнитного поля определенная сила. Эта сила будет такой, как если бы по проводнику двигались свободные заряды, хотя реальная физическая картина сложнее. Взаимодействие проводников определяется слагаемым  $q[v \times B]$  в выражении для силы Лоренца. В данном случае, учитывая, что произведение  $qv$  пропорционально силе тока, текущего по проводнику, можно написать для силы действия магнитного поля на проводник, по которому течет ток  $I$ , выражение

$$F_{\text{Лор}} \sim [I \times B].$$

Вектор  $I$  равен по модулю силе тока, а направлен вдоль проводника по току.

Пользуясь выражениями для индукции магнитного поля прямого тока и магнитной составляющей силы Лоренца, можно сразу записать, что сила, действующая на первый проводник,

$$F \sim \frac{1}{l} [I_1 \times B_2] \text{ или } |F| \sim \frac{|I_1| |I_2|}{l} = \frac{I_1 I_2}{l} \quad (\text{направление силы}$$

$F$  указано на рис. 7.43). Сила, действующая на второй проводник, равна по модулю первой силе и противоположна ей по знаку. Направление действия этих сил можно определить по правилу построения векторных произведений соответственно векторов  $[I_1 \times B_2]$  и  $[I_2 \times B_1]$ . Таким образом, два параллельных проводника притягиваются друг к другу с силой, пропорциональ-

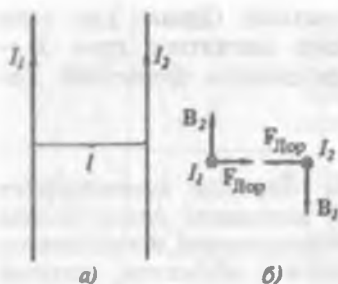


Рис. 7.42. Направление силы Лоренца в случае параллельных токов

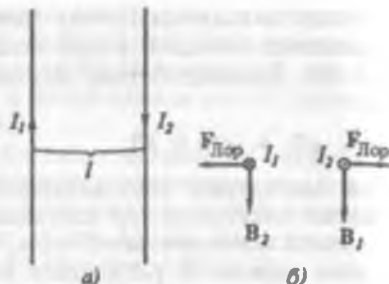


Рис. 7.43. Направление силы Лоренца в случае антипараллельных токов

ной произведению сил токов, текущих по проводникам, и обратно пропорциональной расстоянию между ними.

Если токи текут в разных направлениях, то векторы направлены так, как показано на рис. 7.43. В этом случае силы Лоренца приведут к отталкиванию проводников. Сила отталкивания равна силе притяжения  $F \sim |I \times B|$ .

Воздействие магнитного поля на проводник с током широко используется в технике. Именно на этом принципе работают электродвигатели. Принципиальная схема простейшего электродвигателя заключается в следующем. В некоторой области пространства с помощью двух соленоидов (электромагнитов) создается магнитное поле. В это поле перпендикулярно вектору  $B$  помещается рамка из проводников, соединенная с источником постоянного тока (рис. 7.44).

Если замкнуть цепь источника тока, то по проводникам потечет ток (рис. 7.44, а) и возникнет сила  $F_{\text{Лор}}$ , действующая на проводники.

Силы, действующие на рамку в горизонтальном положении, стремятся сжать ее и для работы двигателя являются бесполезными. Силы, действующие на рамку в вертикальном положении, наоборот, стремятся повернуть ее вокруг оси, а вместе с ней и все устройство. Силы Лоренца во всех случаях действуют в горизонтальном направлении. Вращательный момент создают только компоненты этой силы, перпендикулярные плоскости рамки. Таким образом, в положении б вращающий момент будет наибольшим, в положении а наименьшим (равным нулю), а в промежуточных положениях пропорциональным, очевидно, косинусу угла между направлением  $B$  поля и нормалью к плоскости рамки.

Предположим, что имеются скользящие по кольцу контакты. Тогда после прохождения рамкой горизонтального положения компоненты силы Лоренца, приводящие к вращению, были бы направлены так, как показано на рис. 7.44, и начали бы тормозить первоначальное вращение рамки.

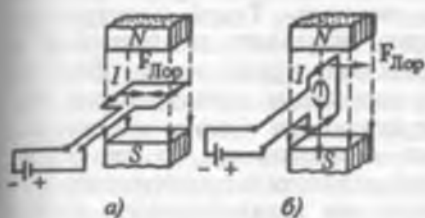


Рис. 7.44. Рамка в магнитном поле



Рис. 7.45. Направление силы Лоренца в случае антипараллельных токов

Чтобы этого не было, надо изменить направление тока в рамке, что достигается благодаря специальному, состоящему из полуколец, коллектору.

Данное устройство может работать и как генератор тока. Для этого с помощью внешнего двигателя надо вращать рамку из проводников. Тогда на положительные и отрицательные заряды в проводниках будет действовать сила Лоренца, направленная согласно правилу векторного произведения по проводнику (рис. 7.45).

Положительные заряды в проводнике двигаться не могут, а отрицательные под действием сил Лоренца начнут перемещаться по проводнику, в результате чего в проводнике возникает электрический ток. Роль коллектора в этом случае заключается в том, что с его помощью ток с рамки снимается и передается во внешнюю цепь.

Возникающий в рамке ток пропорционален силе Лоренца, т. е.  $F \sim [v \times B]$  и  $\cos \alpha$ , где  $\alpha$ —угол между направлениями плоскости рамки и индукцией  $B$  магнитного поля.

Если  $\omega$ —угловая скорость вращения рамки, то угол ее поворота за время  $t$  равен  $\omega t$ . Следовательно, сила наведенного в рамке тока

$$|I| \sim |\omega| |B| \cos \omega t.$$

Даже при очень слабых полях  $B$  можно получить значительные токи, вращая рамку с большой скоростью. Так, например, с помощью такого приема можно обнаружить, что Земля создает вокруг себя хотя и слабое, но все же вполне регистрируемое магнитное поле. Действительно, если в описанном выше приборе отключить ток от соленоидов, создающих магнитное поле, то за счет действия магнитного поля Земли при быстром вращении рамки в ней появится заметный электрический ток, который можно измерить чувствительными приборами.

В науке и технике широко используется действие электрических и магнитных полей на свободные заряженные частицы. В частности, потоки заряженных электронов (электронные лучи)

создаются в телевизорах и осциллографах специальными устройствами (электронными пушками) и фокусируются, подобно световым лучам, электронными линзами. Такой электронный луч, попадая на специальный экран, заставляет его светиться, и след его падения становится виден как яркая точка. Если на пути луча поместить устройства, создающие электрические или магнитные поля, например пропустить луч между пластинами конденсатора или между двумя соленоидами, то он будет отклоняться в зависимости от напряженности электрического или индукции магнитного полей. Меняя эти характеристики, можно добиться того, чтобы электронный луч последовательно пробегал по всему экрану. Если менять интенсивность луча, т. е. яркость пятна на экране, то при достаточно быстром перемещении луча на экране можно получить стойкое изображение, подобное фотографическому.

Особенно часто используется магнитное отклонение луча. Это связано с тем, что сила, действующая на заряды в этом случае, пропорциональна скорости их движения. В частности, это свойство применяется в ускорителях элементарных частиц. Так как для ускорения частицы до больших скоростей ей надо

пробежать громадное расстояние, то стремятся заставить частицы двигаться по окружности, для чего и используется магнитное поле.

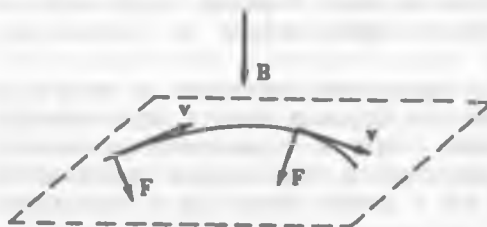


Рис. 7.46. Траектория заряженной частицы в ее скорости  $v$ , испытывает

действие силы Лоренца, равное  $q[v \times B]$  и направленное в любой точке ее траектории (если траектория частицы лежит в плоскости чертежа, а поле  $B$  перпендикулярно ему) перпендикулярно скорости. Модуль скорости частицы вдоль траектории не меняется. Поле приводит лишь к искривлению траектории (рис. 7.46).

Если подобрать такое поле  $B$ , что сила Лоренца равна нормальной силе, т. е.

$$q|v||B| = \frac{mv^2}{r}, \quad \text{или} \quad q|B| = \frac{m}{r}|v|,$$

то частица в магнитном поле будет двигаться по окружности заданного радиуса  $r$ .

Эффектом воздействия поля  $B$  можно воспользоваться, чтобы по заданной скорости частицы  $v$  и ее заряду  $q$  определить ее массу или, наоборот, по скорости и массе найти заряд. И то и другое широко применяется в специальных научных исследованиях. В частности, одним из основных применяемых для анализа строения многоатомных молекул приборов является масс-спектрометр. Его действие основано на способности маг-

прямого пути (обратно пропорционально их массам). В масс-спектрометре под ударами электронов или при другом воздействии происходит распад сложной молекулы на отдельные осколки-радикалы или образование молекулярного иона (молекулярным ионом называют молекулу, потерявшую один электрон). Имеющие разные массы радикалы по-разному отклоняются магнитным полем и попадают в разные участки приемника, где и регистрируются. Зная массы осколков, по эмпирическим правилам можно воссоздать полную исходную систему. Если, например, достаточно точно измерить массу молекулярного иона, то можно с хорошей степенью точности, зная, что каждый из атомов имеет вполне определенную массу, установить брутто-формулу вещества.

Электромагнитная индукция. Уравнения Максвелла. До сих пор мы рассматривали статические электрические и магнитные поля. Для них справедливы соотношения:

$$\left\{ \begin{array}{ll} \oint_S \mathbf{E} d\mathbf{S} = \frac{1}{\epsilon_0} \sum_i q_i = \frac{1}{\epsilon_0} \iiint q dx dy dz, & \oint_S \mathbf{B} d\mathbf{S} = 0, \\ \oint_L \mathbf{E} d\mathbf{L} = 0, & \oint_L \mathbf{B} d\mathbf{L} = \mu_0 I = \mu_0 \iint \mathbf{j} d\mathbf{S}. \end{array} \right.$$

Теперь обратимся к случаю, когда электрические и магнитные поля переменны.

Вспомним некоторые эксперименты, подробно изученные в школьном курсе физики. Первый из них (опыт Фарадея) заключается в следующем.

Имеется соленоид, подключенный к чувствительному гальванометру. Если внутрь соленоида быстро вдвинуть постоянный магнит, то в электрической цепи будет наблюдаться импульс тока. При любом стационарном положении магнита тока не наблюдается.

Тот же эффект, но более слабый, будет наблюдаться, если пронести магнит через один круговой виток проволоки.

Выше мы отмечали, что магнитное поле на неподвижные носители заряда действовать не может. Так как импульс тока тем не менее наблюдается, то это означает, что переменное магнитное поле (магнит движется, и, значит, индукция  $\mathbf{B}$  поля в области витка меняется во времени) порождает электрическое и именно такое, что оно начинает «гнать» электроны по витку.

Значит, это такое поле, для которого циркуляция вектора  $\mathbf{E}$  напряженности по замкнутому контуру уже не равна нулю! Но, в свою очередь, это означает, что линии напряженности такого введенного электрического поля, так же как и поля магнитного, представляются замкнутыми линиями без узлов. Такое поле



называется *вихревым*. Значит, *переменное магнитное поле порождает вихревое электрическое*, причем с тем *большей напряженностью, чем быстрее меняется индукция магнитного поля, пересекающего плоскость витка*. Изменение электрического сигнала при повороте витка по отношению к направлению движения магнита также естественно интерпретировать как результат изменения соответствующего потока вектора  $\mathbf{B}$ .

Обобщение этих и других экспериментов приводит окончательно к следующему соотношению, которое известно как одно из уравнений Максвелла (*закон Фарадея*) для электродинамики:

$$\oint_L \mathbf{E} d\mathbf{l} = - \frac{\partial}{\partial t} \left( \int_S \mathbf{B} d\mathbf{S} \right),$$

где  $S$ —площадь, ограниченная контуром  $L$ .

Обратимся теперь к другому эксперименту. Рассмотрим электрическую цепь с конденсатором (рис. 7.47). Постоянный ток по такой цепи течь не будет: промежуток в конденсаторе представляет для него неодолимое препятствие. Но оказывается, что сразу после включения источника стрелка гальванометра, включенного в цепь, на мгновение отклонится. Значит, в первый момент по контуру пойдет ток. Как это понять? Что будет, когда в цепь включается источник тока? Произойдет зарядка конденсатора так, что на одной его обкладке будет положительный, а на другой отрицательный заряд. Естественно, появится и электрическое поле внутри конденсатора. В первый момент, пока еще картина не установилась и конденсатор полностью не зарядился, это поле будет переменным. Теперь представим, что около конденсатора помещена магнитная стрелка. В момент включения батареи стрелка магнитометра дрогнет. Значит, около конденсатора наряду с переменным электрическим полем возникло и переменное магнитное поле.

Именно это объединенное электромагнитное поле и является переносчиком электрической энергии через конденсатор в момент включения, несмотря на то, что обычный ток проводимости через него не идет. Поэтому и проходят через конденсатор переменные токи. Тогда говорят о появлении *тока смещения*.

Связь между электрическим и магнитным полями в этом случае передается четвертым уравнением Максвелла:

$$\oint_L \mathbf{B} d\mathbf{l} = \mu_0 \int_S \mathbf{j} d\mathbf{S} + \mu_0 \epsilon_0 \frac{\partial}{\partial t} \int_S \mathbf{E} d\mathbf{S}.$$

Поверхность  $S$  заключена внутри контура  $L$ . Совокупность всех четырех уравнений называется *системой уравнений Максвелла* и

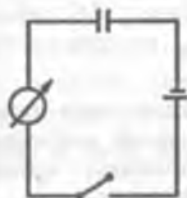


Рис. 7.47. Электрическая цепь с конденсатором



Рис. 7.48. Направление электрического поля, возникающего при наличии переменного магнитного поля

определяет все, что мы сейчас знаем об электромагнитных явлениях. В дополнительных членах вида

$$-\frac{\partial}{\partial t} \left( \int_S \mathbf{B} \, d\mathbf{S} \right) \quad \text{и} \quad \mu_0 \epsilon_0 \frac{\partial}{\partial t} \left( \int_S \mathbf{E} \, d\mathbf{S} \right)$$

во втором и четвертом уравнениях проявляется связь между электрическим и магнитным полями.

Знак минус ставится потому, что индуцированное поле, в свою очередь, препятствует действию причины, вызвавшей его появление (*правило Ленца*).

Рассмотрим некоторое пространство, в котором существует, например, магнитное поле. Проведем перпендикулярно направлению напряженности этого поля небольшую окружность (рис. 7.48). Предположим, что индукция магнитного поля меняется со временем. Тогда второе уравнение Максвелла утверждает, что в этом случае появится не равная нулю циркуляция вектора напряженности электрического поля по выделенной окружности. Другими словами, появится составляющая электрического поля. Эта составляющая перпендикулярна направлению индукции магнитного поля. Можно провести и обратное рассуждение—задать в пространстве электрическое поле и считать его переменным. Тогда согласно четвертому уравнению Максвелла должно появиться переменное магнитное поле. При этом напряженность электрического и индукция магнитного полей взаимно перпендикулярны.

Однажды возникшие переменные электрическое или магнитное поля обязательно создадут переменные магнитное или электрическое поля, причем оба эти поля как единое образование (электромагнитное поле) начнут бесконечно распространяться во все стороны пространства. Так как скорость этого распространения конечная, то возмущение в пространстве будет передаваться в виде некоторого волнового процесса. Такой волновой процесс называется *электромагнитной волной* (см. гл. 8).



Рис. 7.49. Схема трансформатора

Возникновение переменного электрического поля при изменении магнитного называется *электромагнитной индукцией*.

Остановимся на одном важном техническом приложении индукционного эффекта, описываемом уравнением

$$\oint_L \mathbf{E} d\mathbf{l} = - \frac{\partial}{\partial t} \left( \int_S \mathbf{B} d\mathbf{S} \right).$$

Уравнение отражает тот факт, что если взять некоторый кольцевой проводник и менять поток вектора  $\mathbf{B}$  через площадь, ограниченную проводником, то возникнет переменное электрическое поле, направленное по касательной к проводнику. Под действием этого поля электроны в проводнике придут в движение и возникнет электрический ток.

Это свойство объясняет действие *трансформатора*—устройства, которое используется для преобразования переменных токов и напряжений. Трансформатор—это два соленоида, соединенных между собой так, что магнитные линии индукции, выходящие из одного соленоида, пересекают витки другого соленоида. Простейшая схема трансформатора показана на рис. 7.49. Пусть по первому соленоиду трансформатора течет постоянный ток. Он создаст внутри первого соленоида постоянное магнитное поле. Это поле проникает и внутрь второго соленоида. Все заряды внутри второго соленоида в начальный момент покоятся. Магнитное поле, создаваемое первым соленоидом, может воздействовать только на движущиеся заряды. Поэтому при подаче на первый соленоид постоянного тока никакого тока во втором соленоиде не возникает. Однако картина существенно изменяется, если на первый соленоид подавать переменный ток. В этом случае внутри первого соленоида и, следовательно, внутри второго соленоида возникает уже переменное магнитное поле.

Уравнение Максвелла утверждает, что в этом случае появится электрическое поле, зависящее от переменного магнитного поля, направленное перпендикулярно линиям индукции переменного магнитного поля. Следовательно, направление  $\mathbf{E}$  будет способствовать появлению движения электрических зарядов (ранее неподвижных) по виткам соленоида. Возникнет электрический ток внутри проводов второго соленоида.

За счет индукционных эффектов, описываемых дополнительными членами в уравнении Максвелла, во втором соленоиде появится электрический переменный ток. Напряженность наведенного электрического поля пропорциональна числу витков

первого соленоида и силе тока в нем. Индуцированное переменным магнитным полем электрическое поле играет роль сторонних электрических сил для второго соленоида. Работа сил по перемещению единицы заряда по всей длине провода соленоида определяет ЭДС этих сил и равна произведению  $|E|$  на длину провода соленоида. Последняя же пропорциональна числу  $N_2$  витков второго соленоида. Напряженность  $|E|$  наведенного электрического поля пропорциональна числу витков  $N_1$  первого соленоида и силе тока  $I_1$  в нем. Тогда

$$\mathcal{E}_2 \sim |E| N_2 \sim I_1 N_1.$$

Но, как было показано выше в разделе о постоянных токах, сила тока в проводнике пропорциональна  $|E_{\text{тор}}|$ . Следовательно,

$$I_2 \sim |E| \quad \text{и} \quad N_2 I_2 \approx N_1 I_1.$$

Эта формула называется *формулой трансформатора* и показывает, что с помощью соленоидов можно скачком изменять силу тока во вторичной обмотке по отношению к силе тока в первичной обмотке.

Мы привели уравнения для электрических и магнитных полей в вакууме. Если перейти к пространству, заполненному веществом, то в эти уравнения надо ввести изменения, учитывающие влияние среды. Наиболее просто это делается для полей, сравнительно медленно меняющихся во времени и пространстве. Тогда наряду с векторами  $E$  и  $B$  надо ввести дополнительно векторы  $D$  и  $H$ , связанные с первыми следующими соотношениями:

$$D = \epsilon E \quad \text{и} \quad B = \mu H,$$

где  $\epsilon$  и  $\mu$  — скалярные постоянные (диэлектрическая и магнитная проницаемости), характеризующие электромагнитные свойства среды.

Уравнения Максвелла тогда примут вид:

$$\begin{aligned} \oint_S D \, dS &= \sum_i q_i, & \oint_L E \, dl &= - \frac{\partial}{\partial t} \left( \int_S B \, dS \right), \\ \oint_L H \, dl &= I + \frac{\partial}{\partial t} \left( \int_S D \, dS \right), & \oint_S B \, dS &= 0. \end{aligned}$$

В этой записи, кроме того, изменены числовые множители, что соответствует переходу к Международной системе единиц (СИ). В этой системе различие между векторами  $E$ ,  $D$  и  $B$ ,  $H$  сохраняется и в вакууме, которому приписываются свои электрическая и магнитная постоянные  $\epsilon_0$  и  $\mu_0$ , не равные единице.

Тогда напряженность электрического поля в вакууме для точечного заряда равна

$$E = \frac{q}{4\pi\epsilon_0} \frac{r}{|r|^3},$$

а индукция магнитного поля вокруг прямого тока

$$|B| = \frac{\mu_0}{2\pi r} I.$$

Поскольку для вакуума векторы  $D = \epsilon_0 E$  и  $B = \mu_0 H$ , то везде выше в соответствующих уравнениях мы явно записали множители размерности  $\epsilon_0$  и  $\mu_0$ , чтобы не вносить путаницы и сохранить в уравнениях лишь два типа векторов.

#### Вопросы для самоконтроля

1. Какая физическая величина приводит к тому, что в случае движущихся зарядов уже нельзя ограничиваться только кулоновским электростатическим взаимодействием и для полного описания этого взаимодействия приходится вводить понятие индукции магнитного поля в качестве некоторого дополнения?

2. Может ли появиться отрицательный заряд в некоторой области пространства, если при этом нигде не возникает положительный заряд?

3. Записать в векторной форме уравнение движения частицы с массой  $m$  и зарядом  $q$  в однородном магнитном поле.

4. Какие вы знаете способы описания векторных полей? Что такое однородное и неоднородное поля, источник поля или узел? Чем отличается вихревое векторное поле от поля с источниками?

5. Какими основными свойствами обладают линии напряженности и линии индукции электрического и магнитного полей?

6. Найти поток вектора  $E$  однородного электрического поля сквозь боковую поверхность равнобедренной прямоугольной пирамиды со стороной основания, равной  $a$ , если вектор  $E$  нормален к плоскости основания.

7. Показать, что циркуляция вектора напряженности электрического поля вдоль любой окружности в случае однородного поля всегда равна нулю вне зависимости от угла между вектором  $E$  и нормалью к плоскости окружности.

8. Пользуясь каким принципом, можно определить поле в заданной точке пространства, если известно распределение точечных или иных зарядов в пространстве вокруг этой точки?

9. Показать, как реальные тела, состоящие из распределенных в пространстве электронов и ядер, заряженных соответственно отрицательно и положительно, «прячут» свои электрические свойства в том смысле, что уже на сравнительно малых расстояниях они выглядят как электрически нейтральные.

10. Предположим, что в некоторой сложной молекуле на ее периферии расположен атом, обладающий большим положительным зарядом. С какими ионами, положительными или отрицательными, следует в этом случае ожидать наибольшую вероятность прохождения химической реакции присоединения около этого заряда?

11. Известно, что полярные заместители в сложных молекулах способны менять ход ион-молекулярных химических реакций у реакционных центров, расположенных даже иногда достаточно далеко от этого полярного заместителя. Какая причина приводит к изменению хода химических реакций?

12. Приведите примеры использования теоремы Гаусса для вычисления напряженности электростатических полей в различных случаях.

13. Как убедиться в том, что электростатическое поле, создаваемое неподвижными зарядами, является потенциальным?

14. Чему равен потенциал поля, создаваемого протяженным, бесконечно длинным линейным зарядом?

15. Могут ли находиться в состоянии устойчивого равновесия расположенные на некотором расстоянии друг от друга положительные и отрицательные точечные заряды, если никакие силы не удерживают эти заряды в их исходных положениях?
16. Как распределяется заряд на поверхности заряженного проводника?
17. Почему с острого выступа заряженного проводника происходит стекание электрических зарядов?
18. Какое свойство реагирующей среды лежит в основе метода высокочастотного титрования?
19. Какой физический эксперимент позволяет сформулировать первое правило Кирхгофа?
20. Какие физические причины приводят к тому, что при одной и той же напряженности внутреннего электрического поля сила тока в хороших и плохих проводниках оказывается различной?
21. Опишите физическую причину появления тока за счет дрейфа электронов внутри проводников на больших расстояниях от источников зарядов или от полюсов батарей.
22. Какой физический смысл вкладывается в понятие сторонних электрических сил? Как действуют эти силы?
23. Какой физический смысл имеет понятие электродвижущей силы? Как может быть измерена электродвижущая сила гальванического элемента?
24. С помощью каких экспериментов с движущимися зарядами можно определить направление и модуль вектора индукции магнитного поля в данной области пространства?
25. С помощью какого правила можно определить направление вектора индукции магнитного поля вокруг прямого тока?
26. Определить силу взаимного притяжения двух параллельных проводников, по которым протекают токи  $I_1$  и  $I_2$ , расположенные на расстоянии  $R$  один от другого. В каких случаях возникнет отталкивание двух проводников и притяжение их друг к другу?
27. Существовало бы магнитное поле вокруг проводников, по которым протекают токи, если скорость света была бы бесконечно большой?
28. Сохраняется ли закон Кулона при переходе от неподвижных к движущимся зарядам?
29. Какое уравнение для вектора напряженности электрического поля остается неизменным при переходе от неподвижных зарядов к движущимся и почему такое сохранение уравнения возможно?
30. Показать, что электрическое поле, создаваемое заряженной пластинкой, в непосредственной близости от нее меняется в зависимости от того, движется ли эта пластинка или неподвижна. Каков характер этого изменения?
31. Как связано введение магнитного поля по принятым в физике правилам с общим принципом относительности о невозможности отличить равномерное равнолинейное движение от покоя, опираясь на эксперименты, производимые внутри одной из систем?
32. Какой принцип заложен в действии масс-спектрометра и почему с помощью такого прибора удается решать задачи о структурном составе сложных молекул?
33. Что нового появляется в выражении для циркуляции вектора электрического поля по замкнутому контуру в том случае, когда возникают переменные во времени ситуации? Следствием каких экспериментов является появляющийся при этом новый член?
34. Что нового добавляется к выражению для циркуляции вектора индукции магнитного поля по замкнутому контуру в том случае, когда в пространстве возникает переменное электрическое поле? Следствием каких экспериментов является необходимость введения дополнительного члена?
35. К появлению какого нового материального образования приводит наличие дополнительных членов в уравнениях для циркуляции векторов индукции и напряженности магнитного и электрического полей по замкнутому контуру?
36. Почему при переходе к переменным электрическим и магнитным полям сохраняется неизменным уравнение для потока вектора индукции магнитного поля сквозь замкнутую поверхность?

37. Чем отличаются по своему виду и характеристикам линии напряженности электрических полей, генерируемых переменными магнитными полями от линий напряженности электростатических полей?

38. Возможно ли существование электромагнитного поля, если бы линии напряженности наведенного электрического поля по-прежнему имели такой же вид, как линии напряженности, отвечающие статическому электрическому полю?

39. Почему переменный ток проходит по цепи, содержащей конденсатор, а постоянный ток не проходит?

40. Что такое элемент тока и каково содержание закона Био—Савара—Лапласа?

41. Какое дополнительное поле возникает вокруг электрического заряда, движущегося ускоренно, по сравнению с таким же зарядом, но движущимся равномерно и прямолинейно?



## ГЛАВА 8

### ЭЛЕКТРОМАГНИТНЫЕ ВОЛНЫ И СВЕТ

*«И сказал Бог: да будет свет,  
И стал свет. И увидел Бог свет,  
что он хорош...»*

*Бытие*

#### 8.1. Появление и общие свойства электромагнитных волн

В гл. 7 мы остановились на индукционных эффектах. Наличие этих эффектов приводит к необходимости дополнить уравнения для электрических и магнитных полей. Для статики (когда либо заряды неподвижны и неизменны по величине, либо токи постоянны и текут по неподвижным проводам) эти уравнения имеют для вакуума вид

$$\left\{ \begin{array}{l} \oint_S \mathbf{E} d\mathbf{S} = \frac{1}{\epsilon_0} \sum_i q_i = \frac{1}{\epsilon_0} \oint_{V_S} \rho dV, \\ \oint_L \mathbf{E} d\mathbf{l} = 0, \end{array} \right. \quad \left\{ \begin{array}{l} \oint_S \mathbf{B} d\mathbf{S} = 0, \\ \oint_L \mathbf{B} d\mathbf{l} = \mu_0 I = \mu_0 \int_{S_L} \mathbf{j} d\mathbf{S}, \end{array} \right.$$

где  $V_S$  и  $S_L$  — соответственно объем, заключенный внутри замкнутой поверхности  $S$ , и поверхность, ограниченная контуром  $L$ ,  $\rho$  — плотность заряда, распределенного внутри объема  $V_S$ , и  $\mathbf{j}$  — вектор плотности тока, проходящего через поверхность  $S_L$ .

Такие факты, как прохождение переменного тока через цепь, включающую конденсатор, преобразования токов и напряжений трансформаторами и целый ряд других, заставляя учесть связь между векторами напряженностей электрического и магнитного полей и записать уравнения Максвелла в усложненном виде:

$$\oint_S \mathbf{E} d\mathbf{S} = \frac{1}{\epsilon_0} \int_{V_S} \rho dV,$$



$$\oint_L \mathbf{E} d\mathbf{l} = - \frac{\partial}{\partial t} \left( \int_{S_L} \mathbf{B} d\mathbf{S} \right), \quad \oint_S \mathbf{B} d\mathbf{S} = 0,$$

$$\oint_L \mathbf{B} d\mathbf{l} = \mu_0 \int_{S_L} \mathbf{j} d\mathbf{S} + \mu_0 \epsilon_0 \frac{\partial}{\partial t} \int_{S_L} \mathbf{E} d\mathbf{S}.$$

Выше (см. гл. 7) мы говорили о том, что появление в некотором пространстве переменного магнитного поля приведет к появлению переменного электрического, и наоборот. Существенно, что такое переменное электрическое поле будет в отличие от статического вихревым, т. е. его линии напряженности имеют вид замкнутых кривых. В этом смысле наведенное поле подобно магнитному.

Такие свойства индуцированных полей приводят к появлению новой материальной субстанции—*электромагнитного поля*, или *поля излучения*.

Посмотрим, как оно возникает. Для этого разберем явления, происходящие вблизи поверхности плоского тока, т. е. тока, который проходит по однородному плоскому проводнику (рис. 8.1).

Если по плоскости течет постоянный ток, то вокруг нее должно образоваться магнитное поле. Из соображений симметрии или на основании суммирования большого числа магнитных полей от тонких прямых проводников (прямых токов), уложенных плотно один к другому и имитирующих плоскость, по которой идет ток, легко сделать вывод, что линии индукции на участках, достаточно удаленных от границ плоскости и не очень удаленных от самой плоскости, направлены вдоль оси  $Y$ .

В самом деле, выберем точку  $A$  на некотором расстоянии от плоскости (рис. 8.2). Вектор индукции  $\mathbf{B}_0$  магнитного поля в точке  $A$  для прямого тока, расположенного под точкой  $A$ , направлен вдоль оси  $Y$ . Вектор индукции  $\mathbf{B}_1$  в той же точке, но поля, созданного прямым током, проходящим на расстоянии  $O1$  от точки  $O$ , направлен так, как показано на рис. 8.2. Считаем, что элементарные прямые токи текут в направлении снизу вверх по отношению к рисунку. Если плоскость достаточно велика, то всегда найдется сопряженная точке 1 точка 2, такая, что  $O1 = O2$ . Тогда индукция  $\mathbf{B}_2$  поля, создаваемого в точке  $A$  прямым током, проходящим через точку 2, имеет направление, показанное на рис. 8.2.

Полную индукцию  $\mathbf{B}_A$  поля в точке  $A$  можно получить, пользуясь принципом суперпозиции:  $\mathbf{B}_A = \mathbf{B}_0 + \mathbf{B}_1 + \mathbf{B}_2$ ; вектор  $\mathbf{B}_A$  направлен параллельно плоскости по оси  $Y$ .

Линии индукции магнитного поля плоского тока показаны на рис. 8.2. Прямолинейные составляющие этих линий в области,

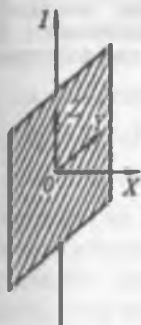


Рис. 8.1. Участок «плоского тока» в электрической цепи — тонкая однородная металлическая пластинка

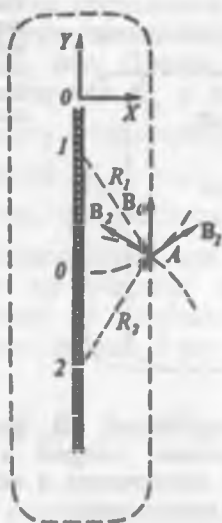


Рис. 8.2. К определению направления вектора  $B$  и силовых линий (штрихпунктирная линия) у поверхности «плоского тока». Кружочками показаны торцы проводников, заполняющих плоскость. Ток идет перпендикулярно чертежу снизу вверх

достаточно близкой к середине плоскости, имеют компоненты только по оси  $Y$ .

Пусть теперь ток в цепи, содержащей проводящую плоскость, меняется со временем, например уменьшается. Тогда должна меняться и индукция магнитного поля, создаваемого этим током. Важно то, что геометрическая форма линий индукции при этом остается неизменной. Это следует из четвертого уравнения Максвелла, в котором сила тока и циркуляция вектора индукции магнитного поля явно не зависят от времени.

Меняющееся во времени вследствие изменения силы тока магнитное поле порождает вихревое электрическое поле согласно уравнению

$$\oint_L E d\mathbf{l} = - \frac{\partial}{\partial t} \left( \int_{S_L} B dS \right).$$

Векторы  $E$  во всех точках поля должны быть перпендикулярны вектору поля  $B$ . Значит, линии напряженности поля  $E$  лежат в плоскости  $XZ$ . Теперь обратимся к эксперименту. Пусть по прямому проводу течет постоянный ток. Вокруг провода существует магнитное поле. Отключим источник тока. Ток при этом идти перестанет, магнитное поле исчезнет. Однако такое прекращение тока произойдет не сразу: в первый момент вслед за размыканием контактов сила тока в проводнике не только не начнет уменьшаться до нуля, как этого можно было бы ожидать, но, наоборот, несколько возрастет и лишь затем начнет уменьшаться. Чем больше была разность потенциалов на контактах источника тока и чем более коротким был промежуток времени, течение которого произошел разрыв цепи, и чем более сложная цепь, тем большим будет импульс тока. Такое явление

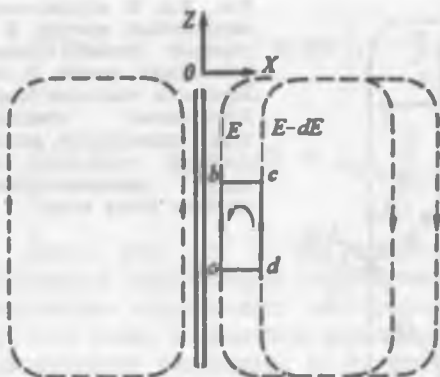


Рис. 8.3. Силовые линии индуцированного переменным магнитным полем электрического поля около пластинки с током. Проекция на координатную плоскость ZX

получило название *экстраток*. В ряде случаев экстраток является причиной серьезных аварий в промышленных сетях электропередач. Поэтому отключение и включение мощных токов желательно производить постепенно.

Появление импульса тока естественно связать с возникновением в результате электромагнитной индукции вектора напряженности электрического поля, направленного вдоль провода.

Принимая это во внимание, легко заключить, что линии напряженности индуцированного переменным магнитным полем электрического поля в случае плоского тока имеют вид, показанный на рис. 8.3.

Выделим близко к середине плоскости прямоугольный контур  $abcd$  (рис. 8.3). Пусть  $ab = cd = l$ ,  $bc = ad = dx$ , где  $dx$ —малая величина (дифференциал). Считая, что на расстоянии  $dx$  модуль вектора  $B_y$  поля, направленного по оси  $Y$  (перпендикулярно плоскости чертежа на рис. 8.3), практически не меняется, находим

$$\int_{S_{abcd}} B dS = B_y l dx$$

и, значит,

$$-\frac{\partial}{\partial t} \left( \int_{S_{abcd}} B dS \right) = -l dx \frac{\partial B_y}{\partial t}.$$

Для определенности примем, что  $B_y$  уменьшается (ток выключается). Тогда  $\frac{\partial B_y}{\partial t} < 0$ .

Определим циркуляцию вектора  $E$  по замкнутому контуру  $abcd$ . Здесь надо учесть, что уменьшение силы тока в цепи будет ощущаться во всем окружающем пространстве не сразу, а в соответствии с требованием конечности скорости распространения

любого сигнала с некоторым опозданием. Это означает, что если около проводника появляется наведенное электрическое поле, то вдали от него его еще нет. Таким образом, должен быть некоторый градиент в величине  $E_z$  в направлении  $X$ . Поэтому если на участке  $ab$  принять составляющую поля равной  $E_{zab}$ , то на участке  $cd$  она равна  $E_{zcd} = E_{zab} - dE_z$  (считаем, что  $dE_z > 0$ ). Учитывая, что на участках  $bc$  и  $ad$  направления напряженности поля и перемещения при обходе контура взаимно перпендикулярны и что на участках  $ab$  и  $cd$  обход контура совершается соответственно по направлению  $E$  и против него, находим, что циркуляция вектора  $E$  по контуру  $abcd$  равна

$$\oint_{L_{abcd}} E dl = (-E_{zab} + E_{zcd})l = -ldE_z.$$

Тогда на основании вышеприведенных рассуждений запишем

$$-ldE_z = l dx \frac{\partial B_y}{\partial t}, \quad \text{или} \quad -\frac{dE_z}{dx} = \frac{\partial B_y}{\partial t}.$$

Так как в переменном по времени процессе величина  $E_z$  есть функция не только  $x$ , но и времени, то надо производную  $dE_z/dx$  заменить на частную производную и переписать последнее соотношение в форме

$$-\frac{\partial E_z}{\partial x} = \frac{\partial B_y}{\partial t}.$$

Теперь обратимся к тому полю  $B$ , которое, согласно последнему уравнению Максвелла, индуцируется переменным электрическим полем  $E$ . Используя свойство взаимной перпендикулярности векторов  $E$  и  $B$  в индуцированных полях, находим, что линии  $B$  индуцированного поля лежат в плоскости  $XY$ . Из опыта с экстратоком можно заключить, что форма линий индукции наведенного магнитного поля не должна меняться по сравнению с формой линии исходного поля. К этому выводу можно прийти и на основании анализа последнего уравнения Максвелла, в котором оба слагаемых в правой части играют одинаковую роль, что и позволило связать слагаемое

$k \frac{\partial}{\partial t} \left( \int_S E dS \right)$  с током смещения. Буквой  $k$  обозначен коэффициент размерности, который для вакуума в СИ равен  $k = \mu_0 \epsilon_0$ . В результате получим картину линий индукции наведенного магнитного поля, показанную на рис. 8.4.

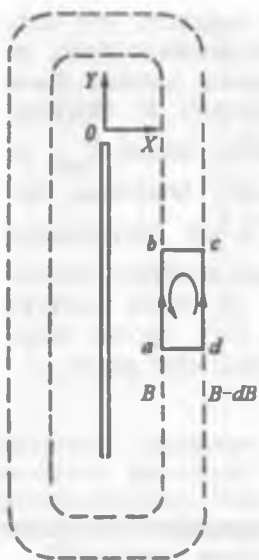


Рис. 8.4. Силовые линии индуцированного переменным электрическим полем магнитного поля. Проекция на плоскость XY

Выделяя контур  $abcd$  (рис. 8.4), учитывая, что теперь уже производная  $\partial E_z / \partial t$  положительна, так как  $|E|$  возрастает, и проводя аналогичные вышеприведенным выкладки, находим

$$-l dB_y = kl dx \frac{\partial E_z}{\partial t} \quad \text{и} \quad -\frac{\partial B_y}{\partial x} = k \frac{\partial E_z}{\partial t}.$$

Теперь мы имеем два соотношения:

$$-\frac{\partial B_y}{\partial x} = k \frac{\partial E_z}{\partial t}, \quad -\frac{\partial E_z}{\partial x} = \frac{\partial B_y}{\partial t}.$$

Продифференцируем их сначала по  $x$ , а затем по  $t$ :

$$\begin{aligned} -\frac{\partial^2 B_y}{\partial x^2} &= k \frac{\partial^2 E_z}{\partial x \partial t}, & -\frac{\partial^2 B_y}{\partial x \partial t} &= k \frac{\partial^2 E_z}{\partial t^2}, \\ -\frac{\partial^2 E_z}{\partial x^2} &= \frac{\partial^2 B_y}{\partial t \partial x}, & -\frac{\partial^2 E_z}{\partial x \partial t} &= \frac{\partial^2 B_y}{\partial t^2}. \end{aligned}$$

Отсюда можно заключить, что

$$\frac{1}{k} \frac{\partial^2 B_y}{\partial x^2} = \frac{\partial^2 B_y}{\partial t^2}, \quad \frac{1}{k} \frac{\partial^2 E_z}{\partial x^2} = \frac{\partial^2 E_z}{\partial t^2}.$$

Таковы соотношения между пространственными и временными производными векторов  $E$  и  $B$  в некоторой области пространства около тока. Эти соотношения характерны для распространяющихся в направлении оси  $X$  со скоростью  $1/\sqrt{k}$  поперечных волн. Значит, поля  $E$  и  $B$  существуют именно в форме волн.

Так как скорости распространения волн  $E$  и  $B$  одинаковы, то везде, где существует волна  $E$ , существует и волна  $B$ . Возникает некоторое единое образование, которое называется *электромагнитной волной*. Электромагнитная волна в вакууме не локализуется в некоторой области пространства, а распространяется во все стороны. Частным случаем электромагнитных волн является свет. Поэтому для вакуума  $1/\sqrt{k} = c$ .

В направлении, перпендикулярном плоскости векторов  $E$  и  $B$  волна переносит энергию, которую можно (учитывая некоторую аналогию с энергией электростатического поля) записать в виде

$$W \sim |E| |B|,$$

где  $W$ —энергия, переносимая волной сквозь нормальную к направлению распространения единичную поверхность в единицу времени.

Обычно для получения электромагнитных волн применяются периодические изменения силы тока в антеннах. В этом случае

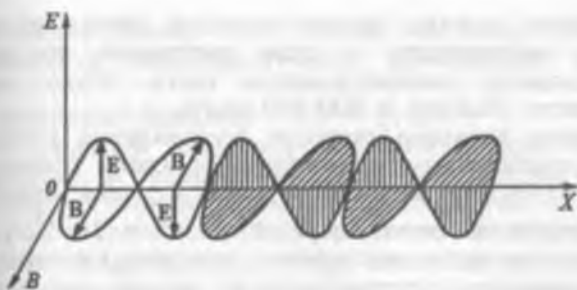


Рис. 8.5. Схематическое изображение электромагнитной волны

в окружающем пространстве возникают разной формы (плоские, цилиндрические, сферические и др.) электромагнитные волны. В случае плоской монохроматической волны имеем

$$E = E_0 \sin \omega(t - x/c), \quad B = B_0 \sin \omega(t - x/c), \quad E \perp B$$

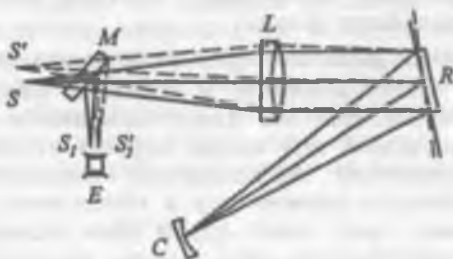
(ось  $X$ —направление распространения волн). Схематически электромагнитная волна изображена на рис. 8.5.

Кроме переменных токов или движущихся зарядов электромагнитные волны порождаются также и внутриатомными и внутримолекулярными процессами (см. гл. 9).

Остановимся на способах измерения скорости распространения электромагнитных волн (в частности, света). В большинстве случаев для измерения скорости света используют один и тот же физический принцип—принцип «эха», основанный на измерении времени распространения волн до экрана и обратно. Зная расстояние до экрана и время распространения, можно определить скорость волны. Одним из первых скорость света измерил Ж. Фуко. Схема его эксперимента представлена на рис. 8.6.

Свет от некоторого источника после прохождения через линзу распространялся параллельными пучками на расстояние порядка одного километра и попадал на вращающееся многогранное зеркало. В зависимости от скорости вращения зеркала свет либо отражался обратно и был зафиксирован в месте испускания, либо под углом уходил в сторону и не был воспринят. Измерив

Рис. 8.6. Определение скорости света методом вращающегося зеркала:  $S$ —источник света,  $M$ —полупрозрачное зеркало,  $L$ —объектив,  $R$ —одна из плоскостей вращающегося барабана с зеркалами,  $C$ —сферическое фокусирующее отражающее зеркало,  $E$ —окуляр. Измеряется смещение «прямого» изображения источника по отношению к изображению за счет отражения от вращающегося зеркала  $R$ . Если скорость вращения известна, то можно найти скорость света  $c$



скорость вращения зеркала, соответствующую случаю отражения света в глаз наблюдателя, и зная расстояние до зеркала, определяют скорость распространения света. Фуко получил значение скорости, близкое к 300 000 км/с.

В дальнейшем значение скорости неоднократно уточнялось. Современные способы измерения скорости света основаны на применении импульсных источников света с очень малой длительностью импульсов света (порядка  $10^{-11}$  —  $10^{-13}$  с). Здесь также производится сравнение времени запаздывания посланного сигнала по сравнению с отраженным от экрана, но при этом из-за малой длительности светового сигнала достигается гораздо большая точность измерения скорости света, чем при использовании непрерывного источника света. Измерения показали, что любая электромагнитная волна независимо от того, является ли она радиоволной или световой, в вакууме распространяется с одинаковой скоростью  $c = 2,997925 \cdot 10^8 \text{ м} \cdot \text{с}^{-1}$ .

## 8.2. Шкала электромагнитных волн

Электромагнитные волны могут генерироваться либо антеннами вследствие колебаний переменного тока (малые частоты), либо вследствие процессов, совершаемых в ламповых или полупроводниковых приборах ускоренно движущимися зарядами (более высокие частоты), либо вследствие внутриатомных процессов или торможения электронов в металлах (свет и рентгеновское излучение) и т. д.

Принято классифицировать электромагнитные волны по интервалам.

Рассмотрим кратко основные интервалы.

### 1. Диапазон длинных радиоволн ( $10^4$ — $10^2$ м).

Этот диапазон в основном используется для радиосвязи. Основной недостаток этих волн состоит в том, что ионосферный слой (или слой с повышенной концентрацией заряженных частиц) над поверхностью Земли хорошо их пропускает. Поэтому эти волны для радиосвязи на большие расстояния не эффективны. Однако волны этого диапазона достаточно хорошо проходят через воду, поэтому применяются для радиосвязи в подводном флоте.

### 2. Диапазон средних и коротких радиоволн ( $3 \cdot 10^2$ — 10 м).

Этот диапазон в основном используется для радиовещания. Объясняется это двумя причинами. Во-первых, по сравнению с длинными эти волны хорошо отражаются ионосферой. Поэтому вследствие многократного отражения короткие волны могут обогнуть земной шар и обеспечить дальнюю радиосвязь. Во-вторых, для того чтобы без искажений передать музыку и человеческую речь, надо выделить для нормальной работы радиостанции участок частот порядка 20 кГц (так как человек воспринимает звуковые волны от 10—20 до 20 000 Гц). Чем

меньше длина волны, тем больше в этом диапазоне можно разместить радиопередатчиков, работающих без взаимных помех.

### 3. Метровый диапазон (10 — 1 м).

Этот диапазон радиоволн используется для телевидения и радиолокации. В телевидении необходимо передавать на большие расстояния не только звуковые сигналы, но и изображения. Поэтому для качественной (без помех и искажений) работы телевизионной станции необходимо выделить полосу частот гораздо большую, чем для работы обычной радиостанции. (В радиотехнике доказана теорема о необходимости выделения тем большей полосы частот, чем больше информации необходимо передать.) Недостаток распространения волн этого диапазона в том, что они хорошо проходят через ионосферу и поэтому не огибают поверхность Земли. В связи с этим для увеличения дальности телевизионного приема приходится размещать излучатели телестанций в очень высоких башнях (например, Останкинская телебашня в Москве) и строить много ретрансляционных станций. В настоящее время для этой цели на специально подобранные околоземные орбиты запускают спутники, постоянно находящиеся над данным районом и служащие ретрансляторами передач центрального телевидения и радиовещания.

Для радиолокации необходимо используемые волны легко фокусировать в определенном направлении. Для этого надо, чтобы размеры отражателя антенны, создающего такой направленный пучок, были одного порядка с длиной волны и даже много больше ее. Поэтому радиолокация в основном осуществляется на волнах метрового и более короткого диапазонов.

### 4. Сантиметровый диапазон ( $1 - 10^{-2}$ м).

Волны этого диапазона используются в авиации для точной близкой радиолокации. Волны этого диапазона сильно поглощаются веществом, поэтому их применяют и для изучения свойств вещества.

### 5. Миллиметровый (микроволновый) диапазон (10 — 1 мм).

Этот диапазон используется для точной радиолокации, а также для научных целей, так как к этому диапазону относятся вращательные спектры многоатомных молекул и по поглощению миллиметровых волн можно судить о строении молекул. С помощью этого метода, в частности, недавно было обнаружено, что в космосе существуют не только простые молекулы, например водород, аммиак и другие, но и сложные молекулы — аминокислоты. Общее число обнаруженных молекул достигает 200. В связи с этим в настоящее время обсуждается гипотеза о внеземном происхождении жизни: за счет занесенных из космоса молекул.

### 6. Инфракрасный диапазон ( $1 - 10^{-3}$ мм).

Этот диапазон используется в научных целях для диагностики и исследования вещества — определения типов взаимодействующих молекул, характера движения ядер в молекулах и



т. д. Кроме того, инфракрасные волны используют в технике для создания приборов «ночного видения», в которых по испущенным или отраженным предметами волнам этого диапазона можно определить расположение и форму предметов даже ночью.

#### 7. Видимый свет (0,7 — 0,45 мкм).

Этот диапазон представляет всю информацию, которую мы воспринимаем органами зрения.

#### 8. Ультрафиолетовый диапазон (0,4 — 0,1 мкм).

Волны этого диапазона способны активно воздействовать на вещество. Ультрафиолетовый загар от солнца или кварцевой лампы может оказывать сильное физиологическое воздействие на человека и даже привести к ожогам. Из-за сильного взаимодействия с веществом ультрафиолетовое излучение Солнца почти полностью поглощается атмосферой (до 99%), и только поэтому возникли условия для существования жизни на Земле—иначе все живое было бы сожжено. В последнее время серьезно обсуждается вопрос о возможном нарушении такого защитного механизма земной атмосферы. Дело в том, что из-за попадания в атмосферу некоторых продуктов химии и запусков спутников и ракет происходит разрушение и образование «окон» в слое озона, который сильно поглощает ультрафиолетовое излучение Солнца. В эти «окна» возможно свободное проникание ультрафиолета до земной поверхности.

#### 9. Диапазон рентгеновского и гамма-излучения (меньше 0,1 мкм).

Диапазон рентгеновского излучения широко используется в медицине, а также в технике для дефектоскопии.

Такова краткая характеристика диапазонов электромагнитных волн. При использовании того или иного диапазона обычно исходят из того, насколько легко получить соответствующее излучение и как оно взаимодействует с веществом. При этом постоянно обнаруживаются новые перспективы. Например, раньше не могли использовать для дальней связи световой сигнал, однако создание волоконной оптики и лазеров сделало это очень перспективным. По специальным стеклянным волокнам (световодам) луч света за счет многократного полного отражения практически без потерь распространяется на большие расстояния. Самое главное преимущество использования света для связи заключается в том, что в этом диапазоне может работать без помех громадное число передатчиков, так как длина световой волны очень мала. В настоящее время предложены межконтинентальные световые кабели. Очевидно, что в будущем светопроводы заменят традиционные средства связи и, в частности, проводную связь.

Выше было показано, что электромагнитная волна возникает вокруг проводников, по которым проходят переменные токи. В § 7.5 было отмечено, что поле излучения создается и любым ускоренно движущимся зарядом. Эта возможность используется в

работающих в сверхвысокочастотном диапазоне электронных лампах (клистролах), в которых генерация обеспечивается за счет модуляции скорости электронов, в рентгеновских трубках, в которых пучок электронов направляется на массивный металлический анод, в поверхностном слое которого и происходит быстрое торможение электронов, в ускорителях электронов для получения так называемого синхротронного излучения. Электромагнитное излучение возникает также в результате внутриатомных, внутримолекулярных (см. § 9.1) и внутриядерных (гамма-излучение) процессов.

### 8.3. Волновые явления в оптике. Интерференция

Представление о свете как о потоке электромагнитных волн используется при рассмотрении таких явлений, как интерференция, дифракция, дисперсия, поляризация и некоторые другие. Однако такое представление совершенно неприемлемо, например, при объяснении фотоэффекта. В связи с этим заметим, что в оптике, как и в некоторых других разделах физики, также можно говорить о макроскопических и микроскопических явлениях. К первым относится интерференция. Ко вторым—явления, в которых четко проявляются особенности процессов на уровне одиночных атомов. Чтобы понять эти процессы, необходимо ввести понятие фотона, или кванта света, и представлять свет как поток таких своеобразных частиц. Мы уже говорили о квантах поля, когда рассматривали вопросы о дальнем действии и близком действии в физике. Если результат эксперимента связан с совокупным действием очень большого числа квантов, то можно производить своеобразное усреднение и переходить к волновому описанию. Амплитуды составляющих  $E$  и  $B$  в электромагнитной волне, т. е.  $|E|$  и  $|B|$ , полагаются тогда пропорциональными плотности фотонов в единице объема.

Сходная ситуация наблюдалась и при изучении свойств газа. Если в нем имеется всего одна частица, то целесообразно говорить о ее траектории и силе, с которой эта частица действует на данный участок стенки сосуда, в котором находится газ. Если частиц в сосуде много, то вся картина усредняется и, где бы ни измерялась сила, с которой частицы действуют на данный участок стенки, результат будет один и тот же независимо от длительности времени измерения. В этом случае вводится понятие давления как результата усредненного действия большого числа частиц и появляется новая физическая характеристика.

Рассмотрим явления, в которых целесообразно использовать понятие волны. К таким явлениям относятся отражение, преломление, интерференция и дифракция света.

При рассмотрении интерференции света будем базироваться на рассуждениях, которыми мы пользовались при изучении механических волн. Пусть в точке  $A$  встречаются волны,

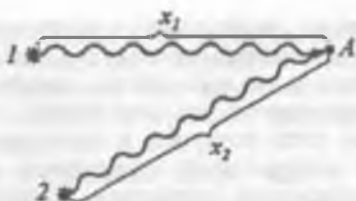


Рис. 8.7. Схема опыта по интерференции двух волн. Волны исходят из источников 1 и 2 и встречаются в точке А

распространяющиеся от источников 1 и 2 (рис. 8.7). Если источники света когерентны (частоты одинаковы и разность фаз волн одна и та же), то в точке А можно наблюдать при одинаковых поляризациях возрастание амплитуды волны или ее уменьшение. При этом следует помнить, что энергия никуда не исчезает, а перераспределяется в пространстве. Записывая волну аналитически в виде

$$y = A \sin \omega(t - x/c)$$

и используя условия когерентности  $\omega_1 = \omega_2$  и  $x_1/c - x_2/c = \text{const}$ , получаем следующие соотношения соответственно для минимума и максимума амплитуды суммарной волны:

$$x_1 - x_2 = (2n + 1) \frac{\lambda}{2} \quad (n = 0, 1, 2, 3, \dots),$$

$$x_1 - x_2 = n\lambda \quad (n = 1, 2, 3, \dots).$$

Для сохранения устойчивости картины необходимо обеспечить сохранение во времени когерентности источников и постоянства разности хода  $\Delta x$ . Раньше создать когерентные источники было нельзя (в лампах накаливания такой когерентности нет). Эта возможность возникла лишь после появления в 1960 г. лазера. До этого для наблюдения интерференционной картины в оптике использовался прием, который может быть назван приемом разделения фронта волны. Этот прием, реализованный в большом числе опытов, из которых мы рассмотрим лишь некоторые, заключается в следующем.

Вспомним принцип Гюйгенса. Согласно этому принципу, каждая точка волнового фронта является источником сферических волн, причем начальные фазы и поляризация этих волн одинаковы. Другими словами, такие точки фронта можно рассматривать как когерентные источники света. В результате интерференции волн от таких источников образуется новый фронт волн. Интерференционные явления в чистом виде удастся наблюдать в том случае, если с помощью какого-либо приема

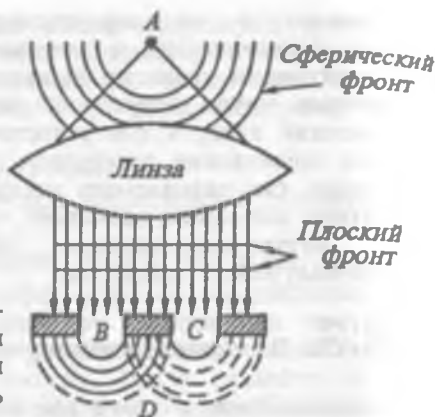


Рис. 8.8. Схема опыта Юнга

выделяется на фронте волны только небольшое число точек. Это можно сделать, например, следующим образом.

Предположим, что источник излучения расположен в точке *A*. Этот источник испускает совокупность сферических волн (рис. 8.8). Поставим на пути сферических волн линзу. Тогда сферическую волну можно превратить в плоскую волну, распространяющуюся дальше так, как показано на рисунке. Разместим на пути волны экран с двумя отверстиями малых размеров. Эти отверстия выделяют на фронте падающей на экран волны два небольших участка *B* и *C*. По принципу Гюйгенса каждая из точек этих участков будет источником электромагнитной волны, причем, поскольку мы выделили с помощью экрана участки одного и того же фронта, начальные фазы волн, исходящих из каждого отверстия, будут одни и те же. Если рассмотреть совокупное действие волн в некоторой точке *D* за экраном, то мы будем наблюдать явление интерференции этих волн. В частности, если в качестве отверстий взять две узкие щели, то на экране при движении точки *D* будет наблюдаться чередование темных и светлых полос. Это чередование возникает благодаря тому, что когда точка *D* движется в горизонтальном направлении, то расстояние до этой точки от каждой из щелей будет все время меняться и, следовательно, будет меняться соотношение фаз за счет разного времени прохождения светом путей *DC* и *DB* от двух щелей. Почему же в этом случае, хотя источник света по-прежнему излучает большой набор различных волн, все-таки удается наблюдать интерференцию света? Причина этого заключается в следующем.

Рассмотрим одну какую-нибудь волну, излученную источником. Эта волна, достигнув экрана, даст две волны, исходящие из щелей *B* и *C*. Щели отделяют на фронте выбранной нами волны два участка, которые служат когерентными источниками света. В результате возникает интерференция волн, исходящих от двух щелей. Если мы возьмем другую волну, исходящую также от первоначального источника, то снова возникнет интерференционная картина за счет того, что интерферируют в данном случае волны, исходящие из различных участков фронта другой первоначальной волны. В результате в среднем по времени будет наблюдаться постоянная картина.

Итак, явление интерференции будет наблюдаться, если между собой интерферируют различные участки фронта одной и той же первоначальной исходной волны. Для этого необходимо осуществить опыт по разделению первоначального фронта волны. Подобные опыты можно поставить самым разнообразным образом, например, с помощью зеркал Френеля. Здесь разделение фронта волны происходит так. Выбираются два плоских зеркала, которые ставятся под небольшим углом друг к другу (рис. 8.9). На зеркала падают плоские электромагнитные волны. При падении плоской волны на зеркала Френеля она разделится на две: одна отражается от одного зеркала Френеля, а вторая—от

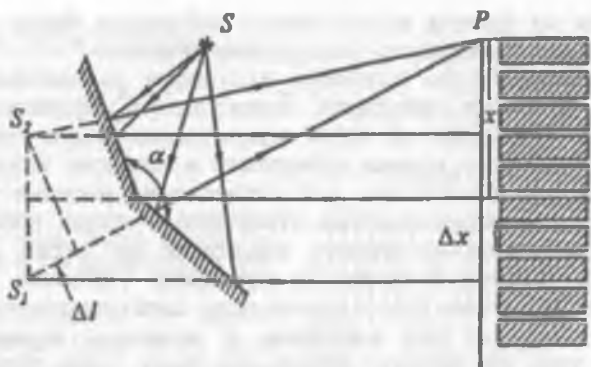


Рис. 8.9. Интерференция на зеркалах Френеля:

$S_1$ ,  $S_2$ —мнимые источники волн,  $S$ —действительный источник,  $\Delta l$ —разность хода лучей от мнимых источников. Справа показана интерференционная картина

второго. В результате получается пара электромагнитных волн, отраженных по различным направлениям. В ряде точек эти две отраженные волны пересекаются. Так как каждая из волн образовалась в результате разделения фронта первоначальной волны, то эти волны обладают свойством когерентности и в области пересечения волн будет возникать интерференционная картина, состоящая из чередующихся темных и светлых полос.

Явление интерференции можно наблюдать и на целом ряде повседневных опытов. В качестве очень распространенного примера интерференции можно привести появление радужной окраски на поверхности воды при растекании по ней слоя нефти или бензина. В этом случае первоначальная электромагнитная волна, попадая на тонкий слой бензина на поверхности воды, частично отражается от него, частично проходит через слой бензина и отражается от поверхности воды. Эти две волны (одна—отраженная от поверхности бензина, а другая—от поверхности воды) распространяются по одному направлению (рис. 8.10).

Поскольку обе волны образуются в результате разделения фронта первоначально падающей волны, возникают постоянная разность фаз и чередование темных и светлых участков. Это чередование темных и светлых участков для разных длин волн наблюдается под разными углами. Если осветить такую поверхность белым светом, состоящим из совокупности электромагнитных волн разной частоты (разного цвета), то появится радужная окраска.

Сходное явление наблюдается, если мы возьмем два плоских стекла и плотно их сожмем. В тех местах, где стекла за счет неровностей соприкасаются, появляются радужные кольца. Их происхождение похоже на происхождение радужной окраски на поверхности воды, покрытой бензином. В этом случае также происходит интерференция волн, отраженных от разных поверхностей (первого и второго стекол). Такой прием широко

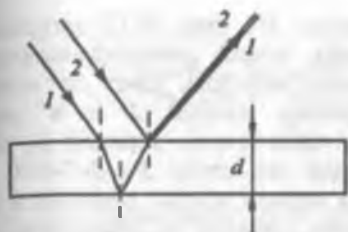


Рис. 8.10. Интерференция на тонкой пластинке. Минимуму отражения отвечает условие  $2d = \lambda/2$

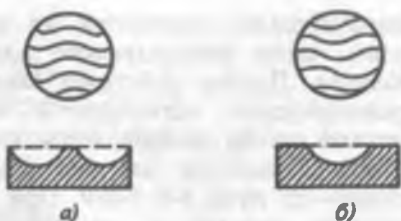


Рис. 8.11. Наблюдение неровностей на поверхности тел

используется в технике для контроля качества полировки стекол. Если одно стекло отполировано очень хорошо и является эталоном, то, прикладывая к нему плотно другое стекло, можно проконтролировать качество полировки. Если полировка плохая и на поверхности контролируемого стекла имеются дефекты, например выпуклости, то при наложении контролируемого стекла на эталонное вокруг этих выпуклостей появятся радужные полосы—кольца Ньютона (рис. 8.11).

Явление интерференции волн, получающихся за счет отражения от разных поверхностей, широко используется и при изготовлении так называемой просветленной оптики. Когда свет падает на стекло, то он частично проходит внутрь него, а частично отражается от его поверхности. Следовательно, только часть света, падающая на объектив, пройдет сквозь него и даст фотографическое изображение. Чем больше света отразится от объектива, тем менее яркая и четкая получится фотографическая картина за объективом. Уменьшить отражение можно следующим образом.

Предположим, что на поверхность стекла нанесена тонкая пленка из специально подобранного вещества, обладающая тем свойством, что лучи, отражающиеся от ее поверхности и от поверхности стекла, после прохождения пленки интерферируют так, что гасят друг друга. В этом

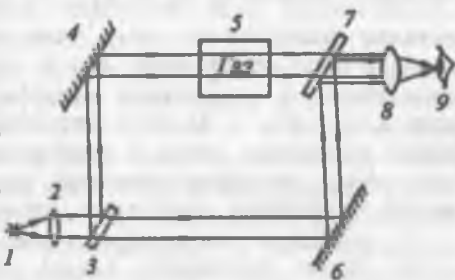


Рис. 8.12. Схема интерферометра

случае можно утверждать, что по закону сохранения энергии доля прошедшего внутрь объектива светового потока увеличивается пропорционально уменьшению отраженного потока. Таким образом удастся увеличить степень пропускания света объективом, т. е. просветлить его.

Явление интерференции и специальные приборы—интерферометры используются для точного измерения показателя преломления веществ, особенно газов, где показатель преломления лишь

незначительно отличается от единицы. На рис. 8.12 показана схема часто используемого для этих целей *интерферометра Жамена*. Прибор действует следующим образом. Свет от монохроматического источника 1 собирается линзой 2 и в виде плоской волны направляется на полупрозрачное зеркало 3, где волна разделяется на две: одна идет по пути 3-4-5-7-8-9, а вторая—по пути 3-6-7-8-9. При этом первая волна, отражаясь от плоского зеркала 4, проходит через кювету с исследуемым газом и попадает на полупрозрачное зеркало, а затем фокусируется линзой 8 на приемник излучения 9. Вторая волна, отражаясь от плоского зеркала 6 и полупрозрачного зеркала 7, также направляется на приемник. На приемнике две волны накладываются друг на друга и наблюдается интерференция. Если вначале кювета с газом отсутствовала и длины плеч интерферометра (расстояния 3-4-7 и 3-6-7) были одинаковыми, то разность хода волн, когда они достигнут приемника излучения, будет равна нулю. В этом случае наблюдается максимум в интерференционной картине.

Если поместить в одно из плеч кювету с веществом, то возникает определенная разность хода, приводящая к изменению интерференционной картины. Почему изменится разность хода, легко понять, если учесть, что скорость света в среде меньше, чем в вакууме (или атмосфере). Тогда  $v = c/n$ , где  $v$ —скорость света и  $n$ —показатель преломления в изучаемой среде.

Если  $d$ —толщина слоя среды с другим показателем преломления, то запаздывание одной волны по отношению к другой составит, очевидно,  $\Delta t = d \left( \frac{1}{v} - \frac{1}{c} \right) = \frac{d}{c} (n - 1)$ . Такое запаздывание равносильно тому, что свет с прежней скоростью  $c$  пройдет еще расстояние  $d(n - 1)$ . Эта величина и составит приводящую к изменению интерференционной картины разность хода  $\Delta x = d(n - 1)$  двух интерферирующих лучей. Зная длину волны источника света и подбирая  $d$  таким образом, чтобы от максимума интерференционной картины перейти к минимуму, можно определить показатель преломления  $n$ .

С помощью интерферометров разной конструкции можно очень точно сравнивать длины эталонов. В частности, таким образом производится сравнение длин национальных эталонов одного метра с международным, хранящимся в Палате мер и весов в Париже.

## 8.4. Фурье-спектроскопия

Не только в перечисленных выше приложениях возможно практическое использование интерферометров. На основе этих приборов были созданы *фурье-спектрометры*. Появление таких спектрометров надо расценивать как одно из крупных достижений в оптике.

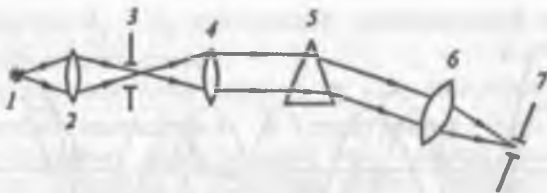


Рис. 8.13. Схема классического спектрометра:

1—источник света; 2—объектив, фокусирующий изображение источника на входную щель 3; 4—объектив; 5—призма, разлагающая исходный свет в спектр; 6—объектив, фокусирующий изображение входной щели на выходную 7

Чтобы понять преимущества фурье-спектрометров, вспомним, что представляют собой обычные спектрометры. Схема простейшего из них изображена на рис. 8.13. Свет от источника 1 собирается линзой 2 и фокусируется на входную щель 3, на которой получается изображение источника. Проходящий через щель 3 свет превращается линзой 4 в нерасходящийся пучок и пропускается через призму 5, где испытывает преломление, и затем линзой 6 фокусируется на фотопластинку или выходную щель 7 так, что в фокальной плоскости линзы 6 получается изображение входной щели 3. Если источник света испускает, например, две волны разной длины ( $\lambda_1$  и  $\lambda_2$ ), то из-за разного коэффициента преломления изображения входной щели окажутся в разных местах в фокальной плоскости линзы 6. Тогда мы зарегистрируем либо два изображения входной щели на фотопластинке (так получаются *линейчатые спектры*), либо обнаружим последовательно два сигнала, если выходная щель 7 будет перемещаться вдоль фокальной плоскости (или поворачиваться вокруг вертикальной оси призмы 5).

С помощью такого приема можно исследовать спектральный состав изучаемого источника света или, ставя на пути света кювету с веществом, изучать, какие волны и в какой степени этим веществом поглощаются. В свою очередь, последнее позволяет изучать строение вещества на атомно-молекулярном уровне.

Преимущество таких спектрометров, называемых классическими, заключается в простоте конструкции и в возможности получения сведений о спектральном составе света практически без всякой дополнительной обработки. Однако эти приборы обладают принципиальным недостатком, связанным с тем, что для извлечения полезной информации о веществе только часть светового потока пропускается через входную щель. Щель эта очень узкая (доли миллиметра), так как при широких щелях изображение щели после линзы 6 на волне  $\lambda_1$  накладывается на изображение на волне  $\lambda_2$  и исчезнет возможность судить о том, с одной или двумя длинами волн мы имеем дело (т. е. уменьшается, как говорят, разрешающая способность спектрометра).



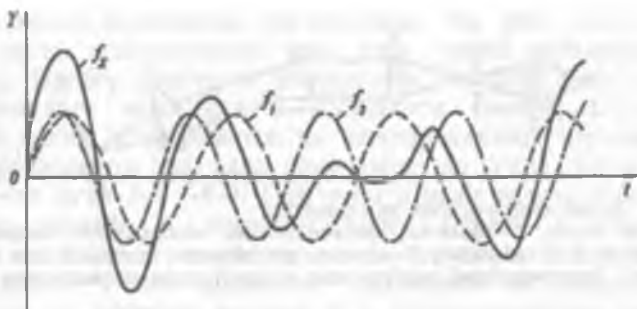


Рис. 8.14. Схема образования результирующего сигнала на регистрирующем устройстве фурье-спектрометра:

$f_1$ ,  $f_2$ —сигналы от каждой из компонент-волн с длинами волн  $\lambda_1$  и  $\lambda_2$ ;  $f_T$ —суммарный сигнал

Через узкую щель пройдет лишь малая доля энергии, испускаемой источником. Поэтому классические спектрометры не приспособлены для анализа слабых излучений или сильно поглощающих сред.

От этого главного недостатка классических спектрометров свободны фурье-спектрометры, где щелей вообще нет.

Чтобы понять их действие, обратимся снова к интерферометру Жамена. Свет в него попадает через широкое окно.

Представим, что либо длина  $d$  кюветы с газом увеличивается линейно ( $d = d_0 t$ ), либо давление газа (известно, что показатель преломления  $n$  пропорционален давлению газа  $p$ ). Тогда линейно во времени меняется и разность хода  $\Delta x$ . На приемнике излучения будет происходить чередование максимумов и минимумов освещенности  $I$ . Эта освещенность, очевидно, меняется периодически с некоторым периодом  $T = \Delta t$ , где  $\Delta t$ —промежуток времени, необходимый для перехода от максимальной освещенности к минимальной. Это время легко определить, зная разность

$$\Delta x_{\min} - \Delta x_{\max} = (2k + 1) \frac{\lambda}{2} - k\lambda = \frac{\lambda}{2} = d_0 \Delta t (n - 1).$$

Значит,

$$T = \Delta t = \frac{\lambda}{2} \frac{1}{d_0 (n - 1)},$$

т. е. период смены «света» и «тьмы» определяется длиной волн  $\lambda$ . Соответствующий сигнал легко преобразуется в синусоидальный!

Если источник света излучает две длины волн, то прибор воспринимает сигнал, равный сумме двух синусоид для волн  $\lambda_1$  и  $\lambda_2$  (рис. 8.14). Сумма двух синусоид периодических сигналов есть также периодический сигнал, который может быть однозначно разложен на составляющие. Это так называемый *фурье-анализ* периодических функций. Регистрируя сумму сигналов, можно

найти параметры  $\lambda_1$  и  $\lambda_2$  для каждой составляющей и произвести спектральный анализ света.

Для определения длин волн  $\lambda_1$  и  $\lambda_2$  необходимо произвести математические вычисления. В спектрометр надо встраивать ЭВМ, что усложняет конструкцию. Это, однако, компенсируется возможностью работать в условиях слабой светимости объектов или сильного поглощения. Сейчас фурье-спектрометры начинают вытеснять классические.

## 8.5. Голография

В 1949 г. Д. Габором в США было открыто также базирующееся на интерференции световых волн явление создания объемных изображений предметов, названное *голографией*. В отличие от других способов получения объемных изображений, голографическое изображение воспроизводит все особенности изображаемого предмета. В частности, наблюдатель не только ощущает, что одни части предмета находятся ближе к нему, чем другие, но и, меняя точку наблюдения (сдвигаясь чуть вправо или влево), может увидеть и те части предмета, которые не видны при центральном и одном из крайних положений глаз. В результате получается полная иллюзия реального объекта.

Чтобы понять принцип построения голографических изображений, рассмотрим следующий упрощенный пример. Пусть в направлении оси  $X$  распространяются от двух когерентных источников две синусоидальные волны  $y_1$  и  $y_2$ , исходящие из точек  $O$  и  $x_1$  на оси  $X$  (рис. 8.15).

Волна  $y_1$  является опорной, а волна  $y_2$ —предметной. На участке за  $x_1$  эти волны могут интерферировать. Будем следить за результатом интерференции в некоторой точке  $x$ . Примем, что

$$y_1 = E_1 \sin \omega(t - x/c)$$

и

$$y_2 = E_2 \sin \omega[t - (x - x_1)/c] = E_2 \sin \omega(t - x/c + x_1/c).$$

Перенеся начало координат в точку  $x$  и обозначив величину  $\omega x_1/c = \varphi$ , получим  $y_1 = E_1 \sin \omega t$  и  $y_2 = E_2 \sin(\omega t + \varphi)$ . В результате интерференции произойдет сложение волн  $y_1$  и  $y_2$ . Пусть теперь в точке  $x$  размещается фотопластинка. Под действием света она засвечивается и после проявления на ней образуется почернение. Плотность почернения определяется интен-

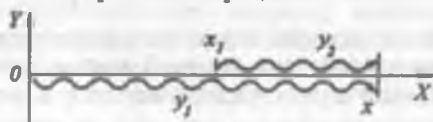


Рис. 8.15. К выводу правила записи голограммы

сивностью света, которая, в свою очередь, пропорциональна квадрату напряженности поля суммарной волны:

$$\begin{aligned} I \sim y^2 &= (y_1 + y_2)^2 = \\ &= E_1^2 \sin^2 \omega t + E_2^2 \sin^2(\omega t + \varphi) + 2E_1 E_2 \sin \omega t \cdot \sin(\omega t + \varphi) = \\ &= E_1^2 \sin^2 \omega t + E_2^2 \sin^2(\omega t + \varphi) + E_1 E_2 [\cos \varphi - \cos(2\omega t + \varphi)]. \end{aligned}$$

После усреднения по времени (пластинка экспонируется очень долго по сравнению с периодом колебаний), принимая во внимание, что средние по периоду значения

$$\overline{\sin^2 \alpha} = \overline{\cos^2 \alpha} = 1/2, \quad \overline{\sin \alpha} = \overline{\cos \alpha} = 0,$$

получим

$$\bar{I} = \frac{1}{2} E_1^2 + \frac{1}{2} E_2^2 + E_1 E_2 \cos \varphi.$$

Предположим, что опорная волна  $y_1$  сохраняется, предметная волна  $y_2$  отсутствует и в точку  $x$  помещается проявленная фотопластинка. Интенсивность прошедшего через пластинку света в непосредственной близости от нее равна  $y^2 = DE_1^2 \sin^2 \omega t$ , где  $D$ —коэффициент пропускания, показывающий, насколько ослабится волна (какую энергию она потеряет) при прохождении засвеченной фотопластинки. Если почернение равно  $k\bar{I}$ , где  $k$ —коэффициент пропорциональности, то  $D = 1 - k\bar{I}$  (чем больше почернение, тем меньше пропускание). Следовательно, сразу за пластинкой

$$y^2 = E_1^2 (1 - k\bar{I}) \sin^2 \omega t$$

или

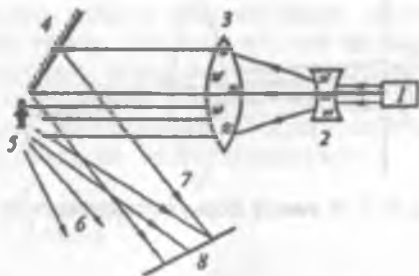
$$y = E_1 (1 - k\bar{I})^{1/2} \sin \omega t = E_1 (1 - \frac{1}{2} k\bar{I}) \sin \omega t.$$

Использовано разложение  $(1 - k\bar{I})^{1/2}$  в степенной ряд и взяты только первый и второй члены. Это вполне допустимо, если  $k\bar{I}$  невелико (в противном случае свет за пластинку практически не пройдет). Воспользовавшись выражением для  $k\bar{I}$  и обозначив постоянные величины коэффициентами  $A$ ,  $B$ , получим

$$\begin{aligned} y &= A \sin \omega t + B \sin \omega t \cos \varphi = \\ &= A \sin \omega t + \frac{1}{2} B \sin (\omega t + \varphi) + \frac{1}{2} B \sin (\omega t - \varphi). \end{aligned}$$

Вклад первого слагаемого регистрируется приемником как результат воздействия опорной волны, второго—в виде света, исходящего от источника, расположенного в точке  $x_1$ , т. е. в своем исходном положении, третьего—в виде еще одного изображения источника

Рис. 8.16. Схема записи голограммы:  
1—лазер; 2—расширитель луча лазера; 3—линза, создающая параллельный пучок лучей; 4—зеркало; 5—предмет; 6—предметная волна; 7—опорная волна; 8—фотопластинка



$x_1$ , но в другом месте. Схема реального опыта показана на рис. 8.16. Монохроматический когерентный пучок света от лазера 1 после расширителя (комбинация вогнутой 2 и выпуклой 3 линз) направляется частично на плоское зеркало 4, а частично на предмет 5. Отраженный от зеркала луч (опорная волна) встречается в месте расположения фотопластинки 8 с волнами, отраженными от предмета. Поскольку как опорная, так и предметная волны происходят от одного лазерного источника и когерентны, то возникнет интерференция. Результат действия интерферирующих волн на пластинку фиксируется в виде потемнений, напоминающих многочисленные налагающиеся отпечатки пальцев. Если убрать предмет 5 и пропустить опорный луч через проявленную и помещенную строго на то же место фотопластинку, то за пластинкой глаз зафиксирует равномерную засветку от опорной волны и волн, как бы исходящих от всех точек предмета, которые ранее были облучены опорной волной и свет от которых в принципе мог попасть на фотопластинку. В результате человек наблюдает объемное изображение предмета.

Применяются и другие схемы регистрации голографических изображений. Голограммы могут быть цветными. Это позволяет делать высококачественные голографические копии музейных экспонатов, например уникальных предметов, найденных при археологических раскопках. При рассматривании таких изображений получается почти полная иллюзия реальных объемных предметов.

## 8.6. Дифракция

Обратимся теперь к дифракционным явлениям. Точной границы между интерференцией и дифракцией провести нельзя. И в том, и в другом случаях наблюдаемая картина возникает в результате сложения волн, идущих от разных участков фронта первоначальной волны, но интерференция—это явление, которое наблюдается в том случае, когда первоначальная волна делится на небольшое число участков и суммируются волны, исходящие от разных участков. В тех случаях, когда складываются волны, исходящие от разных точек одного и того же участка плоской

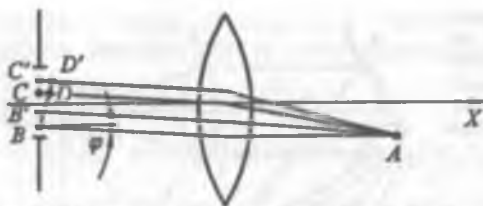
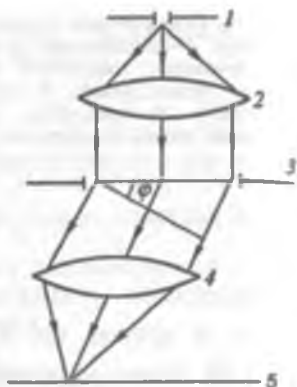


Рис. 8.17. К выводу формулы дифракции на щели

Рис. 8.18. Схема наблюдения дифракции на щели: 1—источник света в форме щели; 2—линза; 3—щель; 4—линза-окуляр; 5—экран



волны, возникает явление, которое принято называть *дифракцией*.

Наиболее простым случаем дифракции является дифракция на щели.

Рассмотрим некоторую плоскую волну, распространяющуюся в направлении  $X$  (рис. 8.17, 8.18). На пути волны поставим экран с щелью шириной  $l$ . Когда волна дойдет до экрана, каждая точка внутри щели будет служить источником сферических волн. Каково результирующее действие этих волн за экраном в некоторой точке  $A$ ? Представим, что за экраном поставлена линза, собирающая лучи в своей фокальной плоскости. Рассмотрим две волны, идущие от точки  $B$  у края щели и некоторой точки  $C$  внутри щели в направлении, определенном углом  $\varphi$ . Если эти две волны собираются с помощью линзы в одной точке, то, поскольку их разность фаз в среднем по времени равна нулю, должны наблюдаться постоянные интерференционные явления. Так как волна, идущая от точки  $B$ , и волна, идущая от точки  $C$ , проходят разное расстояние до места встречи (эта разность равна отрезку  $CD$ ), то в результате интерференции волны складываются так, что интенсивность либо увеличивается, либо уменьшается. Это зависит от пути  $CD$ , определяющего разность хода лучей, идущих от точек  $C$  и  $B$ . Мы считаем, что расстояния от точки  $B$  и точки  $D$  до места встречи одинаковы. Если угол  $\varphi$ , под которым происходит наблюдение, такой, что расстояние  $CD$  равно половине длины волны, то волны, вышедшие из точек  $C$  и  $B$ , должны гасить друг друга.

Пусть точки  $B'$  и  $C'$  сдвинуты вверх от точек  $B$  и  $C$  на одну и ту же величину. Лучи света, вышедшие из точек  $B'$  и  $C'$  под тем же углом  $\varphi$ , снова должны погасить друг друга, если  $C'D' = \lambda/2$ . Это рассуждение можно продолжить, смещая картину вверх параллельно самой себе. Для каждой точки  $B$ ,  $B'$ ,  $B''$  и т. д. может быть найдена сопряженная ей точка  $C$ ,  $C'$ ,  $C''$ , такая, что волны, исходящие от всех этих точек будут гасить друг друга. Если ширина  $l$  щели такова, что удвоенное

расстояние  $CB$  укладывается на ней кратное число раз, то все точки, укладываемые на участке, равном ширине щели, можно разделить на пары сопряженных точек. Это значит, что существует такой угол  $\varphi$ , что в направлении, определяемом этим углом, все волны, исходящие от пар сопряженных точек, погасят друг друга. Световая энергия в таком направлении распространяться не будет. Получается минимум. Условия этого минимума:

$$l = 2 \cdot CB \cdot n = 2 \frac{CD}{\sin \varphi} n = 2 \frac{\lambda}{2 \sin \varphi} n,$$

или

$$l \sin \varphi = n\lambda,$$

где  $n = 1, 2, 3, \dots$

Если длина  $l$  участка не кратна удвоенному расстоянию  $CB$ , то на участке возникает ряд точек, для которых невозможно указать сопряженные им точки. Волны от таких непарных точек уже не гасятся, и в результате наблюдается проходящий поток энергии. Длина отрезка  $CB$  определяется углом  $\varphi$ , следовательно, если менять угол  $\varphi$  при одной и той же длине  $l$ , то будет наблюдаться чередование максимумов и минимумов освещенности. При  $\varphi = 0$  все волны, исходящие от различных точек, лежащих на щели, складываются, усиливая друг друга, поскольку в этом случае нет разности фаз между лучами, исходящими от разных точек. Следовательно, в перпендикулярном направлении за экраном будет максимум интенсивности света. Под любым другим углом наблюдается либо минимум интенсивности света, либо ее максимум, но с интенсивностью значительно меньшей, чем интенсивность центрального максимума, потому что в этом случае только часть точек, лежащих на щели, будут источниками волн, которые друг друга не гасят. Чем больше будет таких точек, тем больше интенсивность света. Наибольшее количество подобных точек возникает в том случае, если участок  $CB$  укладывается на длине  $l$  нечетное число раз. В этом случае всегда остается целый участок ширины щели, все точки которого не имеют пары и волны от которых не гасят друг друга. Следовательно, для появления максимальной интенсивности необходимо выполнение условия

$$l = (2n + 1) \cdot CB = (2n + 1) \frac{\lambda}{2 \sin \varphi},$$

или

$$l \sin \varphi = (2n + 1) \frac{\lambda}{2},$$

где  $n = 1, 2, 3, \dots$

Если по вертикальной оси отложить интенсивность света, а по горизонтальной—координаты расположения точки на экране,

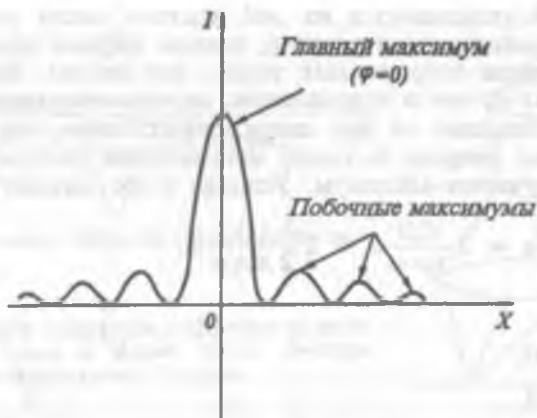


Рис. 8.19. Распределение интенсивности света на экране за щелью при дифракции на ней

где производится наблюдение дифракции, то получится график, показанный на рис. 8.19.

Сделаем ряд выводов.

Если  $l$  значительно меньше  $\lambda$ , то, как следует из формулы для минимума,  $\sin \varphi = n\lambda/l$  возрастает (в пределе до единицы),  $\varphi$  стремится к  $90^\circ$ . Происходит значительное расширение центрального максимума и медленный спад интенсивности центрального максимума к краям.

При  $l$  очень большом  $\sin \varphi = n\lambda/l$  стремится к нулю. Это значит, что получается резкая яркая линия только в центре экрана. Это указывает на условие прямолинейности распространения луча света. В самом деле, когда  $l$  велико и велико количество точек, являющихся источниками сферических волн, в результате интерференции этих волн происходит почти полное гашение отходящих в сторону лучей. Следовательно, свет распространяется прямолинейно. Если отверстие в экране достаточно мало, например порядка долей миллиметра, то наблюдается дифракционная картина и резкая смена границ света и тени не получается, т. е. ярко проявляется волновая природа света.

Заметим, что угол  $\varphi$ , под которым наблюдается максимум или минимум в интерференционной картине, зависит не только от  $l$  и числа  $n$ , но и от длины волны  $\lambda$ . Это значит, что если на экран одновременно падают две плоские волны, имеющие разную длину, то получаются две интерференционные картины, сдвинутые друг относительно друга на величину, зависящую от разности длин волн (рис. 8.20). Другими словами, можно разделить в пространстве волны разной длины и провести анализ спектрального состава света.

Назовем *монохроматическим светом* такой поток электромагнитных волн, в котором все волны имеют одну и ту же длину или частоту. *Немонохроматическим светом* называют

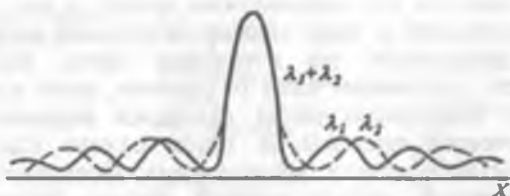


Рис. 8.20. Разделение волн  $\lambda_1$  и  $\lambda_2$  в пространстве при дифракции на щели

поток волн разной длины или частоты. Так как волны одной частоты, воздействуя на наши органы зрения, дают вполне определенное цветовое восприятие, то свет одного цвета является монохроматическим. Свет, состоящий из смеси цветов, в частности солнечный,—белый свет, является немонохроматическим светом. Монохроматический свет в результате дифракции на щели дает определенную дифракционную картину, немонохроматический свет при дифракции на щели дает наложение дифракционных картин и наблюдается радужное отклонение на границе света и тени. Этот эффект можно наблюдать, если взять перо от птицы и посмотреть через него на солнце, слегка покачивая перо перед глазом. Мы увидим радужные блики.

Представим, что имеется круглое отверстие на экране. Общая картина распределения максимумов и минимумов при дифракции на таком отверстии похожа на картину распределения максимумов и минимумов в случае щели (см. рис. 8.19).

Ширина центрального пятна зависит от диаметра  $l$  отверстия экрана. Предположим, что имеется два малых источника света, изображение которых с помощью системы линз проектируется на экран (рис. 8.21). Так как диаметры линз не бесконечны, а имеют оправу, ограничивающую поток света, то должны наблюдаться дифракционные явления на оправках линз. Даже, если источники света малы и имеют четкие границы, их изображения (рис. 8.21) будут несколько размазанными. Другими словами, четкость изображения в результате дифракции на оправках объективов может оказаться значительно меньшей, чем четкость самих источников света. При сближении двух источников света сближаются и их изображения. Поскольку изображения размыты, то начиная с

Распределение интенсивностей  
точечных источников света

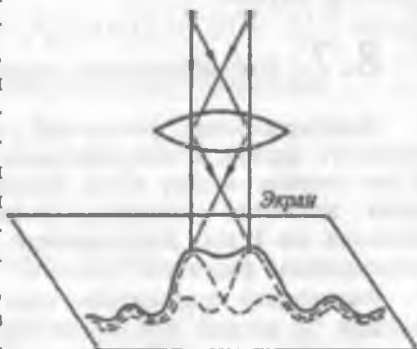


Рис. 8.21. Схема образования дифракционного предела разрешающей способности линзы



некоторого момента уже невозможно судить о том, один или два источника находятся в поле зрения оптической системы. И так, в результате дифракции два источника света, находящиеся на очень близком расстоянии один от другого, даже при абсолютном совершенстве объективов могут оказаться неразличимы в поле зрения оптической системы. В таком случае говорят, что два источника света перестали *разрешаться*. Следовательно, существует определенный дифракционный предел разрешения любой оптической системы.

Так как расстояние между максимумом и минимумом дифракционной картин связано с длиной волны и граница света и тени становится более резкой при уменьшении длины волны ( $\varphi \sim \lambda$ ), то разрешающая способность прибора увеличивается с уменьшением длины волны. Поэтому, например, оптические микроскопы из-за дифракции имеют ограниченный предел разрешающей способности или увеличения, имеющего для лучших систем порядок 2,5—3 тыс. раз. Разрешающая способность увеличивается, если перейти от видимого к ближнему ультрафиолетовому диапазону.

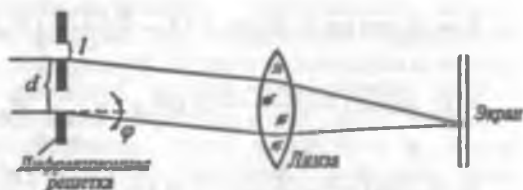
Свойство световой волны дифрагировать на щели позволяет понять и еще один факт, а именно способность параболических антенн концентрировать радиоволны в определенном направлении. Зеркало антенны можно рассматривать как поверхность фронта волны. Фокусировка будет хорошей только тогда (т. е. только тогда будут выполняться правила геометрической оптики), когда размеры антенны существенно больше длины излучаемой волны. Так как на самолетах, например, нельзя поставить большие антенны, то их радиолокаторы работают на длинах волн сантиметрового диапазона. Такие волны заметно ослабляются парами воды в атмосфере. Поэтому обнаружение объектов на больших расстояниях становится затруднительным. Задачу дальнего обнаружения решают наземные радиолокаторы, работающие в метровом диапазоне и имеющие в связи с этим значительные размеры антенн.

## 8.7. Дифракционная решетка

Дифракционной решеткой называется периодическая совокупность щелей в непрозрачном экране. Число таких щелей на 1 мм экрана может быть значительным. Стандартными решетками являются решетки, имеющие 300, 600, 1200 и 1800 штрихов на 1 мм. Изготовление таких решеток осуществимо при современном развитии техники.

Рассмотрим дифракцию сразу на нескольких щелях. Пусть каждая из щелей имеет ширину  $l$  и расстояние между краями щелей (период решетки) равно  $d$  (рис. 8.22). Предположим, что на решетку падает плоская волна. Каждая из щелей вырезает определенный участок фронта плоской волны. Если ширина  $l$  велика по сравнению с длиной волны, то на каждую из щелей

Рис. 8.22. К выводу формулы дифракционной решетки



укладывается большое число точек, являющихся источниками сферических волн. Каждая из щелей дает свет, исходящий под разными углами  $\varphi$  к направлению нормали к решетке.

Выше мы рассматривали явление дифракции на одной щели и видели, что при распространении лучей света под определенными углами, определяемыми соотношением  $l \sin \varphi = n\lambda$ , происходит гашение света и образуется минимум, т. е. в таких направлениях свет от щели не распространяется. Совершенно ясно, что это справедливо и для нескольких щелей. Таким образом, расположение минимумов, называемых главными минимумами, не зависит от того, сколько щелей на экране.

Однако надо еще учесть результат интерференции лучей, исходящих от разных щелей. Если рассматривать интерференцию лучей, исходящих от излучающих точек разных щелей, но расположенных в одном и том же месте внутри каждой щели (рис. 8.23), то могут появиться дополнительные максимумы и минимумы. Если свет распространяется под углом  $\varphi$  к направлению нормали, то между лучами, исходящими от точек  $A, A', A'', A'''$  и т. д. разных щелей, будет существовать постоянная разность фаз. Действие этих лучей в некоторой точке экрана можно определить, вычисляя сумму

$$S = E_0 \cos \omega t + E_0 \cos \omega \left( t - \frac{d}{c} \sin \varphi \right) + \\ + E_0 \cos \omega \left( t - \frac{2d}{c} \sin \varphi \right) + \dots + E_0 \cos \omega \left[ t - (N - 1) \frac{d}{c} \sin \varphi \right],$$

где  $N$ —число щелей. Для этого воспользуемся формулой

$$\sum_{m=0}^{n-1} \cos(x - my) = \cos \left( x - \frac{n-1}{2} y \right) \sin \frac{ny}{2} \operatorname{cosec} \frac{y}{2},$$

положив  $x = \omega t$  и  $y = (\omega/c)d \sin \varphi$ . Тогда

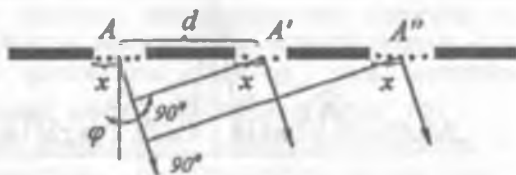


Рис. 8.23. Схема дифракции на решетке

$$\begin{aligned}
 S &= E_0 \cos \omega t + E_0 \cos \omega \left( t - \frac{d \sin \varphi}{c} \right) + E_0 \cos \omega \left( t - \frac{2d \sin \varphi}{c} \right) + \dots \\
 &+ E_0 \cos \omega \left( t - \frac{N-1}{c} d \sin \varphi \right) = \sum_{m=0}^{N-1} E_0 \cos \left( \omega t - m \frac{\omega}{c} d \sin \varphi \right) = \\
 &= E_0 \cos \left( \omega t - \frac{N-1}{2} \frac{\omega}{c} d \sin \varphi \right) \sin \left( \frac{N}{2} \frac{\omega}{c} d \sin \varphi \right) \operatorname{cosec} \left( \frac{\omega d \sin \varphi}{2c} \right).
 \end{aligned}$$

Эта сумма будет обращаться в нуль, если обращается в нуль  $\sin \left( \frac{N}{2} \frac{\omega}{c} d \sin \varphi \right)$ , т. е. при значениях аргумента  $\frac{N}{2} \frac{\omega}{c} d \sin \varphi$ , равных  $0, \pi, 2\pi, 3\pi, 4\pi$  и т. д.

Так как  $\omega/c = 2\pi/\lambda$ , где  $\lambda$ —длина волны, то для углов  $\varphi$ , при которых будут наблюдаться дополнительные минимумы, получим соотношение

$$\frac{N}{2} \frac{2\pi}{\lambda} d \sin \varphi = \pi m \quad \text{или} \quad d \sin \varphi = m \frac{\lambda}{N} \quad (m = 0, 1, \dots).$$

Таким образом, между максимумами и минимумами дифракции от одной щели наблюдаются добавочные максимумы и минимумы, возникающие в результате интерференции света, исходящего от разных щелей (рис. 8.24). Это означает, что чередование максимумов и минимумов в дифракционной картине будет более резким. Следовательно, если на решетку падают две волны с равными длинами волны, то в результате дифракции разделение этих волн будет значительно более четким, что удобнее для наблюдения. Именно поэтому дифракционная решетка и применяется в качестве одного из основных элементов в спектрометрах.

Чем больше штрихов в дифракционной решетке, тем острее максимум (ширина максимума пропорциональна  $1/N$ ). Кроме того, чем больше число штрихов на 1 мм, тем сильнее размыта дифракционная картина, получающаяся в результате действия одной щели, а это значит, что максимумы при разных длинах волн будут разводиться на большую величину. Именно поэтому для увеличения разрешающей способности решетки, т. е. для более легкого наблюдения спектрального состава света, необходимо выбирать решетки, во-первых, с большим числом штрихов на 1 мм (это обеспечивает узость щели) и, во-вторых, больших размеров, т. е. с большим числом  $N$ .

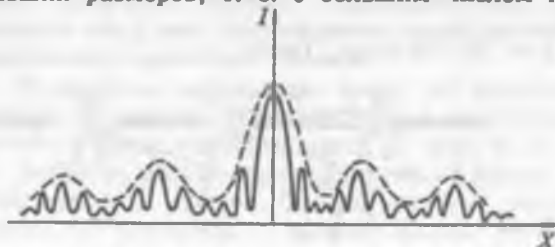


Рис. 8.24. Дифракционная картина за решеткой

Простейшие дифракционные решетки изготовляют, нанося алмазным резцом тонкие штрихи на поверхность стеклянной пластины. В этом случае роль щелей играют промежутки между штрихами. Эти решетки обладают, однако, недостатком: значительная часть света бесполезно теряется, так как в таких решетках происходит распределение света в правую и левую стороны по отношению к первичному пучку. Большая часть светового потока проходит также прямо через дифракционную решетку и не участвует в создании полезного сигнала. Поэтому пользуются *отражательными решетками* или *эшеллетами*, поверхность которых изготавливается в виде чередующихся ступенек определенного профиля. С помощью такой решетки можно добиться, чтобы практически весь световой поток (до 80%) отражался под определенным углом.

Существуют и другие типы решеток. Например, решетки, у которых отражающая поверхность изогнута, позволяют осуществлять и фокусировку лучей.

В настоящее время возрос интерес еще к одному типу дифракционных решеток. Как известно, оптические приборы имеют большое число линз, задача которых состоит в том, чтобы преобразовывать фронт волны для получения изображения. Такие устройства довольно громоздки и велики. Кроме того, линзы из стекла сильно поглощают инфракрасные волны, и поэтому при работе в инфракрасном диапазоне необходимо использовать специальные материалы, которые или дороги, или обладают очень плохими механическими свойствами и требуют специальных условий при хранении и эксплуатации. Оказывается, что если изготовить кольцевую дифракционную решетку (*зонную пластину*), состоящую из прозрачных и непрозрачных колец, то в результате сложной интерференции от этих колец происходит концентрация световой энергии к центру. Эта решетка обладает фокусирующими свойствами и ее можно использовать подобно линзе. Она легка, имеет малые размеры и изготавливается из дешевых материалов, например из полимеров. Такие пластины можно сделать работающими как на прохождение, так и на отражение. Эта новая область специальной оптики, которая сейчас успешно развивается, получила название *дифракционной оптики*. В частности, в радиотехнике такой принцип используется для создания направленных излучателей (*зонные антенны*).

Введем понятие прямых и обратных задач дифракции. Под *прямыми задачами* понимают расчет дифракционных картин (распределение в пространстве максимумов и минимумов и их интенсивностей) по заданной длине волны и заданным свойствам экрана. Например, для данной дифракционной решетки с периодом  $d$  при использовании излучения с длиной волны  $\lambda$  и для данного порядка  $n$  рассчитать угол  $\varphi_n$ , под которым наблюдается соответствующий минимум или максимум.

*Обратная задача*—это определение либо длины волны  $\lambda$ , либо свойств экрана по известному распределению максимумов и

минимумов, т. е. на основе наблюдения дифракционной картины (например, по известным  $\varphi$ ,  $\lambda$  и  $n$  найти период дифракционной решетки  $d$ ).

Как известно, одной из важнейших задач науки является определение расстояния между атомами твердого тела и установление кристаллической структуры. Эту задачу можно решить, если просвечивать кристалл рентгеновским излучением. Так как кристалл представляет собой совокупность атомов, образующих трехмерную структуру, то в результате дифракции на таком трехмерном экране (ее можно зафиксировать на фотопленке) будет также наблюдаться чередование темных и светлых пятен. Измерив положения этих пятен, можно найти параметры исследуемой структуры—длины связей, валентные углы. Длина волны рентгеновского излучения примерно такого же порядка, что и расстояние между атомами в кристаллах. Это является условием четкости дифракционной картины. Решение прямых и обратных дифракционных задач составляет основу математической обработки результатов измерений с целью получения сведений о пространственном расположении атомов в кристалле.

Одно из самых крупных открытий XX в.—определение пространственной структуры молекулы ДНК—было сделано именно на основе рентгеноструктурного анализа. Имеются данные о пространственном строении нескольких сотен белковых молекул. Эти данные послужили основой развития молекулярной биологии и генной инженерии.

Отметим, что дифрагировать на кристаллах могут не только рентгеновские лучи, но и электроны и нейтроны. Поэтому существуют методы электронного и нейтронного микроструктурного анализа.

В качестве еще одного очень интересного примера проявления дифракции и связи принципа Гюйгенса и теории дифракции рассмотрим действие упомянутой выше зонной пластинки Френеля. Одновременно еще раз покажем, при каких условиях выполняются правила геометрической оптики.

Пусть в направлении оси  $X$  распространяется плоская волна, фронт которой совпадает с плоскостью  $YOZ$  (см. рис. 8.25).

Восстановим к этой плоскости перпендикуляр, который примем за ось  $X$ . Волна распространяется в положительном направлении оси  $X$ . Выберем на оси  $X$  точку  $A$  такую, что расстояние  $OA = L \gg \lambda$ , где  $\lambda$ —длина волны. Проведем с центром в точке  $A$  ряд сфер (на рис. 8.25,а показаны кривые пересечения этих сфер с плоскостью  $ZOY$ ; очевидно, что получится ряд окружностей) радиусами  $R_0 = L$ ;  $R_1 = L + l$ ;  $R_2 = L + 2l$ ;  $R_3 = L + 3l$  и т. д., где  $l$ —постоянная величина такая, что  $l \ll L$ .

Нулевая сфера будет касаться фронта волны в точке, принятой за начало координат. Первая сфера пересечет фронт волны и геометрическим местом этого пересечения будет окруж-

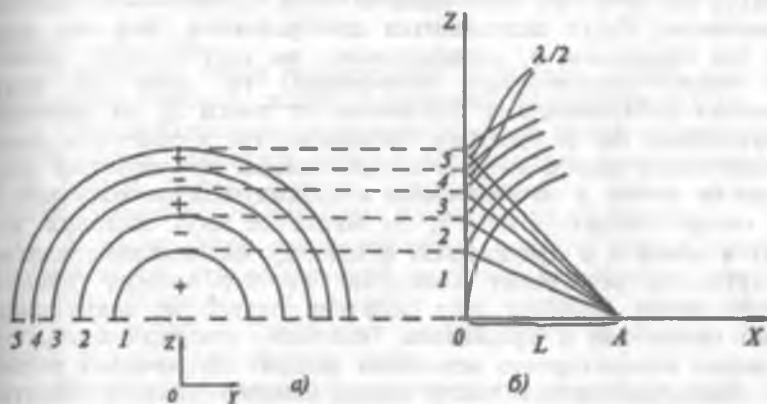


Рис. 8.25. Построение зон Френеля. Цифрами 2, 3,... обозначены соответствующие кольцевые зоны Френеля. Номер 1 относится к центральной зоне. Слева показаны верхние полукольца зон Френеля в плоскости фронта волны YOZ. Справа проиллюстрировано правило построения зон Френеля (плоскость ZOX)

ность с центром в точке  $O$  с радиусом  $r_1 = O1 = \sqrt{R_1^2 - L^2}$ . След второй сферы на плоскости фронта волны также образует концентрическую первой окружность с радиусом  $r_2 = O2 = \sqrt{R_2^2 - L^2}$ . Аналогично для других окружностей на рис. 8.25,а получим:  $r_3 = O3 = \sqrt{R_3^2 - L^2}$ ;  $r_4 = O4 = \sqrt{R_4^2 - L^2}$  и т. д.

Учтем, что  $l \ll L$ . Тогда  $R_k^2 = (L + kl)^2 = L^2 + 2Lkl$  с хорошей степенью точности. Напомним, что здесь использована формула  $f(x + \Delta x) = f(x) + (df/dx)\Delta x$ . Эта формула верна при  $\Delta x$  малом. Производная  $df/dx$  вычисляется в точке  $x$ .

Тогда можно записать  $r_1 = \sqrt{2Ll}$ ,  $r_2 = \sqrt{4Ll}$ ,  $r_3 = \sqrt{6Ll}$  и т. д.

Вычислим теперь площадь центрального круга и каждого из колец (они называются зонами Френеля). Получим

$$s_0 = \pi r_1^2 = \pi \cdot 2Ll; \quad s_1 = \pi r_2^2 - \pi r_1^2 = \pi \cdot 2Ll; \quad s_2 = \pi r_3^2 - \pi r_2^2 = \pi \cdot 2Ll$$

и т. д. Таким образом, площади всех зон оказываются одинаковыми! Это очень важное свойство. Теперь воспользуемся принципом Гюйгенса. Согласно этому принципу каждая точка исходного фронта волны (малый элемент площади) является источником распространяющихся в направлении оси  $X$  сферических волн, причем все такие элементарные источники когерентны. Сферические волны от всех зон Френеля встретятся в точке  $A$  и,

поскольку к тому же они должны быть одинаковым образом поляризованы, будут складываться алгебраически. Так как площади зон одинаковы и, следовательно, им соответствует одинаковое число элементарных источников, то, если бы зоны находились на одинаковом расстоянии от точки  $A$ , их действие просто свелось бы к простому сложению. Но в том-то и дело, что, как легко видеть на рис. 8.25,б, каждая последующая зона удалена от точки  $A$  по сравнению с предыдущей на величину  $l$ . Если теперь выбрать  $l = \lambda/2$ , то волны от двух соседних зон придут в точку  $A$  в противофазе и поэтому максимально ослабят друг друга, но не погасят полностью. То, что полного гашения действия одной зоны за счет действия другой не происходит, связано опять-таки с принципом Гюйгенса, утверждающим, что от каждого элементарного источника исходит сферическая волна. Стало быть, амплитуда такой волны должна убывать обратно пропорционально расстоянию от источника до места наблюдения. Тогда, если вклад в амплитуду волны в точке  $A$  от центральной (нулевой) зоны обозначить как  $a/L$ , то вклад первой зоны будет  $-a/(L + l)$ . Знак минус появляется из-за того, что волна от первой зоны приходит в точку  $A$  в противофазе по отношению к волне от нулевой зоны. Вклад следующей зоны будет передаваться величиной  $a/(L + 2l)$  и т. д. В результате общее действие всех зон в точке  $A$  будет определяться знакопеременным рядом

$$I = \frac{a}{L} - \frac{a}{L+l} + \frac{a}{L+2l} - \frac{a}{L+3l} + \dots = a \sum_{k=0}^{\infty} (-1)^k \frac{1}{L+kl}.$$

Перепишем эту сумму в следующем виде:

$$I = \frac{a}{2L} + \left( \frac{a}{2L} - \frac{a}{L+l} + \frac{a}{2(L+2l)} \right) + \\ + \left( \frac{a}{2(L+2l)} - \frac{a}{L+3l} + \frac{a}{2(L+4l)} \right) + \dots$$

Снова воспользуемся тем, что  $l \ll L$ . Тогда комбинация

$$\frac{a}{2L} - \frac{a}{L+l} + \frac{a}{2(L+2l)} = \frac{a}{2L} - \frac{a}{L} + \frac{al}{L^2} + \frac{a}{2L} - \frac{2al}{2L^2} = 0.$$

Аналогично:

$$\frac{a}{2(L+2l)} - \frac{a}{L+3l} + \frac{a}{2(L+4l)} = \frac{a}{L} \frac{3}{2} \left( \frac{l}{L} \right) \approx 0.$$

Поступая так и далее, найдем, что искомая сумма  $I = \frac{a}{2L}$ .

Это означает, что действие всех зон Френеля сводится к действию лишь половины центральной зоны. Размер этой центральной зоны (радиус) равен, как было показано.

$r_1 = \sqrt{2Ll} = \sqrt{Ll}$ . В обычных условиях наблюдения  $r_1$  имеет порядок  $10^{-2}$  см.

Проведенное выше рассмотрение позволяет сделать ряд очень интересных выводов, подтверждающихся экспериментально.

1. При свободном прохождении плоской волны в пространстве ее действие в некоторой точке в направлении луча определяется лишь малой частью фронта, прилегающей к этому лучу, что вполне соответствует правилам геометрической оптики.

2. Если на пути распространения плоской волны поставить круглый экран, перекрывающий лишь небольшое число центральных зон Френеля, то, вопреки требованиям геометрической оптики, в точке А сохранится освещенность: в центре тени от экрана будет видно светлое пятно.

3. Наиболее интересный результат получится, если на пути волны поставить зонную решетку Френеля, т. е. дифракционную решетку, в которой непрозрачные участки образуют кольца, перекрывающие зоны с номерами 1, 3, 5,... (см. рис. 8.25,а). Тогда остающиеся зоны перестанут гасить друг друга и яркость в точке А возрастет. Зонная решетка работает подобно собирающей линзе! Аналогия простирается еще дальше: с помощью зонной решетки можно получать и фотографические изображения. В принципе, как указывалось выше, это позволяет создавать на основе зонных решеток специальные оптические приборы. В этом случае получается преимущество по сравнению с обычными линзовыми приборами в весе, но проигрывается в светосиле и качестве изображения.

Принцип зон Френеля можно применить и для рассмотрения дифракции на щели: если на щели укладывается четное число зон Френеля, то получается минимум, а если нечетное, то максимум дифракционной картины.

В заключение укажем, что рассмотрение условий интерференции и дифракции проводилось выше для идеализированного случая строго монохроматических излучаемых бесконечно долго волн одинаковой поляризации. В реальных экспериментах эти условия выполняются лишь с той или иной степенью точности, что накладывает ограничения на окончательный результат. В частности, интерференция при очень большой разности хода перестает наблюдаться.

## 8.8. Свет и вещество. Дисперсия света и поглощение

Попытаемся понять, что происходит, когда свет падает на вещество, и, в частности, дадим объяснение таким важнейшим экспериментальным фактам, как отражение и преломление света и зависимости коэффициента преломления от рода вещества и длины волны падающего света. Для простоты будем изучать эти вопросы на примере плоской электромагнитной волны, падающей перпендикулярно на плоскую поверхность.



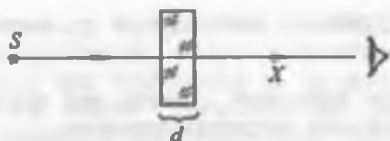


Рис. 8.26. К выводу формулы для дисперсии света

Пусть от источника света вдоль оси  $X$  распространяется электромагнитная волна  $E = E_0 \sin \omega(t - x/c)$  (рис. 8.26). Поставим на ее пути тонкий слой прозрачного вещества. Будем считать, что толщина  $d$  пластинки значительно меньше, чем расстояние как от источника света до места наблюдения, так и от пластинки до места наблюдения. Если свет прошел на пути от источника до места наблюдения через тонкий слой вещества, то, поскольку в веществе он движется со скоростью меньшей, чем в вакууме, должно быть определенное запаздывание по фазе света, попадающего в глаз наблюдателя в том случае, когда свет проходит через вещество, по сравнению с тем случаем, когда свет распространяется от источника до глаза наблюдателя в вакууме.

Это запаздывание легко учесть. В самом деле, свет от источника до наблюдателя пройдет расстояние  $x - d$  со скоростью  $c$  и расстояние  $d$  со скоростью  $v < c$ , причем  $1/v = n/c$ , где  $n$  — показатель преломления вещества.

Тогда колебание вектора  $E$  в точке  $x$  описывается формулой

$$E = E_0 \sin \omega \left( t - \frac{x-d}{c} - \frac{d}{v} \right) = E_0 \sin \omega \left[ t - \frac{x}{c} - \frac{d}{c} (n - 1) \right] = \\ = E_0 \sin \omega \left[ t - \frac{x}{c} - \alpha \right].$$

Здесь  $\alpha = d(n - 1)/c$  — заведомо малая величина. Разлагая функцию  $E$  в ряд по степеням этой малой величины и ограничиваясь только двумя первыми членами разложения, получаем

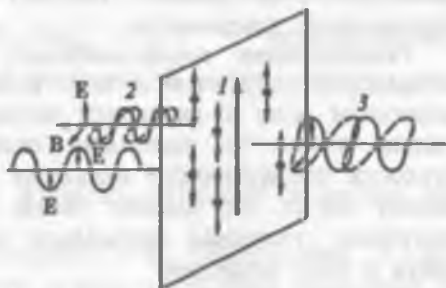
$$E_0 = E_0 \sin \omega(t - x/c) - \omega \alpha E_0 \cos \omega(t - x/c).$$

Таким образом, волну, прошедшую через тонкий слой вещества, можно представить в виде суммы двух волн, одна из которых распространяется подобно волне в вакууме, а другая сдвинута (отстает) относительно первой по фазе на  $\pi/2$  (назовем ее *вторичной волной*), причем амплитуда этой волны пропорциональна  $\alpha$ .

Получается, что наше предположение о том, что свет движется в веществе со скоростью меньшей, чем в вакууме, эквивалентно утверждению о появлении дополнительной волны  $E_0 \cos \omega(t - x/c)$ . Эта дополнительная волна генерируется веществом, составляющим данный слой, под воздействием первоначально падающей волны.

Физическая картина заключается в следующем. Первичная волна, попадая на поверхность вещества, «раскачивает» электроны, входящие в состав атомов данного вещества. При этом ее

Рис. 8.27. Электрический вектор падающей волны 1 «раскачивает» электроны и создает поверхностный переменный электрический ток (жирная стрелка). Этот ток генерирует отраженную 2 и проходящую 3 плоские волны



энергия быстро растрачивается и можно считать, что уже в небольшом слое первоначальная волна исчезает. Однако пришедшие в движение под действием первой волны электроны, двигаясь ускоренно, сами начинают излучать электромагнитную волну (рис. 8.27).

Энергия первоначально падающей волны переходит таким образом в энергию вторичных волн. Эти вторичные волны распространяются во все стороны, и часть волн выходит из вещества снова в вакуум и порождает отраженный луч. Другая часть волн проходит сквозь вещество, порождая преломленный луч.

Какими же свойствами должны обладать вторичные волны? Первичная электромагнитная волна будет достигать поверхность вещества и создавать на этой поверхности направленное вдоль нее переменные электрическое  $E$  и магнитное  $B$  поля. Эти поля должны воздействовать на заряды атомов вещества и приводить их в непрерывное коллективное движение. Так как первоначально на поверхности никакого коллективного направленного движения зарядов не было, то создать его может только поле  $E$  и именно его действие и надо учитывать (магнитное поле  $B$  на неподвижные заряды не действует).

На поверхности вещества возникает переменный генерируемый колебаниями вектора  $E$  первичной падающей волны поверхностный ток с равной плотностью по всей освещенной поверхности. В свою очередь, такой ток должен порождать по обе стороны от поверхности плоские, распространяющиеся вправо и влево, электромагнитные волны. Одна из этих волн будет восприниматься как отраженная волна, а другая—как проходящая. Чем больше сила наведенного поверхностного тока, тем больше интенсивность отраженной волны. В свою очередь, это связано с легкостью приведения электронов в движение. В этом смысле должна наблюдаться резкая разница между проводниками, где поверхностные электроны свободно могут перемещаться вдоль поверхности, и диэлектриками, где они удерживаются около атомов. Поэтому, в частности, металлические поверхности обладают несравненно лучшими отражающими свойствами, чем поверхности диэлектриков. Это действительно и

наблюдается на практике, поэтому зеркала имеют металлические отражающие поверхности.

Генерируемая поверхностным током волна имеет ту же начальную фазу, что и ток. Если электроны колеблются под действием поля  $E$  падающей волны и их смещение пропорционально  $\sin \omega t$ , то колебания силы тока, которая зависит от скорости направленного движения зарядов, должны подчиняться закону  $\cos \omega t$ . Но именно такую дополнительную волну мы и получили, учитывая формально запаздывание распространения света в слое вещества!

Заметим, что после того как облучение поверхности прекратится, распределение электронов в приповерхностном слое вещества вернется к первоначальному. Этот факт можно интерпретировать так, что электроны поверхности не просто свободно следуют за внешним полем, но на них действуют некоторые «возвращающие» силы. Если все это так, то движения электронов поверхности должны подчиняться законам вынужденных колебаний. На основании результатов § 4.4 можно записать для смещений электронов, пренебрегая малым в данном случае по сравнению с разностью  $\omega_0^2 - \omega^2$  коэффициентом  $4\delta^2 \omega^2$ ,

$$y = C_1 \frac{E_0}{\omega_0^2 - \omega^2} \sin \omega t.$$

Здесь  $\omega_0$  — собственная частота колебаний электронов. Эта частота зависит лишь от свойств вещества. Сделаем одну важную оговорку, а именно: будем считать, что  $\omega < \omega_0$ . Именно в этом случае колебания осциллятора следуют за колебаниями побуждающей силы. Значит, колебания электронов повторяют колебания вектора  $E$  падающей волны и лишь их амплитуда зависит от множителя  $(\omega_0^2 - \omega^2)^{-1}$ . Пренебрежем также воздействием движущихся зарядов друг на друга. Это верно, вообще говоря лишь для разреженных сред. Выбирая толщину приповерхностного слоя равной  $d$  (эта величина, по определению, мала), получаем для силы поверхностного тока значение (учитываем, что сила тока пропорциональна скорости смещений электронов и  $d$ )

$$I = C_2 |E_0| d \omega \cos \omega t / (\omega_0^2 - \omega^2).$$

Генерируемая волна «следит» за током, причем направления ее вектора  $E$  и тока должны быть противоположными. В противном случае генерируемое поле совершало бы над зарядами положительную работу и происходило бы поглощение, а не излучение энергии. Значит, около поверхности вещества

$$E_{\text{изл}} = -C_3 E_0 d \omega \cos \omega t / (\omega_0^2 - \omega^2),$$

где  $C_1$ ,  $C_2$  и  $C_3$  — некоторые несущественные константы. Сравнивая коэффициент  $-C_3 E_0 d / (\omega_0^2 - \omega^2)$  с коэффициентом  $-E_0 d (n - 1) / c$  при вторичной проходящей волне, находим, что

$$n = 1 + D / (\omega_0^2 - \omega^2),$$

причем  $D = \text{const} > 0$ . Эта формула позволяет получить ряд важных выводов.

Проследим поведение коэффициента преломления  $n$  при возрастании частоты падающего электромагнитного излучения от нуля. Вначале  $n \approx 1 + D / \omega_0^2$  и есть постоянная величина. При этом обычно отношение  $D / \omega_0^2$  невелико и поэтому  $n \approx 1$  и почти не зависит от  $\omega$ . Затем при приближении  $\omega^2$  к  $\omega_0^2$  (собственная частота колебаний электронов вещества) показатель преломления  $n$  начинает возрастать, особенно сильно в области  $\omega \approx \omega_0$ . Появляется существенная зависимость коэффициента  $n$  от частоты падающей волны, причем  $n > 1$ . Выше мы договорились, что рассматриваем лишь случай  $\omega < \omega_0$ . Именно такая ситуация и наблюдается, как правило, при прохождении света (электромагнитных волн видимого диапазона) через прозрачные для него вещества: стекло, воду, кварц, алмаз и т. д. При  $\omega > \omega_0$  возникает ряд интересных особенностей, приводящих к появлению *аномальной дисперсии*.

Вернемся снова к формуле для коэффициента преломления. В области, где  $\omega$  начинает приближаться к  $\omega_0$ , произойдет возрастание  $n > 1$ , причем растет и производная  $dn/d\omega$  (рис. 8.28). Это значит, что при падении под углом, не равным нулю, на поверхность вещества двух электромагнитных волн с разными частотами  $\omega_1$  и  $\omega_2$  они будут преломляться под разными углами, что позволяет осуществить их разделение в пространстве. Именно это обстоятельство и позволяет воспользоваться призмой для разложения света в спектр. *Явление, которое связано с различием показателя преломления вещества в зависимости от длины волны падающего света, называется дисперсией*. Явление дисперсии оказывается полезным по двум причинам: во-первых, на его основе можно изучить состав падающего на призму излучения и, во-вторых, исследовать свойства самого вещества. В самом деле, по мере приближения  $\omega$  к  $\omega_0$ , т. е. частоты падающего света к собственной частоте колебаний электронов, кривая зависимости показателя преломления от длины волны падающего света идет все круче и круче. Исследуя форму этой кривой, можно получить сведения о величине  $\omega_0$ .

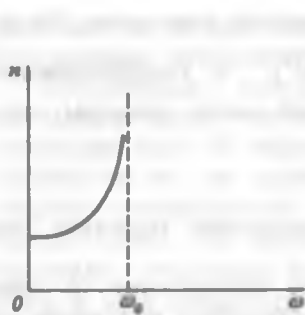


Рис. 8.28. Зависимость показателя преломления от частоты падающей волны

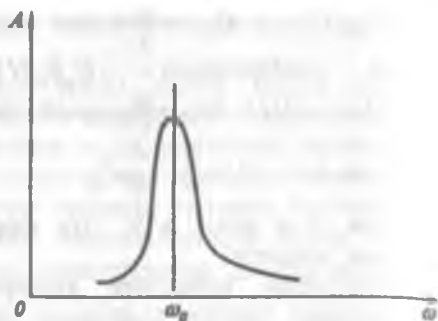


Рис. 8.29. Резонансное поглощение энергии волны при прохождении света сквозь вещество

Рассмотрим явление, происходящее в случае, когда  $\omega$  очень близко по значению к  $\omega_0$  или равно  $\omega_0$ . Вспомним, как ведет себя кривая зависимости амплитуды колебаний маятника, на который действует периодическая внешняя сила. Если этот маятник движется без трения, то в области, близкой к собственной частоте, амплитуда колебаний возрастет до бесконечности. Это случай резонанса. Однако реально такого возрастания до бесконечности никогда не наблюдается, что связано с тем, что во всякой системе существуют потери энергии, которые можно назвать потерями на трение. В случае электронов это трение имеет довольно сложную природу. Реальная кривая зависимости амплитуды от частоты колебаний внешней силы показана на рис. 8.29.

Подобное трение максимально в области, близкой к собственной частоте  $\omega_0$ . В самом деле, потери на трение должны существенно зависеть от амплитуды колебаний, а в области, близкой к  $\omega_0$ , эта амплитуда максимальна и резко увеличивается. Другими словами, именно в этой области первоначальная волна наиболее интенсивно теряет свою энергию, раскачивая электроны вещества. Таким образом, наличие трения при движении электронов должно приводить к падению интенсивности вторичной волны по мере ее проникновения внутрь вещества. Такое явление названо *поглощением света*. При этом наиболее сильное поглощение света происходит при частотах, близких к  $\omega_0$ . Измеряя поглощение света, можно получить сведения о величине  $\omega_0$  и, следовательно, о некоторых свойствах самого вещества.

Поглощение света описывается законом Ламберта—Бугера:

$$I = I_0 e^{-\alpha l}.$$

Здесь  $I_0$ —интенсивность падающего света на слой вещества толщины  $l$ ;  $I$ —интенсивность прошедшего света;  $\alpha$ —коэффициент поглощения света в веществе.

Получается эта зависимость на основании простого рассуждения. Пусть на слой вещества толщиной  $dx$  падает волна интенсивностью  $I$ . При прохождении этого слоя поглощается энергия  $dI$ , пропорциональная толщине слоя и величине падающего потока энергии. В самом деле, чем больший поток энергии направлен на слой, тем большая доля энергии поглотится. Итак,

$$-dI = \alpha l dx$$

(так как энергия поглощается, то перед  $dI$  стоит знак минус). Перепишем это выражение в другом виде и проинтегрируем правую и левую части:

$$\frac{dI}{I} = -\alpha dx \quad \text{или} \quad \ln I = -\alpha x + \text{const.}$$

Выбирая в качестве постоянной  $\ln I_0$ , для поглощения в слое находим  $\ln(I/I_0) = -\alpha l$  или  $I = I_0 e^{-\alpha l}$ . Коэффициент  $\alpha$  (коэффициент пропорциональности между  $-dI$  и  $I dx$ ) называется коэффициентом поглощения и зависит от свойств вещества. Графическая зависимость интенсивности прошедшего света от толщины слоя приведена на рис. 8.30.

Поглощение света широко используется. Подбирая соответствующие вещества, можно осуществить, как и с помощью явления дисперсии, анализ спектрального состава падающего света. На таком принципе работают *светофильтры*—окрашенные стекла, которые пропускают свет определенных длин волн, довольно сильно поглощая свет вне этих спектральных областей. Такой анализ спектра является более грубым, чем анализ спектра с помощью дисперсионных приборов, т. е. с помощью призмы или дифракционной решетки. В целом ряде случаев, однако, этого вполне достаточно для практических целей.

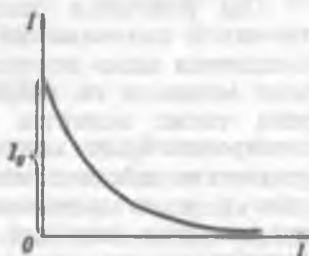


Рис. 8.30. Зависимость интенсивности прошедшего света от толщины поглощающего слоя

## 8.9. Давление света

Исходя из представления о свете как об электромагнитной волне, можно объяснить качественно давление света. Опыт по определению давления света, впервые поставленный П. Лебедевым, известен из школьного курса, и мы его лишь напомним.

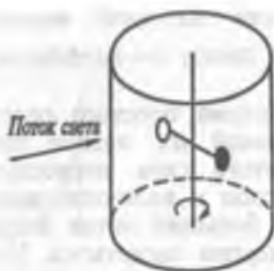


Рис. 8.31. Схема опыта по наблюдению давления света

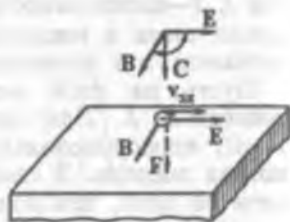


Рис. 8.32. Действие световой волны на поверхность вещества

Если поместить в вакуум прибор, напоминающий мельницу (рис. 8.31) со свободно вращающейся крыльчаткой, и направлять свет на один из лепестков этой крыльчатки, то последняя начинает вращаться. Объяснить это явление можно следующим образом.

Пусть перпендикулярно поверхности вещества (например, металла) падает электромагнитная волна. Свет представляет поперечную электромагнитную волну, а это значит, что электрические и магнитные векторы колеблются, оставаясь все время перпендикулярными направлению распространения волны.

Под действием напряженности электрического поля, электрической составляющей волны, электроны поверхности вещества смещаются вдоль поверхности. Если электрические составляющие волн меняются по синусоидальному закону, то смещение электрона также меняется синусоидально. Скорость же движения электронов будет меняться, очевидно, уже косинусоидально. В результате действия магнитного поля на движущийся со скоростью  $v$  вдоль поверхности электрон действует сила Лоренца, отклоняющая электрон внутрь металла (рис. 8.32). Электрон воздействует на атомы металла, передавая им свое движение. Таким образом, возникает сила, направленная внутрь металла, сила давления света.

## 8.10. Фотоэффект

Возьмем цинковую пластинку, на которую с помощью электрофорной машины нанесены отрицательные заряды. Направим на пластину поток света, присоединив ее предварительно к электроскопу. Пластинка разрядится, следовательно, при воздействии света на пластинку из нее выбиваются электроны. Такое явление называется *фотоэффектом*.

Представить физическую картину фотоэффекта можно следующим образом. Предположим, что в веществе имеются упруго связанные с ядрами атомов вещества электроны. При наложении электрического поля достаточно большой интенсивности амплитуда раскачивания электронов становится столь большой, что в

конце концов они отрываются от атома и переходят в окружающее пространство. Таким образом, сам эффект может быть легко понят в свете волновых представлений. На основании этих же представлений следует ожидать, что количество выбитых электронов (сила фототока) должно расти приблизительно пропорционально интенсивности падающего света.

Это действительно наблюдается в эксперименте, но лишь до известного предела, который связан с длиной волны облучающего света. Если длина волны становится больше определенной (для разных веществ эта величина различна), то сила фототока резко падает почти до нуля, несмотря на то, что интенсивность облучения остается значительной. Это *красная граница фотоэффекта*. Понять этот факт с точки зрения волновой природы света невозможно.

Объяснение прекращения фототока следующее. Электрон удерживается в веществе электрическими силами. Чтобы выбить электрон из вещества, нужно сообщить ему энергию не меньшую, чем энергия, которая необходима на отрыв электрона от атома. Такая энергия называется *работой выхода* электрона из вещества. Обозначим ее  $A$ . Если вылетевший электрон обладает кинетической энергией  $mv^2/2$ , то общая энергия, затраченная на его извлечение из вещества, равна сумме кинетической энергии и работы выхода, т. е.  $W = mv^2/2 + A$ . Оказывается, что выполняется следующее соотношение:

$$\frac{1}{2}mv^2 + A = \hbar\omega \text{ (для одного электрона).}$$

Здесь  $\hbar$ —постоянная Планка;  $\omega$ —частота падающего света;  $\hbar\omega$ —элементарная энергия электромагнитного излучения, пропорциональная частоте излучения. Такая порция энергии  $\hbar\omega$ , распространяющаяся в пространстве, называется *квантом электромагнитной энергии* или для световой волны *квантом света* или *фотоном*.

Картину квантов можно представить следующим образом. Электромагнитная волна никогда не излучается источниками бесконечно долгое время. В частности, процесс излучения или поглощения света атомом происходит за очень короткое время. Поэтому можно считать, что электромагнитная волна имеет определенные начало и конец. Иногда такие отрезки волн называют *цугом волны*\*). Каждый цуг волны несет вполне определенную энергию, которая равна  $\hbar\omega$ . Энергия электромагнитного поля может поглощаться или излучаться только вполне определенными порциями, равными  $\hbar\omega$ . Ни один из процессов не может происходить так, чтобы была поглощена только часть этой энергии. В частности, при фотоэффекте

\*) Здесь и далее мы будем пользоваться понятием цуга, или отрезка волны. Надо, однако, иметь в виду, что это не более чем наглядный образ..



система поглощает энергию, кратную или равную одному кванту. Интенсивность света определяется количеством квантов, падающих на единицу поверхности данного вещества в единицу времени. Если каждый квант несет энергию меньшую, чем работа выхода электрона из вещества, то каково бы ни было количество квантов, попадающих на поверхность металла (фотоэлемента), все равно вырвать электрон из него не удастся (мы не касаемся здесь нелинейных эффектов). Только с ростом энергии кванта, когда выполняется соотношение  $\hbar\omega > A$  для одного электрона, электрон может быть вырван из вещества. Таким образом, гипотеза квантов легко объясняет существование красной границы фотоэффекта. Если энергия квантов больше, чем работа выхода электронов, то число вырванных электронов пропорционально числу падающих квантов, т. е. интенсивности света. В области частот  $\omega > A/\hbar$  можно пользоваться электромагнитной теорией света.

Итак, с одной стороны, свет ведет себя как электромагнитная волна и большое число фактов может быть объяснено с этой точки зрения. С другой стороны, в некоторых случаях свет ведет себя подобно совокупности частиц, обладающих определенной энергией. В частности, явление фотоэффекта можно было бы объяснить, предполагая, что свет есть поток частиц (корпускул), обладающих вполне определенной кинетической энергией, и сделать только дополнительное предположение, что взаимодействие вещества с этими частицами обладает специфическими свойствами, заключающимися в том, что передача энергии от частицы к веществу может быть совершена только определенными порциями. В свое время корпускулярная теория света была выдвинута И. Ньютоном. Волновая теория в этот же период развивалась Х. Гюйгенсом. Впоследствии она получила большое распространение, потому что позволила объяснить большинство наблюдаемых в то время фактов. Интерес к корпускулярной теории света вновь возник на рубеже XIX—XX вв., когда был обнаружен целый ряд фактов, не объясняемых с точки зрения теории электромагнитных волн. Оказалось, что в микромире вообще все явления совершаются так, что в одних случаях проявляются волновые, в других корпускулярные свойства. Также и свет является диалектическим единством, с одной стороны, электромагнитных волн, с другой стороны, корпускул. К сожалению, понятия, которые вкладываются в современной теории света в этот термин, довольно сложны и мы на них подробно останавливаться не будем.

## 8.11. Поляризация света

Представление о свете как о потоке квантов в совокупности с представлением о свете как о наборе электромагнитных волн позволяет объяснить еще одно чрезвычайно важное явление, называемое *поляризацией света*. Воспользуемся наглядным обра-

зом кванта как цуга электромагнитной волны, перемещающейся в пространстве со скоростью света. Электрические и магнитные векторы в таком отрезке волны меняют свое значение по синусоидальному закону, причем взаимное расположение этих векторов всегда перпендикулярно и, кроме того, плоскость, в которой меняется вектор напряженности электрического поля, и плоскость, в которой лежит вектор индукции магнитного поля, во время существования кванта, т. е. от момента его излучения до момента его поглощения веществом, останутся неизменными, т. е. квант света поляризован. Если имеется произвольный источник света, то он излучает всевозможные кванты электромагнитной энергии, т. е. цуги волн, в которых плоскости расположения электрических и магнитных векторов могут быть самые разнообразные. Электромагнитное излучение или свет, состоящий из квантов с произвольной ориентацией плоскостей расположения векторов  $E$  и  $B$ , называется *неполяризованным светом*.

Под действием падающих волн электроны поверхности вещества будут раскачиваться, что приведет к появлению вторичных волн.

Представим, что электрон колеблется вверх и вниз вдоль оси  $Z$ . Тогда изменение вектора  $E$  в волне, порождаемой колеблющимся электроном, также параллельно  $Z$  (рис. 8.33). Это значит, что всякий колеблющийся электрон индуцирует электромагнитную волну вполне определенной поляризации. Если электроны под действием падающего излучения могут смещаться в любом направлении вдоль поверхности и при этом их колебания в двух взаимно перпендикулярных направлениях (по осям  $Z$  и  $Y$ ) совершаются с одной и той же частотой, то мы не будем замечать никаких специфических явлений. Представим вещество, в котором электрон может смещаться только, например, в направлении  $Z$  (рис. 8.33), а в направлении  $Y$  смещаться не может. Тогда при падении на это вещество электромагнитной волны (кванта), электрический вектор которой параллелен оси  $Z$ , электрон придет в колебание и породит электромагнитную волну, которая также будет поляризована вдоль оси  $Z$ . Волны, падающие на вещество и имеющие направление электрического вектора, перпендикулярное оси  $Z$ , не приводят в движение электроны вещества, и вторичные волны с поляризацией, перпендикулярной оси  $Z$  (т. е. по оси  $Y$ ), появляться не будут. В результате наблюдается своеобразное явление фильтрации квантов по признаку их поляризации. Вещество, обладающее таким свойством, что его электроны могут колебаться только в одном заданном направлении, из всех падающих цугов волн выделит только те, которые обладают изменяющимся вектором электрического поля, параллельным направлению, вдоль которого может смещаться электрон. Такие вещества называются *оптически анизотропными*. Таким образом, если пучок фотонов самой разнообразной поляризации падает на оптически анизотропную

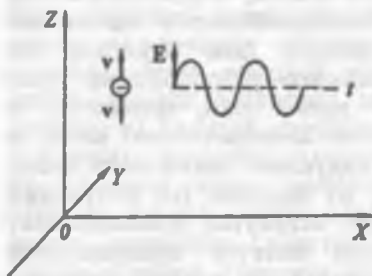


Рис. 8.33. Схема появления поляризованной волны

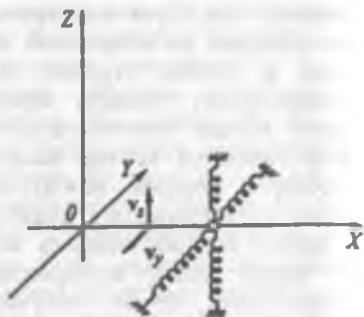


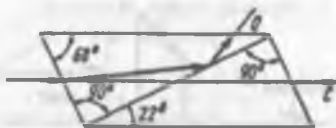
Рис. 8.34. К объяснению двойного лучепреломления

среду, то в этом случае преломленный луч будет иметь преимущественно одну какую-то поляризацию.

Пусть теперь под действием внешнего электромагнитного поля электрон может смещаться в двух взаимно перпендикулярных направлениях, однако собственные частоты  $\omega_{0z}$  и  $\omega_{0y}$  колебаний электрона в этих двух направлениях оказываются неодинаковыми. Коэффициент преломления вещества зависит от собственных частот колебаний электронов, которые ранее предполагались не связанными с направлением колебаний. Было показано, что коэффициент преломления равен  $n = 1 + D/(\omega_0^2 - \omega^2)$ . Если ввести оси координат  $Z$  и  $Y$  и предположить, что частоты колебаний электронов вдоль осей  $Z$  и  $Y$  (рис. 8.34) равны соответственно  $\omega_{0z}$  и  $\omega_{0y}$ , то получим, что вторичные волны, индуцируемые квантами света, поляризованными в двух взаимно перпендикулярных направлениях, распространяются так, как если бы соответствующие коэффициенты преломления, равные  $n_z = 1 + D/(\omega_{0z}^2 - \omega^2)$  и  $n_y = 1 + D/(\omega_{0y}^2 - \omega^2)$ , были бы различны. Если на поверхность такого вещества направить луч света, то после преломления образуются два луча, направленные друг относительно друга под разными углами. Поскольку каждый из лучей индуцируется падающими фотонами двух взаимно перпендикулярных поляризаций, то и сами лучи будут иметь различную поляризацию. Направления векторов электрического поля в волнах взаимно перпендикулярны. Такое деление первоначально падающего луча на два с разной поляризацией называется **двойным лучепреломлением**.

Эффект двойного лучепреломления можно использовать для построения приборов (*поляризаторов*), которые выделяют из всего потока первоначально поляризованного света только те кванты, которые имеют заданную поляризацию. Примером

рис. 8.35. Ход лучей в призме Николя



Поляризатором является *призма Николя* (или просто *николь*), которая изготавливается из материала, называемого *исландским шпатом*. Это кристаллический прозрачный материал. Принцип действия призмы Николя иллюстрируется рис. 8.35. Первоначально падающий луч разделяется на два луча, которые затем падают под разными углами на поверхность раздела двух кусков *исландского шпата*. Можно подобрать угол падения одного из лучей таким образом, чтобы на границе раздела этот луч испытал полное отражение. В результате один из преломленных лучей, отразившись, уйдет в направлении, указанном на рисунке стрелкой. Другой луч пройдет в другую часть призмы и затем выйдет в воздух. Поскольку лучи имеют разное направление поляризации, то вышедший из призмы Николя луч будет иметь вполне определенную поляризацию. Если на его пути поставить второй *николь*, причем повернуть его так, чтобы свет, имеющий поляризацию, заданную ему первым *ником*, не приводил к появлению вторичной волны, направленной по лучу *AB*, то он через призму Николя не пройдет. Такое взаимное расположение двух *николей* называется *скрещенным* (рис. 8.36). Каждый из *николей* пропускает волны только вполне определенной поляризации. Если *никили* расположены перпендикулярно друг другу, то свет не пройдет через такую систему. Если *никили* расположить параллельно друг другу, т. е. таким образом, чтобы они пропускали поляризованный свет одного направления, то наблюдается максимум проходящего света. Если первоначально падающий свет содержит фотоны разнообразной поляризации (например, естественный свет), то, поскольку в естественном свете доля фотонов, поляризованных вдоль любого из направлений, практически одинакова, количество фотонов, прошедших через это вещество, вдвое меньше, чем количество первоначальных фотонов. Таким образом, интенсивность прошедшего сквозь *николь* света равна половине интенсивности первоначально падающего на *николь* естественного света.

Важнейшим эффектом, который можно изучать с помощью поляризаторов, является эффект вращения плоскости поляризации.

Некоторые вещества (обычно биологические, содержащие длинные спиральные молекулы) обладают следующим свойством. Поглощая первоначально падающую электромагнитную волну, они генерируют вторую волну, вектор поляризации которой повернут по сравнению с вектором поляризации первоначально падающей волны. Если через слой вещества, содержащий спиральные молекулы, пропустить поляризованный свет, то на

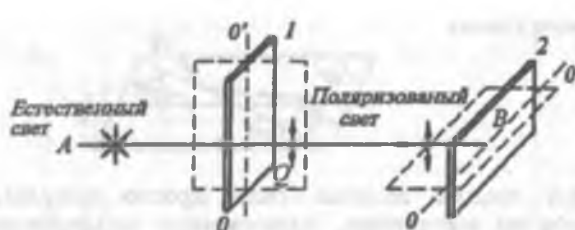


Рис. 8.36. Совместное действие двух поляризаторов

выходе слоя наблюдается поляризованный свет, вектор напряженности электрического поля которого повернут по отношению к исходному на определенный угол. Этот угол пропорционален количеству содержащихся в веществе оптически активных молекул.

Это позволяет производить количественный анализ проб (например, концентрации сахаров). Важно также, что с помощью определения направления угла поворота можно решать вопросы о том, закручены ли спиральные молекулы вправо или влево, что имеет большое значение для диагностики ряда наследственных болезней.

В заключение остановимся на вопросе о поляризации света при отражении. Отраженная электромагнитная волна также является волной, возникающей в результате колебаний электронов поверхности под действием падающей волны.

Каждый из колеблющихся электронов поверхности является источником поляризованной сферической электромагнитной волны. Полный поток волн возникает в результате сложения волн, которые интерферируют между собой. Условия интерференции таких волн при различной поляризации и при различных углах падения волны неодинаковы. В результате при некоторых углах падения волны наблюдается преимущественное отражение света только с одной поляризацией, именно с поляризацией, при которой вектор напряженности электрического поля перпендикулярен плоскости, в которой лежат падающий и отраженный лучи (рис. 8.37,а). Такой угол падения, названный *углом Брюстера*, определяется соотношением

$$\operatorname{tg} \varphi = n,$$

где  $n$  — показатель преломления. Угол между отраженным и преломленным лучами в этом случае составляет  $90^\circ$ . Это легко понять, если учесть, что поляризация любой волны, в частности и преломленной, всегда поперечна и что колеблющийся электрон не может излучать электромагнитную волну в направлении, совпадающем с направлением его движения.

Поэтому электроны, движущиеся перпендикулярно преломленному лучу, в плоскости лучей в том же направлении, т. е. вдоль линии своего движения, волны излучать не будут. Если преломленная и отраженная волны составляют угол  $90^\circ$ , то такие колебания электронов не дадут вторичных волн в направлении отраженной волны. Вклад в отраженную волну могут дать только

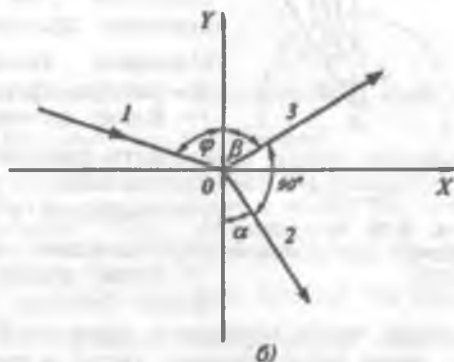
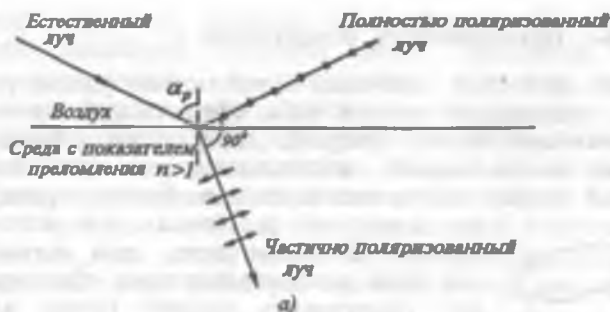


Рис. 8.37. Поляризация света при преломлении на границе двух сред (а). Угол падения равен углу Брюстера. К выводу формулы закона Брюстера (б)

электроны, колеблющиеся перпендикулярно плоскости лучей. Отсюда заключаем, что отраженная при угле Брюстера волна поляризована перпендикулярно плоскости лучей.

Обозначим угол падения  $\varphi$ , угол преломления  $\alpha$  и угол отражения  $\beta$ . Так как  $\varphi = \beta$ ,  $\sin \varphi / \sin \alpha = n$  и отраженный луч составляет при угле полной поляризации (угол Брюстера) угол  $90^\circ$  с преломленным лучом, то  $\alpha = 180^\circ - 90^\circ - \beta = 90^\circ - \varphi$ , т. е.

$$\sin \varphi / \sin \alpha = \sin \varphi / \sin (90^\circ - \varphi) = \operatorname{tg} \varphi = n$$

(рис. 8.37, б).

В той или иной степени поляризация при отражении происходит всегда. Явление это можно наблюдать, если воспользоваться поляризационным светофильтром, представляющим две пластины или две пленки, каждая из которых обладает вполне определенными поляризующими свойствами и действует подобно призме Николя. Если посмотреть, например, через поляризационный фотографический фильтр на блик Солнца, отраженный от поверхности полированного крыла автомобиля, или на солнечные блики на воде, то при вращении фильтра яркость бликов меняется. Это используется в фотографии для того, чтобы на снимке не было ярких бликов, затеняющих основное изображение.

## 8.12. Освещенность поверхности

Каждый точечный источник света можно рассматривать как источник, создающий вокруг себя сферическую световую или электромагнитную волну. Энергия, которую несет волна, определяет освещенность экрана, поставленного на пути волны. Для сферической волны освещенность падает обратно пропорционально

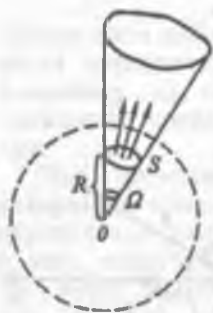


Рис. 8.38. К выводу формул для освещенности

квадрату расстояния от источника до экрана. В самом деле, если источник света окружен концентрическими сферами, то через поверхность каждой сферы или любой телесный угол  $\omega$  (рис. 8.38) в единицу времени проходит одна и та же энергия. Площадь каждой сферы равна  $4\pi R^2$ , где  $R$ —радиус сферы.

Если на поверхности каждой из сфер выбрать участок размером  $1 \text{ см}^2$ , то энергия, проходящая участка на разных сферах, уменьшается с ростом радиуса сферы пропорционально  $R^{-2}$ .

Если волна распространяется в направлении вектора  $e$ , то энергия, переносимая волной через площадку единичной площади, перпендикулярную  $e$ , равна произведению числа  $n$  квантов, проходящих сквозь эту поверхность в единицу времени, на их энергию. Эта энергия определяет интенсивность волны.

Вектор потока квантов в направлении  $e$  равен  $I = n\hbar\omega e$ . Тогда поток квантов, прошедших через элемент поверхности  $dS$  в единицу времени, т. е. освещенность поверхности, равен скалярному произведению  $I dS$ . Если  $I \parallel dS$ , то поток квантов скользит по поверхности, не проходя через нее. Освещенность поверхности в этом случае равна нулю.

Общий поток энергии, переносимой квантами через поверхность  $S$ , равен сумме всех потоков, проходящих через элементы этой поверхности, т. е. равен  $\int I dS$ . Если поверхность замкнута, то общий поток равен  $\oint_S I dS$ . С другой стороны, этот поток

равен мощности источника, т. е. произведению числа квантов, излучаемых им в единицу времени во все стороны, на их энергию, если кванты не поглощаются внутри объема, ограниченного данной поверхностью. Обозначив эту мощность  $W$ , получим  $\oint_S I dS = W$ . Это соотношение по своей математической форме

подобно соотношению  $\oint_S E \cdot dS = \epsilon_0^{-1} q$ , известному из электростатики. Пользуясь этой аналогией, легко найдем (см. § 7.2), что

освещенность единицы площади цилиндрической поверхности, окружающей линейный источник света (например, длинную люминесцентную лампу), на расстояниях, меньших, чем длина лампы, падает обратно пропорционально расстоянию, а освещенность от плоских источников света на расстояниях, меньших, чем размер светящейся плоскости, вообще не зависит от расстояния. Это необходимо учитывать при конструировании тепличных устройств, в которых используются различные источники освещения. Заметим, что совокупность близко расположенных точечных источников (ламп) действует подобно линии или светящейся плоскости.

#### Вопросы для самоконтроля

1. Описать физическую картину появления электромагнитных полей в пространстве, где зарядов или проводников с токами нет.
2. Что такое поле излучения и каким свойством оно обладает?
3. На каких общих принципах базируются различные способы измерения скорости света? Привести примеры экспериментов по измерению скорости света.
4. В каких радиодиапазонах работают радиолокаторы и почему для этой цели выбираются именно такие диапазоны? Каков диапазон длин волн для видимого света? Почему для радиосвязи, локации, исследования веществ и т. д. используют электромагнитные волны разных диапазонов?
5. При каких условиях возникают интерференционные явления? Что такое когерентные источники и разность хода?
6. Приведите примеры опытов по наблюдению интерференции света на основе принципа разделения фронта волн.
7. Что такое просветление оптики и какое физическое явление лежит в основе этого просветления?
8. Как по наблюдению интерференции можно проверить качество полировки поверхности зеркал или металлических деталей?
9. Что такое фурье-спектроскопия? Чем отличаются фурье-спектрометры от спектрометров, построенных по классической схеме, какими преимуществами и недостатками те и другие обладают?
10. Что такое голография, какой физический эффект лежит в ее основе? Чем отличается объемное голографическое изображение от объемного изображения, которое получается, например, на растровых фотооткрытках?
11. Можно ли антенной размером 1 м хорошо сфокусировать электромагнитное излучение с длиной волны 10 м?
12. Известно, что в противовоздушной обороне используются локаторы, которые принципиально можно разделить на два типа. Первая группа локаторов обладает способностью излучать достаточно длинные электромагнитные волны и используется для дальнего обнаружения. Вторая группа локаторов излучает электромагнитные волны малой длины (сантиметровый диапазон) и устанавливается на самолетах. Почему приходится пользоваться двумя типами локаторов и к каким последствиям в отношении размеров антенн приводит использование излучений разных диапазонов (метрового и сантиметрового)?
13. Можно ли с помощью радиолокатора с длиной волны порядка 10—20 м уверенно определять место расположения самолета в пределах летного поля?
14. Что такое дифракционный предел разрешающей способности оптических приборов? Почему при переходе от видимого света к ультрафиолетовому увеличение микроскопов возрастает?
15. В результате чего при использовании дифракционной решетки число максимумов возрастает по сравнению с дифракционной картиной, получаемой от щели?
16. Что такое обратные задачи дифракции? Какую роль играют эти задачи при исследовании строения кристаллов?



17. Как объяснить явления, которые происходят на границе раздела, например воздух—стекло, когда возникает отраженная волна?

18. Какие особенности раскачивания поверхностных и внутренних электронов в среде за счет внешней падающей волны проявляются в эффекте дисперсии при прохождении света через прозрачную среду?

19. Существуют материалы, например кристалл KRS, прозрачные до длин волн порядка 50 мкм и более в инфракрасной области спектра. Почему невыгодно, несмотря на это, делать призму для инфракрасных спектрометров из этого материала и почему для того, чтобы получить спектрометр достаточно высокого качества, используют сменные призмы: из  $\text{CaF}_2$ , которая прозрачна только в области приблизительно до 10 мкм; затем из  $\text{NaCl}$  с прозрачностью в области приблизительно до 20 мкм; затем из  $\text{KBr}$  с прозрачностью в области до 30 мкм.

20. Оценить, как изменится интенсивность прошедшего сквозь стеклянную пластинку света при увеличении толщины этой пластины вдвое и втрое?

21. Какие оптические явления могут быть объяснены с микроскопической (квантовой) и макроскопической (волновой) точек зрения? В каких случаях можно использовать описание световых явлений с позиции теории электромагнитных волн? Как связаны векторы напряженности и индукции в электромагнитной волне с числом и энергией квантов в данной области пространства?

22. Какой эффект при наблюдении фотоэффекта в наибольшей степени показывает ограниченность применения волновой точки зрения?

23. Что такое поляризация света и с помощью каких приемов можно ее наблюдать?

24. Для каких целей в фотографии используются поляризационные фильтры и на каком принципе основано их действие?

25. При каких условиях может наблюдаться двойное лучепреломление? Как этот эффект может быть использован для получения хорошо поляризованных пучков света?

26. Что такое оптическая активность и как это явление может быть использовано в молекулярной биологии?

27. На каком расстоянии друг от друга должны быть размещены под потолком теплицы точечные источники света, чтобы на полу, находящемся на расстоянии  $R$  от потолка, интенсивность освещения нигде не менялась при переходе от одной точки пола к другой более чем в 0,5 раза?



## ГЛАВА 9

### ЭЛЕМЕНТЫ ТЕОРИИ СТРОЕНИЯ ВЕЩЕСТВА

*«Вы, конечно, можете называть это чушью, но я-то встречала чушь такую, что в сравнении с ней эта кажется толковым словом».*

*Л. Кэррал*

#### 9.1. Особенности движения микрочастиц и способы описания их состояний

До сих пор, пользуясь понятиями атома, молекулы, твердого тела, и т. д., мы не касались вопросов о том, как устроены эти объекты, каким законам они подчиняются.

Между тем достаточно вспомнить хорошо известный из школьного курса физики опыт Резерфорда, результаты которого показали, что атомы состоят из тяжелого положительно заряженного ядра размером порядка  $10^{-13}$  см и находящихся вокруг него в области порядка  $10^{-8}$  см отрицательных электронов, чтобы понять, что все изложенные нами до сих пор фундаментальные положения физики не позволяют объяснить, почему атом существует как устойчивое образование.

Если считать, что электроны относительно ядра неподвижны, то, чтобы атом существовал постоянно, надо предположить, что кроме электрических сил взаимного притяжения ядра и электронов, которые по теореме Ирншоу не могут создать устойчивой системы, существуют силы незлектрического происхождения, удерживающие электроны от падения на ядро. Между тем все известные еще до опыта Резерфорда в физике эксперименты совершенно определенно показывали, что таких сил нет.

Не спасает положения и естественное, казалось бы, предположение о том, что электроны вокруг ядра движутся по эллиптическим орбитам, подобно планетам вокруг Солнца, хотя на эту мысль наталкивает сходство сил всемирного тяготения с кулоновскими. Такая планетарная модель атома удержалась в физике до наших дней, но не более как наглядное изображение. Реально атом в форме такой планетарной системы не может

существовать, так как вращающиеся вокруг ядра электроны движутся ускоренно. Но тогда они, как всякие ускоренно движущиеся заряженные частицы, должны излучать электромагнитные волны. Действительно, если заставить электроны двигаться по кругу (как в ускорителях элементарных частиц — синхротронах), то возникает электромагнитное синхротронное излучение.

Атомные электроны в отличие от электронов в ускорителях электромагнитную энергию в обычных условиях не излучают. Возбужденные атомы могут излучать электромагнитные волны (например, светится атомное газовое облако электрической дуги), но при этом они излучают электромагнитные волны лишь с определенным дискретным набором длин волн.

Первым, кто понял, что за этим скрывается новая физика и что надо искать другие средства описания поведения частиц в атоме и вообще атомных объектах, был Н. Бор. Он сформулировал два утверждения, получивших название *постулатов Бора* и лежащих в основе всей современной атомно-молекулярной физики.

1. *Атомная система устойчива только для определенной совокупности состояний (стационарных состояний), которая в общем случае соответствует дискретной последовательности значений энергии атома. Каждое изменение этой энергии связано с переходом атома (или молекулы) из одного стационарного состояния в другое.*

2. *Способность атома (молекулы) поглощать и испускать излучение подчиняется закону, по которому излучение, связанное с переходом, должно быть монохроматическим и иметь частоту  $\omega$ :*

$$\hbar\omega = E_1 - E_2,$$

где  $\hbar$  — постоянная Планка;  $E_1$  и  $E_2$  — энергии стационарных состояний. Это соотношение называется *правилом частот Бора*. Оно показывает, что в процессах излучения и поглощения энергии атомами выполняется закон сохранения энергии.

В дальнейшем Э. Шредингеру, В. Гейзенбергу, П. Дираку и другим физикам удалось развить новую механику и электродинамику, названные *квантовой механикой* и *электродинамикой*. На их основе не только можно объяснить строение и поведение атомно-молекулярных объектов, но и рассчитать спектральные, пространственные и другие характеристики атомов, молекул вплоть до белков, полимеров, кристаллов и т. д.

В отдельных случаях можно детально описать пространственно-временные изменения молекул при их взаимодействиях. Предсказано и подтверждено экспериментально существование новых соединений, например молекул  $\text{H}_3$  и  $\text{NH}_4$  (есть основание ожидать, что существуют молекулы  $\text{H}_4$ ,  $\text{H}_5$ ). Физика вторглась в область химии, и соответствующие ее разделы (квантовая химия)

стали теоретической основой этой области науки. Появилась квантовая биохимия, квантовая фармакология и др.

Обратимся к важному эксперименту. Всем хорошо известна электронно-лучевая трубка, применяемая в осциллографах, в дисплеях и телевизорах. В этих трубках генерируемые катодом и формируемые электрическими и магнитными полями электронные пучки (все устройство называют *электронной пушкой*) направляются на светящиеся под ударами электронов экраны. На пути от источника до экрана электроны ведут себя как классические частицы и их движение можно описать и рассчитывать средствами классической механики и электродинамики.

Однако если на пути летящих в трубке электронов поставить одну или две щели, то после прохождения их через эти щели на светящемся экране наблюдается чередование темных и светлых полос. Получающаяся картина аналогична дифракционной картине, образующейся после прохождения через щели плоской световой волны. Значит, электроны, встречая на своем пути экраны с щелями, способны дифрагировать, т. е. ведут себя подобно волнам. Но волны эти необычны. Если выпускать из пушки по одному электрону, а на месте экрана поставить приборы, способные регистрировать удар лишь одного электрона (электронные умножители), то в каком бы месте за экраном ни был поставлен соответствующий приемник, всегда отмечается действие не части, а всего электрона! Электрон продолжает оставаться частицей, но его попадание в то или иное место экрана (или, более точно, вероятность попадания) определяется свойствами некоторой волны. При этом вероятность попадания, подобно энергии волны, определяется квадратом амплитуды такой гипотетической волны.

Если первоначально плоская световая волна несла энергию только в направлении ее распространения, то за щелью этот энергетический поток разделится по направлениям, отвечающим углам, определяющим максимумы дифракционной картины. Если поток подлетающих к экрану электронов несет энергию, то после дифракции на щели он распадается и поток энергии распределяется в пространстве подобно свету. Каждый отдельный электрон ведет себя как целая частица и лишь вероятность направления его полета подчиняется правилам, аналогичным правилам дифракции плоской волны на щели.

Такое своеобразное поведение электронов можно учесть, если формально сопоставить электрону некоторую волну (назовем ее *волной вероятности*). Соответствующая гипотеза была выдвинута французским физиком Л. де Бройлем, предположившим, что каждой частице может быть сопоставлена волна, длина которой определяется соотношением  $\lambda = h/p$ , где  $p$  — классический импульс частицы. Развивая идеи Л. де Бройля, Э. Шредингер ввел понятие *волновой функции*. Считая, что число электронов в некоторой области пространства пропорционально квадрату ее амплитуды (вспомним, что в обычной волне плотность энергии

также пропорциональна квадрату амплитуды), можно не только объяснить дифракционные эксперименты с электронами, но и количественно описать их.

Волна имеет чисто математический смысл и представление о ней нужно только для того, чтобы предсказать реально наблюдаемое распределение электронов в пространстве после прохождения их через щель или отверстие в экране.

Здесь мы имеем совершенно новую ситуацию. До сих пор предполагалось, что характеристики, вводимые для описания состояния тел (их координаты, скорости, ускорения), могут быть непосредственно измерены приборами и в этом смысле также являются наблюдаемыми. Волна де Бройля (подобно волновой функции) сама по себе принципиально не наблюдаема и служит лишь удобным средством описания.

Это не значит, конечно, что введение представления о волновых свойствах электронов есть чисто математический прием. Наоборот, как раз проявление в соответствующих экспериментах нового физического качества, которое не может быть понято в рамках классических представлений, приводит к необходимости ввести и новые средства математического описания этих представлений.

Если воспользоваться понятием волны вероятности (волновой функции) и считать, что такая волна обладает рядом свойств обычных, например, механических волн, то многие факты сразу же получают простое объяснение.

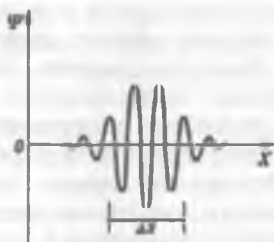
Например, если электрон находится в атоме, т. е. в ограниченной области пространства, то должны существовать своеобразные стоячие волны вероятности. Как известно, в обычных стоячих волнах их энергия сосредоточена в местах пучностей. Значит, должны быть и определенные области, где вероятность пребывания электрона также наибольшая (т. е. существуют стационарные области локализации электронов). Далее, длины стоячих волн и их энергии обладают свойством дискретности. Это сразу приводит к выводу о возможности существования дискретных энергетических состояний электронов в атомах.

Для волн выполняется принцип суперпозиции. Оказывается, что это имеет место и для волн вероятности и объясняет, в частности, появление повышенной электронной плотности между атомами в молекуле водорода, что и приводит к устойчивости молекулы.

При рассмотрении дифракции световых волн отмечалось, что в тех случаях, когда длина волны становится много меньше размеров отверстий в экранах, волновыми свойствами можно вообще пренебрегать и пользоваться представлениями геометрической или корпускулярной оптики.

Длины волн де Бройля становятся очень малыми при больших импульсах частиц, т. е. для тяжелых частиц. Волновые свойства таких частиц заметить практически невозможно. Поэтому можно пользоваться классической механикой.

рис. 9.1. Схематическое изображение волнового пакета



Если предсказать точное местоположение частицы можно только с определенной вероятностью, описываемой квадратом модуля волновой функции  $|\Psi|^2$  (обозначение волновой функции  $\Psi$ ), то безразлично, будет ли математически волна  $\Psi$  описываться действительными или мнимыми (комплексными) величинами. Однако квадрат ее модуля  $|\Psi|^2$  должен быть всегда действительным, так как эта величина определяет наблюдаемый результат. Поэтому в общем случае  $|\Psi|^2 = \Psi\Psi^*$ , где  $\Psi^*$  — комплексно сопряженная функция. Многие волновые функции являются комплексными. Если квадрат модуля волновой функции  $|\Psi|^2 = \Psi\Psi^*$  описывает вероятность пребывания частицы в некоторой области пространства, то в этой области  $|\Psi|^2$  резко возрастает (рис. 9.1). Такое образование называется *волновым пакетом*. Если на пути волнового пакета встретится щель, то пакет должен дифрагировать. Известно, что волновой пакет можно представить как сумму гармонических синусоидальных волн. Приписать определенную частоту и импульс можно лишь одиночной гармонической волне. Значит, для волнового пакета импульс «размыт» в некотором интервале. Нельзя говорить и об определенной области локализации частицы, так как волновой пакет не имеет резких границ.

Чем более узким является волновой пакет, тем больше число гармонических волн надо просуммировать, чтобы описать его. Если уменьшать возможный интервал значений координаты  $\Delta x$ , что соответствует созданию таких условий, когда волновой пакет сужается, то одновременно будет увеличиваться разброс  $\Delta p$  возможных значений импульса, т. е. снижается точность его измерения, и наоборот.

Это правило было сформулировано В. Гейзенбергом в виде соотношения неопределенностей:

$$\Delta p \Delta x \approx \hbar.$$

Это соотношение необходимо для понимания многих особенностей физики микромира. В частности, оно приводит к важнейшему следствию: *микрочастицы не могут покоиться!* В самом деле, если  $\Delta x \rightarrow 0$ , то  $\Delta p \rightarrow \infty$ , т. е. возрастает разброс кинетической энергии и, в частности, возрастает ее максимальное значение. Поэтому, во-первых, в атомах и молекулах и при температуре 0 К сохраняются движения электронов около ядра, ядер друг

относительно друга в молекулах (нулевые колебания). Во-вторых, чем меньше область пространства, в которой находятся частицы, тем большая энергия требуется для их удержания. Поэтому, например, появляющиеся внутри ядра при некоторых внутриядерных процессах электроны не могут удержаться в области ядер и вылетают наружу ( $\beta$ -распад).

Отмеченная выше связь между точностью измерения координат и импульсов микрочастиц не встречается в классической физике, в частности в классической механике. В ней полагается, что каждая из характеристик—координата, скорость, импульс—не зависима одна от другой и что соответствующие величины можно измерить с любой наперед заданной точностью, все дело только в том, насколько совершенны приборы.

Обсуждаемую связь между измерением координат и импульсов можно пояснить следующей наглядной аналогией. Представим себе, что некоторое тело движется вдоль линейки с делениями. Сфотографируем это тело в некоторый момент времени с выдержкой, отвечающей интервалу времени  $\Delta t$ . На фотографии получится резкое изображение неподвижного фона (линейки) и несколько размытое изображение предмета. Эта размытость будет тем большей, чем больше была выдержка, т. е. интервал  $\Delta t$ , и скорость движения предмета. При сильном размытии изображения предмета теряется возможность точно ответить на вопрос, где находился предмет в выбранный момент времени. Увеличивается, однако, точность измерения скорости, которая может быть определена по величине размытия отнесенного к интервалу времени  $\Delta t$ .

Чтобы точно определить местоположение предмета, надо уменьшать выдержку. Если сделать эту выдержку очень малой, то при не слишком больших скоростях получим резкое изображение предмета на фоне линейки. Однако при этом полностью теряется возможность отличить движущийся предмет от неподвижного.

Взаимосвязь между возможностью одновременного измерения координат и импульсов и переход к вероятностному описанию поведения микрочастиц приводят еще к ряду важных следствий. В классической физике вводилось понятие траектории, которая для материальной точки представлялась бесконечно тонкой линией. Для микрочастиц понятие траектории в классическом смысле не применимо. Можно говорить лишь о том, с какой вероятностью будет находиться частица в определенном месте пространства через определенный промежуток времени. Эта вероятность всегда меньше 100%. Вместо траектории (тонкой линии) получается некоторая область, в каждой точке которой с определенной вероятностью может находиться частица. Поэтому при описании поведения микрочастиц не пользуются понятиями скорости и ускорения, поскольку эти понятия теснейшим образом связаны с понятием траектории. Нет смысла сохранять и понятие силы, действующей на частицу, так как нет определенной точки

приложения этой силы. Однако сохраняется понятие потенциала, потенциальной энергии и функции. Понятие кинетической энергии также сохраняется, хотя изменяется соответствующее ей математическое выражение.

Разумеется, все это не может не привести к необходимости ввести новое уравнение движения. Вид этого уравнения был найден Э.Шредингером. Оно названо его именем. Базирующаяся на уравнении Шредингера механика называется *волновой* или *квантовой механикой*. В этой механике уравнение Шредингера играет ту же фундаментальную роль, что и уравнение Ньютона в классической.

## 9.2. Элементы квантовой механики

В квантовой механике состояние системы характеризуется волновой функцией  $\Psi$ , зависящей от координат и времени. Квадрат модуля волновой функции  $|\Psi|^2$ , взятый при данных значениях координат и времени, равен плотности вероятности того, что эти координаты окажутся в результате их измерения равными вычисленным значениям.

Волновая функция удовлетворяет дифференциальному уравнению движения Шредингера. Для совокупности взаимодействующих частиц с потенциальной функцией  $U$  оно имеет вид

$$-\frac{\hbar^2}{2} \sum_{k=1}^N \frac{1}{m_k} \left( \frac{\partial^2 \Psi}{\partial x_k^2} + \frac{\partial^2 \Psi}{\partial y_k^2} + \frac{\partial^2 \Psi}{\partial z_k^2} \right) + U\Psi = i\hbar \frac{\partial \Psi}{\partial t},$$

где суммирование проводится по всем частицам. Символом  $i$  обозначена мнимая единица. Классическая потенциальная функция взаимодействующих частиц при переходе к квантовой механике сохраняется без изменений.

Если потенциальная функция системы от времени не зависит, то уравнению Шредингера удовлетворяет произведение

$$\Psi(x, y, z, t) = \psi(x, y, z) \exp(-iEt/\hbar),$$

причем функция  $\psi(x, y, z)$  зависит только от координат системы. Параметр  $E$  есть *собственное значение* стационарного уравнения Шредингера:

$$-\frac{\hbar^2}{2} \sum_{k=1}^N \frac{1}{m_k} \left( \frac{\partial^2 \psi}{\partial x_k^2} + \frac{\partial^2 \psi}{\partial y_k^2} + \frac{\partial^2 \psi}{\partial z_k^2} \right) + U\psi = E\psi.$$

В результате решения этого уравнения находят собственные (возможные) значения параметра  $E$  и соответствующие ему решения, называемые *собственными функциями*. Физический смысл параметра  $E$  заключается в том, что он передает допустимые для данной системы значения полной энергии, получаемые в результате экспериментов, связанных, в частности, с обменом энергией системы с внешней средой.



Для любого уравнения Шредингера, соответствующего конкретной системе, существует бесконечное множество значений параметра  $E$ . Эти значения могут быть как непрерывными (для свободно движущихся частиц), так и дискретными, если частицы локализованы в малой области пространства.

Дискретные значения  $E$  называются *уровнями энергии*.

Для каждого собственного значения  $E_n$  существует своя собственная функция  $\psi_n$ . Если собственные значения  $E_n$  всегда действительные, то собственные функции могут быть как действительными, так и комплексными. Такое различие  $E_n$  и  $\psi_n$  объясняется тем, что все наблюдаемые характеристики определяются не собственной функцией, а квадратом ее модуля или другими математическими выражениями, в которые обязательно входят либо произведения функции на комплексно-сопряженную ей функцию, либо аналогичные произведения функций на их производные. Соответствующие комбинации всегда оказываются действительными.

Удовлетворяющие уравнению Шредингера волновые функции для разных собственных значений  $n \neq m$  ортогональны. Это

значит, что  $\int \psi_n^* \psi_m dV = 0$ . Интегрирование ведется по всей области задания функций.

Физический смысл ортогональности заключается в том, что система может находиться в состоянии либо с энергией  $E_n$ , либо  $E_m$ , но не в двух состояниях одновременно. Собственные функции непрерывны, причем ни одна из  $\psi_n$  не может обращаться в бесконечность. Собственные функции подчиняются условию нормировки, выражаемому равенством  $\int |\psi_n|^2 dV = 1$ .

Смысл этого равенства заключается в том, что полная вероятность обнаружить частицу или систему частиц во всей области, где они заведомо существуют (так как  $\psi_n \neq 0$ ), равна единице.

Важнейшую роль в квантовой механике играет понятие оператора физической величины. *Оператором* называется символическое изображение совокупности любых математических действий, производимых над функцией. Например, если функция  $f_1(x)$  умножается на число или другую функцию  $f_2(x)$ , то это число или функция является оператором. Если производится дифференцирование, то оператором является символ  $d/dx$  или  $\partial/\partial x$ . Дифференциальные операторы выступают не сами по себе, а лишь в сочетании с соответствующими функциями, на которые они действуют. Такие операторы обладают тем свойством, что результат их действия меняется в зависимости от последовательности расположения оператора и функции. Например,

$$\frac{d}{dx}(f_1 f_2) = f_1 \frac{d}{dx} f_2.$$

В квантовой механике каждой наблюдаемой физической величине сопоставляется свой оператор. Так, оператором  $x$ -компоненты импульса является символ  $\hat{p}_x = -i\hbar \frac{\partial}{\partial x}$ , оператором энергии—оператор  $\hat{H}$  и т. д.

В квантовой механике выполняется принцип суперпозиции, который проявляется в том, что для всех операторов имеется равенство

$$\hat{O}(f_1 + f_2) = \hat{O}f_1 + \hat{O}f_2,$$

где символом  $\hat{O}$  обозначен произвольный оператор. В математике такие операторы называются *линейными*.

Степени операторов определяются как операторы, действие которых равно последовательному действию оператора-основания на функцию столько раз, каков показатель степени оператора. Например, если

$$\hat{O} = \frac{d}{dx}, \quad \text{то} \quad \hat{O}^2 = \frac{d^2}{dx^2}, \quad \hat{O}^3 = \frac{d^3}{dx^3} \quad \text{и т. д.}$$

Пользуясь операторной символикой, можно переписать уравнение Шредингера в сокращенной форме:

$$\begin{aligned} & -\frac{\hbar^2}{2} \sum_{k=1}^N \frac{1}{m_k} \left( \frac{\partial^2 \psi}{\partial x_k^2} + \frac{\partial^2 \psi}{\partial y_k^2} + \frac{\partial^2 \psi}{\partial z_k^2} \right) + U\psi = \\ & = \left[ -\frac{\hbar^2}{2} \sum_{k=1}^N \frac{1}{m_k} \left( \frac{\partial^2}{\partial x_k^2} + \frac{\partial^2}{\partial y_k^2} + \frac{\partial^2}{\partial z_k^2} \right) + U \right] \psi = \hat{H}\psi = E\psi. \end{aligned}$$

Выражение

$$\hat{H} = -\frac{\hbar^2}{2} \sum_{k=1}^N \frac{1}{m_k} \left( \frac{\partial^2}{\partial x_k^2} + \frac{\partial^2}{\partial y_k^2} + \frac{\partial^2}{\partial z_k^2} \right) + U$$

называется *оператором Гамильтона* или *гамильтонианом*. Если воспользоваться выражениями для операторов компонент импульса

$$\hat{p}_{xk} = -i\hbar \frac{\partial}{\partial x_k}, \quad \hat{p}_{yk} = -i\hbar \frac{\partial}{\partial y_k} \quad \text{и т. д.,}$$

то оператор Гамильтона можно представить в виде

$$\hat{H} = \frac{1}{2} \sum_{k=1}^N \frac{1}{m_k} (\hat{p}_{xk}^2 + \hat{p}_{yk}^2 + \hat{p}_{zk}^2) + U.$$

Это выражение формально совпадает с классическим выражением для полной механической энергии совокупности материальных

точек, что указывает путь построения оператора Гамильтона (гамильтониана) и, следовательно, уравнения Шредингера для конкретных случаев: надо записать отвечающее классической механике выражение для полной механической энергии системы в декартовых координатах, применяя для кинетической энергии импульсное представление, а затем все составляющие классических импульсов заменить их операторами ( $p_x \rightarrow \hat{p}_x = -i\hbar \frac{\partial}{\partial x}$  и т. д.).

В этом проявляется связь между классической и квантовой механиками. Заметим, что, как и в классической механике, где выделены механики Ньютона и Эйнштейна, в квантовой механике также существует механика Шредингера (аналог механики Ньютона) и релятивистская квантовая механика Дирака.

Найдем решение многомерного уравнения Шредингера в том случае, когда его гамильтониан представляется суммой членов, каждый из которых является гамильтонианом одномерной задачи. Для простоты ограничимся двумя переменными.

Пусть  $\hat{H}(x, y) = \hat{H}(x) + \hat{H}(y)$ . Покажем, что полная функция  $\psi(x, y)$ , удовлетворяющая уравнению Шредингера с оператором  $\hat{H}(x, y)$ , есть  $\psi(x, y) = \psi(x)\psi(y)$ , а собственное число  $E = E_x + E_y$ , где  $E_x$ ,  $\psi(x)$  и  $E_y$ ,  $\psi(y)$  — решения одномерных уравнений с гамильтонианами  $\hat{H}(x)$  и  $\hat{H}(y)$ . В самом деле, подставляя в исходное уравнение  $\hat{H}(x, y) \psi(x, y) = E \psi(x, y)$  собственную функцию в форме произведения  $\psi(x)\psi(y)$  и учитывая, что оператор  $\hat{H}(x)$  действует только на функцию  $\psi(x)$ , а оператор  $\hat{H}(y)$  только на функцию  $\psi(y)$ , получаем

$$\begin{aligned} \hat{H}(x, y) \psi(x, y) &= \hat{H}(x) \psi(x) \psi(y) + \hat{H}(y) \psi(x) \psi(y) = \\ &= \psi(y) \hat{H}(x) \psi(x) + \psi(x) \hat{H}(y) \psi(y) = \\ &= \psi(y) E_x \psi(x) + \psi(x) E_y \psi(y) = (E_x + E_y) \psi(x) \psi(y), \end{aligned}$$

где  $\hat{H}(x) \psi(x) = E_x \psi(x)$  и т. д.

Перейдем теперь к примерам решения стационарного уравнения Шредингера. Заметим, что в этих примерах, как и во всех последующих, используются лишь действительные собственные функции.

Частица в одномерном потенциальном ящике. Представим, что частица свободно движется вдоль оси  $X$  в интервале от 0 до  $L$ . Вне этого интервала она находиться не может. Свободно движущаяся частица обладает только кинетической энергией, и для нее классическая функция Гамильтона равна  $H = T_{\text{кин}} = \frac{1}{2m} p_x^2$ . Уравнение Шредингера имеет вид  $\left( \hat{p}_x = -i\hbar \frac{d}{dx} \right)$ :

$$-\frac{\hbar^2}{2m} \frac{d^2\psi}{dx^2} = E\psi.$$

Для его решения учтем, что вне интервала  $(0, L)$  частица не может находиться. Это означает, что вне этого интервала, включая и границы, функция  $\psi$  равна нулю. Вид уравнения показывает, что функция после двукратного дифференцирования меняет знак и умножается на постоянную. Такой функцией является  $N \sin ax$  или  $N \cos ax$ , но  $\cos ax$  не подходит, так как тогда при  $x = 0$  функция  $\psi$  не обращается в нуль. Поэтому решение будем искать в виде

$$\psi = N \sin ax.$$

Подставляя  $\psi$  в уравнение, получаем

$$[\hbar^2/(2m)] Na^2 \sin ax = EN \sin ax.$$

Отсюда  $[\hbar^2/(2m)] a^2 = E$ . Параметр  $a$  найдем из условия  $\psi(L) = 0$  или  $\sin aL = 0$ , откуда  $aL = n\pi$ , где  $n = 1, 2, 3, \dots$ . Таким образом,

$$E = E_n = \frac{\hbar^2}{2m} \frac{\pi^2}{L^2} n^2.$$

Следовательно, энергия  $E$  принимает дискретный набор значений. Каждому значению  $E_n$  соответствует своя волновая функция (собственная функция) вида

$$\psi_n = N \sin \frac{n\pi}{L} x.$$

Число  $n$  называется *квантовым числом*.

Низшее значение энергии получается при  $n = 1$ . Ему соответствует функция  $\psi_1 = N \sin (\pi/L)x$ , которая имеет максимум при  $x = L/2$ , т. е. наиболее вероятным будет пребывание частицы в центре интервала. Следующее значение энергии соответствует квантовому числу  $n = 2$  с собственной функцией  $\psi_2 = N \sin (2\pi/L)x$ . Функция  $\psi_2$  достигает экстремума в точках  $L/4$  и  $3L/4$ . При  $x = L/2$   $\psi_2^2 = 0$ . Наиболее вероятной областью локализации частицы является область около точек  $x = L/4$  и  $3L/4$ . Схема уровней энергии для рассматриваемой задачи и вид функции  $\psi$  и  $\psi^2$  показаны на рис. 9.2. Определим теперь постоянную  $N$ . Для этого воспользуемся *свойством нормировки*. Согласно этому свойству интеграл от квадрата собственной функции, взятый по всей области задания функции, равен единице. Следовательно,

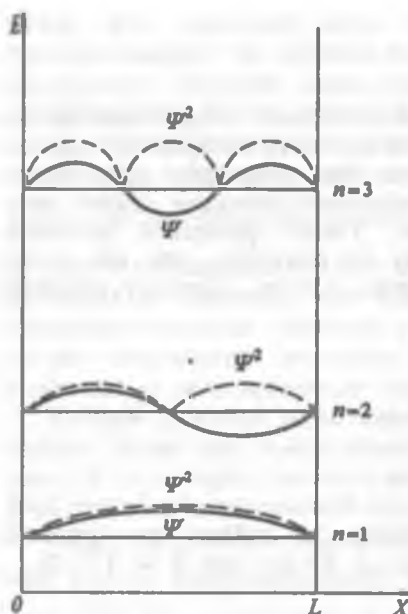


Рис. 9.2. Схема уровней энергии частицы в потенциальном ящике на отрезке  $(0, L)$  и вид функций  $\psi(x)$  и  $\psi^2(x)$ . Функции  $\psi(x)$  и  $\psi^2(x)$  для наглядности вычерчены в разном масштабе

$$N^2 \int_0^L \sin^2 \frac{\pi n}{L} x dx = 1.$$

Отсюда  $N = \sqrt{2/L}$  и

$$\psi_n = \sqrt{\frac{2}{L}} \sin \frac{n\pi}{L} x.$$

Рассмотренный пример позволяет сделать интересный вывод: расстояние между уровнями энергии и энергия самого низкого уровня существенно зависят от области локализации частицы. При очень больших  $L$  уровни энергии располагаются столь близко один к другому, что дискретность возможных значений энергии становится практически незаметной. Частица начинает вести себя подобно классической.

При  $L \rightarrow \infty$   $E_1 \rightarrow 0$ , т. е. ста-

новятся возможными, как в классической механике, нулевые значения энергии. Напомним, что согласно соотношению неопределенностей (см. выше) квантовая частица не может находиться в

состоянии покоя: ее кинетическая энергия всегда отлична от нуля.

Уровни энергии сближаются также при увеличении массы. Поэтому тяжелые частицы ведут себя «классически». Для легких частиц при малых областях локализации, во-первых, растет минимальное значение энергии, а во-вторых, увеличиваются расстояния между соседними уровнями.

Предположим, что частица удерживается в определенной области пространства (электрон около ядра или протон внутри ядра), т. е. находится в потенциальной яме, но не с бесконечно высокими стенками. Если яма не очень глубока и широка, то частица не удержится в ней (внутри ямы нет ни одного допустимого уровня энергии). Частица могла бы существовать в очень узких ямах (могла бы быть локализована в очень малой области), если удерживающие взаимодействия были бы велики. Отсутствие таких аномально сильных взаимодействий (кулоновские силы притяжения не столь уж велики) объясняет отсутствие электронов в непосредственной близости от атомного ядра или внутри него.

Частица на окружности. Рассмотрим свободную частицу на окружности радиуса  $r$ . Уравнение Шредингера в этом случае

легко записать, если вспомнить, что при переходе к движению по окружности вокруг заданной оси соответствующее классическое уравнение получается при замене координаты  $x$  углом поворота  $\varphi$  и массы—моментом инерции  $I$  точки относительно оси вращения. В результате получим

$$-\frac{\hbar^2}{2I} \frac{d^2\psi}{d\varphi^2} = E\psi.$$

Это уравнение аналогично уравнению, рассмотренному ранее.

Решениями его являются либо  $\cos l\varphi$ , либо  $\sin l\varphi$ , либо их линейная комбинация. Здесь  $l$ —некоторый параметр. Какое бы из решений мы ни выбрали, во всех случаях при подстановке в уравнение получим

$$E_l = \frac{\hbar^2}{2I} l^2.$$

Остается определить  $l$ . В данном случае у нас нет граничных условий (частица движется по всей окружности, а не только на ее участке, и для определения  $l$  нельзя воспользоваться приемом, примененным ранее). Согласно смыслу функции  $\psi$ , она не должна иметь разрывов на окружности радиуса  $r$ . Это приводит к требованию, чтобы при полном обходе окружности, когда  $\varphi$  меняется от 0 до  $2\pi$ , угол  $l\varphi$  менялся на величину от 0 до  $2\pi$ ,  $4\pi$ ,  $6\pi$  и т. д. Так как вращения по часовой и против часовой стрелки эквивалентны, то на основании всех рассуждений заключаем, что значения  $l$  должны передаваться положительными или отрицательными целыми числами. Если в качестве решения принять  $\cos l\varphi$ , то при  $l = 0$  получим единицу, т. е. собственная функция станет постоянной. Эта постоянная удовлетворяет уравнению Шредингера, так как соответствующее собственное значение  $E_0 = 0$ . В данном случае возможно нулевое значение энергии (но не функции  $\psi$ !), так как речь идет не о поступательном движении, а о вращении.

Итак,  $l = 0, \pm 1, \pm 2, \pm 3, \dots$

Параметр  $l$  называется *орбитальным квантовым числом*. Снова получаем, что энергия  $E$  частицы принимает дискретный ряд значений:

$$E_l = \frac{\hbar^2}{2I} l^2 = \frac{\hbar^2}{2mr^2} l^2$$

( $m$ —масса частицы). Расстояние между ближайшими уровнями энергии уменьшается с ростом  $r$  как  $r^{-2}$ .

Если энергия движения вдоль прямой квантована только в том случае, когда возможность его ограничена, то энергия движения по окружности всегда квантована.

Электрон в прямоугольной трубке. Рассмотрим частицу (электрон), движущуюся в прямоугольной трубке длины  $L$  с



Рис. 9.3. Электрон в потенциальной трубке; «металлическая модель» линейной молекулы с сопряженными связями

квадратным сечением со стороной  $a$  (рис. 9.3). Такая задача отвечает (в довольно грубом приближении) задаче о движении электрона в молекуле вдоль цепочки сопряженных связей.

Оси декартовых координат совместим с ребрами трубки, а начало с одним из ее углов. Поскольку частица движется свободно, то она обладает только кинетической энергией. Оператор Гамильтона имеет вид

$$\hat{H} = \frac{1}{2m} (p_x^2 + p_y^2 + p_z^2)$$

и ему соответствует уравнение Шредингера

$$-\frac{\hbar^2}{2m} \left( \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} \right) = E \psi.$$

В этом уравнении переменные разделяются, и поэтому

$$\psi(x, y, z) = \psi_x(x) \psi_y(y) \psi_z(z), \quad \text{а} \quad E = E_x + E_y + E_z,$$

причем функции  $\psi_x$ ,  $\psi_y$ ,  $\psi_z$  и собственные значения  $E_x$ ,  $E_y$ ,  $E_z$  определяются из решений соответствующих одномерных уравнений:

$$-\frac{\hbar^2}{2m} \frac{d^2 \psi_x}{dx^2} = E_x \psi_x \text{ и т. д.}$$

с граничными условиями

$$\psi_x(0) = \psi_x(L) = 0, \quad \psi_y(0) = \psi_y(a) = 0, \quad \psi_z(0) = \psi_z(a) = 0.$$

Решение одномерной задачи было найдено, что позволяет сразу написать результат:

$$\psi_{n_x, n_y, n_z} = \frac{2\sqrt{2}}{a\sqrt{L}} \sin \frac{n_x \pi}{L} x \sin \frac{n_y \pi}{a} y \sin \frac{n_z \pi}{a} z,$$

$$E_{n_x, n_y, n_z} = \frac{\pi^2 \hbar^2}{2m} \left( \frac{n_x^2}{L^2} + \frac{n_y^2}{a^2} + \frac{n_z^2}{a^2} \right).$$

Собственные числа и функции зависят от трех квантовых чисел  $n_x$ ,  $n_y$  и  $n_z$  в соответствии с тремя степенями свободы частицы. Вообще, число квантовых чисел равно числу всех ее степеней свободы. Нижнему уровню отвечают квантовые числа  $n_x = n_y = n_z = 1$ . При  $L \gg a$  ближайшими к нижнему будут

уровни с  $n_x \neq 1$ , но  $n_y = n_z = 1$ . Другими словами, длинноволновый спектр поглощения частицы (см. второй постулат Бора) в основном определяется отношением  $n_x^2/L^2$ , т. е. зависит главным образом от длины сопряженной цепи.

Можно заметить еще и следующее: энергия  $E_{n_x, n_y, n_z}$  пропорциональна сумме квадратов квантовых чисел  $n_y$  и  $n_z$ . Поэтому, например, получится одна и та же энергия, если  $n_x = 1$  и  $n_y = 2$  или  $n_x = 2$ , а  $n_y = 1$  и т. д. При этом, однако, собственные функции различны. Такая ситуация, при которой разному набору квантовых чисел  $n$ , следовательно, разным собственным функциям отвечает одна и та же энергия, называется вырождением.

Видно, что вырождение исчезает (снимается), если сечение трубки станет не квадратным, а прямоугольным. Следовательно, вырождение есть результат эквивалентности размеров области локализации вдоль определенных направлений.

Для вырожденных состояний характерно следующее важное свойство: любая линейная комбинация собственных функций вырожденного состояния снова является собственной функцией уравнения Шредингера. В самом деле, пусть  $n$  функций  $\psi_{1k}, \psi_{2k}, \dots, \psi_{nk}$  отвечают  $n$ -кратно вырожденному состоянию с

энергией  $E_k$ . Составим линейную комбинацию:  $\psi = \sum_{m=1}^n c_m \psi_{mk}$ .

Подставив эту функцию в уравнение Шредингера, получим

$$\hat{H}\psi = \sum_{m=1}^n c_m \hat{H}\psi_{mk} = E_k \sum_{m=1}^n c_m \psi_{mk} = E_k \psi.$$

### 9.3. Качественное объяснение основных закономерностей состояний атома водорода

В этом параграфе и ряде последующих рассмотрены некоторые общие вопросы строения и свойств реально существующих физических объектов—атомов и молекул. Основным постулатом в теории этих объектов является постулат о возможности использования потенциальной кулоновской функции в микромире, хотя закон Кулона был сформулирован на основе макроэкспериментов. Естественно возникает вопрос: почему можно так поступать? Точный ответ на этот вопрос возможен именно потому, что в природе имеется уникальный физический объект—атом водорода, состоящий всего из двух частиц: положительного ядра (протона) и электрона. В свою очередь, простота этого объекта позволяет решить квантовую задачу о его уровнях энергии совершенно строго, не прибегая ни к каким упрощающим допущениям.



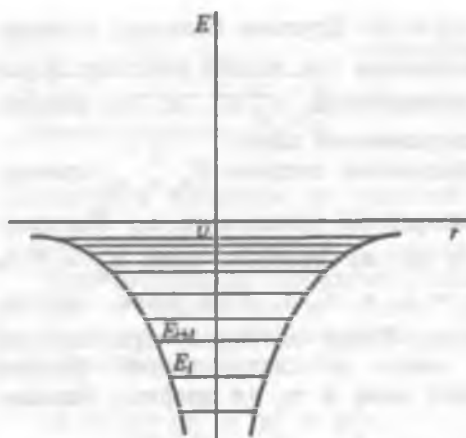


Рис. 9.4. Потенциальная яма, отвечающая кулоновскому притяжению электрона к ядру атома. Горизонтальные прямые в яме характеризуют уровни энергии

Очень хорошее согласие результатов такого решения с использованием кулоновского потенциала и данных спектральных экспериментов и доказывает правильность утверждения, что и в микромире электрические взаимодействия между заряженными частицами имеют тот же вид, что и в макромире.

Попробуем качественно представить некоторые основные закономерности, которые должны наблюдаться для уровней энергии и волновых функций атома. Ограничимся вначале простейшим атомом—атомом водорода. Этот атом имеет всего один электрон, находящийся в малой области пространства около ядра (сфера порядка  $10^{-8}$  см). Это сразу же приводит к выводу о дискретности энергетических состояний электрона. Естественно рассматривать движение электрона вокруг ядра в системе сферических координат, начало которой совмещено с атомным ядром. Тогда движения можно разделить на движения вдоль радиуса сферы и по поверхности сферы. Последние можно охарактеризовать углами  $\varphi$  и  $\theta$ , отсчитываемыми от некоторой оси и плоскости. Так как в атоме никаких взаимодействий (в хорошем приближении) между ядром и электроном, кроме кулоновского притяжения, нет, то потенциальная функция зависит только от расстояния электрона до ядра, т. е. от радиальной координаты.

Если учитывать одну степень свободы, связанную с радиальным перемещением, то электрон будет находиться в потенциальной яме, имеющей форму воронки (рис. 9.4). Эта потенциальная яма отличается от рассмотренной нами выше ямы тем, что она расширяется кверху и стенки ее не бесконечны. Если электрон находится достаточно близко к ядру (в глубине потенциальной ямы), где его движения ограничены стенками ямы, то его состояние будет квантованным. При этом чем ближе находится электрон к ядру, т. е. чем уже яма, тем больше должны быть и расстояния между соответствующими уровнями энергии. По мере увеличения энергии, т. е. появления возможности для электрона находиться на более удаленном расстоянии от ядра,

уровни энергии должны сближаться, практически сливаясь, когда электрон достигнет краев ямы.

Движение электрона по поверхности сферы никак на его потенциальной энергии не отражается, и его можно рассматривать как свободное вращение электрона вокруг ядра. Соответствующие такому движению уровни энергии также будут дискретными, причем эти уровни сближаются по мере того, как растет радиус вращения, т. е. электрон приближается к краям потенциальной ямы, задаваемой радиальной потенциальной функцией.

На основании всех этих рассуждений можно заключить, что уровни энергии электрона в атоме водорода образуют последовательность, сходящуюся к определенному пределу, после которого наблюдается непрерывная область значений  $E$  (непрерывный спектр), отвечающая области свободного полета электрона (отрыв от ядра и распад атома). Расстояния между уровнями монотонно уменьшаются.

Поскольку движение совершается в трехмерном пространстве, то уровни энергии и волновые функции зависят от трех квантовых чисел. Такие квантовые числа получили соответственно название *главного квантового числа  $n$*  (радиальная степень свободы), *орбитального  $l$*  и *магнитного  $m_l$*  (угловые степени свободы). Так как потенциальная функция зависит только от радиальной координаты и не зависит от угловых координат, то следует ожидать, что значения уровней энергии определяются только значениями главного квантового числа.

Действительно, соответствующие вычисления показывают, что если пренебречь всеми взаимодействиями, кроме кулоновского, то дискретные значения уровней энергии  $E$  подчиняются следующей формуле:

$$E_n = -\frac{mZ^2}{2\hbar^2} \frac{1}{n^2}.$$

Здесь  $n = 1, 2, 3, \dots$ —главное квантовое число;  $Z$ —заряд ядра (для атома водорода  $Z = 1$ );  $m$ —масса электрона;  $\hbar$ —постоянная Планка.

Орбитальное и магнитное квантовые числа в эту формулу, как и ожидалось, не входят. Это не значит, однако, что их вообще не надо принимать во внимание.

Чтобы понять это, учтем, что каждый уровень энергии, определяемый приведенной формулой, на самом деле является вырожденным.

Мы уже указали, что при определенных условиях уровни энергии оказываются вырожденными. Если воспользоваться предыдущими результатами и заменить прямоугольную трубку кубом, то для уровней энергии получим:

$$E_{n_x, n_y, n_z} = \frac{\hbar^2 \pi^2}{2mL^2} (n_x^2 + n_y^2 + n_z^2).$$

Тогда, например, для уровней энергии с квантовыми числами  $n_x = 2, n_y = 1, n_z = 1$ ;  $n_x = 1, n_y = 2, n_z = 1$  и  $n_x = 1, n_y = 1, n_z = 2$  получаем одно и то же значение. Возникает трехкратное вырождение. Видно, что вырождение связано с равноправностью движений по всем степеням свободы.

В атоме водорода движения по сфере, связанные с координатами  $\varphi$  и  $\theta$ , также равноправны. Поэтому и здесь следует ожидать вырождения, причем степень вырождения связана с главным квантовым числом. Расчеты показывают, что каждому значению главного квантового числа отвечает следующая совокупность орбитальных квантовых чисел:  $l = 0, 1, 2, \dots, n - 1$ .

Магнитное квантовое число зависит от орбитального и может принимать значения  $m_l = 0, \pm 1, \pm 2, \dots, \pm l$ .

Хотя энергетические состояния при наличии вырождения и оказываются одинаковыми при различных наборах квантовых чисел, однако это не означает, что одинаковыми будут и собственные функции. В самом деле, собственные функции для частицы в кубическом потенциальном ящике для состояний с наборами квантовых чисел, например  $n_x = 2, n_y = n_z = 1$  и  $n_y = 2, n_x = n_z = 1$ , оказываются различными.

Для атома водорода лишь одно состояние, а именно состояние с квантовым числом  $n = 1$  и, следовательно,  $l = m_l = 0$ , является невырожденным. В теории атомов такое состояние получило название  $1s$ -состояния. Символ  $s$  (от латинского слова синглет) обозначает состояние, для которого орбитальное квантовое число  $l = 0$ .

Для главного квантового числа  $n = 2$  уже возможны различные комбинации орбитального и магнитного квантовых чисел:  $l = m_l = 0$  ( $2s$ -состояние);  $l = 1, m_l = 0, l = 1, m_l = +1, l = 1, m_l = -1$  (три  $2p$ -состояния). Вырождение, таким образом, будет четырехкратным.

Состояния для  $n = 3$  характеризуются следующей схемой:

$$n = 3 \begin{cases} l = 0, m_l = 0 & (3s\text{-состояние}), \\ l = 1, m_l = 0, \pm 1 & (3p\text{-состояние}), \\ l = 2, m_l = 0, \pm 1, \pm 2 & (3d\text{-состояние}). \end{cases}$$

Следовательно, имеется девятикратное вырождение.

Соответствующие рассуждения можно продолжать, и мы увидим, что с ростом главного квантового числа степень вырождения соответствующих уровней энергии растет как  $n^2$ .

Появление вырожденных уровней в атоме водорода связано с эквивалентностью движений электрона по поверхности сферы (степени свободы, определяемые координатами  $\varphi$  и  $\theta$ ). Чтобы это вырождение выявить экспериментально, необходимо создать такие условия, когда движения по углам  $\varphi$  и  $\theta$  станут неэквивалентными. Этого можно добиться, если поместить атом во внешнее электрическое или магнитное поле. Тогда движения электрона вдоль векторов напряженности или индукции поля или перпендикулярно им окажутся различными. Вырождение снимется, и вместо одного уровня энергии будет наблюдаться ряд близко расположенных уровней.

Эффекты снятия вырождения и появления близко расположенных уровней или близко расположенных линий в спектрах при наложении электрического или магнитных полей получили названия эффекта Штарка и эффекта Зеемана (по имени физиков, их обнаруживших). Эти эффекты играют важнейшую роль для построения классификации электронных состояний в атомах, что, в свою очередь, лежит в основе современной трактовки периодической системы элементов Менделеева.

Мы рассмотрели атом водорода. Переход к более сложным атомам необходимо сделать поэтапно. На первом этапе не принимаются во внимание взаимодействия электронов. Каждый электрон движется только в поле ядра, а все изменения в уровнях энергии связаны лишь с изменением заряда ядра. Тогда говорят о водородоподобном атоме. В более сложной теории надо учесть взаимное отталкивание электронов. Это приведет к снятию вырождений и образованию более сложной картины уровней и спектральных линий.

## 9.4. Принципы спектрального анализа

При изучении условий устойчивого равновесия в классической механике и принципа плотнейшей упаковки атомов и молекул в кристаллах мы отмечали, что всякая система стремится принять такое состояние, при котором ее потенциальная энергия оказывается наименьшей.

Этот принцип сохраняется и в микромире. Только теперь уже надо говорить не о потенциальной, а о полной энергии, например, электрона в атоме. Это связано с тем, что квантовая частица согласно принципу неопределенности ни при каких условиях не может находиться в состоянии покоя. Поэтому она всегда обладает как потенциальной, так и кинетической энергией (если она даже локализована в пространстве).

Всякая микросистема (атом, молекула), если она обладала дополнительным запасом энергии, самопроизвольно переходит в состояние с наименьшим запасом энергии, т. е. в основное состояние. Если никакого воздействия извне не было, то избыток энергии выделяется в виде электромагнитного излучения.

Для атома водорода основным состоянием является состояние с главным квантовым числом  $n = 1$ . Чтобы атом перешел в состояние с другим квантовым числом, ему надо сообщить дополнительную энергию. Такой процесс перевода атома или молекулы из основного состояния в одно из состояний с большей энергией называется *возбуждением*. Возбудить атом или молекулу можно различными путями: облучением внешним электромагнитным полем с частотой волны  $\omega_{k1} = E_k - E_1/\hbar$  ( $E_1$  и  $E_k$  — энергия основного и возбужденного состояний,  $\hbar$  — постоянная Планка); за счет столкновений с другими атомами или молекулами, когда в энергию возбуждения переходит часть кинетической энергии частиц.

В соответствии с этим существует два способа получения сведений о разностях (по шкале энергии) между уровнями энергии атомов и молекул.

Первый заключается в наблюдении поглощения электромагнитных волн разной длины слоем вещества, размещенным между источником электромагнитных волн и анализирующим прибором. Источник электромагнитных волн обычно берется таким, чтобы он давал непрерывный спектр, т. е. чтобы в его излучении были представлены электромагнитные волны всевозможных длин в достаточно широком интервале. Такими источниками являются нагретые тела (лампы накаливания, нагретые до красного каления керамические стержни и др.). Если поток такого излучения направить на слой вещества, то будет наблюдаться избирательное поглощение электромагнитных волн с частотами, отвечающими требованию второго постулата Бора. На кривой зависимости энергии от длины волны, отвечающей источнику, появятся провалы. Здесь уместно обратить внимание на следующее. При взаимодействии атома или молекулы с электромагнитным полем они могут переходить в возбужденное состояние, отнимая энергию от поля. Однако находиться в возбужденном состоянии микросистема долго не может и самопроизвольно будет возвращаться в основное состояние. Время пребывания в возбужденном состоянии (время жизни в возбужденном состоянии или на уровне энергии) порядка  $10^{-9}$  с. При возвращении в основное состояние возникнет электромагнитное излучение. Общая излученная энергия, взятая за достаточно большой промежуток времени, равна поглощенной. В самом деле, как при поглощении, так и при излучении энергия при одном акте равна  $\hbar\omega = E_k - E_1$ .

Почему при этих условиях наблюдаются спектры поглощения? Если падающее на слой вещества излучение фокусируется на вход регистрирующего устройства (спектрометра), то порождаемые атомами или молекулами при их переходе из возбужденного состояния в основное фотоны «разлетаются» в разные стороны и лишь малая их часть попадает на щель спектрометра.

Второй способ наблюдения спектров состоит в том, чтобы возбудить атомы столкновением. Это можно сделать, воспользовавшись искровым разрядом между двумя электродами, разность потенциалов между которыми достаточно велика. Возникающее между электродами светящееся облако потому и является источником электромагнитных волн, что в нем образуется большое число быстро движущихся атомов и ионов, кинетической энергии столкновений которых хватает для возбуждения атомов.

Как в спектре поглощения, так и в спектре излучения получается дискретный набор полос или линий. Положение этих полос на шкале частот целиком определяется совокупностью уровней энергии атома или молекулы. Для водородоподобного атома совокупность всевозможных переходов между уровнями энергии, приводящих к поглощению или излучению электромагнитных волн с частотами  $\omega$ , передается формулой

$$\omega_{n, n'} = \text{const } Z^2 \left( \frac{1}{n^2} - \frac{1}{n'^2} \right),$$

где  $n$  и  $n'$  — главные квантовые числа для уровней энергии, между которыми совершается переход. Набор  $\omega_{n, n'}$  для каждого атома индивидуален, так как определяется зарядом  $Z$  ядра. Такая же индивидуальность характерна и для поглощения или излучения света молекулами. Поэтому оптический спектр вещества представляет собой своеобразный «отпечаток пальца». Следовательно, если в том или ином спектре наблюдается характерная для данного элемента совокупность линий, то можно утверждать, что этот элемент присутствует в анализируемой пробе. Чем больше в пробе атомов данного сорта, тем более яркими будут, например, линии излучения.

Определяя положения линий в спектрах и их интенсивности, можно решать задачи качественного и количественного анализов. Именно поэтому оптическая спектроскопия является одним из наиболее распространенных методов анализа вещества в самых разнообразных ситуациях. Это анализ газов, металлов, почв и т. д. С помощью спектров можно вести анализ дистанционно, что особенно важно для решения многих проблем контроля за состоянием окружающей среды.

## 9.5. Вид волновых функций атома водорода. Гибридизация

Перейдем теперь к анализу волновых функций атома водорода. Начнем, как и ранее, с некоторой аналогии, связанной с движением частицы в потенциальном ящике. Для наглядности рассмотрим движение частицы на плоскости в некотором квадрате так, что направления  $x$  и  $y$  являются эквивалентными. В этом случае для уровней энергии получим формулу

$$E = \text{const} (n_x^2 + n_y^2).$$

Собственные функции, отвечающие этой же задаче, имеют вид

$$\psi_{n_x, n_y} = N \sin \frac{n_x \pi x}{L} \sin \frac{n_y \pi y}{L},$$

где  $N$ —нормировочный множитель (его величина для нас сейчас не существенна).

Самое низкое энергетическое состояние получается тогда, когда значения квантовых чисел  $n_x$  и  $n_y$  равны единице. В этом случае собственная функция имеет вид

$$\psi_{1,1} = N \sin \frac{\pi x}{L} \sin \frac{\pi y}{L}.$$

Эта функция имеет максимум при значении  $x$  и  $y = L/2$ . Плотность вероятности нахождения частицы в определенном месте, которая определяется квадратом волновой функции, передается некоторым горбом (рис. 9.5), расположенным в центре квадрата.

Мы рассматриваем только действительные волновые функции, поэтому надо брать квадрат волновой функции, а не квадрат ее модуля, что требуется, если функция комплексная.

Поведение функции в направлении  $X$  и  $Y$  совершенно одинаково.

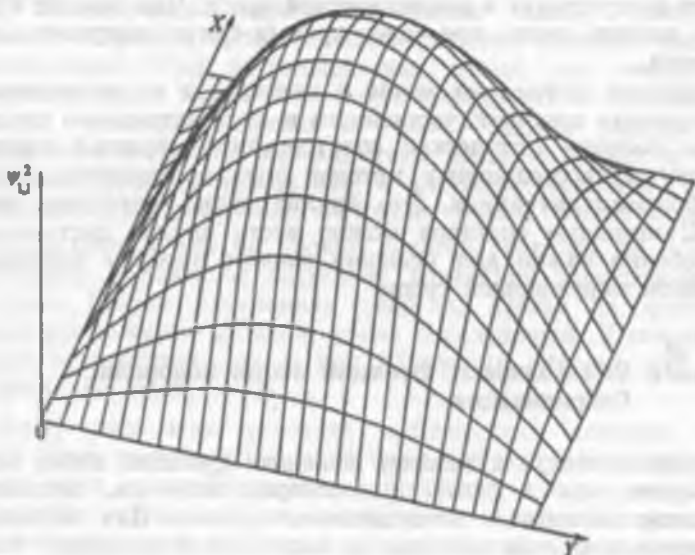


Рис. 9.5. Вид функции  $\psi_{1,1}^2$  для частицы на плоскости  $XY$

Рассмотрим теперь состояние, отвечающее значениям квантовых чисел  $n_x = 2$  и  $n_y = 1$ . В этом случае для волновой функции получаем

$$\psi_{2,1} = N \sin \frac{2\pi x}{L} \sin \frac{\pi y}{L}.$$

Наглядно вид функции и ее квадрата показаны на рис. 9.6, 9.7. В отличие от предыдущего примера квадрат такой функции обращается в нуль в центре плоскости и имеет четко выраженную направленность вдоль оси  $X$ , так что максимально вероятные возможные области расположения частиц оказываются сдвинутыми от центра и отвечают значениям координат  $x = L/4$ ,  $y = L/2$  и  $x = 3L/4$ ,  $y = L/2$ . Сходный вид имеют волновая функция и ее квадрат и в случае сочетания квантовых чисел  $n_x = 1$  и  $n_y = 2$ . Однако при этом функция  $\psi_{1,2}$  уже вытянута вдоль оси  $Y$ . Таким образом, получается некоторая направленность расположения волновых функций и их квадратов. Функции двух рассмотренных состояний как бы вытянуты одна перпендикулярно другой. По значению энергии состояний с  $n_x = 2$ ,  $n_y = 1$  и  $n_x = 1$ ,  $n_y = 2$  одинаковы (двойное вырождение).

Если рассмотреть трехмерную задачу, т. е. задачу о движении частицы в некотором кубе, то самое низкое невырожденное состояние отвечает сочетанию квантовых чисел  $n_x = n_y = n_z = 1$ . Следующим будет трижды вырожденное состояние, у которого  $n_x = 2$ ,  $n_y = 1$ ,  $n_z = 1$ ;  $n_x = 1$ ,  $n_y = 2$ ,  $n_z = 1$ ;  $n_x = 1$ ,  $n_y = 1$ ,  $n_z = 2$ . Для невырожденного состояния наибольшая плотность вероятности будет отвечать центру куба.

В трижды вырожденном состоянии получается как бы вытянутое в трех направлениях по осям  $X$ ,  $Y$  и  $Z$  образование, отвечающее новому распределению вероятностей, такому, что в центре куба вероятность пребывания частицы равна нулю. Для состояния  $n_x = 2$ ,  $n_y = 1$ ,  $n_z = 1$  получаются два наиболее вероятных состояния, когда координата  $x$  имеет значения  $L/4$  и  $3L/4$ . В двух других состояниях получается взаимно перпендикулярное расположение прямых, соединяющих максимумы функций  $\psi_2^2$  (облаков вероятностей распределения возможных состояний частиц).

Если вернуться снова к атому водорода, то окажется, что имеет место близкая аналогия. Для состояния  $1s$ , т. е. для состояния, когда главное квантовое число равно единице и два других квантовых числа равны нулю, получается функция, имеющая следующее аналитическое выражение:  $\psi_{1s} \sim \exp(-\rho/2)$ , где  $\rho = \text{const } (Zr/n^2)$  ( $r$ —радиальная координата



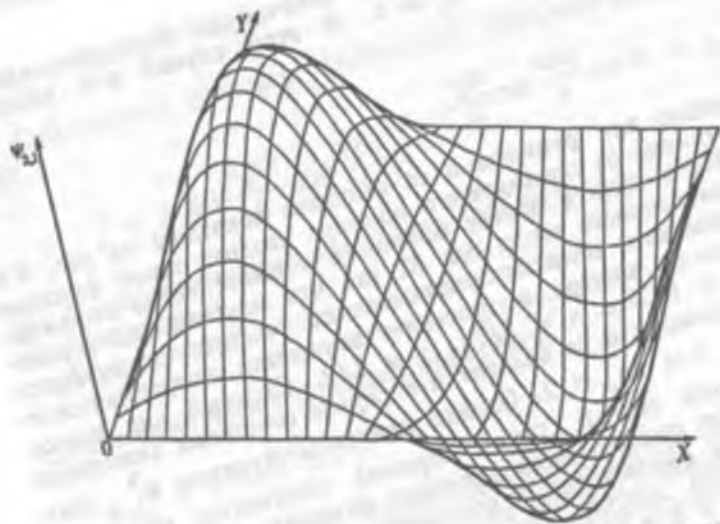


Рис. 9.6. Вид функции  $\psi_{2,1}$  для частицы на плоскости  $XY$

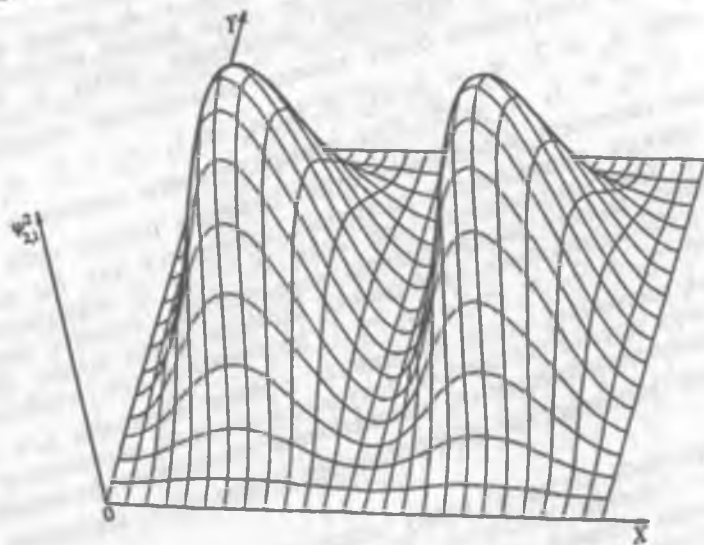


Рис. 9.7. Вид квадрата функции  $\psi_{2,1}^2$  для частицы на плоскости  $XY$

заряда ядра). Такая функция имеет вид сферы с максимумом плотности в начале координат и спадающей плотностью по мере удаления от центра к краям сферы.

Состояние  $1s$  является невырожденным. Состояние с главным квантовым числом, равным двум, является четырехкратно вырожденным. Среди них имеется одно состояние  $2s$  также, подобно  $1s$ , сферически симметричное. Его собственная функция равна

$$\psi_{2s} = \text{const} (2 - \rho) \exp(-\rho/2).$$

Кроме того, состоянию с  $n = 2$  отвечают три  $2p$ -функции. Для них имеется близкая аналогия с функциями для частицы, локализованной в пределах куба, т. е. появится четкая ориентация максимумов функций по осям координат.

Если ввести три оси координат  $X$ ,  $Y$  и  $Z$ , выбрав их взаимно перпендикулярными, и совместить начало координат с ядром атома, то соответствующие волновые функции могут быть записаны в виде

$$\psi_{2p_x} = x \exp(-\rho/2), \quad \psi_{2p_y} = y \exp(-\rho/2), \quad \psi_{2p_z} = z \exp(-\rho/2).$$

Вид квадратов этих функций (электронных облаков) показан на рис. 9.8. Таким образом, наиболее вероятными областями расположения электрона около ядра по осям  $X$ ,  $Y$ ,  $Z$  являются области, сдвинутые по отношению к началу координат на некоторые величины, подобно тому как сдвинуты по отношению к центру квадрата или куба области наиболее вероятного пребывания частицы на плоскости (в квадрате) или пространстве (кубе).

Это обстоятельство является весьма важным, и мы в дальнейшем на нем подробно остановимся. Именно этим объясняется появление вполне определенных пространственных форм молекул.

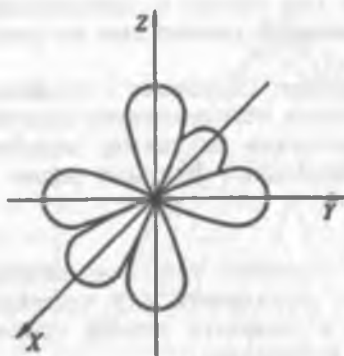
В § 9.2 было показано, что в тех случаях, когда имеется вырождение, линейная комбинация собственных функций уравнения Шредингера является решением этого же уравнения. Если построить линейную комбинацию из таких решений уравнения Шредингера для атома водорода, которые отвечают, например, четырехкратно вырожденным состояниям для главного квантового числа, равного двум, или девятикратно вырожденным состояниям для главного квантового числа, равного трем, то такие линейные комбинации снова будут одними из возможных решений уравнения Шредингера.

Пусть имеется некоторая функция (обозначим ее индексом  $(r)$  и назовем *гибридной*), равная

$$\psi^{(r)} = c_1 \psi_{2s} + c_2 \psi_{2p_x} + c_3 \psi_{2p_y} + c_4 \psi_{2p_z}.$$

Всякое решение уравнения Шредингера должно обладать свойством ортогональности и нормированности. Функция  $\psi^{(r)}$

Рис. 9.8. Квадраты волновых функций для  $p$ -состояния атома водорода



также должна быть нормированной. Это означает, что интеграл  $\int \Psi^{(r)2} dV$ , взятый во всем пространстве изменения соответствующей функции, равен единице.

Так как, в свою очередь, функции первоначального набора  $\Psi_{2s}$ ,  $\Psi_{2p_x}$ ,  $\Psi_{2p_y}$ ,  $\Psi_{2p_z}$  также ортогональным, т. е. для них справедливо  $\int \Psi_{2s}^2 dV = 1$  (аналогично для всех функций с индексом  $2p$ ), то

$$c_1^2 + c_2^2 + c_3^2 + c_4^2 = 1.$$

Если с помощью линейной комбинации выразить через исходный набор собственных функций еще одну гибридную орбиталь и потребовать, чтобы такие орбитали были ортогональными, то для двух гибридных орбиталей должно выполняться

соотношение  $\int \Psi_1^{(r)} \Psi_2^{(r)} dV = 0$ . Обозначая коэффициенты  $c_k$  в

этих функциях соответственно  $c_k^{(1)}$  и  $c_k^{(2)}$ , получаем

$$\sum_{k=1}^4 c_k^{(1)} c_k^{(2)} = 0.$$

Преобразования базисных функций, в которых коэффициенты подчиняются отмеченным выше правилам так, что сумма квадратов коэффициентов  $c_k$  в одной функции равна единице, а сумма попарных произведений коэффициентов в двух различных функциях равна нулю, называются *линейными ортогональными преобразованиями*. Если имеется некоторое решение уравнения Шредингера, являющееся вырожденным, то любое ортогональное преобразование этих собственных функций снова даст ортонормированный набор новых собственных функций, также явля-

Рис. 9.9. Негибридные атом-  
ные  $s$ - и  $p_x$ -орбитали (1) и  
гибридные  $sp$ -орбитали (2)



ющихся решениями уравнения Шредингера для этого же вырожденного энергетического состояния. Поэтому в случае вырождения решение уравнения Шредингера, т. е. собственные функции, находится с точностью до некоторого ортогонального преобразования.

Построенные с помощью ортогональных преобразований собственные функции атома водорода обладают некоторыми важными признаками. Например, рассмотрим пару гибридных функций, построенных из  $2s$ - и  $2p_x$ -функций следующим образом:

$$\Psi_{sp}^{(1)} = (\psi_{2s} + \psi_{2p_x})/\sqrt{2}, \quad \Psi_{sp}^{(2)} = (\psi_{2s} - \psi_{2p_x})/\sqrt{2}.$$

Нетрудно убедиться, что функции  $\Psi_{sp}^{(1)}$ ,  $\Psi_{sp}^{(2)}$  являются ортогональными и нормированными. Вид этих функций и их квадраты для проекции на плоскость, включающую ось  $X$ , показаны на рис. 9.9. Существенное отличие этих функций, например, от функции  $\psi_{2p_x}$  заключается в том, что они

приобретают не только направленность вдоль оси  $X$ , но и резко выраженную несимметрию по отношению к началу координат:

максимум функции  $\Psi_{sp}^{(1)}$  резко сдвинут вправо, в то время как

максимум функции  $\Psi_{sp}^{(2)}$  сдвинут на ту же величину, но влево.

Эти функции обладают тем свойством, что если их зеркально отразить в плоскости, проведенной через начало координат и совпадающей с плоскостью координатных осей  $YZ$ , то функция

$\Psi_{sp}^{(1)}$  перейдет в функцию  $\Psi_{sp}^{(2)}$ , и наоборот. В этом смысле получающиеся функции обладают свойствами симметрии (в данном случае зеркальной).

Приведенные гибридные функции являются не единственными функциями, которые можно построить из набора атомных  $2s$ - и  $2p$ -орбиталей, отвечающих значению главного квантового числа, равного двум. Может быть построен еще, например, следующий набор:

$$\Psi_1^{(r)} = (1/\sqrt{3})\psi_{2s} + (2/\sqrt{6})\psi_{2p_x},$$

$$\Psi_2^{(r)} = (1/\sqrt{3})\psi_{2s} - (1/\sqrt{6})\psi_{2p_x} + (1/\sqrt{2})\psi_{2p_y},$$

$$\Psi_3^{(r)} = (1/\sqrt{3})\psi_{2s} - (1/\sqrt{6})\psi_{2p_x} - (1/\sqrt{2})\psi_{2p_y}.$$

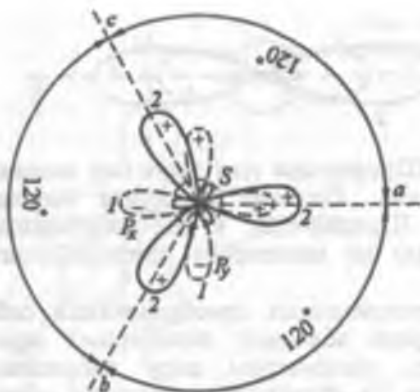


Рис. 9.10. Негибридные атомные  $s$ -,  $p_x$ - и  $p_y$ -орбитали (1) и симметризованные гибридные орбитали (2)  $sp^3$

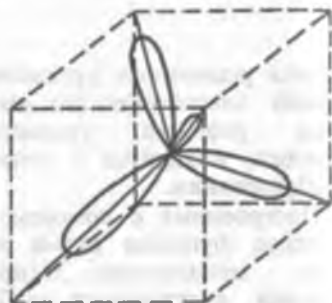


Рис. 9.11. Симметризованные гибридные орбитали типа  $sp^3$

Такие три гибридные функции также ортонормированы и имеют вид, показанный на рис. 9.10. Эти функции, во-первых, имеют резко выраженный сдвиг максимума в сторону от начала координат вдоль прямых, расположенных под углом  $120^\circ$  друг по отношению к другу, а во-вторых, они переходят друг в друга при повороте на этот угол. Еще более сложные гибридные орбитали могут быть получены в том случае, когда используются все четыре функции, отвечающие значению главного квантового числа, равного двум. Явный вид этих функций может быть записан так:

$$\Psi_1^{(r)} = (\psi_{2s} + \psi_{2p_x} + \psi_{2p_y} + \psi_{2p_z})/2,$$

$$\Psi_2^{(r)} = (\psi_{2s} - \psi_{2p_x} - \psi_{2p_y} + \psi_{2p_z})/2,$$

$$\Psi_3^{(r)} = (\psi_{2s} - \psi_{2p_x} + \psi_{2p_y} - \psi_{2p_z})/2,$$

$$\Psi_4^{(r)} = (\psi_{2s} + \psi_{2p_x} - \psi_{2p_y} - \psi_{2p_z})/2.$$

Схематически эти функции показаны на рис. 9.11. Они «вытянуты» вдоль диагоналей куба и каждая из четырех функций симметрична относительно «своей» диагонали. Если провести плоскость через любые две диагонали куба, вдоль которых «направлены» две гибридные волновые функции, то оси двух других функций будут лежать в перпендикулярной плоскости. Оси любой пары гибридных функций образуют друг с другом один и тот же угол  $109^\circ 28'$  (тетраэдрический угол). Все гибридные функции обладают свойством симметрии. В частности, если одну из диагоналей куба, вдоль которой «вытянута» одна из гибридных функций, принять за ось вращения, то при последо-

пительных поворотах вокруг этой оси на  $120^\circ$  три других гибридных функции будут переходить друг в друга.

Приведенные выше гибридные функции, полученные различными сочетаниями собственных  $2s$ - и  $2p$ -функций атома водорода, называются соответственно *атомными  $sp$ -,  $sp^2$ - и  $sp^3$ -орбиталями*. Они играют важнейшую роль для понимания пространственных форм молекул, на чем мы остановимся в § 9.7.

Не следует думать, однако, что если бы мы сумели фотографировать атом, то увидели бы объект с «торчащими» под определенными углами и отвечающими направлениям гибридных атомных орбиталей электронными облаками. Такая направленность появляется только тогда, когда есть взаимодействие данного атома с другим, т. е. в пространстве возникает выделенное направление. Свободный атом всегда сферически симметричен.

## 9.6. Принцип Паули и заполнение электронных оболочек

Выше мы рассматривали одну частицу в потенциальном ящике или один электрон около ядра (атом водорода). Представим теперь, что, как это и бывает в сложных атомах, около ядра размещено несколько электронов. Могут ли они все или хотя бы значительная их часть обладать одной и той же энергией или, что то же самое, находиться на одном и том же уровне энергии?

Если все электроны любого атома могли бы одновременно находиться на одном и том же основном (самом малом по энергии) уровне, то (в приближении невзаимодействующих электронов) частоты линий в оптических спектрах атомов (частоты излучаемых или поглощаемых электромагнитных волн) относились бы друг к другу как квадраты зарядов ядер. Эксперименты показывают, однако, что, даже если исключить такое взаимодействие, ожидаемого в указанном выше смысле сходства спектров не наблюдается.

Обобщение фактов по спектрам атомов и некоторых других позволило сформулировать один из важнейших в современной физике принципов—*принцип Паули* (по имени немецкого физика В. Паули, сделавшего это открытие). Этот принцип утверждает, что *в атоме в одном и том же состоянии, описываемом волновой функцией с квантовыми числами  $n$ ,  $l$  и  $m$ , не может находиться более двух электронов*. Это правило сохраняется и тогда, когда уровень энергии характеризуется любым набором квантовых чисел в соответствии с числом степеней свободы системы. Все равно на одном уровне энергии не может быть более двух электронов.

Оказывается, что «магическое число» два связано с наличием у электрона дополнительного свойства, которое называется *спином*. Спин электрона—это специфическое квантовое свойство,

которое по аналогии (но лишь по аналогии!) с классической физикой может быть названо собственным моментом импульса. Это, конечно, не значит, что электрон представляет частицу, вращающуюся вокруг своей оси. Аналогия с моментом импульса используется не более чем для наглядности.

Подобно моменту импульса движущейся по окружности частицы, возможные значения спина оказываются квантованными. Эксперименты показывают, что проекция спина электрона на любое выделенное направление может принимать лишь два значения:  $\pm \hbar/2$ . Эти величины можно рассматривать как дополнительное квантовое число, которое принимает два дискретных значения:  $\pm 1/2$  (в долях  $\hbar$ ).

Учитывая это, можно несколько иначе сформулировать принцип Паули: *в состоянии, определенном одним набором квантовых чисел, включая и спиновое квантовое число, может находиться не более чем один электрон.*

Принцип Паули и принцип минимального значения энергии в основном состоянии атома позволяют сформулировать закон заполнения возможных энергетических состояний в атомах с большим числом электронов. Если имеется только один электрон, то он должен находиться в состоянии  $1s$ , где главное квантовое число равно единице и значение энергии наименьшее. В этом же состоянии может находиться и второй электрон, причем электроны должны иметь разные спиновые квантовые числа, т. е. их спины должны быть направлены противоположно. Поскольку  $1s$ -состоянию соответствует только один набор пространственных квантовых чисел, то больше электронов в этом состоянии быть не может. Такая ситуация отвечает атому He (гелия). Следующий атом (литий) имеет три электрона. Этот третий электрон должен быть обязательно в одном из состояний с главным квантовым числом, равным двум. При дальнейшем заполнении энергетических уровней, т. е. при переходе от атома Li к атому Ne, надо иметь в виду следующее. Состояние с главным квантовым числом, равным двум, четырехкратно вырождено. Следовательно, этому состоянию отвечают четыре различных пространственных атомных орбитали с различными наборами орбитальных и магнитных квантовых чисел. В этих состояниях может размещаться, очевидно, восемь электронов—по два на каждую атомную орбиталь.

Электроны в следующих атомах заполняют состояния, которые отвечают главному квантовому числу, равному трем. Для этого состояния характерно девятикратное вырождение. Значит, на уровне энергии может быть восемнадцать электронов.

Продолжая эти рассуждения, нетрудно получить следующее общее правило: *на уровне энергии с главным квантовым числом  $n$  максимально можно разместить  $2n^2$  электронов.*

Это правило размещения электронов в атомах по уровням энергии приводит к вполне определенным закономерностям в спектрах, что делает возможным его экспериментальную провер-

ку. Так, например, в спектре неона уже нельзя наблюдать переходы между уровнями, отвечающими изменениям главного квантового числа от единицы до двух, поскольку соответствующие нижние состояния заняты электронами. Для того чтобы переходы между уровнями стали возможными, надо иметь на одном из них хотя бы одно «свободное место».

Здесь необходимо сделать одно принципиальное замечание. Выше при выводе правила заполнения уровней энергии или атомных орбиталей электронами не учитывались взаимодействия электронов. Между тем электроны в атоме испытывают взаимные влияния. Прежде всего это кулоновское отталкивание одного электрона от другого. Есть и другие типы взаимодействий, такие как магнитные.

Все это приводит к тому, что правило  $2n^2$  начинает нарушаться, особенно для атомов, где электронов много. Тем не менее оно продолжает сохраняться как достаточно хорошее приближение к реальности. Наличие нескольких электронов в атоме имеет еще одно следствие: возникают расщепления уровней, отвечающих одному и тому же значению главного квантового числа. В результате в области, соответствующей энергии с заданным значением главного квантового числа, появится целая группа уровней. Разумеется, это приведет к усложнению спектра, который становится гораздо более богатым линиями по сравнению со случаем, когда между отдельными электронами нет взаимодействия.

Современная квантовая теория атомов позволяет не только понять основные закономерности в спектрах, но и с большой степенью точности рассчитать положение отдельных линий. Очень хорошее совпадение рассчитанных и экспериментальных частот подтверждает правильность использованных основных закономерностей и соответствующего математического аппарата.

В заключение заметим, что во всех случаях мы использовали нерелятивистскую квантовую механику.

В квантовой механике существует два вида уравнений, определяющих состояние микрочастиц. Это уравнение Шредингера, которое в какой-то степени является аналогом ньютоновского уравнения движения в классической механике, и другое, более сложное уравнение—уравнение Дирака. Оно является аналогом уравнения для движения классической частицы со скоростями, сравнимыми со скоростью света. Подобно тому как уравнение в форме Ньютона не является инвариантным относительно преобразования Лоренца, уравнение Шредингера также не является релятивистски инвариантным. Релятивистски инвариантно только уравнение Дирака. Характерная особенность этого уравнения состоит в том, что в нем непосредственно учитывается существование спинов электронов.

На этом мы закончим обсуждение атома и перейдем к следующему по сложности объекту—молекуле.



## 9.7. Строение молекул

*Молекулой* называется электрически нейтральное образование, которое состоит из нескольких, иногда очень большого числа, ядер и соответствующего числа, тоже очень большого, электронов. Чтобы понять некоторые основные закономерности, определяющие строение молекул, обратимся к простейшему случаю молекулы водорода. Молекула водорода состоит из двух протонов (ядер атомов H) и двух электронов. Протоны расположены друг от друга на расстоянии приблизительно 1 Å ( $1 \text{ Å} = 10^{-8} \text{ см}$ ). Протоны, так же как и любые ядра в более сложных молекулярных образованиях, должны отталкиваться друг от друга по закону Кулона. Поэтому существование стабильной частицы может быть объяснено только тем, что взаимное отталкивание протонов компенсируется притяжением этих протонов к электронам. Чтобы такое электронно-ядерное притяжение компенсировало ядерно-ядерное отталкивание, должна быть достаточно большая степень вероятности нахождения электронов между ядрами. Поэтому не случайно одной из первых моделей двух-атомной молекулы была своеобразная планетарная модель, в которой электрон, располагаясь в плоскости между двумя ядрами, вращался вокруг оси, соединяющей эти два ядра (рис. 9.12). Эта модель, разумеется, страдает тем же недостатком, что и модель атома, в которой электроны вращаются вокруг ядра. Как в том, так и в другом случае вращающиеся заряженные частицы должны были бы непрерывно излучать электромагнитную энергию, что экспериментально не подтверждается. Тем не менее сама идея о том, что отрицательный заряд должен располагаться между двумя положительно заряженными ядрами, является совершенно правильной. Достаточно точные расчеты, осуществленные на основании методов квантовой механики, показывают, что действительно в области между ядрами должна быть повышенная плотность вероятности нахождения электрона, т. е. должна быть повышенная плотность заряда в электронном облаке, имитирующем задаваемое квадратом модуля волновой функции распределение отрицательного (электронного) заряда.

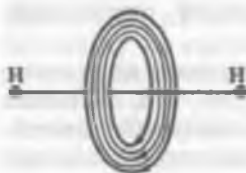


Рис. 9.12. Планетарная модель молекулы водорода

Подобные расчеты можно провести, пользуясь следующим общим приемом. Чтобы понять его, вернемся к уравнению Шредингера для многих частиц.

В § 9.2 мы ввели понятие оператора. Используя его, уравнение Шредингера можно переписать в символической форме:

$$\hat{H}\Psi = E\Psi,$$

где символом  $\hat{H}$  обозначено выражение

$$\left[ -\frac{\hbar^2}{2} \sum_k \frac{1}{m_k} \left( \frac{\partial^2}{\partial x_k^2} + \frac{\partial^2}{\partial y_k^2} + \frac{\partial^2}{\partial z_k^2} \right) + U \right].$$

Составим теперь выражение:

$$\Psi^* \left[ -\frac{\hbar^2}{2} \sum_k \frac{1}{m_k} \left( \frac{\partial^2}{\partial x_k^2} + \frac{\partial^2}{\partial y_k^2} + \frac{\partial^2}{\partial z_k^2} \right) + U \right] \Psi = E |\Psi|^2.$$

В символической форме с использованием понятия оператора это выражение может быть записано следующим образом:

$$\Psi^* \hat{H} \Psi = E |\Psi|^2.$$

Проинтегрируем это выражение по всей области задания функции  $\Psi$ , т. е. вычислим интеграл

$$\int \Psi^* \hat{H} \Psi dV = E \int |\Psi|^2 dV.$$

Теперь представим, что мы предварительно решили уравнение Шредингера и нашли совокупность его собственных значений  $E_n$  и собственных функций  $\Psi_n$ . Подставим эти функции в выражение для записанного интеграла. Тогда

$$\int \Psi_n^* \hat{H} \Psi_n dV = E_n \int |\Psi_n|^2 dV.$$

Если учесть, что функции  $\Psi_n$  должны быть нормированы, то следует положить  $\int |\Psi_n|^2 dV$  равным единице. Окончательно получим выражение

$$E_n = \int \Psi_n^* \hat{H} \Psi_n dV,$$

где  $E_n$  — энергия, которую может иметь частица в одном из стационарных состояний.

Пусть известна не точная функция  $\Psi_n = \Psi_{n \text{ точн}}$ , т. е. не точное решение уравнения Шредингера, а некоторая приближенная функция  $\Psi_{n \text{ пригл}}$ . Тогда

$$E_{n \text{ пригл}} = \int \Psi_{n \text{ пригл}}^* \hat{H} \Psi_{n \text{ пригл}} dV.$$

По мере уточнения решения, т. е. по мере того как функция  $\Psi_{n \text{ пригл}}$  приближается к  $\Psi_{n \text{ точн}}$ , значение  $E_{\text{пригл}}$  стремится к  $E_{\text{точн}}$ . Такое свойство интеграла указывает путь поиска приближенных решений уравнений Шредингера в произвольном случае: надо найти приближенное решение уравнения Шредингера и затем, варьируя его, например, с помощью изменения параметров соответствующей приближенной волновой функции, так менять значение  $E_n$ , чтобы оно стремилось к  $E_{n \text{ точн}}$ . При

этом приближенное значение  $E_0$  для основного состояния, т. е. состояния наименьшей возможной энергии, оказывается всегда большим, чем точное значение энергии, отвечающей основному состоянию:  $E_{0 \text{ прибл}} > E_{0 \text{ точн}}$  (здесь символом «0» обозначено основное состояние). Значит, варьировать значение приближенной волновой функции нужно таким образом, чтобы  $E_{0 \text{ прибл}}$  все время уменьшалось до некоторого предельного значения. Тогда  $E_{0 \text{ прибл}}$  становится практически равным  $E_{0 \text{ точн}}$ . Такой способ математического поиска решения соответствует физическому принципу, согласно которому любая система, предоставленная сама себе, стремится занять положение с наименьшей полной энергией.

Можно провести еще и такую аналогию. Представим, что в молекуле или в атоме электрон первоначально располагался произвольно. Под действием притяжения к ядру он начал бы двигаться, принимая значение с наименьшей энергией. Это фактически и означало бы, что его первоначальная волновая функция постепенно переходила бы в точную волновую функцию основного состояния.

Вернемся теперь к молекуле водорода. Запишем уравнение Шредингера для двух электронов, движущихся в кулоновском поле двух протонов:

$$\left[ -\frac{\hbar^2}{2m} \left( \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial y_1^2} + \frac{\partial^2}{\partial z_1^2} \right) + U_1(1) - \frac{\hbar^2}{2m} \left( \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial y_2^2} + \frac{\partial^2}{\partial z_2^2} \right) + U_2(2) + U_2(1) + U_1(2) + U(1, 2) \right] \Psi = E\Psi.$$

Здесь  $x_1, \dots, x_2, \dots$  — декартовы координаты первого и второго электронов;  $U_1(1)$  и  $U_2(2)$  — потенциальные функции кулоновского взаимодействия первого электрона с первым протоном и второго электрона со вторым протоном;  $U_2(1)$  и  $U_1(2)$  — потенциальные функции кулоновского притяжения первого электрона ко второму ядру и наоборот;  $U(1, 2)$  — потенциальная функция электрон-электронного отталкивания.

Первые два члена в уравнении полностью совпадают с теми, которые отвечали бы уравнению Шредингера для движения электрона в поле одного (первого) протона, т. е. уравнение Шредингера для атома водорода с первым протоном в качестве центра. Аналогичный смысл, но для второго протона имеют вторые два члена.

Если воспользоваться символической операторной записью и обозначить  $\hat{H}(1)$  — гамильтониан задачи о движении первого электрона в поле лишь первого протона (задача об атоме водорода) и ввести аналогично обозначение  $\hat{H}(2)$  для второго

электрона, то уравнение Шредингера запишется в следующей форме:

$$[\hat{H}_1(1) + \hat{H}_2(2) + U_2(1) + U_1(2) + U(1, 2)]\Psi = E\Psi.$$

Такое уравнение можно было бы решить описанным выше приближенным способом. Для того чтобы построить некоторую приближенную волновую функцию, будем рассуждать следующим образом. Молекула водорода образуется в результате объединения двух атомов водорода, электроны которых находятся в основных  $1s$ -состояниях. Пусть два атома водорода находятся друг от друга на достаточно больших расстояниях, когда их взаимодействия еще малы. Тогда около каждого атома будет свое электронное облако, плотность которого будет определяться квадратом функции  $\psi_{1s}^2$ , центрированной на первом или втором атомных ядрах\*). Общее распределение атомной плотности в пространстве, охватывающем два атома, можно передать квадратом волновой функции  $\psi_{1s}^{(1)} + \psi_{1s}^{(2)} = \Psi$ , где  $\psi_{1s}^{(1)}$  и  $\psi_{1s}^{(2)}$  — сферические атомные  $1s$ -волновые функции, центрированные на первом и втором ядрах (рис. 9.13). Такая форма волновой функции удовлетворяет условию, что когда рассматривается область около ядра  $H_1$ , то  $\psi_{1s}^{(2)} = 0$  и все распределение электронной плотности определяется функцией  $\psi_{1s}^{(1)2}$ . В окрестности ядра  $H_2$  эта плотность передается функцией  $\psi_{1s}^{(2)2}$ .

При сближении ядер картина меняется. Электрон первого атома испытывает влияние второго ядра, и наоборот. В результате электронная плотность первого ядра вытягивается в сторону второго, а второго в сторону первого. Следовательно, электрон первого ядра часть времени проводит около второго, и наоборот. Соответственно вероятность пребывания каждого из электронов около «своего» ядра должна уменьшаться. Это можно учесть, если функции  $\psi_{1s}^{(1)}$  и  $\psi_{1s}^{(2)}$  умножить на коэффициенты соответственно  $c_1$  и  $c_2$ , меньшие единицы. Тогда функция  $\Psi$  примет вид

$$\Psi = c_1\psi_{1s}^{(1)} + c_2\psi_{1s}^{(2)}.$$

Такую форму приближенной волновой функции уже всей молекулы принято называть *молекулярной орбиталью* в виде *линейной комбинации атомных орбиталей*.

Коэффициенты  $c_1$  и  $c_2$  можно найти с помощью приема поиска минимума  $E_{\text{приб}} = \int \Psi \hat{H} \Psi dV$ . Мы, однако, поступим здесь иначе, воспользовавшись тем, что вследствие симметрии

\* Еще раз укажем, что здесь и ниже мы используем действительные функции.

молекулы водорода распределение электронной плотности на одинаковых расстояниях в ядерной области справа и слева от ядер  $H_1$  и  $H_2$  должно, очевидно, быть одинаковым.

Так как электронная плотность в молекуле, отвечающая функции  $\Psi = c_1\psi_{1s}^{(1)} + c_2\psi_{1s}^{(2)}$ , равна

$$\Psi^2 = c_1^2\psi_{1s}^{(1)2} + 2c_1c_2\psi_{1s}^{(1)}\psi_{1s}^{(2)} + c_2^2\psi_{1s}^{(2)2},$$

то требуемая симметрия будет иметь место, только если  $c_1^2 = c_2^2$  или  $c_1 = \pm c_2$ . Следовательно,

$$\Psi = c(\psi_{1s}^{(1)} \pm \psi_{1s}^{(2)}).$$

Из условия

$$\int \Psi^2 dV = 1 = c^2 \left( \int \psi_{1s}^{(1)2} dV + \int \psi_{1s}^{(2)2} dV \pm 2 \int \psi_{1s}^{(1)} \psi_{1s}^{(2)} dV \right) = c^2 (2 \pm 2S)$$

может быть найден коэффициент  $c$ :  $c = 1/\sqrt{2 \pm 2S}$ . Символом  $S$  обозначен интеграл  $\int \psi_{1s}^{(1)} \psi_{1s}^{(2)} dV$ . Этот интеграл меньше единицы (но положителен) и вообще невелик, так как область, где обе функции  $\psi_{1s}^{(1)}$  и  $\psi_{1s}^{(2)}$  одновременно отличны от нуля (область перекрывания) сравнительно небольшая.

Таким образом, мы нашли две возможные волновые функции молекулы водорода (рис. 9.13—9.16):

$$\Psi_1 = (\psi_{1s}^{(1)} + \psi_{1s}^{(2)})/\sqrt{2 + 2S},$$

$$\Psi_2 = (\psi_{1s}^{(1)} - \psi_{1s}^{(2)})/\sqrt{2 - 2S},$$

или, пренебрегая  $S$ ,

$$\Psi_{1,2} = (\psi_{1s}^{(1)} \pm \psi_{1s}^{(2)})/\sqrt{2}.$$

Пользуясь соотношением

$$E_{\text{прибл}} = \int \Psi_{\text{прибл}} \hat{H} \Psi_{\text{прибл}} dV,$$

находим два возможных уровня энергии:

$$E_{1,2} = \left[ \int \psi_{1s}^{(1)} \hat{H} \psi_{1s}^{(1)} dV + \int \psi_{1s}^{(2)} \hat{H} \psi_{1s}^{(2)} dV \pm 2 \left( \int \psi_{1s}^{(1)} \hat{H} \psi_{1s}^{(2)} dV \right) \right] / 2$$

(оказывается, что  $\int \psi_{1s}^{(1)} \hat{H} \psi_{1s}^{(2)} dV = \int \psi_{1s}^{(2)} \hat{H} \psi_{1s}^{(1)} dV$ ).

Функции  $\Psi_1 = (\psi_{1s}^{(1)} + \psi_{1s}^{(2)})/\sqrt{2}$  отвечает электронная плотность, распределенная по закону  $\Psi_1^2 = (\psi_{1s}^{(1)2} + 2\psi_{1s}^{(1)}\psi_{1s}^{(2)} + \psi_{1s}^{(2)2})/2$ . Так как значения функций  $\psi_{1s}^{(1)}$  и  $\psi_{1s}^{(2)}$  всюду

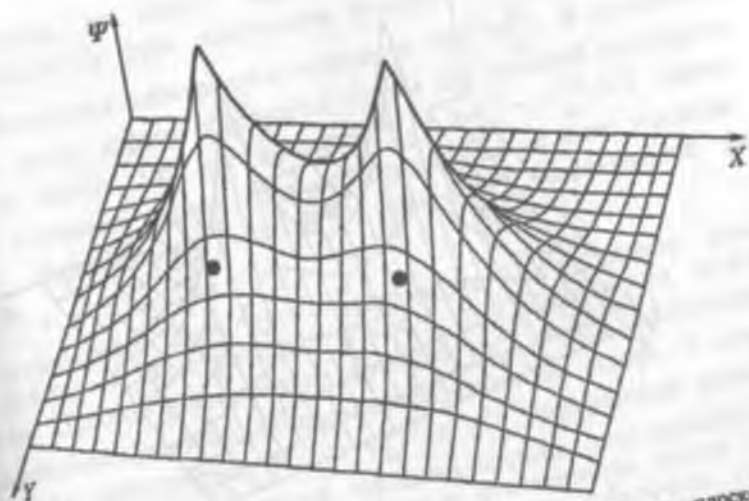


Рис. 9.13. Карта значений функции  $\Psi = (\psi_{1s}^{(1)} + \psi_{1s}^{(2)})/\sqrt{2}$  на плоскости, проведенной через связь в молекуле водорода. Точками обозначены места расположения ядер атомов водорода (протонов)

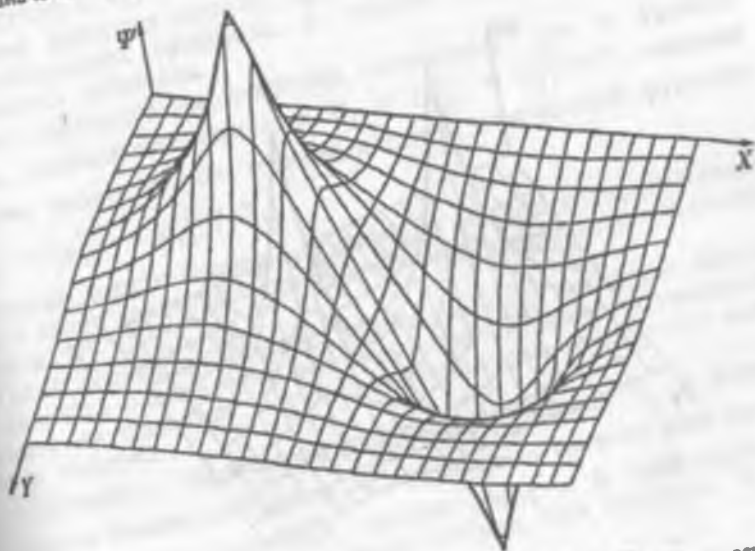


Рис. 9.14. Карта значений функции  $\Psi = (\psi_{1s}^{(1)} - \psi_{1s}^{(2)})/\sqrt{2}$  на плоскости, проведенной через связь в молекуле водорода

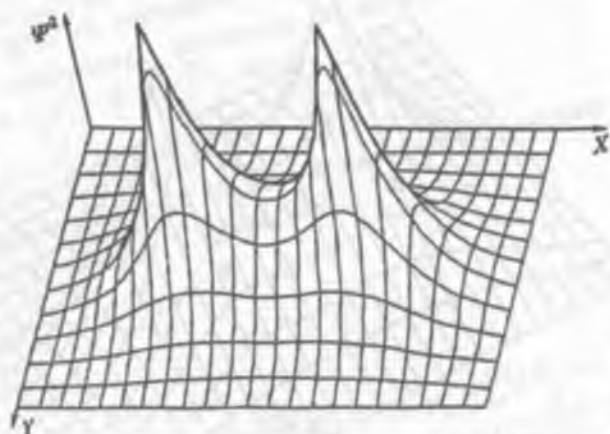


Рис. 9.15. Карта значений функции  $\psi^2 = (\psi_{1s}^{(1)} + \psi_{1s}^{(2)})^2/2$  на плоскости, проведенной через связь в молекуле водорода

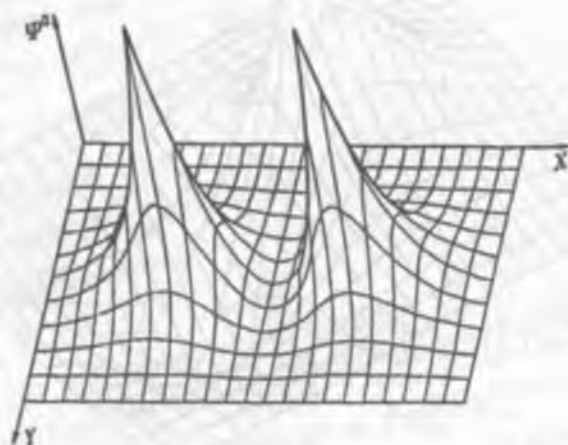


Рис. 9.16. Карта значений функции  $\psi^2 = (\psi_{1s}^{(1)} - \psi_{1s}^{(2)})^2/2$  на плоскости, проведенной через связь в молекуле водорода

положительны, то между ядрами к электронной плотности, получающейся за счет наложения квадратов функций  $\psi_{1s}^{(1)2}$  и  $\psi_{1s}^{(2)2}$ , добавляется электронная плотность  $2\psi_{1s}^{(1)}\psi_{1s}^{(2)}$ . В результате в центре между ядрами образуется сгусток электронной плотности. Нетрудно видеть, что для функции  $\Psi_2 = (\psi_{1s}^{(1)} - \psi_{1s}^{(2)})/\sqrt{2}$  электронная плотность между ядрами, создаваемая за счет перекрывания функций  $\psi_{1s}^{(1)2}$  и  $\psi_{1s}^{(2)2}$ , наоборот, уменьшается на  $2\psi_{1s}^{(1)}\psi_{1s}^{(2)}$ .

Удерживаться ядра друг около друга могут только из-за притяжения к электронному облаку, сосредоточенному между ядрами. Простое перекрывание функций  $\psi_{1s}^{(1)2}$  и  $\psi_{1s}^{(2)2}$  достаточной межъядерной электронной плотности создать не может. В самом деле, оба электронных облака  $\psi_{1s}^{(1)2}$  и  $\psi_{1s}^{(2)2}$  сферически симметричны и вместе с ядерными зарядами создают электронейтральные частицы—атомы. Чтобы возникла стягивающая атомы сила, необходимо, чтобы между ядрами появилась добавочная электронная плотность.

Поэтому мы вправе ожидать, что система будет устойчива в первом случае и неустойчива во втором.

Поэтому функция первого типа получила название *связующей молекулярной орбитали*, а второго типа—*разрыхляющей молекулярной орбитали*.

Более детальное рассмотрение показывает, что и уровень энергии, отвечающий функции  $\Psi_1 = (\psi_{1s}^{(1)} + \psi_{1s}^{(2)})/\sqrt{2}$ , меньше по своему значению, чем уровень энергии, отвечающий функции

$$\Psi_2 = (\psi_{1s}^{(1)} - \psi_{1s}^{(2)})/\sqrt{2}.$$

Эти уровни, в свою очередь, по энергиям находятся соответственно ниже и выше значения энергии для невзаимодействующих атомов водорода.

Такой результат позволяет понять, почему два атома водорода, находясь близко один около другого, стремятся объединиться в молекулу водорода: при таком объединении получается выигрыш в энергии.

Состояние, отвечающее собственной функции  $\Psi_2$  и энергии  $E_2$ , имеет большую энергию, чем суммарная энергия двух атомов водорода. Такое состояние следует отнести к возбужденному состоянию рассматриваемой нами системы. Для молекулы  $H_2$  оно будет неустойчивым.

Еще один весьма важный результат. Он заключается в том, что если первоначально имелись два одинаковых энергетических состояния с функциями  $\psi_{1s}^{(1)}$  и  $\psi_{1s}^{(2)}$ , принадлежащими различным



атомам, различным притягивающим центрам, то объединение их в одну систему приводит к появлению двух уровней энергии, т. е. первоначальное дважды вырожденное состояние расщепилось на два уровня с энергией, большей и меньшей, чем энергия, отвечающая невзаимодействующим атомам.

Вернемся снова к решению, для которого волновая функция равна

$$\Psi_1 = (\psi_{1s}^{(1)} + \psi_{1s}^{(2)})/\sqrt{2}.$$

Это решение, при котором перекрывание атомных орбиталей приводит к тому, что создается повышенная вероятность пребывания электрона между двумя положительными центрами (или между двумя ядрами) или повышенная плотность электронного облака.

Более детальное рассмотрение показывает, что вообще в любых молекулах, если представить некоторый распределенный в пространстве электрический заряд с плотностью заряда, равной  $|\Psi|^2$ , классическое электростатическое взаимодействие ядер атомов с электронным облаком компенсирует ядерные отталкивания. Действующие в молекулах силы притяжения и отталкивания имеют, таким образом, классическую природу, классическое толкование, но, конечно, квантовое поведение электронов отражается в той плотности, которую мы должны приписать для распределения отрицательного заряда внутри общего электронного облака.

Для того чтобы электронное облако за счет электрон-ядерного притяжения было способно компенсировать отталкивание ядер, плотность электрического отрицательного заряда между ядрами должна быть достаточно велика. Такая плотность создается именно тогда, когда атомные орбитали в середине между ядрами или вообще в пространстве между ядрами складываются между собой с некоторыми коэффициентами.

Если имеются одинаковые ядра, то такие коэффициенты оказываются одинаковыми. В случае, если ядра различные (например, в молекуле LiF), то уже не существует симметрия, а наибольшая плотность электронов сдвинута в сторону одного из ядер (атома фтора) (рис. 9.17).

У атома лития, от которого электроны оттягиваются в сторону фтора, возникает некоторый положительный нескомпенсированный заряд. Этот заряд притягивается к тому распределению электронной плотности, которая сосредоточена в основном около атома фтора. В свою очередь, присутствие атома лития приведет к тому, что распределение электронов вокруг атома фтора уже не будет симметричным относительно его центра. Максимум электронной плотности будет смещен к области между ядрами.

В результате на ядро фтора будет действовать кулоновская сила, стремящаяся сдвинуть его в направлении атома Li. Таким

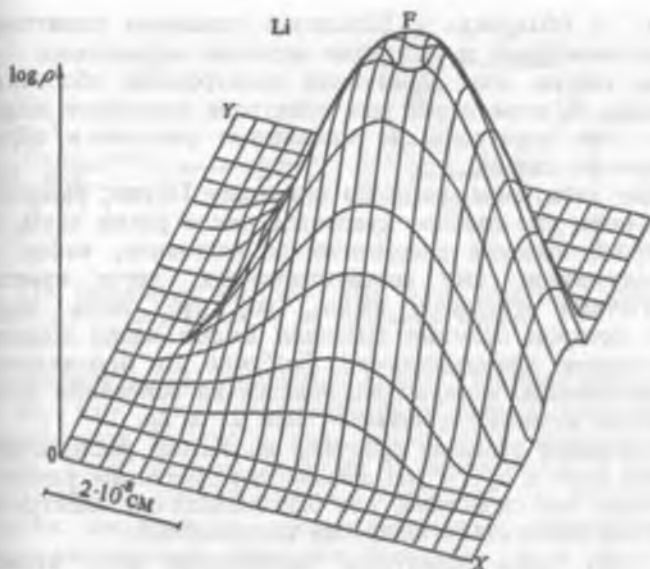


Рис. 9.17. Рельефное изображение валентной электронной плотности в молекуле LiF. Плоскость чертежа содержит связь LiF. Положения атомов Li и F указаны сверху рисунка. Видно, что практически вся электронная плотность сосредоточена вокруг атома F. Это характерно для чисто ионной связи. Стыгивающая молекулу сила (химическая связь) возникает потому, что положительно заряженное ядро атома Li, почти не экранируемое электронами, притягивается к сгустку электронного заряда вокруг атома F. Сила, действующая на ядро F в сторону Li, возникает потому что электронное распределение вокруг F несимметрично и имеет "оттенение" в сторону ядра Li. Это удастся заметить при использовании логарифмической шкалы при изображении электронной плотности

образом и создаются встречные силы, удерживающие два атома от разлета на бесконечные расстояния.

Попробуем теперь объяснить, почему молекулы имеют вполне определенную геометрическую форму или почему в молекулах связь образуется под вполне определенными углами, а не произвольно.

Для того чтобы связь образовалась, необходимо, чтобы электрон имел возможность совершать движение как около одного атома связи, так и около другого атома связи.

Для ядер с большими зарядами внутренние  $1s$ -электроны с очень большой силой притягиваются к своим ядрам. Заставить их двигаться в пространстве, существенно отличающемся от их первоначальной области локализации, за счет появления на некотором расстоянии от своего атома другого атома практически невозможно. Поэтому электроны, находящиеся на внутренних оболочках, обычно не принимают участия в образовании химических связей. Химические связи главным образом образуются за счет электронов внешних оболочек или за счет электронов,

находящихся в оболочках с большими главными квантовыми числами. Эти электроны и получили название *валентных*.

Вспомним теперь, как образуются электронные оболочки в атоме углерода. В этом атоме два электрона занимают нижнее  $1s$ -состояние. Эти электроны не принимают участия в образовании химических связей.

Следующие электроны, согласно принципу Паули, располагаются в состоянии, где главное квантовое число равно двум. Это не только более высокое энергетическое состояние, такое, что движение электронов, ему соответствующих, легче изменить из-за присутствия соседнего атома, но, что очень важно, состояние, в котором атомные орбитали имеют четко выраженную относительно направленность. Особенно это проявляется в гибридных состояниях, которые получаются на основании линейной комбинации атомных орбиталей типа  $2s$  и  $2p$ .

Такие гибридные атомные орбитали не только располагаются по отношению друг к другу под вполне определенным углом, но и обладают еще тем свойством, что отвечающая им электронная плотность четко сдвинута в одном из направлений.

Значит, если связь образуется наложением двух атомных орбиталей двух разных атомов, у которых ядра расположены на прямой, отвечающей направлению гибридной орбитали, то имеются все основания для того, чтобы электронная плотность в промежутке между двумя атомами имела максимальное значение. В этом случае и связь будет наиболее прочной, потому что притяжение двух ядер к сосредоточенному между ними значительному электронному заряду максимально.

Если теперь третий атом попадает в промежуток между двумя направленными гибридными орбиталями, то он не встретит на своем пути участок с достаточно большой электронной плотностью: связь в этом направлении образоваться не может (рис. 9.18). Наиболее выгодным направлением образования связи будет направление, отвечающее другой гибридной орбитали.

Гибридные орбитали могут образовывать друг с другом некоторые типичные сочетания, а именно: они могут иметь противоположное направление ( $sp$ -гибридизация); располагаться под углом  $120^\circ$  в одной плоскости ( $sp^2$ -гибридизация); образовывать некоторую пространственную фигуру, такую, что угол между каждой парой соседних орбиталей составляет  $109^\circ 28'$ , или являться тетраэдрическим ( $sp^3$ -гибридизация).

Значит, углерод с наибольшей вероятностью должен образовывать связи под углами  $180$ ,  $120$  и  $109^\circ$ . Такие связи характерны для органических соединений.

Например, молекула ацетилена имеет линейную структуру, в молекуле этилена связи образуются под углом  $120^\circ$ , в молекуле метана связи расположены так, что они из центра смотрят в вершины куба (рис. 9.19).

Так как образование связей вдоль определенных направлений является наиболее вероятным, то молекула и приобретает вполне

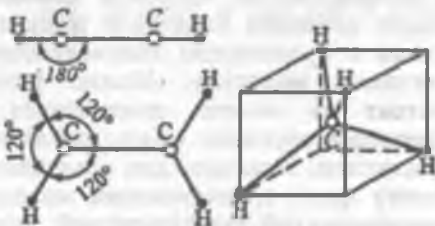
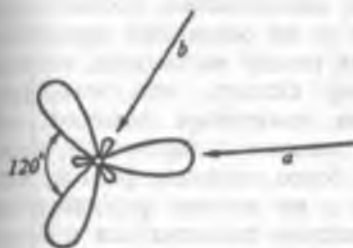


Рис. 9.18. Образование связи возможно вдоль направления *a* и невозможно вдоль *b*

Рис. 9.19. Структура молекул ацетилен, этилена и метана

определенную геометрическую форму. Обратим еще раз внимание на то, что такая геометрическая форма становится вполне закономерной именно в том случае, если выполняется принцип Паули. Если бы этот принцип не выполнялся и все электроны атома могли расположиться, например, в состоянии  $1s$ , то никаких выделенных преимущественных направлений образования химической связи не было бы, и в этом случае не наблюдались бы характерные устойчивые геометрические формы, свойственные органическим многоатомным молекулам.

Условие образования связи в молекулах преимущественно в тех направлениях, при которых атомные орбитали соседних атомов, принадлежащих одной связи, максимально налагаются друг на друга, что и приводит к появлению значительного электронного заряда между атомами, называется *принципом максимального перекрывания*. Он позволяет, зная свойства атомных орбиталей отдельных атомов, во многих случаях с достаточно высокой степенью точности предсказать геометрическую форму сложных молекул, не производя детальных вычислений.

Когда электронов в атомах очень много, то такая относительно простая картина, о которой мы говорили выше, начинает нарушаться. Мы не рассматриваем этот вопрос. Укажем лишь, что эти случаи можно описать не только качественно, но и количественно на базе общего аппарата квантовой механики.

Принципиально, что электронное облако (плотность электронного распределения) вокруг тяжелых атомов обладает почти сферической формой (подобно  $1s$ -состояниям легких атомов). Поэтому выделенных направлений, «выгодных» для образования химических связей, не возникает. Поэтому геометрические структуры соединений с тяжелыми элементами определяются не «направленными» свойствами электронных облаков тяжелых атомов, а в значительной степени электростатическими взаимодействиями между периферийными атомами, окружающими тяжелые.

Мы видели также, что объяснить существование молекул и их пространственных форм удастся на основании представлений

о специфическом волновом движении микрочастиц, соответствующего аппарата квантовой механики и на основании предположения о кулоновских взаимодействиях между частицами, составляющими молекулу. Можно поэтому сказать, что молекулы состоят из частиц, движущихся по квантовым законам, но взаимодействующих между собой по законам классической электростатики. Конечно, для объяснения более сложных фактов, чем просто факт существования молекул с их вполне определенной геометрической конфигурацией, необходимо пользоваться и более сложными представлениями, на которых, однако, мы останавливаться не будем.

Разумеется, ядра в молекулах не могут быть за счет кулоновских отталкиваний друг от друга и кулоновского притяжения к распределенному электрическому заряду жестко закреплены в каких-то относительных положениях. Они могут совершать около этих положений колебательные движения, что и проявляется в эксперименте в форме колебательных или инфракрасных спектров.

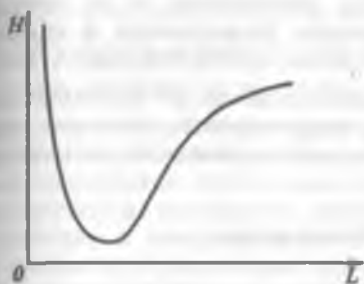
Таким образом, если спектр атома связан только с изменением состояния электронов или с электронными переходами, то в спектре молекул наряду с электронными переходами появляются и определенные спектральные линии, отвечающие ядерным.

Более того, молекула может вращаться как целое. Если такая система имеет дипольный момент, то это приведет к появлению вращательного спектра, наблюдаемого в далекой инфракрасной области или на границе инфракрасной области и диапазона коротких радиоволн.

Молекула существует за счет компенсации сил ядерного отталкивания силами электронно-ядерных притяжений. На первый взгляд создается впечатление, что возникает противоречие с теоремой Ирншоу, согласно которой система, состоящая из электрических зарядов, находится в состоянии устойчивого равновесия не может.

Однако никакого противоречия на самом деле не существует. Вспомним, что теорема Ирншоу справедлива для совокупности зарядов, которые не меняют величины в течение времени. Если расстояния между ядрами изменяются, то электронное облако меняет свои свойства. Например, при сокращении расстояния между ядрами электроны начинают как бы выталкиваться из межатомного промежутка. В результате электронная плотность уменьшается и стягивающая сила уже не может преодолеть ядерного отталкивания. В такой пульсации электронного заряда между ядрами проявляется принцип неопределенностей (см. § 9.1). В самом деле, чем меньше расстояния между ядрами, тем труднее удержать электроны между ними с помощью кулоновских притяжений этих электронов к ядрам. Конечно, чем больше расстояние между ядрами, тем легче попадают электроны в межъядерный промежуток, но до известного предела. Если это

Рис. 9.20. Функция полной энергии системы, состоящей из двух ядер и электронов



расстояние значительно увеличивается, то перекрывание атомных орбиталей, образующих волновую функцию данной связи, начинает обращаться в нуль.

Итак, ядра удерживаются в окрестности положения устойчивого равновесия потому, что изменение их взаимных расстояний приводит к необходимой соответствующей пульсации электронного облака. Такая пульсация не принимается во внимание при формулировке теоремы Ирншоу.

Попробуем теперь использовать некоторые другие результаты, полученные нами при обсуждении поведения электронов в потенциальном одномерном ящике, ограничиваясь двухатомной молекулой.

Электроны, находящиеся между ядрами и осуществляющие притяжение этих ядер, находятся в некоторой локализованной области пространства, связанной с расстоянием между ядрами. Их кинетическая энергия должна быть обратно пропорциональна квадрату расстояния между ядрами. Кроме того, электроны обладают потенциальной энергией притяжения к ядрам. Естественно, по аналогии с законом Кулона, предположить, что эта энергия обратно пропорциональна расстоянию между ядрами. Таким образом, для полной энергии электрона, находящегося в поле двух ядер, получаем, добавляя энергию  $H_{\infty}$  распада молекулы

$$H = a/L^2 - b/L + H_{\infty}.$$

Во второе слагаемое можно включить и подчиняющуюся закону обратных расстояний энергию отталкивания двух ядер. При соответствующих коэффициентах  $a$  и  $b$  функция имеет характерный вид с вполне определенным минимумом (рис. 9.20). Этот минимум и отвечает равновесному состоянию молекулы или ее наиболее вероятной длине.

Если рассмотреть движение ядер как достаточно тяжелых частиц в рамках классической механики, то можно заключить, что эти ядра будут совершать колебания около положения равновесия с частотой, определяемой формулой

$$\omega \sim \sqrt{k(m_1 + m_2)/(m_1 m_2)},$$

где  $k$ —кривизна ямы у положения равновесия, т. е. вторая производная от потенциальной функции, вычисленная в окрестности точки равновесия;  $m_1$  и  $m_2$ —массы соответствующих ядер.

С ростом масс атомов частоты колебаний будут уменьшаться. Этот эффект четко проявляется в инфракрасных спектрах при переходе от обычных молекул к их дейтерированным аналогам.

## 9.8. Особенности строения биомолекул

Остановимся на строении и некоторых важнейших свойствах биомолекул, т. е. крупных молекул, процессы в которых и взаимодействие между которыми определяют существование жизни.

Для того чтобы понять некоторые важнейшие особенности, определяющие специфику химических реакций в таких молекулах, обратимся к принципу максимального перекрывания атомных орбиталей.

Атомы в пространстве молекулы располагаются друг относительно друга в таких положениях, которые обеспечивают по возможности максимальное перекрывание валентных атомных орбиталей каждого из атомов (рис. 9.21).

При изменении взаимного расположения ядер такое максимальное перекрывание может уменьшаться, что является энергетически невыгодным. Это проявляется в том, что возникают определенные силы, стремящиеся вернуть форму молекулы к наиболее выгодной или, как говорят, равновесной.

Однако не все движения атомов друг относительно друга в молекуле могут приводить к изменению взаимного перекрывания атомных орбиталей. Чтобы понять это, рассмотрим расположение молекулярных орбиталей в молекуле этана (рис. 9.22). Видно, что заметно перекрываются атомные орбитали двух атомов углерода у  $\text{CH}_3$ -групп. Атомные орбитали, сосредоточенные

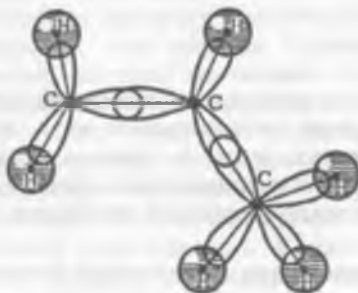


Рис. 9.21. Перекрывание атомных орбиталей в молекуле пропилена

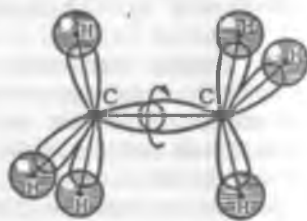


Рис. 9.22. Расположение молекулярных орбиталей в молекуле этана

внутри отдельных  $\text{CH}_3$ -групп, перекрываются для разных групп весьма слабо. Поэтому при таком движении, когда одна  $\text{CH}_3$ -группа вращается по отношению к другой  $\text{CH}_3$ -группе вокруг оси, проходящей через атомы углерода, взаимное перекрывание атомных орбиталей почти не меняется. Соответствующее движение называется *внутренним вращением в молекулах*. Внутреннее вращение в молекулах является квантованным, как и всякое вращение в микромире, и поэтому молекула может находиться в разных энергетических состояниях, которым отвечают различные взаимные расположения групп вокруг оси, проходящей через атомы углерода. Такие внутренние вращения в молекулах, а также другие геометрические перестройки, приводящие к появлению пространственных форм соединений, отличающихся друг от друга очень незначительными приращениями энергии, называются *конформационными превращениями*, а получающиеся при этом различные пространственные формы молекул — *конформационными изомерами* или *конформерами*.

Молекула, имеющая много различных внутренних вращений, может обладать большим числом различных геометрических форм.

Теперь обратимся к химическим реакциям. Химическая реакция произойдет, т. е. совершится определенная перестройка атомов, если реагирующие между собой атомы или атомные группировки окажутся расположенными близко друг к другу.

Если реакция происходит с участием малых молекул, двух-, трех- и четырехатомных, то она затрагивает сразу всю молекулу или все реагирующие пары или тройки молекул.

В крупных молекулах реакции могут происходить в отдельных частях молекулы, не затрагивая остальных.

Атомные группировки в крупных молекулах, способные к определенным реакциям, называются *реакционными центрами*. Реакция, в которой участвуют две крупные молекулы, будет осуществлена, если реакционные центры сблизятся между собой. Возможность такого сближения существенно зависит от геометрических форм реагирующих молекул. Если формы этих молекул не подходят одна к другой, то совместить, приблизить друг к другу реакционные центры не удастся. Тогда реакция происходить вообще не будет. Для того чтобы в крупной молекуле два или, может быть, несколько реакционных центров сблизилась друг с другом и произошла химическая реакция, необходимо соответствие одной формы молекулы другой, т. е. чтобы молекулы были *комплементарны*.

Такая комплементарность возникает при сближении двух крупных биологически активных молекул. В них, как правило, существует большое число внутренних вращений. Такое вращение требует затраты малой энергии. При сближении двух молекул под воздействием весьма слабых межмолекулярных взаимодействий происходят за счет многих вращений такие



взаимные изменения форм двух активных соответствующих друг другу молекул, что они начинают наиболее подходящим образом укладываться друг относительно друга, обеспечивая возможность определенной реакции.

Поэтому реакции между биологическими макромолекулами, в отличие от реакций для малых молекул, характеризуются строгой детерминированностью: в результате таких реакций с большой степенью однозначности создается вполне определенный и единственный продукт.

Такую детерминированность практически невозможно обеспечить, если опираться на реакции, происходящие между малыми молекулами. Именно поэтому жизнь возникла только тогда, когда в результате эволюции в природе образовались достаточно крупные органические и гетероорганические молекулы, обладающие, в частности, способностью к внутреннему вращению. Такое вращение требует, как правило, весьма малой энергии и приводит к тому, что биомолекулы переходят в состояние, обеспечивающее возможность той или иной реакции с весьма малыми внешними энергетическими затратами. Поэтому биохимические реакции идут при достаточно низких температурах, в то время как многие реакции между малыми молекулами требуют повышенной температуры, т. е. довольно существенных затрат внешней энергии.

## 9.9. Особенности спектров молекул. Цветность соединений. Симметрия и правила отбора

Обсудим еще некоторые важные общие свойства молекул. Если молекула не взаимодействует ни с какими другими частицами, то принимать или поглощать энергию она может только в форме электромагнитного излучения.

Частоты поглощаемых или излучаемых волн определяются правилом Бора, из которого следует, что чем больше расстояние между начальным и конечным уровнями энергии, тем больше частота (меньше длина волны) поглощаемого или излучаемого кванта.

Такое излучение или поглощение электромагнитной энергии осуществимо, если одно из энергетических состояний не полностью заселено электронами и электрон, в принципе, может перейти в это состояние.

Теперь рассмотрим два типа молекул: молекулы с несопряженными и сопряженными связями. Типичными представителями молекул с несопряженными связями являются, например, молекулы парафинового гомологического ряда. Типичными представителями молекул с сопряженными связями являются бензол, нафталин, антрацен и целый ряд других молекул. Особенности этих связей и соответствующих структурных элементов является то, что структурные элементы с несопряженными связями

достаточно обособлены и слабо зависят от окружения группировок.

Совокупности сопряженных связей обладают ярко выраженными коллективными свойствами. В случае несопряженных связей электроны разбиваются на группы, локализованные на этих связях. В ароматических соединениях появляются группы электронов, которые могут свободно перемещаться по всем соединенным между собой сопряженным связям, т. е. двигаться, например, по всему бензольному кольцу, по нафталиновым кольцам и т. д.

Таким образом, размеры областей локализации или областей возможных движений электронов в молекулах с несопряженными и сопряженными связями резко различаются между собой.

Расстояние между уровнями энергии, если рассматривать модель свободно движущихся электронов или электронов, находящихся в потенциальной яме, обратно пропорционально квадрату размера этой ямы или квадрату области локализации электронов. Поэтому расстояния между уровнями электронной энергии для молекул с несопряженными связями, где области локализации достаточно малы, будут большими и соответствующие линии излучения или поглощения должны наблюдаться в далекой ультрафиолетовой области спектра.

Расстояния между уровнями энергии в молекулах с делокализованными электронами, т. е. в таких молекулах, где электроны могут перемещаться по довольно значительному пространству, будут значительно меньше. Поэтому соответствующий спектр излучения или поглощения может наблюдаться уже в близком ультрафиолете или даже в видимой области. В последнем случае, очевидно, соответствующее химическое вещество обладает той или иной окраской. Действительно, опыт показывает, что большинство органических красителей обладает резко выраженной системой сопряженных связей.

Зависимость частоты поглощаемого света от области локализации электрона позволяет понять некоторые ключевые вопросы, связанные с одним из важнейших природных процессов — фотосинтезом.

В этом процессе в результате поглощения световой энергии происходит не только рост клеточной массы растений, но и выделяется кислород. При этом поглощаются кванты энергии, соответствующие примерно диапазону видимого света. Почему же природа «отобрала» именно этот диапазон? В самом деле, растение получило бы гораздо большую энергию, если бы в основном поглощались кванты с большей частотой или меньшими длинами волн (ультрафиолет).

Солнечные лучи проходят сквозь толстый слой атмосферы, состоящей из смеси газов, молекулы которых малы (кислород, азот, вода, озон и др.). В малых молекулах электронные уровни энергии вблизи основного состояния лежат довольно далеко друг от друга, и поэтому для таких молекул характерно поглощение

в ультрафиолете. Действительно, атмосфера сильно поглощает радиацию в этой области, выполняя и защитную функцию, так как облучение растений потоком квантов слишком большой энергии может разрушить живую ткань (*ультрафиолетовый ожог*).

Использование инфракрасной области спектра невыгодно не только по энергетическим соображениям, но и потому, что радиация в этой области длин волн сильно поглощается атмосферой (из-за наличия колебаний атомов в молекулах), особенно ее водяными парами.

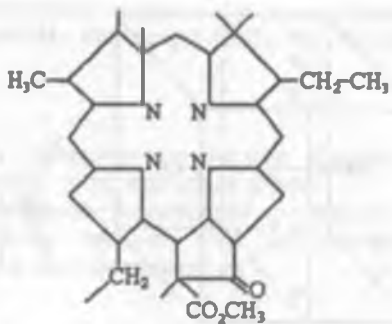
Остается видимая область. Чтобы эффективно поглощать энергию в этой области, молекула должна обладать достаточно большим набором близко расположенных уровней энергии. Значит, это должна быть крупная молекула с хорошо развитой системой сопряженных связей. Последнее требование обычно выполняется в плоских структурах. Действительно, приемником световой энергии в процессе фотосинтеза является молекула хлорофилла (рис. 9.23), обладающая всеми указанными выше свойствами. Эта молекула, однако, даже поглотив квант света, не может сама выделять кислород, да и вообще испытывать заметные превращения в структуре. В самом деле, как показывается в теории спектров молекул, при поглощении одного кванта меняется состояние лишь одного электрона. В результате в сложных молекулах с большим числом электронов получающиеся при этом изменения полной электронной плотности и свойств химических связей столь малы, что их практически можно не учитывать.

Какова же тогда роль молекулы хлорофилла? Роль эта заключается в том, что после перехода в электронно-возбужденное состояние (световая стадия фотосинтеза) в молекуле возникает подвижный электрон (опять-таки благодаря наличию близко расположенных возбужденных уровней энергии, образующих аналог энергетической зоны) и она становится хорошим донором электронов. Затем этот электрон (темновая стадия фотосинтеза) по цепи промежуточных агентов переносится на другой центр. Молекула хлорофилла становится положительно заряженной и превращается в сильный окислитель. Этот окислитель также через промежуточные агенты воздействует на молекулы воды, которые, разлагаясь, выделяют кислород.

Молекула хлорофилла возмещает потерянный электрон за счет молекул окружения. Процесс этот, подобно прохождению тока по проводнику, идет направленно.

В заключение этого раздела остановимся еще на одном важном вопросе, связанном со способностью молекул поглощать или излучать электромагнитную энергию. Атомы в молекулах располагаются друг относительно друга не произвольно, а подчиняясь вполне определенным закономерностям. В результате молекулы представляют достаточно сложные геометрические фигуры.

**Рис. 9.23. Структура молекулы бак-  
риохлорофилла**



Введем теперь понятие *симметрии*.

Некоторое пространство может быть разделено на симметричные подпространства, если каждой точке одного подпространства по вполне определенным правилам может быть сопоставлена эквивалентная (симметричная) точка другого подпространства.

Например, пусть имеется объем в виде куба, разделенный плоскостью  $\sigma$ , проведенной через центр куба параллельно одной из его сторон (рис. 9.24). Эта плоскость разделит весь объем куба на два полупространства—правое и левое. Выберем в левом полупространстве точку  $A$  и охарактеризуем ее положение расстоянием до разделяющей плоскости. Продолжим отрезок, соединяющий точку  $A$  с плоскостью, за плоскость в правое подпространство и отложим на этом отрезке точку  $A'$  на таком же расстоянии от плоскости, что и точка  $A$ , только вправо. Тогда точка  $A$  симметрична относительно точки  $A'$  по отношению к плоскости  $\sigma$ , разделяющей куб на два эквивалентных или симметричных подпространства. Мы вправе говорить о том, что эти два подпространства симметричны, поскольку точка  $A$  является совершенно произвольной, и если перебрать все точки левого подпространства, то можно найти эквивалентную симметричную точку правого подпространства. Плоскость, разделяющая пространство на два симметричных подпространства, называется *элементом симметрии*, в данном случае плоскостью симметрии.

Возьмем в центре того же куба на пересечении его диагоналей точку  $O$ . В пространстве этого куба выберем некоторую точку  $B$  (рис. 9.24) и проведем прямую, соединяющую точку  $B$  с центром куба, а затем продолжим ее за этот центр в сторону от точки  $B$ . На получившейся линии отложим точку  $B'$  на таком же расстоянии, на котором находится от центра точка  $B$ . Тогда точка  $B'$  симметрична точке  $B$  относительно центра. Так как точка выбрана в пространстве совершенно произвольно, то можно выбрать большую совокупность этих точек, составляющих целую область полупространства, таких, что каждая точка из этого полупространства будет иметь симметричную или эквивалентную ей точку в другом полупространстве. Соответственно точка  $O$ , которая как бы разделяет два полупространства друг относительно друга и переводит одно

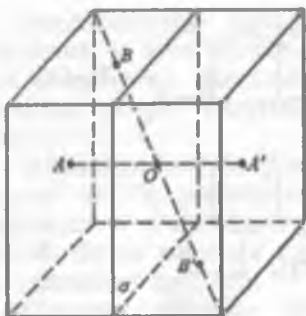


Рис. 9.24. Элементы симметрии в кубе

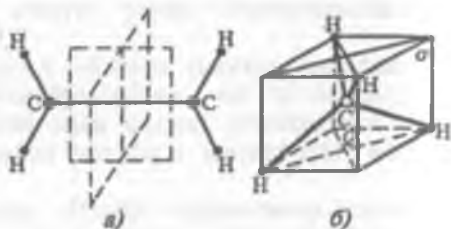


Рис. 9.25. Элементы симметрии молекул этилена (а) и метана (б)

полупространство в другое полупространство, называется *центром симметрии*. Центр симметрии также является элементом симметрии.

Мы не будем останавливаться на других элементах симметрии, к которым относятся оси симметрии разного порядка. Для дальнейших рассуждений нам достаточно этого.

Молекулы как геометрические тела также могут обладать определенной симметрией. Такие симметричные молекулы с их элементами симметрии показаны на рис. 9.25. На рис. 9.25,а, например, показана молекула этилена, разделенная плоскостью, перпендикулярной связи С-С и проведенной через ее центр, на две половины, включающие две  $\text{CH}_2$ -группировки. Такие группировки называются *эквивалентными* или *симметричными группировками*. На рис. 9.25,б показана молекула метана. В этой молекуле плоскость симметрии (их, вообще говоря, в этой молекуле несколько) может быть проведена так, как показано на рис. 9.25,б. Она тоже разделяет пространство молекулы и все ее связи на две симметричные части.

Очевидно, что свойства эквивалентных симметричных группировок должны быть совершенно одинаковы. В частности, если мы рассмотрим левое полупространство в молекуле этилена и в нем выделим некоторую произвольную точку, то плотность вероятности нахождения электрона в этой точке, очевидно, в точности равна плотности вероятности нахождения этого электрона в точке  $A'$  правого полупространства, являющейся симметричной к точке  $A$  по отношению к плоскости, делящей всю молекулу на две части. Вообще, плотность вероятности нахождения электрона в любой точке левого полупространства равна плотности вероятности нахождения в любой точке правого полупространства, если эта точка симметрична относительно выбранной плоскости точке левого полупространства.

Теперь вспомним, что плотность вероятности нахождения частицы в некоторой точке определяется квадратом волновой функции, т. е.  $\Psi^2(x, y, z)$ .

Если  $x, y, z$ —координаты некоторой точки в левом полупространстве, а  $x', y', z'$ —координаты симметричной ей точки в правом полупространстве, то

$$\Psi^2(x, y, z) = \Psi^2(x', y', z').$$

Очевидно, что аналогичное утверждение можно сделать и тогда, когда имеется не плоскость, а центр симметрии.

Из равенства значений квадратов некоторой функции для двух взаимно симметричных точек следует, что для волновых функций выполняется следующее соотношение:

$$\Psi(x, y, z) = \pm \Psi(x', y', z').$$

Это означает, что при переходе из левого полупространства в правое полупространство волновая функция либо должна оставаться неизменной, либо менять свой знак на противоположный. В первом случае волновая функция является *симметричной* относительно плоскости, во втором она *антисимметрична* относительно этой плоскости.

Можно показать, что если в какой-то системе имеется плоскость симметрии или центр симметрии, то волновые функции, отвечающие всем возможным энергетическим состояниям системы, делятся на два класса: симметричные относительно этой плоскости и антисимметричные.

С таким свойством волновой функции мы встречались ранее в § 9.7 при рассмотрении уровней энергии и волновых функций в молекуле водорода.

Напомним, что мы получили решение в виде двух функций:

$$\Psi_1 = (\psi_{1s}^{(1)} + \psi_{1s}^{(2)})/\sqrt{2},$$

$$\Psi_2 = (\psi_{1s}^{(1)} - \psi_{1s}^{(2)})/\sqrt{2}.$$

Первая функция симметрична относительно плоскости, проведенной через середину молекулы  $H_2$  перпендикулярно связи  $H-H$ . Вторая функция антисимметрична относительно этой плоскости. При перестановке функций  $\psi_{1s}^{(1)}$  и  $\psi_{1s}^{(2)}$  местами симметричная функция не меняет своего знака, а антисимметричная функция меняет свой знак на противоположный.

Такое поведение волновых функций в симметричных молекулах чрезвычайно важно и проявляется в целом ряде свойств этих систем.

Мы остановимся только на одном, связанном также со способностью молекул поглощать или излучать электромагнитное излучение.

В классической физике поглощать или излучать электромагнитное излучение может ускоренно движущийся заряд. Молекулы в целом не обладают электрическим зарядом. Это нейтральные системы, так же как и атомы. Однако эти системы могут иметь дипольный момент. В молекуле есть положительные

заряды—ядра и некоторый отрицательный заряд в форме электронного облака с разной плотностью отрицательного заряда в разных его участках. Центры этих положительных и отрицательных зарядов, которые можно определить по правилам, аналогичным правилам нахождения центров масс, могут не совпадать между собой.

Произведение суммарного положительного или отрицательного зарядов на расстояние между центрами суммарных положительных и отрицательных зарядов и дает дипольный момент молекулы.

При различных процессах внутри молекулы дипольный момент может меняться как в результате изменения величины эффективных зарядов, так и расстояний между ними.

В классической физике доказывается, что способность системы поглощать или излучать электромагнитную энергию связана с ускорением, с которым меняется дипольный момент, или со второй производной от дипольного момента системы по времени.

В квантовой механике аналогом этой величины является *матричный элемент дипольного момента*. Если ограничиться одним электроном и электронными переходами, то матричный элемент будет иметь вид

$$\mu = \int \Psi_n^* \Psi_m r dV.$$

Здесь  $r$ —радиус-вектор электрона;  $\Psi_n^*$  и  $\Psi_m$ —волновые функции начального и конечного состояний молекулы. Имеется в виду, что происходит переход из состояния с индексом  $n$  в состояние с индексом  $m$ . В результате такого перехода электромагнитная энергия излучается или поглощается.

Квадрат матричного элемента или квадрат интеграла дает вероятность перехода. Чем больше эта вероятность (предельное ее значение равно единице), тем энергичнее поглощает или излучает электромагнитную энергию совокупность молекул при изменении их состояния из  $n$  в  $m$ .

Если интеграл равен нулю, то система из состояния  $n$  в состояние  $m$  перейти не может, даже если одно из этих состояний и обладает свободным местом для размещения электронов.

Электромагнитные волны с длиной волны или частотой, соответствующей разности между такими уровнями энергии  $E_n$  и  $E_m$ , поглощаться совокупностью молекул не будут.

Представим, что молекула обладает центром симметрии. Через этот центр симметрии произвольно можно провести некоторую плоскость. Все пространство молекулы будет при этом разделено на два симметричных полупространства так, что каждая из точек этого полупространства имеет эквивалентную ей симметричную относительно центра симметрии точку в другом полупространстве. Выберем начало координат в точке, отвечающей центру симметрии. Проведенный в любом направлении

радиус-вектор из этой точки, например в правое полупространство, будет иметь симметричный ему, проведенный в противоположную, в левую, сторону полупространства эквивалентный радиус-вектор.

Теперь вернемся к интегралу  $\int \Psi_n^* \Psi_m r dV$ , определяющему матричный элемент (интегрирование производится по всему пространству, т. е. как слева, так и справа от плоскости). Как функция  $\Psi_n$ , так и функция  $\Psi_m$  должны по отношению к центру симметрии или в двух частях разделенного плоскостью пространства быть либо симметричными, либо антисимметричными.

Допустим, что функции  $\Psi_n$  и  $\Psi_m$  симметричны. Это означает, что при переходе из левой части пространства в правую часть пространства эти функции не меняют своего знака. Следовательно, для любой точки левого полупространства найдется эквивалентная ей симметричная точка правого полупространства, для которой произведение функции  $\Psi_n$  на  $\Psi_m$  будет иметь ту же величину, что и аналогичное произведение этих функций в эквивалентной точке левого полупространства.

В то же время радиус-вектор, проведенный из центра симметрии в выбранную точку в левом полупространстве, при переходе к симметричной точке в правом полупространстве меняет знак на противоположный.

Так как знак произведения  $\Psi_n^* \Psi_m$  при переходе из левого полупространства в правое полупространство не меняется, то все изменения будут связаны только с изменением знака радиуса-вектора. Значит, если при интегрировании по левому полупространству получается некоторое числовое значение интеграла, то при интегрировании по правому полупространству получится то же значение интеграла, но с противоположным знаком.

Полное интегрирование приведет, таким образом, к равенству интеграла нулю.

Аналогичные рассуждения, очевидно, можно привести и в тех случаях, когда функции  $\Psi_n$  и  $\Psi_m$  являются антисимметричными относительно центра симметрии. Тогда произведения этих функций справа и слева от плоскости также будут иметь один и тот же знак.

Равенство нулю интеграла показывает, что оптический переход между двумя соотношениями  $\Psi_n$  и  $\Psi_m$ , т. е. переход, для которого либо излучается, либо поглощается электромагнитная энергия, невозможен (запрещен).

Рассмотрим теперь другой случай. Пусть функция  $\Psi_n$  является симметричной, а функция  $\Psi_m$  антисимметричной относительно центра симметрии. Тогда, проводя аналогичные рассуждения, получим, что при переходе из правого полупространства в левое полупространство произведение функции  $\Psi_n$  на  $\Psi_m$  будет



менять знак. Но одновременно изменит свой знак и радиус-вектор частицы.

Таким образом, при интегрировании по правой и по левой частям мы получим одно и то же по величине и с одним и тем же знаком значение «полуинтеграла». Полный интеграл окажется отличным от нуля, и соответствующий оптический переход (излучение или поглощение электромагнитной энергии) из состояния  $n$  в состояние  $m$  будет разрешен.

На основании сказанного можно заключить, что при наличии центра симметрии оптические переходы разрешены лишь между состояниями, для которых волновые функции имеют разную симметрию.

Рассмотренный нами частный пример является случаем весьма распространенного проявления *правил запрета (отбора)* для излучения и поглощения электромагнитной энергии молекулами в тех случаях, когда эти молекулы имеют вполне определенные элементы симметрии. Именно наличие этих правил и позволяет, наблюдая экспериментально закономерности поглощения или излучения тех или иных молекул, судить о присущей им симметрии и в какой-то степени об их геометрической конфигурации. Поэтому правило отбора составляет одно из фундаментальных положений теории молекулярных спектров сложных систем.

## 9.10. Принцип действия оптических квантовых генераторов (лазеров)

Лазеры имеют очень широкое применение: как средство точного измерения расстояний, для целей связи, для технологической обработки тугоплавких материалов, в медицине, биологии и т. д. Есть попытки использования лазеров в сельском хозяйстве для предпосевной обработки семян и т.п. К сожалению, создано и лазерное оружие.

Лазеры были открыты сравнительно недавно (1960 г.). Правда, общая идея была высказана еще в 1940 г. советским физиком В.А. Фабрикантом, который в теоретической работе сделал вывод, что, по-видимому, такие устройства возможны. Но только через 20 лет соответствующие приборы в радиодиапазоне (*мазеры*) были реально созданы советскими физиками Н.Г. Басовым и А.М. Прохоровым и американцем Ч.Таунсом. Их открытие было быстро оценено и отмечено Нобелевской премией уже через несколько лет, что обычно бывает очень редко. Вслед за мазерами были изобретены оптические квантовые генераторы—лазеры.

Слово «лазер» образовано из первых букв английского названия таких устройств. Эти устройства позволяют получать электромагнитное излучение, в частности в видимом диапазоне, чрезвычайно большой мощности. Кроме того, лазеры способны генерировать свет одной длины волны с очень высокой степенью

монохроматичности, что также нельзя получить при использовании обычных источников света. Лазеры могут работать в разных режимах: непрерывно излучать волны или же излучать в очень короткие промежутки времени (за  $10^{-11}$  —  $10^{-13}$  с). Еще одно полезное свойство лазеров заключается в том, что можно получать очень узкие и мало расходящиеся пучки света, что позволяет производить световую локацию удаленных объектов, например Луны, с очень большой степенью точности. В частности, с помощью лазера расстояние 10 км можно измерить с точностью до 1 мкм ( $10^{-6}$  м). Также можно измерить с помощью лазеров и очень малые скорости движения—до долей мм/с и т. д.

Для понимания физических принципов действия лазеров рассмотрим некоторые общие вопросы, связанные с излучением и поглощением света совокупностью атомов или молекул. Для простоты рассмотрим газ, т. е. совокупность частиц, которые взаимодействуют друг с другом только при столкновениях. Пусть этот газ находится в полости шарового сосуда с зеркальными стенками, способными стопроцентно отражать излучение, которое возникает внутри полости. Представим, что ограничивающий полость шар имеет очень малое отверстие. Тогда часть энергии (хотя и очень малая, так что ее в общем балансе можно не учитывать) в виде слабого светового потока выходит из полости и ее можно зарегистрировать.

Опыт показывает, что если газ в полости находится при очень низкой температуре, то никакого выходящего из отверстия излучения практически не будет. Если газ внутри полости нагревать, то такое излучение начнет наблюдаться вначале в далекой инфракрасной области. Интенсивность излучения будет возрастать с ростом температуры, причем максимум энергии излучения как функции частоты смещается в сторону более коротких длин электромагнитных волн или больших частот (в направлении к видимому диапазону). В конце концов появится характерное для раскаленных тел красное свечение.

Что же происходит с частицами газа в полости и почему возникает излучение? Возрастание температуры газа означает возрастание приходящейся на одну частицу средней кинетической энергии поступательного движения, причем  $T_{кин} = \frac{3}{2} kT$ , где

$T$ —термодинамическая температура газа.

Согласно первому постулату Бора, любая молекула или атом может стационарно, т. е. достаточно длительное по меркам микромира время, находиться лишь в определенных энергетических состояниях (на уровнях энергии  $E_k$ ). В состоянии с минимальным значением энергии (на основном уровне) частица может существовать неограниченно долго.

Для того чтобы перейти из одного стационарного состояния в другое, макрочастица должна принять или отдать энергию,

равную разности между начальным и конечным значениями уровней энергии (например,  $E_n - E_m$ ). Если в результате внешнего воздействия частица с уровня с меньшей энергией перешла в состояние с большей энергией и далее предоставлена самой себе, то «удержаться» на уровне с большей энергией (на возбужденном уровне или в состоянии возбуждения) длительное время (как правило, более 1мкс) она не может и перейдет в одно из состояний с меньшей энергией, вплоть до основного. При этом согласно второму постулату Бора излучается квант электромагнитной энергии (световая волна) с частотой

$$\omega_{n,m} = (E_n - E_m)/\hbar.$$

Если газ первоначально находился при температуре, близкой к 0К, то можно считать, что все его частицы имели внутреннюю энергию, соответствующую основному уровню  $E_0$ . Нагреем газ, но так, чтобы выполнялось неравенство  $E_1 - E_0 \gg \frac{3}{2}kT$ , где  $E_1$  — уровень энергии, ближайший к основному. Частицы при хаотическом движении сталкиваются друг с другом. Если газ по свойствам близок к идеальному, то будет происходить лишь процесс обмена кинетической энергией в результате упругих взаимодействий.

Увеличим температуру газа. Может оказаться, что средняя кинетическая энергия частицы начнет удовлетворять условию

$$\frac{3}{2}kT \approx E_1 - E_0,$$

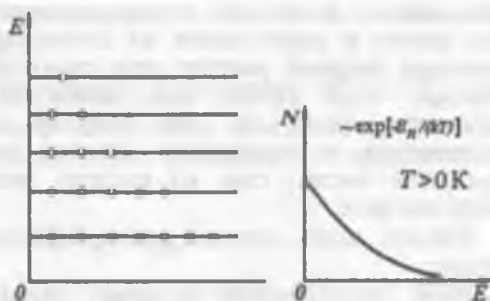
если принять во внимание лишь поступательные степени свободы. В результате столкновения между частицами их кинетическая энергия может израсходоваться на возбуждение частиц, т. е. на переход их в новое, более высокое стационарное энергетическое состояние. Столкновение будет уже неупругим. Большое число таких столкновений приведет к тому, что в любой момент времени наряду с атомами или молекулами газа, находящимися в основном состоянии, будет существовать и некоторое число частиц с энергией  $E_1$ .

Если еще увеличить температуру, то в результате аналогичных процессов станет возможным и переход части атомов или молекул в следующее, ближайшее к  $E_1$ , возбужденное энергетическое состояние  $E_2$ .

Вообще, с ростом температуры будут появляться среди частиц газа такие, которые обладают все большей энергией возбуждения:  $E_3 < E_4 < E_5$  и т. д. (рис. 9.26).

Если газ достаточно разрежен, так что время между соударениями частиц больше того времени, которое частица может находиться в данном возбужденном состоянии, то, получив порцию энергии при одном акте столкновения, она успеет

Рис. 9.26. Распределение атомов по уровням энергии



выделить эту энергию в виде излучения и перейти в одно из нижних энергетических состояний до следующего столкновения. Такое выделение энергии возможно только в форме электромагнитного излучения, которое и начнет накапливаться в полости, приводя к свечению газа. Возникающее в результате таких самопроизвольных процессов излучение называется *спонтанным*, а соответствующие переходы между уровнями энергии *спонтанными переходами*.

Таким образом, причина появления светимости нагретого газа становится понятной. Очевидно также, что общая энергия, которая в любой момент времени выделяется в полость с газом в результате спонтанного излучения, прямо пропорциональна числу частиц, которые к этому спонтанному излучению способны, т. е. числу частиц  $N_n$  с энергией, например,  $E_n$  (числу частиц, находящихся на уровне  $E_n$ ). Это число можно определить, пользуясь формулой Больцмана (см. § 5.2). Это означает, что  $N_n \sim \exp[-E_n/(kT)]$ .

Если в какой-то момент времени на уровне энергии  $E_n$  находится  $N_n$  частиц, то не все они одновременно в этот момент времени «высвечиваются». Этот процесс вероятностный, растянутый по времени, и поэтому в данный момент спонтанные переходы с излучением наблюдаются не у  $N_n$  частиц, а у меньшего их числа. Характеризуется это уменьшение коэффициентом  $A_n^m$ , получившим название *коэффициента Эйнштейна* (именно он ввел этот и другие рассматриваемые ниже коэффициенты) для спонтанных переходов между данной парой уровней. Этот коэффициент зависит лишь от свойств данного атома или молекулы и не зависит, например, от температуры. Энергия электромагнитного излучения, выделяемая в единицу времени в результате таких спонтанных переходов с начального уровня  $E_n$  на конечный  $E_m$ , пропорциональна  $A_n^m N_n$ .

Итак, мы пришли к выводу, подтверждаемому и экспериментом: в полости с хорошо отражающими стенками при нагревании

газа должно появиться электромагнитное излучение всевозможных частот в зависимости от сочетаний значений начальных и конечных уровней энергии при спонтанном излучении. С другой стороны, если атомы или молекулы находятся в электромагнитном поле, то, по второму постулату Бора должно наблюдаться и поглощение квантов электромагнитного поля с переходом частиц газа из низших энергетических состояний в более высокие.

Значит, часть энергии электромагнитного поля уйдет на этот процесс.

Доля поглощенной энергии, связанной с переходами из состояния  $E_m$  в состояние  $E_n$ , зависит от числа частиц на исходном уровне  $N_m$ , от вероятности перехода для поглощения (мы связываем эту вероятность снова с коэффициентом Эйнштейна, но уже для поглощения —  $B_m^n$ , причем  $B_m^n \neq A_n^m$ ) и от плотности энергии излучения подходящей частоты  $u(\omega_{nm})$  или от числа квантов с энергией  $-\hbar\omega_{nm} = E_n - E_m$  в единице объема. Наличие этой величины понятно, так как поглощение вообще невозможно, если нет электромагнитных волн с нужными частотами.

Если бы между двумя этими процессами (спонтанного излучения и поглощения) существовал полный баланс, то должно было бы быть

$$A_n^m N_n = B_m^n N_m u(\omega_{nm}),$$

или, так как  $N_n/N_m = \exp[(E_m - E_n)/(kT)]$ ,

$$u(\omega_{nm}) = (A_n^m/B_m^n) \exp[-\hbar\omega_{nm}/(kT)].$$

Значит, если рассматривать переходы между уровнями  $n$  и  $m$ , то плотность (спектральная) излучения при возрастании  $T$  до  $\infty$  (при очень больших температурах) должна была бы стремиться к конечной величине, определяемой только индивидуальными свойствами частиц газа (коэффициент поглощения  $B_m^n$ , как и коэффициент  $A_n^m$ , зависит лишь от свойств атома или молекулы как изолированной системы).

Результат, однако, противоречит эксперименту, который показывает, что с ростом температуры спектральная плотность излучения непрерывно возрастает. Такое возможно лишь в случае, если имеется не принятый нами во внимание дополнителный процесс, приводящий к накоплению энергии излучения в полости. Это привело Эйнштейна к выводу о том, что наряду со спонтанным излучением должно существовать и индуцированное или *вынужденное излучение*. Энергия этого излучения пропорциональна числу молекул на начальном уровне

$N_n$ , коэффициенту  $B_n^m$  для вынужденного излучения и плотности излучения в единице объема.

Существенно, что коэффициенты Эйнштейна  $B_n^n$  и  $B_n^m$  для поглощения и вынужденного излучения одинаковы. В этом случае при постоянной температуре выполняется соотношение (*детальное равновесие*)

$$A_n^m N_n + B_n^m N_n u(\omega_{nm}) = B_n^m N_m u(\omega_{nm}).$$

Отсюда для плотности излучения данной частоты  $u(\omega_{nm})$ , выходящего из отверстия в полости, получим

$$\begin{aligned} u(\omega_{nm}) &= (A_n^m/B_n^m) \{ \exp [(E_n - E_m)/(kT)] - 1 \}^{-1} = \\ &= (A_n^m/B_n^m) \{ \exp [\hbar\omega_{nm}/(kT)] - 1 \}^{-1}. \end{aligned}$$

С ростом температуры теперь уже  $u(\omega_{nm})$  будет также расти, что вполне отвечает эксперименту.

Дальнейшие рассуждения, на которых мы не останавливаемся, приводят к формуле Планка для спектральной плотности излучения в рассматриваемой полости (излучения *черного тела*):

$$u(\omega, T) = (\hbar\omega^3/(\pi^2 c^3)) \{ \exp [\hbar\omega/(kT)] - 1 \}^{-1}.$$

Эта формула прекрасно описывает эксперимент. В свое время она послужила для обоснования принципа квантового поглощения и излучения энергии атомами и молекулами и становления квантовой теории вообще. Для нас важно, что формула Планка приводит к необходимости введения понятия вынужденного излучения. Именно это излучение становится доминирующим при высоких температурах или при инверсной заселенности уровней энергии (см. ниже) и поэтому делает возможным усиление света при прохождении его сквозь инверсно заселенные среды, что и определяет существование лазеров.

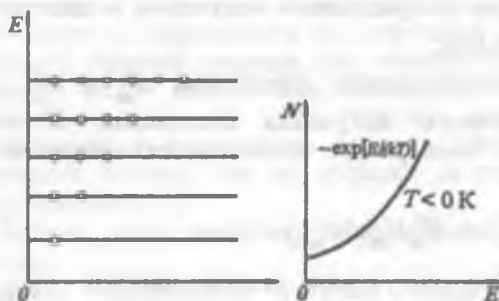
В нагретом газе между всеми атомами и излучением устанавливается энергетическое равновесие и распределение частиц по уровням передается формулой Больцмана.

Чем выше уровень энергии, тем меньше там число частиц. Это распределение в природе обычно и наблюдается. Тогда говорят о нормальной заселенности уровней.

Пусть в результате внешнего взаимодействия большую часть атомов или молекул удалось перевести в возбужденное состояние. Тогда на верхнем уровне будет находиться частиц больше, чем на нижнем. Например, на верхнем уровне находится 60 частиц, на следующем 30 частиц, а на самом нижнем ни одной частицы (рис. 9.27). Такая заселенность уровней называется *инверсной (обратной)*.

Для характеристики такой инверсной среды часто используют понятие отрицательной температуры. В самом деле, в инверсной

Рис. 9.27. Инверсная заселенность уровней совокупности атомов



среде чем выше уровень энергии, тем больше на нем находится частиц по сравнению с основным уровнем. Эту характерную особенность инверсной среды удобно представить, предположив, что температура  $T$  отрицательна. Тогда знак минус в экспоненте исчезает и распределение числа частиц  $N(E)$  будет описываться зависимостью

$$N(E) \sim \exp [E/(kT)].$$

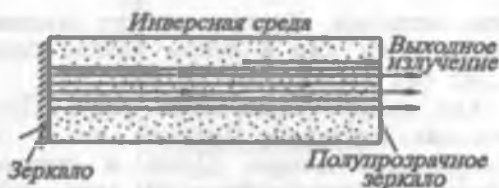
Только для передачи этого свойства и используется понятие отрицательной температуры. Ни в каком другом смысле понятие отрицательной температуры применять нельзя, т. е. нельзя понимать, что это температура какого-то необычного движения ниже температуры 0 К. Отрицательная температура—это всего лишь признак того, что среда инверсная.

Пусть эта среда состоит из частиц, которые возбуждены и находятся в верхнем состоянии. Если на среду никак не воздействовать, то каждый атом или молекула среды спустя какое-то время, излучив электромагнитную энергию, перейдет в основное состояние. Таким образом, инверсная среда должна выделить энергию, но при этом все атомы будут излучать в разное время. Поэтому спонтанно излучающая инверсно заселенная среда, высвечиваясь, создаст растянутый во времени импульс света.

Оказывается, что процессом излучения атомов можно управлять. Если практически одновременно каждый возбужденный атом испытает воздействие электромагнитной волны, то переходы на основной уровень могут произойти синхронно за счет вынужденного излучения. В результате вся инверсная среда мгновенно излучит электромагнитную энергию и вместо растянутого процесса получится острый пик. Таким образом, вся запасенная энергия выделится в очень малый промежуток времени. Мощность, как известно, равна энергии, деленной на время. Поэтому чем за меньшее время произойдет излучение, тем большая мощность выделится в импульсе. Именно этот принцип и используется в лазерах.

Типичный лазер содержит резонатор, состоящий из двух зеркал, причем одно из них полупрозрачное (рис. 9.28). Внутри

рис. 9.28. Принципиальная схема лазера



этой системы создают инверсную среду, когда большинство атомов находится в возбужденном состоянии.

Пусть один из этих возбужденных атомов спонтанно переходит на нижний уровень, излучив при этом квант энергии. Спонтанный переход—это переход вероятностный и поэтому может произойти в самом начале существования инверсной среды. Этот квант энергии распространяется внутри системы из двух зеркал, между которыми инверсная среда. При своем движении квант энергии может встретить другой атом, «столкнуть» его вниз (вынужденное излучение!) и тогда уже будет два кванта. Эти кванты могут встретить другие атомы и тоже «столкнуть» их с верхнего состояния. Образуются дополнительные кванты и т. д. Процесс развивается лавинообразно. Возникающая электромагнитная волна, распространяясь, достигает полупрозрачного зеркала. В результате появления вторичных волн в зеркале часть первоначальной энергии выйдет наружу, а часть отразится обратно и соответствующие кванты начнут опять «сталкивать» атомы с верхнего состояния. Это будет продолжаться до тех пор, пока отраженная электромагнитная волна не достигнет другого зеркала и т. д. Расстояние между зеркалами небольшое, а скорость света очень велика (300 000 км/с). Поэтому весь процесс проходит за очень малый промежуток времени и практически за это время все атомы инверсной среды будут переведены в нижнее состояние. При этом происходит короткая и мощная вспышка. Следует еще подчеркнуть одну особенность, которая способствует образованию такого лавинообразного процесса в инверсной среде. Это то, что вынужденное излучение всегда повторяет направление первоначального кванта.

Таким образом, чтобы работал импульсный лазер, необходимо создать инверсную среду и подходящий резонатор.

Имеются и такие лазерные устройства, в которых создается инверсная среда, затем она высвечивается, опять создается инверсная среда и опять высвечивается и т. д., т. е. получается во времени большое число импульсов или непрерывный процесс. Это лазеры *непрерывного действия*. Схема их устройства примерно такая же.

Лазерный луч обладает одним важным свойством. Он получается практически монохроматическим, так как у всех атомов совершаются одинаковые переходы и излучается электромагнитная волна только одной длины. Поэтому лазерный луч можно хорошо сфокусировать и сконцентрировать большую энергию в чрезвычайно малых областях пространства. Так осуществляется



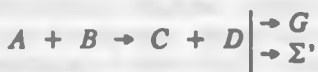
резка металлов. Кстати, следует отметить, что лазерный пучок фокусируется уже в самом начале своего образования за счет зеркал без специальной оптики.

Как создать инверсную среду? Приемов много (например, *световая накачка*). Поясним это с помощью схемы работы *трехуровневого лазера*. Пусть в системе имеется три уровня энергии. Облучим систему. За счет этой энергии частицы среды перейдут в верхнее возбужденное состояние (рис. 9.29).

Система подбирается так, чтобы вероятность обратного спонтанного перехода на нижние уровни была различной, причем переход на самый нижний уровень был менее вероятен, чем переход на промежуточный уровень. Поэтому на промежуточном уровне будут накапливаться частицы.

Время жизни частиц в промежуточном (*метастабильном*) состоянии будет очень большим, но только по сравнению с длительностью всех других процессов, происходящих в системе. За это время на метастабильном уровне можно накопить достаточное число частиц.

Процесс накачки энергии можно осуществить и за счет химической реакции. Пусть, например, происходит химическая реакция, при которой реагируют две молекулы  $A$  и  $B$ , и это приводит к образованию других молекул (частиц)  $C$  и  $D$ . Совершенно необязательно, чтобы частицы  $C$  и  $D$  находились в основных состояниях. Они могут находиться и в возбужденных состояниях, из которых эти частицы стремятся перейти в более выгодные энергетические состояния. Этот переход может происходить по-разному. Например, частица  $D$  может распасться на частицы  $\Sigma$  и  $G$ :



а может, находясь в возбужденном состоянии  $D^*$ , перейти в основное, испустив квант энергии:



Если удастся осуществить такую химическую реакцию, в результате которой появляется частица в возбужденном состоянии, переходящая затем в основное состояние и излучающая квант электромагнитной энергии, то можно создать инверсно заселенную среду и на этой основе построить химический лазер.

Существуют и другие способы создания инверсной среды. Для этого можно, например, применить поток электронов (электронное возбуждение).

Вначале у лазеров нельзя было перестраивать частоту излучения. В дальнейшем было обнаружено, что если пропустить луч лазера сквозь другие специальные среды, то эти среды способны трансформировать первоначальный луч лазера и менять



Рис. 9.29. Схема работы трехуровневого лазера:  
1—накачка, 2—переход на метастабильный уровень, 3—излучение

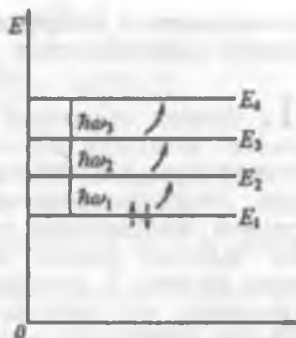


Рис. 9.30. Набор энергетических уровней атома и схема последовательных переходов

его длину волны. На этой основе были созданы *лазеры с перестраиваемой частотой*.

С помощью облучения находящихся в газовом состоянии атомов или молекул одновременно светом от нескольких лазеров можно осуществить очень эффективные способы качественного и количественного спектрального анализа. Широкое распространение получил следующий прием.

Чтобы понять его, представим, что газ состоит, например, из атомов Na, для которых имеется набор энергетических уровней (рис. 9.30). Каждому атому присуща своя совокупность уровней энергии, и нет таких атомов, для которых эта совокупность повторялась бы.

Пусть вначале атомы находились на нижнем уровне  $E_1$ . Облучим газ лучом лазера нужной длины волны, чтобы, поглотив квант с энергией  $\hbar\omega_1$ , какой-нибудь атом перешел в возбужденное состояние с энергией  $E_2$ . Если поток квантов мал, то, пробыв некоторое время в возбужденном состоянии, атом вернется в исходное состояние. Но если плотность потока квантов велика, то, поглотив еще один квант с энергией  $\hbar\omega_2$ , атом перейдет в следующее возбужденное состояние (уровень с энергией  $E_3$ ), а поглотив квант энергии  $\hbar\omega_3$ , — в следующее возбужденное состояние и т. д. Подходящее сочетание энергий  $\hbar\omega_1, \hbar\omega_2, \hbar\omega_3$  квантов характерно только для данного атома, поэтому процесс многократного возбуждения обладает чрезвычайной селективностью.

С помощью такого многократного процесса можно удалить электрон так далеко от атома, что он будет находиться почти в свободном состоянии. Затем накладывается электрическое поле и электрон отрывается от атома. Остается ион, который уже легко зарегистрировать. Построенные на этом принципе атомно-

ионизационные методы спектрального анализа обладают чрезвычайно большой селективностью и чувствительностью.

## 9.11. Зонная структура кристаллических тел

Все твердые тела делятся на две категории: *кристаллические* и *аморфные*. Аморфные тела иногда называют *стеклами*.

Характерной особенностью кристаллических тел является то, что в них имеется вполне определенный порядок взаимного расположения атомов. В аморфных телах такой порядок отсутствует. Оказывается, что свойство кристалла иметь вполне определенный порядок расположения атомов приводит к очень интересным и чрезвычайно важным особенностям в расположении уровней энергии, отвечающих состояниям электронов в этих телах. Перейдем к изучению этих закономерностей.

Чтобы получить наглядный результат, попробуем выяснить основные особенности в структуре электронных уровней энергии в цепочке равноотстоящих друг от друга атомов водорода. Предположим, что расстояния между соседними атомами водорода приблизительно такие же, как и в молекуле водорода.

Если электроны в каждом атоме водорода до образования молекулы находились в состоянии  $1s$  с одинаковой энергией, т. е. описывались функциями  $\psi_{1s}^{(1)}$  и  $\psi_{1s}^{(2)}$ , то после объединения атомов в молекулу ее волновая функция

$$\Psi_{1,2} = (\psi_{1s}^{(1)} \pm \psi_{1s}^{(2)})/\sqrt{2},$$

т. е. возможны два энергетических состояния. Мы ранее сделали этот вывод, опираясь на свойства симметрии электронной плотности и волновой функции. Применим этот прием и сейчас. Для этого представим, что система из четырех атомов водорода, расположенных в линию, получилась в результате сближения друг с другом двух молекул водорода. Будем считать, что эти молекулы первоначально находились в состояниях с волновой функцией  $\Psi_1$ , которую теперь для ясности обозначим  $\Psi_1(\text{H}_2)$  ( $\text{H}_2$  показывает, что функция относится к молекуле водорода).

Тогда, как и для молекулы водорода, приближенную волновую функцию объединенной системы из двух молекул водорода или четырех атомов водорода можно записать в виде

$$\Psi(\text{H}_4) = c_1 \Psi_1^{(1)}(\text{H}_2) + c_2 \Psi_1^{(2)}(\text{H}_2).$$

Верхние индексы относятся к первой и второй молекулам водорода. Так как система из двух молекул водорода имеет центр симметрии, то, пользуясь свойствами симметрии волновых функций и делая дополнительные предположения о малости интеграла  $\int \Psi_1^{(1)}(\text{H}_2) \Psi_1^{(2)}(\text{H}_2) dV$ , получаем

$$\Psi_{1,2}(\text{H}_4) = [\Psi_1^{(1)}(\text{H}_2) \pm \Psi_1^{(2)}(\text{H}_2)]/\sqrt{2}.$$

учитывая вид функций  $\Psi_1^{(1)}(H_2)$  и  $\Psi_1^{(2)}(H_2)$ , находим

$$\Psi_{1,2}(H_4) = [\psi_{1s}^{(1)} + \psi_{1s}^{(2)} \pm (\psi_{1s}^{(3)} + \psi_{1s}^{(4)})]/2.$$

Таким образом, снова получаем две функции: симметричную относительно центра симметрии и антисимметричную. В самом деле, если при знаке минус при втором члене в квадратных скобках поменять местами функции  $\psi_{1s}^{(1)}$  и  $\psi_{1s}^{(4)}$  и одновременно  $\psi_{1s}^{(2)}$  и  $\psi_{1s}^{(3)}$  (эти функции отвечают атомам 1 и 4 и 2 и 3, симметричным относительно центра симметрии), то изменится знак и функции  $\Psi_2(H_4)$ .

Функции  $\Psi_{1,2}(H_4)$  мы получили для двух объединенных молекул водорода, находящихся в основном состоянии, т. е. в состоянии со связывающей орбиталью.

Другие функции для двух молекул  $H_2$ , соединенных между собой и находящихся в возбужденном состоянии, равны

$$\begin{aligned}\Psi_{3,4}(H_4) &= [\Psi_2^{(1)}(H_2) \pm \Psi_2^{(2)}(H_2)]/\sqrt{2} = \\ &= [\psi_{1s}^{(1)} - \psi_{1s}^{(2)} \pm (\psi_{1s}^{(3)} - \psi_{1s}^{(4)})]/2\end{aligned}$$

При знаке плюс перед круглой скобкой получается антисимметричная функция, меняющая знак при перестановке симметричных атомов, а при знаке минус—симметричная.

Интересно отметить, что подобным образом нельзя комбинировать функцию основного состояния одной молекулы  $H_2$  с функцией возбужденного состояния другой, так как получающаяся комбинация не обладает правильными свойствами симметрии.

Итак, первоначальное  $1s$ -состояние атома водорода при образовании молекулы водорода, т. е. при соединении двух атомов, приводит к появлению двух волновых функций и двух уровней энергии, а при образовании цепочки из четырех атомов—к появлению четырех волновых функций и четырех уровней энергии.

Следующая цепочка из восьми атомов водорода получается соединением двух образований из четырех атомов водорода каждая. Возникает уже группа из восьми симметричных и антисимметричных волновых функций и восьми соответствующих им уровней энергии.

Вообще цепочка из  $N$  расположенных на равных расстояниях друг от друга атомов имеет  $N$  разных волновых функций и соответственно  $N$  разных уровней энергии.

Самый низкий уровень отвечает полностью связывающей орбитали

$$\Psi_{cs} = \sum_{k=1}^N \psi_{1s}^{(k)} / \sqrt{N}.$$

Соответствующее этой волновой функции приближенное значение энергии можно найти, если воспользоваться соотношением

$$E_{cs} = \int \Psi \hat{H} \Psi dV.$$

Тогда

$$E_{cs} = \frac{1}{N} \left( \sum_{k=1}^N \int \psi_{1s}^{(k)} \hat{H} \psi_{1s}^{(k)} dV + 2 \sum_{k=1}^{N-1} \int \psi_{1s}^{(k)} \hat{H} \psi_{1s}^{(k+1)} dV \right).$$

При выводе этой формулы учтено, что функции  $\psi_{1s}^{(k)}$  отличаются от нуля только в окрестности своего атома, и поэтому произведение  $\psi_{1s}^{(k)} \psi_{1s}^{(k+2)}$  равно нулю. В очень длинных цепочках можно считать, что все атомы Н находятся в одинаковых положениях. Тогда интеграл  $\int \psi_{1s}^{(k)} \hat{H} \psi_{1s}^{(k)} dV$  не зависит от номера атома  $k$  и является некоторой атомной константой, приблизительно равной энергии электрона в атоме Н в состоянии  $1s$ . Обозначим эту энергию  $E_{1s}$ .

Величины интегралов  $\int \psi_{1s}^{(k)} \hat{H} \psi_{1s}^{(k+1)} dV$  также не зависят от номера атома. Обозначим интеграл  $(-\Delta E)/2$  и назовем эту величину энергией стабилизации или резонанса (энергия стабилизации отрицательна). Тогда для больших  $N$

$$E_{cs} = \frac{1}{N} [NE_{1s} - (N-1)\Delta E] / N \approx E_{1s} - \Delta E.$$

Это означает, что, начиная с некоторого  $N$ , уровень энергии перестанет меняться.

Наименее выгодным состоянием будет состояние с полностью разрыхляющей орбиталью, т. е. такой, при которой при переходе от одной атомной функции  $\psi_{1s}^{(k)}$  к соседней  $\psi_{1s}^{(k+1)}$  знаки меняются. В общем виде

$$\Psi_{\text{разр}} = \left[ \sum_{k=1}^N (-1)^{(k-1)} \psi_{1s}^{(k)} \right] / N.$$

Тогда

$$E_{\text{разр}} = \frac{1}{N} \left[ \sum_{k=1}^N \int \psi_{1s}^{(k)} \hat{H} \psi_{1s}^{(k)} dV - 2 \sum_{k=1}^{N-1} \int \psi_{1s}^{(k)} \hat{H} \psi_{1s}^{(k+1)} dV \right],$$

и для больших  $N$

$$E_{\text{разр}} \approx E_{1s} + \Delta E.$$

Таким образом, все допустимые в нашей задаче значения энергии для электронов, находящихся в поле большого числа расположенных в линию протонов, укладываются в интервал шириной  $2\Delta E$ , причем ширина интервала не зависит от длины цепи.

Чем больше длина цепи, тем больше уровней энергии попадает в интервал. Они все более и более плотно его заполняют. Расстояние между уровнями энергии уменьшается в центре интервала как  $1/N$  и по краям интервала как  $1/N^2$ . Постепенно с ростом длины цепи  $N$  уровни столь плотно заполняют интервал, что экспериментально не удастся обнаружить дискретную структуру. Получается квазиконтиниум или энергетическая зона.

Индивидуальность соответствующего атома и свойства ближайшего окружения при этом в известной степени сохраняются. Они проявляются в местоположении  $E_1$ , центра зоны и ширине ее  $\Delta E$ , достигающей насыщения уже при малых длинах цепи. Это очень важный вывод. Он показывает, что появление зоны связано с увеличением числа частиц в системе, но не с радикальными изменениями свойств в каждой отдельной частице.

Мы рассмотрели только  $1s$ -состояние протонов. Если цепочка состояла бы из реальных атомов, в которых имеется несколько электронов в нескольких возможных состояниях, то было бы несколько зон. Число зон в кристалле или зон в цепи, состоящих из одинаковых атомов, равно числу уровней энергии, которые допустимы для каждого из этих атомов, изолированных в пространстве. Каждый уровень энергии атома как бы расщепился на  $N$  ( $N$ —число атомов в системе) подуровней, образовав энергетические зоны. Ширина каждой из таких зон определяется только ближним взаимодействием с окружающими данный атом другими атомами.

Увеличение размеров кристалла, таким образом, сказывается не в увеличении ширины той или иной зоны, а в увеличении плотности состояний внутри каждой зоны, которая растет пропорционально числу атомов в полной системе.

Таким образом, если в индивидуальных атомах имеются отдельные дискретные уровни энергии, то в кристаллах—отдельные дискретные зоны.

В этом заключается аналогия между атомом и кристаллом, состоящим из этих атомов. Но появляются и некоторые принципиальные различия, связанные с наличием зон.

Обсудим эти моменты. Когда число подуровней в зоне велико и расстоянием между ними практически можно пренебречь, то при переходе с одного подуровня на другой подуровень электрону достаточно приобрести или потерять бесконечно малую порцию энергии.

Любое сравнительно небольшое внешнее воздействие, способное передать в систему энергию, пусть даже небольшую, например наложение внешнего электрического поля, приведет к

тому, что электрон легко изменит свое состояние, т. е. начнет двигаться. Изменение состояния электрона в пределах одной энергетической зоны становится уже не квантовым, а классическим. Не следует путать реальное движение электрона внутри кристалла с «движением» электрона внутри энергетической зоны. Когда электрон движется «внутри энергетической зоны», изменяются его энергетические состояния, но это не значит, что электрон пролетает от одного края кристалла до другого..

Обратимся теперь к собственным функциям. Во всех состояниях волновая функция охватывает всю цепочку атомов, что делает возможным перемещение электрона по всей цепи.

Легко увидеть аналогию между поведением волновой функции для рассматриваемой задачи и видом волновых функций для задачи о движении частицы в одномерной потенциальной яме. Эта аналогия и позволяет для определения уровней энергии частицы, находящейся в поле периодических центров в определенном, достаточно хорошем приближении, пользоваться металлической моделью или моделью свободного движения частицы в потенциальном ящике достаточно большой длины.

Эта аналогия позволяет ввести понятие *электронного газа*, т. е. свободно движущейся в металле совокупности отрицательно заряженных частиц. Такая модель использовалась при описании многих явлений в разделах «Электричество» и «Магнетизм». Теперь становится понятным, почему модель хорошо «работает» в классической физике, несмотря на то, что электроны в металлах на самом деле подчиняются квантовым законам.

Перейдем к следующему вопросу. В согласии с принципом Паули электроны должны располагаться парами на всех возможных уровнях энергии. Число уровней энергии во всей системе, составляющих одну зону, равно  $N$ . Значит, число занятых уровней  $N/2$ , т. е. зона занята электронами до половины.

При большой длине цепи расстояние между верхним занятым уровнем первой половины зоны и нижним свободным уровнем второй половины зоны будет бесконечно малым. Поэтому электрон с верхнего занятого уровня легко может перейти в энергетические состояния, отвечающие не занятой электронами части зоны. Другими словами, при внешних воздействиях электроны легко переходят в состояние свободного движения. Именно это свойственно проводникам. К числу классических проводников, как известно, относятся металлы. В частности, хорошим проводником мог бы быть кристалл, составленный из атомов водорода, или твердый водород.

Если имеется много атомов или уровней энергии в атомах и соответственно много электронов, то получается система зон.

Пусть имеется дискретный набор неперекрывающихся зон. Это может быть тогда, когда первоначальные уровни энергии атомов, которые и порождают зоны, расположены достаточно далеко друг от друга и коэффициенты, описывающие взаимо-



Рис. 9.31. Неперекрывающиеся зоны

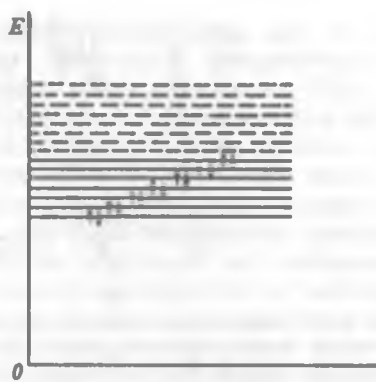


Рис. 9.32. Структура энергетических зон металла

действие атомов друг с другом, сравнительно малы. Такая картина показана на рис. 9.31.

Если эти зоны заполняются по всем подуровням электронами (на каждом подуровне по два электрона), то может оказаться, что часть зон будет заполнена полностью, а другая часть зон окажется свободной (рис. 9.31).

Чтобы перевести электроны из занятой зоны в свободную, необходимо затратить достаточную энергию. Такую энергию не может сообщить внешнее электрическое поле. В результате получается система, обладающая изоляционными свойствами. Это диэлектрик, или изолятор.

Если незаполненная зона близко приближается к верхней заполненной зоне, то система проявляет свойства полупроводников. Наконец, если заполненная зона соприкасается с незаполненной или они перекрываются так, как показано на рис. 9.32, то система проявляет свойства металлов. Разумеется, свойства металлов будет проявлять и система, в которой имеется частично незаполненная зона.

Чем меньше степень заполнения этой зоны и чем больше плотность состояний в ней, тем лучше свойства проводимости этой системы.

Таким образом, нам удалось объяснить различие в свойствах проводников, полупроводников и изоляторов.

Заметим, что верхний, заполненный электронами, уровень получил название *уровня Ферми*. Чем ближе уровень Ферми к ближайшей незаполненной зоне, тем лучше данный кристалл проводит электрический ток.

Для полупроводников характерно следующее интересное свойство, связанное с проводимостью. Если электрон из занятой (валентной) зоны переходит в состояние незанятой зоны, в возбужденное состояние, то в том месте, где он раньше находился, оказывается свободное место (дырка). В присутствии



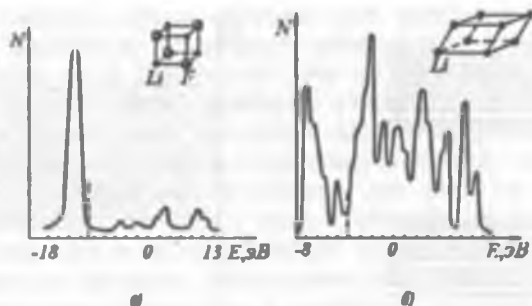


Рис. 9.33. Элементарная ячейка и рассчитанная плотность состояний для фторида лития (а) и металлического лития (б).

Положение уровня Ферми показано пунктиром

внешнего электрического поля ближайшие к дырке электроны валентной зоны могут попасть в эту дырку, оставляя следующее свободное место. Получается, что дырка начинает перемещаться, причем она ведет себя как некоторый положительный заряд. В этом случае возникает *дырочная проводимость*. Такая проводимость объясняет многие важные свойства полупроводников.

Структура энергетических зон проводников, изоляторов и металлов может быть наглядно представлена с помощью *гистограмм плотности состояний*. Для этого выбирается некоторый интервал энергии, например, от  $-E$  до  $+E$ , который разбивается на  $k$  равных частей (классов). На каждый класс приходится порция энергии  $2E/k$ . Одноэлектронные уровни энергии рассортировываются по классам. В результате становится известным, сколько одноэлектронных состояний пришлось на каждый элементарный интервал энергии. Максимальная плотность состояний принимается за единицу и выстраивается относительная плотность состояний во всем интервале энергий. Полученное распределение состояний «заселяется» электронами в соответствии с их общим количеством в рассматриваемой системе. Граница заселения показывает положение уровня Ферми. Ниже уровня Ферми располагается *валентная зона*, выше—*зона проводимости*. На рис. 9.33 представлены гистограммы плотности состояний для фторида лития и металлического лития. Из гистограммы фторида лития видно, что это соединение должно проявлять хорошие изоляционные свойства: между валентной зоной и зоной проводимости имеется *запрещенная зона*, где плотность состояний равна нулю, а плотность свободных уровней в зоне проводимости очень низка. Металлический литий, как и следовало ожидать, обнаруживает типичные проводящие свойства. Из гистограммы лития видно, что валентная зона и зона проводимости перекрываются. Кроме того, плотность свободных состояний в зоне проводимости очень высока. Оба этих фактора обеспечивают практически свободное перемещение электронов из валентной зоны в зону проводимости и высокую электропроводность системы.

Приведенная выше общая схема появления зон и их заполнения электронами требует некоторых дополнительных пояснений, особенно по отношению к полупроводникам. Простое правило заполнения всех уровней парами электронов верно лишь для температуры абсолютного нуля, когда исчезают все тепловые движения частиц. При температурах  $T > 0$  электроны приобретают дополнительную энергию. Если мы имеем дело с характерным диэлектриком или металлом, то при повышении температуры существенно новых физических эффектов не возникает. Однако принципиально иная ситуация наблюдается в полупроводниках. В этом случае энергетическая щель между валентной зоной и зоной проводимости (запрещенная зона) может быть не очень велика. Дополнительной тепловой энергии может оказаться достаточно, чтобы перевести часть электронов из валентной зоны в зону проводимости. В результате в зоне проводимости появляются свободные электроны, а в валентной зоне образуется их недостаток и появляются незанятые или частично (одним электроном) занятые подуровни. Такая картина будет существовать и в отсутствии внешнего электрического поля. Если это поле наложить, то появится дополнительное возмущение, которое и приведет к перераспределению электронов по подуровням как валентной зоны, так и зоны проводимости. Возмущение это, воспринимаемое как направленное движение зарядов, в зоне проводимости и в валентной зоне не будет одинаковым хотя бы потому, что плотности состояний в этих зонах различаются. Кроме того, электроны зоны проводимости обладают большей энергией по сравнению с электронами валентной зоны и поэтому меньше связаны с атомным остовом и более подвижны. Поэтому они более подходят к модели электронного газа. Все это и позволяет говорить о двух типах проводимости: электронной (проводимость  $n$ -типа, от слова negative) и дырочной (проводимость  $p$ -типа, от слова positive). При этом перемещение дырки не есть движение какого-то точечного положительного заряда. Эта проводимость отражает коллективное движение всей массы электронов (перераспределение по подуровням) валентной зоны в присутствии внешнего электрического поля.

За счет хаотического теплового движения и столкновений электроны не только могут приобрести дополнительную энергию и «перескочить» в зону проводимости, но и потерять ее и вернуться в валентную зону, заняв одно из валентных мест на соответствующем подуровне. В таком случае говорят о рекомбинации электрона и дырки (взаимном их уничтожении).

В § 7.3 при обсуждении тока проводимости было показано, что коэффициент проводимости связан с числом способных к перемещению зарядов. Из сказанного выше следует, что это число в полупроводниках самым существенным образом зависит от температуры. Следовательно, и проводимость полупроводников должна сильно зависеть от температуры, причем расти с увеличением  $T$ . Этим полупроводники принципиально отличаются

ся от металлов, проводимость которых с ростом температуры падает, хотя и не сильно.

Для многих полупроводников характерно еще одно специфическое явление, получившее название *внутреннего фотоэффекта*. Явление это заключается в появлении проводимости или сильном возрастании ее при облучении полупроводника светом. При неширокой запрещенной зоне за счет поглощения световой энергии электроны могут перейти из валентной зоны в зону проводимости, в результате чего плотность свободных отрицательных зарядов и дырок возрастет и проводимость увеличится: при приложенном внешнем электрическом поле появится скачок электрического тока.

Все сказанное выше относилось к чистым материалам, состоящим из атомов одного сорта. Если среди атомов одного сорта рассеяны атомы другого (примеси) или если имеются включения одного полупроводника в другой, то возникает ряд дополнительных эффектов, которые и могут быть использованы для создания полупроводниковых выпрямительных и усилительных элементов, составляющих основу современной электроники и процессоров ЭВМ. Мы не будем касаться этих вопросов, ограничившись лишь указанием на то, каким образом атомы примеси могут «выделить» тот или иной тип проводимости в полупроводнике.

Это становится возможным, если уровни энергии примесных атомов попадают внутрь запрещенной зоны чистого полупроводника.

Пусть, далее, эти уровни не являются полностью заполненными, т. е. способны либо принять электроны, либо отдать их. В первом случае создается возможность передачи в результате теплового движения части электронов из валентной зоны на атомы примеси. Это приводит к появлению дырок. Возникает полупроводник *p*-типа. Во втором случае электроны с уровней примесных атомов могут перейти в зоны проводимости. Возникнет полупроводник *n*-типа.

В месте контакта полупроводников этих двух типов возникает так называемый *n-p*-переход, обладающий свойствами преимущественно пропускать ток лишь в одном направлении.

## 9.12. Внутриядерные процессы. Радиоактивность

Как известно из школьного курса физики, атомное ядро имеет размеры порядка  $10^{-13}$  см и состоит из частиц двух сортов—протонов и нейтронов. Обе эти частицы называют также общим термином—*нуклоны*. Протоны несут единичный положительный заряд, нейтроны нейтральны. Массы этих частиц отличаются друг от друга очень мало, но нейтроны более массивны. Как протоны, так и нейтроны достаточно долгоживущи, что и определяет общую стабильность ядер. Однако нейтроны менее стабильны: в свободном состоянии их время

жизни порядка  $10^3$  с, в то время как протон практически стабилен.

Как протон, так и нейтрон имеют спин, равный (в единицах  $\hbar$ )  $\pm 1/2$ . В этом смысле эти частицы подобны электронам. Между нуклонами существуют силы притяжения или ядерные (их принято называть еще сильными) *взаимодействия*. Это самые мощные из взаимодействий.

Благодаря таким взаимодействиям и существуют ядра, т. е. такие малые области пространства, где локализованы нуклоны. Кулоновских сил для этого было бы недостаточно.

Состояния нуклонов в ядрах подчиняются квантовым законам. Ядрам свойственны определенные уровни энергии. Так как спин нуклонов подобен электронному, то их распределение по уровням энергии подчиняется принципу Паули.

Подобно атомам, ядра могут находиться в основном и в возбужденных энергетических состояниях. Также подобно атомам, переходы из возбужденного в основное состояние могут сопровождаться излучением квантов электромагнитной энергии— $\gamma$ -квантов с очень малыми длинами волн. В результате наблюдается  $\gamma$ -радиоактивность. Однако в отличие от атома в ядрах могут происходить и более сложные процессы. При определенных условиях нейтрон может превратиться в протон. В результате образуется электрон, который внутри ядра удерживаться не может (кулоновских сил недостаточно, чтобы локализовать электрон в области порядка  $10^{-13}$  см) и вылетит из него. Такой процесс называется  $\beta$ -распадом или  $\beta$ -радиоактивностью.

Возможен и еще один процесс радиоактивности,  $\alpha$ -распад, при котором из ядра атома вылетает ядро атома гелия (система из двух нейтронов и двух протонов).

На этом процессе мы остановимся более подробно, так как здесь проявляется квантовое явление, называемое *туннельным эффектом*. Чтобы понять, в чем заключается этот эффект, обратимся снова к движению частицы в одномерном потенциальном ящике, но предположим, что движение не свободно, а совершается в потенциальном поле, описываемом следующей функцией (рис. 9.34):

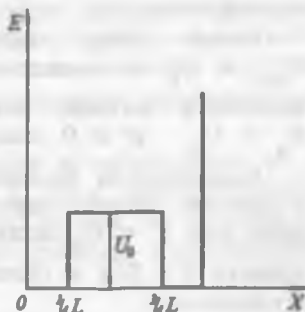


Рис. 9.34. Потенциальный барьер внутри ящика

$$U(x) = \begin{cases} 0 & \text{в интервале } 0 - 1/4, \\ U_0 & \text{в интервале } L/4 - 3L/4, \\ 0 & \text{в интервале } 3L/4 - L. \end{cases}$$

В этом случае имеется потенциальный барьер прямоугольной формы высотой  $U_0$ . Выберем  $U_0 > E_{\min}$  ( $E_{\min}$  — минимальная энергия частицы в ящике). Уровни энергии любой квантовой системы могут быть найдены вариационным методом, если построить

линейную комбинацию базисных функций ( $\psi = \sum_n c_n \varphi_n$ , где  $\varphi_n$  — базисные функции), такую, что найденная энергия  $E$  при вариации коэффициентов  $c_n$  оказывается наименьшей. В данном случае в качестве  $\varphi_n$  целесообразно принять  $\varphi_n = N \sin(\pi n x / L)$ . В результате такого решения обнаруживается, что:

1) существуют уровни энергии, лежащие ниже  $U_0$ , если  $U_0 > E_{\min}$ ;

2) в отличие от ситуации в классической физике, где движения частиц в интервалах  $0 - L/4$  и  $3L/4 - L$  при энергиях  $T_{\text{кин}} < U_0$  совершенно независимы и частица не может перейти, например, слева направо (для этого ей надо приобрести энергию  $T_{\text{кин}} > U_0$ ), квантовая частица, обладающая на любом уровне  $E < U_0$  единой волновой функцией  $\psi = \sum_n c_n \sin(\pi n x / L)$ , ведет себя так, что ее состояния справа и слева от барьера уже не независимы.

Более того, если классическая частица при энергии  $T_{\text{кин}} < U_0$  не может находиться в интервале  $L/4 - 3L/4$ , то квантовая частица, для которой и в этой области даже при  $E_n < U_0$   $\psi \neq 0$ , находиться в этом интервале может!

Классическая частица может перейти из левой части ящика в правую, только перескочив через барьер. Квантовая частица даже при  $E < U_0$  может как бы пронизать барьер и перейти из области  $0 - L/4$  в область  $3L/4 - L$ . Такой переход называется *туннельным*.

Туннельные процессы характерны для микромира и проявляются, в частности, в росте скоростей некоторых химических реакций при очень низких температурах, приводя к нарушению закона Аррениуса, согласно которому с уменьшением температуры уменьшается скорость химических реакций.

Но вернемся к рассматриваемому случаю. Вид волновых функций таков, что в любой точке в интервале  $0 - L/4$  и симметричной ей точке в интервале  $3L/4 - L$  значения  $\psi^2$  одинаковы. Это значит, что частица равновероятно может находиться как справа, так и слева от барьера. Для несимметричного барьера может быть, например,

$$\psi_{(0-L/4)}^2 > \psi_{(3L/4-L)}^2$$

Отношение  $\psi_{(0-L/4)}^2 / \psi_{(3L/4-L)}^2$  вероятностей пребывания частицы слева и справа от барьера характеризует коэффициент прозрачности барьера.

Общий вывод заключается в том, что квантовая частица тем с большей вероятностью может пройти через барьер, чем он уже и чем ближе к его вершине находится один из занятых уровней энергии.

Свойство туннельного перехода позволяет объяснить  $\alpha$ -распад. В самом деле, нуклоны в ядре располагаются по уровням парами, подобно электронам. В стабильных ядрах верхние заполненные уровни должны лежать внутри потенциальной ямы. Если расстояние от верхнего заполненного уровня до края ямы невелико (заметим, что ни одно реальное взаимодействие не приводит к потенциальным ямам с бесконечными стенками) и стенки ямы не очень широки, то появится определенная, не равная нулю, вероятность выхода «верхних» по энергиям частиц за пределы ямы, даже если для них  $E < U_0$ .

Именно это и приводит к  $\alpha$ -радиоактивности и объясняет тот факт, что существуют ядра, выделяющие  $\alpha$ -частицы достаточно редко.

Все процессы в микромире имеют вероятностный характер. Это значит, в частности, что нельзя точно узнать, в какой именно момент то или иное ядро испытает  $\alpha$ -,  $\beta$ - или  $\gamma$ -распад. Если число атомов велико, то определенное их количество испытает такой распад в течение заданного интервала времени. Поэтому принято характеризовать стабильность ядер их периодом полураспада, т. е. интервалом времени, в течение которого, например, из  $10^6$  атомов  $0,5 \cdot 10^6$  испытают радиоактивный распад.

Так как число  $dN$  распадающихся (или испытывающих  $\alpha$ -,  $\beta$ -,  $\gamma$ -распад) в данный интервал времени  $dt$  ядер пропорционально общему числу  $N$  еще не распавшихся атомов и вероятности распада (или обратно пропорционально средней продолжительности жизни  $\tau$ ), то

$$-dN = N \frac{1}{\tau} dt$$

(минус ставится потому, что число нераспавшихся атомов уменьшается). На основании написанного выше получим

$\frac{dN}{N} = -\frac{1}{\tau} dt$  или, интегрируя, с учетом начальных условий имеем

$$N = N_0 \exp(-t/\tau),$$

где  $N_0$  — исходное число атомов. Это соотношение называется законом радиоактивного распада.

При всех внутриядерных процессах выполняются два закона: закон сохранения энергии—массы и закон сохранения заряда. Второй не нуждается в пояснении. Что касается первого, то мы говорим именно о законе сохранения энергии—массы, имея в виду, что при ядерных процессах выделяются не только частицы, но и энергия в виде  $\gamma$ -квантов. Поэтому часть массы исходного продукта согласно формуле  $E = mc^2$  переходит в энергию излучения. В результате появляется *дефект массы*, т. е. отличие суммарной массы исходных продуктов от суммарной массы образующихся.

В заключение напомним, что принято разделять естественную и искусственную радиоактивности. О первой мы уже говорили.

Искусственная радиоактивность появляется тогда, когда в результате внешнего воздействия ( $\gamma$ -облучение, поглощение нейтронов или других частиц) ядра переводятся в возбужденные состояния или образуются новые ядра в состояниях, когда становятся более вероятными туннельные переходы.

Свойство искусственной радиоактивности (меченые атомы) широко используется в различных областях науки и техники, в частности при исследовании процессов миграции вещества в растениях и животных и в медицине для местного облучения больных органов.

#### Вопросы для самоконтроля

1. Какие эксперименты, показывающие особенности поведения микрочастиц по сравнению с классическими частицами, вам известны?
2. Какие противоречия возникают при попытках объяснить строение атома с помощью классической физики?
3. Каково содержание постулатов Бора и какое значение для современной физики они имеют?
4. В чем отличие дифракции электронов на щели от дифракции обычных электромагнитных волн? Каков физический смысл соотношения де Бройля и волновой функции?
5. Какой принципиально новый момент вносит в физику употребление волновой функции? Почему становится возможным пользоваться комплексными функциями?
6. Какой физический смысл вкладывается в описание поведения частиц с помощью волновых пакетов?
7. Какой физический смысл имеет соотношение неопределенностей Гейзенберга?
8. Можно ли привести в обычной жизни какие-либо аналогии принципу неопределенностей, имеющему место для микромира?
9. Каков физический смысл всех членов, входящих в уравнение Шредингера?
10. Что такое оператор, в частности линейный? Как с помощью понятия операторов записать стационарное уравнение Шредингера?
11. Как зависят расстояния между уровнями энергии для частицы в одномерном потенциальном ящике от размеров этого ящика и от массы частицы?
12. Какой должна быть приблизительно минимальная ширина потенциальной ямы и ее глубина, чтобы свободная микрочастица не смогла в ней вообще существовать?
13. При каких условиях наблюдаются дискретные квантовые состояния?
14. Почему электрон в электронно-лучевой трубке, например телевизора, ведет себя как классическая частица и не проявляет отчетливо квантовых свойств, а тот же электрон в атоме ведет себя как квантовая частица с четко выраженными дискретными свойствами?

15. Какова причина появления вырожденных состояний в квантовой системе? Что надо сделать, чтобы снять такие вырождения? Приведите пример вырожденных состояний.

16. Как можно, не прибегая ни к каким вычислениям, сразу же сказать, что электрон в атоме водорода должен обладать вырожденными состояниями? С чем связано появление таких состояний?

17. Как можно, не пользуясь вычислениями, прийти к качественному выводу о том, что расстояния между уровнями энергии по мере возрастания значений соответствующих энергий или удаления электрона от ядра в атоме водорода должны уменьшаться (такой эффект назван сближением уровней энергии к некоторому пределу)?

18. Почему в чисто кулоновском приближении значения уровней энергии в атоме водорода зависят от главного квантового числа и не зависят от орбитального и магнитного квантовых чисел?

19. Что такое эффекты Штарка и Зеемана и с помощью каких экспериментов можно наблюдать эти эффекты?

20. Какова физическая причина того, что каждый из атомов имеет свой индивидуальный спектральный портрет, что и делает возможным проведение спектрального анализа?

21. Что такое гибридизация и какой вид имеют гибридные атомные орбитали в простейшем случае  $sp$ - или  $sp^2$ -гибридизаций?

22. Могли бы существовать определенные пространственные формы, характерные для органических молекул, если бы не выполнялся принцип Паули?

23. Какова физическая природа химической связи? Как можно объяснить из наглядных соображений, почему совокупность положительных тяжелых ядер не распадается, а удерживается при определенных пространственных расположениях электронами?

24. Показать, что при образовании молекулы водорода из двух одинаковых атомов водорода возникает состояние с повышенной электронной плотностью в пространстве между атомами водорода и, следовательно, состояние, которое приводит к устойчивой молекуле водорода.

25. Какие физические соображения лежат в основе рассуждений, приводящих к поиску волновой функции для электронов в поле многих ядер в виде линейной комбинации атомных орбиталей? Какой физический смысл имеет линейная комбинация?

26. Чем отличается с точки зрения распределения электронной плотности ковалентная химическая связь от ионной?

27. Какие изменения в характере химической связи происходят при переходе от сравнительно легких атомов в таблице Менделеева к тяжелым? Как это должно отражаться на формах получающихся молекул?

28. С какими движениями в молекулах связано появление спектров поглощения в ультрафиолетовом и видимом диапазонах, в инфракрасном или далеком инфракрасном диапазонах и миллиметровом диапазоне радиоволн?

29. Какую информацию о строении молекул можно извлечь, наблюдая спектры во всех этих диапазонах?

30. Каковы особенности строения биомолекул и реакций, протекающих в этих молекулах? Почему природа отобрала в качестве биомолекул именно крупные системы, а не системы, состоящие из двух- и трехатомных образований?

31. Чем объясняется цветность органических соединений? Какими характерными особенностями с точки зрения строения химических связей обладают органические красители? Можно ли создать красители на основе органических многоатомных молекул?

32. Какие вы знаете элементы симметрии в молекулах? Как наличие тех или иных элементов симметрии отразится в спектральных свойствах этих молекул?

33. Что такое правило отбора для различных переходов между уровнями энергии в атомах или молекулах под воздействием электромагнитного излучения? Какие вам известны правила отбора?

34. Что такое нормально и инверсно заселенная среда (совокупность атомов или молекул)? Какую роль играют инверсные среды при работе лазеров?

35. Почему с помощью лазеров удается генерировать вспышки света очень большой мощностью? Почему лучи лазеров слабо расходятся в пространстве, что и позволяет производить светолокацию Луны с Земли?



36. Что такое трехуровневый лазер?
37. По какому принципу действуют химические лазеры?
38. Какой принцип лежит в основе атомно-ионизационного многоступенчатого спектрального анализа? Почему в этом случае удастся достигнуть очень высокой степени селективности?
39. Почему при объединении многих атомов в кристалл появляются энергетические зоны? Чем отличается с точки зрения зонной структуры диэлектрик от полупроводника и проводника?
40. Сколько будет подуровней в зоне, если цепочка имеет 100 одинаковых атомов, каждый из которых находится в самом низком  $1s$ -состоянии?
41. Какова физическая причина того, что при объединении атомов в некоторую совокупность—кристалл—ширина зоны стремится к определенной постоянной величине?
42. Как согласовать классические представления о наличии свободных электронов—электронный газ—в металлах с представлениями об уровнях энергии и о зонах, которые вытекают из квантовой механики?
43. При каких условиях структура зон, т. е. расположение подуровней в них, перестанет ощущаться?
44. Что такое уровень Ферми?
45. Почему электроны, возникающие внутри ядер при превращении нуклонов друг в друга, не могут остаться внутри этих ядер и вылетают наружу, приводя к  $\beta$ -распаду?
46. Что такое туннельный эффект и какова его роль в микромире?
47. Почему радиоактивные процессы, в частности скорость радиоактивного распада, не зависят от внешних воздействий на вещество (температуры, давления и т. д.)?



## Г Л А В А 10

### ОТ СУЩЕСТВУЮЩЕГО К ВОЗНИКАЮЩЕМУ\*)

*«И произвела земля зелень, траву, сеющую семя по роду ея, и дерево, приносящее плод, в котором семя его по роду его»*

*Бытие*

*«Мне кажется, что и сам Бог не может знать того, что произойдет случайно и произвольно. Ибо если знает, то это определено произойдет, а если определено произойдет, то не случайно»*

*Цицерон*

#### Введение

Более столетия в естествознании существовала ситуация, которая многим казалась парадоксальной. Эволюционные концепции, сформулированные в середине XIX в. почти одновременно в физике и биологии, имели противоположную направленность и как бы исключали друг друга.

Идея эволюции в физике была провозглашена Р. Клаузиусом в 1850 г. посредством второго начала термодинамики—закона необратимого возрастания энтропии изолированной системы. Л. Больцман, установив связь между энтропией и вероятностью состояния, дал второму началу статистическое обоснование, согласно которому необратимость в термодинамике выражает статистический закон эволюции макроскопической системы к «наиболее вероятному» состоянию, соответствующему состоянию максимального разнообразия. «Классическая термодинамика,—отмечает И. Пригожин,—есть в сущности теория разрушения структур, а производство энтропии можно рассматривать как меру скорости этого разрушения». Термодинамическая концепция

\*) Название этой главы совпадает с названием книги одного из крупнейших ученых нашего времени, лауреата Нобелевской премии 1977 г. И. Пригожина. От существующего к возникающему—эти слова очень точно выражают ряд положений, рассматриваемых в этой главе. По просьбе авторов текст этой главы частично написан проф. Е.М. Поповым. Расчеты для рис. 10.7—10.20 выполнены проф. В.А. Дементьевым.

деградации структур, выравнивания свойств всех составляющих системы не может описать естественным образом феномен жизни, ее возникновение, усложнение и совершенствование.

Эволюционная теория в биологии утверждает диаметрально противоположное направление развития—это теория не разрушения структур, а их создания. Главное в учении Ч. Дарвина, впервые обнародованном в 1859 г., состоит в объяснении движущих сил постепенного перехода по восходящей линии одних состояний живых форм к другим, более сложным и совершенным.

Эволюционные концепции как в физике, так и в биологии базируются на огромном экспериментальном материале. Но поскольку между физическими и биологическими опытными фактами долгое время не было установлено прямой связи и одна концепция эволюции касалась явлений только неорганической природы, а другая—только мира растений и животных, то надолго утвердилось и получило широкое распространение представление о несовместимости законов физики и биологии, об особой сущности явлений органического мира. Предпринятые попытки построить согласованную картину и установить связь между физической и биологической эволюцией, доказать непротиворечивость теории «создания структур» Ч. Дарвина теории «разрушения структур» Р. Клаузуиса не внесли чего-либо качественно нового ни в физику, ни в биологию. Все они оказались бессильными объяснить причины и конкретный механизм спонтанной структурной организации.

Только совсем недавно, на рубеже 70—80-х гг., естествознание сумело приблизиться к пониманию самопроизвольного формирования структур в самых различных физических, химических и биологических системах. Это потребовало осознания существенного различия в природе равновесных и неравновесных процессов, создания теории *открытых систем*, разработки теории особого типа динамических состояний—*диссипативных самоорганизующихся структур*—и введения в термодинамику ряда принципиально новых положений. В результате было строго доказано, что в открытых системах могут образовываться и длительное время существовать упорядоченные структуры без нарушения второго начала термодинамики, причем поведение энтропии в таких системах существенно отличается от поведения энтропии в изолированных системах. Кроме того, обнаружилась всеобщность процессов структурной самоорганизации не только в живой, но и в неживой природе. Для объяснения причин спонтанного возникновения «порядка из беспорядка» потребовалось значительно расширить теоретические основы классической термодинамики, что в результате привело к созданию сначала линейной (30—50-е гг.), а затем (70—80-е гг.) нелинейной термодинамики неравновесных процессов.

Нелинейная термодинамика, созданная главным образом трудами И. Пригожина, в настоящее время бурно развивается;

открывающиеся на наших глазах по мере ее становления перспективы необычайно широки и многообещающи. Уже сейчас ставшие известными закономерности процессов самоорганизации в открытых диссипативных системах, как классических, так и квантовых, используются во всех областях естествознания (физике, химии, биологии), а также в социологии, экономике, языкознании, лингвистике и т.д. Физика Пригожина привела к глубоким, качественным изменениям таких фундаментальных понятий, как время и динамика. Оформившаяся за последнее десятилетие на теоретической основе и методах нелинейной термодинамики новая междисциплинарная область знаний получила, по предложению Г. Хакена, название «синэргетика», что в переводе с греческого означает совместное кооперативное действие.

Тема самопроизвольного воспроизведения структур чрезвычайно актуальна для биологии, поэтому идеи и подходы нелинейной термодинамики имели здесь большой резонанс, получили быстрое и широкое распространение в исследованиях клеточных и надклеточных структур, дифференцирования клеток и ряда других проблем. В работах М. Эйгена, отмеченных, как и работы И. Пригожина, Нобелевской премией, впервые была сформулирована концепция отбора и эволюции биологических макромолекул из неупорядоченных систем на основе качественной теории гиперцикла как одного из принципов структурной самоорганизации открытых систем. Тем самым была доказана непротиворечивость, по крайней мере на элементарном уровне, эволюционных концепций физики и биологии. Таким образом, нелинейная неравновесная термодинамика уже в ранний период своего становления смогла привести к результату, имеющему фундаментальное, общенаучное значение для естествознания. Пожалуй, впервые в истории науки вопросы зарождения жизни, отбора и эволюции были рассмотрены теоретически, хотя и сугубо качественно, с чисто физических позиций.

Современный курс физики уже не может изучаться изолированно от природных явлений, лежащих на грани живого и неживого, поэтому данную главу можно рекомендовать для изучения в качестве последнего раздела либо в качестве введения в биофизику. В ней как бы перекидывается мост от квазистатистических явлений к развивающимся (возникающим) и рассматриваются фундаментальные особенности любых процессов, приводящих к появлению нового качества.

## 10.1. Время в классическом мире

Начнем с обсуждения некоторых проблем, связанных с введением и использованием понятия времени в классической механике Ньютона. Ограничимся при этом для простоты случаем одномерного движения. Вспомним уравнение движения для

материальной точки, движущейся под действием постоянной силы. Имеем

$$\frac{d^2x}{dt^2} = \frac{f}{m}.$$

При начальных условиях  $x(t=0) = 0$  и  $\left.\frac{dx}{dt}\right|_0 = v(t=0) = v_0$  решения его будут следующими:

$$v = \frac{f}{m}t + v_0 \text{ и } x = \frac{f}{2m}t^2 + v_0t.$$

Для того чтобы все стало вполне определенным, надо выбрать начало координат и положительное направление оси  $X$ . Начало координат совместим с точкой начала движения, а положительным будем считать направление вправо по оси  $X$  от начала координат. Время, как и обычно, принимаем положительным.

Теперь зададим себе вопрос: можно ли считать время развивающимся в отрицательном направлении, т. е. допустимо ли, что аргумент может иметь как положительные, так и отрицательные значения? Ведь выбор как начала координат, так и положительных и отрицательных направлений пространственных осей координат есть дело совершенно условное. Мы привыкли всегда считать время положительным, но уместно спросить, есть ли физическое основание для этого? Таким физическим основанием может быть только эксперимент, четко указывающий на то, что развитие события при  $t < 0$  невозможно.

Что же получается в механике?

В уравнение движения дифференциал времени входит как квадрат. Значит, величина  $d^2x/dt^2$  не будет меняться при замене  $dt$  на  $-dt$ . Таким образом, основное уравнение движения инвариантно относительно смены знака времени.

Если в формуле для  $x = x(t)$  поменять знак  $t$ , то тоже ничего не изменится. В самом деле, если считать, что время развивается от 0 до  $-t$ , то  $dt < 0$  и тогда надо и  $v_0$  считать отрицательной. Значит, можно сказать, что движение с отрицательным временем полностью эквивалентно движению с положительным временем. Результаты физического эксперимента, например по измерению длин отрезков, проходимых точкой вдоль оси  $X$  через равные промежутки времени, будут одними и теми же.

Значит, механический эксперимент абсолютного знака времени не устанавливает и с точки зрения классической механики в равной степени возможны события, происходящие как от прошлого к будущему, так и от будущего к прошлому! Поясним это еще одним наглядным примером. Представим себе, что мы сняли кинофильм о движении точки  $M$  вдоль оси  $X$  от начального, принятого нами за нуль, положения до некоторого

положения  $x_M$  максимального. Точка будет перемещаться равно-  
ускоренно по экрану в соответствии с законом  $x = \frac{f}{2m} t^2 + v_0 t$   
слева направо. Запустим теперь тот же фильм в обратном  
направлении. Зритель увидит равнозамедленное движение точки  
справа налево, причем за тот же промежуток времени она  
достигнет крайнего левого положения и будет иметь в этом  
положении скорость, по величине равную  $v_0$ .

Посмотрим, нельзя ли поставить реальный физический экс-  
перимент, в точности воспроизводящий наблюдаемую при таком  
обратном просмотре картину движения точки?

Так как движение во втором случае должно быть, очевидно,  
равнозамедленным и совершаться справа налево, то считая, как  
мы это и привыкли делать, что  $t > 0$ , запишем  $-\frac{d^2x}{dt^2} = \frac{f}{m}$ .

Получим

$$-v_{\text{кон}} = \frac{f}{m} t_M - v_M,$$

где  $v_{\text{кон}}$  — скорость точки в конце пути, пройденного за время  
 $t_M$ ,  $v_M$  — начальная скорость. Время  $t_M$  следует принять равным  
времени достижения точкой крайнего правого положения в первом  
эксперименте. На основании соотношения  $v_M = \frac{f}{m} t_M + v_0$  по-  
лучим, что

$$t_M = \frac{m}{f} (v_M - v_0).$$

Следовательно,

$$-v_{\text{кон}} = \frac{f}{m} \cdot \frac{m}{f} (v_M - v_0) - v_M = -v_0.$$

Не трудно проверить, что если точка при своем движении  
справа налево стартует из положения  $x_M = \frac{f}{2m} t_M^2 + v_0 t_M$ , то  
через время  $t_M$  при таком движении она дойдет до положения  
 $x = x_0 = 0$  (см. рис. 10.1).

Значит, при подходящих начальных условиях можно реально  
осуществить такой эксперимент, который в точности будет  
соответствовать картине при обратном просмотре кино. Отсюда  
ясно, что зритель, не знающий, какой из двух способов показа  
фильма выбран в данный момент, не сможет сказать, видит ли  
он в кино реальный эксперимент или обратное прокручивание  
плёнки. Движение от прошлого к будущему легко спутать с  
движением от будущего к прошлому!

Классические эксперименты в механике не дают для различения  
этих процессов никакой возможности.

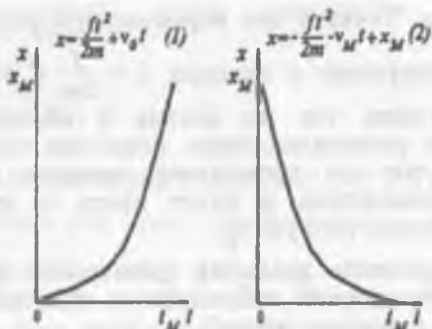


Рис. 10.1. Материальная точка массой  $m$  движется с ускорением по закону (1). Если пустить фильм в обратном направлении, это будет эквивалентно обратному движению точки по закону (2). В классической механике движение обратимо

К сожалению, не спасает положение и переход к механике Эйнштейна, к электродинамике и квантовой механике. Все результаты в этих областях физики тоже обратимы относительно времени. Например, если во временном уравнении Шредингера заменить  $dt > 0$  на  $dt < 0$ , то это не отразится на производной  $d\psi/dt$ , если одновременно заменить  $\psi$  на ее комплексно-сопряженную, которая также удовлетворяет исходному уравнению. Ничто не запрещает событиям развиваться как в направлении от 0 до  $t$ , так и в направлении от 0 до  $-t$ . Во всех этих областях физики время рассматривается просто как некоторый параметр, в известной степени равноправный с пространственными координатами, которому можно придавать как положительные, так и отрицательные значения. Получается, что мы выбираем всегда  $t > 0$  в силу простой договоренности!

Однако так ли это в более сложном случае?

Снова вернемся к кино. Можно снять фильм, и это многократно делалось, о развитии растения из семени. Можно «прокрутить» его в обратном направлении. Если теперь показать «прямой» и «обратный» фильмы зрителю, то он сразу отличит один от другого и поймет, что один способ показа отвечает реальному ходу развития событий, а второй — в природе не существующей, или запрещенной, ситуации. В большинстве из изучаемых нами разделов физики мы с таким запрещением событий при формальной замене  $t$  на  $-t$  не встречались. Значит, соответствующее физическое описание, принятое в этих разделах, неполно и отражает лишь какие-то стороны реальной природы, не затрагивая ее глубинных свойств, учет которых должен автоматически привести к тому, что ситуации, в которых можно было бы считать  $t < 0$ , запрещены. Здесь и должно быть что-то напоминающее известный в квантовой физике принцип запрета Паули. Этот важнейший вопрос, как отмечалось во введении, фактически начал разрабатываться только в наш век, хотя понятием времени (причем всегда интуитивно считая, что  $t > 0$ ) физики пользовались уже не одну сотню лет. Парадокс заключался в том, что хотя весь опыт человечества говорил о существовании направленного развития событий (это можно математически интерпретировать как условие разрешенности

$t > 0$  и запрещенности  $t < 0$ ), однако это не находило никакого отражения в такой высокоразвитой науке, как физика. Такая ситуация, в свою очередь, породила философские споры, трудности и, возможно, отрыв в какой-то степени физики как науки о природе от таких наук, как геология, биология и обществоведение. Это не означает, конечно, что эти науки не пользуются физикой. Как раз напротив. Именно физика лежит в основе молекулярной биологии. Но если во всех упомянутых выше науках идея развития от прошлого к будущему составляет саму их основу, то, если можно так сказать, интерес к истории развития событий был во многом чужд физике. Это ставило барьер между физикой и другими науками, изучающими высшие формы существования материи.

В заключение этого параграфа укажем, что для формирования понятия времени как характеристики развивающейся в одном направлении (от прошлого к будущему) последовательности явлений важнейшую роль играло то, что на непрерывно меняющуюся череду принципиально несовпадающих событий (вспомним древний афоризм: нельзя дважды войти в одну и ту же реку!) накладывались природные периодические процессы, связанные с движением Земли вокруг Солнца, Луны вокруг Земли, сменой дня и ночи. Следствиями высокой стабильности повторяемости этих процессов явилось сначала интуитивное, а затем закрепленное в физике деление времени на одинаковые интервалы и принятие астрономической шкалы времени в качестве единственной. Опыт воспроизведения результатов физических экспериментов, выполненных в разное время иногда с интервалом во много лет, показал, что законы физики инвариантны относительно выбора начала отсчета времени. Математически это возможно, если при изменении начала отсчета значения отрезков времени от начала до данного момента различаются на постоянную величину. Другими словами, должно быть  $t' = t + \Delta t$ , где  $\Delta t = \text{const}$ . Только в этом случае  $dt' = dt$  и все уравнения физики остаются неизменными. Эти же доводы приводят к постулату о равномерном течении времени.

Не следует, однако, забывать, что одних астрономических наблюдений без сопоставления их с развитием биологических, геологических и исторических событий не хватило бы для осознания времени как непрерывно и равномерно меняющейся в одном направлении характеристики. В самом деле, вспомним, как легко сбиться с отсчета времени, если хотя бы в течение часа, следя лишь за полным оборотом стрелки в секундомере (один оборот—одна минута), ограничиться наблюдением неподвижных предметов внутри одной комнаты. Мы не собьемся, если после каждого оборота стрелки будем выкладывать на стол по спичке. Накопление спичек—это уже типичный неперiodический и необратимый процесс.

Такая двойственность факторов, приводящая к общему пониманию времени, является частным проявлением принципа до-



полнительности, на котором мы достаточно полно остановимся в следующей главе.

Эта двойственность (взаимная дополнительность), заложенная в понятии времени, приводит к одному интересному следствию. Оказывается, что некоторые законы физики, возникшие как результат обобщения наблюдений и опытов, «привязанных» к астрономической шкале времени, могли бы быть другими, если бы для отсчета времени использовались другие явления. Например, можно показать, что если воспользоваться нелинейной по отношению к астрономической шкалой времени (например, базирующейся на накоплении продуктов радиоактивного распада), то формулировку первого закона Ньютона (закона инерции) надо изменить. Это не значит, что другими станут результаты экспериментов. Скорректировать надо будет только их трактовку.

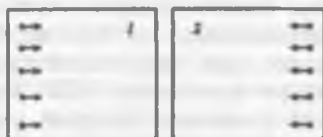
## 10.2. Необратимость процессов и «стрела времени»

Выше мы отметили, что в тех случаях, когда мы имеем дело с явлениями, целиком или в очень хорошем приближении описываемыми уравнениями движения классической, релятивистской или квантовой механики или уравнениями Максвелла, понятие необратимого течения времени возникнуть не может.

Однако где-то оно должно появиться?

Для этого, очевидно, нужно обратиться к явлениям, необратимым по своей природе. С такими явлениями мы встречаемся, когда переходим к изучению процессов, происходящих в системах с очень большим числом частиц, составляющих газ, жидкость и др., и когда для описания их состояний приходится отказаться от обычных средств механики и перейти к другим. Напомним, что для таких систем целесообразно наряду с микроскопическими ввести и понятие макроскопических состояний. Если процессы на микроскопическом уровне подчиняются законам классической механики (например, упругое соударение двух частиц идеального газа можно описать в рамках механики Ньютона) и в этом смысле они все обратимы во времени, то процессы в большом коллективе, т. е. макроскопические, приобретают уже новое качество и в некоторых случаях становятся направленными, т. е. развиваются лишь в одной последовательности во времени. К таким направленным процессам относятся процессы выравнивания температуры на различных участках замкнутых систем, т. е. теплопередача; равномерное заполнение газом всего объема, если газ первоначально находился в одной части этого объема, т. е. диффузия. Интересно при этом отметить следующее. Сейчас, благодаря использованию ЭВМ, реальными стали математические эксперименты, при которых, пользуясь обычными приемами ньютоновской механики и предполагая, что столкновения между шариками, размеры которых много меньше допустимой области движения, являются абсолютно упругими (как и с ограничивающими область движения

Рис. 10.2. В кадре 1 молекулы одновременно с одинаковыми скоростями двинулись от левой стенки идеально отражающего ящика к правой стенке. В кадре 2 они отразились от правой стенки. Так они будут летать вечно, хаос не наступит



стенками), рассчитываются точно траектории многих (до 1000) взаимодействующих таким образом частиц. Все эти эксперименты приводят к одному и тому же результату: независимо от выбора начальных условий, но при правильном физическом задании, конечным, в дальнейшем уже практически не меняющимся состоянием является такое, когда все участвующие в эксперименте шарики распределены в пространстве достаточно равномерно. Таким образом, машинный эксперимент в точности воспроизводит реальную ситуацию.

Возникает вопрос: каким образом, отправляясь от обратимых во времени уравнений механики, но применяя их сразу ко многим частицам, удастся получить новое качество—направленный процесс, который естественно связать с появлением направленного времени, или, как иногда говорят, ввести «стрелу времени»? Понятно, что здесь должно сыграть свою роль именно множество, но только ли это?

Мы знаем, что результат решения одного уравнения движения зависит от начальных условий. Конечно, и для множества это должно быть верным. Если теперь считать, что все частицы первоначально находились на одной плоскости в левой части прямоугольного сосуда и имели одну и ту же скорость, направленную перпендикулярно к правой и левой стенкам, то ясно, что никакого постепенного распространения всех частиц по объему не произойдет: они просто будут летать от одной стенки до другой (см. рис. 10.2). Чтобы получить согласующееся с экспериментом решение, надо задать самые разнообразные как по направлению, так и по величине начальные скорости всех частиц системы (см. рис. 10.3). А это фактически означает, что надо внести в саму постановку задачи принципиальный элемент случайности. Так и поступают: начальные условия генерируют на ЭВМ с помощью датчика так называемых случайных чисел. Теперь в результате решения системы уравнений движения для

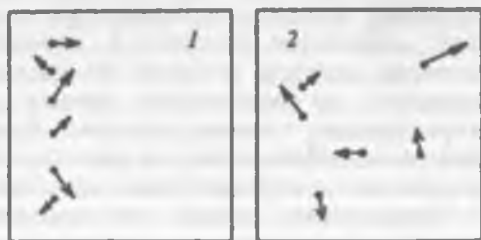


Рис. 10.3. В кадре 1 молекулы стартовали одновременно, но с разными скоростями. В кадре 2 после нескольких столкновений друг с другом и со стенками ящика молекулы движутся хаотически. Процесс перехода к хаотическому движению необратим

всех частиц можно увидеть, как они распределяются равномерно по предоставленному им объему. Однако наблюдаемый на экране ЭВМ процесс все еще является обратимым. Если в какой-то момент вдруг заставить ЭВМ изменить знак времени, то при дальнейшем решении системы уравнений движения ЭВМ снова соберет все частицы в их начальные положения с начальными скоростями (с точностью до ошибок вычислений). На экране дисплея мы увидим, как частицы, расположенные вначале на плоскости, разлетаются по всему объему, а затем снова прилетают к исходной плоскости. Для того чтобы процесс стал необратимым, необходимо внести в расчет дополнительные элементы случайности. Нужно учесть, что частицы имеют конечные размеры и что удары между соударяющимися частицами не обязательно центральные. Нужно учесть, что молекулы стенок сосуда совершают хаотические движения и что налетающие на них частицы сталкиваются с ними под случайными углами. Тогда точный расчет покажет, что процесс разлета частиц по всему объему является необратимым. Как ни меняй знак времени в процессе решения системы уравнений движения, частицы никогда не соберутся на исходные позиции.

Мы получаем важнейший вывод: появление направленного процесса в системе многих сталкивающихся шариков есть следствие случайности! Именно возникающий при этом вероятностный элемент и является тем новым, что целиком отсутствует в классической механике. На самом деле ситуация еще более сложна: в классической механике вообще нельзя точно решить так называемую проблему трех тел. Если одновременно столкнутся несколько частиц, то абсолютно точно предсказать, что в результате получится, невозможно. Так как частицы после каждого столкновения как бы «забывают» свои начальные движения, то ясно, что с течением времени вероятностный элемент будет только возрастать. Частицы реального газа взаимодействуют сложнее, чем по правилам упругого удара, а это еще усилит элемент случайности.

Значит, механический принцип детерминизма, который гласит, что результат механического движения полностью определяется уравнением движения и начальными условиями, оказывается совершенно не применимым к системе многих частиц. Здесь как раз получается, что разные в смысле разной выборки случайных чисел начальные условия приводят к одному и тому же конечному макроскопическому результату. Это важнейший момент: количество перешло в качество! Пойдем дальше. Если состояние системы меняется во вполне определенной последовательности, то естественно связать эту последовательность с направлением течения времени. Если, кроме того, обратный процесс не наблюдается, то понятно, что изменить знак времени уже нельзя: в отрицательном времени не происходят события!

Предположим теперь, что существует некоторая характеристика (функция) макроскопического состояния системы, которая

со временем монотонно возрастает или убывает до предельной величины по мере того, как система приближается к равновесному состоянию: газ полностью заполнил сосуд и давление выравнялось во всех участках. Тогда ясно, что поведение этой функции обладает важнейшим свойством—она также будет отражать направленность процесса. В физике такая функция действительно существует—это энтропия ( $S$ ). В гл. 6 было показано, что при адиабатическом расширении газа в вакууме (это отвечает условиям описанного выше эксперимента на ЭВМ) такая функция может только возрастать. Эта функция вообще возрастает при любом необратимом процессе в замкнутой системе. Напомним, что в физике под обратимыми понимаются такие процессы перехода системы из одного состояния в другое, которым можно сопоставить реально возможный обратный переход, последовательно повторяющий все стадии прямого.

Например, это можно сделать, если газ очень медленно последовательно переходит из состояния с параметрами  $p_1, V_1, T_1$  в состояние  $p_2, V_2, T_2$ , так, что в любой момент времени во всех частях сосуда имеется практически одинаковое число частиц с одной и той же средней скоростью (установившиеся состояния). При необратимом процессе эти последние условия не выполняются. Наглядными примерами такого рода являются диффузия и теплопроводность. Для необратимых процессов производная  $\frac{dS}{dt} > 0$ . Заключение в этом выражении утверждение, которое является одной из формулировок второго начала термодинамики, приобретает, таким образом, фундаментальный смысл, теснейшим образом связанный с утверждением о том, что  $t > 0$ .

Итак, показателем направленных необратимых перемен в системе является возрастание энтропии. Уместно спросить, от какой начальной величины? Этот вопрос был разрешен Клаузиусом, который нашел, что энтропия при температуре абсолютного нуля

$$S(T = 0) = 0.$$

При  $T = 0$  всякое поступательное движение частиц прекращается, т. е. они будут находиться в каких-то фиксированных положениях. Это позволяет высказать утверждение о том, что любое строго фиксированное состояние системы обладает нулевой энтропией. Если назвать это порядком, а переход к хаотическому движению и распределению частиц беспорядком, то можно сказать, что энтропия возрастает при переходе от порядка к беспорядку. Так как энтропия характеризует как пространственное распределение частиц, так и пространственное распределение энергии, т. е. зависит от двух важнейших макроскопических показателей сложной системы, то понятно, что именно эта функция положена в основу новой механики множественных объектов, которая сейчас развивается и представляет одно из интереснейших направлений

современной теоретической физики. Получающиеся при этом уравнения состояний системы уже неинвариантны относительно времени.

### 10.3. Флуктуации, бифуркации и самоорганизация

Выше мы отметили, что для сложных систем со многими частями характерно наличие «стрелы времени». Но не только это отличает данные системы. Вспомним хорошо известный из школьного курса опыт с броуновским движением. Опыт состоит в следующем: в жидкость помещается легкая, наблюдаемая в микроскоп частица, например частица цветочной пыльцы. Если следить за частицей достаточно долгое время, то видно, что она совершает сложные хаотические, но вполне регистрируемые

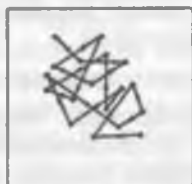


Рис. 10.4. След частицы в поле зрения микроскопа, совершающей броуновское движение

движения (рис. 10.4). Теория броуновского движения была дана А. Эйнштейном, который показал, что оно является результатом возникновения по разные стороны частицы областей с разным числом молекул жидкости, и это приводит к мгновенной разности эффектов соударений молекул о частицу. Получающаяся при этом нескомпенсированная сила и толкает пылцу. Вся ситуация полностью описывается в рамках теории случайных процессов. Так как броуновское движение не прекращается со временем, то это означает, что такие участки неравномерной плотности вещества (размер этих участков, разумеется, мал) существуют всегда и при установившемся в среднем распределении жидкости или газа в объеме. Такие статистические все время возникающие и исчезающие области с несколько большим или меньшим, чем в среднем, числом частиц в малом объеме называются *флуктуациями*. Значит, множество частиц может обладать, и мы должны это принять как эмпирический новый опытный факт, еще одним важным свойством—иметь флуктуации. Во многих случаях они не приводят к заметному макроскопическому эффекту, и поэтому ими вообще можно пренебречь, что и делается обычно при изучении, например, кинетической теории газов и жидкостей.

Но всегда ли это так?

Рассмотрим один абстрактный пример. Вспомним задачу о колебаниях маятника под действием периодической силы. Во всей области, когда частота воздействия этой силы  $\omega \neq \omega_0$ , т. е. частоте собственных колебаний маятника, получается решение, отвечающее сумме двух колебаний—вынужденных и сопутствующих. Однако в области  $\omega \approx \omega_0$  характер решения принципиально меняется—получается лишь одно линейно возрастающее по амплитуде колебание. Такое качественное изменение поведения объекта при некоторых критических значениях определяющих

этот объект параметров называется *бифуркацией*. Возможность появления бифуркаций существенно связана с неустойчивостью объекта при определенных условиях. В общем случае при бифуркациях развитие процесса возможно в разных направлениях. При этом выбор направления развития процесса определяется, как правило, случайными факторами.

Приведем пример. Пусть имеется материальная точка, состояние которой на оси  $X$  описывается уравнением  $md^2x/dt^2 = -Kx$ . Предположим, что коэффициент  $K$  всегда мал и может меняться случайно в области значений  $K$ , близких к нулю. Тогда возможны случаи разного (причем случайного) развития событий: при  $K = 0$  точка будет покоиться или двигаться равномерно (это зависит также от случайных начальных условий); при  $K > 0$  будет наблюдаться хотя и медленный, но экспоненциальный рост скорости движения; при  $K < 0$  возникнут низкочастотные (с собственной частотой  $\omega_0$ ) гармонические колебания.

Пусть, далее, в среду, в которой имеется большое число подобных точек вводится энергия в форме также периодических воздействий с частотой  $\omega$ . Ясно, что в этом случае те элементы общей среды, которые в результате случайного стечения обстоятельств оказались способными совершать колебательные движения (перешли в колебательное состояние), начнут при подходящих условиях (в данном случае при  $\omega = \omega_0$ ) резонировать.

Амплитуда колебаний и, следовательно, локализация энергии от внешнего источника в соответствующих местах среды начнет быстро возрастать.

Флуктуация начнет разрастаться!

Сходный эффект будет наблюдаться в среде, состоящей из большого числа связанных маятников, если случайно один или несколько маятников изменяться таким образом, что их собственные частоты начнут совпадать с частотой внешнего воздействия. «Отсос» резонирующим маятником энергии из внешней среды приведет к тому, что доли энергии, которые будут приходиться на каждый из остальных маятников будут падать. Если увеличивать различие в собственных частотах выделенного маятника и остальных, то при подходящих внешних воздействиях будет колебаться практически только он один, причем с большой амплитудой.

Интересно отметить, что такой эффект будет наблюдаться уже при довольно малых (на 20—30%) отличиях собственной частоты колебаний выделенного маятника от других и быстро нарастать с ростом этого различия.

Можно сказать, что возникает флуктуация. Однако, в отличие от короткоживущей флуктуации плотности частиц в малом объеме жидкости или газа в установившемся состоянии, рассмотренная флуктуация будет возрастать со временем. Такая флуктуация уже не является микроскопической, а приводит к

макроскопическому эффекту. Этот процесс носит название *самоорганизации*.

Рассмотренный простой пример наглядно демонстрирует некоторые характерные черты самоорганизации. Во-первых, должно быть так, чтобы небольшие изменения какого-то элемента системы могли бы привести к принципиально иному характеру его поведения. Во-вторых, и это очень важно, чтобы в результате такого изменения элемент приобрел способность энергично забирать энергию от окружающей среды. Другие флуктуации в среде в достаточно большой области около самоорганизующегося элемента при этом начнут подавляться в силу требования закона сохранения энергии.

Для начала процесса необходимо, чтобы в системе появились микрообразования, склонные к бифуркациям. Тогда подходящая микрофлуктуация может как бы запустить процесс нарастания неустойчивости и привести к макроскопическому эффекту. В этом проявляется один из важнейших и весьма общих законов природы: *всякое развитие сложной системы, выражающееся в появлении нового качества, возможно только при наличии в системе флуктуаций, склонных к разрастанию*. Если в системе отсутствуют подходящие флуктуации или уровень их недостаточен, то такая система к саморазрастанию просто не способна. Это верно для всех систем, включая и общество людей. Это требование, будучи необходимым, однако, не является достаточным. Для эволюционирования какой-то флуктуации так, чтобы в целом можно было говорить о появлении нового качественного макроскопического состояния системы, необходимо выполнение определенных условий, о которых мы будем говорить ниже.

Отметим, что всякое разрастание флуктуаций в среде до макроскопических размеров, или самоорганизация, неизбежно приводит к появлению порядка из беспорядка и является процессом, прямо противоположным тем, которые мы рассматривали выше. Значит, таким процессам свойственно уменьшение энтропии. Получается, что второе начало термодинамики уже не работает?

Попробуем разобраться.

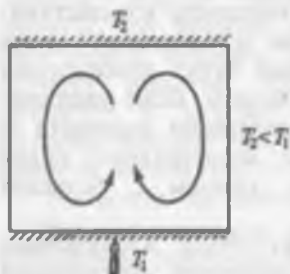


Рис. 10.5. Самоорганизация в прогреваемой жидкости

Остановимся сначала на некоторых простейших примерах самоорганизации в газе и жидкости. Опыт первый. Хорошо известно, что если поместить газ или жидкость в сосуд, дно и «крыша» которого находятся при разных температурах  $T_1$  и  $T_2$ , то в среде при определенных

условиях (горелка помещается в центре дна сосуда, стенки его охлаждаются и др.) возникают устойчивые конвекционные потоки частиц (рис. 10.5). Не только появляется постоянно существующее, пока  $T_1 \neq T_2$ , макроскопическое движение, но и плотность частиц в

конвекционных токах оказывается иной, чем вне их. Это и позволяет заметить сами эти токи; особенно хорошо они видны, если жидкость подкрасить. Сразу укажем, что такая конвекция наблюдается не всегда, а только тогда, когда сосуд становится достаточно большим. То есть существует некоторый критический размер, ниже которого эффект не обнаруживается. Возникает известная аналогия с цепной реакцией в атомной бомбе, где взрыв возникает только тогда, когда масса заряда становится выше некоторой критической.

Второй опыт заключается в появлении устойчивых вихрей в потоке жидкости. Эти вихри, легко наблюдаемые и макроскопические, можно видеть и на поверхности текущей реки (водовороты). Снова мы имеем дело с самоорганизацией.

Не будем приводить примеры, которых очень много. Остановимся на важнейших особенностях такой самоорганизации. Во-первых, она возможна только в средах значительного объема. Во-вторых, для ее поддержания требуется постоянный подвод внешней энергии, иначе, например, не могут существовать долго условия  $T_1 > T_2$ . В-третьих, как показывает специальное исследование, рассмотренные выше процессы могут быть только в реальных, а не идеальных газах, т. е. нужно, чтобы между частицами существовали взаимодействия типа сил Ван-дер-Ваальса или более сильные. Другими словами, самоорганизация в данном случае возникает как результат совокупного действия большого числа короткодействующих межмолекулярных сил.

Мы рассмотрели примеры самоорганизации при наличии внешнего источника энергии. Но самоорганизация может возникнуть не только в этих случаях. Хорошо известно, что при охлаждении газа он сначала превращается в жидкость (за счет межмолекулярных сил получают сгустки молекул, которые постепенно разрастаются в капли), а затем может образоваться и кристалл. В идеальном кристалле все атомы расположены в строго фиксированных положениях. Снова мы имеем процесс самоорганизации. В этом случае не только не надо подводить энергию, но, наоборот, ее надо вывести из системы, так как кристаллическое состояние энергетически более выгодно, чем, например, жидкое.

Во всех рассмотренных нами случаях мы имели дело с необратимыми процессами, при которых энтропия уменьшается (при кристаллизации до минимума), но, и это принципиально, при этом всегда имелся контакт с внешней средой. Наличие такого контакта и является тем существенным фактором, который отличает необратимые процессы в изолированных системах от аналогичных процессов в неизолированных, или открытых, системах.

Второе начало термодинамики и утверждение о возрастании энтропии при любом необратимом процессе было сформулировано именно для изолированных систем. Значит, нельзя просто переносить его на неизолированные. Чтобы перейти к описанию



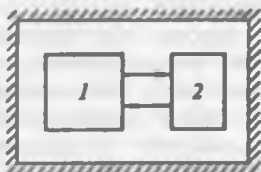


Рис. 10.6. Сообщающиеся системы в изолирующей оболочке

неизолированных систем, надо расширить понятие энтропии и считать, что ее приращение  $dS$  состоит уже из двух частей:  $dS = dS_{\text{сист}} + dS_{\text{пер}}$ , где  $dS_{\text{сист}}$  означает приращение энтропии в системе, а  $dS_{\text{пер}}$  — энтропию, перенесенную

через границы системы. Утверждение о том, что энтропия только возрастает при необратимых процессах, относится лишь к члену  $dS_{\text{сист}}$  и поэтому считается, что

$dS_{\text{сист}} \geq 0$ . Величина  $dS_{\text{пер}}$  может быть как положительной, так и отрицательной, поэтому как положительной, так и отрицательной может быть и полная величина  $dS$ . Если система нацело изолирована, но не находится в состоянии равновесия и в ней происходят какие-то макроскопические изменения, то они могут развиваться самопроизвольно (это означает, что система полностью изолирована) только в направлении, которое приводит к возрастанию энтропии. Процессы с уменьшением энтропии возможны исключительно при контакте с внешней средой. Этот контакт с внешней средой должен быть постоянным в тех случаях, когда необходимо поддерживать устойчивое динамическое состояние системы или усиливать его (случай конвекции или вихрей в потоке жидкости), и временным — до перехода системы в конечное состояние (случай кристаллизации). Таким образом, «антиэнтропийные» процессы могут быть только в неизолированных, или открытых, системах.

Теперь рассмотрим две системы, связанные между собой (рис. 10.6). Обе эти системы вместе будем считать изолированными от внешней среды. Полное изменение энтропии в этой объединенной системе  $dS = dS_1 + dS_2 > 0$  при любых необратимых процессах. Если в одной ее части в результате, например, самоорганизации происходит такой процесс, что  $dS_1 < 0$ , то обязательно должно быть  $dS_2 > 0$ , причем  $|dS_1| < |dS_2|$ . Другими словами, самоорганизация и уменьшение энтропии в одной части системы неизбежно приводят к возрастанию и дезорганизации (возрастание энтропии означает переход от порядка к беспорядку) в другой ее части, причем в большей мере. Вот это последнее чрезвычайно важно и дает нам ключ к пониманию некоторых общих вопросов, прямо относящихся не только к низшим формам организации материи, но и к высшим, т.е. к деятельности живых организмов.

## 10.4. Самоорганизация и условия существования живых организмов

Мы не знаем в деталях, как возникла жизнь на Земле. Можно только предполагать, что она образовалась в результате

длительного процесса и многих флуктуаций, вначале происходивших в так называемом «первичном бульоне». В этом первичном бульоне зародились такие флуктуации, которые были способны к бифуркациям. Этот процесс привел к первичной самоорганизации сначала на молекулярном уровне. Колонии молекул, новые множества, оказавшиеся способными к новым флуктуациям в своей организации и бифуркациям, привели к появлению простейших организмов и т. д. Процесс появления флуктуаций в объектах все более сложной природы в конце длительной эволюции, сопровождающейся естественным отбором, привел к возникновению живых организмов—растений и животных. Такой процесс, во-первых, на начальной стадии носил чисто случайный характер, поскольку появление в этой цепи флуктуаций как некоторых образований, способных к бифуркациям, есть процесс вероятностный, а во-вторых, требовал непрерывного подвода энергии. Энергия—это прежде всего излучение Солнца. Так постепенно возникла Биосфера. Живые компоненты Биосферы функционируют как открытые системы, причем обмен с внешней средой идет прежде всего за счет поглощения живыми системами энергии. Она, в свою очередь, доставляется в организм с помощью обмена веществ. Возьмем такой пример. Взрослый человек, если он здоров, практически не меняется в течение года. Между тем за год он съедает пищу, по массе своей превосходящую его собственную. Постоянство собственного веса взрослого человека в течение длительного времени как раз и является доказательством того, что вещество пищи используется фактически лишь для транспорта энергии. В этом смысле жизнедеятельность человека напоминает опыт с конвекцией жидкости или газа. Человек и внешняя среда в целом есть система в довольно хорошем приближении замкнутая. Поэтому, получая энергию из внешней среды, человек одновременно должен увеличивать энтропию этой внешней среды. Это проявляется, в частности, в том, что, потребляя вещество из внешней среды в виде высокоорганизованной материи (тканей животных и растений), человек выделяет вещество в форме материи менее организованной. Общим биологическим законом является то, что более высокоорганизованный организм способен получать необходимую для жизнедеятельности энергию в основном в виде также достаточно высокоорганизованной материи. Поэтому животные не могли возникнуть раньше растений, а последние—раньше менее высокоразвитых простейших организмов, лежащих на грани живого и неживого.

Вспомним теперь (см. § 6.7), что при прочих равных условиях нарастание энтропии идет быстрее в средах, где уровень энтропии меньше. Следствием высокой степени организации животных является то, что их развитие должно сопровождаться энергичным разрушением также высокоструктурированной среды.

Растительный мир и питающий других животных также животный мир как раз и обладают этим свойством. Непосредственно почва с ее гораздо более высоким уровнем энтропии для этой цели не подходит. Это не значит, конечно, что это общее правило, верное в целом для двух взаимодействующих подсистем, не может нарушаться в частности. Например, можно поддерживать на некоторое время жизнь человека, вводя в организм так называемый физиологический раствор, но очень долгим такое состояние быть не может. В то же время растения можно выращивать методом гидропоники, заменяя обычную почву специальной жидкой питательной средой.

Общий транспорт вещества в Биосфере, приводящий к необходимому транспорту энергии, может быть изображен следующей грубой схемой:



При этом вещество накапливается в растениях и животных при их росте. Это приводило бы к конечному обеднению почвы, если бы Природа предусмотрительно не ввела бы в этот цикл Смерть. Умершее растение или животное уже не в состоянии обмениваться с внешней средой веществом и поэтому не получает энергии в нужной форме. Умерший организм представляет собой замкнутую систему, в нем процессы могут идти только в направлении возрастания энтропии, т. е. от порядка к хаосу. За счет этого и происходит, во-первых, возвращение накопленного вещества в первичный источник, т. е. в почву, и, во-вторых, поддерживается высокий уровень энтропии в этом источнике, что необходимо для существования жизни в течение длительного времени и в большом пространстве. Таким образом, *смерть отдельных представителей живой природы является необходимым условием жизни всей Биосферы.* Именно этим, скорее всего, объясняется биологическая запрограммированность конечности жизненного цикла любых достаточно высокоорганизованных биологических объектов. Например, клетки способны делиться лишь определенное число раз, а потом умирают.

Интересно, что взгляд на Смерть не как на трагедию, а как на необходимый этап Жизни, составляет один из важных элементов многих религиозных учений и в этом смысле согласуется с современными научными воззрениями.

Круговорот вещества как переносчика энергии в Биосфере очень хорошо сбалансирован. Любое нарушение этого баланса чревато трагическими последствиями для животного мира и человеческого общества. Именно поэтому постепенно осознаются как первостепенные для самого факта существования человечества общие и частные экологические проблемы.

## 10.5. Взаимодействие подсистем

Выше мы говорили о возможности разрастания флуктуаций. Пока мы ограничивались только одной системой. Посмотрим, что будет, если систем несколько и они могут взаимодействовать.

Составим следующую математическую модель. Пусть имеется несколько подсистем:  $N_1$ ,  $N_2$ ,  $N_3$  и т. д. Предположим, что каждая из этих подсистем способна развиваться. Степень развития каждой подсистемы в данный момент времени будет характеризоваться величиной  $F_k$ . Приращение степени развития (например, возрастание массы вещества в некоторой области пространства, где происходит разрастание флуктуаций) выражаем дифференциалом величины  $F_k$ , т. е.  $dF_k$ . Величина  $dF_k$  в простейшем случае оказывается, как правило, пропорциональной достигнутой к данному моменту времени степени развития, т. е.  $F_k$ . Обозначим соответствующий коэффициент пропорциональности  $a_k$ . Чтобы учесть взаимосвязь подсистем, введем коэффициенты  $b_{kn}$ .

В результате процесс накопления свойств в каждой из подсистем можно охарактеризовать следующей системой линейных дифференциальных уравнений первого порядка:

$$\frac{dF_k}{dt} = a_k F_k + \sum_{n \neq k} b_{kn} F_n; \quad k = 1, 2, 3, \dots, N.$$

Решения этой системы уравнений имеют в зависимости от значений коэффициентов  $a_k$  и  $b_{kn}$  вид сумм либо экспонент:

$$F_k = F_{0k} \exp(\lambda_k t) + \sum_{n \neq k} F_{0n} \exp(\lambda_n t),$$

либо экспонент и гармонических функций. Величины  $\lambda_k$  и  $\lambda_n$  зависят от всех коэффициентов  $a_k$  ( $a_n$ ) и  $b_k$  ( $b_n$ ) в исходной системе уравнений и в общем случае могут быть как положительными, так и отрицательными, а также комплексными (последнее и приводит к появлению гармонических составляющих).

Результаты соответствующих машинных игр показывают, что быстрее развиваются те подсистемы, в которых с самого начала были большие флуктуации (величины  $F_{0k}$ ). Эти разрастающиеся флуктуации «тянут» за собой другие (или подавляют), причем тем сильнее, чем больше коэффициенты связи  $b_k$ . С течением времени превалирующей по своему развитию может оказаться лишь одна, наиболее быстро разрастающаяся подсистема. В данном случае закон развития, однако, является более сложным, чем тогда, когда подсистема не связана с другими (вместо простейшей экспоненты получается сумма экспонент или коле-

бания). Заметим теперь, что в нашей модели подразумевалось, что сам процесс разрастания связан с наличием какого-то внешнего источника энергии, безотказно поставляющего эту энергию всем подсистемам. На практике чаще всего встречается такой случай, когда одна подсистема не просто связана с другой подсистемой, а черпает из нее энергию, необходимую для саморазвития. Именно такие взаимодействия и приводят к ограничению в саморазвитии флуктуаций.

Для анализа этой проблемы воспользуемся примером двух взаимодействующих подсистем, одна из которых черпает необходимую ей для развития энергию, вещество или другие компоненты из другой. Такая задача называется задачей о *хищниках и жертвах*. Для наглядности представим себе, что в некоторой ограниченной среде обитают лисы и зайцы. Как известно, если просто иметь любую совокупность живых существ, то при благоприятных условиях их популяция будет увеличиваться бесконечно, т. е. разрастание флуктуаций (самоорганизация) ничем внутренне не ограничена. На самом деле внешние факторы, например недостаток энергии или пищи, такой процесс разрастания флуктуации ограничивают. Посмотрим, как это происходит на примере лис и зайцев.

Представим себе, что до определенного момента взаимодействия двух подсистем, т. е. популяций лис и зайцев, было сбалансированным: зайцев (с учетом их естественного пополнения) как раз хватало, чтобы прокормить определенное число лис. Затем в момент, принимаемый за нуль отсчета времени, в силу какой-то флуктуации число зайцев возросло. Это увеличит количество пищи для лис и, значит, их число. Возникает флуктуация лис. Обозначим приращение лис за единицу времени  $dt$  как  $dx$ . Очевидно, что такое приращение должно быть пропорциональным появившемуся в результате флуктуации в данный период времени числу зайцев. Пусть это число равно  $y$ . Учтем далее, что каждая лиса для поддержания своей жизни должна съесть не одного, а несколько зайцев. Это возможно учесть, вводя коэффициент пропорциональности  $K_1$ .

Тогда приращение популяции лис можно математически изобразить так:

$$dx = K_1 y dt, \quad \text{или} \quad \frac{dx}{dt} = K_1 y.$$

С другой стороны, популяция зайцев будет уменьшаться (лисы съедают зайцев) пропорционально числу лис в флуктуации ( $x$ ) и их прожорливости. Этот процесс можно записать так:

$$-dy = K_2 x dt, \quad \text{или} \quad \frac{dy}{dt} = -K_2 x.$$

Мы, таким образом, получаем зацепляющуюся систему уравнений:

$$\frac{dx}{dt} = K_1 y, \quad \frac{dy}{dt} = -K_2 x.$$

Дифференцируя их по времени, находим:

$$\frac{d^2x}{dt^2} = K_1 \frac{dy}{dt}, \quad \frac{d^2y}{dt^2} = -K_2 \frac{dx}{dt}.$$

Подставляя сюда исходные выражения для производных  $dx/dt$  и  $dy/dt$ , получаем:

$$\frac{d^2x}{dt^2} = -K_1 K_2 x, \quad \frac{d^2y}{dt^2} = -K_1 K_2 y.$$

Такие уравнения известны, и можно сразу записать их решения:

$$x = C_1 \sin \omega t + C_2 \cos \omega t,$$

$$y = C_3 \sin \omega t + C_4 \cos \omega t,$$

$$\omega = \sqrt{K_1 K_2}.$$

Чтобы найти постоянные  $C_n$ , примем, что при  $t = 0$  число лис не отклонилось от сбалансированного ( $x = 0$ ), а число зайцев увеличилось на  $y_0$ . Отсюда  $C_2 = 0$ ,  $C_4 = y_0$ . Чтобы

найти  $C_1$ , воспользуемся тем, что  $\frac{dx}{dt} = K_1 y$  и поэтому при  $t = 0$  должно быть  $\left(\frac{dx}{dt}\right)_0 = K_1 y_0$ . С другой стороны,  $\frac{dx}{dt} =$

$= \omega C_1 \cos \omega t$  и  $\left(\frac{dx}{dt}\right)_0 = \omega C_1$ . Значит,  $C_1 = \frac{K_1}{\omega} y_0$ . Остается

определить  $C_3$ . Для этого учтем, что  $\frac{dy}{dt} = -K_2 x$  и что  $x = 0$  при  $t = 0$ . Следовательно,  $\left(\frac{dy}{dt}\right)_0 = 0$ . С другой стороны,

$$\frac{dy}{dt} = \omega C_3 \cos \omega t - y_0 \omega \sin \omega t.$$

Таким образом, если  $\left(\frac{dy}{dt}\right)_0 = 0$ , то  $C_3 = 0$ . Итак,

$$x = \frac{K_1}{\omega} y_0 \sin \omega t = \sqrt{\frac{K_1}{K_2}} y_0 \sin \omega t, \quad y = y_0 \cos \omega t.$$

Значит, количество лис и зайцев будет меняться во времени периодически около некоторого среднего равновесного значения со сдвигом по фазе на  $\pi/2$ . Такие периодические процессы действительно нередко наблюдаются в живой природе.

Рассмотрим более сложный пример—так называемую модель Лотткэ—Вольтерра для сравнительно сложной экосистемы. В данной реализации модели учтены следующие объекты и связи между ними:

имеется популяция кроликов (Rabbits) численностью  $R$ ;  
 имеется популяция лис (Foxes) численностью  $F$ ;  
 имеется неисчерпаемый ресурс травы (Grass)  $G$  для питания кроликов.

Скорость роста популяции кроликов  $dR/dt$  без воздействия лис пропорциональна количеству травы  $G$  и численности популяции кроликов  $R$ :

$$\frac{dR}{dt} = K_{rb}GR.$$

Здесь  $K_{rb}$ —Coefficient of Rabbits Birth—коэффициент рождаемости кроликов. Популяция кроликов должна уменьшаться под воздействием лис. Скорость убыли кроликов пропорциональна количеству лис и численности популяции кроликов  $R$ . Этот процесс передается формулой

$$\frac{dR}{dt} = -K_{rd}FR.$$

Здесь  $K_{rd}$ —Coefficient of Rabbits Death—коэффициент смертности кроликов.

В целом скорость изменения численности кроликов составляет

$$\frac{dR}{dt} = K_{rb}GR - K_{rd}FR.$$

Скорость роста популяции лис  $dF/dt$  пропорциональна количеству кроликов  $R$  и численности популяции лис  $F$ :

$$\frac{dF}{dt} = K_{fb}RF.$$

Здесь  $K_{fb}$ —Coefficient of Foxes Birth—коэффициент рождаемости лис. Популяция лис не только растет, но и уменьшается под воздействием охотников. Скорость убыли лис пропорциональна численности популяции лис  $F$ :

$$\frac{dF}{dt} = -K_{fd}F.$$

Здесь  $K_{fd}$ —Coefficient of Foxes Death—коэффициент смертности лис.

В целом скорость изменения численности лис составляет:

$$\frac{dF}{dt} = K_{fb}RF - K_{fd}F.$$

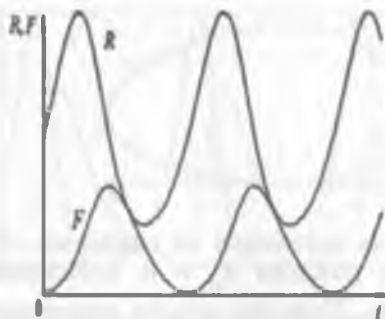
Система дифференциальных уравнений

$$\frac{dR}{dt} = K_{rb}GR - K_{rd}FR,$$

$$\frac{dF}{dt} = K_{fb}RF - K_{fd}F$$

представляет собой математическую модель экологической системы ресурсы—хищники—жертвы.

Рис. 10.7. Динамика популяций кроликов и лис при оптимальном запасе травы



Решение этой системы уравнений с учетом начальных условий при  $t = 0$   $R = R_0$ ,  $F = F_0$  позволяет прогнозировать динамику популяций в экологической системе.

Система уравнений модели Лоттка—Вольтерра отражает достаточно сложные связи между объектами экосистемы, поэтому трудно получить аналитическое решение уравнений и тем самым проанализировать особенности поведения системы при любом сочетании параметров модели. Значительно проще изучить поведение модели, выполнив ряд машинных экспериментов с программой.

Задавшись некоторыми характерными случаями параметризации модели, можно проанализировать эти частные случаи и на их основе получить представление об общих закономерностях поведения системы. В качестве первого примера рассмотрим результат решения при следующих значениях параметров:  $G = 5,0$ ,  $K_{rb} = K_{rd} = 0,10$ ,  $K_{fb} = 0,05$ ,  $K_{fd} = 0,50$ ,  $R_0 = 10,0$ ,  $F_0 = 3,0$  (см. рис. 10.7). Из полученных графиков  $R(t)$  и  $F(t)$  видно, что численности популяций в системе совершают периодические колебания. При этом численность лис отстает по фазе от численности кроликов. В этом смысле данный вариант аналогичен рассмотренному выше. Однако характер колебания сложный: колебания не являются гармоническими, не изображаются синусоидами. Эту особенность колебаний лучше наблюдать, рассматривая не графики  $R(t)$  и  $F(t)$ , а фазовую диаграмму  $F(R)$  (фазовой диаграммой называется график, построенный в данном случае в координатах  $F$  и  $R$ ). Видно, что точка  $F(R)$  периодически движется по кривой яйцеобразной формы, как показано на рис. 10.8. Такая форма диаграммы свидетельствует о негармоническом характере колебаний. В гармонических колебаниях точка двигалась бы по эллипсу.

Рассмотрим другие важные частные случаи, на материале которых можно будет яснее представить себе закономерности поведения экосистемы. Сначала рассмотрим случай, когда в системе вдруг исчезли кролики. В реальной экосистеме это может произойти в результате эпизоотии, быстро скошившей всю популяцию кроликов, либо вследствие гигантской миграции кроликов. Лисы остались без пищи, перестали размножаться, а



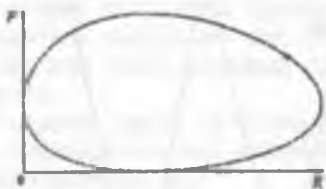


Рис. 10.8. Фазовая диаграмма процесса изменений популяций лис и кроликов

активность охотников не снизилась. В модели мы можем все это отразить, положив  $R_0 = 0$ . Остальные параметры модели можно не менять. Получим вполне ожидаемый результат, изображенный на рис. 10.9: популяция лис будет экспоненциально уменьшаться, стремясь к нулю. Происходит разрушение экосистемы. Этот результат можно получить и аналитически, поскольку в данном случае система уравнений упрощается.

Рассмотрим второй крайний случай. Пусть в системе исчезли лисы (причины: лес обработали с самолета пестицидом, оказавшимся для лис смертельным; грандиозный десант браконьеров). Возможно, что лисы отсутствовали в системе изначально, а завелись в небольшом количестве  $R_0$  кролики. Так было в Австралии, куда европейские переселенцы завезли кроликов. Положим  $R_0 = 3$ ,  $F_0 = 0$ . Получим вполне ожидаемый результат, изображенный на рис. 10.10,—популяция кроликов неудержимо растет по экспоненте. Обратим внимание на то, что в данном случае наша модель адекватна реальной действительности только в ограниченный период времени, пока верно предположение о неисчерпаемости травы.

Как только темп прироста травы перестанет поспевать за прожорливостью кроликов, в их популяции наступит голод и скорость прироста популяции упадет. Постоянное поедание травы, листьев может привести к гибели растений. Снова происходит разрушение экосистемы. Австралийские фермеры устроили кроликам беспощадный и непрерывный отстрел. Экологическая катастрофа была приостановлена.



Рис. 10.9. Кривая вымирания лис при отсутствии кроликов

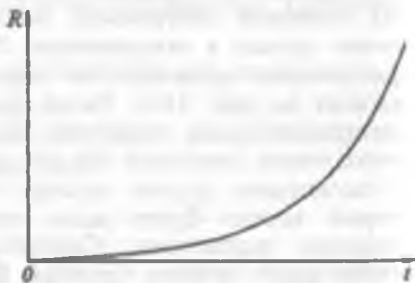
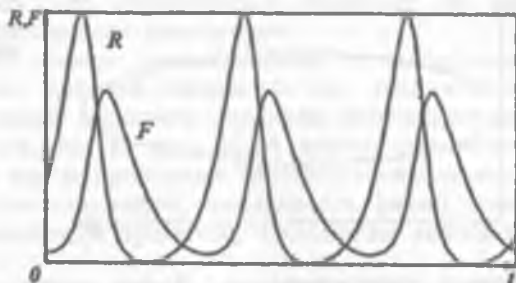


Рис. 10.10. Кривая популяции кроликов при отсутствии лис и при избытке травы

Рис. 10.11. Поведение популяций лис и кроликов при избытке травы



Рассмотрим более сложные случаи. Восстановим первоначальные значения всех параметров, но резко увеличим площадь, занятую травой. Пусть теперь  $G = 9$ . Это почти вдвое больше, чем в первом случае. Хорошо ли это? Можно подумать, что, увеличивая основной неисчерпаемый ресурс, мы придем к благоденствию в системе—больше будет интенсивно питающихся кроликов, значит, будет больше лис. Посмотрим, что будет на самом деле. Решая систему уравнений, получим картину, изображенную на рис. 10.11. Качественно она вполне похожа на первую картину, но количественные характеристики колебаний популяций стали иными. Увеличилась частота колебаний, возросли амплитуды изменений популяций как лис, так и кроликов. Обратим внимание на то, что численность популяции кроликов теперь падает до 2,1 а раньше падала только до 6 единиц. В характере роста и падения популяций явно проступают черты экспонент, отражающих катастрофические изменения популяций, рассмотренные нами в двух частных случаях.

Попробуем изменить площадь, занятую травой, в другую сторону, положим  $G = 2$ . Чего можно ожидать теперь? Ясно, что благоденствия не будет, но кто пострадает в большей степени? Машинный эксперимент приводит к картине, изображенной на рис. 10.12. Видно, что при меньшей мощности неисчерпаемого ресурса частота колебаний численности популяций уменьшилась. Но вот неожиданный результат—в проигрыше оказались не кролики, а лисы. Средняя численность их популяции снизилась. Из этого и предыдущего примера становится ясным, что прогноз отклика сложной экосистемы на внешние воздействия сделать очень трудно, если пользоваться

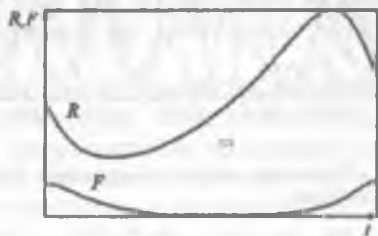


Рис. 10.12. Поведение популяций лис и кроликов при недостатке травы

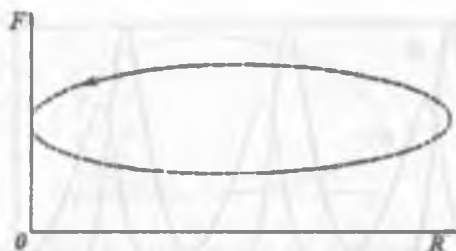


Рис. 10.13. Фазовая диаграмма популяций лис и кроликов от времени вблизи стационарного состояния

общими рассуждениями. Видна явная польза математических моделей и машинного метода прогнозирования поведения сложных многокомпонентных систем.

Исследуем еще один частный случай, называемый стационарным состоянием системы. Анализ системы уравнений показывает, что можно потребовать выполнения условий  $dR/dt = 0$  и  $dF/dt = 0$ , т.е. ни одна из популяций численно не изменяется со временем. Конечно, динамически состав популяций непрерывно обновляется, но лисы поедают ровно столько кроликов, сколько рождается новых, а охотники отстреливают не больше лис, чем их нарождается. Выполнить это условие при заданных коэффициентах представляется возможным только при особых начальных условиях. Соответствующие  $R_0$  и  $F_0$  можно найти из решения системы линейных уравнений, в которую вырождается система дифференциальных уравнений при  $dR/dt = 0$  и  $dF/dt = 0$ . У нас получится  $R_0 = 10$ ,  $F_0 = 5$ .

Попробуем теперь задать небольшое отклонение от стационарного состояния, например будем считать, что в результате внезапной флуктуации популяция кроликов увеличилась до  $R_0 = 11$ . Ясно, что в системе начнутся колебания. Фазовая диаграмма колебательного состояния системы показана на рис. 10.13. Видно, что это правильный эллипс, характерный для гармонических колебаний. Как раз данный случай рассмотрен в

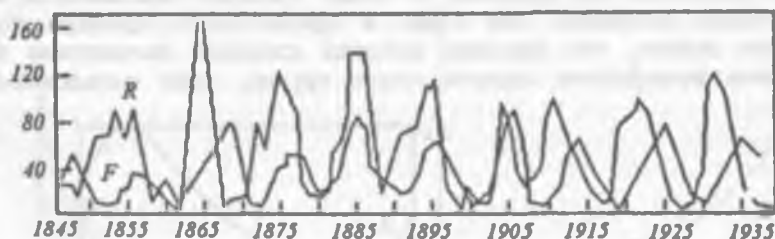


Рис. 10.14. Динамика популяций кроликов (R) и рысей (F) в Канаде по данным меховой компании

начале параграфа. Важно, что стационарное состояние не есть состояние покоя, а динамического равновесия.

В заключение имеет смысл познакомиться со столетними наблюдениями канадской меховой компании за численностью популяций кроликов и рысей. Конечно, реальная жизнь сложнее, чем рассмотренная модель, но из рис. 10.14 видно, что модель воспроизводит основные черты природных процессов—наблюдаются колебания численности популяций, численность рысей всегда отстает по фазе от численности кроликов, амплитуды колебаний взаимосвязаны.

Таким образом, в связанных системах возникает *саморегуляция* в развитии процессов, ограничивающая разрастание флуктуаций и приводящая при подходящих условиях к появлению определенных закономерностей в общем развитии событий.

## 10.6. Связь микроскопического и макроскопического или как «сосчитать» хаос

При изучении вопросов термодинамики и теории газов существенным является различие в средствах описания единичных и множественных объектов. Вводятся понятия о микроскопическом и макроскопическом подходах, а также о величинах, относящихся к целому множеству. Такими величинами, в частности, являются температура, объем и, что особенно важно, энтропия. В § 6.7 и этой главе было показано, что свойство энтропии возрастать при необратимых процессах в замкнутой системе отражает одну из важнейших особенностей окружающего нас мира: развиваться в определенном направлении—от прошлого к будущему, что и позволяет ввести понятие времени как характеристики, принимающей лишь положительные значения.

Несмотря на привлекательность макроподхода, нельзя забывать и о микроподходе, так как на этом уровне достигается наибольшая физическая ясность в понимании всего явления. В самом деле, пользуясь макроскопическими изменениями объема, давления и температуры, можно экспериментально установить, например, газовые законы и выяснить взаимосвязь между соответствующими величинами. Однако природа этой связи остается скрытой до тех пор, пока мы не обратимся к микрорассмотрению, говоря о газе как о совокупности хаотически движущихся частиц, рассматривая их совокупное действие на стенку сосуда и т. д.

Это же характерно и для исследований явлений переноса, макроскопический результат которых заключается в общем уравнивании числа частиц в двух смежных одинаковых объемах, если первоначально в одном из этих объемов частиц было больше (или выравнивании давлений), в выравнивании температур приведенных в контакт нагретых тел и др. Макро-

скопический результат и формулы для соответствующих коэффициентов переноса получаются на основе микрорассмотрения.

Может возникнуть достаточно естественный вопрос: нельзя ли и такую важнейшую функцию, как энтропия, также связать с микросостояниями множественных объектов? Может быть, и в этом случае мы приобретем некоторую ясность, а главное, сможем подсчитать эту величину не только для идеального газа. Ведь понятие энтропии применимо не только к нему.

При успешном решении сформулированной проблемы мы получили бы важнейший результат: нашли связь между микро- и макропараметрами всей системы.

Чтобы подойти к пониманию того, каким образом это можно сделать, остановимся на следующем примере.

Пусть имеется некоторый объем, разделенный, например, на 100 разных небольших частей (клеток). Будем считать, что в этом объеме могут находиться частицы, причем не более одной частицы в каждой клетке. Будем также считать, что имеется макроприбор, который может лишь измерить общее число частиц во всем объеме, но его чувствительности недостаточно, чтобы ответить на вопрос о том, в каких именно клетках объема имеются частицы, а в каких они в данный момент отсутствуют.

Пусть теперь в начальный момент в исследуемом объеме находится ровно 100 частиц. По условию эти частицы должны быть распределены равномерно по всем клеткам, т. е. в каждой по одной. Если провести измерение с помощью выбранного прибора, то он зафиксирует цифру 100.

Предположим теперь, что одна из частиц удалена из объема, т. е. их осталось 99. Ясно, что при этих условиях, как бы мы ни размещали частицы по клеткам, одна из этих клеток останется свободной. Назовем такую пустую клетку «дыркой». Снова проведем измерение выбранным прибором. На шкале его появится цифра 99. Определяет ли эта цифра полностью состояние объекта? Нет, конечно! В самом деле, легко видеть, что «дырка» может быть распределена по всем клеткам 100 способами. Каждый такой способ размещения будем называть микросостоянием. Все эти микросостояния будут различными. Если присутствие частицы в клетке обозначить 1, а отсутствие, т. е. наличие «дырки», пометить как 0, то все состояния могут быть описаны таблицей:

|       |   |   |   |     |
|-------|---|---|---|-----|
| 0     | 1 | 1 | 1 | ... |
| 1     | 0 | 1 | 1 | ... |
| 1     | 1 | 0 | 1 | ... |
| ..... |   |   |   |     |

Первая строка этой таблицы, в которой каждая позиция отвечает одной клетке, соответствует состоянию, когда «дырка» находится в первой клетке; вторая строка относится к состоянию, когда пустует вторая клетка, и т. д.

Если бы имелся очень чувствительный прибор, который мог бы следить за каждой частицей, то разные микросостояния отражались бы в разных показаниях такого микроприбора. Однако мы располагаем только макроприбором, который реагирует лишь на общее число частиц и не может отличить одно микросостояние от другого.

Будем называть такие различные микросостояния, которые приводят к одному и тому же макроизмерению, *вырожденными*, а число всех возможных микросостояний, отвечающих одному и тому же показанию макроприбора, — *степенью вырождения*.

Пусть теперь из системы удалено две частицы, или, что то же самое, в системе имеется две «дырки». Введем дополнительное условие, а именно, будем считать, что частицы, или «дырки», неразличимы, т.е. ничто вообще не изменится даже на микроуровне, если просто поменять в двух заполненных клетках местами частицы и «дырки». Снова будем размещать эти две «дырки» по всем возможным клеткам выбранного объема. Получится следующая таблица микросостояний:

|       |   |   |   |   |     |
|-------|---|---|---|---|-----|
| 0     | 0 | 1 | 1 | 1 | ... |
| 0     | 1 | 0 | 1 | 1 | ... |
| 1     | 0 | 0 | 1 | 1 | ... |
| ..... |   |   |   |   |     |

Число таких микросостояний будет равно  $100 \cdot 99/2 = 4950$ .

Поскольку макроприбор все эти состояния снова различить не может, то заключаем, что степень вырождения, которая при одной «дырке» равнялась 100, существенно возросла.

Для числа «дырок» больше двух общий подсчет степени вырождения можно сделать по формуле  $\frac{N!}{n!(N-n)!}$ , где  $N$  есть

общее число частиц и «дырок» (т.е. число клеток) и  $n$  — число «дырок». В математике такие комбинации объектов называются перестановочными с повторяющимися элементами.

Если число «дырок» нарастает, то число частиц падает. В соответствии с этим в приведенной выше формуле в знаменателе величина  $n!$  растет, а  $(N-n)!$  уменьшается. Есть два предельно минимальных значения степени вырождения: когда имеются только частицы и когда имеются только «дырки», т.е.  $n = N$ . Так как  $0! = 1$ , то и в том и в другом случае степень вырождения равна единице. Во всех промежуточных случаях она отлична от единицы и положительна и, следовательно, должна проходить через максимум. Нетрудно сообразить, что этот максимум отвечает точке  $n = N/2$ . В самом деле, в этом случае число «дырок» равно числу частиц, и процесс постепенного нарастания «дырок» после некоторой границы должен быть симметричным отражением процесса нарастания частиц, если бы мы начали не с удаления частиц, а, наоборот, с помещения частиц последовательно по одной в систему клеток.

Представим теперь, что из первоначального объема частицы уходят не просто в пространство, а в связанный с первой системой другой замкнутый объем. Разделим этот объем на такие же клетки, как и первый. Пусть число этих клеток равно  $k$ . Возможны ситуации, когда второй объем меньше первого, равен ему и больше его. Это означает, что  $k < N$ ,  $k = N$ ,  $k > N$ .

Введем теперь ограничения на возможное «перетекание» частиц из первого объема во второй. По аналогии с обычным давлением, которое внутри объема пропорционально общему числу частиц в объеме и обратно пропорционально величине этого объема, определим давление в изучаемых объемах как отношение числа частиц в системе к числу клеток, которые составляют эту систему. Для первой системы тогда получим  $(N - n)/N$  и для второй  $n/k$ , так как число «дырок» в первой системе равно числу частиц во второй. Как известно, в реальных газах при диффузии частиц из некоторого объема в другой, с ним связанный, в них постепенно устанавливается одинаковое давление, после чего макроскопический процесс прекращается. По аналогии с этим в нашем искусственном случае равновесное состояние устанавливается, когда  $\frac{N - n}{N} = \frac{n}{k}$ , откуда  $n_{\text{равн}} = \frac{kN}{k + N}$ .

Если, далее, обозначить  $k = \alpha N$ , где  $\alpha$  — некоторый коэффициент, который может быть и больше и меньше единицы, то  $n_{\text{равн}} = \frac{\alpha N}{\alpha + 1}$ . При переходе частиц из первого объема во второй будет не только происходить выравнивание давлений, но и меняться степень вырождения состояний. Если по-прежнему считать, что с помощью макроприбора можно определить только общее число частиц соответственно в первом и во втором объемах по отдельности, то степень вырождения для объединенной системы будет равна сумме таких степеней для двух объемов.

Предположим теперь, что  $\alpha = 1$  ( $k = N$ ), т. е. объемы одинаковы. В этом случае  $n_{\text{равн}} = N/2$ , т. е. частицы одинаковым образом заполняют два объема. Как было отмечено выше, при этих условиях степени вырождения как в первом, так и во втором объемах будут достигать максимума.

Пусть, далее,  $\alpha > 1$  ( $k > N$ ), тогда  $n_{\text{равн}} > N/2$  и в первом объеме степень вырождения начнет уменьшаться по сравнению с максимально возможной. Примем для определенности, что  $\alpha = 2$ . Чтобы не оперировать слишком большими числами, положим также, что  $N = 10$ . Тогда  $n_{\text{равн}} = 7$  и степень вырождения будет равна 120 (при максимально возможной 252). Во втором объеме при  $k = 20$  степень вырождения, если в объем перейдет из первого семь частиц, оказывается равной 77520.

Таким образом, когда частицы из малого объема переходят в большой, то хотя и возможно падение степени вырождения в малом объеме по сравнению с максимально возможной, однако это падение с лихвой компенсируется возрастанием степени вырождения во втором объеме.

Если начальное состояние таково, что все частицы находились в первой объеме, то суммарная степень вырождения должна быть равна двум (второй объем целиком заполнен «дырками»).

Таким образом, процесс выравнивания давлений приводит к возрастанию суммарной степени вырождения состояний системы. Разумеется, все сказанное относится и к тому случаю, когда первый объем оказывается больше, чем второй. Все рассуждения сохраняются, если поменять местами частицы и «дырки».

Вспомним теперь, как ведет себя в сходных случаях энтропия системы. Напомним, что при адиабатическом свободном расширении газа энтропия его возрастает по сравнению с исходным значением, когда газ есть только в одной части сосуда. Энтропия оказывается минимальной в том случае, когда все частицы системы находятся во вполне определенных положениях, подобно атомам в идеальном кристалле при температуре, близкой к абсолютному нулю.

Сразу бросается в глаза известная аналогия с поведением степени вырождения в рассмотренном выше примере.

Это наталкивает на мысль о том, не связана ли энтропия с числом возможных вырожденных состояний системы. Вспомнив формулу изменения энтропии идеального газа при его адиабатическом расширении, можно также предположить, что связь энтропии и числа макросостояний должна передаваться логарифмической функцией.

Все это, конечно, не более чем намек на возможный вид решения. Точное же доказательство было дано Л. Больцманом. Действительно, результат оказался очень простым:

$$S = k \ln G.$$

Здесь  $S$ —энтропия,  $k$ —постоянная Больцмана,  $G$ —степень вырождения макросостояния.

Это одна из самых важных в физике формул, устанавливающих связь между макроскопическим параметром—энтропией—и микроскопической картиной состояний этой системы. Независимо от этого, каков ее конкретный вид. Таким образом, энтропия приобретает очень широкое значение, а ее зависимость от  $G$  приводит к колоссальному числу важных следствий. Не случайно эта формула выгравирована на надгробном памятнике Л. Больцману.

Прежде чем перейти к обсуждению некоторых важных следствий формулы Больцмана, приведем один из ее возможных выводов для идеального газа. Это полезно сделать, так как традиционные способы введения понятия энтропии не опираются



«на наглядность», что и приводит ко всем известным затруднениям в понимании этой важнейшей величины.

Будем опираться на макроскопические и микроскопические экспериментальные факты.

Пусть имеется находящийся в равновесном состоянии идеальный газ в сосуде (или реальный газ, но для которого модель идеального достаточно хороша). Предположим, что мы измеряем давление газа с помощью инерционного манометра с большой ( $1 \text{ мм}^2$  или более) площадью мембраны. Тогда, очевидно, будет получаться все время постоянный отсчет (постоянная величина давления). При медленном изменении температуры газа или его объема (вдвигается или выдвигается поршень) давление будет меняться в соответствии с уравнением состояния идеального газа.

Воспользуемся теперь малоинерционным манометром с очень малой приемной площадкой. В этом случае, очевидно, постоянного показания уже не будет, а получится (если нарисовать график зависимости сигнала от времени) зубчатая кривая (стрелка прибора будет дрожать хаотическим образом около некоторого среднего положения). Первый эксперимент отвечает макроизмерению и макросостоянию, а второй—микроизмерениям и микросостояниям.

Меняя объем и температуру газа (последняя измеряется также макроприбором), заметим, что амплитуда пиков и их ширина и расстояния между отдельными пиками будут меняться при изменении как объема, так и температуры. Значит, между какой-то характеристикой (но обобщенной, так как речь идет не об отдельных пиках, а о всей картине их распределения) микросостояний и макропараметрами должна быть связь. Ее и надо установить!

Понять, почему показания микроприбора будут «дрожать», легко. В самом деле, обычное утверждение о том, что в любом достаточно большом «кубике», выделенном внутри общего объема газа, будет все время одно и то же число частиц, верно лишь в среднем по времени. На самом деле из-за хаотических движений молекул число это все время будет хаотически меняться около некоторого среднего. Это изменение при большом объеме «кубика» невозможно заметить, но если этот «кубик» уменьшать и пользоваться безынерционным прибором, то эффект станет заметным (присутствует он, разумеется, всегда). Такое «дрожание» числа частиц в малом объеме и называется флуктуациями. Хаотическое движение частиц газа подчиняется вероятностным законам, что и было использовано Эйнштейном для создания теории броуновского движения частиц. Воспользуемся этим.

Амплитуда и распределение флуктуаций во времени должны быть, очевидно, связаны с вероятностью для частиц быть в определенный момент в данной области пространства внутри общего объема, занимаемого газом. Если выделить малый объем, непосредственно прилегающий к микроприбору, то такая вероятность для одной частицы, очевидно, будет равна  $\omega(V) = V_0/V$ .

где  $V_0$  — рассматриваемый микрообъем и  $V$  — общий (макро) объем газа. Характер флуктуаций будет, очевидно, связан с тем, попадет ли в данный интервал времени частица внутрь выбранного микрообъема или нет. При изменении макрообъема характер флуктуаций будет меняться и можно для описания этого использовать отношение вероятностей попадания микрочастицы в микрообъем при двух разных значениях макрообъема, т. е. величиной

$$\frac{w_1(V_1)}{w_2(V_2)} = \frac{V_2}{V_1}.$$

Далее заметим, что поведение флуктуаций целесообразно описывать разностью значений некоторой величины, с общим характером этих флуктуаций связанной. Обозначим такую, пока еще не известную величину (функцию) символом  $S$ , а ее изменение  $\Delta S = S_2 - S_1$ .

Перейти от отношения двух величин к их разностям можно, применив операцию логарифмирования.

В результате получим:  $\Delta S = C_1 \ln(V_2/V_1) = C_1 \ln V_2 - C_1 \ln V_1$ , где  $C_1$  — некоторая пока неизвестная константа.

Теперь учтем факторы, связанные с температурой. Для этого примем во внимание, что характер пика на кривой, в частности его ширина, зависит от времени существования флуктуации. Это время, в свою очередь, должно быть связано со скоростью движения частиц. В самом деле, если представить себе, что в исследуемом микрообъеме оказалось существенно повышенное по сравнению со средним число частиц (положительная флуктуация), то достаточно очевидно, что она исчезнет (частицы рассеются в окружающее пространство) тем быстрее, чем больше скорость диффузии. Последняя же определяется температурой газа и пропорциональна  $T^{1/2}$ .

Если на основании этих рассуждений использовать для характеристики флуктуаций величину отношения их средней длительности при двух разных температурах газа, то мы получим величину пропорциональную  $(T_2/T_1)^{1/2}$ , причем она не будет зависеть от занимаемого газом объема.

Здесь напрашивается известная аналогия с рассмотренным выше распределением частиц по ячейкам. Пусть имеется всего одна частица, которая может быть случайным образом помещена в одну из 100 ячеек. Пусть микроприбор реагирует на присутствие частицы в ячейке  $N_1$ . Пусть теперь каждое состояние, когда частица находится в одной из ячеек, существует 1 с. Тогда стрелка микроприбора при большом числе испытаний будет в среднем отклоняться раз в 100 с. Интервалы между двумя наблюдаемыми флуктуациями будут определяться степенью вырождения или общим размером пространства возможного нахождения частицы.

Если увеличить время «выдержки» каждого состояния до 2 с, то средний интервал между флуктуациями также возрастет вдвое, но отношение интервалов для двух длительностей «выдержек» не изменится и по-прежнему будет зависеть лишь от объема. Именно поэтому можно такое отношение использовать как не зависящую от времени характеристику флуктуаций.

В свою очередь, отношение длительностей флуктуаций не будет зависеть от пространства, в котором может существовать частица.

Вернемся к идеальному газу. Поскольку характер флуктуаций, как показывает эксперимент, является функцией двух независимых величин—объема и температуры, то окончательно примем приращение искомой функции, связанной с флуктуациями, в виде

$$\Delta S = C_1 \ln \left( \frac{V_2}{V_1} \right) + \frac{1}{2} C_2 \ln \left( \frac{T_2}{T_1} \right).$$

Здесь  $C_2$ —также неизвестная константа. Остается определить константы  $C_1$  и  $C_2$ . Для этого перейдем к бесконечно малым изменениям  $S$ , т. е. к ее дифференциалам. Легко видеть, что

$$dS = C_1 \frac{dV}{V} + \frac{1}{2} C_2 \frac{dT}{T}.$$

Воспользуемся теперь уравнением состояния идеального газа и приведем правую часть выражения для  $dS$  к виду, содержащему в знаменателях слагаемых только температуры. Получим:

$$dS = C_1 p \frac{dV}{RT} + \frac{1}{2} C_2 \frac{dT}{T}.$$

Перепишем это выражение в форме

$$T dS = \frac{1}{R} C_1 p dV + \frac{1}{2} C_2 dT$$

и сравним с выражением для первого начала термодинамики для идеального газа:

$$\Delta Q = p dV + C_V dT.$$

Видно, что эти выражения совпадут, если принять  $C_1 = R$  и  $C_2 = 2C_V$ .

Полученная таким образом функция, дифференциал которой

$$\begin{aligned} dS &= R \frac{dV}{V} + C_V \frac{dT}{T} = \frac{\partial S}{\partial V} dV + \frac{\partial S}{\partial T} dT = \\ &= \frac{\partial}{\partial V} \ln (T^C V^R) dV + \frac{\partial}{\partial T} \ln (T^C V^R) dT, \end{aligned}$$

и есть энтропия. Видно, во-первых, что с ее помощью можно выразить первое начало термодинамики, приняв  $\Delta Q = T dS$ , а, во-вторых, что она характеризует микросостояния газа, хотя и зависит от макропараметров.

Пользуясь выражением для дифференциала энтропии, саму функцию запишем в виде

$$S = \ln(T^C V^R) + \text{const.}$$

Учитывая теперь, что  $R = kN$ , где  $k$ —постоянная Больцмана, и что для идеального газа  $C_V = \frac{3}{2}R$ , получим:

$$S = k \ln(T^{3N/2} V^N) + \text{const.}$$

Сравнивая эту формулу с формулой  $S = k \ln G$ , найдем, что произведение  $G = T^{3N/2} V^N$  и есть мера степени вырождения макросостояния идеального газа.

По теореме Карно изменение энтропии обратимого цикла Карно равно нулю в любом случае и не зависит от рабочего тела. Именно поэтому формула  $S = k \ln G$  годится для любых систем. Если  $G = 1$  (имеется только одно макросостояние, совпадающее с неизменным микросостоянием), то  $S$  должна быть равна нулю. Значит,  $\text{const} = 0$ .

Теперь еще раз вернемся к вопросу: что понимать под микроскопическим состоянием даже для простейшей реальной физической системы—газа в сосуде? Если рассмотреть установившееся состояние газа, то в течение времени его макроскопические параметры (объем, давление, температура) меняться не будут. Однако, как отмечалось, частицы газа не находятся в покое. Они все время движутся, перелетая из одной части объема газа в другую. В результате таких хаотических движений не только возникает устойчивое воздействие частиц газа на стенки сосуда (постоянное давление), но должны существовать и флуктуации, т. е. различия в числе частиц в двух соседних объемах в любой очень короткий промежуток времени. Представим себе, что мы могли бы фотографировать толщу газа с очень большим разрешением так, чтобы на фотопластинке можно было зарегистрировать отдельные маленькие участки объема и чтобы почернения на фотопластинке, отвечающие отдельным участкам, были пропорциональны числу частиц в одной малой области объема. Пусть фотографирование совершается за очень короткий промежуток времени. Тогда мы увидим на одной пластинке неравномерное почернение, отражающее тот факт, что в газе имеются микрообъемы с повышенным (по отношению к среднему) и пониженным содержанием частиц. Такое мгновенное распределение частиц и отвечает одному возможному микросостоянию. Если наложить много полученных таким образом фотографий друг на друга и просветить их, обнаружится равномерный фон. Именно эта ситуация и соответствует макроскопическому эксперименту, при котором флуктуации не проявляются, хотя они реально и существуют.

Возрастание энтропии означает увеличение амплитуды и разнообразия этих флуктуаций, в частности при повышении температуры либо при проникновении газа в какие-то новые

области, где его ранее не было (диффузия или медленное расширение объема).

Степень вырождения будет определяться числом подобных флуктуаций, которые могут возникнуть за единицу времени. Если в результате какого-нибудь процесса в определенном месте пространства системы начинается разрастание некоторой флуктуации так, что эффект уже может быть зарегистрирован макрореприбором, то вырождение частично снимается и степень его убывает, что, в свою очередь, должно привести к уменьшению энтропии. Наличие большого числа хаотических флуктуаций называют *беспорядком* или *хаосом*, а проявление разрастающихся флуктуаций или уменьшение числа возможных распределенных флуктуаций — *переходом к порядку*.

Тот закон, о котором мы говорили выше и согласно которому необратимые процессы в замкнутых системах происходят в направлении возрастания энтропии, означает, что развитие происходит от порядка к беспорядку. При достижении максимального беспорядка достигается устойчивое состояние. Очевидно, при всех возможных направлениях развития событий это состояние оказывается наиболее вероятным. Поэтому можно также сказать, что процессы в сложных системах развиваются от состояний менее вероятных к состояниям более вероятным.

Описанные выше процессы перехода системы частиц к более вероятному состоянию можно имитировать на компьютере с помощью простой модели, которая получила название «демон Больцмана». В этой модели также рассматривается перенос частиц из клеток одного объема в клетки другого, но допускается, что не только частицы из первого объема могут переходить во второй, но и наоборот. Другими словами, модель учитывает то реальное обстоятельство, что частицы газа без ограничений могут двигаться как вправо, так и влево. Все частицы располагаются на плоском прямоугольном поле, разделенном на клетки, которые образуют столбцы и строки.

Как и ранее, в каждую клетку будем помещать не более одной частицы. Однако, в отличие от рассматриваемого выше случая, будем считать, что имеются два сорта частиц, отличающихся друг от друга энергией (запасом классической кинетической энергии или находящихся на разных квантовых уровнях — основном и возбужденном) или массой. Обозначим частицы одного сорта символом + (плюс), а другого символом — (минус). Таким образом, каждой клетке поля может быть приписано три состояния: 0 (дырка), + и —.

В качестве эксперимента рассмотрим процесс диффузии частиц одного сорта (помеченных знаком +) из области, где они все первоначально сосредоточены, в весь объем. Для этого все поле разделим на 8 колонок и разместим все частицы в первой колонке. Для упрощения расчетов число частиц принято небольшим. В процессе диффузии частицы будут переходить во все остальные колонки, постепенно равномерно распределяясь по

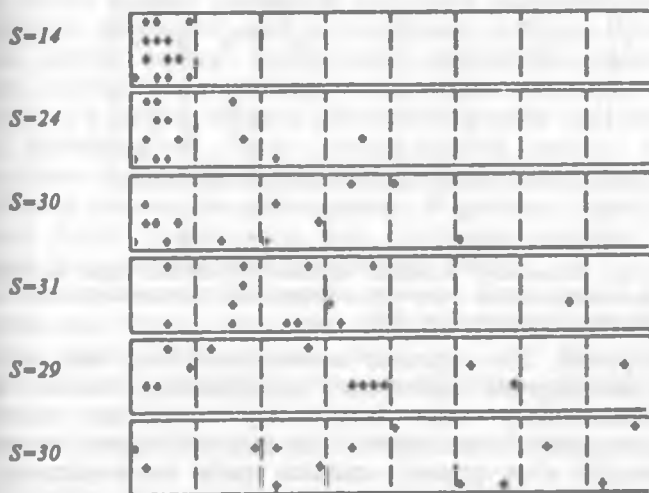


Рис. 10.15. «Демон Больцмана» случайно меняет положение молекул в пределах длины свободного пробега. С экрана ЭВМ снято несколько последовательных кадров. Видно, что в клетках поля число молекул флуктуирует, но с течением времени молекулы распределяются по полю более равномерно. Система в результате диффузии направленно переходит от неравновесного состояния к более равновесному. Энтропия при этом возрастает, а затем флуктуирует. На экране ЭВМ изображены только молекулы примеси (+), молекулы основного газа невидимы

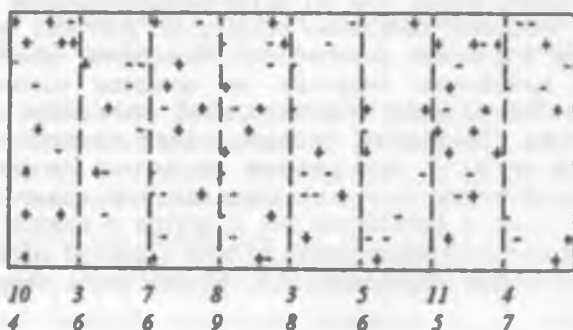


Рис. 10.16 Случайное распределение молекул двух сортов (+ и -) по рабочему полю. Состояние системы близко к равновесному. Энтропия близка к максимальной и равна  $S = 304$ . Внизу под колонками показано число частиц + и -

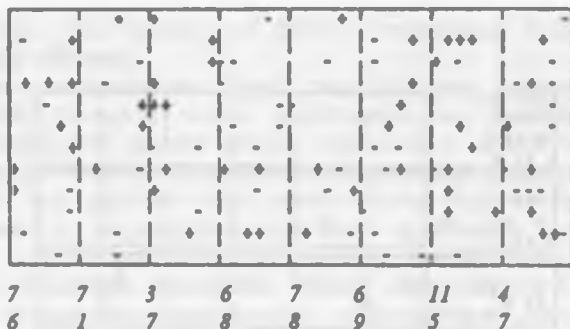


Рис. 10.17. «Демон Больцмана» поменял местами случайные пары молекул. Наблюдаются флуктуации числа частиц в клетках поля, но состояние системы остается равновесным. Энтропия  $S = 303$ .

всему пространству. Этот процесс имитируется тем, что «демон Больцмана» многократно производит перестановку частиц или «дырок», заполняющих пары по-разному выбранных клеток. Притом эта операция выполняется, если в одной клетке из пары имеется частица, а в другой «дырка» либо плюс-частица и минус-частица. Если клетки заполнены одинаковыми объектами, то перестановка не производится (физическое состояние при таком обмене не меняется и поэтому его не надо учитывать). Для того чтобы получить картину, весьма близкую к реальности, введем еще длину свободного пробега и учтем, что перестановка частиц или дырок между клетками возможна на расстоянии не больше длины свободного пробега. Обмен дырками и частицами между клетками определяется датчиком случайных чисел в полном соответствии с тем, что процесс перемещения реальных частиц газа в пространстве есть процесс случайный. На рис. 10.15 показаны расположения частиц по колонкам, найденные через равные промежутки времени, и указаны значения энтропии. Хорошо видно, что по мере распространения частиц в пространстве энтропия постоянно возрастает, достигая некоторого максимума (в состоянии равновесия). Интересно заметить, что около этого максимума энтропия не остается постоянной, а флуктуирует. Это связано с флуктуацией количества частиц в каждой колонке. Специально данный случай продемонстрирован на рис. 10.16 и 10.17, где указана не только энтропия всей системы (подсчитанная в данном варианте как сумма энтропий частиц  $+$  и  $-$ ), но и количество тех и других в каждой колонке. Такие флуктуации энтропии уменьшаются с ростом общего числа частиц и постепенно становятся практически незаметными.

## 10.7. Энтропия и информация

Связь энтропии с вероятностью тех или иных процессов имеет еще более глубокий смысл и позволяет использовать понятие энтропии в области, казалось бы, от физики вообще

далекой. Речь идет об информации. Для того чтобы последующие рассуждения стали понятными, рассмотрим ряд простых примеров. Пусть заданы уравнение движения материальной точки под действием некоторой силы и начальные условия. Известно, что в этом случае можно совершенно однозначно предсказать, где будет находиться частица и с какой скоростью она будет двигаться в любой наперед указанный момент времени. Приписывая материальной точке определенную массу, характеризуя начальное положение и скорость, а также действующую силу, мы сообщаем некоторую информацию. В данном случае эта информация будет полной, так как состояние частицы в будущем может быть предсказано совершенно точно.

Рассмотрим теперь другой пример. Представим себе, что мы производим серию измерений какой-нибудь величины, например массы тела. Если масса мала, то на результат каждого отдельного измерения станут заметно влиять случайные факторы: движение воздуха около чашки весов, особенности помещения грузов сравнения на другую чашку, колебания коромысла и др. В результате ни одно из таких измерений не даст значения истинной массы (либо лишь одно из измерений случайно с ним совпадет, но исследователь об этом подозревать не будет). Если действие всех мешающих факторов будет случайным, то при большом числе измерений приблизительно равное количество их даст значения массы как выше истинной, так и ниже. Чем ближе к истинному будут лежать отсчеты конкретных измерений, тем большее их количество будет оказываться в некоторой небольшой области около истинного значения. Как известно, при таких условиях истинному значению массы будет отвечать среднее арифметическое значений всех одиночных измерений. Можно сказать, что каждое измерение несет информацию об истинной массе объекта, причем информацию неполную, так как по одному измерению установить значение истинной массы нельзя. Информация возрастает, если число измерений увеличивается.

Оказывается, что для характеристики информации также можно ввести понятие энтропии. Делается это следующим образом. Пусть некоторая величина  $X$  в результате каких-то измерений или действий принимает значения  $X_1, X_2, X_3, \dots, X_k$  с вероятностью соответственно  $p_1, p_2, p_3, \dots, p_k$ . В свою очередь, вероятностью значения  $X_1$  называется отношение числа измерений, при которых получено значение  $X_1$ , к общему числу измерений. При этом имеется в виду, что общее число измерений очень велико. Например, вероятность события выпадения герба или цифры при многократном бросании монеты равна  $1/2$ . Вероятность выпадения цифры 1 при бросании игральной кости равна  $1/6$  и т. д.



В теории информации вводится величина  $H(x) = - \sum_n p_n \log_2 (p_n^{-1})$ , которая называется *энтропией случайной величины*.

Величина энтропии указывает среднее число двоичных знаков, необходимых для различения (или записи) допустимых значений случайной величины  $X$ . При использовании двоичного кода (именно поэтому в выражении для энтропии  $\log$  берется по основанию 2) в одном разряде цифры принимают значения лишь 0 и 1. Поэтому, например, изображение числа 2 в двоичном коде имеет вид 10 (цифре 2 отвечает единица во втором разряде), цифра 3 записывается как 11 и т. д.

Вернемся теперь к рассмотренному выше примеру о размещении частиц и «дырок» в системе. Представим себе, что с помощью прибора, способного отличить одно макросостояние от другого, производится большое число измерений. Тогда сколько бы мы ни проводили измерений в том случае, когда все клетки заняты частицами или «дырками», будет обнаруживаться одно и то же состояние. Значит, вероятность его появления будет равна единице. Энтропия информации должна быть, очевидно, равной нулю. Если имеется одна «дырка», то система способна принимать 100 различных микросостояний. Поскольку вероятность обнаружить каждое из этих состояний одинакова и равна  $1/100$ , то для энтропии информации получим

$$100 \cdot 100^{-1} \log_2 100 = 6,65.$$

Таким образом, энтропия информации возрастает.

Бросается в глаза аналогия с обычной (физической) энтропией. Более детальное рассмотрение показывает, что действительно физическая и информационная энтропии фактически эквивалентны и отличаются друг от друга лишь на постоянный множитель. Это, в свою очередь, приводит к утверждению, что за полученную информацию приходится платить увеличением энтропии, а так как новая информация реально добывается с помощью измерений, то любое измерение связано с возрастанием энтропии окружающей среды. Эквивалентность физической и информационной энтропий не должна особенно удивлять, если вспомнить, что под знаком логарифма в формуле Больцмана стоит степень вырождения, которая непосредственно не связана с физическими особенностями объекта. Ясно, что у двух совершенно разных по своему физическому содержанию систем энтропия может быть одинаковой, если у них совпадают числа возможных микросостояний, отвечающих какому-то одному макропараметру. Именно поэтому понятием энтропии и можно пользоваться в самых разнообразных областях.

В свете всего сказанного закон возрастания энтропии в замкнутой системе можно трактовать так: *события самопроизвольно развиваются в сторону увеличения информации, а так как последняя связана с разнообразием, то в сторону увеличения разнообразия. Чем больший при этом получается*

выигрыш в разнообразии, тем более вероятным (выгодным) является процесс.

С этой точки зрения легко понять, почему высокоорганизованные организмы являются двуполыми. Причина в следующем. При каждом акте полового размножения создается новая информация. Встреча пары разнополых объектов есть дело случайности. Родившийся в результате такой встречи новый объект не тождествен ни одному из объектов родителей. В результате уровень разнообразия должен будет увеличиваться, а это как раз и соответствует требованию закона нарастания энтропии.

Если теперь рассмотреть некоторое человеческое общество как часть системы, связанной с окружающей природной средой, а всю Землю—как замкнутую систему, то нетрудно заключить, что возрастающая самоорганизация этого общества с неизбежностью должна привести к возрастанию беспорядка в окружающей среде, что, в частности, и выражается в появлении громадного числа рассеянных по Земле свалок. Таким образом, закон возрастания энтропии оказывается применимым и к исследованию экологических проблем.

Вернемся еще раз к вопросу о роли смерти в круговороте вещества и энергии в природе.

Рассмотрим простой пример двух связанных систем. Пусть из первой вещество переносится во вторую. Снова ограничимся частицами и «дырками», однако теперь будем считать, что во второй системе «дырки» могут располагаться по клеткам не всевозможными способами, а последовательно заполняя клетки во вполне определенном порядке: сначала занимается первая клетка, потом вторая и т. д.

Это создает некоторую аналогию с процессом роста растения, где все развивается детерминированно.

Если принять во внимание отмеченную выше связь энтропии со степенью вырождения, то получим, что, несмотря на перенос частиц из первой системы во вторую, энтропия последней остается равной нулю. Энтропия же первой системы будет возрастать до тех пор, пока число частиц и «дырок» не уравнивается. Потом энтропия должна падать. Так как общая энтропия отвечает сумме двух подсистем, а в одной она продолжает оставаться нулевой, то должна падать и общая энтропия. Это, однако, запрещено глобальным законом возрастания энтропии в замкнутых средах, если в последних процессы совершаются самопроизвольно. Значит, начиная с некоторого момента, процесс переноса частиц из первой подсистемы во вторую должен прекратиться. Этим устанавливается еще один предел возможного нарастания биомассы при росте растений за счет потребления вещества из почвы. Другими словами, в принципе нельзя добиться предельного разрастания растительной массы, полностью истощая почву. Чтобы процесс продолжался достаточно долго, необходимо все время поддерживать энтропию в первой подсистеме на достаточно высоком уровне, возвращая

вещество из второй подсистемы. Это возможно, в свою очередь, только при разрушении строгого порядка во второй подсистеме, что и получается при умирании организма. Если и в процессе жизнедеятельности организм выделяет вещество во внешнюю среду, то наиболее выгодно с точки зрения поддержания высокого значения энтропии во внешней среде производить это так, чтобы во внешнюю среду вещества возвращались в мало организованном виде. Эти закономерности действительно характерны для жизнедеятельности живых организмов.

Таким образом, общие законы, сформулированные в физике, позволяют разобраться в фундаментальных природных процессах. Какова же вообще роль физики в биологии?

Вспомним, что мы получили правильный результат о поведении сложной системы, применив к соударяющимся шарам законы классической механики. Если мы хотим понять, как происходит самоорганизация даже простейших живых образований, надо раскрыть правила взаимодействия отдельных составляющих. А это уже дело физики, которая сейчас в состоянии объяснить и предсказать, например, ход химических реакций, в которых участвуют крупные, так называемые биологически активные, молекулы. Геометрические формы этих молекул, вплоть до белков, изучаются в основном чисто физическими методами.

В физике употребляется громадное число разного рода моделей, поведение которых хорошо изучено. Многие процессы в живом организме по своему характеру оказываются близкими к поведению этих моделей. Тогда эти модели можно с успехом использовать и при описании отдельных проявлений в живом. Так, например, с успехом применяется для описания передачи сигнала по цепи нейронов модель распространения своеобразного волнового пакета—солитона. Мы поняли даже, почему необходима смерть. Можно привести еще много примеров. Вообще, можно сказать, что физические знания и физический способ мышления оказались при изучении живой природы чрезвычайно плодотворны. Не случайно целый ряд крупнейших ученых—лауреатов Нобелевской премии, заложивших основы современной биологии в прошлом, являлись физиками.

Чтобы продемонстрировать некоторые возможности применения физического подхода к решению задач молекулярной биологии с помощью компьютерного эксперимента, проанализируем, каким образом может происходить образование фиксированной формы белковой молекулы за счет большого числа внутримолекулярных движений.

Эта проблема является одной из центральных в естествознании, давно привлекавшей внимание исследователей. При решении ее впервые на фундаментальном молекулярном уровне устанавливается связь между неживой и живой природой и ликвидируется разрыв между физикой как традиционной наукой о неживом и биологией—наукой о живых, развивающихся

организмах. Обнаруживается и решающая роль «нетермодинамически» специфически близкодействующих факторов.

## 10.8. Формирование гибкой молекулярной цепи и механизм образования пространственных структур белковых молекул

Многие замечательные свойства длинноцепочных (полимерных) молекул объясняются почти свободным вращением звеньев вокруг одинарных связей. Например, эластичность таких органических веществ, как каучук, определяется вхождением в их молекулы жестких звеньев, соединенных друг с другом одинаковыми связями  $C-C$ , вокруг которых жесткие звенья могут свободно поворачиваться на любой угол. Вторичная структура белковой цепи также определяется ее временной эластичностью, т.е. возможностью скручиваться и формировать витки спирали, которые затем жестко скрепляются водородными и иными связями.

Эластичность длинноцепочных молекул со свободным вращением их звеньев вокруг одинарных связей имеет энтропийную природу. Поясним это положением на очень простой модели. Пусть молекула представляет собой цепь, состоящую из одинаковых жестких стержней, сочлененных своими концами. Угол  $\alpha$ , образованный двумя соседними звеньями, может быть произвольным в пределах от  $0$  до  $180^\circ$ . Торсионный угол (угол относительного поворота групп вокруг общей оси)  $\beta$ , образуемый звеньями с номерами  $n$  и  $n + 2$ , может быть произвольным в пределах от  $0$  до  $360^\circ$ . Иными словами, новое звено, присоединяясь к цепи, может соориентироваться в любом направлении. Если цепь построена из большого числа ( $N$ ) звеньев, то, как можно показать, длина ее стремится к некоторой фиксированной, равной  $L = a\sqrt{N}$ , где  $a$ —длина одного звена цепи.

Это является следствием закручивания цепи в беспорядочный клубок из-за случайного выбора углов  $\alpha$  и  $\beta$  при добавлении каждого нового звена или при тепловых флуктуациях готовой цепи. Свернутой цепи соответствует максимально возможное разнообразие значений углов  $\alpha$  и  $\beta$  во всем ансамбле звеньев, а следовательно, и максимальная энтропия  $S$  цепи.

Если подействовать на концы цепи, закрученной в клубок, и попытаться ее растянуть, то углы  $\alpha$  и  $\beta$  во всем ансамбле звеньев будут стремиться к нулю либо (только для  $\beta$ ) к  $180^\circ$ . Разнообразие значений углов  $\alpha$  и  $\beta$  уменьшится, при этом уменьшится и энтропия  $S$  цепи. Переход от вытянутой цепи к клубку отвечает переходу от менее вероятного состояния к более вероятному. Следовательно, если растянуть цепь за счет действия внешних сил, а потом предоставить цепь самой себе, то она за счет случайных хаотических (тепловых) взаимодействий с молекулами окружающей среды самопроизвольно вернется к неупоря-

доченному состоянию. Как известно, именно эффект сильного упругого растяжения нитей из некоторых материалов с возможностью возвращения нити к исходному состоянию после снятия внешней нагрузки (такое свойство называется эластичностью) характерен для каучукоподобных полимеров. Таким образом, эластичность привычной нам резины в значительной степени объясняется энтропийным фактором.

Выводы, полученные для этой простейшей модели, не изменяются при ее усложнении. Важно только, чтобы каждое звено цепи имело хотя бы по одной степени свободы с нулевым барьером внутреннего вращения. Так, в молекулярной цепи типа полиэтилена жестко определены плоские углы между звеньями, но торсионные углы вращения вокруг одиночных  $C-C$ -связей могут быть какие угодно, поэтому такая молекулярная цепь должна обладать эластичностью. Однако эластичность цепи может резко снизиться, если ее звенья химически связываются не только друг с другом, но и с удаленными звеньями при их случайном сближении в скрученной цепи или с соседними цепями в веществе. Получаются циклы большей или меньшей протяженностью, а в циклах возможности торсионных поворотов заметно беднее, чем в линейных цепях. Именно поэтому в каучук специально вводят серу и выдерживают его некоторое время при высокой температуре, чтобы атомы серы создали мостики между изопреновыми звеньями и тем самым уменьшили естественную эластичность каучука. Получается вулканизированная резина.

В случае цепи типа полиэтилена или вулканизированной резины внутрицепные и междецепные химические мостики возникают случайным образом. В отличие от них в биополимерах процесс связывания удаленных звеньев и образования устойчивых петель строго детерминирован. Например, при синтезе белковой молекулы фермент считывает с цепи мРНК программу следования аминокислотных остатков в белке и собирает в нужном порядке звенья белковой цепи. В результате возникает определенная первичная структура белка. Только что синтезированный участок белковой цепи является гибким и в результате теплового движения закручивается в случайные петли. Лишь некоторые звенья белковой цепи имеют возможность образовывать друг с другом водородную связь. Какие именно—задано в первичной структуре цепи. Если в процессе флуктуации на сравнительно коротком участке вновь синтезируемой белковой цепи нужные звенья оказываются рядом, то они электрически «узнают» друг друга и образуют водородную связь. После этого эластичность данного участка цепи пропадает, и данный элемент вторичной структуры белка сохраняется в виде устойчивой петли. Так постепенно образуется детерминированная вторичная структура белковой цепи. Мы видим, что малоэластичная белковая молекула обязательно проходит при своем рождении стадию эластичности. Это дает ей возможность с помощью хаотического

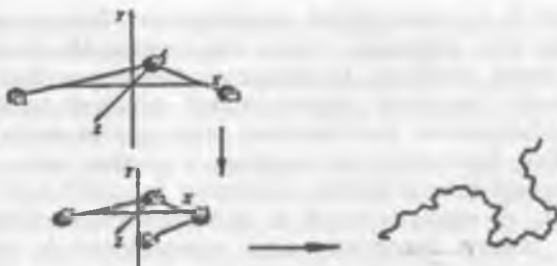


Рис. 10.18. Первые шаги построения на ЭВМ молекулярной цепи и готовая цепь случайной формы. Специфические взаимодействия между звеньями не учтены, циклов нет

теплового движения реализовать строго определенную структурную программу.

Этот процесс также можно смоделировать на компьютере, чтобы проследить за протеканием отмеченных физических явлений и оценить роль флуктуаций в процессе химического превращения гибкой молекулярной цепи в сравнительно жесткое полициклическое образование. Результаты моделирования показаны на рис. 10.18 и 10.19. На первом демонстрируется последовательность состояний синтезируемой цепи без учета тепловых конформационных флуктуаций. Получается линейная цепь, свернутая в рыхлый клубок. На втором видно, как флуктуации формы цепи позволяют ей образовать вторичную химическую структуру в виде жесткой полициклической молекулы.

Ясно, что энтропия цепи во втором случае меньше, но это уменьшение связано с увеличением энтропии окружающей среды за счет выделения тепла, что позволяет таким процессам протекать самопроизвольно.

Рассмотрим теперь более детально механизм формирования трехмерных структур белковых молекул. Понимание движущих сил этого механизма и создание теории структурной самоорганизации белков имеют основополагающее значение для решения последующих задач проблемы белка—его структурной,

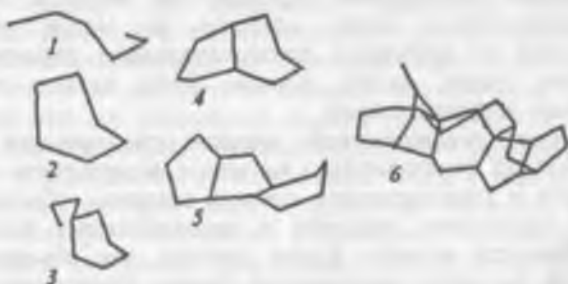


Рис. 10.19. Полициклические структуры, возникающие с ростом цепи при наличии специфических взаимодействий между несоседними участками молекулярной цепи

функциональной и эволюционной организации. Это также чрезвычайно важно для создания основ теоретической биологии и решения множества проблем, касающихся биосистем более высокого ранга. Белки являются элементарной основой всех живых существ и неперенными участниками всех протекающих в них процессов. Любая биологическая система и любая часть каждой системы—это произведение белков. Поэтому без глубокого знания их структурной, функциональной и эволюционной организации нельзя понять общие закономерности морфологии и эволюции более сложных организмов, точно так же, как нельзя объяснить многие физические и химические свойства молекул без детального представления об электронно-ядерной конфигурации атомов и учета их квантовой природы.

Дать трактовку механизма свертывания белковой цепи из состояния статистического клубка в строго детерминированную нативную (физиологически активную) конформацию макромолекулы—означает ответить на следующие вопросы:

1. Какие силы управляют процессом самосборки белковой цепи?

2. Каким образом совершенно беспорядочные и случайные конформационные (вращение вокруг одиночных связей) флуктуации белковой цепи спонтанно и безошибочно переходят во вполне определенный процесс организации трехмерной структуры белка?

3. Почему столь мала продолжительность сборки нативной конформации белка, осуществляемой по беспорядочно-поисковому механизму, единственно возможному в случае самопроизвольно протекающего процесса и, следовательно, неизбежно связанному с перебором огромного массива конформативных конфигурационных флуктуаций?

В течение последних десятилетий было предпринято много попыток ответить на эти вопросы. Они строились на основе эмпирических корреляций или классической равновесной термодинамики. Все попытки оказались неудачными. Причина этого в том, что сборка структуры белка является сугубо неравновесным процессом спонтанного перехода от хаоса к порядку, и потому используемые ранее подходы не могли привести к желанной цели по причинам принципиального характера. Основной решенной таких задач должна быть только нелинейная неравновесная термодинамика.

Обсуждение неравновесной модели свертывания белковой цепи в нативную конформацию начнем с определения минимального фазового и компонентного состава системы, обеспечивающей спонтанное протекание процесса в изолированных условиях. Не нарушая общности модели, будем считать, что объектом нашего рассмотрения является мономерный белок. Имеющиеся опытные данные о свертывании белков позволяют представить укладку линейной полипептидной цепи в трехмерную структуру как внутримолекулярный процесс, который полностью определяется

проявляющимися в соответствующих условиях свойствами природной аминокислотной последовательности. Иными словами, свертывание природной полипептидной цепи не зависит от концентрации белка, поэтому наша модель может включать лишь одну белковую молекулу.

Помимо этого, в систему, безусловно, должна входить водная фаза. Для нашего феноменологического описания процесса не требуется учет конкретных специфических свойств среды. Будем считать, что водное окружение обладает всем необходимым для реализации заложенной в белковой цепи потенции к самоорганизации.

Таким образом, выбранная модельная система состоит из двух подсистем—единичной природной полипептидной цепи и водной среды. Чтобы такая система удовлетворяла требованиям классической равновесной термодинамики, она должна быть макроскопической, а все ее подсистемы представлять собой совокупность множеств, не связанных друг с другом микрочастиц и микросостояний.

Макроскопичность белковой подсистемы определяется большим числом образующих молекулу белка атомов (порядка  $10^3 - 10^4$ ) и значительным количеством конформационных степеней свободы ( $10^2 - 10^3$ ), позволяющих в принципе полипептидной цепи принимать множество микроскопических состояний ( $10^n$ , где  $n$ —число аминокислотных остатков в цепи). И тем не менее эту подсистему нельзя считать строго термодинамической, поскольку она не может быть представлена совокупностью беспорядочно перемещающихся и не взаимодействующих друг с другом микрочастиц. Напротив, все они весьма эффективно связаны между собой посредством как валентных, так и невалентных взаимодействий (дисперсионные, электростатические, торсионные и водородные связи). Тот факт, что разные природные аминокислотные последовательности в одних и тех же условиях свертываются существенно различным образом, неоспоримо доказывает, что совокупность невалентных взаимодействий влияет на траектории движений включенных в уникальную валентную схему атомных групп, делает движения атомов взаимозависимыми, не в полной мере хаотичными, что в конечном счете и определяет процесс структурной организации белков. На это же указывают данные противоположного характера по синтетическим полипептидам. В значительной мере лишенные специфических корреляций между аминокислотными остатками, они ни при каких условиях не образуют идентичных для всех молекул стабильных трехмерных структур. Белковая макромолекула в водном окружении подвержена лишь одному виду истинно хаотического движения—броуновскому движению.

Что касается второй подсистемы—водного окружения, то она состоит из множества малых молекул, склонных в силу своей природы к образованию сильных водородных связей и электро-



статическим взаимодействиям. Ни одно свойство жидкой воды не может быть описано на основе предположения о полностью хаотическом движении отдельных молекул. Эксперименты, в частности колебательные спектры, вообще не обнаруживают в воде свободных молекул  $\text{H}_2\text{O}$ . Дж. Бернал еще в 1932 г. в своих, ставших классическими, рентгеноструктурных исследованиях воды впервые наблюдал в ее жидкой фазе зародышевые формы кристаллов льда. Статистический анализ также свидетельствует о присутствии в воде при нормальных условиях значительных количеств ассоциатов с одной, двумя, тремя и четырьмя межмолекулярными водородными связями.

Интересно обратить внимание, как в этом случае проявляются некоторые важнейшие свойства энтропийных процессов. Выше отмечалось, что при самоорганизации энтропия объекта падает. Процесс этот возможен только потому, что имеется связь с внешней средой, где энтропия, наоборот, должна нарастать, причем быстрее, чем происходит ее уменьшение за счет самоорганизации. Если процесс самоорганизации идет быстро, то еще быстрее должно происходить нарастание энтропии во внешней среде. С другой стороны, быстрое нарастание энтропии невозможно в средах, где уровень энтропии уже высок. Значит, подходящая для процесса образования белковых структур среда должна обладать заметной долей структурной организации. Тогда связанное с ростом белковой молекулы частичное разрушение этой структуры может привести к необходимому быстрому росту энтропии в среде. Вода, со свойственной ей при невысоких температурах квазикристаллической структурой, как раз в этом смысле является очень подходящим объектом.

Специфика взаимодействия воды с природной аминокислотной последовательностью, раскрывающая возможность последней к структурированию, определяется не полностью независимым, хаотическим, тепловым движением молекул  $\text{H}_2\text{O}$ , а сложной многофазной структурой воды, сильным поверхностным натяжением (большой избыточной энергией поверхностного слоя), высокой избирательностью взаимодействий в контактном слое с разными по своей природе атомными группами белка.

Будем считать, что на протяжении всего процесса структурной самоорганизации белка система белок—вода не обменивается с окружающим пространством ни энергией, ни веществом, т. е. принадлежит к замкнутому типу систем. Исходим из того, что в начальный момент белковая молекула находится в состоянии флуктуирующего клубка. Результатом спонтанно возникающего процесса свертывания полипептидной цепи может явиться, как было показано выше, равновесное состояние, в котором белок приобретает высокоорганизованную, сколь угодно долго устойчивую структуру. Процесс идет без нарушения второго начала термодинамики, т. е. протекает с повышением энтропии всей системы и заканчивается тогда, когда она достигает максимума.

Обсуждаемая модельная система не может, тем не менее, быть объектом рассмотрения классической равновесной термодинамики. Последняя оперирует только усредненными характеристиками и функциями состояния, а поэтому ограничена исследованием макроскопических систем с чисто случайным, полностью неупорядоченным движением микроскопических частиц. Равновесная термодинамика в состоянии анализировать коллективное поведение огромного числа частиц без учета деталей их внутреннего строения и без конкретизации механизма равновесного процесса. Особенно важно отметить то обстоятельство, что для классической термодинамики все случайные флуктуации системы неустойчивы, обратимы и, следовательно, не могут оказывать заметного, а тем более конструктивного воздействия на протекающий процесс. Все самопроизвольно совершающиеся явления в изолированной системе должны быть направлены, согласно равновесной термодинамике, на достижение однородности системы во всех возможных отношениях, т. е. на разрушение существующих структур, а никак не на их создание. Равновесные процессы всегда идут с увеличением энтропии не только всей системы, но и каждой ее подсистемы. В классической термодинамике отсутствуют какие-либо следствия из ее фундаментальных начал, которые указывали бы на принципиальную возможность перераспределения энтропии между различными фазами изолированной системы.

Следовательно, выбранная нами модельная система свертывания белковой цепи не отвечает основным положениям классической статистической физики и поэтому исключает саму возможность спонтанной сборки трехмерной структуры белка.

Линейная неравновесная термодинамика изучает процессы, протекающие в окрестности положения равновесия. Для этого к известным началам равновесной термодинамики дополнительно вводится ряд новых постулатов, уже не столь всеобъемлющего и строгого характера. Так, предполагается соблюдение сформулированных И. Пригожиным принципов локального равновесия и минимума производства энтропии. Привлечение этих частных и не соблюдаемых строго постулатов не ведет, однако, к расширению возможностей термодинамического подхода к объяснению механизма самоорганизации высокоупорядоченных структур. Изучение строится здесь путем экстраполяции равновесного состояния, т. е. по-прежнему на основе усредненных термодинамических характеристик системы. Молекулярная специфика объекта учитывается лишь неявным образом. Таким образом, наша модельная система не может рассматриваться не только равновесной термодинамикой, но и линейной неравновесной термодинамикой.

Обратимся теперь к развитой Пригожиным нелинейной термодинамике неравновесных процессов, важнейшими составными частями которой являются теория диссипативных систем и теория бифуркаций. Реакция Белоусова—Жаботинского, турбулентное

(вихревое) движение жидкости, модель Лоттга—Вольтерра, описывающая взаимоотношения между хищником и жертвой,—все это открытые диссипативные структуры. Динамические процессы перечисленных и подобных им макроскопических систем действительно приводят при достижении условий, превышающих соответствующий критический уровень, к спонтанному возникновению из беспорядка высокоорганизованных пространственных, пространственно-временных и просто временных структур. Но во всех случаях поддержание возникшего из хаоса порядка в стационарном режиме оказывается возможным только при постоянном энергетическом или материальном обмене между окружающей средой и динамической системой.

Совершающийся в открытой системе неравновесный процесс вдали от положения равновесия связан с диссипацией, т. е. с производством энтропии или, иными словами, с компенсирующим это производство потреблением *негэнтропии* (отрицательной энтропии) из окружающей среды. Перекрытие внешнего потока негэнтропии автоматически ведет к прекращению системой производства энтропии и, как следствие, к распаду созданной диссипацией структуры. У открытых диссипативных систем «конечной целью» (аттрактором) является не равновесное состояние, а расположенное далеко от него состояние, названное Л. Берталанфи «состоянием текущего равновесия». Такой аттрактор получил специальный термин «странный аттрактор Лоренца».

В отличие от неравновесного процесса диссипативной системы, белковая подсистема спонтанно потребляет негэнтропию из водного окружения внутри изолированной системы, причем только в период своей структурной самоорганизации. Созданная же структура термодинамически стабильна и не нуждается в постоянной инъекции негэнтропии. Результатом процесса здесь служит положение равновесия, к которому стремится двухфазная система и которого она достигает, когда белковая цепь принимает свою нативную конформацию. Таким образом, физиологически активная трехмерная структура белка—это термодинамически устойчивое, равновесное состояние подсистемы в изолированной макроскопической двухфазной системе.

Согласно теории диссипативных систем и теории бифуркаций, возникновение порядка из беспорядка возможно только в ходе истинно неравновесного процесса, который даже в принципе не может быть представлен равновесным. Отличительная особенность истинно неравновесного процесса заключается в наличии у системы особых незатухающих, необратимых флуктуаций, проявляющихся в определенных условиях и вызванных специфическими молекулярными свойствами микроскопических составляющих, т. е. тем, что по определению не учитывается равновесной термодинамикой и линейной неравновесной термодинамикой. Последовательная цепочка таких взаимосвязанных и согласованных друг с другом и благодаря этому неравновесных и эволюционирующих в определенном направлении флуктуа-

ций—непосредственная причина зарождения и развития порядка из хаоса. Пригожиным впервые были разработаны физические принципы перераспределения (диспропорционирования) энтропии при самопроизвольных процессах. Они утверждают, что для понимания и априорного описания спонтанного возникновения порядка из беспорядка необходимо установление количественных взаимоотношений между коллективным поведением макроскопической системы и индивидуальными свойствами ее микроскопических составляющих.

Развитые И. Пригожиным представления нелинейной термодинамики могут служить общей теоретической основой изучения свертывания белковой цепи. В пользу этого свидетельствует наличие у процессов структурной организации диссипативных систем и белков следующих общих черт. Во-первых, оба они эволюционируют в сторону создания высокоорганизованных структур. Во-вторых, и тот и другой самопроизвольны и сопровождаются производством и диспропорционированием энтропии. В-третьих, причина неравновесности процессов заключается в обоих случаях в существовании устойчивых бифуркационных флуктуаций.

Исходным состоянием системы белок—вода является находящаяся в водной среде белковая цепь в состоянии статического клубка, которая совершает броуновское движение. В начальный период все флуктуации—периодические вращения атомных групп вокруг одинарных связей—беспорядочны и не синхронизированы друг с другом. В равновесных процессах они обратимы, поэтому не оказывают существенного воздействия на движение системы в сторону положения равновесия. Некоторые из флуктуаций приводят к сближению определенных атомных групп, которые могут эффективно взаимодействовать между собой. Благодаря этому возникают и сохраняются поначалу случайным образом появившиеся, но не случайным образом стабилизированные конформационные состояния тех или иных участков белковой цепи.

Какие же факторы делают устойчивым определенное сочетание случайных отклонений? Очевидно, возникшее благодаря флуктуациям необратимое конформационное состояние пептидного фрагмента будет обладать наименьшей энергией и, следовательно, явится стабильным, если приведшие к его созданию флуктуации окажутся согласованными между собой. Это означает, что каждая из таких флуктуаций отвечает предпочтительному по энергии взаимному расположению ближайших атомных групп (например, в пределах одного аминокислотного остатка), а определенное сочетание таких флуктуаций приводит к дополнительным стабилизирующим взаимодействиям более удаленных атомных групп, принадлежащих соседним по цепи остаткам.

Что нужно, чтобы такое событие, несмотря на случайность и беспорядочность всех флуктуаций, наверняка имело бы место? Для этого необходимо, чтобы единственно возможным (при самопроизвольности процесса) беспорядочно поисковым механиз-

мом сборка белка осуществлялась бы за короткое время. На первый взгляд, это условие выглядит нереальным, так как число возможных комбинаций чисто случайных и в значительной мере беспорядочных конформационных флуктуаций полипептидной цепи невероятно велико. Возможность появления среди них бифуркационных флуктуаций ничтожно мала. Так, перебор всех микроскопических состояний даже у низкомолекулярного белка занял бы не менее  $10^{30}$  лет. Противоречие между характером процесса и его продолжительностью снимается, если предположить, что актуальные на первом этапе сборки белка бифуркационные флуктуации могут возникать независимо на разных участках аминокислотной последовательности. Это предположение равносильно постулированию возможности образования при определенном сочетании согласованных флуктуаций конформационно достаточно жестких фрагментов только за счет невалентных взаимодействий входящих в него остатков. Время, необходимое для локальной структурной самоорганизации при беспорядочно-поисковом механизме, грубо ориентировочно равно, согласно Д. Уетлауферу,  $10^{(n-14)}$  с ( $n$ —число остатков в цепи). Следовательно, для пептидных фрагментов белковой цепи, например с  $n = 12$ , время сборки составляет всего  $10^{-2}$  с. Процессы структурирования разных участков последовательности могут идти параллельно с неодинаковыми скоростями. Их независимость друг от друга предполагает чередование в белковой цепи конформационно жестких и подвижных фрагментов.

Таким образом, важнейшая особенность белков, обуславливающая неизбежность возникновения необратимых бифуркационных флуктуаций и, следовательно, саму возможность спонтанного возникновения из хаоса высокоорганизованной структуры, согласно этой модели, заключается в специфической конформационной неоднородности природной аминокислотной последовательности. Модель, отражая суть рассматриваемого явления, постулирует наличие неразрывной связи между химическим строением белковой цепи, ее конформационными свойствами и необратимыми флуктуациями. Упорядоченность природной аминокислотной последовательности ответственна за различие в конформационных возможностях отдельных участков. Конформационная упорядоченность чередующихся белковых фрагментов порождает термодинамическую неоднородность флуктуаций, дифференцируя их на обратимые (равновесные) и необратимые (неравновесные). Такая связь обнаруживается только у отобранных эволюцией аминокислотных последовательностей.

В результате неминуемого появления среди множества случайных обратимых конформационных изменений некоторых коротких участков белковой цепи устойчивых необратимых флуктуаций возникает первая бифуркационная точка, являющаяся поворотной в отношении направленности и характера дальнейшего процесса; с этого момента он становится истинно неравновесным. Далее (также за счет возникающих чисто

случайно флуктуаций) в специфические взаимодействия вовлекаются удаленные по цепи остатки, принадлежащие разным участкам белка. Их сближение происходит опять-таки при вполне определенных, строго согласованных и поэтому необратимых флуктуациях. Структурная стабилизация возникает здесь за счет пространственной согласованности избирательно констатирующих конформационно-жестких фрагментов. Появление в течение разумно короткого времени на этом этапе сборки необратимых флуктуаций на значительно больших, но уже частично структурированных участках также неизбежно. Увеличение количества взаимодействующих между собой остатков сопровождается уменьшением конформационных степеней свободы.

Так что и здесь вполне возможно выявление необратимых флуктуаций при самопроизвольном беспорядочно-поисковом механизме сборки. Благодаря избирательности бифуркационных флуктуаций и их строгой согласованности процесс структурной самоорганизации белка нарастает лавинообразно и приобретает детерминистические черты. Из конформационно жестких фрагментов возникают структурные образования, которые через ряд чисто случайных, но развивающихся в одном направлении событий входят в нативную трехмерную структуру белка. Устойчивый и конструктивный характер бифуркационных флуктуаций обусловлен уникальной аминокислотной последовательностью.

Неравновесная термодинамическая модель свертывания белковой цепи—это первый шаг к созданию физической теории структурной самоорганизации белка и соответствующего метода расчета.

Значение модели состоит в том, что она дает принципиальное и непротиворечивое объяснение всем важнейшим особенностям структурной организации белка: беспорядочно-поисковому механизму сборки природной аминокислотной последовательности, самопроизвольному характеру возникновения и протекания всех стадий образования высокоупорядоченной трехмерной структуры, высокой скорости и безошибочности процесса.

Модель указывает направление дальнейшего поиска и раскрывает его содержание. Рассмотренная модель ответила на поставленные выше вопросы, касавшиеся характерных особенностей свертывания белковой цепи в физиологически активную конформацию, а также позволила сформулировать принципы, которым должна отвечать структурная организация белков. Они имеют следующий вид:

1. Нативная трехмерная структура и конформационные динамические свойства белковой молекулы в физиологических условиях полностью определяются аминокислотной последовательностью.

2. Нативная конформация белка отвечает глобальному минимуму внутренней энергии.

3. Нативная конформация белка является наиболее плотно упакованной структурой, и следовательно, на всех стадиях самопроизвольного свертывания белковой цепи внутримолекулярные взаимодействия валентно не связанных атомов превалируют над межмолекулярными взаимодействиями.

4. В нативной конформации белка имеет место согласованность между всеми внутримолекулярными взаимодействиями валентно несвязанных атомов.

5. В белковой аминокислотной последовательности имеет место чередование конформационно-жестких и конформационно-подвижных участков; состояния первых однозначно определяются взаимодействиями входящих в них остатков, а состояния вторых—дополнительными взаимодействиями далеко расположенных остатков.

На основе перечисленных принципов были разработаны физическая конформационная теория и количественный метод расчета трехмерной структуры белковой молекулы по известной аминокислотной последовательности. Справедливость теории и метода, а следовательно, и термодинамической модели самосборки белка подтверждена совпадением результатов априорного расчета конформации нативного белка с данными рентгеноструктурного анализа.

Известные из расчета данные о конформационных возможностях белковой аминокислотной последовательности позволили дать не качественное, как это было сделано выше, а количественное описание всех стадий механизма структурной самоорганизации белковой цепи.

Описанный выше процесс образования упорядоченных белковых структур имеет весьма близкую аналогию с результатами компьютерного моделирования с помощью так называемой программы «Жизнь». Соответствующий алгоритм (правило игры) был предложен кембриджским математиком Дж. Конвеем. Алгоритм игры заключается в следующем. Представим себе разделенное на клетки, подобно шахматной доске, поле. С помощью датчика случайных чисел поместим в одну из клеток фишку. На экране дисплея компьютера это будет выражаться в появлении в какой-то части экрана заштрихованного квадрата. Этот квадратик, если он только не находится на краю поля, окружен восьмью пустыми ячейками: они размещены слева и справа от центрального квадрата, сверху и снизу от него и на концах его диагоналей.

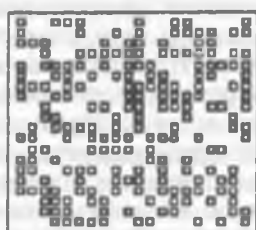
Будем теперь увеличивать число квадратов на экране дисплея. Тогда рано или поздно возникнет такая ситуация, когда новые фишки будут попадать в квадраты, непосредственно окружающие данный. Предположим, далее, что взаимодействия между фишками (заштрихованными квадратами) возможно только тогда, когда они соприкасаются либо вершинами, либо сторонами. Это значит, что будут взаимодействовать между собой все фишки, которые попадут хотя бы в одно из восьми

возможных положений около исходного квадратика. Эти взаимодействия подчиняются правилам «рождения» и «смерти». Правило «рождения» позволяет дополнительно заштриховать любой новый квадратик, если он имеет ровно три ближайших уже заштрихованных соседа. Правило «смерти» предусматривает уничтожение штриховки (квадратик снова свободен от фишки), если соседей менее двух (смерть от одиночества) либо более трех (смерть от перенаселения). На каждом шаге итераций производится компьютерный просмотр всех клеток поля дисплея и те квадратики, которые соответствуют правилу рождения, заштриховываются (в них помещаются новые фишки), а те, которые попали в условия смерти, освобождаются от штриховки (фишки из соответствующих мест удаляются). Этот процесс отвечает «смене поколений». На первый взгляд кажется, что в результате с экран дисплея сразу после первого шага исчезнут все заштрихованные квадраты. Так действительно и будет (во всяком случае, долгое время), если число вводимых на первом шаге случайным образом в клетки экрана дисплея сигналов, приводящих к появлению заштрихованных квадратиков, будет малым по сравнению с общим числом клеток, на которые разделено все поле экрана. В самом деле, в этом случае образование ступка заштрихованных клеток, способного к размножению, маловероятно. Положение, однако, кардинально меняется, если число вводимых в игру фишек увеличить. Тогда ситуация, при которой образуется такая конфигурация близко расположенных фишек, когда уже может сработать правило «рождения», становится не столь редкой. Это можно трактовать как появление флуктуации, способной к саморазрастанию.

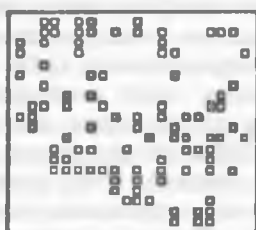
Дальнейший процесс не поддается какому-либо аналитическому описанию, но результаты громадного количества компьютерных экспериментов приводят к одним и тем же конечным результатам: во-первых, несмотря на то, что кроме «рождения» предусмотрена еще и «смерть», довольно быстро появляются способные к разрастанию флуктуации; во-вторых, после смены нескольких поколений может появиться такая устойчивая структура, что ни одна из ее частей не может умереть, но и никакие три соседа не могут породить новую (нет клеток в подходящих положениях); в-третьих, получающиеся устойчивые структуры имеют правильные симметричные формы. Не правда ли, это очень похоже на реальную жизнь.

Очень простые правила, действие которых начинает проявляться на втором этапе после того, как в результате случайного процесса образовались подходящие флуктуации, не только приводят к появлению правильных упорядоченных макроструктур, но и обеспечивают их устойчивость. Таким образом, нечто принципиально новое и организованное рождается из хаоса в результате сочетания случайного (на первом этапе) и закономерного (в том смысле, что взаимодействие подсистем подчиняется вполне определенным правилам) процессов. Это не значит, конечно, что

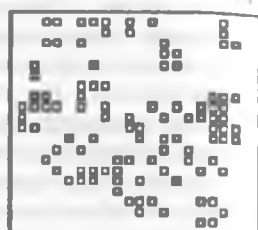




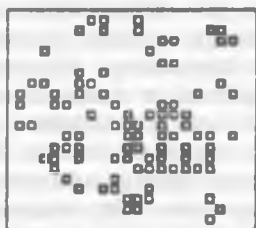
*Начальное состояние*



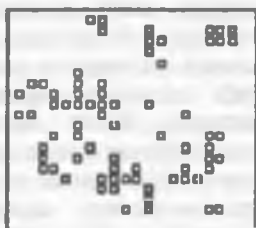
*Шаг 1*



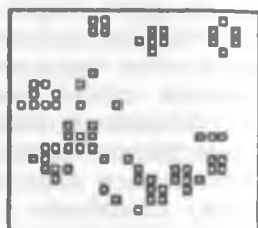
*Шаг 2*



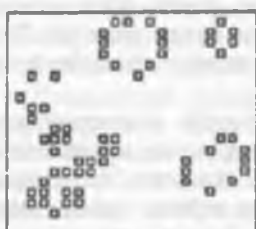
*Шаг 3*



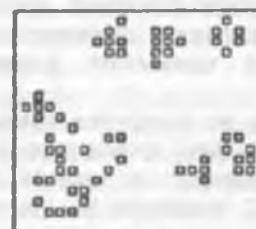
*Шаг 4*



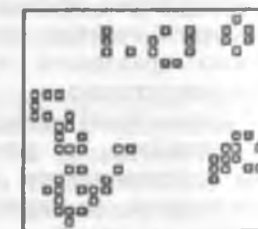
*Шаг 5*



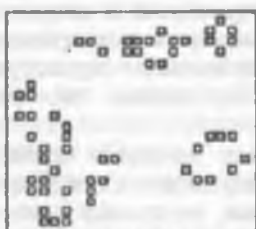
*Шаг 12*



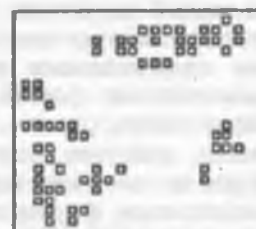
*Шаг 13*



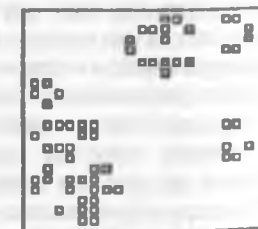
*Шаг 14*



*Шаг 15*



*Шаг 16*



*Шаг 17*

Рис. 10.20. Последовательные стадии (шаги) компьютерной игры «Жизнь»

каждый раз будет получаться одна и та же устойчивая структура. Элемент случайности даже при одинаковых правилах игры может привести к появлению разных, априори не предсказуемых структур.

На рис. 10.20 приведены результаты одной из компьютерных игр. Хорошо видно, как зарождаются и формируются устойчивые в конечном счете геометрически правильные симметричные структуры. Каждый шаг отвечает одной «смене поколений» (одному действию правил «рождения» и «смерти», которые применяются сразу ко всем объектам, в данный момент находящимся в подходящих ситуациях). На рисунках показаны только «живые» клетки поля дисплея. Пустые клетки поля не отмечены. В первый момент игровое поле заполняется достаточно большим числом случайно расположенных «живых» клеток. Многие из них сразу же «умирают» в силу действия правила «смерти». Однако некоторые из клеток оказываются в положениях, приводящих к «рождению» новых. Хорошо видно, что после смены пяти поколений в верхнем правом углу образуется устойчивая структура правильной формы из шести «живых» клеток. Эта структура сама по себе способна сохраняться неопределенно долго (в данном случае до 16-го поколения включительно), если только, опять-таки случайно, не возникнет близко расположенный к ней ступок других «живых» клеток, который и приводит к ее разрушению.

Таким образом, результаты компьютерного эксперимента хорошо демонстрируют как процесс появления порядка из беспорядка, так и возможность разрушения устойчивой правильной структуры под действием вновь появившегося внешнего фактора.

Подводя итог изложенному в этом параграфе, сформулируем общий вывод. Он заключается в том, что сами по себе подчиняющиеся чисто статистическим (случайным) закономерностям процессы, даже сопровождающиеся появлением заметных флуктуаций, ни к каким новообразованиям (новому качеству) привести не могут.

Для этого необходимо еще наличие специфических детерминированных правил взаимодействия объектов, которые в результате флуктуации оказались в положениях, когда такие взаимодействия могут оказать решающее влияние на все дальнейшее развитие событий. Если первая стадия процесса самоорганизации еще может быть описана на языке равновесной или линейной неравновесной термодинамики, то вторая стадия, принципиально связанная со спецификой локальных взаимодействий, в рамках этих разделов термодинамики исследована быть не может, так как в них отсутствует само понятие о специфическом близкодействии.

Несколько дополнительных замечаний в качестве заключения. Мы начали изучение курса физики с простейших объектов и подошли к важнейшим закономерностям появления и существования живого, т. е. к высшей форме организации материи. Это

показывает необычайно широкий круг вопросов, которыми занимается современная физика. По мере усложнения ситуации мы, разумеется, теряем способность углубляться в детали. Но все же уместно поставить вопрос: в какой степени найденные физикой общие закономерности, которые, как мы видели, часто не отражают даже самого факта существования «стрелы времени», что характерно для живой природы, пригодны именно для исследования таких объектов?

Ответ уже дан самой жизнью. Физика необходима особенно тогда, когда исследование ведется на молекулярном уровне или на уровне обобщенных фундаментальных закономерностей.

Но это уже особый разговор. Мы вплотную подошли к биофизике и молекулярной биологии. Не будем вторгаться в область этих специальных наук.

### Вопросы для самоконтроля

1. Можно ли, производя эксперименты в рамках обычной ньютоновской механики, обнаружить, что время всегда принимает только положительные значения? Изменится ли уравнение движения в том случае, если поменять положительные значения времени на отрицательные?

2. Произойдут ли какие-нибудь изменения, приводящие к физическим эффектам, при замене положительного значения времени на отрицательное в том случае, если используется уравнение Максвелла?

3. В обычной классической механике при одном и том же уравнении движения, но при разных начальных условиях через один и тот же промежуток времени состояние системы будет совершенно различным. Что в этом смысле наблюдается для макроскопической системы многих частиц, например для системы молекул газа в некотором сосуде, если первоначально они распределены в разных точках пространства данного сосуда и имели разные скорости?

4. Можно ли применить принцип детерминизма в его классической трактовке к механическим явлениям для системы, состоящей из многих частиц (статистического ансамбля)?

5. Какие физические явления приводят к необходимости введения понятия времени как такой величины, которая меняется от нуля до бесконечности, оставаясь все время положительной, т. е. от прошлого к будущему?

6. Какая из известных физических величин наиболее тесным образом связана с утверждением о том, что время может быть только положительным?

7. Что такое флуктуации и является ли их наличие обязательным признаком существования любой системы, состоящей из очень большого числа объектов?

8. Какой из физических экспериментов наиболее наглядно указывает на наличие флуктуации в жидкостях и газах? Какие условия необходимы для того, чтобы флуктуации начали разрастаться и привели к макроскопическому эффекту? Что такое бифуркация?

9. Возможна ли самоорганизация в системах, не имеющих контакта с внешней средой, т. е. в замкнутых системах?

10. Приведите примеры самоорганизации в простейших объектах—жидкостях и газах.

11. Могут ли существовать устойчивые вихри или конвекционные явления в идеальном газе? Какие свойства частиц реального газа ответственны за появление направленных движений в жидкостях и газах?

12. Пусть в некоторой открытой системе происходит самоорганизация и энтропия убывает. Как поведет себя энтропия в связанной с этой системой наружной среде?

13. Какие основные закономерности наблюдаются при транспорте вещества в природе?

14. Является ли благом появление бессмертных организмов или же смерть является необходимым фактором для существования природных циклов?

15. Возможно ли заимствование вещества биологической системой с очень низкой энтропией от системы с очень высокой степенью энтропии?

16. Можно ли связать законы изменения энтропии с общими законами развития общества и экологическими проблемами?

17. Какова связь энтропии и информации?

18. Какова одна из важнейших причин того, что элементарные биологические процессы проходят именно в водной среде?

19. Как Вы себе представляете различные стадии образования пространственных структур белковых молекул?



# Г Л А В А 11

## УРОКИ ФИЗИКИ

*«Современное положение физики, возможно, лучше всего характеризуется тем, что почти все идеи, которые когда-либо в естествознании показали себя результативными, получили свои права во всеобщей гармонии, не теряя при этом в плодотворности»*

Н. Бор

Настоящая глава написана с целью продемонстрировать, в согласии с эпиграфом, как некоторые идеи, первоначально сформулированные в физике в процессе ее исторического развития, в дальнейшем распространялись на совершенно другие области, вплоть до истории и психологии, позволяя и в этих областях науки и человеческого знания понять некоторые фундаментальные закономерности. Разумеется, число таких примеров столь велико, что даже их краткое изложение потребовало бы специального монографического исследования. Мы ограничимся лишь несколькими примерами, но затрагивающими чрезвычайно важные проблемы. Начнем с так называемого принципа дополнительности.

### 11.1. Принцип дополнительности

Понятие «дополнительность» было введено в науку Н. Бором в 1928 г. в период становления квантовой механики и интенсивных экспериментальных исследований микромира (атомов, молекул и т.д.).

Основоположник этого принципа Н. Бор, отправляясь от решения чисто физических проблем, сразу же понял общность этого принципа и уже в одной из своих первых работ смело перекинул мост от физики к психологии и вообще ко всей теории познания и формирования образа окружающего мира. Именно поэтому принцип дополнительности следует считать одним из важнейших достижений науки, и его знание необходимо для понимания очень многих фундаментальных проблем философии и явлений окружающего нас мира.

Начнем с простейших рассуждений. Уже при исследовании равномерного прямолинейного движения материальной точки обнаруживается, что полное представление об этом движении можно получить только в том случае, когда даны ответы на два вопроса: где в данный момент находится точка и с какой скоростью она движется?

Ответы на эти вопросы в общем случае и прогноз движения точки возможны, если задать уравнение ее движения. Для этого должны быть известны масса точки и действующие на нее во время эксперимента силы. Однако и таких сведений недостаточно: чтобы прогноз был вполне определенным, надо еще указать, где было расположено тело в начальный момент и с какой скоростью оно двигалось. Мы опять сталкиваемся с той же парой величин: координатой (ее задание дает ответ на вопрос, где находилась точка в начальный момент времени, с которого мы следим за движением) и скоростью.

В классической механике, наряду со скоростью, вводится, как известно, понятие импульса. Оказывается, что введение этой величины не сводится просто к перезаписи основных уравнений. На языке импульсов можно описать, например, взаимодействие материальной точки и поля и говорить о передаче телом импульса полю и, наоборот, импульса поля телу. Это нельзя сделать на языке сил, так как, если можно приложить силу к телу, то «приложить» ее к полю никак нельзя. Введение импульса позволяет обобщить понятие силы и называть силой скорость изменения импульса тела. Кроме того, наиболее общие пригодные для характеристики любых механических объектов уравнения движения, предложенные Лагранжем и Гамильтоном, записываются на языке координат и импульсов и их изменений во времени. Например, пара взаимосвязанных уравнений Гамильтона имеет следующий вид:

$$\frac{dq_i}{dt} = \frac{\partial \mathcal{H}}{\partial p_i} \quad \text{и} \quad \frac{dp_i}{dt} = - \frac{\partial \mathcal{H}}{\partial q_i} \quad (i = 1, 2, 3, \dots, n).$$

Здесь  $n$  — число степеней свободы системы (число независимых координат  $q_i$ ),  $q_i$  и  $p_i$  — произвольные координаты и соответствующие им так называемые сопряженные импульсы,  $\mathcal{H} = T_{\text{кин}} + U_{\text{пот}}$  — функция Гамильтона (или гамильтониан), являющаяся суммой кинетической и потенциальной энергий системы. При этом координаты и импульсы, хотя и согласованные между собой, рассматриваются как независимые переменные. Координаты и импульсы встречаются в уравнениях движения и начальных условиях парами, что позволяет называть их каноническими переменными, полностью описывающими изменение состояния системы во времени.

Мы не будем далее углубляться собственно в механику. Для нас сейчас важно усвоить только одно: описание даже простейшей механической системы (материальной точки, движущейся вдоль одной выделенной оси координат в заданном потенциаль-

ном поле, например, в так называемой потенциальной яме определенной формы) требует для своей полноты двух величин, двух составляющих. Обойтись только одной из этих составляющих никак нельзя, поскольку они как бы дополняют друг друга.

Конечно, можно возразить, что, если материальная точка в начальный момент находилась в минимуме потенциальной ямы и не двигалась, то она и дальше будет находиться в состоянии покоя, и в этом смысле для полной характеристики ее поведения достаточно знать координату. Однако не следует забывать, что надо указать, что начальная скорость была равна нулю. Кроме того, это все же крайний частный случай.

Это правило парности сохраняется и при переходе к квантовой механике и к описанию движения микрочастиц. Однако, появляется и существенное различие с классикой. Вернемся к гл. 9. Эксперименты по прохождению, например, потоков электронов через экраны с узкими щелями (реально через тонкую металлическую фольгу или кристаллы) показали, что наблюдается дифракционная картина, подобная случаю прохождения электромагнитных волн через соответствующие экраны. Далее, исследования спектров атомов и молекул указали на существование вполне определенных дискретных энергетических состояний этих объектов, переходы между которыми и приводят к излучению или поглощению электромагнитных волн подходящих длин. Эти новые особенности поведения микрочастиц потребовали для их описания создания адекватной механики, которая получила название квантовой или волновой. Классические уравнения движения были заменены уравнением Шредингера.

Не вдаваясь опять-таки в детали, напомним, что уже решение простейшей задачи о частице в потенциальном ящике с бесконечно высокими стенками (имитация частицы, находящейся в локализованной области пространства, например, электрона в атоме водорода) приводит не только к заключению о том, что энергия такой частицы должна принимать ряд дискретных значений, но и к тому, что она не может обладать нулевой энергией, т. е. покоиться. Значит, необходимость использования пары канонических переменных проявляется во всех возможных случаях. Но и это еще не самое главное. Как в классике, так и в квантовой механике большую роль играют так называемые перестановочные соотношения, т. е., например, комбинации такого рода:

$$x \frac{\partial \varphi}{\partial x} - \frac{\partial}{\partial x} x \varphi.$$

Если теперь воспользоваться так называемой операторной символикой, то от написанного выше выражения можно перейти к следующему:

$$\left( x \frac{\partial}{\partial x} - \frac{\partial}{\partial x} x \right) \varphi, \quad \text{или} \quad x \frac{\partial}{\partial x} - \frac{\partial}{\partial x} x.$$

Последнее выражение включает уже не функции, а операторы, т. е. символы математического действия. Теперь вспомним, что в квантовой механике каждой классической величине сопоставлен вполне определенный оператор. При этом в качестве оператора координаты принимается сама координата, т. е. ничего не меняется по сравнению с классикой, но классический импульс  $p$  заменяется оператором  $-i\hbar \frac{\partial}{\partial x} = p_x$  (если ограничиться только  $x$ -компонентой,  $\hbar$ —постоянная Планка). Теперь операторное выражение

$$-i\hbar x \frac{\partial}{\partial x} - -i\hbar \frac{\partial}{\partial x} x \quad (\text{здесь } i \text{—мнимая единица})$$

можно записать следующим образом:  $\hat{x}\hat{p}_x - \hat{p}_x\hat{x}$ . Такое выражение называется коммутатором. В квантовой теории строго доказывается, что если коммутатор пары операторов равен нулю, т. е. действие пары операторов на некоторую функцию не зависит от порядка их следования, то соответствующие физические величины могут быть измерены одновременно, причем точность измерения одной величины не зависит от точности измерения другой. Если коммутатор не равен нулю, то такие физические величины одновременно измерить с высокой точностью одним прибором в принципе нельзя. Если мы все же захотим прецизионно измерить такие величины, то надо проводить измерения разными способами. При этом в любом случае увеличение точности измерения одной величины приводит к падению точности измерения другой, и наоборот.

Для коммутатора операторов координаты и импульса имеем, как не трудно убедиться с помощью простых выкладок, следующее:  $x p_x - p_x x = -i\hbar$ . Таким образом, коммутатор не равен нулю. Значит, одновременные измерения с достаточно высокой точностью импульса и координаты микрочастицы в принципе невозможны.

Это и позволяет сформулировать принцип дополнителности в следующей достаточно общей форме (формулировка принадлежит известному физiku Д.Бому): *в области квантовых явлений наиболее общие физические свойства какой-либо системы должны быть выражены с помощью дополняющих друг друга пар независимых переменных, каждая из которых может быть лучше определена только за счет соответствующего уменьшения степени определенности другой*. Как указывалось, операторы, отвечающие такой паре величин, должны иметь не равный нулю коммутатор.

Принцип дополнителности не ограничивается только этими моментами. Например, волновые и корпускулярные проявления в поведении микрочастиц также являются взаимодополняющими и отражают реально существующий дуализм микромира. К паре взаимодополняющих понятий надо отнести непрерывность и дискретность и т. д.



А теперь приведем слова Н. Бора, который увидел в приведенных выше чисто физических примерах лишь частный случай чрезвычайно общего правила и написал в одной из своих первых работ по квантовой теории: «...пытаясь анализировать наши переживания, мы перестаем их испытывать. В этом смысле мы обнаруживаем, что между психологическими опытами, для описания которых адекватно употреблять такие слова, как «мысли» и «чувства», существует соотношение дополнительности, подобно тому, какое существует между данными о поведении атомов...» Продолжая эту мысль, Н. Бор отметил, что особенности измерений в атомной механике представляют «близкую аналогию со своеобразными трудностями психологического анализа, проистекающими из того факта, что духовное содержание неизбежно меняется, если внимание сосредотачивается на какой-нибудь его определенной стороне».

Прочитируем теперь А. Пушкина, который о принципе дополнительности не знал, но приведенные выше соображения Н. Бора задолго до него изложил в прекрасной форме (см. «Сцена из Фауста»):

Фауст

Я счастлив был...

Мефистофель

Творец небесный!

Ты бредишь, Фауст, наяву!

.....

Когда красавица твоя

Была в восторге, в упоенье,

Ты беспокойною душой

Уж погружался в размышленья

(А доказали мы с тобой,

Что размышленья—скуки семя).

И знаешь ли, философ мой,

Что думал ты в такое время,

Когда не думает никто?

.....

Ты думал: агнец мой послушный!

Как жадно я тебя желал!

.....

Что ж грудь моя теперь полна

Тоской и скукой ненавистной?..

.....

Потом из этого всего

Одно ты вывел заключение...

Фауст

Сокройся, адское творенье!

Беги от взора моего!

Чувство безрассудно. Анализ убивает чувство!

Вспомним слова, вложенные также А. Пушкиным в уста Сальери: «Звуки умертвив, музыку я разъял, как труп. Поверил я алгеброй гармонию». Чтобы изучить такое явление, как музыка, нужно, с одной стороны, уметь эстетически воспринимать музыку как целое, а с другой, знать законы построения мелодии. Слишком большое увлечение анализом, т. е. одной стороной познания объекта, приводит вообще к потере удовольствия от слушания музыки, от которой остается лишь труп!

А вот что писал П. Чаадаев об истории: «Есть историческая правда факта, есть историческая правда смысла: в истории есть анализ, но есть там и синтез... Без всякого сомнения, наиболее истинным является не то, что она повествует, а то, что она мыслит... В этом смысле поэтические представления могут быть ближе к истине, чем самый добросовестный рассказ...»

Другими словами, архивная полка не есть еще «История государства Российского». Надо, однако, иметь в виду, что и «Три мушкетера» не есть история Франции.

Снова процитируем Н. Бора: «...мы встречаемся здесь с иллюстрацией старой истины, что наша способность анализировать гармонию окружающего мира и широта его восприятия всегда будут находиться во взаимно исключающем, дополнительном соотношении».

Таким образом, сформулированный в физике принцип дополненности является лишь частным случаем весьма общей закономерности, имеющей, как мы увидим ниже, множество весьма важных следствий. Физика здесь сыграла роль катализатора, позволив на относительно простом примере обнаружить нечто гораздо более важное. Мы еще вернемся к вопросам о дополненности и обсуждению следствий, а сейчас обратимся к еще одному принципиальному моменту.

Как известно, в основе физики лежат измерения. Не только измерение, но и любое наблюдение подразумевает, прежде всего, что имеется объект наблюдения и субъект, или наблюдаемое и наблюдатель. В свою очередь, производство измерения возможно только тогда, когда имеется какое-то взаимодействие между наблюдателем и объектом. Это взаимодействие в физике чаще всего происходит через посредника—прибор, измерительное устройство, экспериментальную установку. Естественно, сразу же может возникнуть вопрос: если исследование подразумевает взаимодействие субъекта и объекта, не будет ли такое взаимодействие отражаться на свойствах как того, так и другого?

Исторический опыт астрономии и классической физики не заставлял физиков задумываться над этой проблемой. Опираясь на этот опыт, физики молчаливо предполагали, что если и есть какие-нибудь воздействия прибора на объект, то они, как правило, столь малы, что ими можно пренебречь. Например, хотя и ясно, что подключение вольтметра параллельно исследуемому участку электрической цепи должно изменить силу тока на этом участке по сравнению с тем током, который шел бы в

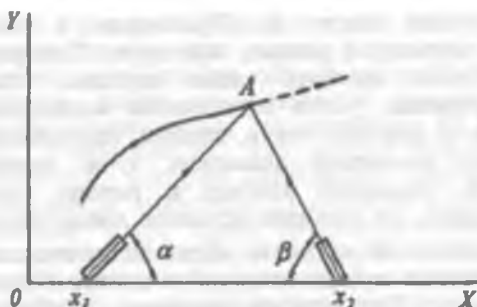


Рис. 11.1 Схема экспериментальной установки для изучения движения материальной точки  $A$ . В точках  $x_1, x_2$  находятся лазеры, подсвечивающие предмет

отсутствии вольтметра, однако, выбирая вольтметр со столь большим внутренним сопротивлением, что его можно считать заведомо во много раз превосходящим сопротивление участка цепи, можно сделать такие изменения пренебрежимо малыми. Науке, правда, были известны во множестве и такие измерения, когда их влияния на объект были столь большими, что уничтожался сам объект. Речь идет о химии, где производство химического анализа вещества сплошь и рядом приводит к разрушению самого исследуемого вещества. Тем не менее, обсуждаемая проблема как-то не привлекала особенного внимания. Считалось, что эксперимент всегда можно поставить так, чтобы он не искажал изучаемого явления.

Положение принципиальным образом изменилось, когда начались исследования микромира и происходило становление квантовой теории и ее математического аппарата. Чтобы разобраться в сути дела, обратимся к наглядному примеру. Представим себе, что мы проводим измерение координат и скоростей движения в разных точках траектории некоторого физического тела, которое можно рассматривать как материальную точку. Измерения будем производить с помощью подсветки этого тела двумя узкими лучами света (например, от двух лазеров), расположенных на заданном расстоянии друг от друга. Движение происходит в одной плоскости  $XY$ . Схема опыта показана на рис. 11.1.

Если оба лазера одновременно освещают объект, то внешний наблюдатель, измеряя для данного момента времени значения углов  $\alpha$  и  $\beta$ , зная расстояние  $l = x_2 - x_1$ , всегда может найти, пользуясь тригонометрическими правилами, координату точки пересечения лучей, т. е. объекта.

Если произвести с помощью того же самого прибора повторный замер координаты тела через малый промежуток времени, то можно найти и среднюю скорость в исследуемой области траектории.

Хотя известно, что свет оказывает давление на облучаемые тела, однако в тех случаях, когда масса материальной точки достаточно велика, а сила света лазеров не так уж значительна, при подобных измерениях вполне можно считать, что прибор на объект не оказывает никакого воздействия. Именно так чаще всего и бывает в классической физике.

Ситуация, однако, принципиально изменится, если масса тела станет очень малой. Тогда уже давлением света никак нельзя будет пренебречь. Более того, это давление может стать очень значительным, если одновременно уменьшать размеры тела и проводить все более и более острую фокусировку лучей на объект. Такая фокусировка потребует уменьшения длины волны света (из-за влияния дифракции), что с неизбежностью приведет к росту значений импульсов квантов света, отвечающих этой волне, и росту светового давления. В конце концов свет, если так можно сказать, «столкнет объект с траектории». Хотя по-прежнему однократное определение координаты точки остается возможным, но уже повторное измерение с целью определения скорости движения теряет всякий смысл, так как само это движение после первого измерения может радикально измениться по сравнению с невозмущенным. Чтобы найти эту характеристику движения с высокой степенью точности, надо видоизменить эксперимент, выбрать другой прибор. При этом окажется, что мы потеряем способность точно измерить координату.

Предположим теперь, что мы изучаем свойства атомов, наблюдая поглощение света слоем вещества, состоящего из этих атомов. Согласно второму постулату Бора, при взаимодействии света с веществом каждый атом может захватить квант света с определенной энергией (частотой). Недостаток соответствующих квантов в прошедшем через слой вещества луче света будет восприниматься как поглощение волн определенных длин.

Атом при поглощении кванта света не остается неизменным, а переходит в возбужденное состояние. Хотя он в этом состоянии долгое время не остается и вновь переходит в основное с наименьшей энергией так, что в целом при обычных экспериментах каких-либо изменений в слое газа на макроуровне не замечается, однако на микроуровне воздействие прибора на объект существует. И на макроуровне такое воздействие вполне может стать существенным, если использовать столь интенсивные пучки света (сейчас это реально возможно), что встреча атома со вторым квантом может произойти до того, как он успеет излучить первоначально поглощенный. Это наблюдается при исследовании так называемых нелинейных явлений при прохождении света через вещество, когда само вещество меняет свои свойства под действием света.

Поскольку весь мир состоит из атомов и молекул, то любые особенности микромира не могут тем или иным способом не проявляться в макроэкспериментах. В этом и заключается причина того, что идея дополнительности, первоначально сформулированная в физике применительно к микромиру, оказывается столь плодотворной в других областях знания. Совершенно прав был Н. Бор, когда писал, что ...«идея дополнительности способна охарактеризовать существенную ситуацию, которая имеет далеко идущую аналогию с общими трудностями образования человеческих понятий, возникающими из разделения

субъекта и объекта». Дополнительность, таким образом, связана с «разграничением объекта и средств наблюдения, свойственным самой идее наблюдения». При этом очень часто попытка более детального изучения одной стороны объекта, в полной аналогии и с соотношением неопределенности Гейзенберга, приводит к потере определенности другой.

Остановимся теперь на некоторых интересных с точки зрения исследования различных сторон окружающего нас мира следствиях из общего принципа дополнительности. Конкретно рассмотрим вопросы познания мира молекул и создания Истории как науки. Казалось бы, совсем разные вещи. Но мы увидим, что эти вопросы объединяет нечто общее. При этом мы будем исходить из еще одной формулировки принципа дополнительности, также данной Н. Бором: «Дополнительной к истине является ясность!»

## 11.2. Природа и модели

Согласно первому постулату Бора, состоящая из электронов и ядер нераспадающаяся микросистема (атом, молекула) может находиться лишь в определенных дискретных энергетических состояниях. Они называются уровнями энергии. Согласно второму же постулату, свободная, не находящаяся в контакте с другой, микросистема может изменить свое энергетическое состояние (перейти с одного уровня энергии на другой) только за счет либо поглощения, либо излучения электромагнитной энергии. И то и другое можно обнаружить, если наблюдать соответственно спектры поглощения или излучения. Таким образом, с помощью атомной и молекулярной спектроскопии удастся исследовать самые общие и фундаментальные свойства микросистем, так как в спектрах, регистрируемых в различных областях при разных условиях постановки экспериментов, проявляются все особенности строения атомов и молекул.

Именно поэтому пример спектральных исследований является не только очень удобным, но и весьма общим для того, чтобы можно было сделать ряд далеко идущих выводов.

Представим себе, что такое исследование ведется с помощью компьютеров (как реально сейчас и делают). Нам важно сейчас обратиться к примеру компьютеров не для того, чтобы популярно рассказать о компьютеризации молекулярной физики и химии, а потому, что компьютерный подход отличается наибольшей объективностью. Недаром считается, что если экзамен у абитуриента принимает человек, то он может быть предвзятым, а компьютер-экзаменатор совершенно объективен!

Каждый, однако, кто имел дело с компьютерами, не может, с одной стороны, не поражаться их мощью, а с другой—тупостью! Тупость эта выражается в том, что компьютеру все нужно подробно растолковывать. Иначе он просто не сможет работать. Это один человек другому может сказать: «Смотри, какой

красивый вид!» И второй, часто мгновенно, ответит: «Да!» Компьютер задаст вопрос, на который вряд ли вообще можно дать полный ответ: «А что значит «красивый»? Попробуйте-ка это объяснить в деталях!

Так и в случае создания компьютерных систем для исследования молекул возникает, на первый взгляд, наивный вопрос: что такое молекула и что значит ее исследовать? Надо дать точные определения этих понятий, причем такие, какие могли бы быть совершенно однозначно переведены на математический язык. Ведь компьютер, в принципе, способен оперировать лишь понятиями, которые, как писали еще древние философы, описываются «числом и мерою». Недостаточно ввести в компьютер просто утверждение: «чай горячий». Надо указать, что его температура такая-то или выше такой-то. Казалось бы, с молекулами проблем не должно возникнуть: каждому школьнику известно, что молекулой называется достаточно устойчивая электрически нейтральная мельчайшая частица вещества—носительница его свойств, состоящая из атомных ядер с данными зарядами и массами и соответствующего числа электронов. Вторая часть этого определения легко переводится на компьютерный язык. В самом деле, в компьютер можно ввести группу цифр: 12,12,12,12,12,1,1,1,1,1. Каждая цифра означает массу (в атомных единицах) одного из ядер. Догадливый читатель сразу же воскликнет: все ясно, речь идет о молекуле бензола. Но только ли? Уже давно было подсчитано, что брутто-формуле  $C_6H_6$  могут отвечать 217 так называемых топологических изомеров (частью показанных на рис. 11.2). Некоторые уже обнаружены. Они отмечены на рис. 11.2 годами их открытия. Так что однозначности, как видим, не получается. И это в простом случае  $C_6H_6$ , а если взять, например, тридцатиатомную систему? Здесь уже число изомеров может исчисляться десятками тысяч. Вполне можно считать, что ясность теряется полностью! Да и вообще на таком языке нельзя обсуждать столь сложный объект, как молекулы, особенно крупные. Химики давно обнаружили, что если знать только атомный состав регистрируемых молекул (брутто-формулы), то прогнозы того, что получится в результате реакций, или того, как поведет себя индивидуальная система в конкретной сложной ситуации, становятся чрезвычайно затруднительными, а то и просто невозможными. Это все равно что пытаться выразить сложную мысль на иностранном языке, обладая знанием лишь десятка слов. Значит, надо расширять богатство и образность языка. В химии это было достигнуто введением в оборот структурных формул, представлений о молекулах как о пространственных фигурах и т. д. Это означало, в сущности, что были введены в употребление разного рода понятия, которые можно отнести к категории физических моделей. Каждая из них может быть вполне строго и однозначно определена набором чисел. Именно благодаря этому можно выделить из всех изомеров желанный и удовлетворить требо-

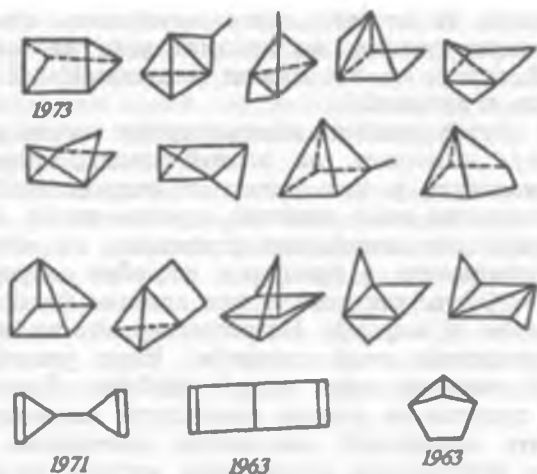


Рис. 11.2. Примеры возможных изомерных структур состава  $C_6H_6$ , с указанием времени открытия или синтеза некоторых из них

ванию о том, чтобы объект был носителем свойств вещества. К этим моделям относятся, во-первых, брутто-формулы (в этой модели указываются лишь составляющие молекулу атомы), во-вторых, формулы, в которых указываются не только атомы, но и отдельные структурные элементы (здесь уже надо частично указать какой атом с каким соединен химической связью), в-третьих, полные структурные формулы, математически описываемые так называемыми матрицами смежности, содержащими всю информацию о связях между атомами, в-четвертых, упругие пространственные модели, в которых допускаются колебания атомов около положения равновесия, в-пятых, модели, в которых атомам—материальным точкам приписываются определенные заряды, и ряд других.

Можно сказать, что понятия, характеризующие те или иные модели составляют лексику языка, на котором и ведется обсуждение тех объектов, которые мы называем молекулами. Вот здесь самое время вернуться к принципу дополнительности. Помните? Дополнительной к истине является ясность. Если под истиной понимать то, что молекула есть устойчивая система, состоящая из электронов и ядер, а это действительно так, то, высказывая эту истину, мы полностью теряем ясность. Чтобы получить эту ясность, надо отойти от истины и оперировать моделями молекул, ни одна из которых истиной не является, а отражает одно из удобных и приближенных представлений об объекте. Что-то напоминающее представление слепых о слоне в известной притче о том, как представили себе слона четверо слепых, когда им дали возможность ощупать части его тела. Одному из слепых достался хобот. Он сказал, что слон похож на змею. Второй ощупал ногу и сказал, что слон похож на дерево.

Третий, обследовав хвост, решил, что слон похож на палку. Наконец, четвертый, дотянувшись до живота, заключил, что слон похож на стену. Самое любопытное заключается в том, что каждый из них был по-своему прав.

На пути использования моделей достигается требуемая ясность, но теряется истина в том смысле, что реально существующему объекту—молекуле уже нельзя дать какое-нибудь одно строгое и информативное определение. Снова обратимся к Н. Бору, написавшему: «...мы должны быть готовы к тому, что всестороннее освещение одного и того же предмета может потребовать различных точек зрения, препятствующих однозначному описанию. Строго говоря, глубокий анализ любого понятия и его непосредственное применение взаимно исключают друг друга».

Нельзя, однако, отсюда делать вывод, что самое общее определение молекулы бесполезно: оно как бы связывает все остальные представления, не давая им «рассыпаться». Одно, таким образом, дополняет другое, причем стремление к ясности одновременно приводит к невозможности дать пригодное для компьютерного использования единственное и строгое определение молекуле как материальному объекту.

А теперь, забегая несколько вперед, перенесемся в область истории. Факты истории есть истина и в этом смысле можно сказать, что полка с архивными папками есть концентрат этой истины. Но ведь собрание архивных документов не есть «История государства Российского»! Чтобы достичь ясности, надо эти архивы обработать и дополнить живой человеческой мыслью (вспомним высказывание П. Чаадаева). Стало быть, чтобы создать целостное представление об истории надо опять-таки удовлетворить двум условиям: с одной стороны, опереться на факты, а с другой, группируя, осмысливая их, прослеживая взаимосвязь, попытаться достичь ясности. При этом чрезмерное увлечение чисто фактической стороной дела с неизбежностью «засушит» любой труд по истории и сделает его малопонятным, а полет фантазии при обработке фактов может привести к историческому роману. Вернемся к молекулам. Выше было сказано, что для работы с компьютером надо дать ответ на два вопроса: что такое молекула и что значит ее исследовать? Оказалось, что ответ на первый вопрос достаточно неопределен, но, как ни странно, это не мешает вполне точно ответить на второй. Ответ будет следующим: исследовать молекулу—это значит построить на количественном уровне совокупность ее моделей разного уровня иерархии; полнота исследований характеризуется степенью сложности и информативности моделей, параметры которых и подлежат определению в результате подходящих экспериментов и последующей обработки результатов измерений. Вот на таком языке уже можно объясняться с компьютером, и он все поймет.

Итак, с строгой математической (логической) точки зрения, единственно понятной компьютеру, исследовать молекулу—это



значит найти численные значения параметров, характеризующих ту или иную модель. Но ведь вообще не существует методов непосредственного измерения, например, длин связей или зарядов на атомах молекулы. Можно измерить спектры молекул, наблюдать дифракционную картину при рассеянии электронов на молекулах и т. д. Другими словами, информацию о численных значениях параметров молекулярных моделей приходится получать только на основании не прямых (как измерение стола линейкой, например), а косвенных наблюдений. Это, в свою очередь, возможно только тогда, когда установлена физическая связь между моделью и ее проявлением (отображением) на множестве тех величин, которые уже поддаются непосредственному измерению. Если обратиться к спектральному анализу молекул, то это означает, что должна быть установлена связь между, например, значением упругости валентного угла и положением частот полос поглощения в инфракрасном спектре.

Реальное положение дел таково, что при исследовании молекул приходится иметь дело с очень сложными соотношениями между параметрами и измеряемыми величинами, требующими при построении соответствующих отображений проведения громоздких вычислений. Вот здесь компьютеры оказываются незаменимыми помощниками.

Варьируя значения параметров модели, можно сблизить, например, ее спектральное отображение с экспериментом и, тем самым, найти величины искомых параметров. Вот на этом пути и подстерегает нас скрытый, но очень опасный подводный камень. Чтобы понять, в чем опасность, представим себе, что мы выбрали модель молекулы, в которой все ядра атомов находятся в определенных фиксированных положениях. Выберем эти положения такими, какие получаются в результате эксперимента по дифракции электронов на молекулах. Решим теперь с любой степенью строгости математическую задачу о движении всех электронов молекулы в поле неподвижных положительно заряженных ядер. В результате мы получим значения электронных уровней энергии и, при желании, вероятностей переходов между ними, т. е. теоретический электронный спектр (он обычно наблюдается в видимой и ультрафиолетовой областях), который можно сопоставить с экспериментом. Совпадут ли такие два спектра? Заведомо можно сказать, что нет. Конкретные расчеты это подтверждают. Дело не в том, что мы плохо решили задачу, а в том, что в реальной молекуле ядра атомов тоже движутся (колеблются около положений равновесия), а в модели мы такое движение отбросили. Поскольку же любая из типичных моделей молекулы не является истинной, а отражает лишь какую-то одну сторону объекта (вроде мнения слепых о слоне), то понятно, что ни одно из получаемых отображений абсолютно с экспериментом совпадать не может.

Все ясно, скажет читатель. Вот мы и возвращаемся на круги своя. Надо брать быка за рога и решать (хотя это и очень

трудно) просто задачу о движениях связанных в общую молекулярную систему электронов и ядер. Если мы все взаимодействия учтем правильно, то теоретическая спектральная картина должна абсолютно совпадать с экспериментальной. Это лишь на данном этапе нужны модели, потому что теория недоразвита и компьютеры слабые, а в будущем... Не торопись, читатель! Если и проделать такую работу, то сопоставить теоретический спектр с реальным экспериментом будет просто практически невозможно по той простой причине, что теоретический спектр отразит всю гамму возможных топологических изомеров (причем их спектры наверняка во многих областях наложатся друг на друга), а в эксперименте мы всегда имеем дело с одним либо очень небольшим числом изомеров. Значит, в соответствующем теоретическом построении надо с самого начала ввести ограничения на возможные движения ядер, чтобы, например, бензол остался бензолом, а не превратился в призманин. В этом случае, однако, соответствующая модель уже не будет отвечать истине. Снова в погоне за усилением одного качества мы безнадежно теряем в другом. Никуда не денешься и никакие теоретические ухищрения и суперкомпьютеры не спасут. Как нельзя бесконечно увеличивать точность определения положения микрообъекта, не потеряв полностью информацию о его импульсе, так и необходимы модели молекул для достижения ясности. Нечего стремиться к истине. Ничего хорошего все равно не получится. С этим надо мириться и все тут!

Итак, мы приходим к важнейшим выводам:

1) математически строгое спектральное отображение точной модели многоатомной молекулы, отвечающее определению этого объекта как стабильной системы из ядер и электронов, невозможно сопоставить с экспериментом;

2) спектральные отображения любых вводимых для достижения ясности моделей молекул могут оказаться лишь в той или иной мере близкими к эксперименту, но никогда не совпадут с ним.

Теперь, когда мы затронули много далеко не тривиальных проблем, самое время поговорить о том, а зачем вообще ясность нужна и что под этим термином на самом деле скрывается. Ведь одного приближенного представления о том, что для понимания некоторых свойств молекулы как системы удобно представить ее в форме жесткой пространственной фигуры, еще совершенно недостаточно для того, чтобы эта модель была пригодной для науки. Ведь цель науки заключается не только в достижении понимания чего бы то ни было, но, и это главное, чтобы такое понимание приводило к возможности прогноза. То, что лунные затмения вызываются тенью Земли, знали египетские жрецы еще в глубокой древности. Стало быть, ясность была. А вот умения предсказать это явление с точностью до минут, насколько известно авторам, не было. Это уже достижение более позднего времени.

Так и при изучении мира молекул. Не сама по себе жесткая пространственная модель молекулы столь уж важна, а то, что дополнительно установлено, что ее отдельные части сохраняются при переносе небольших групп атомов из молекулы в молекулу. А вот это уже дает способ быстро предсказывать пространственную структуру новой молекулы, если известны свойства отдельных ее «кирпичиков» и правила их сложения.

Соображения подобного сорта приводят к новому принципиальному требованию к любым молекулярным моделям: их параметры должны обладать так называемым свойством переносимости, т. е. быть локальными и слабо изменяться в ряду молекул, содержащих повторяющиеся атомные группировки.

К счастью, в процессе эволюции химического знания были отобраны такие модели молекул, которые, обладая, с одной стороны, наглядностью, что и создает ясность, с другой стороны, включают параметры, знание которых позволяет осуществить прогноз свойств новых соединений и, строя соответствующие спектральные отображения, идентифицировать эти соединения и исследовать их весьма подробно.

Стало быть, возникает новая глобальная проблема: как определить значения параметров моделей молекул, опираясь, как уже отмечалось, на косвенные эксперименты.

Чтобы немножко отвлечь читателя от сложных рассуждений, напомним историю о Ходже Насреддине.

Однажды Ходжа Насреддин исполнял обязанности судьи. К нему обратились два человека с просьбой рассудить их. Сначала истец изложил свою версию происшедшего. Ходжа его выслушал и сказал, что он безусловно прав. Тогда выступил ответчик. Ходжа и его выслушал очень внимательно, а потом заявил, что ответчик, разумеется, прав! Тогда один из посторонних слушателей, пораженный таким странным судом, обратился к Ходже и сказал: «Как же так, Ходжа, не может же быть, чтобы люди, утверждающие прямо противоположное, оба оказались правыми!» Ходжа и его спокойно выслушал и потом промолвил: «А ты тоже прав!»

### 11.3. Обратные задачи

При решении задач определения параметров модели по косвенным экспериментам мы встречаемся с еще одним весьма общим положением и попадаем в ситуацию Ходжи Насреддина, одновременно ставя под сомнение безусловную выполнимость главного принципа и гордости науки, а именно принципа воспроизводимости как основы доказательности. Этот научный принцип четко сформулирован О.Писаржевским: «Каждый может воспроизвести описанный в любой научной работе результат, и тот должен подтвердиться».

Класс математических задач, при решении которых восстанавливаются параметры моделей на основании косвенных измерений, получил название *обратных задач*.

Ввиду крайней важности понятия об обратных задачах не только в современной физике, но и во всей теории познания окружающего нас мира вообще (мы уже упоминали об этих задачах во Введении) остановимся на некоторых принципиальных вопросах решения этих задач подробнее.

Рассмотрим вначале предельно простой случай. Пусть некоторая измеряемая величина  $y$  связана с другой измеряемой величиной  $x$  соотношением  $y = a + bx$ , где  $a$  и  $b$  — подлежащие определению параметры модели явления. Пусть, далее, мы произвели ряд измерений величины  $y$  при заданных значениях  $x$ . Это даст совокупность пар чисел  $x_1$  и  $y_1$ ,  $x_2$  и  $y_2$ ,  $x_3$  и  $y_3$ , и т. д. Зная эти пары чисел, составим уравнения:

$$y_1 = a + bx_1,$$

$$y_2 = a + bx_2,$$

$$y_3 = a + bx_3.$$

Сразу бросается в глаза, что параметры  $a$  и  $b$  можно определить, если выбрать любую пару уравнений:

$$\begin{cases} y_1 = a + bx_1 \\ y_2 = a + bx_2 \end{cases}, \quad \begin{cases} y_1 = a + bx_1 \\ y_3 = a + bx_3 \end{cases}, \quad \begin{cases} y_2 = a + bx_2 \\ y_3 = a + bx_3 \end{cases} \quad \text{и т. д.}$$

Все решения этих уравнений были бы совершенно одинаковыми, если бы, во-первых, модель явления, устанавливающая связь между величинами  $x_k$  и  $y_k$ , была бы абсолютно точной, а, во-вторых, если бы абсолютно точными были бы сами измерения характеристик  $x$  и  $y$ . На самом деле эти два условия практически никогда не выполняются. Любая модель потому и называется моделью, что в ней чем-то пренебрегают сознательно или бессознательно (просто потому, что не все влияющие факторы известны) по сравнению с реальным положением дел. Например, при анализе результатов механических экспериментов, производимых на лабораторном столе, почти всегда не учитывается вращение Земли вокруг своей оси. Далее, ни одно измерение нельзя сделать без так называемых случайных ошибок. Все это приводит к тому, что при выборе для построения уравнений разных сочетаний пар величин  $x_k$  и  $y_k$  будут получаться и разные значения искомых параметров  $a$  и  $b$ . Ни одному из таких частных значений  $a$  и  $b$  нельзя отдать предпочтение. Для того, чтобы в этих условиях сделать правильный вывод о параметрах  $a$  и  $b$ , надо учесть все возможные сочетания пар измеряемых характеристик  $x_k$  и  $y_k$  и найти такие

параметры  $a$  и  $b$ , которые наилучшим образом удовлетворяют всем измеренным величинам  $x_k$  и  $y_k$ .

Сразу же возникает вопрос: а что значит «наилучшим образом»? Ясно, что в этом требовании уже содержится некоторая условность. В физике в подавляющем числе случаев в качестве такой меры используется введенный Гауссом так называемый *принцип наименьших квадратов*. Согласно этому принципу, наилучшими признаются такие значения параметров  $a$  и  $b$ , при которых становится минимальной сумма вида  $F =$

$$= \sum_k W_k (y_k - y_{k \text{ выч}})^2. \text{ Здесь } y_k \text{ — измеренная характеристика,}$$

$y_{k \text{ выч}}$  — та же самая характеристика, но вычисленная при значении  $x_k$  и «наилучших» параметрах  $a$  и  $b$ ,  $W_k$  — так называемый статистический вес одного измерения, отражающий взгляд экспериментатора на значимость конкретного опыта. Обычно  $0 \leq W_k \leq 1$ . Ясно, что функция  $F$  есть функция параметров  $a$  и  $b$  и будет непрерывно меняться при их изменении. «Наилучшие» значения параметров найдутся с помощью решения системы уравнений  $\partial F / \partial a = 0$  и  $\partial F / \partial b = 0$  (условия минимума функции  $F$ ).

Критерий наименьших квадратов вполне обоснован, если, во-первых, возможные погрешности при измерениях  $x_k$  и  $y_k$  подчиняются так называемому нормальному закону распределения случайной величины, а, во-вторых, если число измерений очень велико и значительно превосходит число определяемых параметров модели. Нередко, однако, ни то, ни другое не выполняется на практике. В частности, бывает так, что просто невозможно проделать очень большое число экспериментов при разных значениях  $x_k$ . При таких условиях два разных исследователя, повторяя в разных местах и в разное время эксперименты друг друга, обязательно при обработке результатов измерений придут к несколько различающимся значениям параметров  $a$  и  $b$ . Это различие может еще усиливаться, если каждый из них назначит и разные статистические веса  $W_k$ . Таким образом, в окончательном результате появится известная доля субъективизма.

Выше мы отметили, что критерием «наилучшего» является величина  $F = \sum_k W_k (y_k - y_{k \text{ выч}})^2$ . Однако это не единственная возможность. Можно, хотя это и не так удобно, воспользоваться и функционалом

$$F = \sum_k W_k |y_k - y_{k \text{ выч}}|,$$

т. е. суммой модулей отклонений  $y_{k \text{ выч}}$  от  $y_k$ .

Ситуация еще более усложняется, если связь наблюдаемых величин с параметрами модели нелинейная. Чтобы пояснить, к чему это приводит, рассмотрим еще один пример. Пусть имеется система, состоящая из трех, расположенных на одной линии (на оси  $X$ ) шариков 1, 2 и 3 с массами  $m_1$ ,  $m_2$  и  $m_3$ . Пусть шарики 1 и 2 связаны друг с другом легкой пружиной с упругостью  $k_1$ , а шарики 2 и 3 пружиной с упругостью  $k_2$ .

Такая система может совершать колебания вдоль оси  $X$  с частотами  $\omega_1$  и  $\omega_2$  (собственными частотами). Эти частоты можно экспериментально определить с помощью явления резонанса (см. § 4.4), воздействуя на систему внешней периодической силой, с меняющейся в широком диапазоне частот  $\omega$ . При условии  $\omega \approx \omega_1$  или  $\omega \approx \omega_2$  будут наблюдаться резонансы. Задача заключается в том, чтобы, зная  $\omega_1$  и  $\omega_2$ , найти упругости  $k_1$  и  $k_2$ .

Не вдаваясь в подробности соответствующих выкладок, укажем, что между величинами  $k_1$  и  $k_2$ ,  $\omega_1$  и  $\omega_2$  имеют место соотношения:

$$k_1 a + k_2 b = \omega_1^2 + \omega_2^2, \quad k_1 k_2 c = \omega_1^2 \omega_2^2.$$

Здесь  $a$ ,  $b$  и  $c$ —некоторые постоянные, зависящие от масс шариков коэффициенты.

Отсюда получаем, что

$$k_1 a + \frac{\omega_1^2 \omega_2^2 b}{k_1 c} = \omega_1^2 + \omega_2^2,$$

или

$$k_1^2 a c - (\omega_1^2 + \omega_2^2) c k_1 + \omega_1^2 \omega_2^2 b = 0.$$

Таким образом, упругость  $k_1$  находится как корень квадратного уравнения. Формально решения этого уравнения могут быть комплексно-сопряженными. В реальных задачах такой случай не реализуется, и мы его рассматривать не будем. Вполне может получиться, однако, что для корней  $k_1^{(1)}$  и  $k_1^{(2)}$  будут иметь место соотношения  $k_1^{(1)} > 0$ ;  $k_1^{(2)} < 0$  или  $k_1^{(1)} > 0$ ;  $k_1^{(2)} > 0$ , причем  $k_1^{(1)} \neq k_1^{(2)}$ .

Значение  $k_1^{(2)} < 0$  можно исключить по тем соображениям, что упругость пружины не может быть отрицательной. Однако какой корень выбрать при втором варианте? Очевидно, что этот выбор становится возможным только в том случае, когда с самого начала исследователь располагает дополнительной, не зависящей от эксперимента информацией о возможном интервале значений упругости  $k_1$  ( $k_{1\min} \leq k_1 \leq k_{1\max}$ ).

Полученный результат является частным случаем общего правила: *нелинейные обратные задачи могут иметь множество решений, и для выделения физически значимого надо с самого начала наложить дополнительные условия, которые дают возможность выделить желаемое решение.*

Это наложение условий до начала решения задачи отражает исходный взгляд исследователя на свойства окончательного результата и приводит к неизбежному субъективизму при решении очень многих, если не большинства, встречающихся в физике обратных задач. Этим обратные задачи принципиально отличаются от прямых, всегда имеющих единственное решение, при поиске которого субъективный элемент не присутствует вовсе.

Получающиеся при разной постановке обратных задач решения не следует, однако, рассматривать как противоречащие друг другу или несогласующиеся. Их надо воспринимать как взаимодополняющие, устанавливающие известный предел детальности в познании объекта.

Вернемся к конкретным обратным задачам определения численных значений параметров молекулярных моделей: длин связей и величин валентных углов, упругостей связей и т. д. Большинство из них, за исключением простых случаев, решаются по одной и той же схеме. Выбирается некоторое начальное значение параметров, строится соответствующее отображение модели на множестве измеряемых величин и вычисляется (обычно) сумма квадратов отклонений экспериментальных и вычисленных величин (*квадратов невязок*), т. е. величина

$$F = \sum_k (x_{k\text{эксп}} - x_{k\text{теор}})^2,$$

где  $x$ —значения измеряемых характеристик. Затем с помощью вариации параметров итерационным путем производится сближение теоретических характеристик с аналогичными экспериментальными или минимизация  $F$ . Уже на первом этапе при этом возникает неопределенность, связанная с выбором сочетаний пар, например, экспериментальных частот максимумов полос (линий) поглощения и вычисленных. Эти пары должен назначить сам исследователь. Запомним это.

Далее, как оказывается, если даже добиться точного обращения в нуль суммы квадратов невязок, т. е. найти некоторое решение обратной задачи, то оно не будет единственным и часто

не будет иметь физического смысла, если опять-таки с самого начала не наложить некоторые условия, выделяющие физически значимое решение из всего множества допустимых. Математически это означает, что надо от минимума величины  $F$  перейти к минимизации функции  $\Phi = F + \varphi$ . Здесь  $\varphi$  — так называемая штрафная функция, ограничивающая возможную вариацию параметров. Не касаясь других подробностей, обратим внимание на то, что наложение подобных условий означает фактически известное навязывание исследователем своего представления об объекте еще до того, как этот объект начал изучаться. Отмеченные выше общие соображения, а также накопленный обширный конкретный опыт приводят в результате к важнейшему следствию: *параметры молекулярных моделей находятся лишь с известной долей субъективизма*. В этом смысле можно сказать, что ни одна из молекулярных моделей для конкретной сложной системы не может быть построена полностью объективно. Это означает, что два исследователя, располагая даже одним и тем же экспериментальным материалом и строя молекулярные модели одного и того же типа, почти обязательно, работая независимо, разойдутся в деталях, потому что при постановке обратной задачи и задании штрафной функции  $\varphi$  они будут опираться каждый на свое субъективное представление о том, какими свойствами в конечном счете должен обладать анализируемый объект. Вот здесь независимый судья и может сказать, что, по-своему, оба правы, и тоже будет прав! Принцип воспроизводимости, согласно которому любой научный результат должен абсолютно (с точностью лишь до погрешностей эксперимента) воспроизводиться разными независимыми исследователями и только тогда вывод считается доказанным, оказывается, таким образом, нарушенным. Конечно, возникает вопрос: это недостаток наших знаний и порок подхода или опять какой-то знакомый незнакомец?

Вспомним, каким многоликим оказался принцип дополнительности. Внимательное рассмотрение показывает, что мы снова имеем аналогию с физическим выводом о воздействии прибора на анализируемый объект. Выше уже упоминалось о том, что при исследовании микромира было обнаружено, что любая попытка экспериментально проанализировать состояние микрообъекта приводит к изменению его местоположения (координаты), импульса или энергии. А вот что написал специалист в области восточной литературы А. Сыркин: «Опыт показывает, что один и тот же материал эмпирического наблюдения... доставляет обычно широкое поле для самых разных толкований в зависимости от установки исследователя. Применив здесь аналогию из области естественных наук, можно ... сказать, что то «возмущение» прибором исследуемого объекта, которое установлено и применительно к условиям физического эксперимента, действует и в данной области (гуманитарной), проявляясь, в частности, в сознательном и бессознательном отборе иссле-



дователем отдельных элементов и атрибутов объекта. Более того,... объект изучения в свою очередь воздействует на «прибор», определяя его установку. Вряд ли имеет смысл говорить о принципиальной преодолемости этого двустороннего «возмущения» — целесообразнее допустить определенную взаимодополняемость отдельных описаний». Особенно ярко этот факт своеобразного воздействия исследователя (будь то его мозг или прибор) на объект проявляется именно в гуманитарных науках. Когда два известных историка — Е. Тарле и А. Манфред — пишут о Наполеоне, то они опираются практически на один и тот же фактический материал. Однако, стремясь к ясности и создавая образ Наполеона, они по-разному, подчиняясь во многом своей внутренней установке, отбирают и группируют факты, навязывая, в хорошем смысле этого слова, конечно, читателю свою концепцию объяснения исторического явления. Не правда ли, в этом есть немало общего с решением обратных задач в теории молекул. Снова мы оказываемся в положении, когда глупо ставить вопрос о том, кто прав: Тарле или Манфред! Оба, по-своему, правы и их взгляды дополняют друг друга.

Жизнь сложная штука, и, может быть, поэтому Иисус Христос ничего не ответил на вопрос Понтия Пилата: «Что есть истина?» Внимательное рассмотрение показывает, что известная неопределенность в постановке и выводах, привычная для гуманитарных наук, в гораздо меньшей степени проявляется и в естественных науках, особенно в тех случаях, когда объект описывается множеством признаков или параметров (как в молекулярных моделях, например). Человеческий мозг, особенно если не существует строгих критериев отбора, не может охватить сразу все многообразие признаков и четко ограничить их численные значения, что и приводит, при необходимости решать обратные задачи, к разбросу в их постановке и, стало быть, к субъективизму в окончательном результате. Заметим, что обсуждаемая неоднозначность решения обратных задач имеет, если так можно сказать, не математическую (как раз математически обратные задачи, после того как они сформулированы, т. е. задана минимизируемая функция как сумма квадратов невязок, начальные значения параметров и штрафная функция, отражающая взгляд исследователя на то, каким образом должно быть выделено физически значимое решение, решаются вполне однозначно), а «постановочную» причину. Разница между историей и теорией молекул здесь количественная, но не качественная. Тем не менее предел объективному познанию устанавливается и там, и тут.

Конечно, было бы большой ошибкой преувеличивать наличие субъективизма при решении конкретных задач в естественных науках, в частности в физике, химии и биологии. Тем более нельзя скатываться на позицию, выражаемую словами: «Что хотим, то и получим!» Подавляющее большинство фундаментальных положений и выводов в этих науках является следствиями

либо многократно разными способами поставленных экспериментов, либо применения хорошо обоснованных теорий. Например, структурные формулы молекул устанавливают, как правило, вполне однозначно с помощью серии соответствующим образом подобранных химических экспериментов или спектральными методами в диапазоне от ультрафиолета до радиоволн. В объективности представлений о молекулах как о некоторых пространственных геометрических структурах убеждают нас не только тщательно разработанная теория, но и прямые эксперименты по дифракции рентгеновских лучей на молекулярных кристаллах. Сам факт получения достаточно четкой дифракционной картины возможен только тогда, когда имеется некоторое подобие устойчивой во времени «дифракционной решетки». Спектры поглощения в инфракрасной области могут появиться только при наличии колебаний атомов около положений равновесия и т.д. Используемые при решении обратных задач априорные сведения об упругостях химических связей черпаются не только из спектральных, но и из чисто химических экспериментов. Например, давно уже были введены в науку понятия об одинарной, двойной и тройной связях между атомами углерода в углеводородах и было выяснено, что они обладают разной прочностью и т. д.

Таким образом, речь не идет, как правило, о значительном произволе при выборе начальных условий для постановки тех или иных обратных задач и ограничивающих возможные вариации параметров моделей штрафов (обилие самых разнородных сведений, возможность проведения прямых квантовых расчетов, всякого рода аналогии и т.д. резко ограничивают возможность фантазии авторов), а лишь о неизбежной, хотя и небольшой, нечеткости в деталях постановки этих задач и в «размазанности» результата.

Можно поэтому говорить о появлении, наряду с привычными погрешностями собственно эксперимента или округлениями чисел при расчетах, еще одной причины, ограничивающей возможность точного определения параметров тех или иных моделей. Соответствующая погрешность в конкретных случаях может оказаться значительно меньше собственно погрешностей эксперимента. Тогда можно констатировать, что получается вполне объективный результат. Нередко, однако, оказывается (это типично как раз при детальном исследовании молекул спектральными и дифракционными методами), что такая погрешность сравнима или даже превосходит погрешность эксперимента. Тогда, конечно, субъективным фактором пренебрегать нельзя.

В нашу задачу не входит детальный анализ этих важных вопросов, тем более изучение конкретных случаев. Нам сейчас важно только понять и принять во внимание, что субъективный фактор, ограничивающий объективное познание объектов и явлений окружающей среды (сами объекты и явления, конечно, вполне объективны и от нас с вами не зависят), zaczyna

проявляться, хотя и в небольшой степени, уже в одной из точнейших и объективнейших наук.

Таким образом, в этом смысле между естественными и гуманитарными науками нет непроходимой грани: в смысле общих закономерностей формирования их выводов обычное разделение этих наук на две крупные группы отражает не качественную, а количественную сторону. Именно поэтому некоторые самые общие заключения, первоначально сформулированные в естественных науках, оказываются верными и для гуманитарных, и наоборот. То, что, например, принцип дополнительности первоначально был сформулирован в физике, совершенно не означает, что о нем нельзя было догадаться, изучая психологию. Причина здесь в том, что физика как наука об относительно простых явлениях, не сопоставимых, например, с явлениями истории или психологии, позволяет заметить нечто важное и общее на значительно более простых и наглядных примерах. В этом и заключается ее всеобщее значение.

Остановимся теперь на некоторых важнейших смежных вопросах, имеющих громадное значение для понимания вообще основ формирования наших представлений об окружающем мире и отдельных аспектов взаимоотношений между членами человеческого общества.

## 11.4. Объективное и субъективное в науке, или о принципе доказательности

В предыдущем параграфе мы уже коснулись вопроса об объективном и субъективном при формировании представлений об окружающем нас мире в естественных науках и гуманитарных. Вернемся к этой проблеме еще раз, а также обсудим некоторые следствия.

Прежде всего отметим, что в основе физики как науки лежат некоторые фундаментальные принципы. Физика прежде всего является экспериментальной наукой и в своем развитии отправляется от так называемых экспериментальных фактов. В свою очередь, в качестве факта, на основе которого в дальнейшем можно строить те или иные теоретические выводы, признается лишь такое утверждение, такой экспериментальный результат, который воспроизведен в ряде независимых лабораторий и разными людьми. Конечно, не все можно воспроизвести в лаборатории. Например, шаровую молнию пока получить не удалось. Ее, однако, многие реально наблюдали и одинаковым образом описывают. Здесь, в отличие от рассказов об инопланетянах, существенных расхождений в показаниях свидетелей нет.

Важнейшие результаты не только воспроизводятся, но и проверяются по-разному поставленными опытами. Так, например, существовали ранее и существуют сейчас разнообразные способы измерения скорости света. Вошедший в науку под именем Кулона закон взаимодействия двух электрических зарядов при-

дизительно в одно время был открыт Кулоном на основе экспериментов с крутильными весами и Кавендишем с помощью наблюдения электрического поля на внутренней стороне полого заряженного шара. Слава, разумеется, достается первооткрывателям (точнее тем, кто первым опубликовал результаты), но, разумеется, это не означает, что, если бы кто-нибудь вместо Кулона провел бы такой же эксперимент, то он получил бы что-то иное. В таких случаях говорят, что результат объективен, отражает реальную, не зависящую от нас действительность и в этом смысле от личности собственно ученого не зависит. Если один ученый хочет убедить другого ученого в своей правоте, то он должен представить ему возможность полностью воспроизвести свой эксперимент. Это требование закреплено в правиле научных публикаций: хорошая публикация не только должна давать изложение интересного и нового результата, но и содержать столь подробное описание эксперимента, чтобы другой исследователь понял, как его можно воспроизвести, и смог сделать это. Конечно, при этом подразумевается, что сам ученый не воздействует сознательно на измерительный прибор, т. е., например, не подталкивает в крутильных весах заряженные шарики пальцами. Если он даже и подробно опишет, как он это сделал, и другой ученый все это повторит, все равно в качестве физического факта результат признан не будет.

Физика как наука, однако, состоит не только из экспериментальных фактов. Есть и вторая, дополняющая первую сторона — обобщение фактов или построение разного рода теорий.

Подобно тому, как история не сводится к архивной полке, так и физика как наука не сводится к перечислению экспериментальных результатов. Отмеченная П. Чаадаевым «мысль» истории появляется и в физике в виде разного рода теорий и формул, устанавливающих связь между явлениями и измеряемыми величинами. Не трудно заметить и субъективизм в изложении физики как науки, особенно бросающийся в глаза в учебной литературе. Подобно Е. Тарле и А. Манфреду, пишущим о Наполеоне, Р. Фейнман и Д. Сивухин создали несоответствующие курсы физики. Конечно, ни тот ни другой не отклонялись от экспериментальных факторов, не вводили непроверенных представлений, однако каждый из них по-своему отбирал материал, группировал его, расставлял акценты, выделял главное и второстепенное, сообразуясь со своими субъективными соображениями.

Но вернемся к физическим теориям и методам доказательства. Физическая теория проявляется либо как результат обобщения большого числа фактов, либо как результат интуитивного скачка. Однако и в том, и в другом случаях теория признается научным сообществом правильной только тогда, когда ее построение подчиняется строгим требованиям так называемой формальной логики (см. ниже) либо когда строгие логические выводы из интуитивно высказанного утверждения приводят к полностью

согласующимся с многочисленными экспериментами результатам. Так, например, закон Кулона является прямым следствием формально-логической обработки результатов экспериментов с крутильными весами. Второй же закон Ньютона «вывести» на основании какого бы то ни было одиночного или нескольких экспериментов вообще нельзя. Тем не менее, невероятное множество полученных из него строго логическим путем совпадающих с экспериментами следствий убеждают всех в его правильности и адекватности природе.

Заметим, что под формальной логикой понимается (основы ее были заложены еще Аристотелем, почему эта логика и называется аристотелевой) оперирование с такими утверждениями, в отношении которых можно сказать, что это или «истина», или «ложь» (0 или 1 при работе с компьютерами). Соответствующие правила формальной логики четко определены в так называемой булевой алгебре, возникшей в прошлом веке, но только в этом веке в связи с развитием ЭВМ получившей широкое распространение.

Именно по правилам формальной логики доказываются все теоремы, например, известные из школьного курса алгебры, геометрии и тригонометрии. Не вдаваясь в детали этой сложной проблемы, напомним теперь известный, по-видимому, подавляющему числу читателей факт, что убедить другого человека принять свою точку зрения легче всего в двух случаях: когда другому человеку предоставлена возможность увидеть и измерить то, что видел и измерял первый (требование воспроизведения эксперимента), и когда рассуждения первого отличаются строгой логичностью (это, в сущности, тоже подразумевает возможность их абсолютного воспроизведения, если оба субъекта пользуются одними и теми же правилами построения вывода из исходных посылок). Логический метод получения заключений общепринят в математике и теоретической физике и получил название аксиоматического. Ярким примером этого является евклидова геометрия, базирующаяся, как всем хорошо известно, на исходных пяти постулатах. Можно, конечно, изменить эту систему постулатов, что и сделал Н. Лобачевский, развивший так называемую сферическую геометрию. Однако при получении теорем этой геометрии снова применялся метод формальной логики. Как в случае воспроизведения экспериментов, так и при построении формально-логических выводов разные исследователи обязаны прийти к одинаковым результатам, даже видоизменяя эксперимент или следуя другим путем при доказательстве теорем, что и позволяет считать их окончательные заключения объективными. Поскольку именно такие приемы и являются характерными не только для физики, но и вообще для группы так называемых естественных наук, то и возникло стойкое убеждение в том, что выводы этих наук лишены субъективной окраски, в отличие, например, от истории, где субъективный фактор играет очень заметную роль. Именно поэтому считается,

что естественные науки базируются на принципе строгой доказательности (как сказал Л. Берг, «метод науки—это доказательство»). Это, в свою очередь, означает, что только одно мнение может быть правильным. В то же время в медицине и гуманитарных науках принцип доказательности имеет ограниченное хождение и большую роль играют субъективные взгляды исследователя.

Разумеется, и в гуманитарных науках логика применяется очень широко, но, наряду с ней, для убеждения других ученых в своей правоте и привлечения их на свою сторону в неменьшей степени используется и играет громадную роль воздействие на чувственное восприятие, что достигается литературной формой изложения, умело подобранными наглядными примерами и т. д. «Мысль» и «чувство» при формировании представлений об окружающем нас мире отчетливо выступают здесь не как взаимоисключающие, а как взаимодополняющие стороны одного и того же действия.

Мы видели, однако, что субъективный элемент не чужд и естественным наукам и обнаруживается в них при переходе от более простых объектов к более сложным (от атомов к молекулам и т. д.). Этот элемент возрастает в химии, далее в биологии и т. д.

Если ряд наук расположить в следующем порядке: математика, физика, химия, биология, медицина, психология, науки об обществе, философия, то нетрудно видеть, как от математики к философии нарастает субъективный фактор и уменьшается степень доказательности выводов автора.

На крайнем правом фланге этого ряда наук можно расположить и религию. Давно известно, что в сфере разума, т. е. логическим путем, ни доказать, ни опровергнуть религиозное учение нельзя. Принятие или непринятие человеком христианства или мусульманства есть результат его внутреннего чувства, а не логических доводов. Вспомним разговор Берлиоза и Воланда о Христе в романе «Мастер и Маргарита» М. Булгакова.

*Берлиоз:* Но требуется же какое-нибудь доказательство...

*Воланд:* И доказательств никаких не требуется... Все просто: в белом плаще...

Аксиоматическая математика и религия представляют две крайности. Для нас сейчас важно усвоить, что между естественными и гуманитарными науками нельзя провести четкой грани: «мысль» и «чувство» (воздействие исследователя) как средство убеждения присутствуют и в естественных науках, даже в таких, как физика, но, конечно, в гораздо меньшей степени. Это означает, что свойственное многим ученым-естественникам убеждение в том, что в конечном счете только один вывод, одно мнение есть истина, не является абсолютно правильным. К сожалению, такой взгляд стал превалирующим в нашем обществе, воспитанном на привате науки, прежде всего естественной, что порождает характерную для нашего общества нетерпимость к

чужим взглядам. Перенос этих же взглядов на науки об обществе и желание и там выделить единственно правильный взгляд при отмеченной выше принципиальной невозможности строго его обосновать (доказать истинность) с неизбежностью приводят к навязыванию этого взгляда силой. Этому мы все были свидетелями в недалеком прошлом со всеми трагическими последствиями для нашей страны. Уроки физики для понимания всего этого оказываются чрезвычайно важными, так как именно эта область науки очень долгое время воспринималась как образец развития наук вообще и как образец идеала объективного познания окружающего нас мира. Не случайно именно примеры из этой области науки так любили приводить философы-материалисты.

Мы видели, однако, что даже в этой науке проявляется в ряде направлений заметный элемент субъективизма, что приводит к появлению выводов, не имеющих права претендовать на некоторую исключительность и преимущество по отношению к аналогичным выводам других авторов. Такие взгляды надо рассматривать не как противоречащие друг другу, а как взаимодополняющие.

Ясно, что тем более нельзя придавать значение абсолютной истины никакому учению о развитии общества. Любая общественно-политическая теория в лучшем случае способна отразить лишь какие-то отдельные стороны сложного явления. Слепая попытка в практической деятельности следовать только одному «учению» с неизбежностью ведет в тупик, что и произошло в нашей стране.

Множественность способов диагностики и лечения в медицине (европейская медицина, китайская, тибетская, иридодиагностика, аллопатия и гомеопатия и т. д.) также является прямым следствием отмеченного выше неизбежного субъективизма при исследовании сложных объектов и принципа дополнительности. Глупо и вредно поэтому разрешать одно направление и запрещать другое, под воздействием моды кидаться из одной крайности в другую, и т. д. А ведь это постоянно происходит в жизни!

Все это показывает, насколько важно в практической жизни человека, в управлении государством и отраслями производства знать и придерживаться соответствующих общих принципов, весьма часто проистекающих из фундаментальных исследований в физике.

## 11.5. Принцип соответствия и особенности построения физических теорий

Все представления об окружающем мире человек получает с помощью своих органов чувств. Опыт общения с природой закреплялся в виде определенных понятий и находил отражение в языке и образах, складывающихся в мозгу человека. Постепен-

но эти понятия в уточненном и формализованном виде переносились в науку, в частности в физику.

Необходимость ориентироваться в окружающем пространстве и как-то характеризовать его привела к понятию длины. Эта характеристика на первых порах опиралась на «естественные» стандарты. Вспомним названия мер длины: локоть, шаг, фут (размер стопы) и др. Человек поднимал и переносил грузы, сравнивал их. Приобретенный при этом опыт был зафиксирован в понятии массы. Перемещение тел толканием или натягиванием канатов естественным образом трансформировалось в обобщенное понятие силы и т. д. Можно поэтому сказать, что вся механика в конечном счете базируется на фундаментальных понятиях, генетически непосредственно связанных с восприятием органов чувств. Несколько сложнее дело обстоит со световыми явлениями. Глаз может отличить темное от светлого, обладает цветным восприятием. Глазом можно зафиксировать чередование темных и светлых полос за щелью в непрозрачном экране, на который падает параллельный пучок монохроматического света. Эта характерная картина, сравнимая с картиной прохождения волн на поверхности жидкости через щель в стенке, поставленной на пути распространения волн, может натолкнуть на аналогию и привести к заключению о том, что и свет представляет собой распространение колебаний чего-то. Однако для того, чтобы это что-то конкретизировать, недостаточно уже только непосредственного чувственного восприятия, а нужно еще, чтобы проникнуть в суть вещей, дополнительное умственное построение, чтобы по косвенным признакам разного рода догадаться, что свет есть электромагнитная волна. Это не так уж просто было сделать, о чем свидетельствует то, что длительное время в науке бытовала теория колебаний эфира. Мы не можем, если так можно сказать, «ухватить руками световую волну», но наблюдать самые разнообразные проявления ее взаимодействия с веществом, с экранами и др. мы вполне в состоянии. То обстоятельство, что такие экспериментальные факты не только качественно, но и количественно объясняются с одной и той же точки зрения на свет как на электромагнитную волну, заставляет признать, что объективно существует такая форма материи. Сказанное относится и к взаимодействию заряженных частиц через поле и т. д.

Все наши представления об окружающем мире можно, таким образом, разделить на две категории, являющиеся следствием непосредственного чувственного восприятия и умственного усилия, результат которого фиксируется в форме разного рода теорий, если говорить о физике. Опять мы имеем дело с двумя взаимно дополняющими сторонами создания образа окружающего нас мира в мозгу. Не случайно, по-видимому, и мозг наш состоит из двух половин, выполняющих разные функции. Встречаясь с определенными представлениями уже на ранних этапах обучения, люди часто настолько сживаются с ними и привыкают ими пользоваться, что перестают задумываться о том,



что, собственно, послужило основанием для формирования того или иного образа. Такое привыкание к понятиям особенно характерно для профессионалов-ученых.

Возвращаясь к представлению о свете как об электромагнитной волне, укажем снова, что оно базируется не на прямом впечатлении от разглядываний колебаний электрического и магнитного векторов со стороны, а на анализе экспериментов, в которых волна всегда с чем-то взаимодействует и что-то меняет так, что соответствующий эффект регистрируется глазом. Можно поэтому сказать, что сами колеблющиеся векторы  $E$  и  $B$  являются ненаблюдаемыми величинами. Правильная физическая теория, однако, должна перевести эти ненаблюдаемые величины в наблюдаемые (почернение на фотопластинке, радужные кольца за отверстием в экране, на который падает белый свет, и т. д.). Не надо конечно отождествлять термин ненаблюдаемое с придуманным, нематериальным. Слепой от рождения человек не видит света и окружающих предметов, однако отсюда не следует, что они отсутствуют.

С еще более сложной ситуацией встретились физики, когда они перешли к изучению элементарных частиц, атомов и др.

Здесь уже пришлось вводить в обращение понятие о волнах вероятности, совсем не связывая их ни с каким видом колеблющейся материи, а пользуясь этим представлением просто как математическим приемом, позволяющим, решая соответствующее уравнение Шредингера, давать прогноз реально наблюдаемым и фиксируемым органами чувств явлениям.

Несмотря на принципиальное различие в этом смысле между теорией электромагнитных волн и волновой механикой, есть одно общее: и в том, и в другом случае в результате соответствующих теоретических построений можно описать опыты, результаты которых могут непосредственно восприниматься органами чувств.

Это требование в физике было закреплено в так называемом *принципе соответствия*. Согласно этому принципу любая физическая теория сейчас и в будущем может оперировать любыми понятиями как имеющими смысл материальных объектов окружающего мира, так и искусственными, абстрактными, пользоваться любой математикой (действительными или комплексными числами и переменными, так называемыми матрицами, операторами и т. д.), но всегда должна иметься возможность с помощью этих понятий и математических преобразований получить такие величины, которые могут быть прямо сопоставлены с реальными, воспринимаемыми органами чувств человека, экспериментами.

Этот важнейший принцип устанавливает, таким образом, связь между фактом физики и ее мыслью.

Продемонстрируем, как этот принцип «работает» в квантовой, или волновой, механике.

В гл. 9 было указано, что стационарное уравнение Шредингера для частицы с одной степенью свободы (движущейся вдоль оси  $X$ ) имеет вид:

$$-\frac{\hbar}{2m} \frac{d^2\psi}{dx^2} + U(x)\psi = E\psi \quad \text{или} \quad \left( -\frac{\hbar}{2m} \frac{d^2}{dx^2} + U(x) \right) \psi = E\psi.$$

Здесь  $\hbar$ —постоянная Планка;  $m$ —масса частицы;  $\frac{d^2}{dx^2}$ —оператор (символ действия на функцию  $\psi$ );  $U(x)$ —потенциальная функция, передающая, например, вид потенциальной ямы, в которой находится частица;  $E$ —подлежащий определению параметр (в математике это называется собственным числом соответствующего дифференциального уравнения);  $\psi$ —подлежащая определению волновая (или собственная) функция. Будем также считать, что движение частицы ограничено интервалом  $l = x_2 - x_1$ .

Постоянная Планка связана с классическим понятием действия. Это так называемый квант действия. Собственно кванта действия нет в классической механике, но логическая связь легко прослеживается. Масса частицы имеет один и тот же смысл что в классике, что в квантовой механике. Операторами в классической механике не пользуются. Это, в общем-то, математическое понятие. «Классическая» и «квантовая» потенциальные функции тождественны. Собственные числа в классической теоретической механике материальных точек встречаются лишь в задачах о колебаниях. В квантовой механике они отождествляются с возможными для частицы в данном потенциальном поле уровнями энергии.

Собственная (или волновая) функция появляется только в квантовой механике, причем она может быть (и очень часто бывает) комплексной функцией. Например, для свободно движущейся вдоль оси  $X$  частицы  $\psi \sim \exp(\pm ix\sqrt{2mE/\hbar^2})$ .

Как видим, при переходе к уравнению Шредингера кое-что сохраняется и от классики, но появляется нечто существенно новое. Это, прежде всего, величины  $E$  (это всегда действительные числа) и, в особенности, волновая функция  $\psi$ . То, что она может быть и комплексной, уже подразумевает, что ни одному материальному объекту она в принципе сопоставлена быть не может. Это означает, что вообще нельзя поставить никакого эксперимента, в котором она наблюдалась бы непосредственно. Бессмысленно и обсуждать вопрос о том, что же именно колеблется, ибо события развиваются не в реальном, а в абстрактном математическом мире.

А теперь посмотрим, как «язык» квантовой механики «переводится» на «язык» реальных экспериментов.

Прежде всего вспомним второй постулат Н. Бора. Согласно этому постулату атом может менять свои энергетические состояния (переходить с уровня на уровень), но при этом всегда будет поглощаться или излучаться энергия в виде квантов электромагнитного поля. Если одновременно имеется много атомов и

излучается много квантов, то можно говорить об электромагнитных волнах (плотность квантов в единице объема при этом сопоставляется с напряженностью электрического поля электромагнитной волны), частоты которых определяются простой формулой  $\omega_{kn} = E_k - E_n/\hbar$ , где  $k$  и  $n$ —номера комбинирующих уровней. Таким образом, через комбинации собственных чисел непосредственно выражаются длины волн спектральных линий, которые можно наблюдать глазом с помощью спектроскопа (различать цвета).

Далее, каждая спектральная линия характеризуется не только фиксированным положением на шкале частот, т. е. значением  $\omega_{kn}$ , но и интенсивностью. В квантовой теории показывается, что эта интенсивность пропорциональна квадрату определенного интеграла (так называемого матричного элемента дипольного момента), который для рассматриваемого случая одной частицы при одномерном движении имеет вид

$$\int_{x_1}^{x_2} \psi_k^*(x) q x \psi_n(x) dx.$$

Здесь  $q$ —заряд частицы. Функция  $\psi^*$  есть комплексно-сопряженная к функции  $\psi$ . Использование такого сочетания функций под знаком интеграла приводит к тому, что он становится действительным и как таковой может быть сопоставлен с реальным экспериментом.

Другой пример. Если бы мы попытались осуществить эксперимент по определению местоположения частицы, то с наблюдаемыми результатами этого эксперимента надо было бы сопоставить не величину  $x$ , т. е. некоторую координату частицы, а снова интеграл

$$\int_{x_1}^{x_2} \psi^*(x) x \psi(x) dx.$$

Именно с математическими выражениями такого типа и приходится иметь дело при интерпретации экспериментов по дифракции рентгеновских лучей на молекулярных кристаллах или при рассеянии электронов на молекулах. С помощью таких экспериментов определяются в молекулах и кристаллах межатомные расстояния или длины связей и валентные углы.

Таких примеров можно было бы привести много. Сейчас нам, однако, важно обратить внимание на то, что волновая функция нигде не выступает «в одиночку», а только в комбинации с  $\psi^*$ . Таким образом, хотя в теории и используется принципиально ненаблюдаемая функция  $\psi$ , однако это не приводит к каким бы то ни было противоречиям со способностями человека к

ощущениям, так как благодаря выполнению требований принципа соответствия, теория позволяет вычислять такие величины, которые уже можно непосредственно сопоставить с реальными экспериментами (добавим, что всегда классическими и макроскопическими).

Важнейшее значение принципа соответствия заключается в том, что он устанавливает связь между математикой, т. е. миром абстракций, и реальным физическим миром. Математика есть плод деятельности человеческого мозга. В ней используется масса понятий (комплексные числа, операторы, матрицы и т. д.), не имеющих отображений в окружающем нас мире. Оказывается, однако, что различные математические разделы постоянно заимствуются из математики и переносятся в физику и тем самым «связываются» с окружающим миром. Так, аппарат обыкновенных дифференциальных уравнений является фундаментом классической механики, уравнения в частных производных применяются в волновой механике, матрицы (таблицы чисел или функций) широко используются в теории строения и спектров молекул, полимеров, кристаллов, операторы играют важнейшую роль в теории электромагнитных явлений и в квантовой механике, криволинейная геометрия Римана составляет математическую основу общей теории относительности и т.д. Все это становится возможным именно благодаря неукоснительному следованию принципу соответствия.

Никто не знает, что еще завтра почерпнет физика из той богатейшей кладовой, которая называется математикой. Впрочем, развитие физики как науки постоянно стимулирует и появление новых направлений в математике. Таким образом, обогащение тут взаимное.

Обсудим теперь чрезвычайно важный вопрос, являющийся непосредственным следствием как принципа дополнительности и необходимости для анализа сложных микрообъектов перехода к модельным представлениям, так и принципа соответствия.

Рассмотрим один частный, но интересный пример. Как известно, размеры ядер составляют величины порядка  $10^{-13}$  см, а длины химических связей имеют порядок  $10^{-8}$  см. Таким образом, размеры ядер на пять порядков меньше типичных межатомных расстояний в молекулах. Ядра поэтому вполне можно считать материальными точками. Представим себе, что мы хотим измерить длину связи или расстояние между двумя материальными точками. В классической физике считается, что точность соответствующего макроскопического измерения ограничивается лишь погрешностью выбранного для этой цели прибора. На само же понятие расстояния между двумя точками никаких ограничений не накладывается. Другими словами, если с учетом ошибок измерения в одном эксперименте мы получим 1,1 м, а в другом 1,12056 м, то мы просто констатируем, что второй эксперимент гораздо точнее первого, но при этом и в том, и в

другом случае у нас не возникает никаких сомнений, относятся ли эти числа к одному и тому же понятию или нет.

Принципиально иная ситуация обнаруживается в квантовой механике. Выше было сказано, что непосредственно с экспериментом в силу принципа соответствия сопоставляется не длина связи как некоторый отрезок прямой, проходящей через две точки, а матричный элемент типа

$$\int_{x_1}^{x_2} \psi^*(x)x\psi(x) dx.$$

Значение этого матричного элемента будет зависеть от вида волновых функций  $\psi^*$  и  $\psi$ , находящихся под знаком интеграла. Вид последних для молекул целиком определяется выбранной для данного конкретного случая моделью молекулы. Так как разные модели реально различаются друг от друга не только на качественном, но и на количественном уровнях (вспомним замечание о решении обратных задач), то ясно, что, даже если при заданных параметрах модели удастся совершенно точно решить соответствующее уравнение Шредингера, окончательное значение матричного элемента будет нести в себе все те неизбежные погрешности, которые вызваны как несовпадением самой модели с «истиной», так и субъективным элементом при уточнении параметров модели.

Отмеченное выше в п. 2 «размывание» определения понятия молекулы, вызванное требованием достижения «ясности» и переходом к модельным представлениям, будет сопровождаться аналогичным «размыванием» понятия длины связи в молекуле.

В результате при разных экспериментах и при использовании разных моделей молекулы мы будем получать разные взаимодополняющие значения длин связей, разброс которых указывает на естественный предел принципиальных возможностей «измерить» длину связи и на то число значащих цифр, ниже которого нельзя опускаться не только из-за ошибок эксперимента или недостатков модели, а еще и потому, что само понятие длины связи начинает терять свою определенность.

Многолетняя практика показывает, что если ограничиться оценкой длины связи с точностью до  $0,05 \cdot 10^{-8}$  см, то самые разные эксперименты приводят практически к совпадающим величинам. Можно поэтому рассматривать полученное число как вполне объективное. При попытках снизить погрешности до  $0,01 \cdot 10^{-8}$  см начинает обнаруживаться уже вполне заметный разброс измеряемой величины в зависимости от отмеченных выше факторов. Найти длину связи в молекуле с точностью, например, до  $0,0001 \cdot 10^{-8}$  см (эта величина еще на порядок превосходит размеры ядер) нельзя не только потому, что нет

подходящих экспериментов, но и потому, что резко начинает нарастать неопределенность самого понятия.

Снова мы «упираемся» в некоторый предел возможности объективного познания.

## 11.6. Научный прогноз и его ограничения. Физика и История

Понимание процессов, происходящих в окружающем мире, и формирование его образа, конечно, очень важно. Однако целью всякой науки является не только понимание, но и, что главное, прогноз событий. Известно, что физика в этом отношении достигла большого совершенства. Именно на базе установленных в физике общих законов и хорошо развитого математического аппарата создавались соответствующие методы расчетов для обслуживания отдельных крупных направлений в технике: теории машин и механизмов, теоретическая электротехника, методы расчетов летательных аппаратов и т. д. Сейчас ни одно из существующих промышленных изделий от кофеварки до космических ракет не реализуется в натуре до тех пор, пока соответствующая теоретическая проработка не покажет с вероятностью минимум 90, что изделие будет работать именно так, как это хочется создателю. Поражающее воображение космические полеты на Марс, Венеру, к дальним планетам Солнечной системы длительностью до нескольких лет с минимальной коррекцией траектории в пути следования можно смело назвать триумфом прогностической способности физического знания.

Характерной особенностью сегодняшнего периода является бурное внедрение расчетных методов для прогнозирования свойств не только макро-, но и микрообъектов: атомов, молекул, полимеров, кристаллов. Например, с хорошей точностью и сравнительно быстро можно рассчитать геометрическую структуру молекул и молекулярных кристаллов, предсказать их спектральные свойства и т. д. При этом такие задачи решаются не только для малых и средних по величине систем, но даже и для белков. Появилась целая новая область науки, получившая название молекулярного дизайна.

В основе всех этих достижений и нашей уверенности в том, что события будут развиваться именно так, как это предсказывает соответствующая теория, лежит уже обсуждавшийся в § 2.3 принцип детерминизма. Этот принцип впервые был сформулирован как обобщение достижений классической механики. Напомним, что согласно ее основным положениям, если известно уравнение движения или система таких уравнений и начальные условия (значения координат и скоростей в начальный момент времени), то, решая это уравнение или систему, можно точно указать, где будет находиться предмет через определенный промежуток времени и как он будет двигаться и деформироваться. В свою очередь, такое уравнение (или систему уравнений)

всегда можно записать, если заданы массы и моменты инерции отдельных компонентов объекта и действующие силы.

Сразу же бросается, конечно, в глаза, что поскольку прогноз будущего состояния системы делается на основе решения соответствующих уравнений, то это подразумевает, что во всем интервале времени, в пределах которого осуществляется прогноз, уравнение движения не меняется. Можно, конечно, снять это ограничение и учесть возможность появления новых факторов, отсутствовавших в начальный момент. Например, принять во внимание, что тело при своем движении должно проходить область с сильным воздействующим на него полем, которое практически не действует на начальных стадиях движения. Это, однако, все равно подразумевает знание такой возможности в тот момент, когда движение только начинается, т. е. с самого начала. Таким образом, оказывается, что все предопределено (события детерминированы) теми рамками, которые могут быть установлены в начальный момент времени.

Громадное число различных экспериментов, убеждающих в том, что сделанные на такой основе прогнозы событий выполняются, отражены в принципе детерминизма, который выражает нашу уверенность в том, что, пользуясь установленными в физике приемами, мы действительно можем предсказать события и пользоваться этим предсказанием в жизни.

Переход от механики Ньютона к механике Эйнштейна (больших скоростей) и к электродинамике ничего в указанном выше смысле не меняет и только может служить дальнейшим подтверждением обоснованности самого принципа.

Мы не будем здесь касаться философского осмысления и толкования принципа детерминизма, чему посвящена многочисленная литература, а остановимся только на том, каков взгляд на этот принцип современной физики.

Выше было сказано, что если задано уравнение движения и начальные условия, то все развитие дальнейших событий уже предопределено. На самом деле все обстоит, конечно, сложнее, поскольку прогнозирование реального эксперимента требует не идеального, а реального знания масс, сил, начальных условий. Получение таких знаний невозможно без соответствующих измерений. Любые реальные измерения имеют ту или иную степень погрешности. Поэтому, строго говоря, не одну из названных величин мы не можем указать точно, а лишь в определенном интервале возможных значений. Отсюда следует, что и прогноз событий точным быть не может. Более того, начальные ошибки при длительном течении времени могут накапливаться, снижая достоверность прогноза. Именно поэтому при полетах космических аппаратов и требуется периодическая коррекция на промежуточных стадиях их движения к желаемой цели.

Далее, нет полной уверенности, что не появится какой-то неожиданный фактор, например столкновение космического аппа-

рата с метеоритом. Все это приводит к тому, что реальный прогноз никогда не является абсолютно достоверным, а имеет вероятностный характер. Правда, вероятность правильного прогноза, как правило, достаточно велика.

Вероятностный элемент усиливается при переходе к микромиру и при использовании квантовой механики. Аналогия с классикой сохраняется в том смысле, что место классических уравнений движения занимает уравнение Шредингера. Его снова можно решить при подходящих начальных условиях (то, что надо задать уже не координаты и скорости, а начальные значения волновой функции не является принципиальным). Однако существенно новое состоит в том, что в результате решения находятся не координаты и скорости движения частей системы, а значения функции  $\psi$  в заданный момент времени. Мы знаем, что сама по себе она не наблюдаема, однако квадрат этой величины, т. е. произведение  $\psi^* \psi$ , передает вероятность пространственного состояния системы в данный момент времени. Эта вероятность уже не связана с тем, хорошо или плохо мы измерили параметры системы (массы, характеристики потенциальной функции), а с существенно квантовой природой объекта.

С учетом этих оговорок прогноз по-прежнему оказывается возможным, и в этом смысле можно говорить, что принцип детерминизма можно распространить и на область квантовых явлений.

Ситуация не меняется и тогда, когда от описания сравнительно несложных объектов мы переходим к объектам множественным, статистическим, например газам. Если изучать интегральные характеристики этих систем (температуры, давления), то снова можно предсказать их изменения, если указано, как должны меняться внешние воздействия и задать начальные параметры  $p, V, T$ . Ничего неожиданного не происходит и в тех случаях, когда от обратимых процессов мы переходим к необратимым, таким как диффузия, передача тепла и другие.

Если внимательно приглядеться ко всем рассмотренным выше случаям, то можно обнаружить нечто общее—отсутствие качественных изменений в изучаемых системах. Во все время, в течение которого делается прогноз, сохраняются классические и квантовые уравнения, газ остается тем же газом и т. д. Можно поэтому сказать, что все изменения в системе носят количественный, но не качественный характер.

Правда, было отмечено, что может быть и такая ситуация, когда в процессе эксперимента в какой-то момент включится новый фактор, отсутствовавший в начальном уравнении движения. Однако и в этом случае прогноз возможен, если появление нового действия заранее также известно.

Но ведь не только с такими ситуациями имеет дело современная физика.

Вернемся к газу, т. е. сложному объекту, состоящему из многих частиц. Состояние таких систем можно описывать



интегральными параметрами (для газа, например, с помощью температуры и давления), т. е. на макроскопическом уровне, и на микроскопическом уровне, измеряя, например, давление и температуру микроприбором в какой-то малой части объема, занимаемого газом. При этом обнаруживается следующее: если газ находится в равновесии, то его интегральные параметры не меняются. Но отсюда совсем не следует, что будут неизменными и показания микроприбора. Напротив, эксперименты показывают, что его отсчеты будут все время хаотически меняться около некоторой средней величины. Почему это происходит, известно: причиной такого «дрожания» стрелки микроприбора являются флуктуации, т. е. случайным образом возникающие около мембраны микроприбора в небольшом объеме сгустки частиц газа или, наоборот, разрежения. Наличие таких флуктуаций является характерной чертой практически всех сложных статистических систем. В состоянии равновесия сложной системы такие флуктуации возникают и распадаются, но в среднем их уровень остается постоянным.

Теперь представим себе, что реальный, а не идеальный газ находится в состоянии, близком к критической точке. Тогда, если в какой-то микрообласти пространства в результате флуктуации оказалось больше частиц газа, чем в среднем, а это означает, что они расположатся друг относительно друга на меньших, чем в среднем, расстояниях, то за счет действия межмолекулярных сил молекулы могут слипнуться между собой, образуя зародыш капли жидкости.

При дальнейшем снижении температуры и уменьшении общего объема, занимаемого газом, процесс этот будет нарастать и постепенно весь газ перейдет в жидкое состояние. Конечный макроскопический результат, таким образом, предсказуем, но если мы попытаемся заранее указать, где именно в пространстве, занимаемом газом, образуется первая микрокапля жидкости, то обнаружим, что это сделать нельзя. Причина этого заключается в том, что процесс образования флуктуаций в газе не является детерминированным в принципе, а представляет классический пример чисто случайных процессов. Такие процессы сейчас можно легко моделировать с помощью так называемого генератора случайных чисел, легко реализуемого с помощью любого компьютера. Примеры работы такого генератора показаны на рис. 11.3, а, б.

Теперь представим себе, что имеется некоторый участок почвы, разделенный на маленькие клетки. Представим себе, что с помощью генератора, который выдает в случайной последовательности номера клеток, они заполняются семенами какого-то растения. При этом часть семян является здоровыми и способными к прорастанию, а часть мертвыми. Выбирать их из мешка будем также случайным образом. Попадание здорового семени в одну из клеток вполне можно рассматривать как флуктуацию. Теперь подождем, когда семена прорастут и из них разовьются

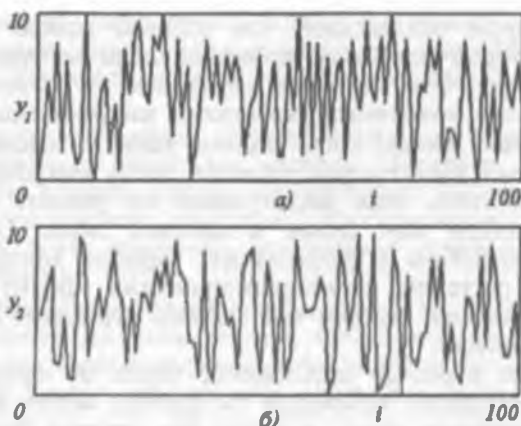


Рис. 11.3. Примеры имитации с помощью компьютера наблюдаемых микроприбором флуктуаций. Генератор случайных чисел выдает 100 чисел в диапазоне 0—10 через один и тот же интервал времени. Приведены результаты двух независимых испытаний. Видно, что «флуктуации» не совпадают между собой, что и должно быть при случайном процессе

растения. Тогда на засеянной площади появится некоторый «узор». Повторим опыт с самого начала. Генератор выдает другие случайные числа. Получится другой «узор» (рис. 11.4, а, б). При многократном повторении опыта все время будут получаться разные «узоры», разные конечные результаты, причем предсказать, когда образуется тот или другой «узор», невозможно в принципе.

Опыт можно еще усложнить, если в клетках размещать семена не одного сорта, а разные и выбирать их из мешка по-прежнему случайным образом.

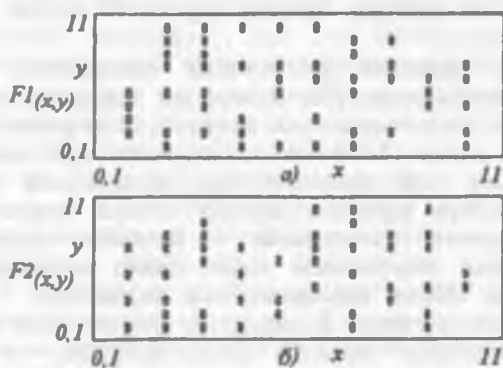


Рис. 11.4. Примеры компьютерного распределения здоровых семян на участке почвы. Приведены результаты двух экспериментов. Видно, что их результаты не повторяются. Генератор случайных чисел работает в интервале 0—10

Эти примеры показывают, что реально можно осуществить процессы, которые включают принципиально неустранимый случайный элемент. Этот элемент не связан ни с погрешностями измерений, ни с волновыми свойствами микрочастиц.

Он возникает всегда, когда система является сложной и в ней могут появиться флуктуации. Обладая лишь макроприбором, мы можем не заметить этих флуктуаций, не увидеть мельчайших капелек жидкости или семян в клетках поля. Однако, если флуктуации способны к разрастанию, подобно каплям жидкости или семенам растений, то микроскопический эффект перерастает в макроскопический, причем вся система переходит в качественно новое состояние.

Для такого перехода необходимо, чтобы не просто в малых областях пространства возникали бы сгустки молекул (это может быть и в идеальном газе), а чтобы эти сгустки были способны образовывать достаточно долгоживущие комплексы связанных между собой частиц.

Процессы качественного перехода в новое состояние в сложных системах развиваются через флуктуации, которые не просто способны увеличивать свои амплитуды, а менять свои свойства либо при небольшом изменении макроскопических внешних условий (малое понижение температуры ниже критической), либо при удачном сочетании факторов вблизи флуктуаций (прорастание семени), либо самопроизвольно. Способность системы при малых изменениях ее свойств или внешних условий развиваться в принципиально разных направлениях (если температура газа чуть выше критической, то сгусток газа распадется и флуктуация исчезнет, если же чуть ниже, то может образоваться микрокапля жидкости) называется, как уже говорилось в гл. 10, бифуркацией. Если характер этих бифуркаций известен, то возможен прогноз развития события на макроскопическом уровне, несмотря на полную непредсказуемость самих флуктуаций.

Если же путь развития флуктуаций неизвестен, никакой прогноз вообще невозможен. Изучаемые до недавнего прошлого ситуации в физике относились, как правило, к первому случаю. Хотя, например, давно было известно, что при медленном охлаждении рассола при наличии так называемой затравки (искусственного центра кристаллизации) можно вырастить монокристалл с заранее известными свойствами, однако при быстром охлаждении получается либо плохо организованный поликристалл, либо стекло (замороженная жидкость).

Если процесс постепенного разрастания флуктуаций и накопления новых признаков назвать эволюционным, а момент бифуркационного скачка в флуктуациях революционным, то можно сказать, что научный прогноз вообще возможен только на этапе эволюционного процесса, т. е. от бифуркации до бифуркации, и невозможен при достижении системой такого состояния, когда возникает и начинает развиваться процесс неизвестного

бифуркационного разрастания. При этом может оказаться, что быстро нарастает только одна флуктуация, постепенно отсасывая энергию из окружающей среды и подавляя все другие флуктуации.

Этот закон подавления хорошо виден как раз на примере биологии: широко известным фактом является то, что для каждого вида имеется вполне определенное значение плотности семян на засеваемой площади, обеспечивающее наилучшие условия для роста растений (достижения наибольшей урожайности, например).

Можно разместить семена сосны через 10 см друг от друга. Они даже могут прорасти и дать небольшие зачатки деревьев. Однако в дальнейшем слишком часто посаженные растения либо погубят друг друга, либо из них выделится одно, подавляющее другие.

Такой процесс подавления разрастающейся флуктуацией других в ее окрестностях является частным случаем проявления общего закона сохранения энергии и вещества применительно к микрообъему.

Заметим еще, что всякий необратимый процесс разрастания флуктуации до макроразмеров, т. е. переход к более упорядоченной системе, обязательно связан с понижением энтропии, что, в свою очередь, возможно только в открытых системах.

Бифуркационные состояния обладают двумя важными свойствами: во-первых, эти состояния неустойчивы, а во-вторых, выход из этих состояний возможен в общем случае по разным путям.

Следствием первого свойства является то, что слабое воздействие может привести в конечном счете к очень большим изменениям в системе. Подобно тому, как вертикально поставленный карандаш можно опрокинуть легким дуновением, ряд случайных малых воздействий могут направить развитие дальнейших процессов по совершенно разным путям.

Интересно проследить, как эти закономерности проявляются совершенно в другой области—в человеческой истории.

В истории всех народов и государств можно четко выделить эволюционные достаточно длительные периоды развития. Конечно, появление отдельных личностей (что можно рассматривать как чисто флуктуационный процесс) с ярко выраженными способностями в той или иной области и в процессе эволюции оказывает воздействие на общее поступательное движение общества. Однако в наибольшей степени это влияние проявляется именно в моменты неустойчивых состояний: военных, околореволюционных и революционных. Как историки, так и писатели, пишущие на исторические темы, неоднократно на это обстоятельство указывали.

Выше мы рассмотрели лишь несколько примеров, показавших, как общие идеи физики распространяются вне ее непосредственной сферы и позволяют обнаружить некоторые важнейшие закономерности, свойственные всему окружающему нас миру и

процессу его познания. Именно в этом, помимо чисто прагматических причин, и состоит смысл изучения физики как образовательной базы при подготовке специалистов любых направлений.

Все законы физики и разного рода принципы, о части которых говорилось раньше, можно разделить на две группы. Одни, подобно законам Ньютона или уравнениям Максвелла, являясь обобщениями собственно физических экспериментов, пригодны для употребления в сфере физики или соответствующих отраслей техники. Значение же других выходит за рамки чисто физических или технических приложений ввиду их применимости и для описания явлений от физики весьма далеких. Просто оказывается, что на примере физики, как науки, занимающейся все же объектами по своей природе более простыми, чем, например, живые организмы, человек и общество, оказалось возможным подметить такие правила, которые при значительно большем числе разнородных обстоятельств в других областях уловить несравненно сложнее.

Примерами таких обобщений как раз и являются принципы, обсуждавшиеся в этой главе.

Характерной особенностью многих сложных систем является наличие в них флуктуаций.

Эти флуктуации, не замечаемые при макроскопических экспериментах, отчетливо регистрируются при микроскопических (опыт с броуновским движением, например) и проявляются в отличном от нуля значении энтропии. Закон возрастания энтропии при необратимых процессах в замкнутой системе можно трактовать как отражение самопроизвольного увеличения разнообразия этих флуктуаций.

В гл. 10 было показано, что могут существовать такие флуктуации, которые способны менять свои свойства и при подходящих условиях не распадаться, а нарастать. Именно этим объясняется сама возможность появления более организованных, чем окружающая среда (почва, например), растительных организмов: травы, деревьев.

Такой процесс сопровождается уменьшением энтропии. Это, в свою очередь, возможно не для всякой системы, а только для такой, которая находится в контакте с другими системами (внешней средой), т. е. является открытой.

Примем теперь в качестве отправных два достаточно очевидных утверждения и посмотрим, нет ли соответствующего аналога в закономерностях развития общества. Утверждения эти следующие: 1) в сложной системе должны присутствовать флуктуации и 2) эти флуктуации при подходящих условиях могут быть способными к разрастанию.

Обратимся к человеческому обществу. Что это система сложная—ясно. Что она состоит из большого числа взаимодействующих частей—тоже не вызывает сомнения. Бросается в глаза и то, что в этой системе действительно имеются флуктуации. В

самом деле, когда на свет Божий появляется новый человек, то это является результатом совершенно случайной встречи и дальнейшего развития отношений какой-то пары женщина—мужчина. Совсем уж непредсказуемым является факт рождения в семье Летиции и Карло Бонапарте будущего знаменитого полководца и императора Франции или то, что через год после женитьбы Сергея Львовича Пушкина на Надежде Осиповне Ганнибал у них не просто родится новый подданный Российской Империи, что, в общем-то, как раз едва ли можно назвать неожиданным, а именно, по выражению Н.Гоголя, русский человек во всем его развитии—Александр Сергеевич Пушкин.

Вообще каждый человек, гений или не гений, уникален и неповторим и поэтому вполне может рассматриваться как флуктуация (как ни странно звучит этот термин в приложении к человеку). Эта флуктуация способна к саморазрастанию (в прямом смысле слова и в смысле развития своего интеллекта) и как таковая нуждается в контакте с ближайшим окружением (семья) и с более удаленным. Группы людей могут объединяться по самым разным признакам (сословным, производства, по интересам и т. д.). Эти объединения приобретают некоторые новые по сравнению с отдельными людьми черты, но по-прежнему могут сохранять некоторые характерные свойства флуктуаций. Они могут разрастаться, подавляться или усиливаться за счет взаимодействий с другими флуктуациями-сообществами и т. д. Такие же выводы можно сделать и по отношению к целым государствам.

Теперь вспомним, что в физических системах любое разрастание флуктуаций и переход системы в новое состояние возможны только в том случае, когда система является открытой, т. е. достаточно энергично обменивается с внешней средой энергией и веществом. Не трудно видеть, что и любое человеческое сообщество также не может существовать замкнуто, т.е. без взаимодействия хотя бы с природной средой обитания. В физических замкнутых системах процессы могут развиваться самопроизвольно только в сторону нарастания энтропии, т. е. разрушения организованной структуры (подобно разложению трупа после смерти живого организма). Ну, а как ведет себя человеческое общество?

Хорошо известным биологическим фактом является то, что небольшое общество людей, вынужденных заключать браки между достаточно близкими родственниками, постепенно вырождается. Впервые это было замечено на примере царствовавших домов, где в силу династических правил круг потенциальных брачующихся пар был невелик. Вырождение коснулось императорского дома Габсбургов, наследственной генетической болезнью страдал сын императора Николая II и т. д.

Значит, замеченное в физике правило деградации замкнутых статистических структур проявляется и в биологии высших организмов.

Если сообщество людей более многочисленно, то физическое вырождение может стать незаметным, но возникает другое.

Чтобы понять это, обратимся к двум примерам: древнегреческим городам Спарте и Афинам. Древняя история Греции в исторической науке играет роль простых моделей в физике. Как на примере идеального газа можно проще всего установить и изучить законы состояния и преобразования множественных объектов, так и на примере древнегреческой истории можно обнаружить и изучить некоторые весьма общие закономерности.

Известно, что Спарта и Афины соперничали друг с другом в установлении гегемонии над другими греческими государствами и даже вели длительные войны, получившие названия Пелопоннесских. В конце концов Спарта одержала военную победу. Однако посмотрим на карту современной Греции. Спарту можно найти лишь на карте достаточно крупного масштаба. Об Афинах с ее великолепными архитектурными памятниками, составляющими бесценное наследие всего человечества, знает каждый. В долгосрочной перспективе Афины, таким образом, оказались несомненными победителями. Мы и культурный-то вклад древней Греции воспринимаем прежде всего как вклад Афин. В чем же дело?

Конечно, невозможно указать на все факторы, которые привели к такому результату. Но взгляд физика сразу же улавливает в истории Афин и Спарты принципиальную разницу: Афины всегда были открытым обществом, с широко развитой торговлей, мощным флотом, обеспечивающим устойчивые связи с окружающим миром на достаточно большом удалении от самого города, интенсивный обмен информацией. Многие выдающиеся люди древнего мира из других его мест подолгу жили в Афинах. Одним словом, Афины можно смело назвать примером открытой системы. В этой системе и шло накопление культуры, интеллекта и т. д. Условия для разрастания соответствующих флуктуаций были вполне подходящими.

Общество Спарты того же исторического периода, наоборот, характеризуется ярко выраженной замкнутостью. Долгое время, следуя законам Ликурга (справедливости ради надо отметить, что в них было много привлекательного), основным требованием которых было сохранение, если так можно выразиться, «спартанской чистоты» и замкнутости, это общество в конце концов деградировало настолько, что стало терпеть поражения даже в военной области, где раньше не имело равных. Показательно, что и создатель новой военной тактики древнего мира полководец Эпаминонд (разгромил спартанцев в битве при Левктрах, впервые применив так называемый косой строй) появился не в Спарте—ярко выраженном военном государстве, а в соседних Фивах, которые много лет этой же Спартой разорялись, но принципа замкнутости не придерживались.

В наше время дошедшие до полного маразма Албания, Куба, Северная Корея обязаны этим не только усиленному навязыва-

ванию нежизнеспособного строя, но и полной изоляции от окружающего мира. Особенно сильно такая изоляция воздействует именно на интеллектуальный уровень общества. Скорее всего именно по этой причине и Спарта, явив миру целый ряд примеров воинской доблести (достаточно вспомнить ставшие нарицательными Фермопилы), не породила ни одного философа, ученого, поэта или скульптора.

Даже и наша собственная очень крупная страна сильнейшим образом пострадала в результате длительной изоляции от остального мира.

Примеры такого рода можно продолжать до бесконечности.

Таким образом, открытый в физике закон, гласящий, что всякое качественно новое развитие возможно только в открытой системе, обладающей достаточным уровнем флуктуаций, причем таких, которые способны к разрастанию, оказывается верным и применительно к человеческому обществу.

Факты физики и факты истории логично объясняются с одной и той же точки зрения. Возникновение более совершенной организации, материальной или интеллектуальной, т. е. процесс с понижением энтропии, в принципе, возможен только в открытых системах и невозможен в изолированных. Этот урок физики полезно усвоить любым государственным деятелям!

Однако одной этой открытости недостаточно. Надо еще, как упоминалось, иметь достаточный уровень специфических флуктуаций. Если их нет вообще или они малы и условия таковы, что разрастание флуктуаций маловероятно, то никакое развитие невозможно. Именно в этом лежит главная причина провала так называемой плановой, жестко детерминированной экономики, которая искусственно насаждалась в странах социалистического лагеря. Рыночная экономика в этом смысле гораздо предпочтительнее, что и доказал прямой опыт истории. К сожалению, он обошелся в десятки миллионов жизней.

Может быть, полезнее было бы изучить физику? Ведь не нарушаем же мы правила  $2 \times 2 = 4!$

## 11.7. Следящие системы и управление

В последнем разделе § 2.2 мы рассмотрели действие так называемых следящих систем. Основываясь на некоторых типичных случаях реакции этих систем на управляющие сигналы (сигналы рассогласования) нам удалось объяснить важнейший факт—прямохождение человека.

В этом параграфе мы убедимся, что анализ следящих систем помогает понять и некоторые общественные явления.

Начнем с простейшей модели управления предприятием или государством. В нормально функционирующем государстве или на предприятии все функции сбалансированы. Пусть теперь какие-то внешние условия изменились: появилась военная опасность, ухудшилось состояние экономики и снизился уровень



жизни населения, увеличился спрос на производимый предприятием продукт и т. д. Как должно тогда отреагировать правительство или директорат предприятия? Очевидно, что попытаться возникший дисбаланс ликвидировать. Предпринимаемые при этом усилия должны быть согласованы, разумеется, с величиной дисбаланса, т. е. с сигналом рассогласования. Например, для предприятия это может быть разность между желаемым в данный момент количеством продукции и реально производимым. Обозначим первую величину символом  $x_1$ , а вторую — символом  $x_2$ . Тогда сигнал рассогласования  $\epsilon \sim x_1 - x_2$ .

Достаточно очевидно также, что ни правительство, ни директорат не смогут отреагировать на сигналы  $\epsilon$  без опоздания. Значит, должна быть аналогия с запаздыванием усилителя, о котором говорилось в § 2.2.

После соответствующей переработки управляющие органы выдадут исполнительным органам некоторый сигнал  $f$ .

В простейшем случае он также будет связан с сигналом рассогласования уравнением:

$$\frac{df}{dt} + a \cdot f = \epsilon(t).$$

Как можно записать действие исполнительных органов и поведение величины  $x_2$ ? Если  $x_2$ , например, количество производимой предприятием продукции, то, очевидно, можно ввести понятия скорости и ускорения производства как прироста продукта и изменения такого прироста в единицу времени.

Ускорение прироста должно быть, очевидно, пропорциональным требованию, т. е. величине управляющего сигнала  $f$ . Значит, можно записать  $d^2x/dt^2 \sim f$ . Обозначим коэффициент пропорциональности между этими величинами символом  $1/m$ ,

так что  $\frac{d^2x}{dt^2} = \frac{1}{m} f$ . Выбор коэффициента пропорциональности

именно в такой форме вызван хорошо известным фактом своеобразной инертности любого предприятия: крупное предприятие перестраивается гораздо медленнее, чем мелкое. Явно проглядывается аналогия с массой тела, которая определяет способность тела разгоняться под действием заданной силы.

По соображениям, высказанным в § 2.2, надо предусмотреть и торможение, чтобы не возникали ненужные колебания.

В результате получим уравнение, полностью аналогичное уравнению, описывающему простую механическую следящую систему:

$$m \frac{d^2x_2}{dt^2} - k \frac{dx_2}{dt} = f.$$

Конечно, это очень упрощенная модель управления сложными процессами. Естественно, возникает вопрос: пригодна ли она для объяснения чего-нибудь реального?

Обсудим некоторые примеры.

Прежде всего, на основании результатов § 2.2 можно заключить, что, если управляющий орган способен лишь отслеживать ситуацию, но не предвидеть ее развитие, реакция всей системы будет всегда запаздывающей. Это запаздывание будет возрастать с ростом «неповоротливости» органа управления (величина, обратная коэффициенту  $a$  в первом уравнении). В свою очередь, эта неповоротливость, очевидно, возрастает с ростом штата чиновников и числа виз, которые надо получить для выдачи сигнала на производство.

Опыт истории прекрасно этот вывод подтверждает: все государства с гипертрофированной бюрократией вне зависимости от экономического и политического строя всегда с трудом и большим опозданием реагировали на изменяющуюся обстановку, что нередко приводило к их гибели.

Правило единоначалия в армии, эмпирически отобранное как наилучшее, лишь отражает тот факт, что приемлемое функционирование соответствующей следящей системы в быстро меняющейся боевой обстановке возможно только тогда, когда запаздывание в принятии решения (в выдаче сигнала  $f$ ) минимально.

Идеальный орган управления должен так отслеживать обстановку, чтобы не давать сигналу рассогласования достигать больших величин. Если это не выполняется, то из-за наличия инертности (а она всегда есть и в производстве, и в обществе) около желаемого нового уровня возникнут колебания. При непрерывном сложном изменении сигнала  $x_1$  поведение желаемой величины  $x_2$  также будет сопровождаться колебаниями.

Если это учесть, то становится понятным, почему всем революциям свойственны «перегибы» и смена прогресса и реакции. В самом деле, любой революции предшествует появление сильного сигнала рассогласования между желаемым и действительностью (высокая ступень на рис. 2.4). Система начинает стремительно переходить в новое состояние в условиях несовершенных (переходных) органов управления и исполнения. Органы управления, состоящие из новых людей, не имеющих, как правило, никакого опыта, но зато желающих добиться радикальных перемен как можно быстрее, приводят к тому, что режим реакции следящей системы, показанный на рис. 2.5, становится практически невозможным. Тем более нереализуемой оказывается и идеальная реакция (см. рис. 2.4). При таких условиях наиболее вероятной является ситуация, изображенная на рис. 2.6: переход в новое состояние происходит через стадию сильных затухающих колебаний.

Обратимся теперь к вопросам более частным. Представим себе, что какое-то предприятие, учитывая возникший и нарастающий спрос на некоторое изделие (узкопланочные любительские кинокамеры, например) начинает их выпуск. Это можно интерпретировать так: следящая система пришла в движение и начинает реагировать на внешний сигнал. Теперь пусть спрос на

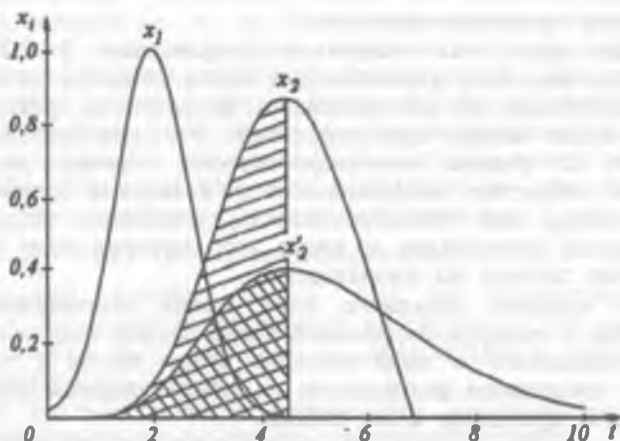


Рис. 11.5. Иллюстрация реакции производства на спрос изделия при разных параметрах простейшей следящей системы «управляющий орган—производство»

данное изделие достиг предела и начал затем падать вплоть до нуля. Так и получилось с любительскими камерами, когда началось производство относительно дешевых портативных видеокамер. Тогда зависимость спроса от времени можно изобразить в виде колоколообразных кривых ( $x_1$  на рис. 11.5) типа гауссовых экспонент. Какова типичная реакция предприятий в этих условиях? Сначала предприятие будет стараться наращивать выпуск нужных изделий (кривая произведенного продукта растет). Когда спрос уменьшится и упадет до нуля, дальнейший рост выпуска изделий невыгоден и его надо прекращать. Мгновенно это сделать, в силу инерции любого производства, нельзя. Поэтому период сокращения выпуска продукции будет довольно продолжительным.

Ну, а теперь посмотрим на кривые  $x_2$  и  $x'_2$  на рис. 11.5, показывающие реакцию следящей системы на входной сигнал  $x_1$  при разных параметрах системы. Сразу же бросается в глаза, что функции  $x_2$  и  $x'_2$  ведут себя как раз так, как и ожидалось. Видна, однако, и разница в форме кривых  $x_2$  и  $x'_2$ . Кривая  $x_2$  отклика следящей системы на входной сигнал гауссовой формы получена при тех же параметрах системы, что и на рис. 2.6. Система с выбранными параметрами достаточно быстро реагирует на изменение входного сигнала.

Кривая  $x'_2$  реакции системы отвечает значениям параметров одинаковых с теми, которые использованы при построении соответствующей кривой на рис. 2.5. Система в этом случае обладает большой инерционностью и торможением. Хорошо видно, что если в первом случае величина  $x_2$  достигает значения

0,8 от максимального значения сигнала  $x_1$ , то во втором случае лишь значения 0,4.

Заштрихованные слева на кривых  $x_2$  и  $x_2'$  площади отвечают производству в условиях хотя и спадающего, но все же спроса. Поэтому величины этих площадей можно сопоставить с общей прибылью, получаемой предприятием. Видно, что более подвижная система обеспечивает не только значительно большее насыщение рынка, но и большую прибыль. «Неповоротливая» система в этом смысле гораздо хуже.

Площади под кривыми  $x_2$  и  $x_2'$  справа от заштрихованных отвечают работе предприятия при отсутствии сбыта. Поэтому они характеризуют чистый убыток. Хорошо видно, что в случае кривой  $x_2$  «левая» площадь хотя и не намного, но меньше правой. В случае кривой  $x_2'$  все наоборот. Поэтому во втором случае предприятие понесет гораздо больше убытков. Отсюда можно сделать вывод, что предприятия первого типа способны гораздо оперативнее следить за спросом и насыщать рынок товарами и менее убыточны.

Экономика, которая базируется на предприятиях второго типа, всегда будет дефицитной и убыточной.

Особенно это различие будет проявляться при производстве продукции, спрос на которую возникает довольно быстро и быстро угасает. Это характерно, например, при производстве одежды и обуви, где спрос во многом диктуется модой, как известно, недолговечной.

Теперь вспомним, что один из важных факторов торможения возникает уже в управляющем элементе следящей системы. Он отражает способность элемента быстро преобразовывать входной сигнал  $x_1$ . В реальной экономике управление осуществляют люди.

Переработка системой управления входной информации в значительной степени зависит от желания людей это делать, а это, в свою очередь, зависит от стимулирующего фактора, или, как принято говорить, от мотивации поступков. В качестве такого фактора чаще всего выступает вознаграждение за труд. В этом смысле существует принципиальная разница между государственными и частными предприятиями: если в государственных предприятиях вознаграждение за труд определяется должностью и поэтому не зависит в течение достаточно долгого времени от качества и количества труда работника, то в частных предприятиях, особенно средних и мелких, оплата за труд почти напрямую связана с деятельностью управленца в данном отрезке времени. Именно поэтому частные предприятия гораздо быстрее реагируют на изменение внешней ситуации.

Отсюда получается следующий вывод, полностью подтверждаемый как прежней, так и новейшей историей развития стран с разными политической и экономической системами: чрезмерное превалирование государственного сектора, особенно доведение его

до абсолюта, с неизбежностью сопровождается появлением хронического дефицита, прежде всего в сфере так называемых товаров народного потребления; во всех странах с благополучной в этом смысле экономикой решающую роль играет частный сектор.

Интересно отметить, что это правило действует объективно, совершенно независимо от государственного строя: во второй половине нашего века не только наблюдается повсеместный провал так называемого социалистического способа производства, но даже, например, в Великобритании попытка лейбористов усилить государственное регулирование привела к неконкурентности этой страны на мировом рынке. Основой внутренней политики М. Тэтчер как раз и явилось «разгосударствление», что и привело, постепенно, к оздоровлению британской экономики.

Как видим, от простой механики мы перешли к некоторым принципиальным выводам в экономике, т. е. в области, от физики очень далекой. Это стало возможным потому, что мы попытались поискать аналоги простым механическим системам и явлениям. Именно возможность находить такие аналоги (а их много, и не только в рассмотренной сфере) и определяет ценность физики как общей базы научного образования. Владеть физикой—это не значит уметь решать только физические проблемы. Это подразумевает и умение смотреть на мир глазами исследователя, привыкшего искать в окружающих явлениях причинно-следственные связи и формулировать их на точном математическом языке.

Важнейшая отрасль современной науки—кибернетика—в значительной степени выросла как раз на основе физики. В сущности, в своих рассуждениях в данном параграфе мы и подошли к проблемам, изучаемым в этой области.

Этим параграфом нам хочется закончить главу и учебник. Как уже указывалось, главной целью главы было показать на ряде выходящих за рамки собственно физики примеров, как общие принципы, первоначально сформулированные в физике как науке, при более внимательном рассмотрении оказываются в высшей степени полезными и в других областях человеческого познания. Процесс такого переноса открытий в физике в другие сферы идет постоянно. Никто не может сказать, что еще обнаружится впереди и к каким выводам придут специалисты, в равной мере владеющие, например, физикой и медициной или физикой, историей и психологией. К сожалению, объединение в одной голове глубоких знаний в столь различных областях—очень редкое явление. Полезность же такого взаимодействия знаний несомненна. Именно поэтому достаточно углубленное изучение физики, как впрочем, и гуманитарных дисциплин, необходимо каждому человеку вне зависимости от его узкой специализации и конкретного рода деятельности.

#### Вопросы для самоконтроля

1. Что такое принцип дополнительности? Приведите примеры проявления этого принципа в различных областях знания.

2. Почему, изучая природу, мы вынуждены пользоваться теми или иными модельными представлениями?

3. Чем отличаются по своей постановке прямые и обратные задачи? Почему при решении последних приходится использовать так называемые штрафные функции?

4. Почему возникает элемент субъективизма при исследовании сложных природных объектов и в чем этот субъективизм заключается? Приведите примеры.

5. Какое условие должно быть выполнено при математическом оформлении любой физической теории, чтобы ее выводы могли быть сопоставлены с экспериментом?

6. Что ограничивает в принципе возможности научного прогноза? На каких этапах развития сложной системы прогноз реален и на каких нет?

7. Приведите примеры открытых в физике законов, которые оказываются справедливыми и в других областях человеческого знания.

8. Приведите примеры связи физики и кибернетики.



## Оглавление

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| <b>Предисловие</b> .....                                          | 3   |
| <b>Введение</b> .....                                             | 5   |
| 1. Что такое физика? .....                                        | 5   |
| 2. Вопросы методологии физики .....                               | 10  |
| 3. О размерностях физических величин .....                        | 21  |
| Вопросы для самоконтроля .....                                    | 24  |
| <b>Глава 1. Кинематика</b> .....                                  | 25  |
| 1.1. Системы координат и степени свободы .....                    | 25  |
| 1.2. Траектория и мировая линия .....                             | 31  |
| 1.3. Скорость движения .....                                      | 34  |
| 1.4. Ускорение движения .....                                     | 40  |
| 1.5. Релятивистская кинематика .....                              | 44  |
| Вопросы для самоконтроля .....                                    | 59  |
| <b>Глава 2. Динамика</b> .....                                    | 61  |
| 2.1. Сила и масса, второй закон Ньютона .....                     | 61  |
| 2.2. Уравнение движения .....                                     | 66  |
| 2.3. Принцип детерминизма .....                                   | 80  |
| 2.4. Третий закон Ньютона, импульс .....                          | 81  |
| 2.5. Поле .....                                                   | 83  |
| 2.6. Кинетическая энергия материальной точки .....                | 89  |
| 2.7. Движение точки по окружности .....                           | 90  |
| 2.8. Работа .....                                                 | 92  |
| 2.9. Потенциальная энергия. Потенциальные кривые и поверхности .. | 97  |
| 2.10. Сохранение энергии в механической системе .....             | 107 |
| Вопросы для самоконтроля .....                                    | 109 |
| <b>Глава 3. Элементы динамики тела и больших скоростей</b> .....  | 111 |
| 3.1. Движение твердого тела .....                                 | 111 |
| 3.2. Момент инерции тела .....                                    | 120 |
| 3.3. Закон сохранения момента импульса .....                      | 121 |
| 3.4. Динамика больших скоростей .....                             | 124 |
| 3.5. Принцип эквивалентности .....                                | 130 |
| Вопросы для самоконтроля .....                                    | 132 |

|                                                                                   |            |
|-----------------------------------------------------------------------------------|------------|
| <b>Глава 4. Колебания и волны</b>                                                 | <b>133</b> |
| 4.1. Общие положения                                                              | 133        |
| 4.2. Физический и математический маятники                                         | 138        |
| 4.3. Энергия гармонических колебаний                                              | 140        |
| 4.4. Вынужденные колебания                                                        | 141        |
| 4.5. Связанные маятники. Волны                                                    | 149        |
| 4.6. Типы волн                                                                    | 152        |
| 4.7. Сложение волн. Интерференция                                                 | 157        |
| 4.8. Принцип Гюйгенса и принцип Ферма                                             | 163        |
| 4.9. Волны в замкнутой среде. Стоячие волны                                       | 169        |
| Вопросы для самоконтроля                                                          | 173        |
| <b>Глава 5. Элементы статистической физики</b>                                    | <b>175</b> |
| 5.1. Кинетическая теория газов, элементы статистической физики                    | 175        |
| 5.2. Распределения Максвелла и Больцмана                                          | 185        |
| 5.3. Нестационарные процессы                                                      | 197        |
| 5.4. Явления переноса                                                             | 199        |
| 5.5. Реальные газы                                                                | 211        |
| Вопросы для самоконтроля                                                          | 217        |
| <b>Глава 6. Термодинамика</b>                                                     | <b>219</b> |
| 6.1. Первое начало термодинамики                                                  | 219        |
| 6.2. Теплоемкость                                                                 | 221        |
| 6.3. Адиабатический процесс и уравнение Пуассона                                  | 224        |
| 6.4. Термодинамический потенциал — энтропия                                       | 225        |
| 6.5. Тепловые машины                                                              | 228        |
| 6.6. Теорема Карно и обобщение понятия энтропии как термодинамического потенциала | 238        |
| 6.7. Энтропия в необратимых процессах                                             | 240        |
| Вопросы для самоконтроля                                                          | 245        |
| <b>Глава 7. Электромагнитные явления</b>                                          | <b>247</b> |
| 7.1. Заряды. Электрические и магнитные поля. Характеристики поля                  | 247        |
| 7.2. Электростатика                                                               | 256        |
| 7.3. Постоянный ток                                                               | 280        |
| 7.4. Магнитостатика                                                               | 290        |
| 7.5. Электродинамика                                                              | 300        |
| Вопросы для самоконтроля                                                          | 312        |



|                                                                                     |            |
|-------------------------------------------------------------------------------------|------------|
| <b>Глава 8. Электромагнитные волны и свет</b>                                       | <b>315</b> |
| 8.1. Появление и общие свойства электромагнитных волн                               | 315        |
| 8.2. Шкала электромагнитных волн                                                    | 322        |
| 8.3. Волновые явления в оптике. Интерференция                                       | 325        |
| 8.4. Фурье-спектроскопия                                                            | 330        |
| 8.5. Голография                                                                     | 333        |
| 8.6. Дифракция                                                                      | 335        |
| 8.7. Дифракционная решетка                                                          | 340        |
| 8.8. Свет и вещество. Дисперсия света и поглощение                                  | 347        |
| 8.9. Давление света                                                                 | 353        |
| 8.10. Фотоэффект                                                                    | 354        |
| 8.11. Поляризация света                                                             | 356        |
| 8.12. Освещенность поверхности                                                      | 362        |
| Вопросы для самоконтроля                                                            | 363        |
| <b>Глава 9. Элементы теории строения вещества</b>                                   | <b>365</b> |
| 9.1. Особенности движения микрочастиц и способы описания их состояний               | 365        |
| 9.2. Элементы квантовой механики                                                    | 371        |
| 9.3. Качественное объяснение основных закономерностей состояний атома водорода      | 379        |
| 9.4. Принципы спектрального анализа                                                 | 383        |
| 9.5. Вид волновых функций атома водорода. Гибридизация                              | 385        |
| 9.6. Принцип Паули и заполнение электронных оболочек                                | 393        |
| 9.7. Строение молекул                                                               | 396        |
| 9.8. Особенности строения биомолекул                                                | 410        |
| 9.9. Особенности спектров молекул. Цветность соединений. Симметрия и правила отбора | 412        |
| 9.10. Принцип действия оптических квантовых генераторов (лазеров)                   | 420        |
| 9.11. Зонная структура кристаллических тел                                          | 430        |
| 9.12. Внутриядерные процессы. Радиоактивность                                       | 438        |
| Вопросы для самоконтроля                                                            | 442        |
| <b>Глава 10. От существующего к возникающему</b>                                    | <b>445</b> |
| Введение                                                                            | 445        |
| 10.1. Время в классическом мире                                                     | 447        |
| 10.2. Необратимость процессов и «стрела времени»                                    | 452        |
| 10.3. Флуктуации, бифуркации и самоорганизация                                      | 456        |
| 10.4. Самоорганизация и условия существования живых организмов                      | 460        |
| 10.5. Взаимодействие подсистем                                                      | 463        |

|                                                                                                                         |     |
|-------------------------------------------------------------------------------------------------------------------------|-----|
| 10.6. Связь микроскопического и макроскопического, или как «считать» хаос? . . . . .                                    | 471 |
| 10.7. Энтропия и информация . . . . .                                                                                   | 482 |
| 10.8. Формирование гибкой молекулярной цепи и механизм образования пространственных структур белковых молекул . . . . . | 487 |
| Вопросы для самоконтроля . . . . .                                                                                      | 502 |

|                                                                                    |            |
|------------------------------------------------------------------------------------|------------|
| <b>Глава 11. Уроки физики . . . . .</b>                                            | <b>504</b> |
| 11.1. Принцип дополнительности . . . . .                                           | 504        |
| 11.2. Природа и модели . . . . .                                                   | 512        |
| 11.3. Обратные задачи . . . . .                                                    | 518        |
| 11.4. Объективное и субъективное в науке, или о принципе доказательности . . . . . | 526        |
| 11.5. Принцип соответствия и особенности построения физических теорий . . . . .    | 530        |
| 11.6. Научный прогноз и его ограничения. Физика и История . . . .                  | 537        |
| 11.7. Следящие системы и управление . . . . .                                      | 547        |
| Вопросы для самоконтроля . . . . .                                                 | 552        |

**Книги издательства «Гардарика»**

**можно купить и заказать**

*по адресу:*

107005, Москва, ул. Ф. Энгельса, д. 3/5, стр. 6

(ст. метро «Бауманская»)

Тел.: (095) 267-3405. Факс: (095) 267-8252

Санкт-Петербург, Невский проспект, д. 85/11

(вход с ул. Гончарной) Тел.: (812) 168-4928

---

**Учебное издание**

*Грибов Лев Александрович*

*Прокофьева Нина Ивановна*

## **ОСНОВЫ ФИЗИКИ**

**Учебник**

**Переплет художника *А.Л. Бондаренко***

**ЛР № 064323 от 23.11.95**

**Подписано в печать 02.04.98. Формат 60 × 90/16.  
Бумага офсетная. Печать офсетная. Усл. печ. л. 35.  
Тираж 30 000 экз. (1-ый завод – 10000). Заказ 487**

**Фирма «Гардарика»**

**107005, Москва, ул. Ф. Энгельса, д. 3/5, стр. 6**

**Изготовлено в полном соответствии  
с качеством предоставленных диапозитивов  
в ОАО «Можайский полиграфкомбинат»  
143200, г. Можайск, ул. Мира, 93**



ИЗДАТЕЛЬСКО-КНИГОТОРГОВАЯ ФИРМА

# Гардарика

**Н о в и н к а !**



П.П.Бочаров, А.В.Печинкин

## ТЕОРИЯ ВЕРОЯТНОСТЕЙ. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

М.: Гардарика, 1998. Переплет, 280 с.

ISBN 5-7762-0035-0

140x210 мм(60x90/16)

### О книге

В первой части пособия рассматриваются основные понятия теории вероятностей. При этом авторы избегают сложных математических конструкций, но тем не менее ведут изложение на основе аксиоматического построения, предложенного академиком А.Н. Колмогоровым. Во второй части излагаются основные понятия математической статистики. Рассматриваются наиболее часто встречающиеся задачи оценки неизвестных параметров и проверки статистических гипотез и описываются основные методы их решения. Каждое приведенное положение иллюстрируется примерами. Излагаемый материал в целом соответствует государственному образовательному стандарту.

### Рекомендуется

Студентам, аспирантам и преподавателям вузов, а также научным работникам различных специальностей, не владеющим теорией меры и другими специальными математическими дисциплинами в достаточном объеме и желающим получить первое представление о теории вероятностей и математической статистике.

Книги Вы можете приобрести или заказать и получить по почте в издательско-книготорговой фирме «Гардарика»:

107005, Москва, ул. Ф. Энгельса, 3/5, стр. 6

Тел.: (095) 267-6483, 267-3405,

факс: (095) 261-8297

Оптовым покупателям — скидка.

Л.А.Грибов Н.И.Прокофьева

# ОСНОВЫ ФИЗИКИ

Учебник охватывает все основные разделы, традиционно включаемые в учебные курсы физики для высших учебных заведений: классическую, релятивистскую и квантовую механику, термодинамику, включая элементы неравновесной термодинамики, элементы статистической физики, учение об электромагнитных явлениях, оптику и др. Характерной особенностью учебника является обсуждение ряда важнейших проблем, ранее в отечественной учебной литературе по общей физике не рассматривавшихся. Это относится, прежде всего, к главам, посвященным строению вещества на молекулярном уровне, появлению "стрелы времени", флуктуациям как необходимому условию качественного развития, самоорганизации, сочетанию случайного и закономерного и др.

