

УДК 501

ББК 87.22

К 12

Каганов М. И., Любарский Г. Я. **Абстракция в математике и физике.** — М.: ФИЗМАТЛИТ, 2005. — 352 с. — ISBN 5-9221-0410-1.

Математическая часть книги представляет собой собрание эпизодов по истории математики, поскольку история абстрактных понятий от нее неотделима. В ней рассказано о цепи связанных друг с другом задач и соответствующей цепи абстрактных понятий–инструментов, созданных для решения этих задач.

Главное содержание физической части — рассказ о проблемах и достижениях теоретической физики, подчеркивающий роль абстрактных понятий, которые помогают описать многообразие окружающего нас мира.

Книга предназначена широкому кругу читателей, интересующихся математикой и физикой.

Введение

В настоящей книге мы надеемся рассказать много интересного о математике и физике.

Первоначально было задумано написать статью в защиту абстрактных понятий в этих науках. Постепенно стало ясно, что мы не можем ограничиться столь узкой задачей. Мы испытали странное ощущение, будто в защите нуждаются сами науки. Правда, по разным причинам.

То, что физики добились огромных успехов в развитии своей науки, понимают все. И все понимают, что успехи эти коренным образом изменили условия жизни человека, создав технологическую основу промышленности. Менее известны успехи физиков в развитии картины мира. Каждое экспериментальное открытие, пока не создана объясняющая его теория, остается изолированным удивительным фактом. Каждая новая теория неизбежно разрушает, хотя бы частично, прежнее представление о мире и создает новое. Если теория достаточно глубока, возникает возможность предсказать большое количество других экспериментальных фактов, которые в головах инженеров синтезируются в новые технологии и изобретения.

Нам кажется, именно эта сторона деятельности ученых — развитие физической картины мира — часто является объектом необоснованных упреков и нападок. Наиболее часто физику обвиняют в абстракции, в оторванности от жизни именно тогда, когда обсуждаются теоретические работы. В свое время упрекам пытались придать политическую окраску и обвиняли физиков в идеализме.

В математике все по-другому. Многие считают, что время ее успехов закончилось лет 200–300 тому назад, а теперь математика — сухая наука. Даже те, кто знает, что это не так, часто не могут преодолеть страха перед непривычным уровнем абстрактности ее методов.

Упомянутые нами обстоятельства могут помешать кому-то выбрать в качестве своей профессии теоретическую физику или математику. Особенно негативную роль играет двусмысленность слова *абстракция*.

Поясним, что мы имеем в виду.

Рассмотрим существительное *неверный*. Оно может иметь резко отрицательный смысл. Это человек, на которого нельзя положиться, которому нельзя доверять, перебежчик, коварный человек. Данное

слово имеет и вполне нейтральный смысл: *человек, не исповедующий мусульманскую религию*. В умах немусульман это два разные слова, но для многих мусульман оба столь различные смысла сливаются, образуя силлогизм: «немусульманин — человек, на которого нельзя положиться. . .»; как правило, они перечисляют много отрицательных характеристик.

Вернемся к слову *абстрактный*.

Если тема разговора или содержание текста непосредственно не связаны с наукой, оно — это слово — имеет, или лучше сказать, ему придают резко отрицательный смысл — *оторванный от жизни, ненужный, заумный*. Оно эмоционально окрашено. Часто можно слышать выражение «*абстрактная наука*», значит говоривший считает, что эта наука не нужна. . . Кому не нужна? Простым людям — таким как мы с вами. Иногда вместо возражения по существу можно услышать: «У меня нет времени на *абстрактные рассуждения*». Часто это слово используют как пугало: «Все это *чересчур абстрактно*, мне не понять». И непонимание пытаются превратить в достоинство.

Это одно значение слова *абстрактный*. Другое значение мы нашли в Энциклопедическом словаре (М., «Советская энциклопедия», 1989): «*Абстракция* (от лат. *abstractio* — отвлечение) — форма познания, основанная на мысленном выделении существенных свойств и связей предмета и отвлечении от других, частных его свойств и связей. . . Основные типы абстракции: изолирующая (вычленяющая исследуемое явление из некоторой целостности); обобщающая (дающая обобщенную картину явления); идеализация (замещение реального эмпирического явления идеализированной схемой). Понятие абстрактное противопоставляется конкретному». Слово «отвлечение» в этой цитате можно, по-видимому, заменить более понятным словом «игнорирование».

Итак, в этом значении слово абстракция не несет на себе никаких негативных признаков. Оно не принадлежит ни к физическим, ни к математическим терминам, хотя используется и в физике, и в математике. Абстракция — философское понятие.

Мы не ставим себе целью изгнать из употребления слово абстракция ни в одном из двух его смыслов. Даже составители словарей не борются с языком, а фиксируют, как употребляют те или иные слова, обороты, выражения. Конечно, и мы не надеемся избавить язык от обвинительного употребления слова *абстракция*. Нам представляется, что наилучший способ защиты абстракции — это, не рассыпая похвал по ее адресу, рассказать о некоторых интересных эпизодах истории физики и математики, неразрывно связанных с применением абстракции.

Пытаясь перенести на полотно или бумагу понравившийся пейзаж, художник или фотограф выбирают ракурс. Ракурс необходим и популяризаторам. Нам, кажется, удалось найти необычный ракурс,

позволяющий рассказать не вполне традиционно о математике и физике. Рабочим названием рассказа было: «Об абстракции — конкретно на примере физики и математики». Решая поставленную перед собой задачу, мы будем пользоваться и философским, и житейским значениями слова абстракция. Надеемся, из контекста будет понятно, какое значение мы имеем в виду.

Современная наука необозрима. Энциклопедистов нет. Даже наиболее эрудированные люди знают на профессиональном уровне, кроме своей области, в лучшем случае лишь смежные. Именно поэтому мы надеемся, математики-профессионалы найдут для себя кое-что интересное в главах, посвященных физике, а физики — в математических главах.

* * *

Во время размышлений и разговоров о содержании будущей книги мы обнаружили статью В. Гейзенберга «Абстракция в современной науке». Она опубликована в его книге «Шаги за горизонт» (Москва, «Прогресс», 1987).

Гейзенберг — не только великий физик. Один из создателей квантовой механики, он — крупный, интересный философ. Его мысли о роли науки, о методах познания не потеряли интереса и сегодня, хотя они были высказаны более пятидесяти лет назад.

В названной статье В. Гейзенберг защищает абстракцию. Естественно, он не одинок. Даже не занимаясь поисками, а просто посмотрев на полку домашней библиотеки, открыв почти наугад несколько книг, мы обнаружили высказывания о пользе абстракции и у Анри Пуанкаре, и у Альберта Эйнштейна, и у Нильса Бора. Остановились мы на статье Гейзенберга, во-первых, потому что она была первой, которую мы прочли на эту тему, а во-вторых, она очень подходит к содержанию задуманного нами сочинения.

В первом абзаце статьи Гейзенберга сказано: «... часто выдвигается следующее утверждение: наука в процессе своего развития становилась все более и более абстрактной, а в наше время во многих отраслях достигла прямо-таки *пугающей степени абстрактности...*» (выделено нами). Чем пугает абстрактность? Гейзенберг, по смыслу статьи, говорит *только* о физике: об *уходе от конкретности*, понимаемой как совокупность явлений, доступных органам чувств человека, об уходе в атомный и субатомный мир, объекты которого не только непосредственно не наблюдаемы, но и не подчиняются законам привычного нам мира. Из-за того что явления, происходящие в микромире, недоступны нашим чувствам, они как бы удалены от нас и воспринимаются как абстрактные.

Для того чтобы понять тех, кого *пугала абстрактность атомной физики*, надо вспомнить: еще сравнительно незадолго до того, как Гейзенберг написал свою статью, многие ведущие ученые вообще

сомневались в существовании атомов. Сегодня, конечно, никто в существовании атомов и субатомных частиц не сомневается. При этом ощущение «абстрактности» микромира исчезло не полностью. Оно питается истинной непохожестью законов микромира на привычные законы макромира.

Абстракция в философском смысле слова и в физике, и в математике, естественно, встречается и используется весьма часто, буквально на каждом шагу. Приемы абстрагирования разнообразны, никого они не пугают, когда применяются осознанно и оправданно. И при знакомстве с работами, областями, направлениями исследований, использующими абстрактные приемы, упрек в абстракции редко входит в формулу обвинения.

Стоит отметить два обстоятельства. Наша статья не выходит за пределы физики и математики. Так вот, первое: чаще всего обвинение в абстракции высказывается математиками по отношению к физическим работам, а физиками — по отношению к математическим работам. Ни физикам, ни математикам в подавляющем числе случаев свои работы не кажутся абстрактными. И второе: когда обвинение произносится, то в большинстве случаев слово *абстракция* используется в житейском, а не в научном смысле.

Правда, иногда взгляд со стороны выявляет нечто, что не замечают те, кто привык к применяемому методу или к используемому подходу. Например, недостаточную обоснованность используемых методов или неправомерность приема.

В теоретической физике абстрагирование часто сводится к упрощению задачи путем пренебрежения второстепенными деталями. Если взгляд математика различает лишь несоответствие канонам математической строгости, то физик легко игнорирует обвинение, но если выясняется, что абстрагирование не так безобидно, как казалось, то нередко следствием оказывается существенное развитие теории именно за счет переосмысления возможного уровня абстрагирования.

Можно привести конкретный пример непродуманного абстрагирования в преподавании математики. Изложение правил преобразований алгебраических выражений без ссылок на арифметику (такие попытки предпринимались) превращает алгебру в трудную абстрактную (!) науку.

Наша книга — не просто защита абстракции. Она — апология абстракции во всех тех смыслах, которые упомянуты в Энциклопедическом словаре. Мы будем довольны, если прочтя ее, читатель поймет, что абстракция — одна из движущих сил науки. На одной из заключительных страниц своей статьи В. Гейзенберг приводит фразу известного немецкого философа Фридриха Ницше (1844–1900): «Абстракция — мучка для многих, а для меня в мои лучшие дни — праздник и упоение». Это высказывание Ницше нам понятно и близко.

При всем разнообразии методов и приемов, используемых современной наукой, все они принадлежат единой *научной методологии*. Хотя методология и есть совокупность методов и приемов, в выражение *научная методология* вкладывается нечто большее — то, что в последнее время принято именовать *парадигмой*. Здесь не место формулировать, что есть современная научная парадигма. Научная методология меняется с развитием науки, предъявляет к каждой конкретной работе свои очень серьезные требования.

В увлечении конкретными достижениями науки часто забывают, что создание научной методологии, формулировка научной парадигмы и расширение области ее применимости, возможно, представляют собой наиболее важное достижение науки, под воздействием которого отступают, правда, очень медленно, иногда с большими задержками, мракобесие, вера в чудеса, укрепляется вера в науку у людей, далеких от нее.

Абстракция на полном основании входит в арсенал научной методологии.

Уделив много внимания содержанию понятий, объединенных словом абстракция, отметим, что следующие два выражения (оба из научного арсенала) несут совершенно различную смысловую нагрузку:

- использование абстракции как метода;
- создание абстрактных понятий.

Абстракция как метод в той или иной мере используется почти в каждой физической и математической работе. Создание полезных и глубоких абстрактных понятий — искусство, доступное немногим. Подобно тому, как географическая карта — памятник, запечатлевший имена первооткрывателей, наука зафиксировала имена тех, кто ввел и укоренил новые понятия. Вспомним *решето Эратосфена, интеграл Лебега, пространство Гильберта, функции Бесселя...*

Некоторые новые физические понятия и имена физиков, их разработавших, будут названы в тексте.

Рассказ об использовании абстрактных приемов и введении абстрактных понятий в математике и в физике затронет несколько тем. Каждая из тем, о которых пойдет речь, могла бы быть содержанием отдельной статьи.

* * *

В понимании, а особенно в ощущении понимания, привычка играет важную роль. От привычки или от ее отсутствия во многом зависит, считать ли то или иное явление, метод, прием исследования абстрактным или конкретным. Ниже будут приведены соответствующие примеры. Не только привычка приводит к конкретизации абстракции. Бывает, что превращение какой-либо области науки из абстрактной в конкретную происходит в общественном сознании на наших глазах «революционным путем». Один, но очень яркий, пример. За несколько

лет до открытия распада урана под действием медленных нейтронов — открытия, послужившего началом атомного, точнее, ядерного века, не одни только далекие от физики люди осуждали многих физиков за их дорогостоящий интерес к *абстрактным* (читай, бесполезным) исследованиям, прежде всего к ядерной физике.

Говоря с осуждением о том, что какая-то работа — сплошная абстракция, не следует забывать этот и подобные примеры.

* * *

В процессе совместной работы авторы с некоторым удивлением обнаружили, что их взгляды на физику и математику не всегда полностью совпадают. Указанное расхождение было устранено в той мере, в какой это необходимо при написании двух частей одной книги. В результате возник труд, который можно рассматривать как совместный манифест физика и математика в защиту абстракции. По необходимости манифест подчеркивает наличие в отношении к абстракции у физика и математика как сходства, так и отличий. Последние наиболее заметны в принятом уровне строгости. Мы надеемся, что наш текст позволит не только отметить подобные различия, но и понять их причину.

Позволим себе поделиться сделанными выводами.

История создания современной техники и ее технологической базы легко просматривается. Отчетливо видна ее многоступенчатость. Для создания новых инструментов и технологий как минимум необходимо иметь инструменты и технологии предыдущей ступени.

Создание абстрактных понятий — также многоступенчатый процесс: каждое новое понятие для своего определения требует обращения к ранее введенным. Возникает язык понятий, встроенный в литературный язык. Овладеть этим языком нелегко, поскольку переход со ступеньки на ступеньку требует усилий и своеобразной сноровки (обучения).

Это первая причина непонятности, а в результате — кажущейся заумности новых абстрактных понятий и использующих их теорий. Все сказанное пока относится как к физике, так и к математике.

Обе науки развивались в тесном контакте друг с другом. Они питали друг друга идеями, методами, представлениями. Многие понятия, возникнув в недрах одной науки, становились фундаментальными понятиями другой. Надо признать, что процесс взаимовлияния и взаимного обогащения имеет место и в настоящее время. И все же сегодня мы довольно четко отличаем физику от математики, а математику от физики.

По-видимому, мы сильно отклонимся от темы нашего рассказа, если начнем выяснять связь математики с Миром, с Природой. Есть ощущение, что связь эта — одно из величайших чудес (А. Эйнштейн). Мы не знаем, *почему возможно описать* свойства Мира, Природы строгим

языком математики, созданным Человеком. Мы столь к этому привыкли, что готовы утверждать: «Сколько математики, столько науки».

Но... даже самые смелые математические теории, пока они не используются для описания свойств Мира, Природы, не могут вступить в противоречие с нашим представлением о внешнем и внутреннем мире ¹⁾. Критерием истинности математической теории служат ее безукоризненная логичность и продуктивность. Бессмысленно создавать теорию, из которой ничего не следует.

С физическими теориями дело обстоит иначе. Физика имеет своим объектом изучения и понимания реально существующий мир вещей. Создание тончайших приборов, необозримо расширяющих возможности наших органов чувств, позволяющих чувствовать то, чего без прибора человек почувствовать не может, не ликвидирует представления о существовании (вне нас и хоть в какой-то мере независимо от нас) объекта исследования. За многие тысячелетия общения Человека с внешним миром накопился огромный чувственный опыт понимания, опыт, дающий человеку возможность не только познавать свойства Мира, но и использовать результаты познания для совершенствования своей жизни.

Познание, проникновение в тайны Природы, к сожалению, не происходит лишь путем накопления новых сведений. Новые теории требуют пересмотра *наших представлений*. Выясняется, что реально существующие объекты Природы имеют совершенно непривычные черты, черты не только плохо доступные нашему пониманию, но и буквально противоречащие нашему многовековому чувственному опыту. Всякий раз, когда создается новая фундаментальная теория, Человек стоит перед выбором: либо «отвернуться», признать новую теорию «пугающе абстрактной» (Гейзенберг), либо кардинально изменить свое мировоззрение. И механика Ньютона, и электродинамика Фарадея—Максвелла, и молекулярная физика, и обе теории относительности Эйнштейна, и квантовая механика, и физика элементарных частиц, и современная космогония — все они и каждая в отдельности были

¹⁾ В действительности, это не совсем так. Мы увидим, что в первой половине двадцатого века представление о таком основополагающем понятии, как система аксиом, подверглось столь же драматическим изменениям, что и представление о пространстве и времени в теории относительности. Более того, был найден пример, подтверждающий новые представления о системах аксиом.

Но и это еще не все! Выяснилось, что для одного из наиболее употребительных в математических исследованиях понятия — бесконечного множества объектов (какова бы ни была их природа) — не существует удовлетворительной системы аксиом. Зато существуют две противоречащие друг другу системы, и математическое сообщество не знает, какой из них отдать предпочтение.

причиной кардинальной ломки наших представлений о Море. Это одна из особенностей физической абстракции. Скорее всего, именно в этом причина затруднений в приятии новых абстракций в физике.

Естественно, и математика не лишена *специфических* особенностей при создании новых абстрактных понятий. Одно из ведущих направлений развития математики — создание теорий, которые объединяют ряд ранее существовавших теорий и, главное, делают последние более прозрачными, простыми и, по сути дела, отмирающими. Тем самым это направление упрощает изучение математики и противостоит тенденции безграничного расширения минимума знаний, необходимых исследователю и математику-прикладнику. Кроме того, объединение облегчает открытие новых фактов и решение новых задач. Разумеется, каждая объединенная теория щедро вводит новые понятия и тем самым добавляет новую ступеньку к лестнице понятий.

Важная особенность большинства абстрактных понятий — их способность развиваться, включая в себя все более широкий круг объектов. Примеры нетрудно найти и в физике, и в математике: энергия или элементарные частицы (физика); число или интеграл (математика).

Перейдем к пресловутой строгости рассуждений.

Принято считать, что физики подтрунивают над математиками по поводу их неумеренной любви к строгости, а математики только пожимают плечами, когда при них говорят о строгости и логичности в статьях физиков. Бывает. Но отнюдь не всегда. Как правило, претензии физиков к математикам и математиков к физикам не имеют серьезных оснований.

В каждой из обеих наук используется несколько разных уровней строгости. Существуют работы, к которым понятие строгости вообще неприменимо. Что же это за работы?

Для уборки снега математическая строгость, естественно, не нужна, хотя выполняя такую простую работу необходимо придерживаться определенных правил. Точнее было бы сказать так: к процессу уборки снега понятие строгости неприменимо. А есть ли в науке работы, к которым неприменимо понятие строгости? Есть!

Начнем с работ, о строгости которых можно судить. Что, собственно, есть строгость? Это свойство рассуждения, не позволяющее опровергнуть его чисто логическим путем. Возьмем научную статью, которая явно или молчаливо исходит из некоторых общепринятых положений и выводит из них некоторые интересные следствия. Если при получении выводов законы логики не нарушены, о строгости такой статьи можно судить по тому, какие предположения лежат в ее основе — лишь правдоподобные или истинные.

Но как быть, если весь пафос статьи заключается в формулировке нового принципа в новой области науки, к которой некоторые из старых принципов оказались неприменимыми? Оценивая статью, надо интересоваться не ее строгостью, а ее *правильностью*. Подобную

статью может опровергнуть либо эксперимент, либо ее противоречие с теми из старых принципов, которые заведомо применимы и в новой области.

Хрестоматийным примером может служить открытие Дираком квантово-механического уравнения, описывающего релятивистское движение электрона. Для его написания Дираку нужно было представить квадратный корень из суммы квадратов трех операторов в виде линейной суммы этих же операторов. Простейшее рассуждение показывает, что *сделать этого невозможно*. Но Дирак это сделал! Он «позволил» искомым коэффициентам быть не числами, а матрицами! Однако не в этом проявилось его величие. Дирак понял, что данный математический фокус имеет глубокий физический смысл. Поняв, поверил, несмотря на то что полученное им уравнение привело к парадоксальным выводам. Это его не смутило и, во всяком случае, не остановило. Каждая трудность превращалась в победу, в открытие. Наличие четырех компонент у волновой функции электрона объяснило наличие у электрона спина и магнитного момента. Магнитный момент перестал быть феноменологическим понятием. Кажущаяся непреодолимая трудность теории — существование бесконечного числа уровней с отрицательной энергией — привела к открытию позитрона и идее античастиц.

В математике тоже можно найти подобные примеры. Например, работы Н. Лобачевского, Дж. Буля, Э. Галуа, а из более поздних — Г. Кантора, Д. Гильберта.

Основной вид нестрогости в фундаментальных физических работах как правило связан с умолчанием. Некое качественное утверждение, часто кажущееся очевидным, не упоминается вовсе.

Наглядным примером может служить построенная Ландау теория фазовых переходов. Несколько вводных страниц его статьи 1937-го года, общепризнанно сыгравшей фундаментальную роль в понимании фазовых переходов и критических явлений, посвящены физической стороне вопроса. По существу, автор вводит, но не называет важнейшую характеристику фазовых переходов — величину, потом получившую наименование *параметр порядка*. Без параметра порядка не обходится ни одна работа о фазовых переходах. Вводный раздел статьи завершается словами: «Таким образом, начало разложения [термодинамического потенциала] можно записать следующим образом...» Ландау даже не упоминает, что принятое разложение предполагает аналитичность термодинамического потенциала как функции параметра порядка. В дальнейшем выяснилось, что именно это неупомянутое предположение не всегда справедливо; внимание теоретиков было привлечено к тем случаям, когда теория расходится с экспериментом. Последнее и способствовало построению современной теории фазовых переходов.

Существенное различие физики и математики — следствие различия объектов, которые они изучают. Строго говоря, всякое определение физического понятия всегда приближенно, так как всякий физический объект обладает, по-видимому, бесконечным числом независимых (различных) свойств, а любое взаимодействие включает в себя неограниченное число объектов. Поэтому физики изначально решают приближенные задачи, что оправдывает использование приближенных методов. Искусство физиков-теоретиков состоит в умении находить простые, но не слишком переупрощенные модели, исследование которых по возможности не требует владения виртуозной математической техникой.

В противоположность сказанному каждый математический объект имеет ограниченное количество независимых свойств, оговоренных в его определении. Все остальные свойства логически вытекают из определения.

По мнению многих математиков, математика проще физики. Тем не менее и математики любят работать, не надевая на себя «корсет излишней строгости». В работе математика интуиция играет не меньшую роль, чем в работе физика.

Повышенная требовательность к строгости часто характерна для начинающих математиков. Усвоив характерные ошибки, они действуют по пословице: «Обжегшись на молоке, дуют на воду».

Сделаем вывод. И математики, и физики выработали, каждый для себя, оптимальный уровень строгости, который обеспечивает надежность получаемых результатов и не тормозит работу (последнее не менее важно, чем надежность) — ситуация, хорошо описанная гномом в стихотворении Джона Хендрика Богса. В ответ на вопрос, почему он так мал ростом,

Гном произнес, нахмурившись,
Мой рост как раз по мне.
И твой, вероятно, тоже
Подходит для тебя.

* * *

Достаточно перелистать нашу книгу, чтобы убедиться: внешний вид текстов первой (математической) и второй (физической) частей отличается количеством формул. Хочется предупредить читателя: обилие формул в математической части не есть свидетельство ее большей сложности. Мы думаем, что обе части одинаково сложны (или одинаково просты — кому как хочется считать). Обнаружив различие, мы вспомнили, что любой номер журнала «Квант» в этом смысле похож на нашу книжку: в математических статьях значительно больше формул, чем в физических. Что это? Просто различие традиций при

написании научно-популярных статей по математике и физике, или за этим нечто кроется?

Когда в статье речь идет о результатах, полученных в эксперименте, или о самом эксперименте, очевидна необходимость обычных слов. Как иначе описать устройство прибора, перечислить полученные результаты?! Следует подчеркнуть: экспериментальные работы, публикующиеся в научных журналах, тоже содержат значительно меньше формул, чем теоретические.

Научные статьи, в которых изложено содержание работ, посвященных решению задач (проблем) теоретической физики, полны формул. Часто язык этих статей предельно беден: «Подставляем... Заменяем... Получим...» Вместе с тем, в научно-популярных статьях на теор-физические темы как правило не так много формул и много слов. Это происходит потому, что каждый автор научно-популярной статьи, пытающийся рассказать о проблемах и результатах теоретической физики, должен прежде всего объяснить читателю, каково соотношение между физическими понятиями и формулами, призванными их описать; объяснить, что и каким образом формулы выражают. В большинстве популярных статей по физике конкретный метод решения задачи остается вообще за пределами статьи. Наше изложение не нарушило эту традицию.

Математическая часть книги — это по сути собрание эпизодов по истории математики, так как история абстрактных понятий от нее неотделима. В ней рассказано о цепи связанных друг с другом задач и соответствующей цепи абстрактных понятий — инструментов, созданных для решения этих задач.

Главное содержание физической части — рассказ о проблемах и достижениях теоретической физики, рассказ, подчеркивающий роль абстрактных понятий, которые помогают описать многообразие окружающего нас Мира.

* * *

Мы назвали свою книгу «Беседы об абстракции в математике и в физике». Действительно, все начиналось с бесед. К сожалению, чаще по телефону, чем при встречах: живем мы в разных городах. Подумали: не записать ли свои диалоги? Но как-то хлопотно в письмах задавать вопрос, ждать ответ, задавать следующий. Поэтому решили написать две разные части: одну — об абстракции в математике и вторую — об абстракции в физике. Конечно, каждый из нас читал и то, что написал соавтор. Все обсуждали.

С содержанием книги можно познакомиться по оглавлению. Оно достаточно подробно. Хотелось бы, чтобы читатель ознакомился со всей книгой. Однако можно читать и избранные главы.

Не следует думать, что наши разговоры об абстракции были лишь поводом поделиться своей любовью к математике и физике. Не только.

На разных примерах мы хотим убедить будущего читателя: встретившись с трудностью, не слишком доверяйте своему первому ощущению. Мысль «это для меня слишком абстрактно», скорее всего, просто означает, что требуется более тщательное продумывание. Вполне вероятно, после этого слова «прочитанная работа», «введенное понятие», «используемый метод» обретут вполне конкретный и полезный смысл.

Не зная своих читателей, мы очень надеемся, что среди них будут молодые люди, раздумывающие над выбором профессии. Если кое-кто из молодых читателей под воздействием написанного нами текста попытает свои силы на научном поприще, мы будем обрадованы и горды.

*Авторы
Бостон, Чикаго*

Часть I
ЭВОЛЮЦИЯ
АБСТРАКТНЫХ ПОНЯТИЙ
В МАТЕМАТИКЕ

Введение

Настоящая часть книги обращена к двум категориям читателей: к тем, кто боится математики и считает, что не любит абстракций, и к тем, кто не боится абстракций и любит математику. Таким образом, авторы ничем не ограничивают свою предполагаемую аудиторию. Они надеются избавить от неприязни к абстракциям одних читателей и познакомить всех читателей с некоторыми этапами новейшей истории математики, многие из которых известны только специалистам. Обе цели могут быть достигнуты, если читатель прочтет хотя бы несколько глав (не обязательно подряд). Предоставив читателю эту свободу, авторы чувствуют и себя вправе говорить с ним, варьируя от темы к теме предполагаемый уровень его квалификации. Как правило, чем ближе к концу, тем сложнее рассматриваемые вопросы. Пусть каждый самостоятельно решает, что прочесть, что пропустить.

Мы не пытаемся внушить читателю робость перед грандиозностью системы абстрактных понятий математики двадцатого века. Наоборот, мы стараемся показать их естественность и внутреннюю простоту. Все известные нам абстрактные понятия служат делу и приносят пользу. Мы не упускали случая это продемонстрировать. Вместе с тем, нельзя было выйти за пределы разумного объема книги и уровня подготовки читателя. Последнее ограничение пришлось в какой-то мере нарушить в заключительных главах в расчете на добрую волю и готовность читателя к сотрудничеству. Нам было жаль пожертвовать этими главами.

Несколько слов об абстракциях. В общем введении приводится определение абстракции как метода. Применение этого метода в математике выглядит проще и, может быть, прямолинейней, нежели в физике. Не очень уклоняясь от истины, можно сказать, что в математике (в отличие от физики) метод абстракции сводится к созданию и использованию абстрактных понятий. При этом абстрактные понятия в математике и физике очень отличаются друг от друга. Приведем классический пример. Понятие волновой функции в квантовой механике требует для своего усвоения веры, воображения и готовности к сотрудничеству с автором учебника или учителем. Понятие предела требует «только» осознания роли каждого слова в его определении.

В настоящей части книги мы будем пользоваться следующим определением.

Понятие, означающее группу объектов, является абстрактным, если оно описано не поименным перечислением объектов, а указанием их определяющих свойств.

В силу этого определения одно и то же абстрактное понятие может описывать не одну группу объектов, а много разных групп различной природы. Например, стадо — это группа пасущихся вместе животных (каковы бы ни были их вид и число). Понятие «четное число» является абстрактным, поскольку оно описывается как число, обладающее свойством делиться без остатка на два, а не перечислением всех четных чисел. Абстрактными являются такие разные понятия, как прямоугольный треугольник, сумма двух чисел, школьник. Абстрактным является число три, поскольку оно обозначает тройку любых объектов. Приведенные примеры показывают, что наряду со сложными существуют и простые, хорошо нам знакомые абстрактные понятия.

Понятия можно сравнивать по степени абстрактности. Чем более общим является понятие, тем выше оно расположено на иерархической лестнице. Вот простой пример такой цепочки понятий: люди, позвоночные, живые существа. Мы увидим, что по мере своего развития математика переходит от менее абстрактных понятий к более абстрактным. Как ни странно, именно этот процесс облегчает ее усвоение.

Как рождается понятие, и могло ли оно появиться раньше, чем появилось в действительности? Возьмем два физические термина: «энергия» и «работа». Как физические термины они были придуманы только тогда, когда выяснилось, что в механических процессах увеличение величины, названной энергией системы, равно произведенной работе. Можно привести множество других примеров, показывающих, что в науке жизнеспособные понятия возникают с возникновением новой теории. Они создают язык этой теории.

Приведем еще один пример: понятие «группа крови». Оно появилось, когда обнаружилось, что все люди могут быть разбиты на пять групп по свойствам их крови. Эти группы перенумеровали. По группам крови пострадавшего и донора созданная теория определяет, допустимо ли ее переливание. Данный пример не только демонстрирует прикладное значение абстрактных понятий, но и указывает на их классифицирующую способность: понятие «группа крови» выделяет из совокупности всех людей тех, которые имеют, скажем, вторую группу крови. Подобным же образом понятие «квадратное уравнение» выделяет строго определенную группу уравнений. Зачем нужно такое выделение? Существует теория квадратных уравнений, и если мы установили, что данное уравнение сводится к квадратному, то к нему можно применить все выводы теории, например решить его с помощью известной формулы.

Теперь скажем несколько слов о предрассудках, окружающих математику и мешающих ее изучению. Очень распространенный предрассудок выражается формулой «математика — сухая наука». Скорей

всего, здесь речь идет об алгебраических или тригонометрических преобразованиях, которые, как известно, следует делать в соответствии с определенными правилами. Если не знать, откуда взялись эти правила и зачем нужны преобразования, то, конечно, можно возненавидеть и алгебру, и тригонометрию. Впрочем, если Вы имели несчастье попасть к плохому учителю, то можно возненавидеть и любую другую науку (и музыку «с ее гаммами»).

Упомянем еще один предрассудок. Иногда говорят, что математика представляет собой язык. Думается, это ошибочная точка зрения. Вряд ли кто-нибудь приравнивает художественную литературу языку. Считать математику языком столь же нелепо. Другое дело, что существует язык математики. Освоить его нелегко, но все же легче, чем изучить в том или ином объеме математику, не пользуясь ее языком.

Любой из языков изучить трудно. В каждом из них огромное количество слов, и выучить хотя бы самые нужные нелегко. Да и грамматику сложно выучить. Именно поэтому человек почти неизбежно попадает в какую-нибудь языковую ловушку при попытке выразить свою мысль на чужом языке.

Язык математики труден по совершенно другой причине. Вы не можете начать изучение его словарного запаса с любого слова, например со слова «интеграл». Чтобы объяснить это понятие, необходимо использовать множество других понятий, таких как функция, предел, сумма и т. д. Да и последние в свою очередь нуждаются в определениях. Это создает известные трудности при изучении математики. В настоящей книге мы постарались преодолеть подобные трудности. Рассказывая о той или иной теории, мы сначала объясняем, зачем она нужна, потом говорим о наиболее интересных ее результатах и, наконец, переходим к приложениям теории. Что же узнает читатель, прочтя первую часть книги? Он познакомится с некоторыми яркими моментами истории математики, увидит известную логику в ее развитии, поймет, что знакомые ему школьная и, возможно, высшая математика очень отличаются от современной. После прочтения этой книги читателю будет легче оценить назначение и важность различных разделов курса математики в высшей школе; он станет видеть в абстрактных понятиях инструменты, позволяющие за короткое время узнать и усвоить многое. Возможно, читателю будет интересно ознакомиться со всей книгой, возможно — с отдельными ее частями.

Считаем своим долгом предостеречь читателя от возможного заблуждения. Математика практически необъятна. Если сравнить ее с неизученным континентом, то чтение этой книги подобно путешествию по одной из его рек. Не больше! Следует также иметь в виду, что наш рассказ о математике доведен до первой половины двадцатого века и почти не затронул второй половины. С примерным содержанием книги можно ознакомиться по оглавлению.

Глава 1. ПРОСТЕЙШИЕ СТРУКТУРЫ

Между математическими абстракциями и абстракциями языка существует важное различие. Абстрактные понятия в математике, в отличие от языковых абстракций, не допускают различных толкований: все люди, знакомые с определением данного понятия, понимают его совершенно одинаково. В языке это, конечно, не так. Многие понятия, в том числе и абстрактные, не нуждаются в определениях, хотя при желании и для них можно подыскать определение. В качестве примера можно рассмотреть слово «лошадь».

Другое важное отличие: в математике абстрактные понятия организованы в коллективы или, как их принято называть, структуры. С точки зрения математика каждая структура — это изолированный мир со своей системой аксиом, который можно изучать, игнорируя до поры, до времени его связи с другими структурами.

§ 1.1. Целые числа сами по себе

Поясним сказанное на примере трех простейших структур. Все они связаны с целыми числами. Каждое целое число является абстрактным понятием. Отметим своеобразие этих понятий.

Если мы говорим, что на столе стоят три чашки, то мы ничего не сказали о каждой чашке в отдельности. Наше высказывание относится только к коллективу чашек и обнаруживает много важных свойств этого коллектива; например, он достаточен для чаепития трех персон.

Самой простой из трех структур является совокупность всех целых чисел, рассматриваемых без каких-либо арифметических действий. Теория этой структуры, создававшаяся веками, выработала различные способы счета и записи целых чисел. Вероятно, самый примитивный способ записать число — сделать соответствующее количество зарубок (способ Робинзона Крузо). Труднее дать числам названия. Один из методов: «один, два, много» годится далеко не всегда. Римляне ввели два принципиальные усовершенствования в метод зарубок. Во-первых, число пять — пять «зарубок» — они заменили одним значком V. Для чисел десять, пятьдесят, сто и пятьсот они ввели обозначения X, L, C, и D соответственно. Во-вторых, величина числа, изображаемого комбинациями этих символов, зависит от их расположения. Например, комбинация CX означает, как и следовало ожидать, сто десять, а XC — девяносто.

Числа, записанные принятым в настоящее время способом, называются арабскими, поскольку именно арабские математики позаимствовали их у индийцев и познакомили с ними европейцев. Данная система записи использует только десять цифр (от нуля до девяти). Для того чтобы прочесть вслух любое число, нужно знать названия этих десяти цифр и еще несколько слов типа «десять», «сто», «тысяча», «миллион».

§ 1.2. Целые числа со сложением и умножением

Более сложная структура получается, если целые числа рассматриваются не сами по себе, а вместе с операцией сложения. Здесь перед теорией стояла задача выработать практические методы сложения (и вычитания) чисел, особенно больших. Было ясно, что возможности устного счета весьма ограничены. Что касается письменного счета, то использование римских чисел вряд ли заметно облегчало бы выполнение этих действий. Совсем по-другому проявили себя арабские числа. Они будто специально созданы для того, чтобы сложение даже очень больших чисел сделать детски легкой задачей. Впрочем, вероятно, они действительно были созданы именно с этой целью.

Еще более сложная структура — целые числа вместе с двумя операциями: сложением и умножением. Арабские числа оказались идеально приспособленными и для выполнения умножения. Теория этой структуры называется теорией чисел.

Теория чисел имеет интересную историю. Она замечательна полным отходом от запросов практики и бурным развитием теории в соответствии с ее внутренней логикой.

В рамках этой структуры можно рассматривать разложение чисел на множители. Отсюда один шаг до введения простых чисел, то есть чисел, которые не могут быть представлены в виде произведения двух не равных единице целых чисел. Простые числа обладают огромным количеством удивительных свойств. Часть этих свойств твердо установлена (то есть доказана), часть — угадана и впоследствии доказана, часть свойств не установлена до сих пор, хотя и подтверждается множеством примеров.

§ 2.1. Две составляющие теории чисел

Скажем несколько слов о самой теории чисел. Она состоит из двух частей. Одна часть содержит важные, фундаментальные результаты. Среди них теорема Евклида о существовании бесконечного множества простых чисел и теорема о единственности разложения любого целого числа на простые множители.

Еще одним примером является теорема о двух простых числах.

Пусть p и q — два фиксированные простые числа. Тогда существует пара целых чисел m и n такая, что

$$mp - nq = 1.$$

Эта теорема может пригодиться, если Вам нужно заплатить кассиру один рубль. У Вас в кармане есть только пятирублевые купюры, а у кассира — только трешки. Вы легко подбираете фигурирующие в теореме множители m и n . Подобные теоремы сравнительно легко доказываются.

Другая часть теории чисел представляет собой собрание большого количества задач, очень простых по формулировке и очень трудных для решения. Эти задачи обладают невероятной притягательной силой. Каждая такая задача — вызов, брошенный математикам, и многие, в том числе великие, математики не могли удержаться от соблазна поднять перчатку (при этом они принесли немалую пользу и другим разделам математики). Напрашивается аналогия с покорением неприступных горных вершин. Здесь похоже многое — и упорство в достижении цели, и затраченные усилия, и слава в случае успеха, и,

боюсь, малая объективная значимость одержанной победы. Конечно, и в том, и в другом случае победа показывает, на что способен Человек.

Приведем две подобного рода теоремы. Первая из них — теорема Ферма о двух квадратах: если простое число после деления на 4 дает в остатке 1, то оно равно сумме квадратов двух целых чисел.

§ 2.2. Великая теорема Ферма

О второй задаче поговорим более подробно. Это великая, или последняя, теорема Ферма (Пьер Ферма, 1601–1665). Она состоит в том, что никакие три целые числа a , b и c не могут удовлетворять равенству

$$a^n + b^n = c^n$$

ни при каком целом значении числа n , за исключением $n = 1$ и $n = 2$. Простые рассуждения, которые мы опустим, показывают, что теорема Ферма была бы доказана полностью, если бы удалось установить ее справедливость в тех случаях, когда n равно четырем либо является простым числом. Случай $n = 4$ легко поддается анализу и был рассмотрен Ферма. Поэтому в дальнейшем число n предполагается простым. Дадим краткое перечисление попыток доказать эту теорему. Казалось бы, сама длина приводимого ниже списка показывает, что он — список неудач. В действительности это летопись побед (список неудач — ошибочных доказательств — мы опускаем).

Начнем со знаменитой женщины-математика Софии Жермен (1776–1831). Она рассмотрела частный случай теоремы Ферма, при котором ни одно из чисел a , b , c не делится на n . Для этого случая она доказала, что теорема верна, если оба числа n и $2n + 1$ являются простыми. Адриен Лежандр (1752–1833) усилил предыдущий результат, показав, что условие С. Жермен можно ослабить, заменив его условием «хотя бы одно из чисел

$$4n + 1, \quad 8n + 1, \quad 10n + 1, \quad 14n + 1, \quad 16n + 1$$

является простым». Последнее условие охватывает все простые числа меньше ста и много других простых чисел. В 1849 г. Эрнст Куммер (1810–1893) еще более ослабил это условие. Одновременно он положил начало теории идеалов¹⁾ — одного из краеугольных камней современ-

¹⁾ Простейшим идеалом является набор всех чисел, кратных некоторому числу, например

$$0, 5, -5, 10, -10, 15, \dots$$

Отличительное свойство подобных наборов состоит в том, что при умножении на любое целое число каждое число набора переходит в некоторое другое число из *этого же* набора (или остается на месте при умножении на единицу). Впоследствии понятие идеалов было обобщено на более сложные

ной алгебры (так что увлечение Куммера теоремой Ферма пошло на пользу не только теории чисел).

Пропустим сто лет, в течение которых ряд математиков улучшал эти результаты. Заметим только, что в их числе был и Фробениус, один из математиков, внесших существенный вклад в теорию групп. Общими усилиями рассматриваемый случай Софи Жермен был доказан для всех показателей степени вплоть до астрономического числа 714 591 416 091 398 (1991), но все же не для абсолютно всех показателей.

В общем случае (без ограничения Софии Жермен) теорема была доказана Леонардом Эйлером (1707–1783) для показателя $n = 3$ и, как мы уже говорили, Пьером Ферма — для показателя $n = 4$. Лежен Дирихле (1805–1859) и Жозеф Лагранж (1736–1813) установили ее справедливость для случая $n = 5$, затем опять Дирихле — для $n = 14$ (1832). Случай $n = 7$ был доведен до конца Ляме (1839). В течение последующих 150 лет было получено много дополнительных результатов, но в полном объеме теорема Ферма доказана не была, хотя выяснилось, что она справедлива для всех показателей степени вплоть до четырех миллионов.

В 1995 году теорема Ферма была, наконец, доказана в полном объеме Уайлсом и Тейлором.

Закончилась эра, начавшаяся в момент, когда после смерти Ферма на полях одной из принадлежавших ему книг обнаружили формулировку знаменитой теоремы и слова «я нашел поистине удивительное доказательство этого предложения, но оно не уместится на полях»¹⁾.

А еще говорят, что математика — сухая наука!

алгебраические структуры. Так, совокупность полиномов, численное значение которых в некоторой фиксированной точке равно нулю, образует идеал в множестве всех полиномов.

¹⁾ Конечно, интересно было бы узнать, действительно ли Ферма в одиночку совершил то, что после него не смогли осуществить несколько поколений выдающихся математиков. Одни считают, что запись, сделанная Ферма — результат заблуждения, другие верят в то, что ему удалось найти строгое доказательство. Однако многие просто не знают, какой версии отдать предпочтение.

Усложнение численных структур пошло и по другому пути — по пути расширения понятия числа. Основные этапы этого пути — нуль, отрицательные числа, дробные (они же рациональные), иррациональные и комплексные числа. Мы остановимся на трех последних видах чисел.

§ 3.1. Рациональные числа

Все знают, что такое простые дроби, как их сравнивать по величине, по каким правилам производить над ними арифметические действия. Однако откуда взялись эти правила? Попытаемся показать, что они являются результатом абстрагирования.

Если Вы хотите что-либо пересчитать, то дробные числа Вам совершенно не нужны. Совсем другое дело, если требуется что-либо измерить. В этом случае прежде всего нужно выбрать единицу измерения, желательно общепринятую. Пусть для определенности речь идет о длине, а в качестве единицы выбран метр. Вы приступаете к измерению и обнаруживаете, что измеряемая длина D , скажем, больше пяти метров, но меньше шести. В ряде случаев может потребоваться более определенный результат. Чтобы его получить, Вы делите метр на N равных частей и обнаруживаете, что длина D больше чем d таких частей, но меньше чем $d + 1$. Если такая точность Вас устраивает, Вы пишете $D = d/N$, прибегая тем самым к дробям.

Зачем может понадобиться сложение дробей? Есть много прикладных задач, которые для получения ответа требуют сложения дробей. Например, в стакан насыпали две ложки сахара; в одной было $15/7$ грамма, а в другой — $13/6$ грамма. Сколько всего насыпано сахара в стакан? Ответ, разумеется, получается сложением соответствующих дробей. «Где же здесь абстракция?» — спросит читатель. Абстракция заключается в том, что одно и то же арифметическое действие — сложение дробей — обслуживает большой класс задач самого разного содержания и в момент, когда дело доходит до фактического подсчета, конкретное содержание задачи, если она принадлежит этому классу задач, оказывается несущественным. Именно класс задач определяет правило сложения дробей и, конечно, всегда одно и то же правило. В нашем случае для вывода этого правила можно рассуждать следующим образом.

Насыпать $15/7$ это все равно, что насыпать 15 раз по одной седьмой грамма. Если пользоваться в 6 раз меньшими порциями, то чтобы насыпать такое же количество сахара, следует произвести не 15 операций, а в 6 раз больше, то есть 90 операций. Значит, $15/7 = 90/42$. Аналогично $13/6 = 91/42$. Ответ задачи: $181/42$. Мы видим, что правило сложения абстрактных дробей диктуется любой из конкретных задач, для которых это правило создано. Кстати, выделение указанного класса задач является хорошим примером удачной абстракции.

Зададим читателю вопрос, который на первый взгляд кажется простым. Вот этот вопрос. Пусть последовательность

$$a_1, a_2, \dots, a_n, \dots$$

является некоторой бесконечной совокупностью рациональных чисел.

Спрашивается:

- а) *могут* ли все члены последовательности быть меньше единицы?
- б) *обязательно ли* среди членов последовательности существует наибольшее число, то есть число, которое не меньше ни одного из других чисел последовательности, но может быть равно некоторым из них?

Оказывается, на первый вопрос следует ответить положительно, а на второй отрицательно. Для того чтобы в этом убедиться, достаточно взглянуть на последовательность

$$\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots, \frac{n-1}{n}, \dots$$

Действительно, все числа этой последовательности меньше единицы, и среди них нет наибольшего.

Глава 4. ИРРАЦИОНАЛЬНЫЕ ЧИСЛА

Нельзя ли обойтись без иррациональных чисел?

Мы не утверждаем, что без них невозможно обойтись. Вероятно, можно, если примириться с потоком громоздких утверждений, который буквально захлестнет все учебники математики. Думается, что одного примера окажется достаточно для того, чтобы на всю жизнь убедить читателя не сторониться иррациональных чисел.

В физике часто существенную роль играют различного рода «особые точки». Простейшим примером является набор координат, определяющий положение устойчивого равновесия механической системы. Теория малых колебаний отвечает на вопрос, как будет двигаться система, если вывести ее из положения равновесия, придав малые скорости ее точкам. Теория обнаружила существование собственных частот у такой системы, явление резонанса, возникновение обертонов и многое другое. Математический аппарат теории малых колебаний — ряды Тейлора. Для простоты ограничимся рядом Тейлора для функции одной переменной. Он выглядит следующим образом:

$$F(x) = F(a) + F'(a)(x - a) + \frac{1}{2}F''(a)(x - a)^2 + \dots,$$

где $F'(a)$, $F''(a)$ — производные от функции $F(x)$ первого и второго порядка. Спрашивается, как записать эту формулу, если a — иррациональное число, а мы решили не пользоваться иррациональными числами? Заменить a близким к нему рациональным числом? Непонятно. Близким к чему? Ведь числа-то a нет!

Фактически мы столкнулись с лингвистической трудностью и ощутились в положении человека, который забыл половину всех слов родного языка: мы даже не имеем права пользоваться буквами для обозначения иррациональных чисел.

Можно было бы привести множество подобных примеров, но не будем «ломиться в открытую дверь».

Все эти трудности исчезают при простом признании иррациональных чисел и знакомстве с некоторыми свойствами совокупности вещественных чисел, то есть объединения рациональных и иррациональных чисел. Речь об этом пойдет в гл. 16. Проследим основные этапы истории иррациональных чисел.

§ 4.1. Иррациональные числа и греческая математика

«Трагедия» рациональных чисел состоит в том, что их недостаточно много. Уже греческие математики знали, что рациональных чисел недостаточно для обозначения всех отрезков. Так, диагональ квадрата со стороной, равной единице, согласно теореме Пифагора обладает следующим свойством: квадрат ее длины равен двум. Но среди рациональных такого числа нет¹⁾. Поэтому следовало либо отказаться от использования чисел для обозначения длин, либо расширить понятие числа.

Античные математики пошли по первому пути. Легче всего понять суть этого пути на примере теоремы Пифагора. Ее современная формулировка имеет вид

$$a^2 + b^2 = c^2,$$

где a и b — длины катетов прямоугольного треугольника, а c — длина его гипотенузы. Но что такое a^2 ? Это a , умноженное на a . Однако перемножать можно только числа. А если хотя бы одна из длин не является числом? Тогда приведенная формулировка становится бессмысленной! Поэтому греки формулировали теорему Пифагора следующим образом: площадь квадрата, построенного на гипотенузе, равна сумме площадей квадратов, построенных на катетах. Доказательство велось соответственно, то есть без использования чисел для выражения длин. Площади двух прямоугольников считались равными, если одна сторона первого прямоугольника была во столько же раз меньше стороны второго, во сколько другая сторона первого больше другой стороны второго прямоугольника. В этом определении используется пропорция. Установить наличие пропорции не прибегая к числам можно, если воспользоваться чисто геометрическим понятием подобия треугольников. Мы видим, на какие ухищрения приходится идти, если отказаться от иррациональных чисел. В известном смысле это тупиковый путь.

¹⁾ Это легко доказать (еще одна задача по теории чисел). Рассмотрим произвольную дробь вида m^2/n^2 . Числа m и n можно разложить на простые множители. Например, если $m = 60$, то $m = 2 \times 2 \times 3 \times 5$. При переходе от числа m к его квадрату m^2 число одинаковых сомножителей удваивается: $m^2 = 2 \times 2 \times 2 \times 2 \times 3 \times 3 \times 5 \times 5$, так что каждое простое число, которое входит в это разложение, входит в него четное число раз. То же, разумеется, относится и к числу n^2 . Поэтому после сокращения дробь m^2/n^2 содержит каждый простой множитель четное число раз либо в числителе, либо в знаменателе. Если эта дробь равна целому числу, то она представляет собой квадрат какого-либо простого числа или произведение нескольких таких квадратов. Во всех случаях она не может равняться никакому простому числу, и, в частности, не может равняться двум.

Он не позволяет сформулировать ни одного соотношения, которое содержит произведение более трех чисел. Кроме того, невозможно придать геометрический смысл выражениям типа $a^2 + b$, поскольку a^2 — площадь, b — длина, а как сложить площадь с длиной?

Тем не менее греческая математика отказалась от иррациональных чисел.

§ 4.2. Первое признание иррациональных чисел

Прошло более полутора тысяч лет. Многое изменилось. Если говорить о математике, то появились дифференциальное исчисление и связанное с ним интегральное исчисление (связанное, как деление с умножением, как любая обратная задача с прямой). Вместе они составили мощнейший математический аппарат — математический анализ. Перед математиками открылись панорама разнообразных задач, которые теперь поддавались решению, и безбрежная перспектива совершенно новых проблем, которые, чувствовалось, будут рано или поздно решены. Последнее было тем более увлекательно, что многие из этих задач и проблем являлись прикладными. Грань между математикой и физикой почти полностью исчезла. Великие ученые того времени — Исаак Ньютон (1642–1727), Готфрид Лейбниц (1646–1716), Леонард Эйлер (1707–1783), Яков (1645–1705), Иоганн (1667–1748) и Даниил (1700–1782) Бернулли, Даламбер (1667–1748) и другие — и в наши дни одинаково почитаются как физиками, так и математиками. Если учесть еще, что новый математический аппарат жадно впитывал в себя различного рода важные усовершенствования, нуждался в них, то становится ясным, что математики того времени должны были жить счастливой и полнокровной духовной жизнью.

Новые задачи привлекли внимание математиков к некоторым новым числам; наиболее известны среди них основание натуральных логарифмов — число Эйлера e и постоянная Эйлера ν .

В 1760 г. Иоанн Ламберт (1728–1777) доказал, что отношение длины окружности к ее диаметру — число π — является иррациональным (точнее, не является рациональным). Он же годом позже установил иррациональность числа e . Следует сказать, что открытие Ламберта не побудило математиков ни тогда, ни потом перестать пользоваться этими числами. Были найдены близкие к ним рациональные числа, и этого оказалось достаточно, чтобы молчаливо присоединить их к совокупности всех рациональных чисел и тем самым обогатить множество используемых в математике чисел. Впоследствии была установлена иррациональность и многих других чисел. По сути дела, подобные исследования были ближе к теории чисел, чем к математическому анализу. Если некоторое число, например π , возникало в процессе

решения некоторой задачи и требовалось его численное значение, то оно вычислялось с необходимой точностью, причем процесс вычисления не зависел от того, является ли искомое число рациональным.

Долгое время не требовалось никакой теории иррациональных чисел. Поясним, в сколь «бесправном положении» находились в то время иррациональные числа. У них не было определения! Ведь отрицательное определение не является определением. Разве можно определить лису словами «лиса — это не медведь»? Хороша была бы систематика животного мира, если бы она строилась с помощью подобных «определений»! А чем лучше определение «иррациональное число — это не рациональное число»? Вспомним, как определяется число π : π — это отношение длины окружности к ее диаметру. Говорит ли данное определение что-нибудь о природе других иррациональных чисел? Абсолютно ничего! Что же, для каждого такого числа придумывать свое определение? Проблему нужно решать как-то иначе.

§ 4.3. Новое действие — переход к пределу

Начнем с вопроса, зачем естествоиспытателям новейшего времени понадобилась теория иррациональных чисел? Думается, что причиной явилось вторжение в науку понятия предела. Это одно из фундаментальных понятий, лежащих в основе современной математики. Неизбежность его возникновения вытекает из его поразительной способности определять фундаментальные понятия многих прикладных наук. Чтобы не «ломиться в открытую дверь», ограничимся несколькими примерами. Без понятия предела невозможно определить, что такое мгновенная скорость точки или вообще мгновенная скорость любого процесса, нельзя определить касательную к кривой. Для определения производной функции также не обойтись без понятия предела и т. д. и т. п.

Но это только одна сторона медали. Не менее важным свойством пределов является их незаменимость при решении подавляющего большинства трудных (и не очень трудных) задач. Рассмотрим более подробно это очень важное свойство пределов и заодно дадим определение самого предела.

Предположим, что нужно решить задачу, состоящую в отыскании неизвестного числа A . Пусть это будет прикладная задача и число A нам требуется не само по себе, а например, для определения необходимой прочности деталей проектируемой конструкции или, скажем, для расчета параметров радиосхемы. В таком случае нет необходимости искать точное значение числа A . Достаточно найти приближенное значение, лишь бы его погрешность δ была достаточно мала. Кстати, величина δ допустимой погрешности должна быть отдельно оговорена. Пусть нам удалось найти способ получать сколь угодно точные

приближенные решения. Это означает, что какова бы ни была требуемая (положительная) погрешность δ , мы можем найти приближенное решение (обозначим его A_δ), погрешность которого меньше чем δ :

$$|A - A_\delta| < \delta.$$

В частности, мы можем вычислить любой член бесконечной последовательности приближенных решений

$$A^1, A^2, \dots, A^n, \dots, \quad (4.1)$$

обладающих следующим свойством: погрешность $|A - A^n|$ стремится к нулю с ростом n , например

$$|A - A^n| < 10^{-n}. \quad (4.2)$$

Число A называют *пределом последовательности* (4.1), если разности (4.2) стремятся к нулю.

Итак, каждый приближенный метод порождает последовательность, стремящуюся к решению. И наоборот, каждая стремящаяся к решению последовательность позволяет находить сколь угодно точные приближенные решения.

§ 4.4. Итерационный метод и теория пределов

Здесь нам придется сделать очень важное отступление для короткого рассказа об итерационном методе. Это удивительно универсальный метод, но без понятия предела он как птица с одним крылом. Поясним на возможно более простом примере несложную сущность данного метода. Пусть требуется найти корень уравнения

$$x^7 = 1 + x, \quad (4.3)$$

лежащий между единицей и двойкой. Эту задачу можно решить следующим образом. Возьмем для начала число $x_1 = 1$ и рассчитаем число x_2 по формуле

$$x_2 = (1 + x_1)^{1/7}.$$

Подобным же образом можно определить число x_3 : $x_3 = (1 + x_2)^{1/7}$, а затем и любой член бесконечной последовательности чисел $x_1, x_2, x_3, x_4, \dots$:

$$x_{n+1} = (1 + x_n)^{1/7}. \quad (4.4)$$

Всякая последовательность чисел, задаваемая каким-либо соотношением, позволяющим вычислить каждый последующий ее член по предыдущему, называется итерационной последовательностью. В частности, последовательность x_n ($n = 1, 2, \dots$) является итерационной. Оказывается, предел этой последовательности есть точное решение нашего уравнения!

Остановимся! Мы пропустили самое главное. Что скрывается за сделанным утверждением? Как его доказать? Вот один из способов.

Первый шаг. Исследуем последовательность (4.4).

Заметим, что все ее члены не меньше единицы и меньше двух. Это справедливо как для чисел x_1 и x_2 , так, в силу соотношений (4.4), и для всех остальных членов последовательности. Заметим еще, что x_2 больше x_1 . Из (4.4) следует, что такое же соотношение сохраняется для всех остальных членов рассматриваемой последовательности:

$$x_{n+1} > x_n, \quad n = 1, 2, 3, \dots$$

Итак, последовательность (4.4) является ограниченной (все ее члены меньше одного и того же числа) и монотонно возрастающей (каждый ее член больше предыдущего).

Второй шаг. Пользуясь теоремой: *Всякая ограниченная монотонно неубывающая числовая последовательность имеет предел*, заключаем, что последовательность (4.4) имеет предел, и этот предел может быть вычислен с любой степенью точности, если в качестве приближенного значения взять число x_n с достаточно большим номером n .

Если взять в качестве x_1 не единицу, а двойку, то получится другая последовательность чисел. Обозначим ее члены через

$$y_1 = 2, y_2, \dots, y_m, \dots$$

Эта последовательность является монотонно убывающей и ограниченной снизу и тоже имеет предел. Можно показать, что оба предела равны друг другу и совпадают с искомым корнем уравнения (4.3). Зачем же нам понадобилась вторая последовательность? Для строгой оценки погрешности! Дело в том, что общий предел двух построенных последовательностей больше любого члена первой последовательности и меньше любого члена второй. Следовательно, искомым корень x уравнения (4.3) заключен в интервале

$$x_n < x < y_m$$

при любом выборе индексов m и n .

Что демонстрирует приведенный пример? Мы с самого начала отказались от попытки найти точное значение корня рассматриваемого уравнения и действительно не нашли его. Нас интересовали приближенное значение корня и строгая оценка погрешности. Мы отыскивали обе эти величины!

Описанная схема использования понятия предела применима для широкого круга разнообразных задач, но не является единственной.

§ 4.5. Рациональные числа и рациональные точки

Первую строгую теорию иррациональных чисел создал Рихард Дедекин (1831–1916)¹⁾. Для ее изложения удобно, хотя и необязательно, вернуться к отрезкам, для обозначения длины которых у греков (да и у нас тоже) не хватило рациональных чисел. Будем представлять эти отрезки отложенными на прямой от одной и той же точки O , называемой началом отсчета. От этой же точки отложим всевозможные отрезки, длины которых в выбранном масштабе равны рациональным числам. Совокупность концов этих отрезков назовем рациональными точками. Все остальные точки прямой условимся называть иррациональными. Прямая, рассматриваемая как совокупность всех рациональных и иррациональных точек, обычно называется числовой прямой.

Итак, у нас есть рациональные числа, рациональные и иррациональные точки, но нет пока иррациональных чисел!

Рациональные точки образуют плотное множество в том смысле, что на числовой прямой нет ни одного отрезка, свободного от них. Это, конечно, не означает, что на числовой оси нет места для других точек²⁾.

§ 4.6. Разрезы Дедекинда

Возьмем совершенно произвольную иррациональную точку A на числовой прямой. Она делит числовую прямую на две части: первая состоит из всех точек, лежащих слева от нее, вторая содержит все точки, лежащие справа. Тем самым все рациональные точки разделены на два класса: левый, состоящий из всех рациональных точек, расположенных на левой части числовой прямой, и правый, имеющий аналогичный смысл. Поскольку любой рациональной точке соответствует

¹⁾ Позволим себе привести подлинный случай из жизни этого ученого. Последние годы жизни он не печатался в математических журналах. Возможно, поэтому разнесся слух о его смерти и в одном из журналов появился посвященный ему некролог, разумеется, с указанием даты кончины. Дедекин дождался конца года и послал опровержение примерно такого содержания: «Не могу утверждать, что число или месяц моей смерти указаны неправильно, что же касается года смерти, то тут, безусловно, вкралась ошибка».

²⁾ Сама по себе плотность некоторого точечного множества не исключает присутствия на числовой прямой других точек, то есть точек, ему не принадлежащих. Подтвердим это примером. Точки, изображаемые несократимыми дробями с нечетными знаменателями, «мирно уживаются» на числовой прямой с точками, соответствующими несократимым дробям с четным знаменателем. Между тем, каждое из этих двух множеств плотно на числовой прямой.

рациональное число, то тем самым все рациональные числа разбиты на два класса. Назовем их нижним и верхним (рис. 4.1). Каждый из этих классов обладает важным для дальнейшего и на первый взгляд парадоксальным свойством: в левом классе нет самой правой точки, а в правом классе нет самой левой точки! В действительности в этом нет ничего удивительного. Если бы в левом классе существовала самая правая точка, то интервал между ней и исходной иррациональной точкой не содержал ни одной рациональной точки, что противоречило бы свойству плотности множества рациональных точек. Подобное рассуждение легко провести и применительно к правому классу.



Рис. 4.1. Деление числовой прямой произвольной иррациональной точкой A ; a — рациональное число из нижнего класса, b — рациональное число из верхнего класса

Вспомним, что каждую рациональную точку можно представить как рациональное число. Таким образом, любая иррациональная точка порождает разбиение совокупности всех рациональных чисел на два класса, обладающие следующими свойствами:

- 1) каждое рациональное число из нижнего класса меньше любого рационального числа из верхнего класса;
- 2) каждое рациональное число принадлежит одному из этих классов;
- 3) в нижнем классе нет наибольшего числа, а в верхнем классе нет наименьшего.

Условимся называть *разрезом* любое разбиение множества рациональных чисел на два класса, если оно обладает указанными тремя свойствами. Мы получили абстрактное определение разреза — основного понятия теории Дедекинда.

Подчеркнем, что в этом определении не используется понятие иррациональной точки и, тем более, не упоминаются иррациональные числа.

Идея Дедекинда состоит в том, что понятие разреза и является *определением* иррационального числа! В качестве иллюстрации рассмотрим разрез, соответствующий числу $\sqrt{2}$. Нижний класс этого разреза состоит из всех отрицательных рациональных чисел и тех положительных рациональных чисел, квадрат которых меньше двух; верхний класс состоит из всех остальных рациональных чисел.

§ 4.7. Теория Дедекинда

Вряд ли современники серьезно отнеслись бы к этой идее, если бы Дедекинд не создал с ее помощью теорию действительных чисел. При соединяя к совокупности всех рациональных чисел совокупность всех разрезов, мы получим совокупность действительных чисел. Иными словами, каждое действительное число является либо разрезом, либо рациональным числом, или, возвращаясь к привычному словоупотреблению, рациональным либо иррациональным. На первый взгляд мы вернулись к наивному определению иррационального числа как числа, не равного никакому рациональному. В действительности это не так, поскольку теперь мы имеем однозначное определение иррационального числа, а его наивное определение — простое следствие.

Дедекинд показал, что отношения «равно», «меньше» и «больше» естественно распространяются на все действительные числа, также как и все четыре арифметические действия (с сохранением известных из арифметики свойств этих действий). Более того, оказалось возможным достроить теорию пределов, установив несколько полезных признаков их существования. Оказалось также возможным построить теорию так называемых непрерывных функций. Говоря образно, удалось выйти на оперативный простор.

Вот почему математики высоко ценят иррациональные числа и их теорию.

На первый взгляд предложение представлять иррациональные числа как разрезы множества всех рациональных кажется чудовищным. Но очень скоро становится ясно, что задание разреза, соответствующего числу α , попросту определяет место последнего в «строю» рациональных чисел. С другой стороны, определение разреза является логически безукоризненным. Поэтому все свойства иррациональных чисел можно вывести, отождествляя их с разрезами. Продемонстрируем это, показав, как вводятся отношения порядка, суммы и произведения.

Два иррациональных числа равны друг другу, если их нижние классы совпадают. Число α меньше числа β , если все рациональные числа нижнего класса α содержатся в нижнем классе β , но не исчерпывают его (рис. 4.2).

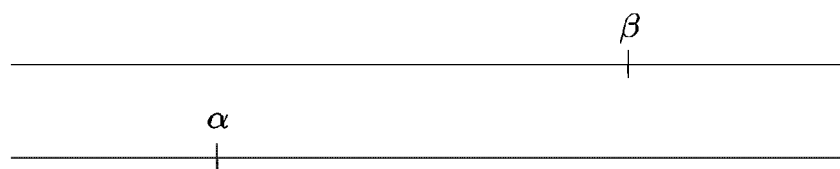


Рис. 4.2. Сравнение иррациональных чисел: $\alpha < \beta$

Подобным же образом можно ввести сумму $\alpha + \beta$ любых двух иррациональных чисел. Определим ее как разрез, нижний класс которого состоит из всех сумм вида $a + b$, где a — любое число из нижнего

класса α , а b — любое число из нижнего класса β . Аналогично определяется верхний класс разреза $\alpha + \beta$. Столь же просто вводится произведение $\alpha\beta$ двух положительных иррациональных чисел: это разрез, верхний класс которого состоит из всевозможных произведений чисел из верхних классов α и β соответственно. Умножение отрицательных чисел сводится к умножению положительных, если воспользоваться правилом знаков: $(-3) \times (-5) = 15$, $(-3) \times (+5) = -15$. Нетрудно определить сумму $a + b$ двух чисел и в том случае, когда число a является рациональным, а b — иррациональным. Нужно прибавить число a ко всем числам нижнего класса b . Мы получим нижний класс суммы $a + b$. Подобным же образом определяется и ее верхний класс.

§ 4.8. Основная теорема теории пределов

Понятие предела числовой последовательности является весьма тонким. Хотим еще раз напомнить, что это одно из основополагающих понятий математики, а значит, и всего современного естествознания. Мы не будем предполагать, что оно знакомо читателю, и дадим его определение. С этой целью предварительно введем два более простые понятия: «окрестность числа» и «почти вся последовательность». Определяя первое из них, мы используем теорию Дедекинда.

Пусть α — произвольное действительное число, a и b — два произвольные рациональные числа, взятые из нижнего и верхнего классов числа α соответственно. Совокупность всех действительных чисел, больших a и в то же время меньших b , называется окрестностью числа α (рис. 4.3). Принято говорить, что все эти числа *лежат* в окрестности (a, b) . Поскольку числа a и b произвольны, у всякого числа α существует бесконечное множество окрестностей.



Рис. 4.3. Окрестность (a, b) действительного числа A

Договоримся понимать под числовой последовательностью любую пронумерованную совокупность чисел. Условимся говорить, что *почти все* члены данной последовательности лежат в данной окрестности, если отбрасывая достаточно большое, но конечное число ее первых членов, можно получить последовательность, *все* члены которой лежат в указанной окрестности (рис. 4.4).



Рис. 4.4. Почти вся последовательность лежит в окрестности (a, b)

Теперь легко определить, что такое предел последовательности.

Определение. Число α является пределом данной последовательности чисел, если, какова бы ни была окрестность (a, b) числа α , почти все члены последовательности лежат в этой окрестности.

Сама последовательность в этом случае называется сходящейся.

Заметим, что в силу данного определения предел последовательности можно с любой степенью точности приближенно заменить одним из ее членов, если номер этого члена достаточно велик. В этом смысле можно сказать, что решение задачи найдено, если найдена последовательность, предел которой является искомым решением. Сам предел вычислять не обязательно, даже если он равен какому-либо «знакомому» числу, например двум или $\sqrt{2}$.

Основная теорема теории пределов состоит в следующем.

Теорема. *Всякая ограниченная монотонно возрастающая бесконечная последовательность действительных чисел имеет предел.*

Условия теоремы означают, что все члены последовательности меньше некоторого числа M , а каждое последующее входящее в нее число не меньше предыдущего.

Докажем эту теорему. Мы имеем дело с теоремой существования, в данном случае — существования предела последовательности. Как многие доказательства подобных теорем, оно состоит из двух частей. Первая часть — конструкция. Мы построим число, являющееся, как выяснится во второй части, тем самым пределом, существование которого требуется доказать.

Итак, конструкция. Разобьем все рациональные числа на два класса. К первому классу отнесем все числа, которые больше всех членов рассматриваемой последовательности, ко второму — все остальные рациональные числа. Если среди чисел первого класса нет наименьшего, а среди чисел второго класса нет наибольшего, мы получаем разрез. Обозначим через α действительное число, которое определяется этим разрезом. Если же в первом классе есть наименьшее число, обозначим его α . Аналогично поступим, если во втором классе есть наибольшее число. Конструкция закончена; остается доказать, что построенное число является пределом.

Ограничимся случаем, когда α — иррациональное число. Покажем, что α является пределом нашей последовательности. Выберем произвольно число a из нижнего класса и число b — из верхнего. Число a не может превосходить все члены последовательности, иначе оно принадлежало бы верхнему классу. Значит, среди членов последовательности найдется число большее a . Пусть k — номер этого числа. Отбросим первые $k - 1$ членов последовательности. Все оставшиеся числа и подавно больше a , поскольку последовательность является монотонно возрастающей. С другой стороны, все эти числа меньше b (по определению b). Мы видим, что почти все члены последовательности принадлежат окрестности (a, b) . Так как (a, b) — совершенно произвольная окрестность числа α , это число оказывается пределом

рассматриваемой последовательности в полном соответствии с определением. Теорема доказана.

Хотелось бы обратить внимание читателя на изящество доказательства: никаких преобразований, никаких сложных логических умозаключений. Все сделано как бы в белых перчатках.

§ 4.9. Бесконечные десятичные дроби

Очень удобным способом записи действительных чисел являются бесконечные десятичные дроби. Продемонстрируем это на примере. Рассмотрим следующую монотонно возрастающую последовательность рациональных чисел:

$$1,2; 1,27; 1,271; 1,2718; \dots \quad (4.5)$$

Каждый ее элемент получается из предыдущего добавлением еще одного произвольно выбранного десятичного знака. Это позволяет записать последовательность в виде одной бесконечной десятичной дроби — $1,2718\dots$ Подчеркнем, что эта дробь пока является не числом, а только удобной формой записи нашей бесконечной числовой последовательности. Вспомним, однако, теорему о существовании предела у всякой монотонно возрастающей ограниченной числовой последовательности. Последовательность (4) является именно такой — все ее члены меньше чем $1,3$. Поэтому последовательность (4) имеет предел — некоторое вещественное число α . Удобно и совершенно естественно записывать это число в виде приведенной выше бесконечной десятичной дроби.

§ 4.10. Обратная задача теории пределов

Теория пределов содержит некоторые примеры и приемы вычисления пределов. Однако основную роль с точки зрения приложений играет обратная задача: по заданному условию задачи, но неизвестному пределу найти сходящуюся к нему последовательность. Поясним эту мысль с помощью двух примеров.

1. *Вычисление числа π* , то есть отношения длины окружности к ее диаметру. В высшей математике доказывается, что последовательность чисел, определяемая соотношениями

$$S_0 = 0, \quad S_{k+1} = S_k + 1/[(4k+1)(4k+3)], \quad k = 0, 1, 2, \dots,$$

имеет своим пределом число $\pi/8$. Эта формула демонстрирует решение обратной задачи теории пределов. Она позволяет найти число π со сколь угодно высокой точностью. Числа S_k можно рассматривать как приближенные значения числа $\pi/8$.

Если какой-либо приближенный метод позволяет найти некоторую величину со сколь угодно большой точностью, то это, собственно,

означает, что с его помощью можно построить последовательность приближений, имеющую своим пределом данную величину. Разность между рассчитываемой величиной и ее приближенным значением называется погрешностью приближения или его точностью. Строго говоря, без оценки погрешности польза приближенного метода сомнительна. Можно показать, что в рассмотренном выше примере имеет место следующая оценка погрешности:

$$0 < \frac{\pi}{8} - S_k < \frac{1}{4k+5}.$$

Мы видим, что погрешность стремится к нулю очень медленно. Существуют значительно лучшие методы приближенного вычисления числа π .

Качество двух разных приближенных методов естественно сравнивать по трудозатратам, необходимым для достижения с их помощью заданной точности. Поэтому, если при некоторой требуемой точности один метод лучше другого, то при другой он может оказаться хуже.

2. *Вычисление квадратного корня из двух.* Предлагаемый метод состоит в построении последовательности чисел с помощью следующего соотношения:

$$u_{k+1} = \frac{1}{2}u_k + \frac{1}{u_k}, \quad k = 0, 1, 2, \dots,$$

при заданном значении $u_0 = 1$. Потратим несколько минут на анализ приведенной последовательности. Прежде всего, заметим, что если она имеет предел (скажем, u), то этот предел должен удовлетворять соотношению

$$u = \frac{1}{2}u + \frac{1}{u}$$

и, следовательно, является квадратным корнем из двух. Поэтому стоит попытаться установить сходимость последовательности u_k . Предположим, что одно из чисел u_k не меньше единицы и не больше двух. Тогда каждое из двух слагаемых правой части рекуррентного соотношения не больше единицы и не меньше половины. Отсюда следует, что свойство $1 \leq u_k \leq 2$ является «наследственным»: оно передается последующему члену последовательности от предыдущего и является, таким образом, общим свойством всех ее членов, поскольку $u_0 = 1$. Следующий шаг состоит в том, что переписав определяющее соотношение в виде

$$u_{k+1} - \sqrt{2} = \frac{(u_k - \sqrt{2})^2}{2u_k},$$

мы получим неравенство

$$u_{k+1} - \sqrt{2} < \frac{(u_k - \sqrt{2})^2}{2}.$$

Отсюда легко вывести оценку

$$0 < |u_{k+1} - \sqrt{2}| < 2^{-m}, \quad m = 1 + 2^k.$$

Поскольку при $k = 4$ показатель степени m равен 17, а $2^{17} = 131\,072$, уже четвертый член последовательности имеет погрешность меньшую одной сотысячной.

Глава 5. ВЕКТОРЫ

В начале двадцатого века стало ясно, что изложение многих разделов физики упрощается и становится более наглядным, если пользоваться векторами. Для нас важно, что вектор является новым объектом, имеющим почти столь же очевидное право на существование, как и число. Между тем, еще в начале двадцатого века некоторые математики считали, что рассуждения, оперирующие векторами, не могут считаться строгими. Из последующих глав будет ясно, насколько эти математики ошибались.

Появление векторов — это явление в истории математики; оно достойно того, чтобы попытаться в нем разобраться. В настоящей главе мы рассмотрим вопрос: что такое вектор? Окончательный ответ получится в результате нескольких последовательных сужений интуитивного представления о векторе.

Если говорить коротко, то вектор это стрелка. Еще можно сказать «вектор — это величина, имеющая длину и направление». В действительности, однако, один-единственный, рассматриваемый сам по себе вектор нельзя определить сколько-нибудь полно. Подобно тому, как, скажем, нельзя определить понятие «гражданин», не упоминая о существовании других граждан. Таким образом, мы будем говорить не об отдельных векторах, а о структурах, состоящих из стрелок-векторов. Без этого нельзя, например, вести речь о сложении векторов.

Мы поступим следующим образом. В качестве исходного возьмем интуитивное представление о стрелке и построим три структуры, содержащие всевозможные стрелки.

§ 5.1. Сложение векторов и умножение на числа

Данная структура описывается с помощью простых геометрических понятий. Условимся называть два вектора равными, если соответствующие им стрелки параллельны и имеют одинаковую длину. Совпадение начальных точек стрелок для равенства векторов не обязательно. Если хотят подчеркнуть последнее обстоятельство, то векторы называют свободными. В дальнейшем вместо слова «стрелка» будем говорить «вектор».

Заметим, что не все встречающиеся в физике векторы являются свободными. Сила, например, не является свободным вектором. При ее описании необходимо указывать не только ее величину и направление, но и точку приложения. При умножении на положительное

число направление вектора не изменяется, а его длина умножается на это число. Векторы можно складывать друг с другом по известному правилу параллелограмма. Обращаем внимание читателя на то, что оба эти свойства являются свойствами *совокупности* векторов и не содержатся в интуитивном образе стрелки.

Два вектора называют коллинеарными, если они параллельны друг другу. Три вектора называют компланарными, если их можно поместить в одной плоскости, переместив при необходимости их начальные точки. Всякая тройка некопланарных векторов $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ обладает важным свойством: любой вектор \mathbf{x} представим в виде суммы

$$\mathbf{x} = c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + c_3\mathbf{e}_3$$

с подходяще подобранными коэффициентами.

Впоследствии мы увидим, что описанная здесь структура является одной из конкретных реализаций абстрактной структуры — линейного конечномерного пространства.

§ 5.2. Скалярное произведение векторов

Добавим к уже введенным действиям еще одно — скалярное произведение двух векторов. Для этого введем в обращение еще одно геометрическое понятие — угол между векторами. Скалярное произведение двух векторов \mathbf{a} и \mathbf{b} обозначают символом (\mathbf{a}, \mathbf{b}) и определяют равенством

$$(\mathbf{a}, \mathbf{b}) = |\mathbf{a}||\mathbf{b}| \cos \varphi,$$

где $|\mathbf{a}|$ и $|\mathbf{b}|$ — длины векторов, а $\cos \varphi$ — косинус угла между ними. Скалярное произведение двух взаимно перпендикулярных векторов равно нулю. В физике скалярное произведение часто выступает в роли работы dA , произведенной силой \mathbf{f} при перемещении $d\mathbf{r}$: $dA = (\mathbf{f}, d\mathbf{r})$.

Выявим некоторые важные свойства скалярного произведения и с их помощью научимся обходиться без угла φ . Ясно, что $(\mathbf{a}, \mathbf{b}) = (\mathbf{b}, \mathbf{a})$ и что скалярное произведение двух векторов равно произведению длины одного вектора на проекцию на него другого. Отсюда сразу следует, что скалярное произведение линейно зависит от проектируемого вектора:

$$(\mathbf{a}, \mathbf{b} + \mathbf{c}) = (\mathbf{a}, \mathbf{b}) + (\mathbf{a}, \mathbf{c}); \quad (\mathbf{a}, k\mathbf{b}) = k(\mathbf{a}, \mathbf{b}) \quad (k — \text{число}).$$

В силу симметрии скалярное произведение линейно зависит и от первого сомножителя. Для вывода этих свойств мы использовали свойство линейности проекций. Зато теперь можно получить чисто алгебраическим путем некоторые геометрические теоремы. Так, тождество

$$(\mathbf{a} + \mathbf{b}, \mathbf{a} - \mathbf{b}) = (\mathbf{a}, \mathbf{a}) - (\mathbf{a}, \mathbf{b}) + (\mathbf{b}, \mathbf{a}) - (\mathbf{b}, \mathbf{b}) = |\mathbf{a}|^2 - |\mathbf{b}|^2$$

показывает, что диагонали ромба $(a + b)$ и $(a - b)$ взаимно ортогональны, поскольку длины его сторон равны. Еще одно тождество

$$(\mathbf{a} + \mathbf{b}, \mathbf{a} + \mathbf{b}) = (\mathbf{a}, \mathbf{a}) + 2(\mathbf{a}, \mathbf{b}) + (\mathbf{b}, \mathbf{b}),$$

доказывает теорему Пифагора, если применить его к прямоугольному треугольнику, построенному на катетах a и b . Применительно ко всем другим треугольникам это равенство дает известную теорему косинусов.

Пусть $\mathbf{i}, \mathbf{j}, \mathbf{k}$ — три взаимно перпендикулярные вектора единичной длины. Направим оси координат OX, OY, OZ параллельно этим векторам и запишем очевидное тождество:

$$\mathbf{a} = a_x \mathbf{i} + a_y \mathbf{j} + a_z \mathbf{k},$$

где a_x, a_y и a_z — проекции вектора \mathbf{a} на оси координат. Из последнего равенства вытекает другое определение скалярного произведения:

$$(\mathbf{a}, \mathbf{b}) = a_x b_x + a_y b_y + a_z b_z,$$

не содержащее угла между векторами.

Рассмотренная здесь структура является одним из конкретных воплощений абстрактного понятия «евклидово пространство». Оно будет рассмотрено в § 16.1.

§ 5.3. Векторное произведение

Введем векторное произведение двух векторов. Оно обозначается символом $[\mathbf{a}, \mathbf{b}]$ и определяется как вектор, перпендикулярный каждому из векторов-сомножителей и направленный в ту сторону, откуда поворот от первого сомножителя ко второму на меньший угол происходит против часовой стрелки. Длина этого вектора определяется как площадь параллелограмма, построенного на векторах-сомножителях. В частности, справедливы равенства

$$[\mathbf{i}, \mathbf{j}] = \mathbf{k}, \quad [\mathbf{j}, \mathbf{k}] = \mathbf{i}, \quad [\mathbf{k}, \mathbf{i}] = \mathbf{j}.$$

Из определения следует, что векторное произведение антисимметрично относительно своих сомножителей, а «векторный квадрат» $[\mathbf{a}, \mathbf{a}]$ равен нулю:

$$[\mathbf{a}, \mathbf{b}] = -[\mathbf{b}, \mathbf{a}], \quad [\mathbf{a}, \mathbf{a}] = 0.$$

Основным свойством векторного произведения является его линейность по отношению к каждому из сомножителей. Не будем злоупотреблять вниманием читателя и примем на веру это свойство. Оно позволяет вычислить все три компоненты векторного произведения:

$$[\mathbf{a}, \mathbf{b}]_x = a_y b_z - a_z b_y, \quad [\mathbf{a}, \mathbf{b}]_y = a_z b_x - a_x b_z, \quad [\mathbf{a}, \mathbf{b}]_z = a_x b_y - a_y b_x.$$

Мы получили второе, эквивалентное определение векторного произведения.

Векторная алгебра содержит ряд полезных тождеств. Для наших целей они не очень существенны.

§ 5.4. Физические величины как векторы

Векторы, о которых мы говорили до сих пор, представляют собой геометрические векторы. Их свойствами являются свойства, которыми мы их наделили, и это было наше право. Рассмотрим какие-либо физические векторы (например, скорость или силу). Будут ли все они обладать свойствами, которыми мы наделили геометрические векторы? Для ответа на этот вопрос нужно обратиться к физике. Подчеркнем, что другого пути нет. К счастью, физика дает на него положительный ответ. Ничего удивительного — скажет читатель — математические векторы были введены по образцу физических.

Это верно, но сама возможность позаимствовать свойства у физических векторов предполагает, что подобные свойства у них одинаковы. Кстати, тот факт, что в классической механике скорости складываются по правилу параллелограмма, хотя и справедлив, но не является очевидным, поскольку в теории относительности это не так.

Обратимся к операции вращения системы координат вокруг заданной оси на заданный угол. Естественно, один и тот же геометрический вектор будет иметь разные координаты в старой и новой системах координат. Элементарный расчет приводит к формулам перехода от старых координат к новым. При выводе этих формул из всей физики нам нужна только евклидова геометрия как наука о физическом пространстве. Спрашивается, почему векторы совсем другой природы, такие как скорость или электрическое поле, должны подчиняться тем же формулам?

Как ни удивительно, на поставленный вопрос можно внятно ответить. Попробуем это сделать. С этой целью обратимся к трем системам координат — K_1 , K_2 и K_3 , получающимся одна из другой соответствующими вращениями. Рассмотрим координаты некоторого физического вектора в каждой из них. Переход от координат вектора в системе K_1 к системе K_3 можно совершить двумя разными способами: либо непосредственно, либо в два шага, сначала найдя координаты в системе K_2 , а затем перейдя из нее в систему K_3 . Мы пока не знаем формул перехода между системами координат, но убеждены в том, что они существуют. Более того, эти формулы должны обладать следующим свойством: *оба способа перехода от системы координат K_1 к системе K_3 всегда должны приводить к одному и тому же результату.*

Можно поставить перед собой следующую задачу: найти все мыслимые формулы перехода, обладающие указанным свойством. Это

чисто математическая задача. Ее решение показывает, что существует только два типа искоемых формул перехода. Первый из них тривиален: набор трех координат вовсе не меняется при переходе от одной системы координат к другой. Второй тип формул перехода такой же, как и для математических векторов. Последнее означает, что при всех вращениях любой физический вектор ведет себя точно так же, как и вектор-стрелка, то есть как математический вектор.

Сделаем вывод. *Вектор является не просто совокупностью трех произвольных чисел. Это математический объект, изображаемый в каждой системе своей тройкой чисел. Задание такой тройки в одной системе координат однозначно определяет соответствующие тройки во всех системах координат, получаемых вращением исходной системы. Фактический переход от одной системы координат к другой совершается по тем же формулам, что и для геометрических векторов.*

Наконец-то, мы получили полное определение вектора как физической величины. Впрочем, остался еще один шаг. Он связан с существованием систем прямоугольных координат, которые нельзя получить путем вращения исходной. Такова, например, система, у которой оси направлены в стороны, противоположные одноименным осям исходной системы. Для удобства назовем ее инверсной. При переходе к инверсной системе все три координаты вектора-стрелки умножаются на минус единицу. Означает ли это, что и все физические векторы ведут себя также? Нет, не означает!

Здесь нет логического противоречия. Мы расширили класс рассматриваемых систем координат, и если в старом, более узком классе все физические векторы вели себя как векторы-стрелки, то это не значит, что такое послушное поведение сохранится и в расширенном классе координатных систем. Анализ показывает, что для вектора существует только два варианта перехода к инверсной системе: либо все его координаты умножаются на минус единицу, либо не меняются вовсе. В первом случае мы будем по-прежнему называть его вектором, во втором присвоим ему название псевдовектора. Напряженность магнитного поля является псевдовектором, а электрического — вектором. Легко видеть, что векторное произведение двух векторов — псевдовектор. Последнее можно сказать и о векторном произведении двух псевдовекторов. Аналогично каждое число является либо скаляром, либо псевдоскаляром. Скаляр во всех системах координат имеет одно и то же значение. Псевдоскаляр умножается на минус единицу при переходе к инверсной системе координат и не изменяется при любом вращении системы координат. Скалярное произведение любого вектора на псевдовектор является псевдоскаляром.

По сути дела, мы пришли к выводу, что полное определение вектора требует рассмотрения структуры, включающей не только векторы, но

и всевозможные вращения прямоугольной системы координат, а также переход к инверсной системе координат.

Если речь идет о конкретном виде физического вектора, например о напряженности электрического поля, то физика *знает*, как он выглядит в разных системах координат. Это знание получено экспериментальным путем и может быть подтверждено анализом системы уравнений электромагнитного поля. Последняя, как и всякая другая система физических законов, должна иметь одинаковый вид во всех *неподвижных* друг относительно друга системах координат. Не нужно думать, что этого можно достичь, просто потребовав, чтобы проекции напряженностей электрического и магнитного полей не менялись при переходе от одной системы к какой-либо другой, повернутой относительно нее системе. Ничего не получится! Дело в том, что содержащиеся в уравнениях Максвелла производные по координатам от этих полей преобразуются друг через друга по своим правилам. Чтобы компенсировать происходящие при этом изменения в системе уравнений, электромагнитное поле тоже должно преобразовываться. Можно проверить, что требуемые преобразования как раз такие, какие испытывают векторы и псевдовекторы при указанных преобразованиях систем координат.

Мы подошли к черте, за которой угадываются теория четырехмерных векторов специальной теории относительности и инвариантность уравнений Максвелла относительно перехода от одной инерциальной системы координат к другой, движущейся относительно нее системе.

Об этом и многом другом читатель узнает, ознакомившись с четвертой главой второй части книги.

Кто бы мог подумать, что введение числа i — корня квадратного из минус единицы — приведет к таким последствиям? Впрочем, известные ожидания были бы естественны. В самом деле, если появление дробей сделало возможным деление всякого числа на любое другое (кроме нуля) число и обнаружило необходимость введения иррациональных чисел, а появление последних открыло дорогу понятиям предела, производной и интеграла, то почему бы в таком случае и числу i не внести свою лепту в развитие математики?

Первое, что бросается в глаза после введения комплексных чисел, это возможность разложить на линейные множители любой полином второй степени и найти решения любого квадратного уравнения. Однако оказалось, что это же можно проделать с полиномом и уравнением произвольной степени! Гораздо более неожиданным явилось то, что все элементарные свойства тригонометрических функций можно теперь получить чисто алгебраическим путем, если понять, как изменяются координаты вектора, лежащего в плоскости XOY , при повороте на 90 градусов. Впрочем, это только методический прием. Самое главное и самое поразительное применение комплексных чисел — выделение нового класса функций, получивших название аналитических. Этот класс обладает двумя, казалось бы, взаимоисключающими особенностями: его теория изобилует нетривиальными полезными фактами и в то же время охватывает почти все встречающиеся на практике функции.

Для многих математиков, включая и автора, есть нечто загадочное в комплексных числах, и вовсе не потому, что они содержат «мнимую единицу». Нет, дело в другом. Как получилось, что комплексные числа проникли во все уголки и закоулки математики?

Комплексные числа помогли создать фантастически эффективные методы исследования и решения математических задач. Разумеется, это не осталось незамеченным ученым миром инженеров, математиков и физиков. Комплексные числа уже давно «приняты ими на вооружение».

Что же такое комплексные числа?

§ 6.1. Появление комплексных чисел

Все начинается с введения числа i — «корня квадратного из минус единицы». Беда заключается в том, что такого числа не существует.

Впрочем, иррациональных чисел тоже не существовало, пока их не ввели. Введение всякого нового числа представляет собой целую церемонию, напоминающую включение в семейный клан нового члена. И в том и в другом случае необходимо распространить на новичка систему внутриклановых отношений. Если речь идет о числах, то необходимо прежде всего определить умножение нового числа на все старые. Это приводит к появлению бесконечного множества новых чисел. Затем вводятся в рассмотрение всевозможные суммы новых и старых чисел. Наконец, определяется умножение указанных сумм друг на друга, и притом так, чтобы при этом не появились какие-либо «сверхновые» числа. Продемонстрируем подобный процесс на простом примере.

Пример. Рассмотрим число $\sqrt{3}$. Присоединим его к совокупности всех рациональных чисел. Определяющим свойством этого числа является равенство $\sqrt{3}\sqrt{3} = 3$. Сначала введем произведения числа $\sqrt{3}$ на всевозможные рациональные числа: $x\sqrt{3} = \sqrt{3}x$, а также суммы $x\sqrt{3} + y$ (где x и y — рациональные числа). Определим естественным образом сложение новых чисел:

$$u\sqrt{3} + v + (x\sqrt{3} + y) = (u + x)\sqrt{3} + (v + y).$$

Теперь остается задать произведения таких сумм. Если мы хотим сохранить в силе дистрибутивный и коммутативный законы арифметических действий, то существует только одна возможность это сделать:

$$(u\sqrt{3} + v)(x\sqrt{3} + y) = (3ux + vy) + (uy + vx)\sqrt{3}.$$

Процедура присоединения числа $\sqrt{3}$ к множеству всех рациональных чисел завершена.

§ 6.2. Комплексные числа

Совершенно аналогично происходит расширение совокупности действительных чисел с помощью символа i . Сначала вводим произведения $bi = ib$ всевозможных действительных чисел b на число i . Затем вводим в рассмотрение всевозможные суммы: $a + bi = bi + a$. Совокупность всех чисел такого вида образует множество всех комплексных чисел. Число a называется *вещественной частью* числа $a + bi$, число bi — его *мнимой частью*; числа $a + bi$ и $a - bi$ носят название *комплексно сопряженных*.

Сложение комплексных чисел производится путем сложения их одноименных частей. Умножение определяется однозначно, если постулировать дистрибутивный и коммутативный законы и учесть, что $i^2 = -1$:

$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i.$$

Итак, мы получили новую систему чисел с действиями сложения и умножения, которые не приводят к появлению еще более новых чисел. Эти два действия обладают всеми свойствами обычных сложения и умножения. Применительно к вещественным числам они совпадают с обычными сложением и умножением.

Читатель может спросить: «На каком основании вы определили сложение и умножение именно так, а не как-либо иначе?» Вот откровенный ответ. Оказывается, что введенные подобным образом сложение и умножение превращают комплексные числа в очень полезное орудие в руках исследователя.

§ 6.3. Три «лика» комплексных чисел

Можно представлять себе комплексные числа и по-другому — как точки на координатной плоскости. Условимся изображать число $a + bi$ точкой с координатами (a, b) (рис. 6.1). Саму плоскость в этом случае

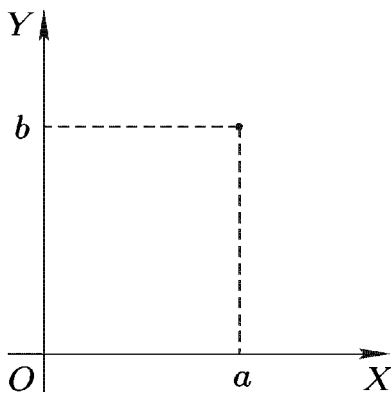


Рис. 6.1. Комплексное число $a + ib$ как точка на комплексной плоскости

принято называть комплексной. Кроме того, комплексные числа часто представляют с помощью векторов; вектор, проекции которого равны a и b , изображает комплексное число $a + bi$.

Итак, у нас есть три образа («три лика») комплексных чисел. Все они — число, вектор и точка — взаимозаменяемы. Мы говорим «точка», а подразумеваем «число» и т. д. Зачем это нужно? Опишем возникшую ситуацию. По сути дела, мы имеем три разные, хотя и эквивалентные, модели, или формы, комплексных чисел. Назовем их алгебраической, векторной и точечной формами. Подобное положение вещей, когда один и тот же математический объект, а точнее, математическая структура, выступает в нескольких очень разных «ликах», встречается в математике довольно часто. И каждый раз, когда это случается, исследователь выбирает тот «лик», в котором рассматриваемая проблема выглядит всего проще. Иногда бывает полезно несколько раз переходить от одного «лика» к другому.

§ 6.4. Пример: аналитическая геометрия

Классическим примером может служить аналитическая геометрия, суть которой состоит в возможности рассматривать соотношения между двумя переменными как кривые на плоскости.

Скажем несколько поясняющих слов. Мы знаем, что положение каждой точки на плоскости можно задать ее прямоугольными координатами. Рассмотрим какое-нибудь соотношение, связывающее эти

две величины, например $x^2 + xy + y^3 = 1$. Сопоставим ему кривую, состоящую из тех точек, координаты которых удовлетворяют этому соотношению. Данное правило однозначно определяет кривую: о любой точке плоскости мы можем сказать, лежит ли она на этой кривой. Точка $(x = 1, y = 1)$ на кривой не лежит, поскольку для нее значение левой части уравнения равно трем, а не единице. А вот точка с координатами $(1, 0)$ принадлежит кривой. Нетрудно ответить и на более сложный вопрос: в скольких точках рассматриваемая кривая пересекается прямой, параллельной оси иксов? Уравнение такой линии имеет вид $y = y_0$. Подставляя это значение y в наше соотношение, превратим его в квадратное уравнение относительно x . Как известно, квадратное уравнение имеет либо два вещественные решения, либо одно, либо ни одного. Значит, число точек пересечения равно двум, единице или нулю, и легко выяснить, при каких значениях y_0 реализуется каждая из перечисленных трех возможностей.

Рассмотрим противоположный в интересующем нас смысле пример, когда переход к кривым облегчает исследование чисто алгебраической задачи.

Зададимся вопросом о том, сколько различных решений имеет система уравнений:

$$\begin{aligned}(x - 1)^2 + (y - 5)^2 &= u^2, \\ (x - 5)^2 + (y - 3)^2 &= v^2.\end{aligned}$$

Будем рассуждать следующим образом. Согласно теореме Пифагора каждое из этих уравнений соответствует окружности (рис. 6.2). Центр первой окружности находится в точке $(1, 5)$, центр второй — в точке $(5, 3)$. Их радиусы равны u и v .

Если пара x, y является решением нашей системы, то точка (x, y) должна лежать на каждой из этих окружностей, то есть быть точкой их пересечения! И наоборот, координаты каждой точки пересечения дают решение системы. Пусть для определенности оба радиуса меньше расстояния между центрами. Тогда ясно, что две окружности не пересекаются, если сумма их радиусов также меньше этого расстояния; в противоположном случае они пересекаются в двух точках; при равенстве этих двух величин окружности касаются друг друга. Итак, все зависит от знака величины

$$[(1 - 5)^2 + (5 - 3)^2]^{1/2} - (u + v).$$

Наше рассуждение предполагало, что каждый из радиусов меньше расстояния между центрами. Если мы имеем дело с противоположным

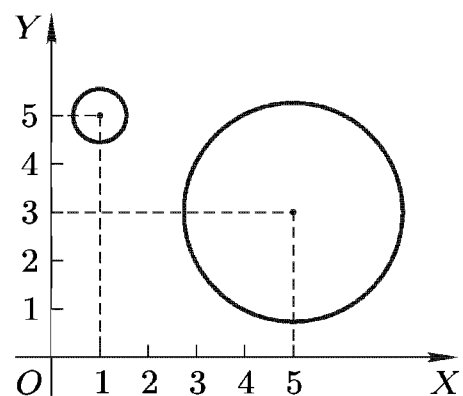


Рис. 6.2. Иллюстрация к задаче о двух окружностях

случае, то в приведенном критерии следует заменить сумму радиусов на модуль их разности. Разумеется, в этом параграфе x и y — действительные числа.

Мы попытались показать, как можно использовать два разных «лика» (два представления) одного и того же объекта.

§ 6.5. Три лика комплексных чисел (продолжение)

Вернемся к трем формам комплексных чисел. Установим между ними простейшие связи. Длина вектора, отвечающего числу $a + bi$, то есть число $\sqrt{a^2 + b^2}$, и угол φ между этим вектором и осью OX называются модулем числа $a + bi$ и его аргументом соответственно. Их принято обозначать следующим образом: $|z|$ и $\arg z$ (рис. 6.3). Легко видеть, что

$$\cos \varphi = \frac{a}{|z|}, \quad \sin \varphi = \frac{b}{|z|}.$$

Единица изображается вектором единичной длины, параллельным оси OX , мнимая единица — число i — представляет собой единичный вектор, параллельный оси OY (рис. 6.4)

Из представления в векторной форме следует, что всякое комплексное число z может быть записано в виде

$$z = |z|(\cos \varphi + i \sin \varphi).$$

Как выглядят сложение и умножение комплексных чисел в их различных формах? Сложение чисел в алгебраической форме сводится к нахождению суммы их вещественных частей и суммы их мнимых частей. Сложение этих же чисел в векторной форме выполняется по известному правилу параллелограмма (рис. 6.5). Наконец, если числа заданы в виде точек комплексной плоскости, то складываются их одноименные координаты. Пожалуй, сложение в векторной форме наиболее наглядно. Зато две другие формы удобнее для численных расчетов.

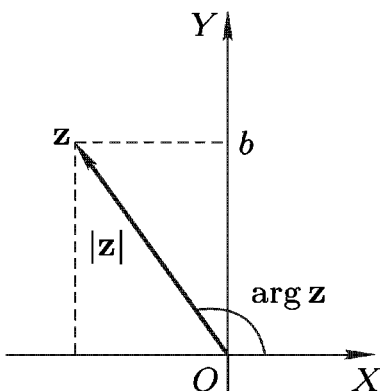


Рис. 6.3. Модуль и аргумент комплексного числа

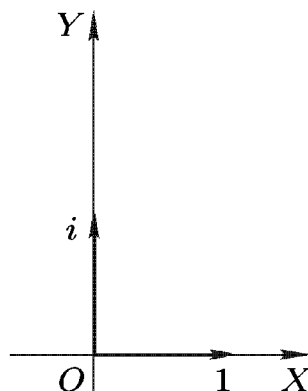


Рис. 6.4. Вещественная и мнимая единицы на комплексной плоскости

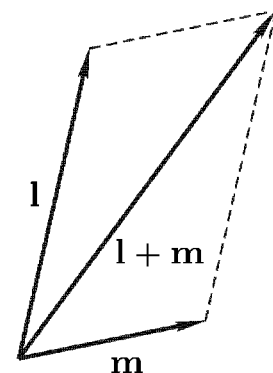


Рис. 6.5. Сложение векторов по правилу параллелограмма

Перейдем к умножению комплексных чисел. Это действие очень наглядно изображается при использовании векторной формы. Как именно? Для того чтобы ответить на этот вопрос, рассмотрим предварительно несколько простых примеров. Возьмем произвольное комплексное число и умножим его на i . Получим

$$(a + bi)i = -b + ai.$$

«Ну и что!» — подумает нетерпеливый читатель. Перейдем от чисел к векторам. При повороте на 90° против часовой стрелки вектор, отвечающий числу $a + bi$, перейдет в вектор, отвечающий числу $-b + ai$, так как единица перейдет в i , а i — в минус единицу. Значит, умножение на i в векторной форме равносильно повороту на прямой угол (рис. 6.6)! Рассмотрим, что происходит с векторами, соответствующими единице и i , при умножении их на произвольный вектор z . Единичный вектор переходит в вектор z , то есть поворачивается на угол, равный аргументу z , и удлиняется в $|z|$ раз, где $|z|$ — длина вектора z .

Обратимся теперь к вектору iz . Как мы уже видели, он получается из вектора z путем поворота на угол $\pi/2$. Следовательно, аргумент вектора $z i$ больше аргумента вектора z на $\pi/2$, то есть равен $\varphi + \pi/2$. Это означает, что вектор $z i$ получается из вектора i поворотом на тот же угол φ , разумеется, с последующим удлинением в $|z|$ раз. Таким образом, и единичный вектор, и вектор i ведут себя одинаково при умножении на z : оба они поворачиваются на один и тот же угол, равный аргументу вектора z , а их длины умножаются на длину последнего (рис. 6.7).

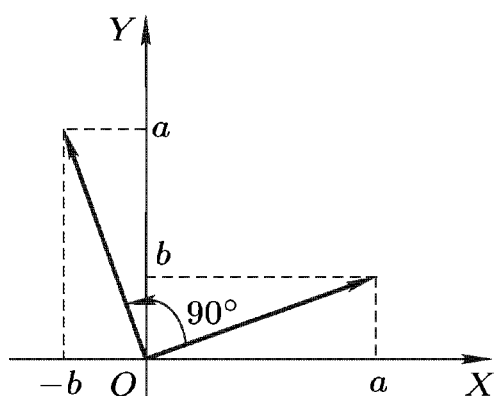


Рис. 6.6. Умножение числа z на i

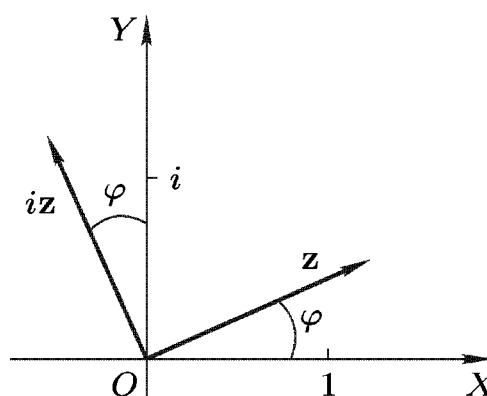


Рис. 6.7. Умножение 1 и i на z

Отсюда, пользуясь дистрибутивным законом, легко заключить, что то же происходит с любым вектором w при умножении на вектор z . Это можно записать следующим образом:

$$\arg(zw) = \arg z + \arg w, \quad |zw| = |z||w|,$$

где $\arg z$ — аргумент z . Итак, при умножении комплексных чисел их аргументы складываются, а модули перемножаются. Мы получили

очень наглядную картину для произведения векторов. Одновременно мы вывели удобную формулу:

$$|z|(\cos \varphi + i \sin \varphi) \cdot |w|(\cos \psi + i \sin \psi) = |z||w|[\cos(\varphi + \psi) + i \sin(\varphi + \psi)].$$

Назовем ее векторным правилом умножения комплексных чисел. Существует два ряда очень интересных следствий, вытекающих из сказанного.

Начнем со следствия, наводящего на некоторые философские размышления. Рассмотрим умножение некоторого фиксированного комплексного числа a на комплексное число b , которое пробегает всю комплексную плоскость. При этом вектор c , отвечающий произведению ab , тоже пробегает всю комплексную плоскость и получается из вектора b поворотом на некоторый угол (равный аргументу числа a) и умножением его длины на модуль числа a . Мы видим, что умножение на комплексное число a можно рассматривать как преобразование плоскости. Итак, *каждое комплексное число можно рассматривать как преобразование плоскости.*

Таким образом, мы имеем четвертый по счету «лик» комплексного числа. При этом положительное число представляет собой умножение длин векторов с сохранением их направлений, а число, модуль которого равен единице, — поворот всех векторов с сохранением их длин. Число i является поворотом плоскости на 90 градусов.

Остановимся на минуту! Мы получили «прозаический» образ числа i , который не бросает вызов нашему воображению, не требует «вообразить невообразимое». Число i — это поворот плоскости как целого на прямой угол против часовой стрелки. Произведенный дважды, он поворачивает каждый вектор на 180 градусов, то есть умножает его на минус единицу, подтверждая тем самым правило $i^2 = -1$. Означает ли это, что в дальнейшем мы должны мыслить комплексные числа как элементы некоторого класса преобразований плоскости? Думаю, не означает. В конце концов, к числам мы привыкли, и нам легче иметь дело с мнимой единицей, чем представлять себе все числа преобразованиями плоскости.

Подойдем к этому вопросу с другой стороны. Древним скотоводам было не под силу представить себе три четверти овцы. Античные математики отказались иметь дело с иррациональными числами (кстати, и название у этих чисел отпугивающее). А мы? Думаю, математики давно не видят ничего особенного в мнимой единице.

§ 6.6. Комплексная плоскость

До сих пор мы ничего не сказали о достоинствах точечной формы комплексных чисел. Она удобна при рассмотрении функций. Функциям посвящена следующая глава, однако настоящий параграф можно читать независимо.

У каждой функции $w = f(z)$ есть область определения и область значений. Первая является совокупностью всех тех значений z , для которых можно подсчитать соответствующее значение w , вторая — совокупностью значений, принимаемых w , когда z пробегает всю область определения функции $f(z)$.

Область определения функции естественно представлять себе в виде некоторой части комплексной плоскости. Это же относится и к области значений функции. Последнюю удобно представлять себе расположенной на другом экземпляре комплексной плоскости. Сама функция переводит каждую точку из области определения в соответствующую ей точку в области значений. При этом каждая кривая внутри области определения переходит в некоторую кривую внутри области значений (рис. 6.8).

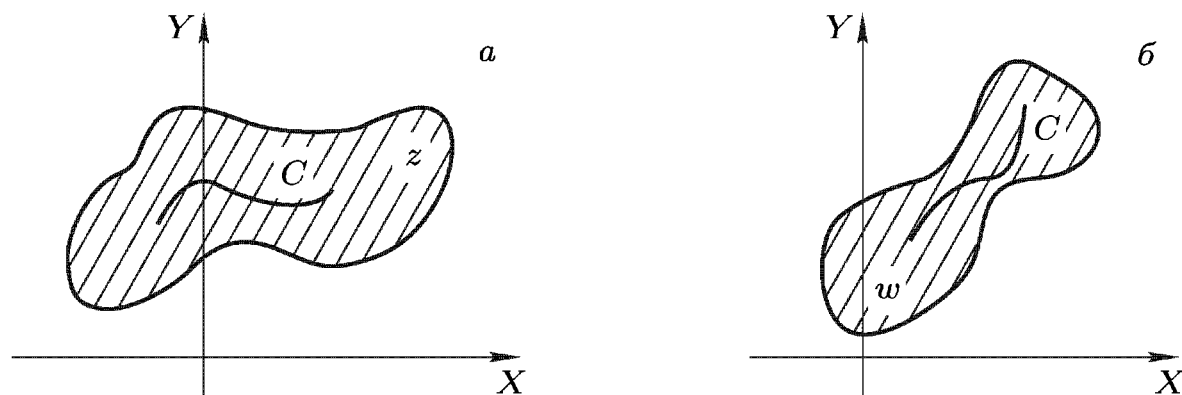


Рис. 6.8. Область определения (а) и область значений (б) функции $w = f(z)$; кривая C (а) и ее образ (б)

Поясним сказанное на примере. Рассмотрим функцию

$$w = 1/z.$$

Ее область определения — вся плоскость z без точки $z = 0$. Область значений — вся плоскость w без точки $w = 0$. Вспоминая геометрический смысл умножения, можно записать

$$|w| = 1/|z|, \quad \arg w = -\arg z.$$

Эти равенства показывают, что совокупность концентрических окружностей плоскости z переходит в систему таких же окружностей плоскости w , а каждый луч, исходящий из точки $z = 0$, в луч, исходящий из точки $w = 0$ (рис. 6.9). Заметим еще, что когда точка z движется по окружности по часовой стрелке, точка w движется против часовой стрелки. Когда точка z движется по лучу, удаляясь от его начала, точка w движется по своему лучу, приближаясь к его началу. Таким образом, мы получаем довольно наглядную геометрическую картину рассматриваемой функции и легко можем представить себе, как движется точка w , если известно движение ее прообраза — точки z .

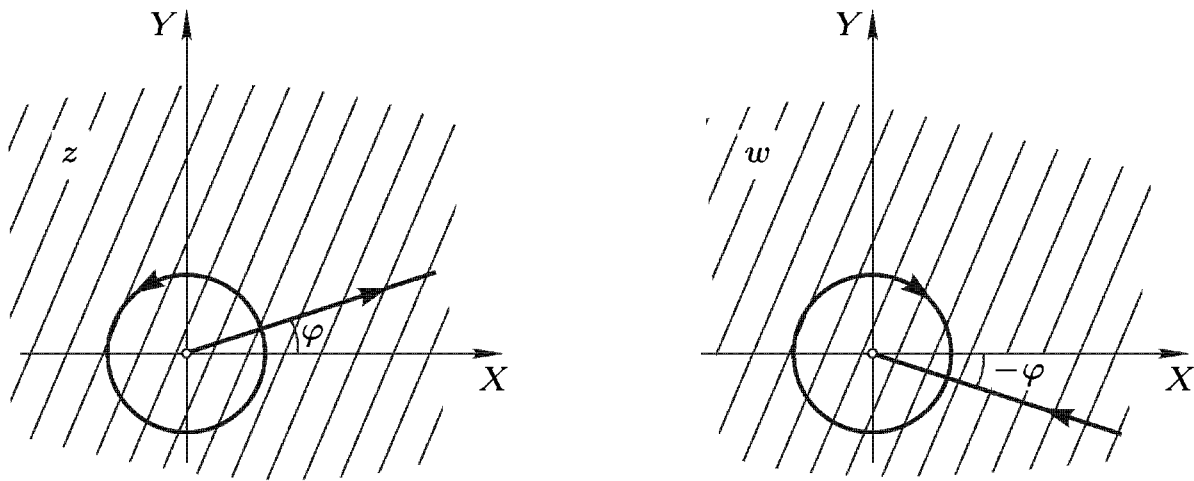


Рис. 6.9. Отображение функцией $w = 1/z$ окружности и луча из области определения в область значений

§ 6.7. Формула Муавра

Вернемся к векторному правилу умножения комплексных чисел. Обычно его записывают, предварительно сократив на произведение модулей. Прежде всего заметим, что определение произведения в алгебраической форме дает «бесплатно» два тригонометрические тождества:

$$\cos(\varphi + \psi) = \cos \varphi \cos \psi - \sin \varphi \sin \psi,$$

$$\sin(\varphi + \psi) = \sin \varphi \cos \psi + \cos \varphi \sin \psi.$$

Столь же просто получается известная из школьного курса формула Абрахама Муавра (1667–1754):

$$(\cos \varphi + i \sin \varphi)^n = \cos(n\varphi) + i \sin(n\varphi).$$

Формула Муавра записывается гораздо проще, если воспользоваться удачным обозначением, которое было введено Эйлером:

$$\cos \varphi + i \sin \varphi = e^{i\varphi}.$$

Она принимает легко запоминающийся вид:

$$(e^{i\varphi})^n = e^{in\varphi}.$$

Заметим также, что в соответствии с определением

$$e^{i\varphi} e^{i\psi} = e^{i(\varphi+\psi)}.$$

Мы видим, что с символом Эйлера можно обращаться как со степенью. На этой стадии бессмысленно спрашивать, чему равно e . Пока символ Эйлера представляет собой только удобное обозначение. Позже, когда будет введена функция e^z , окажется возможным отождествить символ Эйлера со значением этой функции при подстановке $z = i\varphi$ и вычислить число Эйлера e , принятое, кстати, в качестве основания

натуральных логарифмов. Функции $\cos \varphi$ и $\sin \varphi$ можно рассматривать как вещественную и мнимую части функции $e^{i\varphi}$:

$$\cos \varphi = \operatorname{Re} e^{i\varphi}, \quad \sin \varphi = \operatorname{Im} e^{i\varphi}.$$

Любое не равное нулю комплексное число может быть представлено в «Эйлеровой форме»:

$$a + bi = Ae^{i\alpha},$$

где

$$A = (a^2 + b^2)^{1/2}, \quad \cos \alpha = a/A, \quad \sin \alpha = b/A.$$

Формула Муавра позволяет с помощью одной только алгебры выразить косинусы и синусы кратных углов через косинус и синус исходного угла. Так, например, вспоминая бином Ньютона, можно написать

$$\begin{aligned} \cos 4a + i \sin 4a &= (\cos a + i \sin a)^4 = \\ &= \cos^4 a + 4 \cos^3 a (i \sin a) + 6 \cos^2 a (i \sin a)^2 + 4 \cos a (i \sin a)^3 + (i \sin a)^4. \end{aligned}$$

Отсюда

$$\begin{aligned} \cos 4a + i \sin 4a &= \\ &= \cos^4 a - 6 \cos^2 a \sin^2 a + \sin^4 a + i4(\cos^3 a \sin a - \cos a \sin^3 a). \end{aligned}$$

Приравнивая в последнем равенстве вещественные и мнимые части, находим известные выражения для косинуса и синуса четырехкратного угла. При наличии таблиц синусов и косинусов формула Муавра позволяет легко извлекать корни любой степени из комплексных чисел. В самом деле, если положить в ней $\varphi = \psi/n$, то получится

$$\cos \psi + i \sin \psi = [\cos(\psi/n) + i \sin(\psi/n)]^n.$$

Это равенство показывает, что число

$$\cos(\psi/n) + i \sin(\psi/n)$$

является корнем степени n из комплексного числа $\cos \psi + i \sin \psi$.

Сколько разных корней есть у числа z , если оно не равно нулю? Заметим, что левая часть последнего равенства не изменится, если заменить ψ на $\psi + 2\pi k$, где k — любое целое число. Придавая ему значения $k = 0, 1, 2, \dots, n-1$, получим n различных корней из числа z :

$$Z_k = \cos[(\alpha + 2k\pi)/n] + i \sin[(\alpha + 2k\pi)/n].$$

Значения $k = n, n+1, n+2, \dots$ не дают ничего нового вследствие периодичности функций косинус и синус. Полагая в этих формулах $\psi = 0$, получим n корней из единицы. Изображающие их точки комплексной плоскости лежат на единичной окружности в вершинах вписанного правильного n -угольника. Заметим, что каждый корень n -й

степени из единицы является в то же время корнем уравнения $z^n = 1$. Поэтому неудивительно, что их ровно n штук.

В заключение получим формулы разложения на множители суммы косинусов:

$$\begin{aligned}\cos \alpha + \cos \beta &= \operatorname{Re}(e^{i\alpha} + e^{i\beta}) = \operatorname{Re}[e^{i(\alpha+\beta)/2}(e^{i(\alpha-\beta)/2} + e^{-i(\alpha-\beta)/2})] = \\ &= \operatorname{Re}[e^{i(\alpha+\beta)/2}] \cdot 2 \cos[(\alpha - \beta)/2] = 2 \cos[(\alpha + \beta)/2] \cos[(\alpha - \beta)/2].\end{aligned}$$

Отметим еще одно важное свойство комплексных чисел. Очевидное тождество

$$(Ae^{i\varphi})(A^{-1}e^{-i\varphi}) = 1$$

означает, что у каждого комплексного числа, если оно не равно нулю, есть обратное число.

§ 6.8. Метод комплексификации

Приведенные примеры показывают, насколько полезна может быть «двуликость» умножения. Оказывается, она используется и в применении в физике методе, который мы позволим себе назвать методом комплексификации. В чем состоит этот метод? Во многих физических задачах возникает ситуация типа «причина–следствие», или «вход–выход». Поясним, о чем идет речь, на двух примерах.

Рассмотрим произвольный радиотехнический контур, состоящий из нескольких омических сопротивлений, емкостей и индуктивных элементов. Под влиянием переменной разности потенциалов $V(t)$, приложенной к двум точкам контура, скажем, точкам A и B , между этими точками начинает течь ток $I(t)$. Требуется определить связь между приложенной разностью потенциалов и током. Здесь приложенная разность потенциалов является причиной (или входным сигналом), а ток — следствием (или выходным сигналом). В качестве второго примера можно рассмотреть магнитное поле $H(r, t)$, создаваемое током, описанным в первом примере. Здесь роли сигналов играют электрический ток в контуре и создаваемое им магнитное поле.

Кроме того, можно было бы рассмотреть колебания гармонического осциллятора под действием вынуждающей силы, а также множество других примеров типа вход–выход. Если обратиться к физике, то оказывается, что во многих случаях связь между входным и выходными сигналами выглядит наиболее просто, когда входной сигнал является гармоническим, то есть зависит от времени следующим образом:

$$V(t) = V_0 \cos(\omega t + \alpha)$$

(для определенности возьмем обозначения из первого примера). При этом и ток зависит от времени гармонически:

$$I(t) = I_0 \cos(\omega t + \beta).$$

Далее выясняется, что отношение амплитуд $I_0/V_0 = \sigma$ и разность фаз $\beta - \alpha = \Delta$ не зависят от амплитуды V_0 и фазы α приложенной разности потенциалов. Обе эти величины являются физическими характеристиками контура. С их помощью можно записать закон Ома (обобщенный):

$$I_0 = V_0 \sigma, \quad \beta = \alpha + \Delta;$$

он отличается от привычного закона Ома «предписанным» сдвигом фаз. Последний связан с присутствием в контуре емкостей и индуктивностей.

Можно ли и в данном случае придать закону Ома обычную форму? Оказывается, можно! Только для этого необходимо воспользоваться комплексными числами. Делается это так. Прежде всего введем комплексные разность потенциалов $V_{\text{КОМ}}$ и силу тока $I_{\text{КОМ}}$:

$$U_{\text{КОМ}} = U_0 e^{i(\omega t + \alpha)}, \quad I_{\text{КОМ}} = I_0 e^{i(\omega t + \beta)}.$$

Ясно, что их реальные части совпадают с одноименными вещественными величинами. С другой стороны, их отношение равно

$$I_{\text{КОМ}}/U_{\text{КОМ}} = \sigma e^{i\Delta}$$

и, следовательно, не зависит от времени. Оно играет роль комплексной проводимости:

$$\sigma_{\text{КОМ}} = \sigma e^{i\Delta}$$

и позволяет записать обобщенный закон Ома в привычной форме:

$$I_{\text{КОМ}} = \sigma_{\text{КОМ}} V_{\text{КОМ}},$$

что и было нашей целью.

Обратная величина $R_{\text{КОМ}} = \sigma_{\text{КОМ}}^{-1}$ называется комплексным сопротивлением. Как воспользоваться такой формой закона Ома? Пусть, например, задана переменная разность потенциалов $V(t)$. Это означает, что известны амплитуда V_0 и фаза α . Тем самым известна и комплексная разность потенциалов. Закон Ома в комплексной форме позволяет определить комплексную силу тока путем перемножения двух комплексных чисел! Это в свою очередь определяет амплитуду и фазу реальной силы тока. Таким образом, закон Ома в комплексной форме содержит полную информацию о реальных физических величинах, разности потенциалов и силе тока. Этот пример показывает, что каждая из величин, фигурирующих в комплексной форме закона Ома, представляет собой конкретную реализацию комплексных чисел.

Отметим еще одно обстоятельство. Если у нас есть цепочка параллельно соединенных n контуров с комплексными сопротивлениями R_1, R_2, \dots, R_n , то ее результирующая комплексная проводимость K дается соотношением

$$K = 1/R = 1/R_1 + 1/R_2 + \dots + 1/R_n.$$

Таким образом, сопротивления не только являются комплексными числами, но и подчиняются арифметике комплексных чисел: вычисление комплексного сопротивления R производится с помощью ранее определенных арифметических действий.

Использованный способ записи физического закона путем комплексификации физических величин и параметров, характеризующих физический прибор, применим при рассмотрении многих гармонических процессов.

§ 6.9. Алгебраические уравнения

Перейдем к другим приложениям комплексных чисел. Рассмотрим уравнение

$$x^2 + 1 = 0.$$

Оно не имеет ни одного вещественного решения, но имеет два комплексные: $x = i$ и $x = -i$. Существует множество других алгебраических уравнений, которые не имеют действительных решений. Возникает вопрос, имеют ли они решения в комплексной области? Ответ на него дает следующая теорема, поразительная по простоте формулировки и трудности доказательства.

Основная теорема алгебры. Всякое уравнение

$$x^n + a_1x^{n-1} + \dots + a_{n-1}x + a_n = 0$$

имеет по крайней мере одно (вещественное или комплексное) решение.

Это удивительная теорема. В самом деле, почему присоединения к клану вещественных чисел только одного числа — корня простого уравнения $x^2 + 1 = 0$ — достаточно для того, чтобы все мыслимые алгебраические уравнения имели по крайней мере один корень? Заметим еще, что теорема остается верной и в случае, когда коэффициенты уравнения — произвольные комплексные числа. В дальнейшем мы очень просто докажем эту теорему, используя один из результатов теории аналитических функций (докажем, но не объясним, то есть не сделаем ее очевидной).

Почему важна эта теорема? Не только потому, что она вселяет надежду в душу ищущего решение. Она показывает, что все полиномы имеют общую и притом простую структуру: каждый из них представим в виде произведения линейных сомножителей:

$$P_n(x) = a(x - x_1)(x - x_2) \dots (x - x_n),$$

где x_1, x_2, \dots, x_n — нули полинома (не обязательно различные). В самом деле, вспоминая деление с остатком полинома на полином, можно представить $P_n(x)$ в виде

$$P_n(x) = (x - X)Q(x) + R,$$

где X — произвольное число; $Q(x)$ — некоторый полином степени $n - 1$ (частное); R — число (остаток), которое, естественно, зависит от выбора числа X . Если в качестве X взять один из нулей x_1 полинома $P_n(x)$ (такой нуль существует согласно теореме), то остаток будет равен нулю и последнее равенство примет вид

$$P_n(x) = (x - x_1)Q(x).$$

Применяя это рассуждение к полиному $Q(x)$, находим

$$Q(x) = (x - x_2)Q_1(x), \quad P_n(x) = (x - x_1)(x - x_2)Q_1(x).$$

Проделав n таких шагов, получим требуемое представление полинома $P_n(x)$.

В области действительных чисел такое представление возможно только в случае, когда все нули полинома — вещественные числа. Подчеркнем, что в рассмотренном фрагменте теории полиномов дважды используются комплексные числа: без них нельзя сформулировать приведенные утверждения, а без теории аналитических функций последние невозможно (точнее, трудно) доказать.

Покажем на примере, с какой легкостью из фактов теории комплексных чисел можно получать нетривиальные следствия в области чисел действительных. Пусть x_1, x_2, \dots, x_n — полный набор корней некоторого алгебраического уравнения с действительными коэффициентами. Легко проверить, что все невещественные корни этого уравнения попарно комплексно сопряжены. Пусть $u + iv$ и $u - iv$ — два такие корня. Полином, стоящий в левой части уравнения, делится на $x - (u + iv)$ и на $x - (u - iv)$, а значит, и на их произведение

$$[x - (u - iv)][x - (u + iv)] = (x - u)^2 + v^2.$$

Итак, каждый вещественный полином выше первой степени делится без остатка на некоторый вещественный полином второй степени. Если для читателя мнимая единица до сих пор представлялась лишь забавной абстракцией, то рассмотренные примеры должны пошатнуть это убеждение. Они показывают, какую незаменимую помощь могут оказать абстрактные теории при изучении конкретных объектов. В этой связи хочется привести знаменитую фразу выдающегося французского математика Жака Адамара (1865–1963): «Кратчайший путь между двумя истинами в вещественной области проходит через комплексную область».

Зачем нужны функции? Функции можно дифференцировать. Никакой другой математический объект не поддается дифференцированию. То же можно сказать и об интегрировании. Значит, без функций не было бы ни производных, ни интегралов, не было бы ни дифференциальных, ни интегральных уравнений. Что еще? Вся физика, как и все естествознание, определяет зависимость одних величин от других, то есть пользуется функциями. Без функций не обойтись!

§ 7.1. Эволюция понятия функции

С течением времени понятие «функция» меняло свой смысл. Точнее, совокупность объектов, описываемых этим термином, становилась все шире и шире. Сначала это были полиномы от одного или нескольких переменных, например

$$y = x^3 - 3x + 1.$$

Затем появились дробные степени, показательные функции, логарифмы и тригонометрические функции.

В отличие от полинома, который является одновременно и обозначением функции и одним из алгоритмов ее вычисления, обозначения функций второго эшелона не содержат ни их определения, ни способа их вычисления, то есть алгоритма. Например, $\sin x$ — это только обозначение. По определению синуса это отношение катета прямоугольного треугольника, лежащего против угла x , к гипотенузе. Одним из алгоритмов служит таблица синусов. То же можно сказать и о других функциях второго эшелона.

На этой стадии произошло отделение понятия «функция» от понятия «алгоритм», поскольку одну и ту же функцию можно вычислять с помощью различных алгоритмов. Так, для вычисления синуса можно вместо таблиц пользоваться бесконечным степенным рядом (рядом Тейлора). По сути, каждое тождество предлагает на выбор два алгоритма вычисления одной и той же функции. Например, тождество

$$1 + x + x^2 + \dots + x^n = (1 - x^{n+1}) / (1 - x)$$

дает два различных алгоритма для вычисления суммы геометрической прогрессии. Итак, функция не является алгоритмом. Что же такое функция? Обратимся к интуиции и спросим себя, в каком случае мы скажем, что два разные алгоритма определяют одну и ту же

функцию? Ответ очевиден. Если эти алгоритмы выдают совпадающие значения y , какое бы значение x мы ни взяли, то они определяют одну и ту же функцию $y(x)$. Теперь можно дать определение функции, не определяя, что такое алгоритм.

Определение. Пусть задана некоторая совокупность значений величины x . Обозначим ее буквой X . Пусть каждому x из этой совокупности сопоставлено (отнесено, отвечает, соответствует) некоторое число y . Соответствие « x переходит в y » и называют функцией. Совокупность X называют областью определения функции, совокупность соответствующих значений y — областью ее значений.

Теперь можно «оторваться» от чисел x и y . Пусть X — совокупность элементов любой природы, например совокупность всех полиномов $P(x)$. Если каждому элементу из X сопоставлен некоторый элемент y из некоторой, вообще говоря, другой совокупности Y , то и такое соответствие называют функцией. При этом элемент y называется значением функции на элементе x .

Это очень широкое обобщение первоначального понятия функции. Возможно, максимально широкое. Приведем простой пример — векторное произведение двух векторов. Эта функция определена на совокупности всевозможных пар векторов, а ее область значений — совокупность всех векторов.

Мы получили абстрактное определение функции. Оно не перечисляет все мыслимые функции; вместо этого оно фиксирует определяющее свойство этого понятия, а именно соответствие между элементами двух множеств. Заметим еще, что в тех случаях, когда значением функции служит число, а область ее определения не совпадает с некоторой совокупностью чисел, то функцию принято называть функционалом. Например, если каждому многоугольнику сопоставлена его площадь, то мы имеем дело с функционалом, область определения которого есть совокупность всех плоских замкнутых многоугольников. Часто употребляемыми функционалами являются среднее, минимальное и максимальное значения рассматриваемой функции. Так, для функции $\sin x$ на интервале $(0, 2\pi)$ эти числа равны нулю, минус и плюс единице.

Другой частный класс функций составляют операторы. Функцию принято называть оператором, если область ее определения не содержит чисел и то же можно сказать о ее области значений. Если из учебника квантовой физики вычеркнуть все упоминания об операторах, то в нем останется только описание некоторых экспериментов. Квантовая теория исчезнет! Наиболее часто встречающимися являются дифференциальные и интегральные операторы. Так, зависимость плотности электрического тока $j(x)$ в проводящем полупространстве $x > 0$ от электрического поля $E(x)$ при учете ненулевой длины «свободного»

пробега электронов задается интегральным оператором вида

$$j(x) = \int_0^{\infty} K(x, s)E(s) ds.$$

Главы 21 и 22 будут посвящены операторам.

Расширим понятие алгоритма. Алгоритм это не только один из способов вычислить функцию, это еще и один из способов решить ту или иную задачу. Если Вы не знаете алгоритма решения интересующей Вас задачи и не можете его найти, то, естественно, задача останется нерешенной. Любопытно, что конкретный алгоритм решает, как правило, не одну задачу, а некоторый класс задач. Так, решение всех квадратных уравнений может быть получено с помощью одного и того же алгоритма. Принято классифицировать задачи по алгоритмам, дающим их решения.

Очень важно понимать, что как задачи, так и алгоритмы имеют иерархическую структуру. Поясним с помощью примера, что мы имеем в виду. Решение дифференциального уравнения часто удается получить в виде интеграла или ряда. Затем возникает задача вычисления полученного интеграла (или ряда). Ее решение может иметь вид формулы или итерационного процесса. Для получения окончательного результата — числа, таблицы, графика или программы для компьютера — требуется ручной или машинный счет или программирование. В рассматриваемом примере участвуют задачи трех разных уровней. Верхний уровень — решение дифференциального уравнения. Следующий уровень — вычисление интеграла или ряда. Самый нижний уровень в этом примере — получение численных результатов. Можно надстроить еще один уровень — сведение физической (лучше сказать, прикладной) задачи к дифференциальному уравнению.

Как обучение математике, так и реальная организация процесса исследования строятся с учетом иерархии задач. В чем это выражается? Теория дифференциальных уравнений «умывает руки», как только ей удалось свести исследование уравнения к какой-либо задаче менее высокого уровня. Аналогично поступают теории интегралов, рядов и приближенных вычислений. Подобная «спихотехника» характерна для всех математических дисциплин. Она позволяет учащимся изучать новый материал и знакомиться с новыми алгоритмами без бесконечных повторений пройденного ранее материала. Как сказывается иерархия задач на работе, скажем, физика-теоретика? Вначале он только теорфизик. Потом он «переодевает халат» и становится специалистом по дифференциальным уравнениям. Под конец его можно увидеть «в халате» программиста или дружески с программистом беседующим. Разумеется, на всех стадиях он остается физиком, который не выпускает из-под контроля весь процесс исследования.

Соответственно строится и система алгоритмов. Каждый алгоритм сводит рассматриваемую задачу к одной или нескольким задачам менее высокого уровня, после чего уступает место своим «младшим братьям». Если этого не понимать, то любая математическая теория будет казаться оторванной от жизни, поскольку сама по себе она не дает окончательного численного результата. Исключением является, пожалуй, только элементарная арифметика.

Если читатель заметит, что описанная схема относится практически ко всем видам человеческой деятельности, он поймет, насколько тривиальным, а значит, и справедливым, является все сказанное об иерархии алгоритмов и задач.

§ 7.2. «Хорошие» и «парадоксальные» функции

У многих слово функция (как математический термин) ассоциируется с кривыми. Между тем, о функции можно сказать очень многое, не обращаясь к кривым. С другой стороны, далеко не каждую функцию можно изобразить кривой. Приведем простой пример. Определим функцию $y(x)$ следующим образом. Если x — рациональное число, то соответствующее значение y равно единице, в противном случае y равен нулю. Невозможно построить график этой функции, так как невозможно изобразить на графике «просветы» между рациональными точками; то же, разумеется, относится и к иррациональным точкам. Приведем еще один пример неизображаемой функции. Пусть если x — несократимая дробь $x = m/n$, то $y = 1/n$, если же x — иррациональное число, то $y = 0$. Существуют элементарные функции, которые нельзя изобразить, не отрывая карандаша от бумаги. В качестве примера приведем функцию (рис. 7.1)

$$\operatorname{sign} x = \frac{x}{|x|} \quad (x \neq 0), \quad \operatorname{sign} 0 = 0.$$

Такие функции называются разрывными.

На рис. 7.2 изображены части графика функции

$$y = \frac{1}{x}, \quad x \neq 0;$$

при $x = 0$ эта функция не определена, иными словами, ее область определения состоит из всех тех значений x , которые не равны нулю. Эту функцию нельзя полностью изобразить на бумаге, поскольку среди ее значений есть сколь угодно большие числа.

Выдающийся немецкий математик Карл Вейерштрасс поразил своих современников, построив пример непрерывной функции, график которой ни в одной точке не имеет касательной.

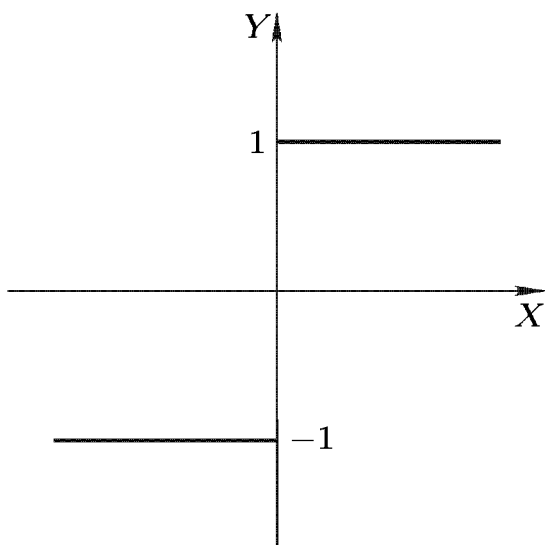


Рис. 7.1. График функции
 $y = \text{sign } x$

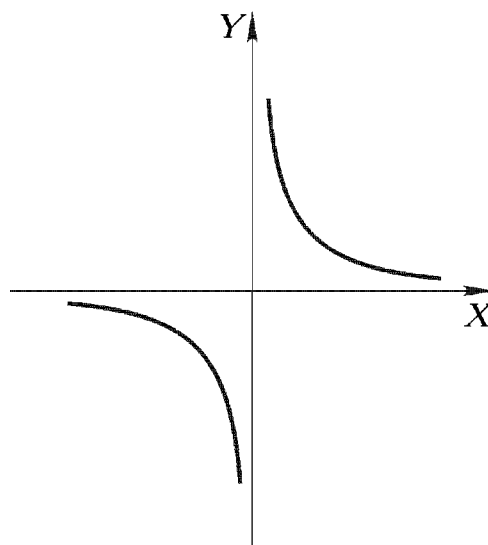


Рис. 7.2. График функции
 $y = 1/x$

Пора заняться «хорошими» функциями. Речь пойдет о непрерывных функциях. Чтобы понять, что такое непрерывная функция, рассмотрим следующую практическую задачу.

Задача. Две физические величины связаны некоторой зависимостью

$$y = f(x).$$

Путем регулирования значения величины x требуется придать y значение y_0 . Расчеты показывают, что для этого необходимо выбрать $x = x_0$. Трудность состоит в том, что практически невозможно сделать x в точности равным x_0 . Поэтому вместо того, чтобы требовать выполнения точного равенства, следует удовольствоваться выполнением приближенного равенства

$$|y - y_0| < \varepsilon, \quad (7.1)$$

где ε — допустимое отклонение y от заданного значения y_0 . Возникает вопрос, существует ли такой допуск для x

$$|x - x_0| < \delta, \quad (7.2)$$

который гарантирует выполнение условия (7.1)? Положительный ответ означает принципиальную возможность выполнить это условие. В последнем случае возникает еще один вопрос (очень важно понять, что он действительно является другим вопросом). Пусть x можно сделать сколь угодно близким к x_0 , но, разумеется, не равным этому значению. Означает ли это, что и y можно сделать сколь угодно близким к y_0 ? Оказывается, не всегда! Если не избегать искусственных примеров, то достаточно посмотреть на функцию $\text{sign } x$, описанную в начале настоящего пункта. Поэтому важно выделить класс функций $y = f(x)$ таких, что из ограничения (7.2) вытекает условие (7.1) при любом фиксированном значении ε , если только допуск δ достаточно

мал. Чем меньше ε , тем меньше должно быть δ . Всякая функция, обладающая этим свойством, называется непрерывной в точке $x = x_0$.

Как ни удивительно, это определение хорошо согласуется с интуитивно-геометрическим образом непрерывной функции. Возьмем простейшую разрывную функцию, например $y(x) = 0$ при $x \leq 0$, $y(x) = 1$ при $x > 0$.

Эта функция моделирует работу плохого водопроводного крана, который либо вовсе не пропускает воду ($x \leq 0$), либо дает максимальный поток ($x > 0$). В точке $x = 0$ функция $y(x)$ не является непрерывной: каким бы малым интервалом мы не окружили точку $x = 0$, среди значений функции в точках этого интервала будут и нуль, и единица.

Многие не любят рассуждений и определений, использующих обороты типа «для всякого положительного ε найдется такое δ , что ...». Хорошо помню, что при первом знакомстве с этим оборотом я долго не мог понять его смысл, его логическую структуру. Сейчас мне непонятно мое тогдашнее замешательство.

Позже мы увидим, что как само понятие непрерывной функции, так и многие свойства подобных функций поддаются плодотворным и нетривиальным обобщениям. Например, если функция непрерывна во всех точках некоторого интервала, включая его концы, то найдется точка, в которой функция принимает свое наибольшее на интервале значение. То же, разумеется, относится и к наименьшему значению.

Спросим себя, стоит ли требовать включения концов интервала в приведенное утверждение? Оказывается, это необходимо. В качестве примера рассмотрим функцию

$$y = \frac{1}{x}, \quad 0 < x \leq 1.$$

Она непрерывна во всех точках интервала $0 < x < 1$. Тем не менее она принимает на этом интервале сколь угодно большие значения. Причина? Рассматриваемая функция не непрерывна в точке $x = 0$, то есть на левом конце интервала.

Отметим еще одно свойство непрерывных функций. Если такая функция в каких-либо двух точках $x = a$ и $x = b$ принимает значения разных знаков, то где-то между ними она обязательно принимает равное нулю значение (если только она определена во всех промежуточных точках)! Это свойство часто используется при вычислении корней уравнений. Поясним, как это делается. Пусть, например, нужно решить уравнение

$$2^{2x} + x - 6 = 0.$$

Функция $y = 2^{2x} + x - 6 = 0$ является непрерывной и принимает значения -5 при $x = 0$ и $+12$ при $x = 2$, то есть значения разных знаков. Поэтому где-то внутри интервала $0 < x < 2$ лежит точка, в которой эта функция принимает равное нулю значение. Возьмем

середину интервала — точку $x = 1$. В этой точке $y = -1$. Иными словами, и на концах вдвое более короткого интервала $1 < x < 2$ наша функция принимает значения противоположных знаков. Следовательно, и внутри этого меньшего интервала находится по крайней мере один корень рассматриваемого уравнения. После, скажем, десяти подобных шагов искомый корень оказывается зажатым в интервале, длина которого в 1024 раза меньше длины исходного интервала. Его середина дает приближенное значение корня уравнения, а края — нижнюю и верхнюю оценки корня.

В главе 17 мы познакомимся с некоторыми полезными обобщениями понятия непрерывности функции.

Глава 8. О ТЕОРИИ АНАЛИТИЧЕСКИХ ФУНКЦИЙ

Приведенные в главе 4 примеры дают лишь слабое представление о значении комплексных чисел в современной математике. Область применения комплексных чисел значительно расширилась после того, как в поле зрения математиков попали так называемые аналитические функции. Теория аналитических функций неразрывно связана с именем выдающегося французского математика Огюста Коши (1789–1857).

§ 8.1. Аналитические функции

Каждую комплекснозначную функцию комплексной переменной можно рассматривать как пару вещественных функций $u(x, y)$, $v(x, y)$ двух вещественных переменных:

$$w(z) = u(x, y) + iv(x, y), \quad z = x + iy.$$

Однако при определении производной с ней принято обращаться как с функцией одной переменной, не используя производные по направлению:

$$w'(z) = \lim \frac{\Delta w}{\Delta z}, \quad |\Delta z| \rightarrow 0,$$

где

$$\Delta w \equiv w(z + \Delta z) - w(z).$$

Теперь можно разбить все функции комплексной переменной на два класса. К первому относятся те из них, которые дифференцируемы во всех точках некоторого, безразлично какого, круга комплексной плоскости. Функции этого класса называются аналитическими. Остальные функции составляют второй класс. Если функция принадлежит второму классу, то любой круг содержит точки, в которых она не дифференцируема.

Все неисчерпаемые потенциальные возможности теории аналитических функций содержатся в определении производной. Поэтому стоит присмотреться к нему повнимательнее. Возьмем простейшую функцию

$$w(z) = x - iy, \quad z = x + iy.$$

Придадим z приращение, равное $\Delta z = \Delta x + i\Delta y$. Соответствующее приращение w равно $\Delta w = \Delta x - i\Delta y$. Отношение этих приращений

равно единице, если приращение Δz вещественно, и минус единице, если оно чисто мнимое. Если устремить приращение z к нулю так, чтобы оно попеременно принимало вещественные и мнимые значения (что разрешается), то рассматриваемое отношение будет попеременно принимать значения плюс единица и минус единица и, следовательно, не будет стремиться к какому-либо пределу. Поэтому рассматриваемая функция, несмотря на ее незамысловатость, не является дифференцируемой ни в одной точке комплексной плоскости! Мы видим, что попасть в «клуб» аналитических функций совсем не просто.

Рассмотрим этот вопрос более подробно. Воспользуемся известной формулой для дифференциала функции двух переменных:

$$dW(x, y) = W_x dx + W_y dy;$$

добавив и отняв в правой части произведение $iW_x dy$, перепишем ее следующим образом:

$$dW = W_x(dx + idy) + (W_y - iW_x)dy.$$

Отсюда

$$\frac{dW}{dz} = W_x + \frac{(W_y - iW_x)dy}{dx + idy}.$$

Как мы уже видели, отношение $dy/(dx + idy)$ не стремится ни к какому пределу, когда $dx + idy$ стремится к нулю. Поэтому необходимым условием аналитичности функции в некоторой области является равенство

$$W_y = iW_x,$$

которое должно выполняться во всех точках этой области. Впрочем, во многих случаях аналитичность функции видна «невооруженным глазом». Мы еще вернемся к этому вопросу.

§ 8.2. Первое удивительное свойство аналитических функций

Удивительным является обилие свойств, общих для всех аналитических функций. Все они, естественно, вытекают из определения этого класса функций. Но само определение необычайно лаконично, оно выражается четырьмя словами: *функция должна быть дифференцируемой*. Ничего больше не требуется! Откуда же огромное количество следствий, изложение которых для студентов-математиков требует семестрового курса? Мы не можем ответить на этот вопрос. Остается лишь удивляться.

Здесь мы упомянем только одно свойство каждой аналитической функции. Оно состоит в следующем. Если Вы знаете точные значения функции на произвольной, даже очень малой, дуге, Вы можете вычислить ее во всех точках области ее определения. В этом отношении

аналитическая функция подобна живому организму: по одной его косточке (а теперь и по одной его клетке) ученый, говорят, может описать весь организм.

Вернемся к точным формулировкам. Предположим, что есть две аналитические функции $f_1(z)$ и $f_2(z)$, определенные в областях D_1 и D_2 соответственно. Пусть эти области имеют общую часть D_0 , содержащую некоторую дугу L и пусть во всех точках этой дуги функции $f_1(z)$ и $f_2(z)$ совпадают. Имеет место следующая теорема: *функции $f_1(z)$ и $f_2(z)$ совпадают во всех точках области D_0 .*

Эта теорема интересна и сама по себе. Однако значительно важнее одно из ее простых следствий. Будем рассуждать следующим образом. Рассмотрим функцию $f(z)$, которая в области D_1 совпадает с $f_1(z)$, а в области D_2 — с $f_2(z)$. В области D_0 она, разумеется, совпадает с каждой из этих функций. Поэтому функция $f(z)$ аналитична во всех точках области D , получаемой путем объединения областей D_1 и D_2 . Мы доказали, что в условиях теоремы существует функция $f(z)$, аналитическая в области D и совпадающая в области D_1 с исходной функцией $f_1(z)$. Функция $f(z)$ называется аналитическим продолжением функции $f_1(z)$. Что касается областей D_1 и D_2 , то они слились в одну область D подобно двум каплям жидкости на поверхности стекла. Заметим, что в описанном процессе аналитического продолжения математику отведена пассивная роль наблюдателя. Обнаружив, что две аналитические функции совпадают на некоторой дуге, он узнает, что область D_1 является частью более широкой области, в которой обе функции аналитичны (и совпадают друг с другом). Он только узнает об этом, он ничего не конструирует! Описанная ситуация делает почти очевидным следующее утверждение.

У каждой аналитической функции есть своя естественная область определения, за пределы которой она не может быть аналитически продолжена.

Именно это свойство аналитических функций позволило нам сравнить каждую из них с живым организмом.

§ 8.3. Второе свойство. Ряды Тейлора

Среди дифференцируемых функций вещественной переменной принято выделять классы дважды дифференцируемых, трижды дифференцируемых и т. д. вплоть до класса бесконечно дифференцируемых функций. Последний особенно почетен, поскольку каждой функции этого класса можно сопоставить степенной ряд (так называемый ряд Тейлора). Коэффициенты ряда Тейлора выражаются через производные исходной функции. Однако далеко не всегда этот ряд сходится и не всегда его сумма равна этой функции. Самый «аристократический» класс составляют функции, которые равны сумме своего ряда Тейлора (Брук Тейлор, 1685–1731).

В царстве аналитических функций все по-другому, демократичней: если функция $f(z)$ дифференцируема хотя бы один раз, она дифференцируема сколько угодно раз. Более того, она раскладывается в ряд Тейлора! Это замечательное утверждение является одним из краеугольных камней теории аналитических функций.

Скажем несколько слов о ряде Тейлора. Он записывается следующим образом:

$$f(z) = f(z_0) + f'(z_0)(z - z_0) + \\ + \frac{1}{2}f''(z_0)(z - z_0)^2 + \dots + \frac{1}{n!}f^{(n)}(z_0)(z - z_0)^n + \dots$$

Член этого ряда с номером n пропорционален степени $(z - z_0)^n$, где z_0 — любая точка из области определения функции $f(z)$. Выбирая различные точки z_0 , можно получить для одной и той же функции сколько угодно различных рядов Тейлора. В теории аналитических функций доказывается, что область комплексной плоскости, в которой ряд Тейлора сходится, представляет собой внутренность круга с центром в точке z_0 . Этот круг называется кругом сходимости. Лишь внутри него функцию $f(z)$ можно заменить рядом Тейлора. С другой стороны, этот ряд и нужен только для замены функции $f(z)$. Поэтому очень важно уметь вычислять радиус круга сходимости.

В вещественной области роль круга сходимости играет интервал, во всех точках которого ряд Тейлора сходится. Этот интервал называется областью сходимости. Найти область сходимости — непростая задача. Мало того, что нужно вычислить все производные рассматриваемой функции, привлечь подходящий признак сходимости степенных рядов, нужно еще проверить, что сумма равна исходной функции! В комплексной области эта же задача решается до смешного просто. Радиус круга сходимости равен расстоянию от точки z_0 до ближайшей точки, в которой функция не имеет производной. Покажем, как это делается. Пусть, например, $f(z) = 1/(1 + z^2)^n$ и $z_0 = 2$, а n — любое целое положительное число. Выбранная функция аналитична во всех точках комплексной плоскости за исключением двух точек, $z = \pm i$, в которых знаменатель обращается в нуль. Они и составляют границу области аналитичности (последняя представляет собой всю комплексную плоскость с двумя «выколотыми» точками). Интересующий нас радиус круга сходимости равен расстоянию от точки z_0 до любой из точек $\pm i$, то есть корню квадратному из пяти. Если нас интересует только область сходимости ряда Тейлора, то она представляет собой тот интервал вещественной оси, который размещается внутри круга сходимости. В рассматриваемом случае это интервал, заключенный между точками $2 - \sqrt{5}$ и $2 + \sqrt{5}$.

§ 8.4. Контурные интегралы

Еще одно удивительное свойство аналитических функций связано с контурными интегралами. Мы не будем объяснять, что это такое. Для нашей цели достаточно знать, что контурные интегралы могут отличаться друг от друга либо выбором кривой в комплексной плоскости (контура), по которой они «берутся», либо подынтегральной функцией, либо тем и другим. Это понятие играет важную роль в теории вещественных функций нескольких переменных и переносится на комплекснозначные функции двух вещественных переменных, поскольку последние можно рассматривать как пару вещественных функций.

Неожиданности начинаются, когда под знаком контурного интеграла оказывается аналитическая функция. В этом случае интеграл по замкнутому контуру равен нулю, если функция аналитична во всех точках контура и внутри него! Благодаря указанному свойству интеграл по незамкнутому контуру можно деформировать в широких пределах, не изменяя его величины. Во многих случаях это позволяет вычислять или исследовать обычные вещественные интегралы, например оценивать их или находить для них асимптотические формулы. Можно также получать неожиданные равенства между вещественными интегралами, совершенно не похожими друг на друга.

Вместо доказательства обсуждаемой теоремы «укажем пальцем» на то единственное место, где используется аналитичность функции. Если рассмотреть произвольный замкнутый контур очень малого диаметра, то во всех его точках в первом приближении подынтегральную функцию можно заменить линейной функцией

$$f(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0),$$

где (x_0, y_0) — произвольная точка внутри контура. При этом интеграл по контуру принимает вид

$$f_x \int (x - x_0)(dx + i dy) + f_y \int (y - y_0)(dx + i dy).$$

Последнее выражение можно упростить, если учесть следующие четыре элементарные тождества:

$$\int (x - x_0) dx = 0, \quad \int (y - y_0) dy = 0,$$

а интегралы

$$\int (x - x_0) dy = - \int (y - y_0) dx = \pm \Delta S,$$

то есть с точностью до знака равны ограниченной контуром площади ΔS . Выбор знака зависит от направления обхода. В результате

приходим к равенству

$$\int f(x, y) dz = \pm i(f_x - i f_y) \Delta S.$$

Уже упоминалось, что если функция $f(x, y)$ аналитична, то

$$f_x = f'(z), \quad i f_y = f'(z).$$

Таким образом, в рассматриваемом приближении интеграл по контуру равен нулю! Если функция не аналитична, то данный интеграл, вообще говоря, отличен от нуля. Это и есть то единственное место в доказательстве теоремы, где используется аналитичность подынтегральной функции.

§ 8.5. Аналитические функции встречаются на каждом шагу

Перейдем к следующему удивительному свойству аналитических функций: они встречаются всюду, их очень много. Почему это удивительно? Из геометрии известно, что чем больше свойств у некоторого класса фигур, тем, вообще говоря, этот класс уже. Так, у всех параллелограммов диагонали делят друг друга пополам, но только у ромбов они еще и взаимно перпендикулярны. Поэтому обилие общих свойств у класса аналитических функций может навести на мысль, будто этот класс очень узок. Трудно подобрать слова, которые с достаточной силой охарактеризовали бы ошибочность подобного предположения. Пожалуй, лучше дать краткое описание совокупности всех аналитических функций.

Во-первых, это все полиномы, все тригонометрические функции и логарифм. Во-вторых, это все суммы, произведения и отношения двух аналитических функций при условии, что у них есть общий круг, в котором обе функции аналитичны. В-третьих, это все сложные функции, составленные из аналитических, например $\sin \ln(1 + x^2)$.

Следующий ярус составляют решения линейных дифференциальных уравнений, коэффициенты которых являются аналитическими функциями. Сюда входят так называемые специальные функции: функции Лежандра, Бесселя (Фридрих Бессель, 1784–1846), гипергеометрические и многие другие. О важности этих функций свидетельствует то, что почти все они тщательно изучены и табулированы. Заметим еще, что во всех перечисленных до сих пор случаях никакой проверки на аналитичность не требуется. Существуют общие теоремы, охватывающие все эти случаи и гарантирующие аналитичность рассматриваемой функции.

Последний, самый загадочный ярус можно описать следующим образом. Пусть условие некоторой задачи содержит параметр. Решение задачи, естественно, зависит от этого параметра, то есть является

его функцией. Так вот, эта функция большей частью оказывается аналитической! Собственно, именно последнее обстоятельство и объясняет, почему область приложений теории аналитических функций настолько широка. Где обнаруживаются аналитические функции, там нужна их теория, а аналитические функции объявляются практически всюду!

После всего сказанного читатель с пониманием отнесется к шутливому признанию некоего математика. Оно начинается словами «От друзей я слышал, что не все функции являются аналитическими». Возможно, и физик Редже, создатель одного из направлений физики элементарных частиц, тоже «не знал», что бывают неаналитические функции. Во всяком случае, его теория основана на предположении о том, что некоторая очень важная, но, к сожалению, в каждом конкретном случае неизвестная функция является аналитической. Очень показательно, что из такого, казалось бы, малоинформативного предположения удалось извлечь множество важных выводов. И это не единственный случай.

§ 8.6. Теорема Лиувилля

Сделаем последнее замечание. В вещественной области для нахождения неизвестной функции необходимо знать, какому уравнению она удовлетворяет. В комплексной области часто удается идентифицировать аналитическую функцию не по уравнению, которому она удовлетворяет, а по ее свойствам! В качестве простейшего примера приведем знаменитую теорему Жозефа Лиувилля (1809–1882).

Если функция аналитична во всех точках комплексной плоскости, а ее модуль ограничен некоторым числом, то она принимает одно и то же значение во всех точках. В этом случае для полной идентификации функции достаточно знать ее значение только в одной-единственной точке.

Теория аналитических функций изобилует неожиданными следствиями и связями между теоремами. В качестве эффектного примера покажем, как из теоремы Лиувилля вытекает основная теорема алгебры.

Что можно сказать о полиноме $P(z) = a_n z^n + a_{n-1} z^{n-1} + \dots + a_1 z + a_0$, если он не принимает равного нулю значения ни в одной точке комплексной плоскости? Присмотримся к нему повнимательнее. Функция $Q(z) = 1/P(z)$ аналитична во всех точках (это функция из второго яруса). При $|z|$ стремящемся к бесконечности $Q(z)$ стремится к нулю, если степень полинома положительна. В таком случае вне некоторого круга C достаточно большого радиуса с центром в точке $z = 0$ функция ограничена по модулю; скажем, $|Q(z)| < 1$. Обозначим через k наименьшее значение модуля полинома $P(z)$ в точках этого круга. Оно не равно нулю, поскольку полином $P(z)$ ни в одной точке

не принимает нулевого значения. Следовательно, функция $Q(z)$ ограничена и во всех точках круга ее модуль не превышает $1/k$. Итак, функция $Q(z)$ аналитична и ограничена во всей комплексной плоскости. Вспоминая теорему Лиувилля, заключаем, что она является константой. Последнее означает, что $P(z)$ не может быть полиномом положительной степени. Значит, любой полином положительной степени принимает равное нулю значение по крайней мере в одной точке комплексной плоскости. Это и есть основная теорема алгебры!

Хочется обратить внимание читателя на необычный характер приведенного доказательства: оно не содержит выкладок! Это одна из особенностей современной математики — исследователь в значительной степени освобождается от необходимости производить громоздкие преобразования, на первый план выступает умение обнаруживать связь между, казалось бы, далекими друг от друга утверждениями. Последнее, в свою очередь, облегчает для молодых математиков выход к той границе, которая отделяет учебу от творчества. Заметим, что выделение теоретической физики в самостоятельную специальность сыграло аналогичную роль для молодых физиков.

Глава 9. КОНЕЧНЫЕ И БЕСКОНЕЧНЫЕ МНОЖЕСТВА

Новейшая история математики, а значит, и история развития абстрактных математических понятий неразрывно связаны с появлением теории множеств. Термин «множество» не имеет ничего общего с русским словом того же звучания или со словом «много». С математического языка на русский он переводится как «совокупность». Совокупность чего? Чего угодно, любая совокупность. Однако первоначально речь шла о совокупности чисел или о совокупности точек на числовой прямой. В качестве примера можно указать множество всех целых чисел или, скажем, множество, состоящее только из одного числа.

Зачем понадобилось изучать точечные множества? Ни чистые математики, ни прикладники по разным причинам не могли ограничиться миром элементарных функций. Функции, лежащие за пределами этого уютного мира, отличались большим количеством неприятных особенностей. У них могли существовать точки разрыва непрерывности, точки, в которых они не имели производных или обращались в бесконечность, и другие особые точки. Для классификации возникшего очень широкого круга функций было естественно привлечь классификацию точечных множеств, состоящих из особых точек той или иной функции. Последняя задача упрощается, если сначала научиться классифицировать произвольные точечные множества. Так родилась теория точечных множеств.

Вероятно, в то время никто не представлял себе драматических последствий этого события.

§ 9.1. Первые шаги теории

Сначала все шло хорошо. Были обнаружены и классифицированы точечные множества с удивительными свойствами. Эта классификация была перенесена на функции. Основным достижением данного периода можно считать выделение нескольких полезных классов функций, достаточно широких с точки зрения приложений и в то же время удобных для работы с ними. В частности, был выделен класс так называемых измеримых функций, которые можно интегрировать, если должным образом расширить понятие интеграла. Мы имеем в виду интеграл Лебега. Зная имя этого ученого, читатель при желании легко найдет полную теорию как измеримых функций, так и интеграла Лебега. Это красивая страница в истории математики.

Переход от точечных множеств к множествам элементов любой природы сначала происходил спокойно и почти незаметно, но потом привел к неконтролируемому «размножению» множеств. Появились множества, элементами которых являлись другие множества, например множество, состоящее из трех элементов: числа $\sqrt{2}$, совокупности всех чисел, меньших квадратного корня из двух, и совокупности всех остальных чисел. Были построены множества с неожиданными свойствами, например множество всех множеств, содержащих более двух элементов. Особенностью последнего является то, что в качестве одного из элементов оно содержит само себя.

Множества очень естественно интерпретируют законы и выводы классической логики. Поясним суть дела с помощью двух простых примеров.

§ 9.2. Теория множеств и логика

Рассмотрим следующую задачу. В классе 18 учеников. Каждый из них изучает либо французский язык, либо испанский, либо оба языка сразу. Французский язык изучают 12 учеников, а испанский — 10. Сколько учеников изучают одновременно оба эти языка?

Задача решается просто. Сложим 12 и 10. Получим 22, то есть на 4 ученика больше чем есть в классе. Следовательно, четырех учеников при сложении мы посчитали дважды. Ясно, что это те ученики, которые изучают оба языка. Задача решена. Теперь «наведем наукообразие» на это простое рассуждение и убедимся в том, что подобное малопочтенное занятие может оказаться очень полезным.

Обозначим через A множество учеников, изучающих французский язык, а через B — испанский. Число учеников в этих множествах обозначим через $N(A)$ и $N(B)$. Множество всех учеников обозначим через $A+B$, так как оно получается, если объединить в одно множество все элементы множеств A и B . Наконец, множество учеников, каждый из которых изучает два языка, обозначим символом $A \cap B$ и назовем пересечением этих двух множеств (рис. 9.1).

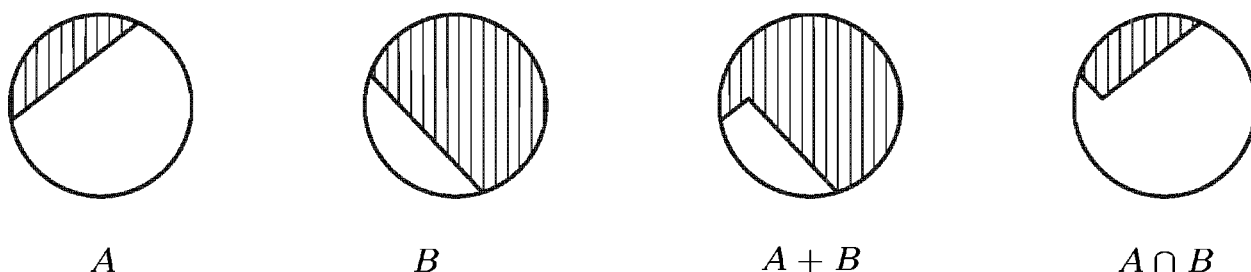


Рис. 9.1. Множества A и B , их сумма и пересечение

При решении нашей простой задачи мы, по сути дела, вывели логическое тождество

$$N(A + B) = N(A) + N(B) - N(A \cap B), \quad (9.1)$$

которое записали, используя две операции «алгебры множеств» — сложение множеств и их пересечение.

Теперь мы предложим читателю чуть более сложную задачу и покажем, что ее проще решить, пользуясь алгеброй множеств, нежели руководствуясь словесной логикой. Вот эта задача.

В том же классе ввели факультативное изучение немецкого языка. На него записалось 9 учеников, семь из которых изучали (и продолжают изучать) французский, а пять — испанский. Сколько учеников изучают все три языка?

Решение. Применяя формулу (9.1) к множествам A и $B + C$, получим

$$N(A + B + C) = N(A) + N(B + C) - N(A \cap (B + C)). \quad (9.2)$$

Легко убедиться в том, что $A \cap (B + C) = A \cap B + A \cap C$. Опять применяя тождество (9.1), находим

$$N(A \cap B + A \cap C) = N(A \cap B) + N(A \cap C) - N((A \cap B) \cap (A \cap C)). \quad (9.3)$$

Символ

$$(A \cap B) \cap (A \cap C)$$

в правой части последнего равенства означает пересечение всех трех множеств, то есть совокупность элементов, каждый из которых содержится одновременно в A , в B и в C . Обычно его записывают как

$$A \cap B \cap C.$$

В нашем случае это совокупность учеников, каждый из которых изучает все три языка. Из соотношений (9.2) и (9.3) следует, что

$$N(A + B + C) = N(A) + N(B) + N(C) - N(A \cap B) - \\ - N(A \cap C) - N(B \cap C) + N(A \cap B \cap C).$$

Подставляя числа, получим

$$18 = 12 + 10 + 9 - 4 - 9 - 5 + N(A \cap B \cap C).$$

Отсюда число учеников, изучающих все три языка, равно пяти. Задача решена. Этот простой пример доказывает существование естественной связи между алгеброй множеств и элементарной логикой.

События, о которых пойдет речь ниже, поднимают на поверхность вопросы совершенно нового типа и приводят к возникновению математической логики.

§ 9.3. Сравнение множеств по Кантору

Зададимся простодушным вопросом, каких чисел больше, рациональных или иррациональных? Поскольку и тех, и других бесконечно много, поставленный вопрос не так прост, как кажется с первого взгляда.

Вообразим группу кавалеров и дам на балу, ожидающих начала танцев. Можно ли, не подсчитав тех и других, узнать, кого больше, кавалеров или дам? Можно! Как только заиграет музыка и кавалеры пригласят дам, надо посмотреть, кто остался без пары — несколько кавалеров или несколько дам. Если танцуют все, то дам столько же, сколько кавалеров. Этот прием, казалось бы, можно использовать и для сравнения двух бесконечных множеств.

Предположим, что все элементы двух заданных множеств A и B удастся разбить на пары так, чтобы в каждой из них первый элемент был взят из A , а второй — из B и чтобы ни один элемент из A или B не входил в несколько разных пар. Конечно, не всегда такое разбиение возможно. В тех случаях, когда его удастся осуществить, говорят, что *между элементами множеств A и B можно установить взаимно однозначное соответствие*. Кажется естественным принять следующий принцип: всякие два множества имеют одинаковое количество элементов, если между их элементами существует взаимно однозначное соответствие. Однако следующий пример бросает тень сомнения на этот принцип.

Пусть A — множество всех положительных чисел, больших единицы, а B — множество всех положительных чисел, больших трех. образуем из них пары: $(x, x + 2)$, по одной паре на каждое число x из A . При этом второе число — $x + 2$ — автоматически оказывается принадлежащим множеству B и входит, очевидно, только в одну из пар. Тем самым между элементами этих множеств установлено взаимно однозначное соответствие. Согласно предложенному принципу получается, что оба множества, A и B , имеют одинаковое число элементов. Вместе с тем, интуиция подсказывает, что в A больше элементов чем в B . Вообще, если одно множество получилось из другого в результате изъятия части элементов, то естественно считать, что новое множество меньше исходного. Однако эти два принципа несовместимы.

Для доказательства ложности некоторого положения (в данном случае, предположения о совместимости двух принципов) достаточно привести только один пример. Все же мы приведем дополнительный пример. Рассмотрим множество всех целых чисел, начиная с единицы. В качестве второго возьмем множество целых чисел, которые делятся на 3. Ясно, что второе множество является частью первого. Последнее, однако, не мешает установить взаимно однозначное соответствие между их элементами. Для этого достаточно сопоставить каждому числу из первого множества втрое большее число из второго множества!

Получилось противоречие. Конечно, в этом противоречии виноваты мы сами. Мы молчаливо предположили, что между любыми двумя бесконечными множествами можно установить соотношения «больше», «меньше», обладающие желательными свойствами. Приведенный пример показывает, что это не всегда возможно.

§ 9.4. Мощность множества

Заслуга Георга Кантора (1845–1918) состоит в том, что для сравнения множеств он ввел новое, более гибкое, понятие «мощность множества». Всякие два множества, удовлетворяющие условиям первого принципа, он назвал множествами одинаковой мощности. Второй принцип Кантор модифицировал, приняв следующее определение. Будем говорить, что мощность множества A меньше мощности множества B , если оно не равномощно B и имеет ту же мощность, что и некоторое подмножество множества B . В такой форме второй принцип не противоречит первому. Из него, в частности, следует, что мощность множества не может быть меньше мощности какого-либо его подмножества.

Заметьте, что Кантор не пользуется понятием «одинаковое количество элементов». Как мы видели, для бесконечных множеств оно бесплодно.

Кантор доказал отнюдь не тривиальное утверждение: любые два множества либо равномощны, либо мощность одного из них меньше мощности другого. Он показал, что все рациональные числа можно перенумеровать, то есть присвоить каждому рациональному числу целочисленный номер, причем никакие два числа не будут иметь одинаковый номер. Это означает, что множество всех рациональных чисел имеет такую же мощность, как и множество всех целых чисел. Множества такой мощности Кантор назвал счетными.

Легко видеть, что всякое бесконечное подмножество счетного множества само является счетным. Отсюда следует, что среди бесконечных множеств наименьшую мощность имеют счетные множества.

Кантор также обнаружил, что действительные числа перенумеровать невозможно: как бы мы ни использовали целые числа для нумерации некоторого набора иррациональных чисел, полученный набор никогда не охватит все множество этих чисел. Поэтому мощность множества всех действительных чисел, называемого континуумом (от латинского слова *continuous* — непрерываемый), больше мощности множества рациональных чисел, которое «то и дело прерывается» иррациональными числами. Воспроизведем рассуждения Кантора.

Рассмотрим всевозможные не равные нулю дроби m/n . Объединим в одну группу все дроби с одной и той же суммой $|m| + n$. Если эта сумма равна k , то группа содержит ровно $2(k - 1)$ дробей. Из каждой группы удалим все имеющиеся в ней сократимые дроби. Теперь перенумеруем все оставшиеся дроби. Числу 0 присвоим первый номер. Второй и третий номер присвоим дробям из группы с $k = 2$. Затем присвоим номера дробям из группы с $k = 3$. Какую бы дробь мы ни

взяли, процесс нумерации дойдет и до нее и она получит свой номер. Мы видим, что все дроби действительно можно перенумеровать.

Как уже говорилось, любое множество, элементы которого (не обязательно числа) можно перенумеровать, то есть присвоить каждому номер, Кантор назвал счетным множеством. Он показал, что всякое множество, распадающееся на два счетные подмножества, является в свою очередь счетным. Прием, которым воспользовался Кантор, состоит в следующем. Возьмем первый элемент первого подмножества и сохраним за ним первый номер. Первому элементу второго подмножества присвоим номер два. Номер три присвоим второму элементу первого подмножества, а номер четыре — второму элементу второго подмножества. Продолжим этот процесс, вызывая попеременно очередные элементы из двух подмножеств и присваивая им очередной еще не использованный номер. В результате каждый элемент исходного множества получит свой номер. Это и означает, что множество, распадающееся на два счетные подмножества, само является счетным.

Применим этот результат к множеству всех действительных чисел. Оно состоит из двух подмножеств: рациональных и иррациональных чисел. Первое, как мы видели, является счетным. Если бы и второе подмножество являлось счетным, то счетной была бы и их сумма, то есть множество всех действительных чисел. Мы, однако, покажем, что последнее не является счетным. Это будет означать, что, как и утверждалось, множество всех иррациональных чисел нельзя перенумеровать, то есть его мощность больше мощности множества рациональных чисел. Что же представляет собой доказательство, найденное Кантором? Оно ведется «от противного». Предположим, что нам удалось перенумеровать все действительные числа. Каждое действительное число можно записать в виде бесконечной десятичной дроби. Построим бесконечную десятичную дробь, у которой число целых рано нулю, а первый знак после запятой не равен первому знаку после запятой у действительного числа, получившего первый номер. Второй знак у числа, которое мы строим, выберем произвольно, лишь бы он не совпадал со вторым десятичным знаком у действительного числа с номером два. Подобным же образом, при выборе k -го десятичного знака позаботимся о том, чтобы он не совпадал с k -м десятичным знаком действительного числа с номером k . Такой метод построения нашего числа гарантирует, что оно не совпадет ни с первым действительным числом, ни со вторым, и, вообще, ни с одним из перенумерованных чисел, то есть этого числа нет в списке. Следовательно, как бы мы ни нумеровали действительные числа, занумеровать все их невозможно. Они образуют несчетное множество. Это же относится и к множеству иррациональных чисел. Итак, мощность счетного множества меньше континуума.

Кантор также сформулировал следующую проблему: «существует ли множество, мощность которого меньше континуума, но больше мощности счетного множества?» Ниже мы вернемся к этой проблеме.

Кантор сделал и еще одно парадоксальное утверждение: множество всех точек плоскости с рациональными координатами равномощно множеству всех рациональных точек прямой. К сожалению, рамки книги не позволяют обрисовать сколько-нибудь полно роль работ Георга Кантора в становлении современной математики. «Кантор создал рай, из которого никто не сможет нас изгнать», — сказал Давид Гильберт.

Глава 10. ПОХВАЛЬНОЕ СЛОВО ЭЛЕМЕНТАРНОЙ АЛГЕБРЕ

Как известно, алгебра оперирует буквами. Каждая буква обозначает число. Не всегда это любое число. Часто значения буквы ограничиваются некоторым классом чисел, например целыми или положительными числами, что в каждом случае принято оговаривать. Буква никогда не используется для обозначения конкретного числа, если только это число не является знаменитым, вроде числа π или скорости света c . Неизбежное исключение составляют неизвестные числа. Употребляемые в алгебре буквы представляют собой более высокую степень абстракции, поскольку буква означает любое число, а, скажем, соотношение

$$a^2 - b^2 = (a - b)(a + b)$$

является абстрактной записью бесконечного множества конкретных численных соотношений. Равенство $9^2 - 5^2 = (9 - 5)(9 + 5)$ — одно из них.

Сама по себе абстрактность не является достоинством. Она превращается в наукообразие, если не приносит пользы. В чем же польза введения букв? Рассмотрим следующую часто встречающуюся ситуацию. Пусть две величины (скажем, a и b) жестко связаны друг с другом. Это могут быть, например, время и путь, проделанный за это время камнем, который брошен вертикально вниз с известной начальной скоростью; число участников шахматного турнира и число сыгранных ими партий, если любой из них играет по две партии с каждым другим участником. Можно привести еще много подобных примеров. Во всех этих примерах величины a и b с человеческой точки зрения неравноправны: если зная одну из них, например a , легко вычислить другую, то вычислить a по известному b гораздо труднее. Обычно более легкую задачу называют прямой, а более трудную — обратной задачей.

На наш взгляд, основная задача алгебры состоит в разработке методов решения обратных задач в тех случаях, когда решение прямой задачи известно. Единая схема решения содержит два этапа. Сначала мы выражаем величину b через величину a , то есть решаем прямую задачу. Полученное соотношение мы рассматриваем как уравнение, в котором величина a является неизвестной. Чтобы ее найти, нужно решить это уравнение.

Итак, сначала составляем уравнение, потом его решаем. Разработка методов решения уравнений — составная часть алгебры.

Величайшим вкладом алгебры в современную математику является введение понятия «уравнение». В арифметике этого понятия не было. Сейчас оно распространилось на всю математику. Достаточно вспомнить дифференциальные, интегральные, операторные и другие виды уравнений.

Начнем с первого этапа. Он описывает, как из величины a получить величину b или, точнее, какие арифметические действия следует произвести для этой цели над величиной a . Что значит, произвести действия?

До сих пор помню, каким шоком были для меня слова учителя: «Для того чтобы сложить или перемножить две буквы, достаточно поставить между ними знак плюс или знак умножения». «Господи — подумал я — а где же все тонкости искусства умножения столбиком?» Однако я принял новые правила игры. Я не понимал тогда, что знак плюс является командой, а ее фактическое выполнение не входит в задачу алгебры. Говоря современным языком, с помощью подобных команд строится алгоритм перехода от числа a к числу b . В физических задачах этот алгоритм обычно является записью соответствующего физического закона. Решить уравнение — значит найти алгоритм вычисления величины a по известной величине b .

Чем этот метод лучше арифметического? Многим.

1. Очень непохожие друг на друга арифметические задачи приводят к одному и тому же уравнению. Решив его, мы «одним махом» решаем все эти задачи. Например, к квадратному уравнению

$$ax^2 + bx + c = 0$$

(где x — искомая величина, a , b и c — известные числа) сводятся все задачи о свободном падении камня, если пренебречь сопротивлением воздуха. К этому же уравнению сводится задача об определении сторон прямоугольника по его площади и периметру. Еще одна из многих задач: сумма первых n целых чисел равна 210, чему равно число n ?

2. Уравнения легко поддаются классификации, а арифметические задачи очень трудно или невозможно классифицировать.

3. В алгебре разработаны методы точного и приближенного решения уравнений.

4. Уравнение (иногда вместе с некоторыми дополнительными ограничениями) содержит всю необходимую информацию и только ее. Это избавляет исследователя от анализа несущественных сведений.

5. Уравнение позволяет качественно исследовать зависимость решения от параметров задачи.

6. В подавляющем большинстве случаев алгебраический метод проще арифметического.

Многое из сказанного относится и к решению неравенств. Мы не будем на этом останавливаться.

Глава 11. МАТЕМАТИКА И ФИЗИКА. НЕМНОГО ИСТОРИИ

Предмет наших бесед — эволюция абстрактных понятий — далеко не исчерпан. Пройдено около половины намеченного пути. Однако если пройденный путь сравнить с течением реки, в которую время от времени впадают притоки, то теперь нам предстоит знакомство с устьем реки и сетью ее многочисленных рукавов. Чтобы увидеть порядок в кажущемся хаосе, понадобится небольшой экскурс в историю.

17-й век. Для математиков и физиков это век Исаака Ньютона (1642–1727) и Готфрида Лейбница (1646–1716). Мы опускаем рассказ о математике предшествующего периода. Известная фраза Ньютона — «Если мне удалось проникнуть взглядом дальше других, то это потому, что я стоял на плечах гигантов» — не только свидетельство его честности, но и очень точное описание положения дел.

Что же сделал Ньютон?

Основываясь на известных наблюдениях Тихо де Браге, Иоганн Кеплер обнаружил, что планета Марс движется по эллипсу, причем Солнце находится в одном из его фокусов. Кеплер нашел также просто формулируемые закономерности движения Марса по эллиптической орбите. Эти эмпирические закономерности носят название законов Кеплера. Именно эта титаническая работа двух астрономов позволила Ньютону сделать два замечательные открытия: обнаружить силы всемирного тяготения и установить связь между ускорением материальной точки и действующей на нее силой.

Он нашел способ записать эту связь в математической форме, для чего ему пришлось изобрести операцию дифференцирования (эту же операцию изобрел и Лейбниц).

Ньютон поставил перед собой задачу определить, как движется материальная точка под действием известной силы, то есть в случае, когда ее ускорение известно. В результате он написал первое (и много последующих) дифференциальных уравнений.

Он нашел способ решить целый ряд таких уравнений и тем самым решил множество важных задач земной и небесной механики. В частности, Ньютон теоретически вывел эмпирические законы Кеплера. После этого изменился характер астрономических наблюдений за движением планет. Объектом наблюдений стали отклонения от предсказаний модели двух тел «Солнце–планета». Их причиной является влияние других планет.

Он сравнил полученные результаты с экспериментальными данными и после уточнения массы Земли пришел к следующему выводу: закон всемирного тяготения и исходное предположение о том, что ускорение пропорционально массе, справедливы.

Сделав все это, Ньютон обессмертил свое имя.

Его великими предшественниками в области механики были Тихо де Браге (1546–1601), Иоганн Кеплер (1571–1630) и Галилео Галилей (1564–1642).

Лейбниц, как и Ньютон, изобрел дифференциальное и интегральное исчисления. Однако в отличие от Ньютона он затратил много сил на пропаганду новой теории. Лейбниц привлек большую группу европейских математиков, в том числе братьев Иоанна и Даниила Бернулли, к использованию новых методов для решения многих ранее нерешенных задач. В результате образовалась мощная группа энтузиастов, быстро раздвинувшая область применения теории. Если бы не усилия Лейбница, начало этого процесса могло затянуться на десятилетия. Немалую роль в несомненном успехе Лейбница сыграла созданная им и сохранившаяся до настоящего времени удачная система обозначений.

18-й век. Одновременный прорыв в физике (механика Ньютона, закон всемирного тяготения) и в математике (анализ, то есть соединение дифференциального и интегрального исчислений) создал чрезвычайно благоприятные условия для совместного бурного развития этих двух наук. Развитие механики и математики в последующие 100–150 лет производит впечатление взрыва.

В механике была продвинута теория твердого тела и, в частности, теория волчка, создана теория малых колебаний системы материальных точек, найден метод записи уравнений Ньютона в произвольных (обобщенных) координатах, что особенно важно при наличии жестких связей между точками. Система материальных точек и абсолютно твердое тело перестали быть единственными объектами изучения механики: на основе механики Ньютона (и закона Гука) начато изучение упругих тел, сжимаемых и несжимаемых жидкостей. Рассмотрены задачи о движении твердых тел в жидкости.

В математике начато изучение рядов как инструмента для решения дифференциальных уравнений, включая и уравнения в частных производных. Выделены линейные уравнения как объект специального изучения. Появились первые задачи, относящиеся к вариационному исчислению, и открыт метод вариаций для решения таких задач. Изучено большое количество кривых, возникающих в конкретных задачах механики.

Имена многих ученых этого времени, таких как Леонард Эйлер (1707–1783), Иоанн, Якоб и Даниил Бернулли, Жан Даламбер (1667–1748), Жозеф Лагранж (1736–1813), Пьер Лаплас (1749–1827) навсегда останутся в истории европейской культуры.

Что же восемнадцатый век оставил в наследство веку девятнадцатому? Если говорить о том, что ближе всего к теме настоящей книги, то следует коротко упомянуть два обстоятельства.

Во-первых, огромный объем решенных прикладных задач и освоенная математическая техника позволили физикам вернуться к своим специфическим задачам и, сохраняя на первых порах уважение к работе математиков, постепенно отдалиться от них. Со своей стороны математики успели удовлетворить свой аппетит к приложениям анализа и, говоря образно, стали готовиться к свободному плаванию. Итак, физика и математика несколько отделились друг от друга и каждая из этих наук занялась выбором своего пути.

Во-вторых, пока математики (Ньютон, Эйлер и другие) занимались прикладными вопросами, у них не возникало сомнений в истинности получаемых выводов. Теперь положение изменилось, и они не могли более игнорировать недоговоренности, замену аналитических доказательств обращением к геометрии и механике, отсутствие строгих определений основных понятий. Ограничимся только одним примером зыбкости основных понятий. Вот как Лейбниц объясняет понятие «бесконечно малая величина». Он пишет, что эти величины не являются истинными числами, они — фиктивные числа, однако они подчиняются тем же законам, что и обычные числа. Математики понимали, что основания анализа не прочны, но, как говорится, «руки не доходили». Впрочем, Эйлер и Лагранж сделали многое для того, чтобы освободить анализ от геометрии. Так, Эйлер изложил теорию тригонометрических функций без привлечения геометрии. Лагранж в своем фундаментальном курсе «Аналитическая механика» использовал чисто аналитические методы, не прибегая к геометрическим или физическим соображениям (после формулировки основных законов Ньютона). Читатель, вероятно, будет прав, если усомнится в педагогической ценности этих начинаний. Следует, однако, учесть их полемическую или, скорее, пропагандистскую направленность.

Итак, к концу 18-го века перед математиками возникла задача приведения в порядок основ их науки и в то же время появилась свобода выбора направления дальнейшей деятельности.

В девятнадцатом веке для обоснования анализа было сделано многое. Огюст Коши ввел понятие предела. Это является очень важным шагом, поскольку позволяет строго определить такие понятия анализа, как производная, определенный интеграл, непрерывная функция и сходимости ряда. Вместе с тем, предел по Коши не нуждается в понятии движущейся точки. Это чисто аналитическое понятие (мы о нем писали в § 4.8). Бесконечно малые величины утратили свой мистический характер и стали стремящимися к нулю переменными величинами.

Коши и Лагранж установили множество важных свойств непрерывных и дифференцируемых функций (см. § 6.2). Карл Вейерштрасс

существенно развил теорию степенных рядов и теорию непрерывных функций. Его работы закрепили уровень строгости, достигнутый Коши в определении предела. Как уже говорилось, Рихард Дедекин (1831–1916) создал строгую теорию иррациональных чисел.

В целом, в 19-м веке основания анализа достигли той же степени прочности, что и основания арифметики. На втором международном конгрессе в 1900 году один из признанных ведущих математиков — Анри Пуанкаре (1854–1912) — заявил: «Можно сказать, что сегодня в математике достигнут абсолютный уровень строгости». Дальнейшие события показали, насколько он ошибался.

Впрочем, и события прошлого не оправдывали оптимизма Пуанкаре. Речь идет о получивших к тому времени признание неевклидовых геометриях. Можно было спорить о том, каким является реальное физическое пространство, но бесспорным оставался факт, что умозрительно установить справедливость геометрии Евклида и, в частности, его пятого постулата невозможно. Для математиков это означало, что они в течение двух тысячелетий принимали на веру недоказуемый факт. Впору было задать вопрос, сколько еще недоказуемых утверждений содержит математика и нет ли среди них ошибочных? Девятнадцатый век оставил в наследство двадцатому все тот же вопрос об основах математики. Но не только его.

Математики предыдущего века накопили огромный фактический материал, создали много неожиданных направлений и открыли множество задач совершенно нового типа. К сожалению, при этом не только вся математика, но даже отдельные ее части стали необозримыми. Неудовлетворенность подобным положением усиливалась тем, что в различных, казалось бы, далеких друг от друга теориях угадывалась общая логика рассуждений. Возникла надежда заменить несколько таких теорий одной более абстрактной. Работы в этом направлении создали новую область — функциональный анализ.

Ниже, продолжая рассказ об абстрактных понятиях, мы ограничимся основаниями математики и функциональным анализом. Разумеется, мы затронем только отдельные и чаще всего наиболее простые достижения в этих двух направлениях.

Глава 12. АКСИОМАТИЧЕСКИЙ МЕТОД ГИЛЬБЕРТА

Немецкий математик Давид Гильберт (1862–1943), пожалуй, как никто другой имеет право называться архитектором математики XX века. Созданный им аксиоматический метод во многом определяет лицо современной математики.

§ 12.1. Сущность метода

Аксиома по Гильберту вовсе не является «истиной, не требующей доказательства», самоочевидной истиной. По Гильберту, система аксиом — это описание свойств тех объектов, которыми занимается данная теория, причем природа объектов для теории безразлична, важны только свойства. Возьмем, например, следующую простую систему аксиом.

1. Каждым двум различным объектам типа A соответствует не более одного объекта типа B .

2. Каждым двум различным объектам типа B соответствует не более одного объекта типа A .

Можно ли что-нибудь сказать об объектах типа A и B , не зная, что они собой представляют? Оказывается, можно! Например, можно доказать следующую теорему.

Если a_1, a_2, a_3 — три объекта типа A , и каждая из двух пар (a_1, a_2) и (a_2, a_3) определяет один и тот же объект b_0 типа B , то пара (a_1, a_3) определяет тот же объект b_0 .

Существуют ли реальные объекты, для которых выполняется рассматриваемая система аксиом? Да, существуют. Это, например, множество прямых в роли объектов A и множество точек в качестве объектов B , если соответствие точки двум прямым определить как принадлежность точки каждой из этих прямых, а соответствие прямой двум точкам — как ее прохождение через эти две точки. Можно получить еще одну реализацию рассматриваемой системы аксиом, если всюду в описании объектов A и B заменить слово «прямая» словом «плоскость», слово «точка» — словом «прямая», а соответствие в п. 1 понимать как принадлежность прямой одновременно двум плоскостям, а в п. 2 — как принадлежность двух прямых данной плоскости.

Приведенная система аксиом, несмотря на ее тривиальность и незамысловатость «приложений», дает первое представление о системе

аксиом по Гильберту. Каковы же основные черты аксиоматического метода Гильберта?

Первое, что бросается в глаза — разделение труда (и ответственности) между естествоиспытателем (физиком, биологом, статистиком и пр.) и математиком. Математик выбирает важный объект исследования, описывая его с помощью системы аксиом, а естествоиспытатель, имея дело с реальной системой, решает, достаточно ли хорошо она описывается этими аксиомами и можно ли пользоваться вытекающими из них следствиями. Поэтому, например, вопрос о справедливости геометрии Евклида, возникший после открытия неевклидовых геометрий, должен решаться и действительно был решен физиками. Этот вопрос находится вне «юрисдикции» математики.

Второй особенностью метода Гильберта является анонимность объектов, подчиняющихся данной системе аксиом. Она вытекает из самой сути метода: вместо того чтобы выводить аксиомы из физических свойств рассматриваемых объектов, метод предоставляет пользователям самим подобрать систему аксиом, которая удовлетворительно описывает объект их исследования. Это обеспечивает высокий уровень абстрактности метода — одна и та же система аксиом охватывает множество разных теорий, создатели которых, имея дело с совершенно различными понятиями, могли и не подозревать, что они, по сути, занимаются одной и той же задачей. И еще, если две различные системы аксиом обладают некоторым важным общим свойством, то это может спроецироваться на следствия из этих систем. Поясним последнее утверждение на простом примере.

Принцип двойственности. В разное время в разных областях математики было замечено, что многие аксиомы и теоремы остаются справедливыми, если в их формулировках поменять местами некоторые термины.

Например, если в аксиоме «одна и только одна прямая проходит через каждые две точки» поменять местами слова «точка» и «прямая», а выражение «проходит через» заменить на «лежит на пересечении», то получится аксиома «одна и только одна точка лежит на пересечении двух прямых». Это явление называется принципом двойственности. Заметим, что обнаружить принцип двойственности в конкретной системе аксиом не всегда просто. Для этого нужно догадаться, какие замены терминов не изменяют систему аксиом. Однако если это удалось сделать, то найденная замена оставляет справедливыми все полученные следствия и позволяет «бесплатно» получать новые следствия.

Попробуем внести скромный «вклад» в принцип двойственности. Обратимся к приведенной выше системе двух аксиом. Она совершенно симметрична относительно множеств A и B . Поэтому любое следствие из указанных аксиом остается справедливым, если в его формулировке поменять местами эти два множества. Мы получили частный случай принципа двойственности.

Вернемся к прямым и точкам и воспользуемся утверждением «если одна и та же прямая проходит через точки a_1 и a_2 и через точки a_2 и a_3 , то она совпадает с прямой, проходящей через точки a_1 и a_3 ». Легко видеть, что двойственное утверждение звучит так. «Если точка, лежащая на пересечении прямых b_1 и b_2 , совпадает с точкой пересечения прямых b_2 и b_3 , то прямые b_1 и b_3 пересекаются в этой же точке». Пусть читателя не разочаровывает тривиальный характер данных утверждений. Они приведены только для иллюстрации принципа двойственности.

Сделаем вывод.

Аксиоматический метод Гильберта упрощает изучение математики, поскольку сводит изучение нескольких различных математических теорий к изучению одной, более абстрактной и потому более понятной теории. В то же время он выделяет роль исходной системы аксиом и требует явного ее описания.

В двадцатом веке аксиоматический метод получил всеобщее признание. Его практическое применение явилось причиной бурного развития математики и появления нового направления — функционального анализа. Все последующие главы — это рассказ о функциональном анализе. В настоящей главе мы коснемся математической логики.

§ 12.2. Непротиворечивость системы аксиом

Аксиоматический метод снял вопрос об истинности аксиом. Однако если аксиомы не являются истинами, то прежде чем строить математическую теорию на базе той или иной системы аксиом, следует присмотреться к этой системе и выяснить, не является ли она порочной. У системы аксиом могут быть два порока: она может быть противоречивой и может быть неполной. Противоречивой системой аксиом пользоваться нельзя, даже если она полна. Такая система подобна справочному бюро, которое сообщает номер абонента иногда правильно, а иногда неправильно.

Непротиворечивой системой аксиом можно пользоваться, даже если она неполная. Такую систему можно сравнить с Вашим блокнотом, в котором, может быть, и нет всех нужных Вам номеров телефонов, но имеющиеся номера записаны правильно.

Следует иметь в виду, что не всегда легко с первого взгляда уличить систему аксиом в противоречивости. Приведем следующий известный пример.

«Парадокс» парикмахера. Рассмотрим поселок, в котором единственный парикмахер а) бреет тех, кто сам не бреется и б) не бреет тех, кто бреется сам.

Данное описание можно рассматривать как систему двух аксиом, позволяющую выяснить, кто бреется в парикмахерской, а кто дома. На первый взгляд это хорошая система аксиом, не содержащая никаких

противоречий. Но это только на первый взгляд. В действительности, описание противоречиво. Спросим себя, бреется ли парикмахер? Если он бреется сам, то это нарушает аксиому б. Если он сам не бреется, то это нарушает аксиому а. Что же получилось? Мы доказали несовершенство человеческой логики? Ничуть не бывало. Мы доказали, что поселок с указанными двумя свойствами не может существовать! Или, другими словами, что предложенная система аксиом противоречива.

Если система аксиом противоречива, то ни один объект, реальный или идеализированный, не может удовлетворять этой системе. Противоречивая система аксиом бессодержательна.

Пора дать определение непротиворечивости. *Система аксиом непротиворечива, если не существует такого утверждения, которое одновременно можно как доказать с ее помощью, так и доказать отрицание этого утверждения.*

Приведем еще один пример (*парадокс Расселла*).

Когда книга «Основания арифметики, логико-математическое исследование концепции числа» была готова к печати, ее автор, Готлоб Фреге, получил письмо от английского философа Бертрانا Расселла (1872–1970), в котором тот сообщал о найденном им парадоксе. Вот этот парадокс. Очевидно, что каждое множество либо содержит само себя среди своих элементов, либо не содержит. В первом случае назовем его особым, во втором случае — обыкновенным.

Рассмотрим совокупность всех обыкновенных множеств. Обозначим ее буквой R . Является ли множество R обыкновенным? Нет! Если бы R было обыкновенным, оно содержалось бы в R , то есть в себе самом, что недопустимо для обыкновенных множеств. Парадокс состоит в том, что множество R не является и особым, поскольку в этом случае его не было бы среди элементов R и оно оказалось бы обыкновенным множеством.

Какие выводы следует сделать из парадокса Расселла? Первый вывод: в теории множеств не все благополучно. Как написал Фреге в заключительном замечании к своей книге: «Вряд ли для ученого есть что-либо менее желательное, чем обнаружить шаткость основания, на котором покоится его труд, в момент его завершения; именно в таком положении я оказался, получив письмо от мистера Расселла».

Второй, более конструктивный вывод, сделанный математиками, был таков. Нужно немного исправить теорию множеств, с тем чтобы парадокс Расселла был в ней невозможен и, вообще, чтобы не было никаких парадоксов. Исключить парадокс Расселла оказалось нетрудно. Просто исключили из рассмотрения все особые множества. «Позвольте — возмутится читатель — на каком основании? Вы бы еще запретили упоминать само имя Расселла». Однако основания были. Во-первых, имела место критика самого понятия «особое множество». Исходным пунктом послужило положение: для описания множества необходимо описать все его элементы. Значит, если L — какое-либо

особое множество, то для его описания нужно иметь описание всех его элементов и, в том числе, описание его самого. Итак, для описания особого множества L нужно иметь его описание. Где же его взять?

Второе соображение заключается в следующем. Как показал тщательный анализ всех существовавших применений теории множеств, во всех случаях особые множества либо не используются, либо тот же результат можно получить и без них.

Итак, было решено (беспрецедентная ситуация в истории современной математики) отказаться от понятия «особое множество». Однако старое доброе время уже нельзя было вернуть. Парадокс Расселла подорвал доверие математиков к их коллективной интуиции. Разумеется, интуиция по-прежнему сохраняет свою главенствующую роль на эвристической стадии исследования, но доказательства должны опираться на соответствующую систему аксиом и проводиться методами математической логики.

Была создана система аксиом и в теории множеств. Точнее, было создано несколько систем аксиом, но, пожалуй, никто не мог отдать безоговорочного предпочтения какой-либо из них. Впрочем, чаще всего они приводили к одинаковым следствиям; однако когда этого не случалось, трудно было сказать, какое из следствий правдоподобней.

Вернемся к парадоксу Расселла. И здесь мы имеем дело с противоречивой системой аксиом. Трудность раскрытия этого парадокса объясняется тем, что используемая в рассуждении Расселла система аксиом не формулируется явно: отсутствует определение самого понятия «множество». Если, например, определить его как совокупность любого конечного числа любых объектов, то парадокс Расселла исчезнет. Но это слишком дорогая цена. Очень многие приложения теории множеств оперируют бесконечными множествами. Как мы уже говорили, были предложены компромиссные определения, не исключающие бесконечные множества.

Вернемся к Фреге. В основе предложенной им аксиоматики теории множеств лежит аксиома, которую кратко можно сформулировать следующим образом. *Всякое свойство определяет совокупность (множество) элементов, обладающих этим свойством.*

Так, например, свойство делиться без остатка на два выделяет множество всех четных чисел, а множество всех сирот определяется свойством не иметь живых родителей. Казалось бы, вполне разумно связать эти два понятия: множество и свойство. Однако парадокс Расселла показал, что свойство быть обыкновенным множеством не определяет какую-либо совокупность множеств.

Можно ли установить непротиворечивость данной системы аксиом? Для ответа на этот вопрос вспомним, как была доказана *непротиворечивость геометрии Болиаи–Лобачевского*.

Как известно, в двадцатые годы девятнадцатого столетия два математика — Николай Лобачевский (1792–1856) и Янош Болиаи (1802–

1860) — независимо друг от друга опубликовали систему геометрии, отличающуюся от геометрии Евклида. Если по Евклиду через точку вне прямой можно провести только одну прямую, которая не пересекается с исходной прямой, то в новой геометрии такие прямые образуют пучок, причем угол раствора этого пучка тем больше, чем больше расстояние d между точкой и прямой. Тем самым новая геометрия вводила абсолютную единицу длины d , которой следовало пользоваться для измерения расстояния между параллельными прямыми. Каждой теореме евклидовой геометрии соответствовала должным образом измененная теорема в новой геометрии. Это обстоятельство в глазах создателей новой геометрии делало обе геометрии — евклидову и неевклидову — одинаково правдоподобными. «Спор» между ними должен был решить физический эксперимент, поскольку, по сути дела, теория Лобачевского–Болиаи является физической теорией. Стоит упомянуть, что Лобачевский пытался привлечь данные из астрономии для проверки геометрии Евклида. К сожалению, их точность не позволила обнаружить какие-либо отклонения от геометрии Евклида. Только в двадцатом веке наблюдения за ходом лучей вблизи Солнца подтвердили правильность общей теории относительности, согласно которой геометрия физического пространства не совпадает с евклидовой. Правда, она не совпадает и с геометрией Лобачевского–Болиаи, однако их основная идея — необязательность геометрии Евклида — получила окончательное подтверждение.

Заметим, что Кант считал евклидову геометрию единственно возможной и потому выводимой чисто логическим путем без обращения к эксперименту. Для него евклидова геометрия была примером априорного знания, то есть знания, не нуждающегося в экспериментальном подтверждении своих положений. Само по себе появление новой геометрии не опровергает теорию Канта, хотя и подрывает веру в нее. Для окончательного опровержения требуется доказательство того, что геометрия Лобачевского–Болиаи непротиворечива.

Доказать непротиворечивость какой-либо теории представляется на первый взгляд очень трудной, почти невыполнимой задачей. В самом деле, как проследить все мыслимые следствия из аксиом данной теории и убедиться, что никакие два из них не противоречат друг другу? К счастью, имеется простой общий метод, который упрощает эту задачу и часто приводит к цели. Вспомним, что противоречивая система аксиом не имеет ни одного реального или мыслимого объекта, который бы ей удовлетворял. Поэтому теория непротиворечива, если удалось найти хотя бы один подобный объект. Именно таким образом в 1860 г. итальянский математик Эженио Бельтрами (1835–1899) доказал непротиворечивость геометрии Лобачевского–Болиаи. Он открыл поверхность, на которой действительно выполняются все аксиомы этой геометрии. Существование такой поверхности и ее свойства вытекают из аксиом геометрии Евклида!

В этом нет парадокса, поскольку из геометрии Евклида вытекает не истинность геометрии Лобачевского–Болиаи, а только ее непротиворечивость. Поясним, что непротиворечивость является свойством системы аксиом. Непротиворечивыми могут быть одновременно обе системы геометрии. Истинность той или иной геометрической системы аксиом — свойство окружающего нас физического пространства. Поэтому невозможно, чтобы обе системы были истинными, они не могут обе правильно описывать физическое пространство. Работа Бельтрами принесла новой геометрии признание математического сообщества. Все согласились с тем, что геометрия Евклида не единственно мыслимая и не обязательно именно она правильно описывает свойства физического пространства.

§ 12.3. Полнота системы аксиом

О непротиворечивости систем аксиом мы уже говорили. Что же такое полнота? Начнем с примера. Если в геометрии Евклида отбросить аксиому о параллельных прямых, то, опираясь на остальные аксиомы, невозможно ни доказать ее, ни опровергнуть. В этом смысле система оставшихся аксиом — неполная. Дадим определение полноты системы аксиом.

Система аксиом называется полной, если с ее помощью можно судить об истинности любого утверждения.

Это неплохое определение, но у него есть очевидный серьезный недостаток. Приведем возможно более тривиальный пример. Вернемся к геометрии Евклида (включая и постулат о параллельных прямых). Абсурдно ожидать, что, опираясь только на геометрию, можно судить, скажем, о возрасте Земли. Поэтому в данном определении следует как-то ограничить круг рассматриваемых утверждений.

Будем исходить из того, что всякая система аксиом описывает некоторое множество M объектов и некоторое множество P отношений между ними, например совокупность всех прямых на плоскости и два отношения: наличие или отсутствие параллельности между двумя прямыми. Всякое высказывание, использующее только элементы из M и отношения из P , будь оно истинным или ложным, назовем допустимым. В действительности, круг высказываний, называемых допустимыми, значительно шире. Например, высказывание «допустимое высказывание A является ложным» тоже является допустимым. Теперь ясно, как следует улучшить определение полноты. Нужно слово «любое» заменить выражением «любое допустимое», хотя, как мы видели, понятие допустимости высказывания не является элементарным. Мы не будем пытаться дать определение этого важного понятия.

Рассмотрим еще один пример из истории математики 19–20 веков.

§ 12.4. Проблема континуума

Перейдем к событиям, происходившим вокруг восходящей к Георгу Кантору проблемы. Она состоит в доказательстве или опровержении *континуум-гипотезы Кантора*: «не существует множества более мощного чем счетное множество, но менее мощного чем континуум».

В 1940 году Гедель опубликовал доказательство непротиворечивости общепринятой системы аксиом теории множеств, дополненной гипотезой Кантора. Иными словами, он доказал, что гипотезу Кантора невозможно опровергнуть, исходя из обычной системы аксиом теории множеств. Через двадцать три года Паул Коэн обнаружил, что, исходя из этой же системы аксиом, невозможно доказать гипотезу Кантора. Из этих двух замечательных результатов вытекает важное следствие общего характера: принятая в настоящее время система аксиом теории множеств не является полной, то есть она недостаточна для того, чтобы судить об истинности или ложности каждого мыслимого утверждения, сделанного в терминах теории множеств. Иными словами, имеются утверждения, которые нельзя ни доказать, ни опровергнуть с помощью принятой системы аксиом теории множеств. Итак, существующую систему аксиом следует дополнить. Подчеркнем, что это нужно сделать не столько для решения проблемы континуума, сколько для получения прочной основы теории множеств.

По-видимому, есть много способов дополнить существующую систему аксиом теории множеств. Например, достаточно присоединить к ней в качестве аксиомы либо гипотезу Кантора, либо ее отрицание. В результате мы получим две разные и притом непротиворечивые теории множеств и, соответственно, две разные математики.

§ 12.5. Теория Геделя

Существование двух математик — вещь удивительная. Если и удастся когда-либо выбрать одну из них, то это будет сделано не с помощью логических рассуждений, а с помощью эксперимента. В самом деле, ни одна из предлагаемых гипотез, как доказано, не противоречит принятой системе аксиом.

Задолго до описанных событий, а именно в 1931 г., было сделано не менее фундаментальное открытие. Математик Курт Гедель (1906–1978) доказал, что ни одна непротиворечивая система аксиом, включая систему аксиом арифметики, не может быть полной.

Это явилось не только полным поражением наивной веры Канта в возможность априорного познания, но и крахом программы Гильберта по построению безупречной математики с помощью аксиоматического метода. В ряду открытий, каждое из которых нанесло ощутимый удар по самолюбию человечества (таких как гелиоцентрическая

система Коперника, теория Дарвина о происхождении видов и учение Фрейда о подсознании), место и теореме Геделя.

Заметим, что при расширении системы аксиом, то есть при добавлении новых аксиом, некоторые недоказуемые ранее положения могут перейти в разряд истинных. Поэтому теорема Геделя не означает, что существуют неразрешимые вопросы. Стоит только добавить несколько аксиом, и вопрос решится с помощью традиционных математических методов. Но спросим себя, откуда возьмутся эти новые аксиомы? Ответ очевиден: отсюда же, откуда появились старые аксиомы, то есть из наблюдений, накопленного опыта, из эксперимента. Что же, математика является экспериментальной наукой? Вероятно, нет. Вернемся к идее Гильберта. Если существует несколько взаимоисключающих систем аксиом, то математик волен работать с любой из них, а пользователь (возможно, другой математик) решит, подходит ли ему именно эта система.

Подойдем к рассматриваемому вопросу по-другому. Пусть речь идет о проблеме Кантора. До тех пор, пока не возникнет практическая задача, ответ на которую зависит от справедливости его гипотезы, человечеству достаточно знать об этой гипотезе то, что оно уже знает сегодня. Как только появится подобная задача, проблему Кантора можно будет решить экспериментально!

Как бы то ни было, после работ Геделя интерес к основаниям математики несколько уменьшился. Тонкие вопросы математической логики разрабатываются, разумеется, с прежним упорством. Однако теорема Геделя показала, какими вопросами не следует заниматься, подобно тому как в свое время закон сохранения энергии показал, что не стоит изобретать вечный двигатель.

Вторая проблема Гильберта. В 1900 г. на съезде математиков Гильберт в ряду других актуальных проблем поставил задачу, которая мало кому показалась актуальной. Речь идет о знаменитой второй проблеме Гильберта — доказать состоятельность системы аксиом арифметики. Система аксиом называется состоятельной, если она непротиворечива и полна. В свете теории Геделя следует сузить формулировку второй проблемы Гильберта и исследовать непротиворечивость системы аксиом арифметики вместо ее состоятельности. Если система непротиворечива, то она не полна. Последнее не означает, что такой системой аксиом нельзя пользоваться. Поскольку она непротиворечива, то все сделанные с ее помощью выводы справедливы для любых объектов, описываемых этой системой. Коль скоро необходимые выводы получены и цель исследования достигнута, исследователю совершенно не существенно то, что среди не относящихся к исследованию положений скрываются и недоказуемые.

§ 12.6. Краткие выводы

Итак, следует примириться с тем, что любая непротиворечивая система аксиом, а только такими и можно пользоваться, не может дать ответ на все возникающие вопросы. Нечто похожее имеет место в физике элементарных частиц, которая не отвергает предположения о бесконечной сложности материи. Как показывает история физики, расширение наших знаний происходит путем добавления новых физических законов, дополняющих и видоизменяющих старые. Этот процесс аналогичен (а по сути дела, тождественен) добавлению новых аксиом к той или иной системе «старых». Новые аксиомы позволяют однозначно ответить на некоторые ранее неразрешимые вопросы. Теорема Геделя показывает, что сколько бы новых аксиом мы ни добавляли, всегда останутся или возникнут новые неразрешимые вопросы. Важно только, что выводы, получаемые на каждом шаге данного процесса, соответствуют действительности, если этим качеством обладают присоединяемые аксиомы.

Еще раз подчеркнем, что все выводы, которые удастся получить с помощью неполной системы аксиом, являются истинными, коль скоро истинна каждая аксиома используемой системы. Теорема Геделя не может опровергнуть ни одной из уже добытых истин. Совокупность законов современной физики можно рассматривать как ее систему аксиом. К счастью, исследования, проводимые на базе этих законов-аксиом, не используют строгих канонов математической логики. Аксиоматизация физики, по крайней мере в настоящее время, представляется ненужной.

Глава 13. ПРОСТРАНСТВА, ОПЕРАТОРЫ И ФУНКЦИОНАЛЬНЫЙ АНАЛИЗ

Как уже говорилось, к началу двадцатого века математики рассмотрели и решили множество задач различных типов и создали теории, обосновывающие использованные методы. Настало время привести в порядок и осмыслить все это «хозяйство». Аксиоматический метод Гильберта оказался как нельзя более подходящим для данной цели. Его зримой заслугой является создание начал функционального анализа. Грубо говоря, функциональный анализ для математики представляет собой то же, что теоретическая физика для физики. Конкретнее, это успешная попытка сделать математику более обозримой, а также способ получать новые результаты.

Построим рассказ о функциональном анализе, исходя из представления о математике как о науке, занимающейся поиском методов решения задач. Будем говорить о задачах, в которых неизвестными являются функции в широком смысле этого математического термина. Условие таких задач обычно состоит из двух частей: *общей* и *специальной*.

Общая часть содержит описание того класса функций, среди которых следует искать решение. Например, нас могут интересовать только непрерывные функции, либо только ограниченные функции, либо функции, стремящиеся к нулю, когда одна из координат или время стремятся к бесконечности. Часто встречающиеся в задачах математической физики граничные условия (например, равенство нулю температуры на поверхности охлаждаемого шара) также заранее ограничивают круг рассматриваемых функций. Назовем *пространством* совокупность всех функций, выделяемых общей частью условия данной задачи. Выбор этого термина вызван стремлением по возможности более полно использовать имеющиеся аналогии с трехмерным геометрическим пространством. Элементы пространства часто называют точками, тем самым придавая новый смысл этому термину.

Понятие «пространство» как математический термин близко к понятию «множество», но относится, как правило, к совокупности объектов, среди которых следует искать решение данной задачи. У каждой задачи, вообще говоря, свое пространство. Однако если описывать пространства аксиоматически, то оказывается, что целые классы различных пространств сливаются в одно (абстрактное) пространство.

Специальную часть обычно можно записать в виде уравнения

$$Ay = f,$$

где f — известная функция, y — искомая функция, A — заданный оператор. Понятие *оператора* было введено в главе «Функции и алгоритмы». Функциональный анализ рассматривает различные классы операторов. Как и пространства, многие из них при абстрактном описании сливаются в один класс. Это обстоятельство обеспечивает большую общность получаемых выводов и слияние в одну нескольких ранее независимых теорий.

Специальная часть задачи может выглядеть и по-другому. В качестве примера сошлемся на задачи поиска оптимального управления. Воспользуемся следующим схематическим описанием подобного рода задач. Тремя обязательными элементами каждой из них являются описание процесса (объекта) управления, имеющиеся возможности управления и критерий качества. Поясним, что может скрываться за этими терминами. Выплавка стали или выпечка хлеба, равно как воспитание ребенка или лечение болезни являются процессами. Во всех случаях существенно то, что мы можем вмешиваться в эти процессы. Каждое такое вмешательство принято называть управлением. Подобное вмешательство часто выражается в том, что в течение процесса мы обеспечиваем определенное изменение во времени одного или нескольких параметров.

Пусть для простоты мы управляем только одним параметром. Тогда управление характеризуется выбранной функцией времени. Естественно, что набор всех доступных для реализации управлений всегда ограничен по техническим причинам. Будем называть его пространством управлений. Задача состоит в том, чтобы в каждом конкретном случае выбрать наилучшее управление. Вот тут-то появляется критерий качества. Например, требуется при заданной стоимости выплавки тонны стали минимизировать затраченное на выплавку время. В этом примере заданная стоимость выделяет пространство (допустимых) управлений, а затраченное время служит критерием качества управления.

Для нас существенно, что каждому выбранному управлению отвечает свое значение критерия качества. Иными словами, критерий качества является функцией, заданной на пространстве управлений. Мы пришли к задаче нахождения управления, которое обеспечивает наибольшее (или наименьшее) значения критерия качества. Это один из частных случаев более общей задачи определения наибольшего значения функции, заданной на элементах некоторого пространства. Имея в виду эту общую задачу, полезно выделить некоторые классы функций, для которых она решается более или менее просто.

Наша цель — познакомиться с некоторыми классами пространств и некоторыми классами операторов и функций и показать, как могут

быть использованы полученные при этом сведения. Не все мыслимые пространства играют одинаково важную роль. Функциональный анализ выделил и изучил много полезных классов пространств.

§ 13.1. Основные типы пространств

Для перехода от конкретных теорий к абстрактным потребовалось ввести в «обиход» и изучить целый ряд абстрактных пространств и различных (опять-таки абстрактных) классов операторов и функций. Начнем со знакомства с коллективом пространств. По необходимости это знакомство будет беглым.

Три основные типа математических пространств имеют общую прародительницу — числовую прямую. Она содержит главные черты каждого из этих типов пространств, хотя и в простой, незамысловатой форме.

Мы уже говорили, что у каждой точки числовой прямой есть система окрестностей. Множества, обладающие этим свойством, называются *топологическими пространствами*. Следующим важным свойством точек числовой прямой является их упорядоченность: если два числа не равны друг другу, то одно из них больше другого. Перенос этого свойства в абстрактную область привел к появлению второго типа структур, *упорядоченных* и *частично упорядоченных множеств*. Наконец, простое понятие расстояния между точками легко может быть описано абстрактно, то есть перечислением его свойств, и использовано в таком виде для введения структур третьего типа — *метрических пространств*.

Понятие непрерывности, введенное в главе «Функции и алгоритмы», довольно естественно распространяется на функции, определенные в топологических и метрических пространствах, а некоторые важные свойства отрезка числовой оси переходят в свойства так называемых компактных подпространств.

Сложение и умножение чисел также может быть описано абстрактно и использовано для введения *линейных пространств*. Еще один важный класс пространств ведет свое происхождение от обычных трехмерных векторов, обобщая их свойство образовывать так называемое скалярное произведение. Далее мы коротко рассмотрим все упомянутые структуры.

Глава 14. ЛИНЕЙНЫЕ ПРОСТРАНСТВА

Линейными называются пространства, элементы которых можно складывать друг с другом и умножать на вещественные или любые комплексные числа. Ниже приводятся примеры линейных пространств. Их элементы принято также называть векторами, вероятно, потому что и векторы можно складывать друг с другом и умножать на числа. Надеемся, что каждый раз при упоминании этого термина читателю будет ясно, в каком смысле он употребляется — специальном или общем.

§ 14.1. Сложение и умножение как абстрактные операции

Облегчим себе дальнейшую работу, надолго забыв, что под элементом пространства мы подразумеваем функцию. Это позволит нам до поры до времени не говорить об области определения функции или области ее значений. Правда, одновременно мы теряем возможность свести сложение элементов к сложению чисел так же, как и умножение элемента на число. Однако обойтись без указанных действий невозможно, поэтому не остается ничего другого, как ввести их аксиоматически. Делается это так.

Начнем с того, что не только числа можно складывать и умножать на другие числа. Аналогичным свойством обладают векторы, полиномы от одной или нескольких переменных, поля скоростей движущейся жидкости, электромагнитные поля, волновые функции в квантовой механике и множество других совокупностей физических и математических объектов. При этом сумма двух векторов является вектором, а сумма двух электромагнитных полей, конечно, опять представляет собой электромагнитное поле. То же относится и к остальным перечисленным видам величин. Существенно, что какие бы однотипные величины мы ни взяли, их сложение и умножение на число приводят к величине того же типа. Эти два действия подчиняются коммутативному, двум ассоциативным и двум дистрибутивным законам:

$$x + y = y + x, \quad x + (y + z) = (x + y) + z, \quad \lambda(\mu x) = (\lambda\mu)x,$$

$$(\lambda + \mu)x = \lambda x + \mu x, \quad \lambda(x + y) = \lambda x + \lambda y.$$

Здесь скобки указывают на порядок действий; λ и μ — любые числа; x , y и z — произвольные три однотипные величины.

Условимся в дальнейшем называть любые два действия (каковы бы они ни были) умножением на число и сложением, если они обладают указанными двумя свойствами. Таким образом, два действия, умножение элемента на число и сложение элементов, определены абстрактно путем перечисления их свойств, но без указания конкретного вида (то есть без указания алгоритма, позволяющего, скажем, по двум элементам найти их сумму). Теперь можно определить понятие «линейное пространство».

§ 14.2. Линейные пространства

Пространство называется линейным, если на нем определены два действия: абстрактные сложение любых двух элементов и умножение любого элемента на произвольное число. Результаты указанных действий являются элементами того же пространства (т. е. сложение и умножение не выводят элементы пространства за его пределы). Кроме того, дополнительно предполагается, что умножение на число нуль дает нуль-элемент. Определяющим свойством нуль-элемента является выполнение равенства

$$0 + x = x$$

для любого элемента x .

Мы получили исчерпывающее определение понятия «линейное пространство». Для построения его теории остается ввести несколько простых понятий и отметить их очевидные или легко доказуемые свойства. Цепочку новых понятий начнем с *подпространства, натянутого на данные s элементов*: x_1, x_2, \dots, x_s . Так называется совокупность M тех элементов пространства, которые представимы в виде суммы

$$\lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_s x_s,$$

где $\lambda_1, \lambda_2, \dots, \lambda_s$ — всевозможные комплексные числа.

Отметим простое свойство совокупности M . Если сложить два элемента такого вида, то получится элемент того же вида. Иными словами, совокупность M инвариантна относительно сложения. Она также инвариантна относительно умножения на любые числа. Эти два свойства по определению означают, что совокупность M сама по себе является линейным пространством. Если M представляет собой часть некоторого другого линейного пространства, то ее называют подпространством, натянутым на элементы

$$x_1, x_2, \dots, x_s. \tag{14.1}$$

Перейдем к следующему понятию. Элементы (14.1) называются *линейно-независимыми*, если натянутое на них подпространство M не может быть натянуто на меньшее количество элементов. Очевидно, что линейная независимость является свойством набора элементов,

а не каждого из них в отдельности. Легко доказать от противного следующее важное свойство линейно-независимого набора элементов x_1, x_2, \dots, x_s : выполнение равенства

$$\lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_s x_s = 0$$

возможно только в тривиальном случае, когда все числовые коэффициенты λ равны нулю.

Третье и последнее понятие — размерность пространства. Пространство называется бесконечномерным, если оно содержит сколько угодно много линейно-независимых элементов. Говорят, что *размерность пространства* равна n , если максимальное число линейно-независимых элементов в нем равно n .

Нетрудно доказать следующее свойство. Если число элементов в наборе линейно-независимых элементов меньше размерности пространства, то к этому набору, не нарушая его линейной независимости, можно присоединить еще один элемент.

Заметим, что введенные понятия являются, если можно так выразиться, более абстрактными, чем те, с которыми мы встречались до сих пор. Если раньше абстрактными были объекты рассматриваемых множеств (целые числа, рациональные, иррациональные и комплексные числа, буквы), то на этой ступени сами действия над объектами становятся абстрактными: они определены перечислением свойств, т. е. аксиоматически, а не с помощью алгоритма. Среди линейных особенно полезными оказались пространства, введенные двумя выдающимися математиками — Давидом Гильбертом и Стефаном Банахом (1892–1945). Они называются, соответственно, пространствами Гильберта и Банаха. Исследование многих задач существенно упрощается, если удастся показать, что они связаны с одним из названных пространств. Ниже мы разъясним сказанное на примерах.

Начнем со знакомства с гильбертовыми пространствами, предварительно описав метрические пространства, частными случаями которых они являются.

Напомним, что метрическим называется произвольное множество объектов x, y, \dots , если задано расстояние $\rho(x, y)$ между любой парой объектов этого множества. Одновременно предполагается, что расстояние от любого объекта до него самого равно нулю, а расстояние от x до y равно расстоянию от y до x . Кроме того, должно выполняться очень важное «неравенство треугольника»: сумма расстояний от x до z и от z до y не может быть меньше расстояния от x до y . Как следует из определения, любая совокупность элементов метрического пространства сама является метрическим пространством. Если требуется подчеркнуть, что данная совокупность не совпадает со всем пространством, ее называют подпространством.

Важное понятие предела числовой последовательности очень естественно переносится на последовательности, состоящие из элементов любого абстрактного пространства: элемент z называется пределом последовательности

$$z_1, z_2, \dots, z_n, \dots,$$

если последовательность расстояний $\rho(z, z_n)$ стремится к нулю.

Заметим, что совокупность всех действительных чисел можно превратить в метрическое пространство. Для этого нужно определить расстояние между любой парой этих чисел. В качестве последнего принято брать модуль их разности. Определенное таким образом расстояние равно длине отрезка, соединяющего соответствующие числам точки числовой оси.

§ 15.1. Два примера метрических пространств

Часто встречающимся метрическим пространством является пространство всех непрерывных функций, рассматриваемых на заданном отрезке $[a, b]$ (включая его концы). Расстояние между двумя функциями $f(x)$ и $g(x)$ определяется как максимум модуля их разности:

$$\rho(f, g) = \max |f(x) - g(x)|, \quad a \leq x \leq b. \quad (15.1)$$

Это метрическое пространство принято обозначать символом $C(a, b)$. Можно определить расстояние и иначе, положив

$$\rho(f, g) = \left(\int |f(x) - g(x)|^2 dx \right)^{1/2} \quad (15.2)$$

(интегрирование производится по интервалу (a, b)). Однако это будет уже другое расстояние, или, как говорят, другая метрика. Метрика, определяемая расстоянием $\rho(f, g)$, называется равномерной; она является квадратичной метрикой. Пространство непрерывных функций со среднеквадратичной метрикой обозначают через $l(a, b)$.

Заметим, что два пространства с одинаковым составом элементов считаются разными пространствами (и обозначаются разными буквами), если у них различные метрики. Существует очень наглядное различие между пространствами $C(a, b)$ и $l(a, b)$. Пусть c — какое-либо число из интервала (a, b) . Введем в рассмотрение *разрывную* функцию $h(x)$, принимающую значение -1 между точками a и c , и значение $+1$ между точками c и b . Определим ее расстояние d до любой непрерывной функции $f(x)$ сначала с помощью формулы (14.1), а затем с помощью формулы (14.2).

Тем самым мы включаем эту функцию в оба пространства, $C(a, b)$ и $l(a, b)$. Легко видеть, что в $C(a, b)$ расстояние d не меньше единицы — разрывная функция $h(x)$ остается «чужаком» в пространстве $C(a, b)$ непрерывных функций, у нее нет близких соседей. Совсем другое положение данная разрывная функция занимает в пространстве $l(a, b)$ тех же непрерывных функций: здесь существуют сколь угодно близкие к ней непрерывные функции. Причина очевидна, она кроется в различии метрик двух рассматриваемых пространств. Сделаем вывод: присоединение разрывной функции $h(x)$ к пространству $C(a, b)$ непрерывных функций бесплодно, оно не превращает ни одну последовательность в сходящуюся! В пространстве $l(a, b)$ присоединение этой же функции полезно, она является пределом многих последовательностей функций из этого пространства.

§ 15.2. «Хорошие» (полные) и «плохие» (неполные) пространства

Начнем с того, чем мы закончили предыдущий параграф. Мы увидели, что присоединение функции $h(x)$ к пространству $l(a, b)$ превращает некоторые несходящиеся последовательности в сходящиеся. Ясно, что не одна лишь функция $h(x)$ обладает подобным свойством. Любая кусочно-непрерывная функция после присоединения ее к этому пространству порождает новые сходящиеся последовательности. Исчерпывает ли рассмотренный метод все возможности превращения несходящихся последовательностей в сходящиеся? Оказывается, не исчерпывает. Не вдаваясь в подробности, скажем только, что в рассмотрение был введен новый класс функций, названных измеримыми;

на них была распространена операция интегрирования. Как было показано, только после присоединения всех функций данного класса к функциям пространства $l(a, b)$ можно считать процесс присоединения полностью законченным. Полученное пространство принято называть пространством квадратично-интегрируемых функций и обозначать символом $L(a, b)$.

Указанная работа была проделана Лебегом и является составной частью образования всех профессиональных математиков. Упомянутый интеграл называется, естественно, интегралом Лебега. Мы лишены возможности подробно рассказать, как Лебег решил конкретную задачу дополнения пространства с непрерывной метрикой до пространства $L(a, b)$ квадратично-интегрируемых функций. Вместо этого мы познакомим читателя с гораздо более простой и значительно более общей конструкцией Г. Кантора.

Начнем с выяснения двух принципиальных вопросов.

1. Что, собственно, должно означать присоединение нового элемента к имеющемуся метрическому пространству? В ответе на этот вопрос кроется некоторая неожиданность.

2. Какие последовательности элементов любого (т. е. абстрактного) метрического пространства можно превратить в сходящиеся последовательности путем добавления нового элемента? Как ни удивительно, ответ на этот вопрос очень прост и его легко получить.

Начнем с первого вопроса. Вспомним, что мы знаем об элементах абстрактного пространства. Нам известны *только* расстояния между всевозможными парами элементов (и ничего больше!). Поэтому «обряд» присоединения нового элемента состоит в определении расстояний между ним и всеми старыми и этим исчерпывается.

Займемся вторым вопросом. Предположим, что при добавлении элемента h обнаружится некоторая последовательность элементов метрического пространства

$$x_1, x_2, \dots, x_n, \dots,$$

имеющая своим пределом этот новый элемент. До прибавления элемента h она не имела, разумеется, никакого предела (иначе после присоединения h последовательность имела бы два различных предела, что невозможно). По определению предела последовательность расстояний должна стремиться к нулю:

$$\lim \rho(x_n, h) = 0.$$

Согласно неравенству треугольника можно написать

$$\rho(x_n, x_{n+k}) \leq \rho(x_n, h) + \rho(h, x_{n+k}).$$

Правая часть последнего неравенства стремится к нулю, когда n стремится к бесконечности, а k принимает произвольные положительные

значения. Поэтому

$$\lim \rho(x_n, x_{n+k}) = 0, \quad n \rightarrow \infty, \quad k > 0. \quad (15.3)$$

Всякая последовательность, обладающая подобным свойством, называется *фундаментальной* или *последовательностью Коши*.

Мы получили следующий ответ на второй вопрос: *только последовательность Коши может превратиться в сходящуюся при добавлении нового элемента*.

Покажем, что заданную последовательность Коши всегда можно превратить в сходящуюся, присоединяя подходящий элемент x . Предположим, что это возможно. Как вычислить расстояние от x до произвольного элемента f ? Оказывается, неравенство треугольника допускает только одну возможность задать это расстояние. В самом деле, согласно неравенству треугольника можно записать

$$\rho(x, f) \leq \rho(x, x_n) + \rho(x_n, f).$$

Первое слагаемое в правой части стремится к нулю, поскольку по предположению x является пределом последовательности x_n . Поэтому

$$\rho(x, f) \leq \lim \rho(x_n, f), \quad n \rightarrow \infty.$$

Подобным же образом из другого неравенства треугольника, имеющего вид

$$\rho(x_n, f) \leq \rho(x_n, x) + \rho(x, f),$$

следует неравенство

$$\lim \rho(x_n, f) \leq \rho(x, f).$$

Значит, расстояние между присоединяемым элементом x и любым элементом f определяется однозначно:

$$\rho(x, f) = \lim \rho(x_n, f), \quad n \rightarrow \infty.$$

Полученное равенство можно рассматривать как исчерпывающее описание присоединяемого элемента x . Мы видим, что сам факт превращения последовательности Коши в сходящуюся при присоединении элемента x однозначно определяет все расстояния $\rho(x, f)$.

Легко проверить, что введенный таким образом идеальный элемент x не нарушает неравенства треугольника и, главное, действительно является пределом рассматриваемой последовательности Коши.

Итак, *путем добавления новых элементов можно превратить все последовательности Коши в сходящиеся*. Если все последовательности Коши некоторого метрического пространства являются сходящимися, то такое пространство принято называть *полным*. Все прочие пространства называются *неполными*.

Мы показали, что *всякое неполное пространство можно рассматривать как подпространство полного, получаемого путем добавления новых элементов*. Условимся называть эти новые элементы идеальными. Таким образом, мы показали, что *всякое неполное пространство можно дополнить до полного путем присоединения к нему некоторого множества идеальных элементов*. Последнее позволяет без ограничения общности считать каждое метрическое пространство полным. Почему это удобно?

Рассмотрим следующую просто формулируемую задачу: среди всех замкнутых гладких дуг, проходящих через вершины данного треугольника, найти дугу наименьшей длины.

В качестве первого шага... изменим условие задачи. Расширим класс кривых, среди которых следует искать кратчайшую. Включим в него все кусочно-гладкие замкнутые кривые, т. е. кривые, имеющие некоторое количество точек излома. Примером подобной кривой является упоминаемый в задаче треугольник. Он же дает решение видоизмененной задачи. Итак, мы решили видоизмененную задачу. Что мы узнали об исходной? Практически все, что можно узнать. Мы выяснили, что *искомой гладкой кривой не существует, но имеются гладкие кривые, длины которых сколь угодно мало отличаются от периметра рассматриваемого треугольника*. Именно поэтому среди гладких кривых нет кривой, имеющей наименьшую длину. Нетрудно представить себе подобные кривые.

Приведем менее элементарный пример. Уравнения магнитной гидродинамики допускают решения типа одиночной волны. Если вязкость магнитогидродинамической жидкости очень мала, то возникает мысль исключить из уравнений все члены, пропорциональные вязкости. При этом система уравнений настолько упрощается, что становится очевидным отсутствие у полученной таким образом идеализированной системы решений типа одиночной волны. Что же делать? Физики нашли оптимальную стратегию для подобных ситуаций. Они включили в рассмотрение разрывные функции и *среди них* нашли решение для случая нулевой вязкости. Это так называемые ударные волны. «Включение» малой вязкости приводит к сглаживанию фронта ударной волны и может быть вычислено как небольшая поправка. Идеальные элементы, т. е. разрывные решения, дают превосходное представление о реальных ударных волнах.

Возвращаясь к пространству $C(a, b)$ непрерывных функций с равномерной метрикой, заметим без доказательства, что оно является полным. Всякая последовательность Коши элементов этого пространства имеет предел.

Стоит упомянуть еще одну пару пространств: неполное пространство, состоящее из всех рациональных чисел, и соответствующее полное пространство действительных чисел.

Итак, каждое метрическое пространство можно сделать полным путем добавления идеальных элементов. Поэтому в дальнейшем без ограничения общности будем считать все упоминаемые метрические пространства полными.

§ 15.3. Непрерывные функции и компакты

Как уже говорилось, непрерывные функции на отрезке обладают тремя замечательными свойствами.

Напомним, что такое отрезок действительной оси и чем он отличается от интервала. Пусть a и b — какие-либо две точки числовой оси. Все точки оси, лежащие *между* ними, образуют *интервал*. Подчеркнем, что сами точки a и b интервалу не принадлежат. Если к интервалу присоединить эти крайние точки, получится *отрезок*. Интервал принято обозначать, используя круглые скобки, отрезок — с помощью квадратных скобок: (a, b) — интервал; $[a, b]$ — отрезок.

Казалось бы, трудно найти два различных понятия, которые столь мало отличаются друг от друга, как интервал и отрезок (всего двумя точками!). В действительности это совсем не так. Поясним, в чем тут дело.

Представим себе два города, окруженные каждый крепостной стеной. В первом городе компетенция полиции не распространяется на площадь, занимаемую крепостной стеной. Во втором городе полиция может задерживать подозрительных лиц не только на его улицах и площадях, но и на крепостной стене. Легко понять, что для скрывающихся от полиции граждан эти два города отнюдь не равноценны. И все потому, что крепостная стена входит в черту второго города и не входит в черту первого.

Вероятно, подразумеваемая в приведенном примере аналогия с интервалом и отрезком покажется многим читателям неубедительной, даже надуманной. Подождем с выводами. Рассмотрим какое-либо бесконечное множество точек, принадлежащих интервалу $(0, 1)$. Обозначим его буквой B и спросим себя, можно ли утверждать, что среди точек интервала найдется такая, в любой окрестности которой имеется бесконечное подмножество точек из B ? Оказывается, нельзя. А вот на отрезке $[0, 1]$ подобная точка (она называется *точкой сгущения*) обязательно найдется! Приведем пример. Пусть B представляет собой совокупность всех положительных дробей с целыми знаменателями, у которых в числителе стоит единица. Ясно, что в любой окрестности точки нуль содержится сколь угодно много чисел из множества B . Однако нуль не принадлежит интервалу $(0, 1)$, а отрезку $[0, 1]$ принадлежит! Значит, бывает так, что бесконечная последовательность точек интервала не имеет на нем ни одной точки сгущения. У отрезка же такая точка всегда найдется. В § 19.7 это доказывается методом поимки льва в пустыне.

Вернемся к замечательным свойствам непрерывных функций. Во-первых, они всегда ограничены на любом данном отрезке. Во-вторых, в одной или нескольких точках они достигают своего наибольшего на отрезке значения. Наконец, если среди значений функции встречаются как положительные, так и отрицательные, то в некоторой точке отрезка она обязательно принимает равное нулю значение. Заметим, что все перечисленные свойства вытекают из теоремы о точке сгущения.

— Что же тут замечательного? — удивится читатель. — Это самые естественные свойства, даже странно, что математики их доказывают.

Читатель был бы прав, если бы не существовало функций, не обладающих подобными свойствами. Однако они существуют. Поэтому выделение класса непрерывных функций является необходимой и полезной работой. С ними можно чувствовать себя уверенно. Позволительно руководствоваться интуицией, не вспоминая о тех случаях, когда эта интуиция может подвести.

А как быть с функциями, которые заданы не на числовой прямой или ее части, а на элементах некоторого пространства? Хорошо бы и среди них выделить класс непрерывных и выяснить, обладают ли и они уже известными нам свойствами непрерывных функций числового аргумента.

Начнем с того, что покажем, как распространить понятие непрерывности функций на произвольные метрические пространства. Впрочем, этого будет мало. Вспомним, что первые два из перечисленных свойств могут исчезнуть, если непрерывная функция рассматривается не на отрезке, а на интервале. Поэтому нам понадобится найти в метрических пространствах подходящий аналог отрезка числовой оси.

Итак, мы хотим обобщить на произвольные метрические пространства два важных понятия: непрерывная функция и отрезок.

Легко ввести понятие непрерывности функции. Функция $y = f(x)$ называется *непрерывной в точке x_0* , если предел $\lim f(x_n)$ существует и равен $f(x_0)$, какова бы ни была стремящаяся к x_0 последовательность x_n .

Не намного сложнее обобщить должным образом и понятие отрезка. Как мы только что упомянули, первые два из перечисленных трех свойств непрерывных на отрезке функций вытекают из существования по крайней мере одной точки сгущения у любого бесконечного множества точек, размещенных на этом отрезке. Так вот, если некоторая часть M_1 метрического пространства также содержит хотя бы одну точку сгущения любой бесконечной последовательности, размещенной на M_1 , то ее (эту часть) называют *компактом* и считают естественным обобщением понятия отрезка применительно к метрическим пространствам.

Как оказывается, любая определенная на компакте вещественнозначная непрерывная функция ограничена и достигает в некоторых

его точках наибольшего (и наименьшего) значения. Последнее свойство используется при решении задач оптимального управления.

Не следует думать, что внутренность любой сферы метрического пространства является компактом. Это, конечно, не так. Однако, как легко докажет читатель, каждое компактное подпространство находится внутри сферы достаточно большого радиуса. Во многих случаях найдены критерии, позволяющие судить, компактно ли данное подпространство. Мы не будем останавливаться на данном вопросе.

Глава 16. ГИЛЬБЕРТОВЫ И ЕВКЛИДОВЫ ПРОСТРАНСТВА

Гильбертовы и евклидовы пространства отличаются от всех прочих линейных метрических пространств тем, что их элементы можно скалярно умножать друг на друга. При этом гильбертово пространство является бесконечномерным, а евклидово — конечномерным. Сказанное может служить определением двух рассматриваемых пространств, если будет определено понятие «скалярное произведение». Именно этим мы займемся в следующем параграфе.

§ 16.1. Скалярное произведение

Если в трехмерном пространстве заданы какие-либо два вектора \mathbf{a} и \mathbf{b} , то, как известно, их можно скалярно перемножить. Полученное при этом число (\mathbf{a}, \mathbf{b}) называется их скалярным произведением. Оно может быть вычислено по формуле

$$(\mathbf{a}, \mathbf{b}) = a_x b_x + a_y b_y + a_z b_z,$$

где (a_x, a_y, a_z) и (b_x, b_y, b_z) — координаты соответствующих векторов. Если координаты вектора являются комплексными, то последнюю формулу нужно переписать следующим образом (черта над буквой означает переход к комплексно-сопряженной величине):

$$(\mathbf{a}, \mathbf{b}) = a_x \bar{b}_x + a_y \bar{b}_y + a_z \bar{b}_z.$$

Очевидно, что приведенные формулы не годятся для определения скалярного произведения в произвольном линейном пространстве. Есть, однако, общий метод перехода от конкретного понятия к абстрактному. В нашем случае он выглядит следующим образом. Прежде всего посмотрим, какими свойствами обладает конкретное скалярное произведение. Это хорошо известные свойства симметрии, линейности и положительности:

$$(\mathbf{a}, \mathbf{b}) = \overline{(\mathbf{b}, \mathbf{a})}, \quad (\lambda \mathbf{a}, \mathbf{b}) = \lambda(\mathbf{a}, \mathbf{b}),$$

$$(\mathbf{a}, \mathbf{b} + \mathbf{c}) = (\mathbf{a}, \mathbf{b}) + (\mathbf{a}, \mathbf{c}), \quad (\mathbf{a}, \mathbf{a}) \geq 0.$$

Здесь λ — произвольное число; знак равенства в последнем соотношении имеет место только при $\mathbf{a} = 0$. Легко проверить, что в векторной алгебре не используются никакие другие свойства скалярного произведения.

Возникает надежда обобщить векторную алгебру на произвольное линейное пространство путем введения абстрактного скалярного произведения. Введем следующее определение.

Если каждой паре элементов $\{x, y\}$ некоторого линейного пространства сопоставлено число (x, y) , и если это соответствие обладает перечисленными тремя свойствами, то число (x, y) называется скалярным произведением элементов x и y , а действие перехода от x и y к числу (x, y) — скалярным умножением.

Данное определение позволяет, как мы увидим, построить теорию гильбертовых пространств, не интересуясь алгоритмом вычисления скалярного произведения. В каждом конкретном случае этот алгоритм можно выбирать по собственному усмотрению, лишь бы он обеспечил все необходимые свойства. Например, можно образовать любое количество гильбертовых пространств H_k ($k = 0, 2, 4, \dots$), взяв в качестве соответствующих скалярных произведений интегралы по отрезку $[-1, 1]$:

$$(f, g)^k = \int f(t) \overline{g(t)} t^k dt. \quad (16.1)$$

Отметим, что если k — нечетное число, то интеграл (16.1) удовлетворяет не всем требованиям, предъявляемым к скалярным произведениям.

Подчеркнем, что переход от одного скалярного произведения к другому — очень радикальная операция. Прежде всего, такой переход может изменить состав элементов гильбертова пространства. Например, функция

$$F(t) = t^{-1}$$

не содержится в гильбертовом пространстве H_0 , но содержится в H_2 .

Кроме того, следует помнить, что если мы ищем приближенное решение некоторой задачи среди функций гильбертова пространства H , то сколь бы мала ни была погрешность найденного приближения, она мала только в метрике данного пространства. В метрике другого пространства эта погрешность может оказаться большой. Еще один эффект: последовательность функций, сходящаяся в одной метрике, может стать расходящейся в другой. Например, последовательность

$$Y_n(x) = \begin{cases} 1 - nx & \text{при } 0 < x < 1/n, \\ 0 & \text{при } 1/n < x < 1, \end{cases} \quad \text{где } n = 1, 2, \dots,$$

стремится к нулю в метрике пространства H_0 и не имеет предела в равномерной метрике пространства $C(0, 1)$.

§ 16.2. Простейшие теоремы и новые понятия

Познакомимся с геометрией пространств со скалярным произведением и покажем, что не только можно, но и очень удобно работать с абстрактно определенными понятиями. Введем несколько определений.

Норма элемента x определяется следующим образом:

$$\|x\| = \sqrt{(x, x)}.$$

Два элемента, x и y , называются *параллельными*, если они отличаются друг от друга только некоторым численным множителем:

$$x = \lambda y.$$

Два элемента называются *ортогональными*, если их скалярное произведение равно нулю:

$$(x, y) = 0.$$

Перейдем к простейшим теоремам.

Теорема (Пифагора). *Норма суммы двух взаимно ортогональных элементов равна квадратному корню из суммы квадратов их норм.*

Если взаимно ортогональные элементы назвать катетами, а их сумму — гипотенузой, то данная теорема будет звучать так же, как и классическая одноименная теорема. Доказательство не содержит никаких хитростей:

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

(мы дважды применили дистрибутивный закон и учли ортогональность элементов x и y). Это и есть теорема Пифагора.

Зададимся вопросом, почему это доказательство проще, чем любое доказательство классической теоремы Пифагора? Если попытаться перенести его на случай векторов в плоскости и определить скалярное произведение как сумму произведений одноименных координат, то придется доказывать, что скалярное произведение двух перпендикулярных векторов действительно равно нулю. Если же определить скалярное произведение как произведение длины одного вектора на проекцию на него другого, то придется доказывать, что оно симметрично относительно своих сомножителей. Конечно, и то, и другое можно осуществить, используя свойства подобных треугольников (аналогично доказывается и теорема Пифагора). Из сделанного замечания следует, что благодаря введению абстрактного скалярного произведения теория гильбертовых пространств избавляется от необходимости

проверять правильность введения скалярного произведения в каждом конкретном случае.

Продолжим знакомство с теорией гильбертовых пространств. По аналогии с понятием проекции одного вектора на другой введем понятие проекции элемента x на элемент y . Элемент z называется проекцией элемента x на элемент y , если он параллелен элементу y , а разность $x - z$ ортогональна элементу y .

Теорема. Норма проекции не может быть больше нормы проектируемого элемента.

Доказательство исходит из очевидного тождества:

$$x = z + (x - z).$$

По определению элемент z параллелен y , а разность $x - z$ ортогональна y , так что x представлен в виде суммы двух взаимно ортогональных элементов.

Применяем теорему Пифагора:

$$\|x\|^2 = \|z\|^2 + \|x - z\|^2.$$

Учитывая, что норма $x - z$ положительна или равна нулю, получаем

$$\|x\|^2 \geq \|z\|^2,$$

что и доказывает теорему. Заметим, что ни определение проекции, ни доказанная теорема сами по себе не устанавливают факт существования проекции. Неясно также, может ли какой-либо элемент иметь несколько различных проекций на заданный элемент. Попытаемся ответить на эти вопросы.

Предполагая существование проекции z , докажем ее единственность. По определению $z = \lambda y$ и $(x - z, y) = 0$. Исключая из этих равенств z , получим

$$0 = (x - \lambda y, y) = (x, y) - \lambda(y, y).$$

Отсюда $\lambda = (x, y)/(y, y)$ и

$$z = y \frac{(x, y)}{(y, y)}.$$

Итак, если проекция существует, то она выражается последней формулой. С другой стороны, элемент z , определенный данной формулой, является проекцией x на y : он параллелен элементу y , а разность $x - z$ ортогональна этому элементу. Следовательно, каждый элемент можно спроектировать на любой другой.

Из последнего равенства сразу же получается в абстрактной форме знаменитое неравенство Коши–Буняковского:

$$|(x, y)|^2 \leq \|x\|^2 \|y\|^2.$$

В самом деле, учитывая предыдущую теорему, находим

$$\|x\|^2 \geq \|z\|^2 = \left| \frac{(x, y)}{(y, y)} \right|^2 \|y\|^2.$$

Это и есть требуемое неравенство.

Неравенство треугольника. Данное неравенство записывается следующим образом:

$$\|x + y\| \leq \|x\| + \|y\|.$$

Приведем его простое доказательство. Справедливо выражение

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

(мы применили неравенство Коши–Буняковского). Отсюда

$$\|x + y\|^2 \leq \|x\|^2 + 2\|x\| \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2,$$

что и доказывает неравенство треугольника.

Введем еще одно «геометрическое» понятие — *расстояние* $\rho(x, y)$ между элементами x и y . Положим

$$\rho(x, y) = \|x - y\|.$$

Это определение хорошо согласуется с нашим представлением о расстоянии в физическом пространстве. Оно обладает следующими очевидными свойствами:

- 1) расстояние от любого элемента до него самого равно нулю;
- 2) расстояние от x до y равно расстоянию от y до x ;
- 3) расстояния подчиняются правилу треугольника:

$$\rho(x, y) \leq \rho(x, z) + \rho(z, y).$$

Для приложений очень важно понятие проекции элемента на подпространство. Проекцией элемента f на подпространство L_s называется элемент f_L , обладающий следующими двумя свойствами: он принадлежит L_s , а разность $f - f_L$ ортогональна всем элементам из L_s .

Возьмем в качестве L_s подпространство, натянутое на s элементов: e_1, e_2, \dots, e_s . Если все они взаимно ортогональны, то для f_L справедливо простое выражение:

$$f_L = \left[\frac{(f, e_1)}{(e_1, e_1)} \right] e_1 + \left[\frac{(f, e_2)}{(e_2, e_2)} \right] e_2 + \dots + \left[\frac{(f, e_s)}{(e_s, e_s)} \right] e_s.$$

В самом деле, этот элемент, очевидно, принадлежит подпространству L_s , а скалярное произведение (f_L, e_p) равно (f, e_p) . Поэтому разность $f - f_L$ действительно ортогональна любому из элементов e_1, e_2, \dots, e_s .

Последнее равенство принимает более простой вид, если элементы не только ортогональны, но еще и нормированы, т. е.

$$\|e_1\| = \|e_2\| = \dots = \|e_n\| = 1.$$

Тогда

$$f_L = (f, e_1)e_1 + (f, e_2)e_2 + \dots + (f, e_n)e_n. \quad (16.2)$$

Коэффициенты в правой части равенства называются коэффициентами Фурье функции f относительно элементов e_1, e_2, \dots, e_s . Если элемент f ортогонален всем элементам подпространства L_s , то его проекция на это подпространство равна нулю. Данный интуитивный факт тривиально вытекает из соотношения (16.2).

Понятие проекции часто применяется в приложениях благодаря экстремальному свойству f_L : из всех элементов подпространства L_s наиболее близким к элементу f является его проекция на L_s . Это можно записать следующим образом:

$$\|f - f_L\| \leq \|f - x\|, \quad x \in L$$

(знак равенства имеет место только в тривиальном случае — $x = f_L$). Данное свойство проекции будет использовано, когда мы дойдем до приложений теории. Приведем его простое доказательство. В очевидном тождестве

$$f - x = (f - f_L) + (f_L - x)$$

первое слагаемое ортогонально всем элементам из L_s , а второе принадлежит L_s . Значит, эти два слагаемые взаимно ортогональны и можно применить теорему Пифагора:

$$\|f - x\|^2 = \|f - f_L\|^2 + \|f_L - x\|^2.$$

Второе слагаемое в правой части неотрицательно. Отбрасывая его, получаем экстремальное свойство проекции.

Базис в пространствах со скалярным произведением. Конечная или счетная совокупность взаимно ортогональных элементов

$$e_1, e_2, \dots, e_n, \dots$$

называется базисом в пространстве со скалярным произведением, если в этом пространстве только нуль ортогонален всем элементам данной совокупности. В дальнейшем все элементы базиса предполагаются нормированными на единицу:

$$(e_n, e_n) = 1.$$

Если базис содержит ровно n элементов, то пространство называется n -мерным евклидовым пространством. Если базис содержит бесконечное множество элементов, пространство называется гильбертовым.

Как и всякое метрическое пространство, гильбертово пространство можно без ограничения общности считать полным.

Получим удобное выражение для нормы $f - f_L$. Ее можно рассматривать как погрешность, возникающую при замене элемента его проекцией. Напомним, что элемент f_L определен соотношением (16.2). Имеем

$$\|f - f_L\|^2 = (f - f_L, f - f_L) = (f, f - f_L) - (f_L, f - f_L).$$

Второе слагаемое в последнем выражении равно нулю. Поэтому

$$\|f - f_L\|^2 = \sum |(f, e_k)|^2,$$

где суммирование производится по всем элементам базиса с индексом $k > n$. Как и следовало ожидать, квадрат нормы погрешности может только уменьшаться с увеличением числа n используемых для приближения элементов.

Этим мы ограничим наше знакомство с общей теорией пространств со скалярным произведением.

§ 16.3. Абстрактное описание гильбертова пространства

Воспользуемся гильбертовым пространством для того, чтобы сделать наглядным различие между абстрактной структурой и одной из ее реализаций. Начнем с перечисления понятий, которыми мы оперировали, говоря о гильбертовых пространствах.

1. Исходные элементы:

- а) элементы гильбертова пространства (коротко: элементы или векторы);
- б) комплексные числа.

2. Действия:

- а) сложение элементов;
- б) сложение и умножение чисел;
- в) умножение элемента на число;
- г) скалярное умножение, ставящее в соответствие каждой упорядоченной паре элементов некоторое число.

Из всех перечисленных лишь действия над числами были описаны конкретно; они позаимствованы из арифметики. Остальные действия характеризовались только их свойствами.

3. Свойства действий:

- а) коммутативность и ассоциативность сложения элементов;
- б) дистрибутивность сложения элементов относительно умножения на числа;
- в) линейность, коммутативность и положительность скалярного произведения.

О перечисленных действиях мы знаем лишь их свойства, об элементах — то, какие действия можно производить над ними. Читатель легко убедится, что приведенный ранее фрагмент теории гильбертовых пространств не использовал никаких других сведений, кроме перечисленных выше. Совокупность этих сведений и составляет полное описание абстрактного гильбертова пространства.

Для сравнения приведем описание какого-либо конкретного гильбертова пространства, например пространства l_2 . Каждый элемент x из l_2 представляет собой бесконечную последовательность чисел:

$$x = (x_1, x_2, \dots),$$

удовлетворяющих условию сходимости ряда

$$\sum |x_k|^2 < \infty.$$

Сумма двух элементов и умножение элемента на число определяются естественным образом:

$$(x + y)_k = x_k + y_k, \quad (\lambda x)_k = \lambda x_k.$$

Столь же естественно определяется скалярное произведение:

$$(x, y) = x_1 \bar{y}_1 + x_2 \bar{y}_2 + \dots$$

Абсолютная сходимость последнего ряда вытекает из «школьного» неравенства:

$$2xy \leq x^2 + y^2.$$

Из него же следует и сходимость ряда

$$\sum |x + y|^2 < \infty,$$

означающая, что сумма любых двух элементов из l_2 также представляет собой элемент l_2 . Приведенное описание является конкретным: каждый элемент описан как последовательность чисел. Столь же конкретно, а именно с помощью формул, описано сложение элементов и их скалярное произведение. Если читатель на минуту вернется к описанию абстрактного гильбертова пространства, он почувствует разницу между конкретным и абстрактным описаниями.

Глава 17. БАНАХОВЫ ПРОСТРАНСТВА И НОРМИРОВАННЫЕ КОЛЬЦА

Скажем несколько слов об одном важном и плодотворном обобщении гильбертовых пространств.

§ 17.1. Банаховы пространства

Линейное метрическое пространство называется банаховым, если расстояние между его элементами зависит только от их разности. Аналогичным свойством обладают и все гильбертовы пространства. Отличие от гильбертова пространства состоит в том, что в банаховом пространстве при определении расстояния не используется понятие скалярного произведения; расстояние задается абстрактно, то есть перечислением его свойств. Само же скалярное произведение при описании банаховых пространств не упоминается. Каждое гильбертово пространство является банаховым, однако обратное утверждение неверно.

Нормой элемента f называют его расстояние до нулевого элемента:

$$\|f\| = \rho(f, 0).$$

Приведем пример банахова пространства. Совокупность всех непрерывных функций, определенных на заданном отрезке $a \leq x \leq b$ вещественной оси, превращается в пространство Банаха, если определить норму элементов $f(x)$ формулой

$$\|f\| = \max |f(x)|, \quad a \leq x \leq b$$

(для читателей, которые знакомы с понятиями максимума и минимума из курса высшей математики, скажем, что в данном определении нормы под максимумом подразумевается нечто другое, а именно наибольшее значение). Напомним, что каждая непрерывная на отрезке функция достигает своего наибольшего значения по крайней мере в одной из его точек. Именно это наибольшее значение мы обозначили символом «max» с указанием соответствующего отрезка.

Легко проверить, что введенная норма действительно обладает всеми положенными ей свойствами. Пространство, описанное в приведенном примере, принято обозначать символом $C(a, b)$. Оно называется пространством с непрерывной метрикой и, как оказывается, представляет собой полное пространство.

Теория банаховых пространств, говоря фигурально, гораздо тоньше теории гильбертовых пространств. Она была создана Стефаном Банахом и его учениками (представителями знаменитой Львовской школы) в период, непосредственно предшествующий началу Второй Мировой войны.

§ 17.2. Нормированные кольца

Вернемся к пространству $C(a, b)$, состоящему из функций $f(x)$, непрерывных на отрезке $a \leq x \leq b$. То обстоятельство, что оно является банаховым, отражает не все важные свойства его элементов. Оказывается, элементы $C(a, b)$ можно умножать друг на друга! Как выяснилось, это свойство достойно того, чтобы его аксиоматизировать. Так возникла еще одна абстрактная структура — нормированное кольцо.

В абстрактной алгебре кольцом называется любая совокупность элементов, на которой определены два действия: сложение и умножение. Результатами этих действий являются элементы из той же совокупности. Умножение обязано быть дистрибутивным относительно сложения и ассоциативным:

$$a(bc) = (ab)c, \quad a(b + c) = ab + ac.$$

Обычно элементы кольца допускают также умножение на числа. Совокупности всех чисел и всех полиномов дают нам два примера конкретных колец. Существует множество других колец, играющих важную роль в чистой математике и ее приложениях. Имеется развитая алгебраическая теория колец.

Сопоставим две структуры — банаховы пространства и кольца. Первая из них характеризуется наличием нормы и действия сложения, вторая включает сложение и умножение, но не использует понятий нормы. Сливая их в одну, мы получим новую структуру — с нормой, сложением и умножением. Эта структура имеет два названия: банахова алгебра и нормированное кольцо. Для ее описания необходимо дополнительно ввести одну естественную аксиому, определяющую свойства нормы произведения двух элементов. Мы этого делать не будем.

Простейшим приложением теории нормированных колец оказались некоторые классы линейных интегральных уравнений.

Одним из наиболее важных достижений данной теории является распространение понятия аналитичности на функции, аргументы которых представляют собой линейные операторы. Это чрезвычайно расширило множество рассматриваемых функций от операторов и облегчило их изучение. Теория нормированных колец используется при решении операторных уравнений, а также имеет значительное количество других красивых и важных приложений.

Создателем теории нормированных колец является И. М. Гельфанд, выдающийся представитель российской школы функционального анализа. Он ввел в употребление нормированные кольца и использовал для их изучения множество тонких и очень плодотворных результатов теории аналитических функций.

Глава 18. ПРИЛОЖЕНИЯ ТЕОРИИ ГИЛЬБЕРТОВЫХ И ЕВКЛИДОВЫХ ПРОСТРАНСТВ

Теория гильбертовых пространств является фундаментом одного из наиболее важных направлений математики двадцатого столетия — теории линейных операторов. Сейчас эта теория чрезвычайно расширила первоначальный круг своих интересов, включив в него изучение многих важных классов операторов и выйдя за пределы гильбертовых пространств. Ее составная часть, спектральная теория операторов, оказала на физику несомненное, хотя и косвенное влияние.

Однако в настоящей главе, демонстрируя приложения теории гильбертовых пространств, мы ограничимся примерами, не связанными с использованием операторов.

§ 18.1. Неравенство Коши

В различных гильбертовых пространствах неравенство Коши принимает различный вид. Для примера рассмотрим пространство функций $x(t)$, определенных на интервале $(0, 2\pi)$ оси t . В качестве скалярного произведения выберем интеграл

$$(x, y) = \int x(t)y^*(t)\mu(t) dt, \quad (18.1)$$

где $\mu(t)$ — какая-либо функция, принимающая только положительные значения. В этом конкретном случае неравенство Коши имеет вид

$$\left| \int x(t)y^*(t)\mu(t) dt \right|^2 \leq \int |x(t)|^2 \mu(t) dt \int |y(t)|^2 \mu(t) dt. \quad (18.2)$$

Сделаем несколько замечаний по поводу вывода формулы (18.2). Он представляет собой замечательный пример разделения труда. Пользователь, которому нужно данное неравенство, может и не подозревать о существовании гильбертовых пространств; ему требуется конкретное неравенство. Зато он знает, что такое интеграл и каковы его свойства. С другой стороны, специалист по гильбертовым пространствам может (по крайней мере теоретически) иметь самое смутное представление об интегралах. Однако ему известно неравенство Коши в абстрактной форме. Схематически их диалог можно представить себе следующим образом.

Пользователь. Справедливы ли подобные неравенства?

Специалист. Они напоминают абстрактное неравенство Коши. Попробуем определить скалярное произведение двух функций формулой (18.1). Если это удастся, то интегралы в правой части окажутся квадратами норм этих функций и неравенство (18.2) будет частным случаем неравенства Коши.

П. Что же мешает нам определить скалярное произведение формулой (18.2)?

С. Мы должны проверить, обеспечивает ли данная формула все те свойства, которые по определению должно иметь скалярное произведение и которые использовались при выводе абстрактного неравенства Коши.

П. А именно?

С. Линейность, коммутативность и положительность.

П. Из теории интегралов вытекает, что эти свойства действительно присущи интегралу (18.1), если функция $\mu(t)$ положительна.

С. Последнее утверждение на вашей ответственности. Если оно справедливо, то неравенство (18.1) при оговоренном условии доказано.

§ 18.2. Приложения к теории аппроксимации

Перейдем к еще одному приложению теории. Рассмотрим следующую задачу. Пусть требуется подобрать коэффициенты a_0, a_1, \dots, a_n тригонометрического полинома

$$T_n(\varphi) = a_0 + a_1 \cos \varphi + a_2 \cos 2\varphi + \dots + a_n \cos n\varphi$$

степени n так, чтобы он наилучшим образом приближал на интервале $(0, 2\pi)$ заданную функцию, например

$$x(\varphi) = \varphi^2.$$

По сути дела, это не одна задача, а семейство задач, и притом различной трудности. Наилучшее приближение можно понимать по-разному в зависимости от цели, ради которой оно ищется. Чем меньше расстояние между функциями, тем приближение лучше. Однако само понятие «расстояние» отнюдь не однозначно. Его можно ввести бесконечным числом неэквивалентных способов.

Приведем два из них.

1. Расстояние $\rho_0(T_n, x)$ определяется следующим образом:

$$\rho_0(T_n, x) = \max |T_n(\varphi) - x(\varphi)|, \quad 0 \leq \varphi \leq \pi.$$

2. Расстояние $\rho_k(T_n, x)$ определяется с помощью интеграла:

$$\rho_k(T_n, x) = \left[\int_0^\pi |T_n(\varphi) - x(\varphi)|^k d\varphi \right]^{1/k},$$

причем различным выборам целого числа k соответствуют различные, то есть неэквивалентные, расстояния.

Рассмотрим наиболее простую версию поставленной задачи, выбрав расстояние $\rho_2(T_n, x)$. В качестве первого шага сформулируем абстрактную задачу, одним из воплощений которой является рассматриваемая. Вторым шагом будет, естественно, решение этой абстрактной задачи. Функцию φ^2 будем считать элементом гильбертова пространства, состоящего из функций, определенных на интервале $(0, 2\pi)$ со скалярным произведением

$$(x, y) = \int x(\varphi)y^*(\varphi) d\varphi.$$

Искомая функция $T_n(\varphi)$ является элементом подпространства, натянутого на функции

$$1, \cos \varphi, \cos 2\varphi, \dots, \cos n\varphi.$$

Минимизируемая величина представляет собой расстояние между функциями φ^2 и $T_n(\varphi)$. Поэтому искомая абстрактная задача выглядит следующим образом: *найти в подпространстве, натянутом на заданные элементы e_1, e_2, \dots, e_n , элемент, наиболее близкий к данному элементу f .*

Подобная задача была решена нами для пространства со скалярным произведением. Доказано, что таким элементом является проекция элемента f на подпространство, порожденное элементами e_1, e_2, \dots, e_n . Получена формула для этой проекции в наиболее простом случае, когда указанные элементы взаимно ортогональны. К счастью, мы имеем дело с аналогичным случаем:

$$(\cos m\varphi, \cos n\varphi) = 0, \quad m \neq n,$$

так что можно воспользоваться формулой (18.2). В нашей конкретной задаче

$$(f, f) = \frac{1}{5}\pi^5, \quad \|e_k\|^2 = \frac{1}{2}\pi, \quad k \neq 0, \quad \|e_0\|^2 = \pi.$$

Для коэффициентов получаем следующие выражения:

$$a_k = \frac{4}{3}(-1)^k, \quad a_0 = \frac{1}{3}\pi^2.$$

Таким образом, искомый тригонометрический полином равен

$$T_n(\varphi) = \frac{1}{3}\pi^2 - 4 \left\{ \cos \varphi - \frac{1}{4} \cos 2\varphi + \frac{1}{9} \cos 3\varphi - \right. \\ \left. - \frac{1}{16} \cos 4\varphi + \dots + (-1)^{n+1} \frac{1}{n^2} \cos n\varphi \right\},$$

а квадрат нормы соответствующей погрешности можно вычислить по формуле

$$\Delta_n^2 \equiv \|\varphi^2 - T_n(\varphi)\|^2 = \frac{4}{45}\pi^5 - 8\pi \left[1 + \frac{1}{2^4} + \frac{1}{3^4} + \dots + \frac{1}{n^4} \right].$$

Из теории рядов Фурье известно, что погрешность стремится к нулю, когда n неограниченно возрастает. Пользуясь этим, можно получить удобное представление π^4 в виде быстро сходящегося ряда и, сверх того, следующую двустороннюю оценку погрешности:

$$1 \leq 3(n+1)^3 \Delta_n^2 \leq 1 + \frac{3}{n+1}.$$

§ 18.3. Обработка результатов ядерного эксперимента

Перейдем к третьему примеру использования теории гильбертовых пространств. Рассмотрим одну из задач, возникающих при обработке результатов эксперимента, типичного для ядерной физики.

Эксперимент состоит в следующем. Мишень обстреливается потоком частиц. Столкновение этих частиц с ядрами мишени с некоторой вероятностью приводит к ядерной реакции. Продукты реакции разлетаются в разные стороны, часть из них улавливается счетчиками, расставленными вокруг мишени. В результате физик узнает, сколько частиц (продуктов реакции) зарегистрировано каждым счетчиком.

Цель эксперимента — найти функцию $f(\theta)$, характеризующую распределения продуктов реакции по углу θ между скоростью частиц-снарядов и скоростью вылетающей частицы (продукта реакции).

Размеры как рабочей поверхности счетчиков, так и мишени обычно очень малы по сравнению с расстояниями от мишени до счетчиков, поэтому можно говорить о точечной мишени и считать известным направление скорости всех частиц, зарегистрированных в данном счетчике. Число частиц, регистрируемых в единицу времени счетчиком, расположенным под углом θ , равно $K f(\theta) dO$, где dO — телесный угол, под которым виден счетчик; K — коэффициент его эффективности, т. е. вероятность того, что частица, попавшая в счетчик, будет зарегистрирована.

Из теоретических соображений следует, что искомая функция $f(\theta)$ может с хорошей степенью точности быть представлена в виде суммы

$$f(\theta) = c_0 + c_1 \cos \theta + c_2 \cos 2\theta + \dots + c_m \cos m\theta.$$

Число m обычно невелико. Если бы теория давала не только вид функции, но и значения коэффициентов, то эксперимент был бы не нужен, разве что для проверки теории. Однако описанный эксперимент часто является практически единственным средством узнать (разумеется,

приближенно) требуемые коэффициенты. Собственно, их вычисление и является его непосредственной целью.

Если $\theta_1, \theta_2, \dots, \theta_N$ — углы, под которыми расставлены счетчики, а n_1, n_2, \dots, n_N — их показания, то объединяя эксперимент с теорией, можно записать

$$n_i = K(c_0 + c_1 \cos \theta_i + c_2 \cos 2\theta_i + \dots + c_m \cos m\theta_i) + \delta_i, \quad (18.3)$$

где δ_i — статистическая ошибка, вызванная случайным характером изучаемого процесса. Для уменьшения роли случайности время эксперимента и число счетчиков выбирают по возможности бóльшими; последнее, во всяком случае, должно существенно превышать число подлежащих определению коэффициентов. Однако сколь бы малыми ни были ошибки, их нельзя отбросить уже хотя бы потому, что после этого выписанная система уравнений (18.3) становится неразрешимой, ведь число счетчиков больше числа искомых коэффициентов, а ошибки хотя и малы, но все же не равны нулю.

Итак, ошибки неизвестны, но пренебречь ими нельзя. Стандартный путь, позволяющий обойти эту трудность, следующий. Условия разрешимости системы налагают некоторые ограничения на погрешности δ_i . Поставим задачу найти по возможности малые значения погрешностей, которые совместимы с данными ограничениями. Однако в таком виде задача все еще не является корректной, поскольку нельзя проводить минимизацию сразу по нескольким параметрам.

Обычно в качестве минимизируемого параметра выбирают сумму квадратов

$$\Delta^2 \equiv \delta_1^2 + \delta_2^2 + \dots + \delta_N^2 = \min.$$

Собственно, только теперь рассматриваемая физическая задача обрела форму математической. Мы не будем обсуждать, насколько удачен сделанный выбор математической задачи, и приступим к ее решению.

Сформулируем абстрактную задачу, частным случаем которой она является. Перепишем минимизируемое выражение следующим образом:

$$\sum |n_i - K(c_0 + c_1 \cos \theta_i + c_2 \cos 2\theta_i + \dots + c_m \cos m\theta_i)|^2 = \min.$$

Его можно рассматривать как квадрат расстояния между двумя векторами в N -мерном евклидовом пространстве: вектором

$$n = (n_1, n_2, \dots, n_N)$$

и вектором

$$r = (r_1, r_2, \dots, r_N),$$

координаты которого равны

$$r_k = K(c_0 + c_1 \cos \theta_k + \dots + c_m \cos m\theta_k),$$

если определить скалярное произведение обычным образом:

$$(n, r) = \sum n_k r_k.$$

Вектор n нам известен. Его координатами являются показания счетчиков. Вектор r — искомый. О нем известно только то, что он принадлежит подпространству, натянутому на $m + 1$ векторов:

$$C^0, C^1, \dots, C^m,$$

где C^j — вектор с координатами

$$C^j = (\cos j\theta_1, \cos j\theta_2, \dots, \cos j\theta_N), \quad j = 0, 1, 2, \dots, m.$$

Мы видим, что наша задача представляет собой частный случай следующей абстрактной задачи: в данном подпространстве найти вектор, ближайший к заданному вектору a . Мы ее уже решили — искомый вектор является проекцией вектора a на подпространство. Еще раньше было показано, как находить проекцию вектора на конечномерное подпространство, натянутое на несколько взаимно ортогональных векторов.

К сожалению, в рассматриваемом случае векторы C^j не ортогональны. Эта трудность не является принципиальной, надо только уметь заменить данный набор векторов равноценным набором взаимно ортогональных векторов. Два набора векторов называются равноценными, если натянутые на них подпространства совпадают. Не будем останавливаться на этом вопросе.

Элементарная теория гильбертовых пространств дает полное решение не только рассмотренной, но и многих других задач обработки результатов ядерного эксперимента.

Глава 19. РЕШЕНИЕ ЗАДАЧ И ТЕОРЕМЫ СУЩЕСТВОВАНИЯ

§ 19.1. О роли компьютеров

Огромное количество задач, связанных с дифференциальными уравнениями, а также прикладных задач нового типа требует создания численных методов. Существует очень распространенное заблуждение по поводу последних. Многие думают, что решив справиться с задачей с помощью численных методов, можно, забыв о математике, тут же приняться за вычисления, если под рукой есть компьютер. Помнится, была даже статья «Дело о разводе» в одной из газет. Суть ее состояла в том, что с появлением компьютеров физики наконец освобождаются от необходимости обращаться за помощью к математикам, они «разводятся» с математиками.

В действительности это не так. Прежде чем приниматься за вычисления, нужно знать, с чего начинать, что на что умножить, что делать потом. Говоря «попросту», нужен алгоритм. Как правило, задачи решаются приближенно, а приближенных алгоритмов гораздо больше, чем точных, которых тоже немало. Каждый из них имеет свои недостатки, а некоторые — свои достоинства. Программисты начинают с просмотра имеющихся программ. В преамбуле к каждой программе указывается, в каких пределах должны лежать параметры задачи, как оценить погрешность получаемого решения и объем работы, необходимый для достижения желательной точности. Это в лучшем случае. Однако часто нет ни одного готового алгоритма, его приходится создавать и затем исследовать, хорошо ли он работает. Это неизменная схема создания численных методов.

Что же изменилось после появления компьютерной техники? Очень многое. Появилась возможность использовать огромное число новых алгоритмов, применение которых ранее исключалось из-за чудовищного количества необходимых арифметических действий. Эта возможность реализована пока далеко не полностью. Здесь хочется отметить одно любопытное обстоятельство. Оказалось, что многие так называемые теоремы существования содержат превосходные численные методы. И только отсутствие компьютеров не позволяло ими пользоваться. Поясним в двух словах сказанное. Возьмем теорему существования решения некоторого уравнения. Как правило, доказательство такой теоремы состоит из двух частей: построения некоторой итерационной

последовательности чисел или функций вместе с ее пределом и доказательства того, что построенный объект (предел) действительно удовлетворяет рассматриваемому уравнению. Взятые вместе они доказывают теорему существования, причем первая часть представляет собой алгоритм решения задач, а вторая — собственно доказательство эффективности предложенного алгоритма.

§ 19.2. Два лика задачи о предсказании будущего

Думается, мы не разочаруем читателя, сразу же отмежевавшись от гадалок и астрологов. Поговорим о предсказании поведения той или иной физической системы на основе надежно установленных законов физики. Пусть для определенности речь идет о движении артиллерийского снаряда, а мы хотим узнать, попадет ли он в цель и когда именно. Прежде всего бросается в глаза, что согласно законам физики координаты и компоненты скорости снаряда, рассматриваемые как функции времени, удовлетворяют системе дифференциальных уравнений. Это позволяет рассматривать задачу о движении снаряда либо как чисто физическую, либо как чисто математическую. Мы получили два разные лика одной и той же задачи! Попробуем извлечь пользу из этого наблюдения.

С физической точки зрения очевидно, что один и тот же снаряд может лететь по-разному в зависимости от угла наклона ствола и скорости в момент вылета. Именно поэтому нужны умелые артиллеристы. Этот же факт означает, что соответствующая система дифференциальных уравнений имеет не одно, а бесчисленное множество различных решений. Этим она отличается от систем алгебраических уравнений, и это позволяет ей описывать все возможные движения снаряда.

Для того чтобы получить решение, описывающее реальное движение снаряда, нужна дополнительная информация. Опять обратимся к физической интерпретации задачи. Как воспроизвести в точности один сделанный уже выстрел? Нужно, чтобы при повторном выстреле снаряд имел в момент вылета из ствола те же величину и направление скорости, что и при первом, и, разумеется, находился в той же точке пространства. Применительно к математической формулировке сказанное означает, что задание начальных значений координат и проекций скорости (задание начального состояния) выделяет из множества возможных решений дифференциальной системы уравнений одно-единственное решение.

Мы видим, что полная формулировка всякого физического закона должна отвечать на два вопроса: как записать закон в виде системы дифференциальных уравнений и как задать начальное состояние. Обе части формулировки имеют одинаковую прочность, и правильность

принятых в физике ответов на оба эти вопроса гарантируется всем авторитетом физической науки.

Что же в физике подразумевается под состоянием системы? В каждом конкретном случае легко объяснить, что такое состояние, однако дать общее (то есть абстрактное) определение этого понятия довольно трудно, хотя и возможно. Начнем с двух примеров.

Во-первых, рассмотрим падающую материальную точку. В любой момент времени ее состояние характеризуется положением и вектором ее скорости. Их можно задать тремя координатами и тремя проекциями скорости.

Во-вторых, рассмотрим электромагнитное поле в пустоте, т. е. в отсутствие зарядов и токов. Его состояние в любой момент времени характеризуется величинами электрического $\mathbf{E}(r)$ и магнитного $\mathbf{H}(r)$ полей в каждой точке пространства, т. е. двумя вектор-функциями. Заметим, что, в отличие от первого примера, здесь производные по времени (т. е. скорости изменения полей) в определении состояния не содержатся. Это объясняется просто: уравнения Максвелла сопоставляют каждому состоянию совершенно определенные значения производных по времени. Пожалуй, лучше сказать так: уравнения Максвелла показывают, с какой скоростью изменяется данное состояние.

Что общего у двух приведенных примеров? Какие свойства их объединяют? Оказывается, таких свойств два.

1. Задание состояния системы в некоторый момент времени определяет ее состояние во все последующие моменты времени; поэтому в первом примере координаты точки сами по себе, без указания скорости не задают ее состояния. Аналогично, задание одних только скоростей тоже не задает состояния.

2. Параметры, определяющие состояние, являются независимыми переменными — их можно задавать независимо друг от друга, по крайней мере в известных пределах. Поэтому в первом примере нельзя считать состоянием совокупность координат точки, вектора скорости и вектора ускорения. Иными словами, задание состояния не должно содержать лишней информации.

Описание состояния является правильным, если законы физики позволяют по состоянию системы в некоторый момент времени рассчитать ее состояние во все последующие моменты или по крайней мере до тех пор, пока не произойдет какое-либо неконтролируемое вмешательство.

Все используемые в физике состояния установлены правильно.

В математике совокупность системы дифференциальных уравнений и начальных условий называется задачей Коши. Если ей соответствует некоторая физическая задача, то задача Коши имеет решение, притом единственное. Иными словами, для такой задачи справедлива теорема существования и единственности. Стоит ли в этих случаях ее доказывать?

§ 19.3. Почему математики считают важным доказательство теорем существования?

Мы видим три главные причины необходимости доказательства теоремы существования.

Установившаяся в математике традиция приветствует использование различного рода соображений, взятых из геометрии, физики, теории вероятностей, теории игр и т. д. Эта же традиция запрещает использовать подобные соображения при доказательствах теорем. Только доказательство, опирающееся исключительно на постулированные или ранее доказанные свойства объектов исследования, является приемлемым.

Сказанное подводит нас ко второй причине. Отсутствие строгого доказательства (или опровержения) всякого весьма правдоподобного утверждения рассматривается математиками как вызов. Всегда найдется группа математиков, принявшая этот вызов. Что касается теорем существования, то зачастую прежде, чем их доказывать, приходится искать условия, при которых теорема действительно имеет место. Иными словами, прежде чем доказывать теорему существования, нередко приходится искать правильную ее формулировку.

На третьей причине мы остановимся подробнее, хотя наиболее важной считаем первую. На каждом шагу физики прибегают к идеализациям. При этом они отходят от строгих физических законов и заменяют некоторые из них интуитивно правдоподобными. Начинает действовать принцип: если пациент нарушил хотя бы одно предписание своего врача, врач освобождается от ответственности за возможные нежелательные последствия. Применительно к нашему случаю этот принцип означает: если физик для упрощения своей задачи отошел от точных соотношений, то он не может более ссылаться на авторитет физической науки и рискует прийти к неправильным выводам. Соответствующая теорема существования уберегла бы его от возможных ошибок. Приведем два примера «из жизни».

Ударные волны в магнитной гидродинамике. Ограничимся схематическим описанием ситуации. Уравнения магнитной гидродинамики сильно упрощаются, если в них пренебречь «вязкими» членами. При этом среди решений появляются разрывные (ударные волны). Ударные волны представляют собой идеализацию гладких волн с очень крутым фронтом. Обо всем этом мы уже говорили.

Одна из задач теории ударных волн состоит в расчете столкновения двух из них. Естественно ожидать, что в итоге появится несколько новых волн, в том числе и ударных. Можно также предположить, что анализ системы уравнений магнитной гидродинамики позволит узнать, какие именно ударные волны образовались. Оказалось, однако,

что уравнения магнитной гидродинамики допускают бесконечное множество различных исходов столкновения двух конкретных ударных волн! Как это может быть, ведь эксперимент, если его поставить, должен всегда приводить к одному и тому же результату?

Несколько раньше был получен еще более удивительный парадокс. Оказалось, что задачи о столкновении ударной волны с обычной плоской волной малой интенсивности во многих случаях не имеют ни одного решения. Как будто с системой происходит нечто непредвиденное, не поддающееся описанию при принятой идеализации. Между тем, идеализация была самой обычной. Она состояла в отбрасывании в уравнениях квадратов и более высоких степеней всех малых величин. К последним были отнесены амплитуды образовавшихся волн и изменение амплитуды первичной ударной волны. Иными словами, была произведена линеаризация рассматриваемой задачи — самый распространенный метод упрощения задач с малым параметром.

Решение загадки состояло следующем. Хотя линеаризация оказалась оправданной по отношению ко всем малым волнам, поведение первичной ударной волны было угадано неправильно: она расщепилась на две ударные волны конечной амплитуды! Полученный вывод заключается в том, что в рассматриваемом случае ударная волна неустойчива относительно малых возмущений и не может наблюдаться в эксперименте. Из общего числа ударных волн, совместимых с уравнениями магнитной гидродинамики, часть должна быть исключена в связи с их неустойчивостью.

Теперь разрешается и первый парадокс. Оказывается, расчет столкновения двух ударных волн приводит к единственному возможному исходу, если при рассмотрении задачи учитывать лишь устойчивые ударные волны.

Часто ненужность для физиков теорем существования аргументируется примерно так. «Зачем мне нужна теорема существования, если я и так знаю, что искомая функция существует, тело-то как-то движется, жидкость-то как-то течет и т. д. Чем заниматься пустяками, лучше бы помогли мне найти искомое решение».

Мы не беремся судить, в какой мере физикам необходимы теоремы существования. Это вопрос сложный. Скажем только, что данная мотивировка неубедительна.

Первая половина приведенной филиппики содержит ошибку. Как уж упоминалось, физики часто (точнее будет сказать — всегда) при теоретическом исследовании явлений пользуются идеализированной математической моделью. И хотя интересующие их величины, безусловно, существуют, и нет необходимости доказывать их существование, следует, однако, помнить, что уравнения приближенной математической модели могут вообще не иметь решений. Повторим: вполне возможна ситуация, когда идеализация является настолько грубой, что соответствующая система уравнений вовсе не имеет решений. Чем

раньше узнать об этом, тем лучше. В следующем параграфе будет рассмотрен элементарный пример, показывающий, что поиски хорошего алгоритма — дело тонкое.

§ 19.4. Пример: между Сциллой и Харибдой

Пусть требуется выяснить, как изменяется сопротивление проводника при росте температуры от нуля до ста градусов. Для этого измеряется сопротивление проводника при разных температурах и составляется таблица; в ее первой строке содержатся значения температуры, при которых производились измерения, а во второй — полученные значения сопротивлений. Если измерять сопротивления через каждые пять градусов, то в таблице будет двадцать один столбец. Цель измерений — получить возможность вычислять сопротивление проводника при любой температуре между нулем и ста градусами, не прибегая к дополнительным измерениям. Для этого нужно иметь приближенную формулу, выражающую сопротивление в виде функции от температуры. Ниже приводятся три способа получения такой формулы.

Начнем с никуда не годного способа «решения» этой задачи.

1. «Погоня за точками». Как известно, коэффициенты полинома степени n всегда можно подобрать так, чтобы его график проходил через заданные $n+1$ точек. Формула, полученная Лагранжем, дает явное выражение для такого полинома. Построим с ее помощью полином Лагранжа двадцатой степени. Единственное достоинство последнего заключается в том, что он в точности воспроизводит экспериментальные данные. К сожалению, он не дает сколько-нибудь надежных сведений о сопротивлениях при других значениях температуры. Приведем одну поразительную цифру. Предположим для простоты, что при измерении сопротивления при температуре пятьдесят градусов допущена ошибка в одну сотую ома, а все остальные сопротивления измерены точно. Какова же окажется ошибка в величине сопротивления при температуре в полтора градуса? Как выясняется, при использовании формулы Лагранжа эта ошибка будет превосходить десять тысяч ом! При других температурах ошибка меньше, но все же непомерно велика. Вот к чему может привести некритичное использование точных алгоритмов.

2. «Прокрустово ложе». Второй способ несколько лучше. Он тоже не практичен, но по крайней мере не вводит в заблуждение и показывает, что сама постановка задачи нуждается в уточнении.

Обжегшись на молоке, будем дуть на воду. Если должным образом ограничить степень полинома, то получится более плавная кривая. Будем искать сопротивление y в виде полинома, скажем, третьей степени:

$$y = a_3x^3 + a_2x^2 + a_1x + a_0,$$

где x — температура. Подставляя в правую часть поочередно двадцать одно значение температуры и приравнивая каждый раз полученное число соответствующему экспериментальному значению температуры y , найдем двадцать одно уравнение с четырьмя неизвестными (a_0, a_1, a_2, a_3). Если экспериментальные данные не являются абсолютно точными (чего не бывает никогда), то эта система не имеет решения. Однако идеализируя результат эксперимента и игнорируя погрешности измерений на том основании, что они «пренебрежимо малы», мы получим пример физической задачи, не имеющей решения!

Нужно искать какой-то третий способ.

3. «От идеализма к реализму». Откажемся от идеализации экспериментальных данных. Признаем их приближенный характер. Одновременно откажемся от попытки получить точную формулу для y . Какую бы формулу мы ни выбрали, она будет приводить к случайным ошибкам, поскольку исходные ошибки в экспериментальных данных являются случайными. Возникает задача — найти вероятностные характеристики ошибок при вычислении y и по возможности оптимизировать их путем наилучшего выбора рабочей формулы. В § 18.3 достаточно подробно рассмотрен конкретный пример подобной ситуации.

Существует и еще одна опасность. Не нужно думать, что если идеализированная задача не имеет решений, то в процессе счета, «решая» ее численными методами, мы обязательно обнаружим, что задача неразрешима. К сожалению, это не так. Предположим, что мы ищем решение каким-либо итерационным методом и дошли до третьего приближения. Разве это будет надежным свидетельством того, что приближенное уравнение имеет решение? Разумеется, нет.

Нарисованная картина будет односторонней, если не упомянуть, что теоретическая физика выработала свои каноны строгости, которые предохраняют опытных теоретиков от описанных неприятностей. Обращаясь к молодым теорфизикам, хочется сказать: «Думайте сами, решайте сами». Разумеется, всегда лучше иметь под рукой нужную теорему существования решения рассматриваемой задачи, приближенный алгоритм ее решения вместе с оценкой погрешности, вносимой этим алгоритмом, и оценкой трудоемкости.

Ниже мы познакомимся с некоторыми видами доказательств теорем существования.

§ 19.5. Операторы и уравнения

Если понятие пространства было введено как абстрактный образ той или иной совокупности функций, то для абстрактного описания уравнений требуется понятие оператора. Простейшим примером оператора является функция одной независимой переменной, $y = f(x)$,

заданной на некотором множестве чисел E . Она ставит в соответствие каждому значению величины x из E некоторое числовое значение величины y . Для фактического задания соответствия нужен, конечно, алгоритм, т. е. рецепт нахождения по данному значению x соответствующего ему значения y . Тем не менее эти два понятия — соответствие и алгоритм — не совпадают: одно и то же соответствие может быть задано с помощью различных алгоритмов, например таблицей, кривой или одной из многих эквивалентных алгебраических формул.

Если в определении функции заменить числовое множество E каким-либо пространством, а под y подразумевать элемент того же или некоторого другого пространства, то мы получим определение оператора. Операторы принято обозначать большими латинскими буквами. Всякое уравнение можно представить как одну из реализаций абстрактного уравнения $Ax = f$. Остается выяснить, какую пользу можно извлечь из такого представления.

Сделаем маленький экскурс в историю. Рассмотрим некоторые этапы развития методов решения уравнений. Начнем с алгебраического подхода. Он имеет дело с алгебраическим уравнением и состоит в том, чтобы с помощью допустимых преобразований привести это уравнение к виду « x равняется известной величине». Данный подход, естественно, был распространен и на дифференциальные уравнения. При этом в число допустимых операций пришлось включить интегрирование. Вскоре стало ясно, что возможности подобного подхода весьма ограничены. В алгебре была доказана знаменитая теорема; в силу нее уравнения выше четвертой степени, которые можно разрешить в радикалах (а других операций, не считая арифметических, в алгебре не существует), являются скорее исключением, нежели правилом. Аналогичное положение вскоре обнаружилось при поисках решений дифференциальных уравнений. Практика заставила переключиться на поиски приближенных решений, которые с ее точки зрения ничуть не хуже точных.

Появилось множество приближенных методов, а вместе с ними возникла проблема оценки качества полученных приближенных решений. Собственно, для получения количественной оценки следует сделать две вещи. Во-первых, нужно ввести расстояние между любыми двумя элементами пространства, в котором мы ищем решение, причем так, чтобы достаточно малое расстояние ρ между приближенным и точным решениями гарантировало удовлетворительное качество приближенного решения. Во-вторых, нужно уметь оценить это расстояние, не зная, разумеется, точного решения.

Напомним, что не всякую функцию $\rho(a, b)$ принято считать расстоянием между элементами a и b данного пространства P . Необходимо, чтобы расстояние от любого элемента a до него самого равнялось нулю и чтобы расстояние от a до b было положительным и равнялось расстоянию от b до a . Однако это еще не все. Должно выполняться

«неравенство треугольника»:

$$\rho(a, b) \leq \rho(a, c) + \rho(c, b).$$

Смысл последнего условия можно проиллюстрировать следующим образом: если по дороге из a в b нужно заехать в c , то это не может укоротить минимально необходимый путь.

После введения расстояния $\rho(a, b)$ пространство «повышается в звании»: оно становится метрическим пространством. Заметим, что все гильбертовы пространства являются метрическими.

§ 19.6. Рассказ о принципе Дирихле

Этот короткий рассказ можно начать с задачи о брахистохроне. Ее автором является Яков Бернулли, а решил ее, согласно математическому фольклору, сам Ньютон, отвлекшись на один вечер от повседневных забот директора монетного двора. В задаче требуется найти форму кривой скорейшего спуска в вертикальной плоскости, предполагая, что по этой кривой скользит без трения тяжелая точка. Метод, которым воспользовался Ньютон, оказался применимым к обширному кругу задач и положил начало вариационному исчислению и теории оптимального управления. Для нас, однако, важно, что Ньютон свел задачу о брахистохроне к решению некоторого дифференциального уравнения. Возникла ситуация, которую можно описать следующим образом. Были обнаружены задачи об оптимальном выборе функции, эквивалентные задачам о решении системы дифференциальных уравнений. Если основным объектом исследования являются дифференциальные уравнения (или их системы), то полезно помнить, что может существовать эквивалентная оптимизационная задача. Так, Лагранж показал, что в отсутствие трения все уравнения механики можно свести к одному типу оптимизационных задач. Это открытие получило название принципа наименьшего действия. Впоследствии данный принцип был распространен на уравнения Максвелла и на многие другие разделы физики. Таким образом, мы столкнулись с еще одним классом «двуликих» задач.

Французский математик Дирихле предложил применять связь между задачами оптимизации и дифференциальными уравнениями для доказательства существования решений у последних. «Принцип Дирихле» состоит в следующем: каждая оптимизационная задача имеет решение; это может быть использовано для доказательства существования решения у многих дифференциальных уравнений путем сведения их к оптимизационным задачам.

Конец триумфальному шествию принципа Дирихле положил немецкий математик Карл Вейерштрасс (1815–1877). Он построил простейший пример оптимизационной задачи, которая не имеет решения! Вот эта задача: найти гладкую кривую наименьшей длины,

проходящую через три заданные точки. Напомним, что кривая называется гладкой, если она имеет касательную в каждой точке. Совершенно ясно, что если отбросить требование гладкости, то решением задачи будет ломаная, состоящая из двух отрезков. Среди гладких кривых есть сколь угодно близкие по длине к ломаной. Поэтому какую бы гладкую кривую мы ни выбрали, найдется более короткая гладкая кривая, проходящая через заданные три точки. Последнее и означает, что искомой в задаче кривой в природе нет!

Однако это не конец истории. Гораздо позже немецкий математик Давид Гильберт показал, как выделить более узкий класс оптимизационных задач, всегда имеющих решение! Тем самым он реабилитировал принцип Дирихле. Более того, он успешно применял данный принцип в своей работе. В дальнейшем мы остановимся на этом вопросе более подробно.

§ 19.7. Как поймать льва в пустыне?

Математический фольклор знает несколько способов поимки льва в пустыне. Нам понадобится только один из них, удачно пародирующий вполне серьезный метод поиска решений задачи и одновременного доказательства их существования. Решим вынесенную в заголовок задачу. Поделим пустыню на две примерно равные по площади части. По крайней мере в одной из них окажется лев. Исключим из рассмотрения другую часть, а оставшуюся опять поделим на две. После достаточного числа таких делений лев окажется на площадке, лишь немногим превышающей его размеры, и его можно будет взять голыми руками.

Мы уже пользовались описанным методом для поиска нуля непрерывной функции внутри интервала, на концах которого она принимает значения разных знаков (см. §7.2). Рассмотрим более сложный случай. Пусть C — некоторый замкнутый несамопересекающийся контур в комплексной плоскости, а $f(z)$ — какая-либо функция, аналитичная на этом контуре и во всех точках внутри него. Необходимо узнать, есть ли внутри контура C нули данной функции и если есть, то сколько. Теория аналитических функций дает ответ на оба вопроса. Вот предлагаемое ею решение. Заставим точку z обегать контур C против часовой стрелки. При этом вектор $w = f(z)$ опишет в комплексной плоскости некоторую замкнутую кривую, сделав n ($n = 0, 1, 2, \dots$) полных оборотов. Искомое число нулей и равно числу n оборотов вектора w . Для более точного определения положения рассматриваемых нулей вспомним о льве и разделим на две части область, ограниченную контуром C . Вторично применяя описанный прием, найдем число нулей в каждой из них.

Поймаем «еще одного льва». Напомним, что точкой сгущения множества M называют всякую точку t , обладающую следующим свойством: любой содержащий ее интервал содержит бесконечное множество точек из M . Пусть M — произвольное множество точек, расположенных на некотором отрезке. Докажем справедливость утверждения: если множество M является бесконечным, то оно имеет по крайней мере одну точку сгущения.

Приступим к «охоте». Поделим исходный отрезок пополам. По крайней мере в одной из половинок окажется бесконечное множество точек из M . Назовем эту половинку отрезком номер один. Разделив ее на две части, найдем отрезок номер два, который тоже содержит бесконечное множество точек из M . Продолжая процесс, получим бесконечное множество вложенных друг в друга отрезков, каждый из которых содержит бесконечное подмножество множества M . Их левые концы образуют монотонную последовательность точек. Эта последовательность имеет предел. Он-то и является искомой точкой сгущения. Один из «львов» пойман!

§ 19.8. Принцип неподвижной точки

Прежде всего поясним, о какой неподвижной точке будет идти речь, и свяжем данное понятие с теоремами существования. Это позволит нам сделать краткий обзор еще одной группы доказательств подобных теорем.

Начнем издалека. Пусть на некотором множестве элементов M определен оператор A , отображающий это множество само на себя. Действие такого оператора можно представлять себе как перемещение всех элементов M , не выводящее ни одного из них за его пределы. Случается, что при таком перемещении некоторый элемент x остается на месте. Естественно назвать его неподвижной точкой оператора A . У оператора может быть несколько неподвижных точек. Согласно введенному определению

$$A(x) = x. \quad (19.1)$$

Последнее соотношение можно рассматривать как уравнение относительно x . Мы видим, что две разные задачи — решение уравнения и поиск неподвижной точки оператора — по сути дела являются двумя «лика́ми» одной и той же задачи. Всякое абстрактное уравнение $F(x) = 0$, где F — некоторый оператор, принимает вид уравнения (19.1), если положить

$$A(x) = F(x) + x$$

(и если определена операция сложения элементов множества M).

В дальнейшем будут рассмотрены некоторые простые признаки существования неподвижных точек (см. § 19.10). Они относятся к тем

случаям, когда на множестве M определена метрика, т. е. введено расстояние $\rho(x, y)$ между любой парой его точек, а оператор A является сжимающим. Смысл последнего условия таков: каковы бы ни были две точки x_1 и x_2 из M , под действием оператора A они сближаются:

$$\rho(Ax_1, Ax_2) < \rho(x_1, x_2).$$

Здесь же мы остановимся на случаях, когда оператор A не является сжимающим, а неподвижная точка тем не менее существует. Следовательно, нужны методы обнаружения неподвижных точек, не использующие понятие расстояния.

Теории, не использующие понятие расстояния, называются топологическими. Задачи, рассматриваемые этими теориями, и теоремы, устанавливаемые в них, тоже принято называть топологическими. Вот пример простой топологической задачи. Можно ли соединить девятью непересекающимися дорогами каждый из трех хуторов с магазином, почтой и ж/д станцией?

Знаменитой топологической задачей является задача о четырех красках. Она формулируется очень просто. Превратим произвольный квадрат в географическую карту воображаемой планеты. (Мы имеем в виду политическую карту, никаких гор или рек, только границы государств.) Карту нужно раскрасить так, чтобы никакие два государства, имеющие общий участок границы, не были окрашены в один и тот же цвет. Опыт картографов показывает, что для этой цели всегда достаточно иметь под рукой четыре краски, а тремя зачастую обойтись нельзя. Возникла математическая задача: нарисовать карту, требующую для раскраски пять или более красок, или доказать, что четырех красок достаточно для всех мыслимых карт.

В конце двадцатого столетия эта задача была решена. Было доказано, что четырех красок достаточно!

Вернемся к теоремам существования. В двадцатых годах прошлого века польский математик Шаудер исследовал возможности использования топологических теорий для доказательств теорем существования. Следует сказать, что к тому времени уже был известен отнюдь не тривиальный результат Брауэра (1888–1966) о неподвижных точках отображений в евклидовых пространствах. Простейшая его формулировка такова. *Каждое непрерывное отображения n -мерного шара на него самого имеет по крайней мере одну неподвижную точку.*

Приведем еще два впечатляющих результата, вытекающих из теории Брауэра и получаемых топологическими методами.

1. Любое непрерывное отображение сферы на самое себя либо содержит неподвижную точку, либо переводит каждую точку сферы в диаметрально противоположную.

2. На поверхности сферы нельзя построить два семейства непрерывных кривых так, чтобы через любую ее точку проходила ровно

одна кривая из каждого из этих семейств. Заметим, что пример меридианов и параллелей не опровергает данное утверждение, поскольку через северный и южный полюса проходит бесконечное множество меридианов.

По сути дела, программа Шаудера состояла в таком обобщении теории Брауэра, которое позволило бы находить неподвижные «точки» (в данном случае — функции) среди бесконечномерных пространств функций. Работы И. Шаудера (1930), Ж. Лере (1934) и А. Н. Тихонова (1935) содержат много важных результатов; они создали новый мощный метод доказательств теорем существования.

Сделаем одно замечание. Выше мы упоминали непрерывные кривые и непрерывные отображения. Читатель может думать, что всякое утверждение, содержащее упоминание о непрерывности, автоматически предполагает введение расстояния. Действительно, понятие непрерывности связано с понятием предела, а предел последовательности — это тот элемент пространства, расстояния до которого от остальных элементов последовательности стремятся к нулю. Однако одно из достижений топологии — введение понятия предела, не использующего понятия расстояния.

§ 19.9. Ограниченные операторы

Введение расстояния позволяет выделить отдельные важные классы операторов. Пусть X — некоторое метрическое пространство. Выделим из совокупности операторов, переводящих его элементы в элементы этого же пространства, три класса операторов. Прежде всего введем класс ограниченных снизу операторов. Оператор A называется *ограниченным снизу* числом $m > 0$, если для любой пары элементов x_1 и x_2 из X выполняется неравенство

$$\rho(Ax_1, Ax_2) \geq m\rho(x_1x_2).$$

Аналогично будем говорить, что оператор A ограничен сверху числом M , если выполняется условие

$$\rho(Ax_1, Ax_2) \leq M\rho(x_1x_2).$$

Назовем ограниченный сверху оператор A сжимающим, если число M в последнем условии меньше единицы.

Докажем две теоремы, касающиеся уравнения

$$Ax = f,$$

где x — искомый элемент из X ; f — заданный элемент. Каждая из них доказывается в одну строчку.

Теорема 1. Если оператор A ограничен снизу, то рассматриваемое уравнение не имеет двух разных решений x_1 и x_2 .

Доказательство. Имеем

$$m\rho(x_1, x_2) \leq \rho(Ax_1, Ax_2) = \rho(f, f) = 0.$$

Значит $\rho(x_1, x_2) = 0$ и теорема доказана.

Теорема 2. Если оператор A ограничен снизу, x_0 — одно из точных решений уравнения, а x — произвольный элемент пространства, то имеет место оценка

$$\rho(x, x_0) \leq m^{-1}\rho(Ax, Ax_0) = m^{-1}\rho(Ax, f).$$

Это очень полезная теорема. Выбирая в качестве x найденное тем или иным способом приближенное решение, мы можем оценить расстояние от него до любого точного решения x_0 . Заметим, что $\rho(Ax, f)$ называется *невязкой приближенного решения*, а $\rho(x, x_0)$ — *погрешностью решения* x . Вычисление невязки, в отличие от погрешности, не составляет труда; ее всегда можно считать известной. Последняя теорема позволяет оценить погрешность по невязке, если оператор ограничен сверху.

§ 19.10. Сжимающие операторы

Обратимся к сжимающим операторам. Отдельное рассмотрение именно этого класса оправдывает следующая теорема.

Теорема 3. Если уравнение

$$x = Ax \tag{19.2}$$

имеет решение x , то оно может быть вычислено со сколь угодно малой погрешностью при условии, что оператор A является сжимающим.

Опишем приближенный метод решения уравнения (19.2), называемый итерационным. Возьмем произвольный элемент x_1 и построим итерационную последовательность:

$$x_2 = Ax_1, \quad x_3 = Ax_2, \quad \dots, \quad x_{n+1} = Ax_n, \quad \dots,$$

рассчитывая на то, что при достаточно большом числе итераций мы получим удовлетворительное приближение. Будет ли это справедливо в действительности, зависит, разумеется, от свойств оператора. Если оператор A является сжимающим, то можно записать

$$\rho(x_{m+1}, x) = \rho(Ax_m, Ax) \leq k\rho(x_m, x),$$

откуда

$$\rho(x_m, x) < k^{m-1}\rho(x_1, x).$$

Последнее неравенство и доказывает теорему.

Покажем, как данная теорема используется в теории возмущений. Вот одна из типичных задач этой теории. Нужно решить или исследовать уравнение вида

$$(B + C)x = f,$$

где B и C — заданные операторы; f — заданная функция; x — неизвестная функция. Предполагается, что так называемое невозмущенное уравнение

$$Bx = f \tag{19.3}$$

имеет единственное решение и что мы умеем находить его при любом выборе функции f . Последнее условие означает существование и известность оператора B^{-1} , обратного B .

Попытаемся решить рассматриваемую задачу следующим образом. Подействуем на обе части уравнения (19.2) обратным оператором B^{-1} . Получим

$$x + B^{-1}Cx = B^{-1}f.$$

Или, окончательно,

$$x = Ax \quad (Ax = -B^{-1}Cx + B^{-1}f). \tag{19.4}$$

Сделанное преобразование показывает, что уравнение (19.2) имеет решение, если оператор $B^{-1}C$ — сжимающий. Это решение является единственным и может быть найдено с помощью итерационного метода.

Хочется обратить внимание читателя на прозрачность приведенного рассуждения и подчеркнуть, что ее причина — абстрактное изложение метода.

Глава 20. СПЕКТРЫ В ФИЗИКЕ И МАТЕМАТИКЕ

В настоящей главе мы познакомим читателя с отдельным типом задач, о котором раньше не упоминалось.

§ 20.1. Немного истории

Впервые спектр как физическое понятие возник в оптике. Пропуская белый свет через треугольную прозрачную призму, каждый может заметить, что луч не только отклоняется, но и расщепляется на несколько лучей разного цвета. Естественный вывод таков. Луч белого света представляет собой смесь многих цветных лучей, а причиной расщепления является различие их коэффициентов преломления. Как выяснилось, свет, излучаемый различными источниками, расщепляется по-разному. Особенно интересной оказалась картина расщепления света, излучаемого чистыми химическими элементами. Вместо непрерывной радуги, порождаемой белым светом, наблюдаются несколько изолированных линий, каждая своего цвета. Этот набор цветов (или, если хотите, линий) получил название спектра (латинское слово «*spectrum*» означает «видимое»; в современном русском языке это слово имеет совсем другой смысл, вспомните, например, выражение «широкий спектр»).

Разным элементам соответствуют разные спектры. Спектр света, излучаемого химическими соединениями, оказался наложением (суммой) спектров входящих в них элементов. Это позволяет определять химический состав образца неизвестной породы, изучая его спектр. Так в химии появился принципиально новый метод анализа — спектральный анализ.

Подобное состояние дел вполне устраивало химиков, однако было совершенно неудовлетворительным с точки зрения физиков. Именно они должны были объяснить, как перейти от измерения спектров к их расчету. Физики понимали, что в этом вопросе у них нет точки опоры — приемлемой теории или хотя бы рабочей модели. Планетарная модель атома не годилась. Конечно, она предсказывала излучение света вращающимися электронами, но именно это предсказание демонстрировало отсутствие у физиков сколько-нибудь удовлетворительного представления об атоме. Планетарная модель приводила к выводу о чрезвычайно короткой жизни атома и о непрерывности его спектра. И то, и другое вопиющим образом противоречило эксперименту и здравому смыслу.

Любопытно, что оба эти противоречия устранялись одновременно и, так сказать, шаг за шагом, хотя шаги были гигантскими. Требовалось изменить представление и о свете, и об атоме, притом так, чтобы связать спектр излучаемых атомом частот с процессами, которые могут в нем происходить. Связующей цепью должен был служить закон сохранения энергии.

Первые шаги, сделанные в данном направлении, преследовали совершенно другие цели. Макс Планк выдвинул гипотезу: свет излучается дискретными дозами, или порциями. Он назвал их квантами. Энергия одного кванта пропорциональна его частоте; коэффициент пропорциональности получил название постоянной Планка и удостоился собственного обозначения: его всегда обозначают символом \hbar . Излучение представляется в виде большого числа элементарных актов. Каждый акт — излучение одного фотона, реже — двух фотонов. Последнее происходит, когда излучение одного фотона нарушает закон сохранения импульса. Закон сохранения энергии требует, чтобы излучающий атом терял энергию, равную излученной. В этом месте дорога обрывается, поскольку неизвестно, какие процессы могут происходить в атоме. Нужно начинать с другого конца — с атома.

Французский физик Луи де Бройль предположил, что с каждой материальной частицей связана некоторая волна, длина которой обратно пропорциональна ее импульсу. Коэффициентом пропорциональности является постоянная Планка. Де Бройлю удалось извлечь из этого предположения некоторые следствия, нашедшие экспериментальное подтверждение. Волны де Бройля получили признание и послужили ступенькой на пути к квантовой теории атома.

Мощным прорывом в разгадке спектров излучения явилась теория Нильса Бора. Она впервые позволила рассчитать спектр атома водорода и была с энтузиазмом принята как первая удовлетворительная феноменологическая теория. Основой теории Бора служат два постулата.

1. Электрон в атоме может двигаться только по тем орбитам, длина которых кратна длине его дебройлевской волны.

2. При таком движении электрон не излучает.

По Бору, таким образом, существует только дискретное множество разрешенных орбит и их можно перенумеровать. Применение к этим гипотетическим орбитам законов классической механики позволило определить энергии движущихся по ним электронов. Разности последних очень точно соответствовали энергиям излучаемых квантов! Это был потрясающий успех теории. Читатель узнает о том удивительном периоде становления современной физики гораздо больше, обратясь ко второй части книги.

Теория Бора не могла считаться окончательной, законченной теорией излучения хотя бы из-за того, что она не объясняла, почему не

действует вытекающий из электродинамики значительно более мощный механизм излучения. Заметим, что у созданной ранее теории относительности не было никаких расхождений с электродинамикой.

Поиски теории атома не прекращались. В 1926 г. такая теория была создана сразу в нескольких эквивалентных вариантах. Ее называли квантовой механикой. Сейчас эту теорию чаще называют квантовой физикой. Значительная доля второй части настоящей книги посвящена квантовой физике. Здесь мы ограничимся только двумя замечаниями.

Согласно квантовой физике понятия координаты частицы и ее траектории являются приближенными; они размываются по мере уменьшения массы частицы, о которой идет речь, и совершенно непригодны для описания состояния электрона внутри атома. Состояние электрона описывается так называемой волновой функцией (см. вторую часть книги). Поэтому подсчет интенсивности излучения, исходящий из представления об электронных орбитах, должен быть отвергнут с порога как базирующийся на неправильных представлениях об атоме. Что же остается от теории Бора? Остается главное: представление о дискретном наборе возможных состояний электрона в атоме; при этом в каждом из возможных состояний электрон имеет вполне определенную энергию и, следовательно, не излучает. Процесс излучения есть переход электрона из одного допустимого состояния в другое, с меньшей энергией; освободившуюся энергию уносит квант света; частота последнего определяется законом Планка.

Как искать разрешенные уровни энергии? Для ответа на этот вопрос необходимо ввести новое понятие — *собственное число оператора*. Ему и посвящена настоящая глава.

Временно оставим атом и его излучение и обратимся к классической механике. В механике спектр имеет очень простой физический смысл: это набор собственных частот механической системы, совершающей малые колебания вблизи положения устойчивого равновесия. Подобные колебания передаются окружающей среде (воздуху); если их частота лежит между шестнадцатью и сорока тысячами колебаний в секунду, то они воспринимаются нами как звук. Для того чтобы заранее рассчитать частоты звука, порожденного натянутой струной или колебаниями воздушной массы внутри духового инструмента, нужно решить задачу о малых колебаниях этих механических систем. Аналогия с излучением атома очевидна. В терминах акустики атом — это миниатюрный «музыкальный» инструмент, только его излучение воспринимается не ухом, а глазом и порождает он живопись, а не музыку.

В математике задача о спектре сначала возникла как неотъемлемая часть метода Фурье, но затем заинтересовала математиков сама по себе, а также в связи с акустикой, электродинамикой и квантовой физикой.

§ 20.2. Спектр оператора и физические спектры

Современное понятие спектра оператора возникло в результате длительного эволюционного процесса. Оно относится только к линейным операторам, то есть операторам, действующим в том или ином линейном пространстве и обладающим следующими двумя свойствами:

$$A(x_1 + x_2) = Ax_1 + Ax_2, \quad A(\lambda x) = \lambda Ax,$$

где λ — произвольное число; x_1 и x_2 — любые элементы пространства.

Простейшим примером элемента спектра является собственное число.

Определение. Число μ называется собственным числом оператора A , если в пространстве L , где действует этот оператор, существует отличный от нуля вектор x , который под действием оператора A умножается на число μ :

$$Ax = \mu x.$$

Элемент x называется собственным вектором оператора A , принадлежащим собственному числу μ .

Какие еще числа, кроме собственных, входят в спектр оператора? Оказывается, число ν не входит в состав спектра оператора A и называется обыкновенным, если оператор $A - \nu E$ (где E — единичный оператор) имеет обратный оператор $(A - \nu E)^{-1}$ и этот последний является ограниченным. Число ν принадлежит непрерывному спектру, если оператор $A - \nu E$ имеет обратный оператор, но последний не является ограниченным. Поясним сказанное на примерах. Рассмотрим три разные оператора, связанные с одним и тем же формальным оператором. Оператор называют формальным, если его понимают в максимально широком смысле, то есть без каких-либо дополнительных ограничений области определения. В качестве примера формального оператора можно привести оператор D^2 двукратного дифференцирования, определяемый равенством $D^2 f(x) = d^2 f / dx^2$.

Чем плохи формальные операторы? Как правило, они не годятся для вычисления физических спектров и бесполезны для метода Фурье, о котором речь пойдет ниже. Причина может показаться парадоксальной: у них слишком много собственных чисел и собственных векторов. Так, любое вещественное или комплексное число является собственным для оператора D^2 . Это следует, например, из тождества

$$D^2 \cos lx = -l^2 \cos lx,$$

где l , а значит, и l^2 — любое комплексное число. Перейдем от формального оператора D^2 к «хорошо определенному». Для этого сузим область определения D^2 , ограничив ее функциями, рассматриваемыми на сегменте $[0, 1]$ и принимающими на его концах нулевые значения.

Набор всех собственных чисел и собственных функций данного оператора будет следующим:

$$\lambda_n = -\pi^2 n^2, \quad \varphi_n(x) = \sin \pi n x, \quad n = 1, 2, \dots$$

Других точек спектра у рассматриваемого оператора нет.

Какова же связь между спектром оператора и различного рода физическими спектрами? Любой физический спектр совпадает со спектром некоторого линейного оператора. Одна из задач физики — в каждом случае найти соответствующий оператор.

Рассмотрим, например, малые колебания механической системы с конечным числом степеней свободы. Замечательной особенностью подобной системы является то, что она может совершать так называемые собственные колебания. Это движения, при которых зависимость координат всех точек системы от времени выражается гармоническим законом:

$$Q_j = q_j \cos \omega t + s_j \sin \omega t \alpha_j, \quad j = 1, 2, \dots, n,$$

с одной и той же частотой. Подставив последние выражения в уравнения механики, получим систему линейных уравнений относительно амплитуд q_j и s_j . Ее коэффициенты линейно зависят от частоты ω . Схематически эту систему уравнений можно записать в виде

$$AQ = \omega Q,$$

где A некоторый линейный оператор, а Q — $(2n)$ -мерный вектор:

$$Q = (q_1, q_2, \dots, q_n; s_1, s_2, \dots, s_n).$$

Если частоту выбрать наугад, то практически с единичной вероятностью полученная система уравнений будет иметь только тривиальное решение — вектор амплитуд равен нулю. Собственное колебание с произвольно выбранной частотой невозможно. У возможных собственных колебаний квадраты частот совпадают с собственными числами оператора A .

Укажем на основное свойство спектра оператора, действующего в каком-либо m -мерном пространстве L . Такой оператор имеет от одного до m собственных чисел. Данное утверждение напоминает основную теорему алгебры: каждое алгебраическое уравнение степени m имеет от одного до m корней. Последнее не случайно; именно эта теорема используется для его доказательства. Дело в том, что поиск собственного числа оператора сразу же приводит к уравнению степени m , корнем которого и является искомое число. Это уравнение называется характеристическим. Число различных собственных чисел равно числу различных корней характеристического уравнения. Если это число равно m , то оператор имеет m собственных чисел и столько же собственных векторов. Последние образуют базис в L .

Противоположный случай, когда число корней характеристического уравнения меньше m , значительно сложнее. При этом можно сказать только, что число линейно независимых собственных векторов не больше m .

§ 20.3. Эрмитовы операторы

В дальнейшем мы будем говорить только о так называемых *эрмитовых операторах*. Это поразительный класс операторов. Их теория сравнительно проста, но тем не менее изобилует неожиданными фактами.

Эрмитовы операторы встречаются на каждом шагу как в физике, так и в математике. Постараемся дать хотя бы неполное представление о роли, которую они играют в этих науках. Возможно, читатель более внимательно отнесется к определению данного понятия.

Вот некоторые обстоятельства, характеризующие роль эрмитовых операторов.

1. В отсутствие сил трения квадраты собственных частот механических систем являются собственными числами эрмитовых операторов.

2. В квантовой механике каждой физической величине соответствует эрмитов оператор; ее возможные значения принадлежат спектру этого оператора.

3. Уравнения теплопроводности, электростатики, уравнения Максвелла и уравнения Шредингера содержат эрмитовы операторы.

4. Спектральная теория эрмитовых операторов в гильбертовых пространствах на несколько десятков лет старше спектральной теории неэрмитовых операторов. Она проще, а используемый ею далеко не простой математический аппарат все же не столь сложен, как аппарат спектральной теории неэрмитовых операторов.

Читатель, вероятно, будет удивлен, узнав, что столь фундаментальное понятие определяется коротким равенством и несколькими словами.

О п р е д е л е н и е. *Оператор A называется эрмитовым, если выполняется условие*

$$(u, Av) = (Au, v),$$

каковы бы ни были элементы u и v гильбертова или евклидова пространства. Можно сказать и так: оператор является эрмитовым, если в скалярном произведении его можно перебрасывать от одного сомножителя к другому.

Познакомимся с простыми свойствами эрмитового оператора. Из его определения следует, что квадратичная форма (Ax, x) всегда вещественна. В самом деле, согласно одному из свойств скалярного произведения величины (Ax, x) и (x, Ax) комплексно сопряжены. С другой стороны, в силу эрмитовости оператора A они равны друг

другу. Значит, они вещественны. Обратимся к собственным числам оператора A . Умножая обе части равенства

$$Ax = \lambda x \quad (x \neq 0)$$

скалярно на x , получаем

$$(Ax, x) = \lambda(x, x).$$

Отсюда следует, что все собственные числа λ всех эрмитовых операторов всегда вещественны.

Займемся собственными векторами эрмитовых операторов. Пусть x и y — два собственные вектора, причем соответствующие им собственные числа не равны друг другу. Оказывается, и это легко доказать, что векторы x и y взаимно ортогональны:

$$(x, y) = 0.$$

И последнее. Если x — какой-либо собственный вектор оператора A , то совокупность всех векторов, ортогональных x , образует подпространство, которое инвариантно относительно оператора A . Подробней это означает, что если какой-нибудь вектор y ортогонален собственному вектору x , то и вектор Ay ортогонален x . Указанное свойство позволяет при необходимости рассматривать оператор A только на подпространстве, ортогональном данному собственному вектору. Мы перечислили простые, почти очевидные свойства эрмитовых операторов.

Укажем три главные проблемы теории эрмитовых операторов.

1. Имеет ли данный эрмитовый оператор хотя бы один собственный вектор?

2. Существует ли хотя бы один не равный нулю вектор, который ортогонален всем собственным векторам данного эрмитова оператора?

3. Существует ли у данного эрмитова оператора спектральное разложение?

Смысл последнего вопроса будет разъяснен чуть ниже.

Если эрмитов оператор действует в евклидовом пространстве, то ответить на три поставленные вопроса нетрудно. Мы это сделаем в следующем параграфе.

Ответ на эти же вопросы применительно к некоторым более широким классам операторов был получен только в первой половине двадцатого века. С тех пор спектральная теория операторов продолжает развиваться, охватывая все новые их классы. Мы попытаемся рассказать об успехах спектральной теории, сделанных ею в первой четверти прошлого столетия.

§ 20.4. Эрмитовы операторы в конечномерном пространстве

Начнем с эрмитового оператора A , который действует в конечномерном пространстве. Если это пространство n -мерно, то, оказывается, оно содержит ортогональный базис

$$e_1, e_2, \dots, e_n,$$

состоящий исключительно из собственных векторов эрмитова оператора A . Разумеется, в подобном случае не существует ни одного не равного нулю вектора, ортогонального всем собственным векторам. Заметим еще, что количество различных собственных чисел может быть меньше n , если несколько собственных векторов соответствуют одному и тому же собственному числу.

Из собственных векторов можно составить базис пространства E ; любой вектор u можно разложить по этому базису, то есть представить его в виде суммы

$$u = c_1 e_1 + c_2 e_2 + \dots + c_n e_n. \quad (20.1)$$

Обратим внимание на первое слагаемое правой части; оно представляет собой проекцию вектора u на вектор e_1 . В самом деле, оно параллельно вектору e_1 , а разность $u - c_1 e_1$, очевидно, ортогональна ему. Подобным же образом все остальные слагаемые в разложении (20.1) являются проекциями того же вектора на соответствующие собственные векторы. Базисные векторы удобно нормировать так, чтобы скалярный квадрат каждого равнялся единице. В этом случае коэффициенты последнего разложения легко вычисляются: умножая скалярно обе его части на e_k , получим

$$c_k = (u, e_k). \quad (20.2)$$

Полезно ввести в рассмотрение операторы P_1, P_2, \dots, P_n , положив

$$P_j u = (u, e_j) e_j, \quad j = 1, 2, \dots, n. \quad (20.3)$$

Эти операторы называются проекционными. С их помощью можно переписать разложение (20.1) следующим образом:

$$u = P_1 u + P_2 u + \dots + P_n u.$$

Последнее равенство показывает, что сумма введенных проекционных операторов равна единичному оператору E :

$$E = P_1 + P_2 + \dots + P_n.$$

Это соотношение называют разложением единицы.

Обратимся к оператору A . Применяя его к обеим частям равенства (20.3), получим

$$AP_j = \lambda_j P_j,$$

где $\lambda_1, \lambda_2, \dots, \lambda_n$ — соответствующие собственные числа оператора A . Отсюда вытекает соотношение

$$A = \lambda_1 P_1 + \lambda_2 P_2 + \dots + \lambda_n P_n.$$

Оно называется *спектральным разложением оператора A* . Смысл последнего соотношения таков. Всякий эрмитов оператор в конечномерном пространстве является суммой проекционных операторов, умноженных на его собственные числа. В векторной форме это же соотношение можно записать как

$$Au = c_1 \lambda_1 e_1 + c_2 \lambda_2 e_2 + \dots + c_n \lambda_n e. \quad (20.4)$$

Столь же просто вычисляются все степени оператора A :

$$A^k u = c_1 \lambda_1^k e_1 + c_2 \lambda_2^k e_2 + \dots + c_n \lambda_n^k e. \quad (20.5)$$

Равенство (20.5) можно рассматривать как спектральное разложение оператора A^k .

Спектральное разложение позволяет распространить всякую числовую функцию $f(x)$ на совокупность эрмитовых операторов, положив по определению

$$f(A) = f(\lambda_1) P_1 + f(\lambda_2) P_2 + \dots + f(\lambda_n) P_n.$$

При этом, разумеется, необходимо, чтобы все собственные числа оператора A лежали в области определения функции $f(x)$.

Естественно, что спектральная теория не могла ограничиться операторами, действующими в конечномерных пространствах. Первая экспансия в область бесконечномерных пространств произошла в теории интегральных уравнений. Следующий впечатляющий шаг в развитии спектральной теории был сделан при замене интегральных операторов вполне непрерывными. Эта глава спектральной теории операторов заслуживает особого внимания; она запечатлела победу аксиоматического подхода и методов функционального анализа. Посвятим ей несколько следующих параграфов.

Обращение к читателю

Пять последних параграфов могут оказаться трудными для читателя. Однако они включены в книгу. Дело в том, что именно эти параграфы показывают функциональный анализ в действии, притом в обстановке, приближенной к реальной исторической. Более того,

сама методика перехода от конкретной задачи к обобщающей ее абстрактной, демонстрируемая в этих параграфах, характерна для многих других приложений функционального анализа. Показано также, что решение одной абстрактной задачи является решением целого ряда конкретных задач, на первый взгляд лишь отдаленно напоминающих исходную.

§ 20.5. Интегральные операторы Фредгольма в абстрактной форме

В этом и двух следующих параграфах мы надеемся продемонстрировать своеобразие и мощь методов функционального анализа, получив в абстрактной форме теорию интегральных уравнений Фредгольма.

Первые серьезные шаги теории операторов в гильбертовых пространствах тесно или, скорее, неразрывно связаны с теорией интегральных уравнений вида

$$x(t) = \int K(t, s)x(s) ds + f(t), \quad (20.6)$$

где $x(t)$ — искомая функция; $K(t, s)$ — известная непрерывная функция двух переменных; $f(t)$ — известная функция; интеграл берется по интервалу $(0, 1)$. Ядро $K(t, s)$ предполагается вещественным и симметричным:

$$K(t, s) = K(s, t). \quad (20.7)$$

Очень поучительно рассмотреть, как в абстрактном гильбертовом пространстве H строится теория уравнения

$$x = Kx + f, \quad (20.8)$$

включающая в себя теорию уравнений (20.6). Этот вопрос стоит того, чтобы остановиться на нем подробнее.

В настоящем параграфе мы ограничимся тем, что выделим класс абстрактных уравнений (20.8), который обладает всеми существенными свойствами уравнения (20.6). Теория таких уравнений будет построена «пунктирно» в последующих параграфах. Первый шаг прост: под $x(t)$ и $f(t)$ будем подразумевать элементы конкретного гильбертова пространства $L_2(0, 1)$ со скалярным произведением

$$(x, y) = \int x(t)y(t) dt$$

и введем оператор K , положив

$$(Kx)(t) = \int K(t, s)x(s) ds. \quad (20.9)$$

Уравнение (20.6) примет вид

$$x(t) = (Kx)(t) + f(t)$$

и станет похожим на уравнение (20.8),

Тут мы сталкиваемся с первой трудностью: как записать в абстрактной форме условие (20.7)? Очевидно, следует придать ему форму, поддающуюся обобщению. Поиски такой формы привели к следующему условию: равенство

$$(u, Kv) = (Ku, v)$$

должно выполняться, каковы бы ни были функции $u(t)$ и $v(t)$. Мы уже знаем, что операторы, удовлетворяющие этому условию, называются *эрмитовыми*. Итак, при переходе к абстрактному уравнению условие (20.7) заменяется условием эрмитовости оператора K . Само понятие эрмитовости неразрывно связано с тем, что уравнение (20.6) трактуется как уравнение в гильбертовом пространстве.

Осталось перевести на абстрактный язык условие непрерывности ядра $K(t, s)$ и то обстоятельство, что интегрирование в (20.6) производится по интервалу конечной длины. Как показал анализ, оба эти обстоятельства используются в теории интегральных уравнений *только* для доказательства следующей леммы.

Лемма. Интегральный оператор

$$(Kx)(t) = \int K(t, s)x(s) ds \quad (20.10)$$

переводит всякое ограниченное по норме подмножество пространства $L_2(0, 1)$ в компактное подмножество (см. § 15.3).

Оказалось целесообразным ввести специальный термин, который выделяет все операторы, обладающие этим свойством, даже если они неэрмитовы. Их называют *вполне непрерывными операторами*.

Будем считать оператор K вполне непрерывным. Этим завершается переход от конкретного интегрального уравнения (20.6) к абстрактному уравнению (20.8), в котором x и f — элементы абстрактного гильбертова пространства, а K — вполне непрерывный эрмитов оператор.

Хочется обратить внимание читателя на сжатость описания задачи (20.8). В нем нет ничего лишнего. Исследуя полученное абстрактное уравнение, мы даже не обязаны знать, что такое интеграл. Ничто не отвлекает нашего внимания от сути дела.

Второе важное обстоятельство — полученное абстрактное уравнение имеет множество других конкретных реализаций помимо интегрального уравнения (20.1). В частности, можно перейти к интегральным уравнениям, в которых интеграл берется по интервалу бесконечной длины, если одновременно потребовать, чтобы ядро $K(x, y)$ достаточно быстро убывало с ростом x и y . Нужно только проверить,

остался ли вполне непрерывным полученный при такой замене интегральный оператор. Можно перейти к интегральным уравнениям, в которых искомой является функция на поверхности или внутри некоторого объема. Вместо одного интегрального уравнения можно взять систему таких уравнений и т. д. Это показывает, как аксиоматический метод осуществляет объединение «под одной крышей» различных, казалось бы непохожих друг на друга задач. Приступим к знакомству со спектральной теорией вполне непрерывных эрмитовых операторов.

§ 20.6. Спектр вполне непрерывного эрмитового оператора не пуст

Постараемся показать, что любой вполне непрерывный эрмитовый оператор имеет хотя бы один собственный вектор. Исходной идеей является принцип Дирихле. В нашем случае он выглядит так.

Введем в рассмотрение A -образ какого-либо шарового слоя, т. е. совокупность M всех точек вида Ax , $r \leq |x| \leq R$, где r и R — два произвольные, но фиксированные положительные числа. Множество M (A -образ) является компактным, поскольку шаровой слой представляет собой ограниченное множество, а оператор A вполне непрерывен. Рассмотрим на множестве M вещественнозначную функцию

$$F(x) = \frac{(Ax, x)}{(x, x)}.$$

Нетрудно показать, что она непрерывна и, как всякая непрерывная функция, рассматриваемая на компактном множестве, достигает своего наибольшего значения по крайней мере в одной из точек x_0 этого множества:

$$F(x_0) \geq F(x), \quad x \in M. \quad (20.11)$$

Если векторы Ax_0 и x_0 параллельны, т. е.

$$Ax_0 = \lambda x_0, \quad (20.12)$$

то x_0 — собственный вектор и теорема верна. Предположим противное. Тогда в двумерной плоскости, натянутой на пару векторов Ax_0 и x_0 , имеется не равный нулю вектор y , ортогональный вектору x_0 , такой что

$$Ax_0 = \mu x_0 + y, \quad (x_0, y) = 0$$

(собственно, μx_0 есть проекция вектора Ax_0 на вектор x_0). Скалярное произведение $(Ax_0, y) = (y, y)$, разумеется, положительно:

$$(Ax_0, y) = (Ay, x_0) > 0. \quad (20.13)$$

Возьмем в (20.11) в качестве x вектор $x = x_0 + \varepsilon y$. Получим

$$\frac{(Ax_0, x_0)}{(x_0, x_0)} \geq \frac{(Ax_0 + \varepsilon Ay, x_0 + \varepsilon y)}{(x_0 + \varepsilon y, x_0 + \varepsilon y)},$$

или

$$(Ax_0, x_0)(x_0 + \varepsilon y, x_0 + \varepsilon y) \geq (Ax_0 + \varepsilon Ay, x_0 + \varepsilon y)(x_0, x_0).$$

После простых преобразований приходим к неравенству

$$\varepsilon(Ax_0, x_0)(y, y) \geq [(Ax_0, y) + (Ay, x_0) + \varepsilon(Ay, y)](x_0, x_0).$$

При достаточно малом значении произвольного параметра ε последнее неравенство противоречит (20.13); следовательно, векторы Ax_0 и x_0 параллельны.

Утверждение доказано. В процессе доказательства были использованы такие понятия функционального анализа, как компактное множество, вполне непрерывный оператор и теорема о непрерывной функции, заданной на компактном множестве.

Полученная теорема существования замечательна еще и тем, что она не содержит какого-либо малого параметра или параметра, ограничивающего оператор A . При доказательстве решающую роль сыграл принцип Дирихле, с помощью которого задачу о собственном числе удалось свести к задаче о наибольшем значении функции. Последняя оказалась «легкой добычей» мощных методов функционального анализа.

§ 20.7. Спектр вполне непрерывного эрмитового оператора

Продолжим изучение спектра вполне непрерывного эрмитового оператора. Чтобы не делать постоянных оговорок, исключим из рассмотрения почти тривиальный случай, когда область значений оператора представляет собой конечномерное подпространство (в этом случае оператор называется вырожденным).

Оказывается, каждый невырожденный оператор имеет счетное множество взаимно ортогональных собственных векторов и соответствующих им собственных чисел:

$$e_1, e_2, \dots, e_n, \dots, \quad (20.14)$$

$$\lambda_1, \lambda_2, \dots, \lambda_n, \dots \quad (20.15)$$

Собственные числа выписаны в порядке убывания их абсолютных величин; собственные векторы, принадлежащие собственному числу нуль, невозможно включить в этот список — до них «не дойдет очередь».

В настоящем параграфе будет показано, что всякий вектор x , ортогональный всем собственным векторам (20.14) оператора A , если такой имеется, переводится этим оператором в нуль:

$$Ax = 0.$$

Поясним, как найти последовательность собственных векторов с перечисленными свойствами. Сначала находим вектор e_1 , как это было сделано в § 20.6. Затем рассматриваем подпространство H_1 , ортогональное e_1 . Оператор A не выводит векторы этого подпространства за его пределы. Следовательно, можно рассматривать A как оператор, отображающий H_1 в H_1 . При этом он остается вполне непрерывным эрмитовым оператором. Пользуясь уже доказанной теоремой, заключаем, что и в H_1 существует собственный вектор e_2 . Он ортогонален e_1 , а его собственное число λ_2 не превосходит по абсолютному значению λ_1 . Повторяя счетное число раз описанный процесс, получим последовательность взаимно ортогональных собственных векторов, причем абсолютные значения соответствующих собственных чисел образуют монотонно невозрастающую последовательность.

Разумеется, мы не «нашли» эти векторы. Собственно, мы и не ставили перед собой такой задачи. Была решена задача более высокого уровня. Установлено существование искомым векторов и обнаружены некоторые их свойства. Это позволяет строить алгоритмы, использующие собственные функции, в том числе алгоритм метода Фурье. По отношению к этим алгоритмам фактическое нахождение собственных векторов представляет собой задачу более низкого уровня.

Закончим доказательство теоремы. Что можно сказать о векторах, которые ортогональны всем векторам (20.14)? Если такие векторы существуют, то они, очевидно, образуют подпространство. Обозначим его символом H_0 . Данное подпространство инвариантно относительно оператора A и, следовательно, содержит по крайней мере один его собственный вектор e_0 . Соответствующее собственное число равно нулю, поскольку в противном случае вектор e_0 вошел бы под некоторым номером в число векторов (20.14) и был бы сам себе ортогонален. Разумеется, если число $\lambda = 0$ не является собственным числом оператора A , то подпространство H_0 пусто. Ограничимся случаем, когда подпространство H_0 конечномерно. При этом в нем есть ортогональный базис

$$e_{-1}, e_{-2}, \dots, e_{-k},$$

состоящий из собственных векторов оператора A ; все соответствующие им собственные числа равны нулю.

Для полного описания спектра вполне непрерывного эрмитового оператора осталось упомянуть еще одно его свойство. Последовательность (20.15) его собственных чисел стремится к нулю.

Для доказательства этого утверждения рассмотрим последовательность (20.14) собственных векторов, предполагая, что каждый из них

нормирован на единицу:

$$(e_k, e_k) = 1, \quad k = 1, 2, \dots,$$

так что последовательность (20.14) оказывается ограниченной. Оператор A переводит ее в последовательность

$$\lambda_1 e_1, \lambda_2 e_2, \dots, \lambda_n e_n, \dots \quad (20.16)$$

Если предположить, что все собственные числа по абсолютной величине больше некоторого положительного числа a :

$$|\lambda_k| > a, \quad k = 1, 2, \quad (20.17)$$

то расстояние между любой парой элементов этой последовательности

$$\rho(\lambda_m e_m, \lambda_n e_n) = \sqrt{(\lambda_m^2 + \lambda_n^2)} > a.$$

Последовательность (20.16) не содержит ни одной сходящейся подпоследовательности. Мы получили противоречие, поскольку оператор A , будучи вполне непрерывным, переводит всякое ограниченное множество в множество, содержащее сходящуюся последовательность. Следовательно, предположение (20.17) неверно, а последовательность (20.15) стремится к нулю.

§ 21.1. Теория Фурье

Основная идея теории Фурье очень проста. Объясним ее в двух словах. Пусть требуется решить уравнение вида

$$Ax - \lambda x = f, \quad (21.1)$$

где x — неизвестная функция; A — некоторый линейный оператор; f — заданная функция; λ — число. Уравнение (21.1) легко решается, если f совпадает с одним из собственных векторов оператора A :

$$f = f_n, \quad Af_n = \lambda_n f_n.$$

В таком случае решение записывается просто:

$$x = \frac{f_n}{\lambda_n - \lambda}$$

($\lambda \neq \lambda_n$). Будет ли это решение единственным? Поставим вопрос шире. Может ли уравнение (21.1) иметь больше одного решения? Пусть x и y — два решения одного и того же уравнения (21.1). Ясно, что их разность, $z = x - y$, является решением однородного уравнения

$$Az - \lambda z = 0.$$

Последнее равенство возможно только в двух случаях: либо $z = 0$, либо λ — собственное значение оператора A . В последнем случае z представляет собой соответствующий собственный вектор оператора A . Итак, если λ не является собственным числом оператора A , то уравнение (21.1) не может иметь двух разных решений. В противном случае оно имеет сколько угодно решений (их разность — собственный вектор, соответствующий числу λ).

Рассматриваемое уравнение решается столь же просто, если его правая часть представляет собой сумму двух или нескольких собственных векторов оператора A :

$$f = c_1 f_1 + c_2 f_2 + \dots + c_n f_n. \quad (21.2)$$

Соответствующее решение дается формулой

$$x = c_1 x_1 + c_2 x_2 + \dots + c_n x_n,$$

где

$$x_k = (\lambda_k - \lambda)^{-1} f_k, \quad k = 1, 2, \dots, n.$$

К сожалению далеко не всегда (скорее, в виде исключения) правую часть уравнения (21.1) можно представить в виде конечной комбинации нескольких собственных векторов оператора A . В общем случае требуется бесконечный набор собственных векторов с тем, чтобы представить функцию f в виде (21.2), заменив конечную сумму бесконечным рядом. При этом и решение принимает вид бесконечного ряда:

$$x = c_1 x_1 + c_2 x_2 + \dots + c_n x_n + \dots$$

Возникает несколько вопросов. Первый (практический) вопрос состоит в следующем. Предположим, что f можно разложить в ряд по собственным векторам. Как найти коэффициенты этого ряда? Второй, тоже практический, вопрос: сколько членов ряда надо учесть, чтобы получить удовлетворительное приближенное решение? И, наконец, два теоретических вопроса: можно ли в принципе представить правую часть уравнения в виде искомого ряда и будет ли сходиться ряд, дающий формальное решение задачи?

В следующем параграфе мы дадим ответы не все четыре вопроса в предположении, что оператор A является эрмитовым и вполне непрерывным. Тем самым мы представим читателю строгую теорию метода Фурье для данного класса операторов.

Обратим внимание на своеобразную зависимость решения задачи от параметра λ . Когда этот параметр приближается к одному из собственных значений λ_n , одно из слагаемых, а именно x_n , неограниченно возрастает. Это широко известное явление резонанса. Как быть, если параметр λ в точности совпадает с одним из собственных чисел λ_k ? Легко дать исчерпывающий ответ на этот вопрос. Если функция f , стоящая в правой части уравнения (21.2), не ортогональна собственной функции f_k , то уравнение не имеет ни одного решения. В подобном случае следует присмотреться к постановке прикладной задачи, которая привела к уравнению (21.2). В противоположном случае уравнение (21.2) разрешимо, но имеет бесконечно много решений. Очевидно, и тогда необходимо присмотреться к его выводу.

На очереди стоят дифференциальные операторы. С самого начала было ясно, что они представляют собой «твердый орешек». Во-первых, дифференциальные операторы определены не на всех функциях, а только на дифференцируемых (достаточное число раз). Во-вторых, ни один формальный дифференциальный оператор не является эрмитовым. Для того чтобы быть эрмитовым, дифференциальный оператор должен рассматриваться вместе с некоторыми граничными условиями. Однако не при всяком выборе этих условий он оказывается эрмитовым. Наконец, у дифференциальных операторов возможен непрерывный спектр. Из-за отсутствия собственных векторов, соответствующих

непрерывной части спектра, построение необходимых проекционных операторов превращается в проблему.

Здесь следует остановиться и сделать небольшое отступление. Существует важное различие в употреблении терминов «собственное число» и «собственный вектор» применительно к непрерывному спектру. Рассмотрим простой пример на вычисление производной:

$$-i \frac{d}{dx} e^{ipx} = p e^{ipx}, \quad -\infty < x < \infty.$$

Данное соотношение показывает, что функция e^{ipx} представляет собой собственный вектор оператора $-i d/dx$; соответствующее собственное число равно p . Однако математику трудно этим воспользоваться, поскольку функция e^{ipx} не является квадратично-интегрируемой; ее норма в пространстве $L_2(-\infty, \infty)$ равна бесконечности. Перейти в иное пространство? Но тогда исчезает понятие эрмитовости. Математическая спектральная теория пошла по другому пути. Его начало связано с тем обстоятельством, что операция усреднения по любому интервалу длины Δ возвращает функцию e^{ipx} в лоно гильбертова пространства.

Посмотрим на то же соотношение глазами физика. Оператор $-i\hbar d/dx$ представляет собой оператор импульса; e^{ipx}/\hbar — плоская волна, описывающая состояние частицы с импульсом p . Как и положено, значение импульса оказывается собственным значением его оператора. Рассматриваемое соотношение является одним из основных соотношений квантовой физики. Совершенно немислимо игнорировать его. Целесообразней игнорировать тот факт, что плоские волны лишены обязательного свойства быть квадратично-интегрируемыми, и рассматривать его как результат некоторой идеализации. Последняя, как и всякая идеализация, требует при исследованиях определенной осторожности.

В соответствии с духом времени создаваемая теория должна включать спектральную теорию дифференциальных операторов, но не сводиться к ней. Создание такой теории было в основном завершено к концу первой половины двадцатого века. Здесь мы не будем пытаться рассказать о ней.

§ 21.2. Обобщенные ряды Фурье

Классическая теория рядов Фурье состоит из двух частей. Первая из них использует так называемую квадратичную метрику. Во второй полученные результаты переносятся на некоторые другие метрики. В настоящем параграфе мы познакомимся с абстрактной теорией, обобщающей первую часть классической теории рядов Фурье.

Исходными понятиями являются абстрактное гильбертово пространство, действующий в этом пространстве вполне непрерывный эрмитов оператор A , а также его собственные векторы и соответствующие им собственные числа, введенные в предыдущем параграфе. Для простоты ограничимся случаем, когда нуль не является собственным числом оператора A . В этом случае не существует отличного от нуля вектора, ортогонального всем собственным векторам e_k оператора A .

Введем понятия *коэффициентов Фурье* и *ряда Фурье*. Пусть f — произвольный элемент гильбертова пространства H . Числа

$$c_k = (f, e_k) \quad (21.3)$$

называются коэффициентами Фурье вектора f , а формальный ряд

$$c_1 e_1 + c_2 e_2 + \dots + c_n e_n + \dots \quad (21.4)$$

носит название ряда Фурье этого вектора. Здесь и в дальнейшем мы предполагаем, что собственные векторы нормированы условием $(e_k, e_k) = 1$. Усеченные суммы ряда Фурье, имеющие вид

$$S_n = c_1 e_1 + c_2 e_2 + \dots + c_n e_n,$$

обладают некоторыми полезными очевидными свойствами, а именно:

$$\begin{aligned} (S_n, S_n) &= |c_1|^2 + |c_2|^2 + \dots + |c_n|^2, \\ (f - S_n, e_k) &= 0, \quad k = 1, 2, \dots, n, \end{aligned}$$

и, как следствие,

$$(f - S_n, S_n) = 0.$$

Применяя теорему Пифагора к взаимно ортогональным векторам $f - S_n$ и S_n :

$$(f, f) = (f - S_n, f - S_n) + (S_n, S_n) \geq (S_n, S_n),$$

получим так называемое неравенство Бесселя:

$$|c_1|^2 + |c_2|^2 + \dots + |c_n|^2 \leq (f, f).$$

Оно демонстрирует, что числовой ряд

$$|c_1|^2 + |c_2|^2 + \dots + |c_n|^2 + \dots \quad (21.5)$$

сходится и что усеченные суммы

$$S_1, S_2, \dots$$

образуют последовательность Коши. В самом деле, справедливо равенство

$$|S_{m+n} - S_m|^2 = |c_{m+1}|^2 + \dots + |c_{m+n}|^2.$$

Оно показывает, что сходимость числового ряда (21.5) влечет за собой сходимость ряда Фурье (21.4) в метрике H . Поэтому сумму S ряда

Фурье можно почленно скалярно умножать на любой вектор. В частности, получаем

$$(S, e_k) = c_k.$$

Отсюда заключаем, что разность $S - f$ ортогональна всем собственным векторам и, следовательно, равна нулю. Мы доказали основную теорему абстрактной теории рядов Фурье: *каждый вектор гильбертова пространства равен сумме своего ряда Фурье.*

С помощью этой теоремы легко получить спектральное разложение любого вполне непрерывного эрмитова оператора A и обосновать метод Фурье решения уравнений типа

$$Ax - \lambda x = f \quad (21.6)$$

(см. § 20).

Согласно основной теореме векторы x и Ax равны своим рядам Фурье:

$$\begin{aligned} x &= c_1 e_1 + c_2 e_2 + \dots + c_k e_k + \dots, \\ Ax &= C_1 e_1 + C_2 e_2 + \dots + C_k e_k + \dots \end{aligned} \quad (21.7)$$

Коэффициенты C_k и c_k связаны соотношением

$$C_k = \lambda_k c_k. \quad (21.8)$$

Оно является следствием эрмитовости оператора A :

$$C_k = (Ax, e_k) = (x, Ae_k) = \lambda_k (x, e_k) = \lambda_k c_k.$$

Теперь можно записать

$$Ax = \sum \lambda_k c_k e_k.$$

Заметим, что вектор $c_k e_k = (x, e_k) e_k$ является проекцией вектора x на e_k . Обозначая соответствующий оператор проектирования через P_k , перепишем последнее равенство следующим образом:

$$Ax = \sum \lambda_k P_k x,$$

или в операторной форме:

$$A = \sum \lambda_k P_k.$$

Это и есть спектральное разложение оператора A .

Обратимся к уравнению (21.6). Скалярно умножим обе его части на e_k . Учитывая соотношения (21.7), получаем $C_k - \lambda c_k = f_k$ и

$$c_k = \frac{f_k}{\lambda_k - \lambda}.$$

Мы нашли все коэффициенты ряда Фурье решения уравнения (21.6) и можем записать само решение:

$$x = \sum \frac{e_k f_k}{\lambda_k - \lambda}.$$

Осталось оценить погрешность при замене данного ряда его усеченной суммой. Последнее сделать нетрудно, и мы не будем на этом останавливаться.

На очереди неограниченные эрмитовы операторы. Однако пора ставить точку...

Часть II

АБСТРАКЦИЯ В ФИЗИКЕ

Понятие плоской волны, подобно многим физическим понятиям, есть не больше как абстракция, которую мы можем осуществить лишь с известной степенью точности. Тем не менее это полезное понятие.
А. Эйнштейн, «Эволюция физики»

Введение

Вторая часть книги целиком посвящена физике.

Достижения современной физики огромны. Они дают возможность непротиворечиво описать свойства элементарных частиц и твердых тел, плазмы и нейтронных звезд, сверхпроводников и солнечного вещества. Появилась надежда построить сценарий развития Вселенной от Большого взрыва до наших дней. Хотя познанное пространство огромно, нас не покидает ощущение, что и теперь мы стоим на берегу океана незнания. Или, в лучшем случае, чуть переступили через береговую черту, если сравнивать наше время с теми далекими временами, когда Ньютон использовал этот образ. Впереди — новые неожиданные открытия. То, что движение от незнания к знанию, похоже, бесконечно, не приводит в уныние, а воодушевляет. Этому способствует преемственность: новое знание не отменяет старое. Новые открытия не требуют зачеркнуть созданную картину Мира, а расширяют ее и совершенствуют.

Можно представить себе огромное строительство. Вырастают новые этажи. Одновременно на всех этажах строящегося здания идет работа. Улучшают отдельные детали, иногда перестраивают один-два этажа.

Тех, кто внимательно следит за строительством, особенно интересует, что происходит на уровне фундамента. Они испытывают волнение, когда выясняется, что фундамент требует усовершенствования. Особенно остро воспринимается необходимость создания нового, более глубокого этажа. Проходит некоторое время, у фундамента новый этаж появляется, волновавшиеся привыкают. На верхних этажах продолжается совершенствование постройки. Многие строители попросту не заметили изменений, происшедших с фундаментом. Конструкция

такова, что строительство одного этажа почти не зависит от происходящего на других.

Никто не знает плана постройки. Никто не может предсказать, какой этаж потребует в ближайшее время концентрации усилий. Строительство идет, повинувшись скрытому от строителей плану. Когда вспоминают прошедшие этапы, возникает впечатление, что удалось уловить черты плана, по которому веками осуществлялось строительство. Но попытка руководствоваться старым планом для продолжения работ, как правило, ни к чему хорошему не приводит. При, казалось бы, суматошном, бесплановом строительстве постройка продолжается, здание не только растет, но и укрепляется, отдельные его этажи совершенствуются, обретая черты завершенности.

В сравнении истории физики со строительством есть некое лукавство. Собственно говоря, что строят физики? Мир, который они изучают, существует вне зависимости от того, изучают его или нет. Мы в этом уверены. Но. . . Давайте мир, который мы непосредственно ощущаем, которым часто любимемся или который мысленно проклинаям, когда он оборачивается своей неприглядной стороной, сравним с тем миром, который различается за строгими законами, формулируемыми на страницах учебника физики, который описывается изящными уравнениями, допускающими не менее изящные решения. Думаю, возникал у Вас крамольный вопрос: «Какое отношение к реальному миру имеет мир физики, впрочем, и любой другой науки?»

Строгость, порядок, красоту привнесли в Мир художники и ученые-естествоиспытатели. Правда, привнесли — неточное слово. Обнаружили, обратили внимание, подчеркнули, показали. И именно тем привнесли.

Превращение беспорядочного, неорганизованного, чувственного мира в гармоничный мир, упорядоченный пониманием, — заслуга науки. Среди средств, используемых для гармонизации, важное место занимает *абстракция*. Хочется верить, что нам удастся показать это на нескольких примерах из разных областей физики.

Глава 1. КЛАССИЧЕСКАЯ МЕХАНИКА

Создание почти любого понятия требует абстрагирования. В этом легко убедиться на примере образования даже такого обиходного понятия, как стол. Любая наука, в том числе физика, оперирует бесконечным количеством абстрактных понятий. Процесс создания понятий в науке, ее научных терминов отличается от процесса создания понятий в языке, предназначенном для обычного общения, только большей осознанностью.

Мы не предполагаем заниматься этимологией тех или иных научных терминов, хотя это очень интересная тема. Интересовать нас будет использование в физике абстрактных методов и приемов, получение абстрактных результатов, приносящих вполне конкретную пользу.

Напомним: написать настоящую книгу нас побудило то, что эпитет «абстрактный» часто воспринимается как несущий негативный оттенок. Именно против такого восприятия абстракции мы выступаем.

Начнем с основы основ — с классической механики, с механики Ньютона.

Если попросить привести примеры абстрактных утверждений или построений, то мало кто приведет законы Ньютона. Законы Ньютона, служащие основой классической механики, воспринимаются, скорее, как пример конкретного знания. Все хорошо знают, или думают, что знают, что такое масса, ускорение, сила. Известно, что уравнение

$$\text{масса} \times \text{ускорение} = \text{сила}$$

достойно всякого уважения, поскольку оно является основой наших представлений о движении. Под его юрисдикцию попадает любое движение: от движения небесных тел до движения разнообразных деталей технических устройств. Классическая механика воспринимается не просто как конкретная наука, а как образец конкретной науки.

Важным аргументом в пользу конкретности классической механики служит убеждение, что она соответствует нашему житейскому опыту. Мне кажется, что в данном контексте эпитет «конкретный» воспринимается нами как «очевидный», а «абстрактный» — как «неочевидный», и даже «плохо доступный пониманию».

Имеем ли мы право считать классическую механику конкретной наукой? Если да, то всегда ли она воспринималась как конкретная, доступная наука? Заведомо, так было не всегда. Когда законы механики формулировались (конец XVII века), они казались совершенно

абстрактными. Во-первых, создание классической, ньютоновской механики потребовало создания новой математики — дифференциального и интегрального исчислений. Кроме того, новые законы противоречили житейскому опыту. Ведь все знали и видели на каждом шагу, что силе, приложенной к телу, пропорциональна скорость тела, а отнюдь не ускорение (!). Существовала проверенная веками практики механика Аристотеля (384–322 гг. до нашей эры), узаконившая утверждение о пропорциональности между силой и скоростью.

Между прочим, и в наши дни человек, далекий от физики, скорее согласится с положениями механики Аристотеля, чем Ньютона. Отнюдь не все понимают существо дела. Согласовать механику Ньютона с житейским опытом, естественно, можно, но для этого необходимо учесть силу трения. Коротко говоря, для многих классическая механика и сейчас — весьма абстрактная наука.

Стоит задуматься. Что мы понимаем под житейским опытом? У большинства ощущение житейского опыта основано отнюдь не на собственном опыте. Мало кто совершал кругосветное путешествие. Но все знают, что Земля — шар. То, что классическая механика соответствует житейскому опыту, мы знаем из школьной программы по физике. Не нужно думать, что приобретенный в школе опыт неглубок. Думаю, он не менее глубок чем истинный собственный опыт, а возможно, и глубже. Я могу ошибиться, а учителя не ошибаются. За ними авторитет знания, накопленного человечеством.

Согласно Гейзенбергу, как мы уже цитировали, одна из форм абстракции заключается в рассмотрении «предмета или группы предметов под одним углом зрения, *отвлекаясь от всех других свойств*» (выделено нами). Приведенное выше основное уравнение классической механики именно в этом смысле представляет собой очевидный пример абстракции. «Угол зрения» — рассмотрение движения предметов. Все предметы, имеющие одинаковую массу, подчиняются одному и тому же уравнению (оно выписано). Этим они объединены, выделены из всех остальных предметов. Другие свойства тела, кроме его массы, роли не играют: ни его состав, ни цвет, ничто другое. А влияние на тело всего окружающего мира сведено к одному вектору — силе.

Надо сказать, что привычная абстракция редко воспринимается как абстракция.

Не все отчетливо представляют себе степень абстракции уравнения Ньютона. Надо помнить, что по траектории, определяемой с помощью решения этого уравнения, движется некая точка. Обычно ее называют центром масс тела. Иногда — центром тяжести. Есть строгое правило, по которому можно найти центр масс любого тела. Нередко встречается ситуация, когда он находится вне тела (см. ниже). Только представьте себе: по законам конкретной классической, ньютоновской механики движется, оказывается, некая абстрактная точка, не всегда даже расположенная внутри тела.

Уравнение Ньютона в написанной выше форме недостаточно конкретно до тех пор, пока не сказано, что представляет собой сила, действующая на тело. Конкретизация возникает лишь тогда, когда записано выражение для силы. Однако практически невозможно учесть все силы, действующие на тело. Незначительные, слабо влияющие на движение тела силы просто опускают. Высоким стилем можно сказать: «Мы производим абстрагирование типа идеализации».

Если сила не изменяется при движении тела (частицы), то говорят, что движение происходит под действием постоянной внешней силы. И это практически всегда — идеализация. Постоянной сила бывает в исключительных случаях. Например, при падении тела с небольшой высоты на Землю, или при ускорении заряженной частицы под действием постоянного электрического поля.

Используя уравнение Ньютона, можно изучить не только движение центра тяжести тела, но и вращение — перемещение тела вокруг центра тяжести. Вывод соответствующего уравнения движения не обходится без абстракции: как правило, движущееся тело предполагается *твердым*. В механике в результате абстрагирования понятие «твердое тело» приобретает строгий смысл: твердое тело представляет собой систему материальных точек, расстояние между которыми неизменно. Естественно, это условие выполняется лишь приближенно. Однако чаще всего изменение расстояний между точками твердого тела действительно очень мало, то есть тело прекрасно сохраняет свою форму и величину, и абстракция совершенно оправдана. Наверное, следует упомянуть, что стабильность формы и размеров деталей машин — одно из условий нормальной работы механизмов.

Давайте вернемся к движению материальных точек, немного задержимся и в виде примера более внимательно чем, казалось бы, положено для популярной статьи, рассмотрим конкретную задачу о движении двух частиц, между которыми действует сила, зависящая от расстояния между ними. Слово «частица», естественно, является условным наименованием. Как оказывается, результат применим, скажем, к движению планеты Меркурий вокруг Солнца или Луны вокруг Земли; кроме того, это могут быть реальные частицы (например, электрон и протон, если их движение можно было бы описывать с помощью классической механики).

Ясно, что нам надо записать два векторные уравнения. Каждое описывает движение одной из частиц. Поэтому величины, относящиеся к частицам, будем снабжать индексом 1 или 2. Векторы выделим жирным шрифтом. Итак:

$$m_1 \frac{d^2 \mathbf{r}_1}{dt^2} = \mathbf{F}_{12}(\rho), \quad m_2 \frac{d^2 \mathbf{r}_2}{dt^2} = \mathbf{F}_{21}(\rho), \quad \rho = |\rho| = |\mathbf{r}_1 - \mathbf{r}_2|.$$

Третий закон Ньютона утверждает, что действие равно противодействию, то есть $\mathbf{F}_{21} = -\mathbf{F}_{12}$. Следовательно, индексы при букве \mathbf{F}

вовсе можно опустить:

$$m_1 \frac{d^2 \mathbf{r}_1}{dt^2} = \mathbf{F}(\rho), \quad m_2 \frac{d^2 \mathbf{r}_2}{dt^2} = -\mathbf{F}(\rho), \quad \rho = \mathbf{r}_1 - \mathbf{r}_2.$$

Пара строк арифметики. Сложим два уравнения и вычтем одно из другого. В результате, введя новые обозначения, получим

$$\frac{d^2 \mathbf{R}}{dt^2} = 0, \quad \mathbf{R} = \frac{m_1 \mathbf{r}_1 + m_2 \mathbf{r}_2}{m_1 + m_2},$$

$$m_{\text{эф}} \frac{d^2 \rho}{dt^2} = \mathbf{F}(\rho), \quad \rho = \mathbf{r}_1 - \mathbf{r}_2, \quad m_{\text{эф}} = \frac{m_1 m_2}{m_1 + m_2}.$$

Первое из этих уравнений описывает свободное движение центра тяжести двух частиц (сила равна нулю). Выражение для вектора \mathbf{R} представляет собой определение центра тяжести двух частиц, легко обобщающееся на общий случай тела, состоящего из многих частиц.

Второе уравнение описывает движение частиц относительно друг друга. Однако выглядит оно довольно странно. Рассматриваемое уравнение описывает движение несуществующей частицы с массой, составленной из масс обеих частиц. То, что это ненастоящая масса, подчеркнуто индексом «эф» (сокращение от прилагательного *эффективная*). Если одна из частиц имеет массу, во много раз превосходящую массу второй частицы, то эффективная масса с большой точностью равна настоящей массе легкой частицы. Можно убедиться в том, что тяжелая частица, если она очень тяжела, практически покоится, а движется только легкая. Однако отнюдь не всегда так бывает.

Мы привели этот пример, чтобы показать: абстрактная ситуация часто и очень естественно возникает при решении сравнительно простых задач классической механики.

Все же нельзя не отметить, что большинство понятий, которые использует классическая механика, воспринимаются нами как естественные. Мы их называем наглядными (координата частицы, ее скорость, ускорение). Правда, механика не может обойтись без силы. Если речь идет о движении поезда или автомашины, то понять, а лучше сказать, почувствовать, что из себя представляет сила, нетрудно. Вместе с тем, сила всемирного тяготения при первом знакомстве с фактом ее существования должна вызвать удивление. Как, каким способом одно тело влияет на движение другого тела, находясь от него на большом расстоянии? Вы, конечно, знаете, что ответ на этот вопрос нашел Альберт Эйнштейн в начале XX-го века, почти через 300 лет после того, как Ньютон создал классическую механику, введя в нее абстрактное понятие — дальное действие — силу, действующую между телами и не требующую ни их непосредственного соприкосновения, ни передающей среды.

Теперь мы знаем: дальное действие в природе нет. Всегда существует переносчик взаимодействия. Между заряженными частицами

взаимодействие осуществляется электромагнитным полем, а в роли переносчика взаимодействия масс — их притяжения друг к другу — выступает сама метрика пространства, в котором движутся тела.

Математический аппарат теории гравитации сложнее математического аппарата классической механики. Но для тех, кто его усвоил, движение под действием гравитационных сил стало нагляднее. В каждой точке, где находится тело, есть нечто, благодаря чему скорость тела изменяется, а ускорение отлично от нуля; так, перемещаясь от точки к точке, тело движется.

С другой стороны, развитие математического аппарата классической механики привело к выводу о возможности формулировки ее основных положений в виде принципа наименьшего действия. Согласно последнему движение осуществляется так, что построенный некоторым образом функционал имеет минимум. Этот функционал и носит название действия. Действие — интеграл, зависящий от функций, определяющих движение.

Из принципа наименьшего действия, используя математические приемы, можно вывести локальные дифференциальные уравнения движения. Но все же в его существовании есть нечто загадочное. Будто тело заранее знает, как оно должно двигаться, чтобы некая величина была минимальной. Некоторые, впрочем, считают, что никакой загадки нет, а есть парадоксальная, неожиданная формулировка закона Ньютона.

Принцип наименьшего действия можно успешно обобщать и использовать не только в механике, но и в других физических теориях. Это делает его принципиально абстрактным. Предпринимаются попытки всю физику вывести из принципа наименьшего действия.

Задумывая статью, автор обычно планирует поделиться своими мыслями, своими знаниями. В некоторый момент выясняется, что ему нужно кое-что проверить, на кое-что сослаться. Приходится обращаться к книгам. Как легко убедится читатель, чаще, чем другими, я пользовался пятитомной Физической энциклопедией. И сейчас, заканчивая параграф о классической механике, решил прочесть статью «Механика» (см. Т. 3. С. 126). Статья относительно небольшая, занимает 4 столбца. Ниже приведен абзац, имеющий, как мне кажется, прямое отношение к теме нашего рассказа.

«При изучении движения материальных тел в механике вводят ряд абстрактных понятий, отражающих те или иные свойства реальных тел; ими являются:

- 1) материальная точка — объект пренебрежимо малых размеров, имеющий массу; это понятие применимо, когда тело движется поступательно или когда в изучаемом движении можно пренебречь вращением тела вокруг его центра масс;

- 2) абсолютно твердое тело — тело, расстояние между любыми двумя точками которого всегда остается неизменным;

3) сплошная изменяемая среда; это понятие применимо, когда при изучении изменяемой среды (деформируемого твердого тела, жидкости, газа) можно пренебречь молекулярной структурой среды».

Механика — один из наиболее разработанных разделов современной науки. Успехи и достижения ее необозримы. Перечисление, приведенное выше, убедительно показывает, что абстракция является важным методическим приемом, применяемым при изучении движения различных и вполне конкретных материальных тел окружающего нас мира.

Глава 2. ПРИЧИНА РЕВОЛЮЦИЙ

Нам предстоит познакомиться с элементами новой физики. Будем откровенны: не очень новой. Новой по сравнению с классической ньютоновской механикой. К концу XIX-го века физика, казалось, была построена на прочном фундаменте, «сложенном» из классической механики и электродинамики. Многие считали, что впереди лишь уточнение, детализация. Мало кто ожидал, что новый век принесет революционные преобразования не только в обществе, но и в науке. С другой стороны, если бы не было тех, кого не удовлетворяло положение дел в физике, то и революции произойти не могли бы. Что же волновало создателей новой физики? Почему они пытались нарушить спокойствие? Что заставило их искать новые законы?

Сейчас, когда мы знаем, к каким последствиям привело открытие новых законов Природы, как изменилась жизнь всех людей под влиянием научных достижений, приходят на ум слова «генеральный путь развития науки» или «потребности техники — тогда, когда исчерпали себя открытия предыдущей эпохи».

Готовясь писать эту главу, я открыл Собрания научных трудов Эйнштейна и Бора. Перечитываю много раз читанные статьи, пытаюсь найти *серьезные* основания, оправдывающие желание их авторов разрушить стройный порядок, воцарившийся в Природе в результате «законотворческой» деятельности Ньютона и Максвелла. Ну, не смог Планк описать излучение черного тела, основываясь на законах классической физики. Или не удастся понять, почему спектры имеют яркие линии вместо того, чтобы быть непрерывными, как того требуют законы электродинамики. Правда, требуют в том случае, если электроны, излучая, как им положено по законам механики и электродинамики, постепенно сваливаются на ядро.

Последнее обстоятельство, похоже, по-настоящему серьезно. Планетарную модель атома — непосредственное, вроде бы, следствие экспериментов Резерфорда — невозможно согласовать с электродинамикой Фарадея–Максвелла. Однако как от электродинамики отказаться? Ведь она удивительно результативна. Электродинамика не только показала, что нет отдельно электричества, магнетизма и оптики, что все три науки — части одной, всеобъемлющей, но и предсказала существование (впоследствии открытых!) радиоволн, заставив почувствовать дыхание будущего — Мира, связанного вне зависимости от расстояний

и границ. Стоит ли думать о таких мелочах, как устойчивость атомов? Еще не вполне доказано, что атомы — реальные объекты, а не абстрактные домыслы теоретиков, призванные помочь понять законы химии. Исходя из классической механики и электродинамики, построена картина Мира. Вот-вот удастся описать все наблюдаемые явления. А возможно, будут получены ответы и на философские вопросы: существует ли свобода воли, или это иллюзия и господствует фатализм; верующему, похоже, удастся понять замысел Творца и описать его уравнениями механики и электродинамики.

Саркастичность предыдущего абзаца является, конечно, только литературным приемом.

Фактически ситуация была следующей: лишь небольшое число ученых ощущало неблагополучие. Оно относилось к областям физики, очень далеким от потребностей практики. В последнем и состоит одна из причин того, что только выдающиеся умы почувствовали это неблагополучие. Кроме того, надо было прекрасно ощущать, сколь едина наука, сколь невозможны нарушения ее законов.

Все началось с попыток Макса Планка описать излучение абсолютно черного тела. Абсолютно черное тело — тело, которое не отражает, но излучает. На первый взгляд, такого не бывает. Оказывается, его можно создать искусственно. Для этого надо сделать небольшое отверстие в полости, окруженной веществом, температуру которого можно устанавливать по своему желанию. Так вот, само отверстие и есть абсолютно черное тело. Падающий на отверстие луч света имеет нулевую вероятность вырваться наружу, а имеющееся в полости излучение ее стенок может выйти наружу и выходит. Классическая физика предсказывает странный результат, или, точнее, ничего не может сказать, поскольку расчет приводит к бесконечной интенсивности излучения.

Попытки Планка описать излучение черного тела заставили его признать несостоятельность классической механики и предположить, что гармонический осциллятор может иметь не произвольную энергию, а только кратную частоте колебаний ω . Уровни энергии осциллятора пришлось записать следующим образом:

$$\varepsilon_n = n\hbar\omega,$$

где $n = 0, 1, 2, \dots$ — целые числа.

Именно в этой формуле впервые появилась новая мировая константа — *постоянная Планка*, — ознаменовавшая начало квантовой революции. Букву \hbar принято называть «*h перечеркнутое*»; это постоянная Планка, деленная на 2π : $\hbar = h/2\pi$, $h \approx 6,63 \cdot 10^{-27}$ эрг · с; $\hbar = 1,05 \cdot 10^{-27}$ эрг · с. В настоящее время постоянная Планка известна

с точностью до семи значащих цифр, если не точнее. От года к году ее значение уточняется.

По научным статьям редко можно установить, что двигало ученым в процессе его работы. Помочь могут высказывания его современников. Особенно в тех случаях, когда они выдают собственные устремления говорившего. Вот что написал о Максе Планке в связи с его 60-летним юбилеем младший современник мэтра Альберт Эйнштейн: «Храм науки — строение многосложное. Различны пребывающие в нем люди и приведшие их туда силы. Некоторые занимаются наукой с гордым чувством своего интеллектуального превосходства; для них наука является тем подходящим спортом, который должен им дать полноту жизни и удовлетворение честолюбия. Можно найти в храме и других: плоды своей мысли они приносят здесь в жертву только в утилитарных целях. Если бы посланный Богом ангел пришел в храм и изгнал из него тех, кто принадлежит к этим двум категориям, то храм катастрофически опустел бы. Все-таки кое-кто из людей как прошлого, так и нашего времени в нем бы остался. К числу таких людей принадлежит наш Планк, потому мы его любим». (Собрание научных трудов. Т. 4. С. 29. Статья называется «Мотивы научного исследования». Написана она в 1918-м году. Издательство «Наука», Москва, 1967).

Последние слова цитаты не оставляют сомнений в том, какой позиции придерживается сам Эйнштейн. Через 12 лет Эйнштейн почти дословно повторил высказанную мысль.

И еще: несомненно, «кое-кто из людей . . . нашего времени», оставшихся в храме после изгнания, — это истинные творцы новой физики. На следующей странице той же статьи четко сформулировано, что ими движет: «Человек стремится каким-то адекватным способом создать в себе простую и ясную картину Мира».

Теперь мы знаем, каковы последствия подобного стремления. До простой и ясной картины Мира по-прежнему далеко. Ответ на один вопрос порождает другие. Но несомненно, многое разъяснилось. Область познанного фантастически расширилась. Методы теоретической и экспериментальной физики усложнились. Последнее вполне оправдано. Ведь туда, где проходит граница между познанным и непознанным, трудно «добраться».

Однако в какие бы глубины материи ни погружался исследователь, из какого бесконечного далека он ни ловил бы сигналы, какие бы сумасшедшие теории ни строил, им движет абстрактный принцип: создать простую и ясную картину Мира. Не может быть простоты и ясности, если существуют противоречия.

Мы не имеем возможности перечислить все возникшие противоречия, поставившие под сомнение возможности классической физики и тем самым послужившие причиной революций.

Упомянем два вопроса (из многих), на которые не могла ответить классическая физика. Ответы на них потребовали поистине революционных преобразований, столь глубоких, что изменили научное мировоззрение. Наш выбор вопросов не случаен. В большой мере он продиктован содержанием последующих глав.

Итак.

1. Как совместить независимость скорости света от скорости источника с правилом сложения скоростей, принятым со времен Галилея?
2. Почему устойчивы атомы?

Ответ на первый вопрос потребовал пересмотреть такое фундаментальное понятие, как *одновременность*, отказаться от абсолютного времени и привел к построению специальной теории относительности (Альберт Эйнштейн, 1905 г.).

До сих пор, кажется, нет полной ясности, использовал ли Эйнштейн результат опыта Майкельсона, не обнаружившего зависимости скорости света от скорости источника, или знал, что так должно быть. В 4-м томе Собрания трудов А. Эйнштейна, на который я уже ссылался, есть изложение его выступления в 1931-м году «Памяти Альберта А. Майкельсона» (С. 149), где сказано: «Как известно, опыт [. . .] дал отрицательный результат. Именно это негативное открытие в значительной мере способствовало признанию правильности специальной теории относительности». Трудно не выразить удивления и восхищения по поводу последствий ответа на волновавший Эйнштейна вопрос. Напомним только, что одним из результатов специальной теории относительности было открытие формулы $E = mc^2$, которую вполне можно сделать символом XX-го века.

Ответ на второй вопрос (об устойчивости атомов) был получен после создания квантовой механики.

Планетарная модель атома «родилась» в 1911-м году, позволив объяснить опыты Резерфорда по рассеянию α -частиц веществом. С тех пор прошло около 20 лет до создания квантовой теории атома. По сути лишь небольшая группа физиков-теоретиков в эти годы понимала, сколь неблагоприятна ситуация, и пыталась выйти из тупика. О данном периоде написано и издано много книг. Авторы, как правило, сосредоточивают свое внимание на тех, кто ощущал неблагоприятие, поскольку именно они искали выход из существовавшей ситуации. А остальные? Продолжали заниматься «своим делом» — решать задачи, которые поддавались решению, исследовать то, что давало возможность разобраться в изучаемом явлении.

Нет никаких оснований осуждать тех, кто не занимался подготовкой революции. Отдавая должное «революционерам», хочу подчеркнуть роль консерваторов. Без них наука не могла и не может обойтись.

Они обеспечивают преемственность, сохранность полученных в предреволюционную эпоху результатов. Слово «консерватор» происходит от латинского слова *conservator* — охранитель.

Но, думаю, нельзя не подчеркнуть степень неблагополучия. Давайте вычислим, сколько мог «прожить» простейший атом — атом водорода, если бы в мире микрочастиц господствовали классические законы. Состав его известен: протон и электрон. Без движения такая система существовать не может: электрон попросту свалится на протон. Вращение должно компенсировать силу притяжения e^2/r^2 (где r — расстояние от электрона до протона). Поскольку ускорение равно v^2/r (где v — скорость движения по орбите), из равенства $mv^2/r = e^2/r^2$ следует, что $mv^2/2 = e^2/2r$, а энергия электрона — сумма потенциальной ($-e^2/r$) и кинетической ($mv^2/2$) энергий — равна

$$\varepsilon = \frac{mv^2}{2} + \left(-\frac{e^2}{r}\right) = -\frac{e^2}{2r}.$$

Выпишем также значение ускорения: $w = v^2/r = e^2/mr^2$.

Согласно законам классической электродинамики, если заряженное тело движется с ускорением, то оно излучает энергию. В единицу времени электрон излучает $2e^2w^2/3c^3$ эрг. Составим естественное уравнение — скорость изменения энергии ε равна интенсивности излучения:

$$\frac{d\varepsilon}{dt} = -\frac{2e^2w^2}{3c^3}.$$

Воспользовавшись значениями энергии ε и ускорения w , имеем

$$\frac{dr}{dt} = -\frac{4e^4}{3c^3m^2r^2}.$$

Считая, что начальное расстояние от электрона до протона равно a , находим время жизни атома τ — время падения электрона на протон:

$$\tau = \frac{4a^3m^2c^3}{e^4}.$$

Для оценки удобно рассчитать число оборотов N , которое совершит классический электрон прежде, чем упадет на ядро:

$$N = \frac{\tau v}{a} = 4 \left(\frac{c}{v}\right)^3,$$

где $v = (e^2/ma)^{1/2}$.

Примерно 10^8 раз электрон успеет обернуться вокруг протона. Как ни странно, число оборотов велико потому, что скорость электрона мала по сравнению со скоростью света. Подставив значения входящих в формулу величин ($a = 0,5 \cdot 10^{-8}$ см, $m = 10^{-27}$ г, $e = 4,8 \cdot 10^{-10}$ CGSE; в начале XX-го века они уже были известны, во всяком случае по

порядку величины), мы убедимся, что время жизни τ приблизительно равно нескольким единицам, умноженным на 10^{-8} с. Надо сказать, что с точки зрения атомных масштабов это время не так уж и мало (число-то оборотов велико). Но, конечно, совершенно невозможно понять, почему атом «бессмертен», а ведь никто никогда не наблюдал спонтанную гибель атомов.

Мы специально провели весь расчет, чтобы продемонстрировать: он предельно прост, ошибки быть не может, а противоречие столь существенно, что от него нельзя отмахнуться.

Квантовая механика — один из впечатляющих примеров использования абстракции. Не зная ее хотя бы поверхностно, очень трудно понять примеры использования абстракции в физике: квантовая механика — основа современной физики.

Выше мы признались: задача настоящей книги — апология абстракции. Часто абстракцией именуют всю физику атомного и субатомного мира, аргументируя последнее тем, что свойства атомных и субатомных частиц очень непохожи на свойства привычных макроскопических предметов (это утверждал Гейзенберг). Ранее были перечислены виды абстракции. Легко убедиться, что необычность свойств (их непривычность) — не предлог для отнесения целого раздела знаний к абстракции. Чтобы избавиться от ощущения дискомфорта при пользовании непривычными понятиями и от комплекса неполноценности, с неизбежностью возникающего при этом, есть только один способ: надо непривычные понятия сделать привычными, понять их суть.

Изучить квантовую механику можно, скажем, в рамках университетского курса. Однако понимание приходит и при чтении научно-популярной литературы. Нам, естественно, придется ограничиться таким уровнем. Мы попытаемся описать кардинальное отличие квантовой механики от классической. Отбор материала зависит от воли автора и определяется его вкусом. Но не только, так как нам придется обходиться почти без математики, а это совсем непросто даже при том, что наша главная задача — физическая суть.

В предыдущей главе рассказано о необходимости ревизии классической физики. Те из ученых, кто понял неспособность классической физики описать микромир, мучительно пытались найти новую основу для «адекватного способа создать в себе простую и ясную картину Мира» (как сказал Эйнштейн).

Квантовая механика — результат коллективного творчества. Мы не можем рассказать о вкладе каждого из ее творцов. Приводить перечень фамилий, не объясняя, кто что сделал, бессмысленно, поэтому фамилии будут упоминаться только «по ходу сюжета». В процессе изложения мы ограничимся только нерелятивистской квантовой механикой. Ей «подвластны» движения микрочастиц со скоростями, малыми по сравнению со скоростью света. Подобное ограничение возникает не потому, что релятивистской квантовой механики не существует. Нами руководит мудрое высказывание Козьмы Пруткова: «Нельзя объять необъятное». Ограничение необходимо.

В классической физике сложно подобрать два объекта, столь разительно отличающиеся друг от друга, как материальная точка и волна. Оба их нетрудно себе представить. Это позволяет избежать определений. Материальная точка ассоциируется с частицей, корпускулой, а волна — с волной произвольной природы (с волной на водной поверхности, с волной звука, света). Любая волна имеет обязательные характеристики: частоту ω и волновой вектор \mathbf{k} . Мы будем пользоваться так называемой круговой частотой: $\omega = 2\pi\nu$, где ν — число колебаний в единицу времени, $\nu = 1/T$, T — период колебаний. Длина волнового вектора \mathbf{k} есть $k = 2\pi/\lambda$, где λ — длина волны; направление вектора \mathbf{k} совпадает с направлением распространения волны ¹⁾.

Оказывается, электроны и другие микроскопические частицы проявляют не только корпускулярные, но и волновые свойства, а волны любой природы (даже световые) — корпускулярные. С необходимостью квантовая механика должна уметь описывать, казалось бы, несочетаемое — корпускулярные и волновые свойства одного и того же объекта, например электрона. Как ни удивительно, она сумела преодолеть эту трудность.

В 1905-м году Альберт Эйнштейн высказал идею о двойственной природе света, ввел понятие *квант света*, а в 1923-м году французский физик-теоретик Луи де-Бройль распространил корпускулярно-волновой дуализм и на движение частиц вещества, сопоставив каждой движущейся частице *волновой процесс*. Используя постоянную Планка \hbar , Луи де-Бройль связал энергию ε и импульс \mathbf{p} частицы с частотой ω и волновым вектором \mathbf{k} :

$$\varepsilon = \hbar\omega, \quad \mathbf{p} = \hbar\mathbf{k}. \quad (3.1)$$

Сопоставляя утверждения Эйнштейна и де-Бройля, можно сказать, что выписанные равенства (их называют соотношениями де-Бройля) можно читать как справа налево, так и слева направо: *волны обладают корпускулярными свойствами, частицы — волновыми*.

Соотношения де-Бройля получили экспериментальное подтверждение еще до создания квантовой механики. Введение квантов света позволило объяснить закономерности фотоэффекта, а также сдвиг частоты рентгеновского излучения при рассеянии электронами ²⁾. Волновой процесс, сопровождающий движение частиц, стал объективной реальностью, когда была открыта дифракция электронов на кристаллической решетке.

¹⁾ Точнее было бы сказать «с направлением распространения фазы волны».

²⁾ Явление получило название эффекта Комптона по фамилии открывшего его американского физика. Теория эффекта основана на анализе законов сохранения импульса и энергии, причем и электрон, и квант света рассматриваются как сталкивающиеся частицы.

Надо подчеркнуть, что соотношения де-Бройля оказались справедливыми не только качественно, но и количественно. И все же, наверное, главное значение соотношений де-Бройля заключается в том, что они послужили отправной идеей для Эрвина Шредингера при формулировке уравнения, носящего его имя и ознаменовавшего создание *волновой механики* — одного из вариантов квантовой механики.

Квантовая механика допускает различные способы описания движения микроскопических частиц. Открытые независимо, они оказались эквивалентными. Однако в первое время (в первой четверти прошлого века) все были уверены, что используемые математические абстракции несовместимы. К сожалению, мы не сможем сопоставить механику Гейзенберга, оперировавшую некоммутирующими величинами, с волновой механикой, созданной Шредингером. Мы ограничимся только волновой механикой.

В квантовой механике очень важную роль играет термин *состояние*.

Этот термин, конечно, используется и в классической физике, но, как правило, в тех случаях, когда речь идет о многих частицах.

В квантовой механике термин «состояние» применяется часто, в том числе и по отношению к отдельной частице. Вполне законен вопрос: «В каком состоянии находится электрон в атоме водорода?» Ответ звучит следующим образом: «Электрон находится в основном состоянии» либо «Электрон находится в состоянии . . .», далее следует обозначение одного из состояний. Признавая планетарную модель атома и исповедуя классическую механику, то есть считая, что электроны — «планеты», вращающиеся вокруг ядра-«солнца», мы бы так вопрос не задали. Наверное, мы спросили бы, что собой представляет траектория электрона при том или ином значении его энергии. В классической физике энергия частицы может быть любой. Она лишь не должна быть меньше ее потенциальной энергии.

То, что квантовая механика запрещает атомным и субатомным частицам двигаться по траекториям, часто подчеркивается. Отсутствие траектории — следствие существования у микрочастиц волновых свойств. Когда какое-либо тело движется по траектории, его центр тяжести в каждый момент времени имеет определенные координату и скорость ¹⁾. Для волновых свойств попросту нет места.

Поэтому отсутствие траектории, невозможность как для электрона, так и для любой другой атомной или субатомной частицы одновременно иметь определенные координату и скорость возведены в один из основных принципов квантовой механики, называемый *принципом неопределенности Гейзенберга* (последний установил его в 1927 году).

Но негативным принципом, конечно, нельзя ограничиться.

¹⁾ Чаще принято использовать не скорость, а импульс, равный скорости, умноженной на массу частицы.

Как же описывают состояние в квантовой механике? Каким образом определить состояние электрона так, чтобы иметь возможность непротиворечиво описать и корпускулярные, и волновые его свойства?

Способ, принятый в волновой механике, состоит во введении так называемой *волновой функции*. Обычно волновую функцию обозначают греческой буквой Ψ (пси). Поэтому ее часто называют *пси-функцией* или *Ψ -функцией*.

В квантовой механике слова «волновая функция» и «состояние» — синонимы.

Естественен вопрос: от каких переменных зависит Ψ -функция? Пусть речь идет об одной частице. Тогда можно считать, что волновая функция является функцией координат ¹⁾ и времени. В классической физике нередко фигурируют функции координат. От координат зависит электромагнитное поле; температура различна в зависимости от расстояния до нагревателя; скорость течения жидкости по трубе в центре нее больше, а ближе к стенке поменьше. Координата отмечает точку, а значения соответствующих функций регистрируют значения определенных физических величин в этой точке (одной или нескольких, например плотности жидкости и ее температуры).

С Ψ -функцией дело обстоит иначе. Надо учитывать, движение какого количества частиц мы изучаем: Ψ -функция зависит от координат всех частиц. Для одного электрона волновая функция — просто функция координаты r , для двух электронов — функция координат r_1, r_2 двух электронов и т. д. Введен специальный термин — *конфигурационное пространство*. Его размерность равна $3N$, где N — число частиц, состояние которых описывает Ψ -функция. О спине мы пока не упоминаем (см. гл. 7).

В 1926-м году Шредингер сформулировал уравнение, которому должна удовлетворять волновая функция. Это уравнение носит его имя. Следует подчеркнуть, что формулируя уравнение, Шредингер не знал, какова физическая суть вводимой им волновой функции. Заслуга выяснения физического смысла последней принадлежит Макс Бору.

Определение состояния (волновой функции) не есть самоцель при решении задач квантовой механики. Аппарат квантовой механики содержит рецепты предсказания на основании волновой функции результатов всевозможных экспериментов. Результаты фиксируются классическими приборами, поэтому ответы любой квантово-механической задачи должны быть получены на языке классической физики. Именно благодаря этому волновая функция приобретает конкретный смысл, поскольку с ее помощью можно получить конкретные ответы на конкретно поставленные вопросы о поведении частицы.

¹⁾ Можно выбрать и другие переменные. Выбор переменных называют выбором представления. Мы ограничимся *координатным представлением*.

Однако для человека, не воспитанного на квантовой механике, способ получения ответов с помощью волновой функции, а главное, характер ответов в большинстве случаев весьма необычны.

Действительно, как относиться человеку, хорошо знающему классическую механику, к тому, что существует туннельный эффект и атомная частица может пройти через область пространства, где потенциальная энергия больше полной энергии частицы, а ее скорость, следовательно, является мнимой величиной? Или к тому, что энергия частицы в потенциальной яме принимает только дискретный набор значений (их называют *энергетическими уровнями*)?

Сколько таланта и труда пришлось затратить создателям квантовой механики, чтобы понять самим и убедить других в том, что квантовая механика выдает разумные и окончательные ответы, не противоречащие логике и не приводящие к противоречиям!

Но, пожалуй, труднее всего было привыкнуть к тому, что во многих задачах квантовой механики *окончательный ответ не имеет достоверного характера*. Попытка теоретически выяснить, через какое отверстие в экране с двумя отверстиями должен пройти электрон от источника до приемника, приводит к следующему результату: с такой-то вероятностью — через одно отверстие, а с такой-то вероятностью — через другое. При сравнении теории с экспериментом необходимо иметь дело с большим числом электронов (как иначе определить вероятность!?). Расчет возможных вероятностей производится на основании решения уравнения, описывающего движение одного электрона с максимально возможной точностью. Эксперимент, поставленный специально для проверки результатов теории, покажет, что отдельные электроны (их, естественно, много) ведут себя по-разному. При этом электроны не взаимодействуют друг с другом. Каждый из них находится, с точки зрения описания его поведения, в совершенно тождественных условиях. Уточнить что-либо, чтобы получить однозначный ответ, невозможно.

«Абстракция, да и только!» — возмутится «непосвященный».

Большинство физиков-теоретиков с таким высказыванием не согласится. В слове «абстракция» они услышат нотку осуждения квантовой механики, некоего недовольства ею.

Зададим себе вопрос: «Почему так трудно принять вероятностный характер результатов, получаемых с использованием волновой функции и не допускающих уточнения никаким способом — ни фактически, ни даже с помощью мысленного эксперимента?». В классической физике причиной вероятностного описания всегда является либо неполное знание, либо невозможность воспользоваться полным знанием, даже если принципиально это возможно, но практически неосуществимо. Яркий пример — макроскопические системы. Невозможность описать движение всех частиц, из которых состоит система, заставляет прибегать к теории вероятности. Так возникает особая наука — *статисти-*

ческая физика. Обладай мы бесконечными возможностями, не было бы необходимости выходить за пределы классической механики: мы могли бы следить за движением каждой частицы.

Отсюда ощущение: раз есть вероятность, значит возможно уточнение. А логика квантовой механики утверждает, что *уточнение невозможно*. Именно уверенность в том, что данный вывод окончателен, что квантово-механическое описание принципиально невозможно усовершенствовать, избавившись от вероятностей, и является главным источником трудности восприятия квантовой механики. Это психологическая трудность. Но она столь велика, что даже Эйнштейн, будучи одним из создателей квантовой физики, до конца жизни не мог полностью «примириться» с квантовой механикой, хотя признавал ее успехи и не отрицал ее логической непротиворечивости.

Теперь нам предстоит познакомиться (весьма поверхностно, конечно) с математическим аппаратом квантовой механики. Читателю придется преодолеть возможный страх перед математикой. Надеемся, что мы не воздвигаем перед ним непреодолимое препятствие. Будем использовать лишь несложные математические операции и ограничимся максимально упрощенными задачами для одной частицы. Но и в этом случае нам придется прибегнуть к «рецептурному подходу»: делай так, результат будет означать то-то.

Главным объектом математических операций является комплексная волновая функция $\Psi = \Psi(\mathbf{r}, t)$.

Приведем перечень квантовомеханических рецептов.

1. Физический смысл волновой функции: $|\Psi(x, t)|^2 dx$ есть вероятность обнаружить частицу в интервале $(x, x + dx)$ в момент времени t (определение легко обобщается на трехмерный случай).

2. Каждой физической величине (координате, импульсу, моменту количества движения и т. д.) ставится в соответствие оператор, указывающий, какое действие надо произвести над Ψ -функцией.

3. Оператором импульса является оператор $(\hbar/i)d\dots/dr$, где $i = (-1)^{1/2}$ — мнимая единица. Оператор координаты \mathbf{r} — сам вектор \mathbf{r} (можно сказать иначе: оператор координаты \mathbf{r} есть оператор умножения). Сказанное о координате — следствие *координатного представления* (см. примечание на с. 182).

4. Важную роль играют такие состояния (Ψ -функции), действие на которые оператора физической величины сводится к умножению Ψ на константу. Они называются *собственными функциями оператора*, а получающиеся в результате действия оператора константы — его *собственными значениями*.

Пример. Собственные функции оператора импульса — плоские волны:

$$\Psi_p(\mathbf{r}) = A \exp(i\mathbf{p}\mathbf{r}/\hbar). \quad (3.2)$$

Вектор \mathbf{p} имеет действительные компоненты. Подействовав оператором импульса на Ψ_p (продифференцировав и умножив на \hbar/i), получим $\mathbf{p}\Psi_p$. Следовательно, в состоянии Ψ_p импульс частицы равен \mathbf{p} .

5. Если Ψ -функция не является собственной функцией оператора физической величины \hat{C} (шляпка над буквой означает, что \hat{C} — оператор), то физическая величина \hat{C} в состоянии Ψ не имеет определенного значения.

При измерении физической величины \hat{C} будут получаться различные значения. Какие? Такие, каковы собственные значения оператора \hat{C} . Аппарат квантовой механики позволяет выяснить, с какой вероятностью будет получаться то или иное значение. Для этого Ψ -функцию надо разложить по собственным функциям оператора физической величины. Последнее всегда возможно, поскольку собственные функции каждого оператора физической величины представляют собой *полный набор функций*. Квадраты модулей коэффициентов разложения пропорциональны вероятностям того, что при измерении физической величины будут получены величины, равные соответствующим собственным значениям оператора.

6. Введение Ψ -функции и знание операторов физических величин позволяет конкретизировать принцип неопределенности, записав его в виде *соотношений неопределенности*.

Возможность вычисления вероятностей значений физических величин позволяет определить их средние значения (будем обозначать их большими буквами или угловыми скобками).

Сформулируем соотношения неопределенности для координат и импульса.

Координата частицы вдоль одной из осей может иметь определенное значение одновременно с компонентами импульса по двум другим осям.

Координата и компонента импульса вдоль одной и той же оси одновременно не существуют. Пусть X и P_x — средние значения координаты и импульса в каком-либо из состояний. Мерой неопределенности координаты и импульса могут служить следующие средние величины:

$$\delta x = [\langle (x - X)^2 \rangle]^{1/2}, \quad \delta p_x = [\langle (p_x - P_x)^2 \rangle]^{1/2}. \quad (3.3)$$

Соотношение неопределенностей утверждает, что

$$\delta x \delta p_x \geq \frac{\hbar}{2}. \quad (3.4)$$

Наименьшее значение произведения неопределенностей равно $\hbar/2$.

7. Фундаментальное уравнение волновой механики — аналог уравнения Ньютона в классической механике — уравнение Шредингера. Для его формулировки используется гамильтониан H — оператор,

построенный из функции Гамильтона:

$$H = \frac{p^2}{2m} + U(\mathbf{r}), \quad p^2 = p_x^2 + p_y^2 + p_z^2. \quad (3.5)$$

Оператор H представляет собой энергию частицы, выраженную через импульс \mathbf{p} и координату \mathbf{r} . Потенциальная энергия $U(\mathbf{r})$ определяет силу $\mathbf{F} = -dU(\mathbf{r})/d\mathbf{r}$, действующую на частицу. Знание оператора импульса (его компоненты: $p_x = (\hbar/i)d/dx$, $p_y = (\hbar/i)d/dy$, $p_z = (\hbar/i)d/dz$) позволяет записать гамильтониан частицы как

$$H = -\frac{\hbar^2}{2m}\Delta + U(\mathbf{r}), \quad \text{где } \Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \quad (3.6)$$

Уравнение Шредингера имеет следующий вид:

$$-\frac{\hbar}{i} \frac{d\Psi}{dt} = \hat{H}\Psi. \quad (3.7)$$

Поскольку в уравнение входят производные от Ψ -функции, необходимо добавить граничные и начальные условия. Постановка задачи помогает их формулировке (см. ниже). Отметим только, что задание волновой функции в начальный момент времени (при $t = 0$) позволяет определить ее во все последующие моменты времени.

Рассмотрим несколько конкретных задач.

Свободная частица. Согласно классической механике у свободной частицы сохраняется импульс. Импульс сохраняется и в квантовой механике. Поэтому решение уравнения (3.7) будем искать в виде

$$\Psi(\mathbf{r}, t) = \varphi(t)\Psi_p(\mathbf{r}).$$

Произведя дифференцирование и сократив на $\Psi_p(\mathbf{r})$, получим выражение $(-\hbar/i)d\varphi(t)/dt = \varepsilon\varphi(t)$, где $\varepsilon = p^2/2m$ — энергия частицы. Отсюда $\varphi(t) = \exp(-i\varepsilon t/\hbar)$, а

$$\Psi(\mathbf{r}, t) = A \exp[(-i/\hbar)(\varepsilon t - \mathbf{p}\mathbf{r})]. \quad (3.8)$$

Таким образом, волновая функция свободно движущейся частицы представляет собой плоскую волну, имеющую в строгом соответствии с соотношениями Луи де-Бройля следующие характеристики:

$$\omega = \frac{\varepsilon}{\hbar}, \quad \mathbf{k} = \frac{\mathbf{p}}{\hbar}. \quad (3.9)$$

Квантовая свободная частица полностью делокализована, как того требуют соотношения неопределенности (3.4). Поскольку частица имеет определенный импульс, координата ее не определена вовсе (формально это следует из независимости от координаты плотности вероятности $|\Psi(\mathbf{r}, t)|^2$).

Стационарные состояния. Рассмотренная выше задача о свободной частице — пример определения ее стационарного состояния, то есть состояния, энергия которого сохраняется. Данный класс задач

очень важен. Электроны в атоме имеют стационарные состояния, поэтому построить теорию атома можно, только умея решать задачи по их определению. Итак, в стационарных состояниях энергия сохраняется. Последнее справедливо, если потенциальная энергия не зависит от времени. Такие ситуации мы и будем рассматривать.

Согласно общим правилам (см. 4-й пункт «Перечня рецептов») значения энергии частицы являются собственными значениями оператора Гамильтона H , а волновые функции — собственными функциями оператора H , удовлетворяющими стационарному уравнению Шредингера:

$$H\Psi(\mathbf{r}) = \varepsilon\Psi. \quad (3.10)$$

Это означает, что волновая функция стационарного состояния имеет вид $\Psi(\mathbf{r}, t) = \exp(-i\varepsilon t/\hbar)\Psi(\mathbf{r})$. Отсюда немедленно следует вывод: в стационарных состояниях вероятности значений физических величин, а также их средние значения не зависят от времени (см. «Перечень рецептов»).

Набор возможных значений энергии называют *энергетическим спектром*. В зависимости от вида потенциальной энергии спектр может быть дискретным или непрерывным. В большинстве случаев полный спектр при заданной потенциальной энергии является комбинацией дискретного и непрерывного спектров. В одних интервалах значений энергии спектр дискретный, в других — непрерывный. Обычно характер спектра зависит от того, совершает частица *инфинитное* (непрерывный спектр) или *финитное* (дискретный спектр) движение. Далее будут приведены примеры как дискретного, так и непрерывного спектров, а в гл. 17 мы расскажем о важном примере *зонного спектра*, когда зоны (полосы) разрешенных непрерывных значений энергии разделяются зонами (полосами) запрещенных значений.

Туннельный эффект. Пусть потенциальная энергия имеет вид, изображенный на рис. 3.1, а частица, энергия которой $\varepsilon < U_0$, летит

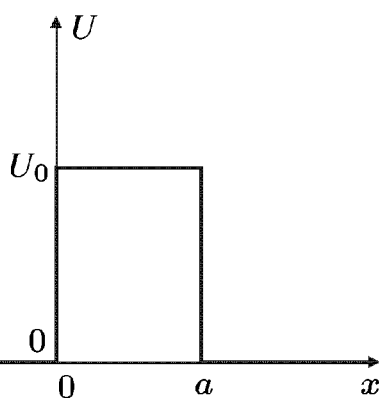


Рис. 3.1. Прямоугольный потенциальный барьер

слева направо. Классическая частица с необходимостью отразится от потенциального барьера и вернется обратно — туда, откуда прилетела. Как будет вести себя квантовая частица?

Может показаться, что прямоугольный потенциальный барьер — надуманная абстракция. Конечно, в рис. 3.1 присутствует идеализация. Однако если вспомнить, что сила, действующая на частицу, есть $-dU_0/dx$, то легко понять, что рисунок изображает силовой диполь — два узкие участка, разнесенные на расстояние a друг от друга, пролетая через которые, частица испытывает действие сил, равных по величине и направленных строго

в противоположные стороны. При $\varepsilon < U_0$ тормозящая сила «завернет» налетающую классическую частицу.

Решение квантовомеханической задачи требует формулировки граничных условий в точках $x = 0$ и $x = a$. Мы не имеем возможности строго объяснить, почему в этих точках Ψ -функция и ее производная $d\Psi/dx$ должны быть непрерывны. Для того чтобы подчеркнуть важность абстрактных принципов, добавим, что условие непрерывности волновой функции и ее производной является следствием требования сохранения числа частиц: если на потенциальный барьер падает одна частица, то в результате взаимодействия с барьером она не может исчезнуть и не могут возникнуть новые частицы. К граничным условиям необходимо отнести и требование о структуре волновой функции вне барьера. Считая, что частица налетает на барьер слева, мы понимаем, что при $x < 0$ волновая функция — линейная комбинация падающей и отраженной волн де-Бройля, а при $x > a$ имеет место только одна волна — прошедшая.

Что собой представляет волновая функция в интервале $(0, a)$? Уравнение (3.10) принимает здесь следующий вид:

$$\frac{d^2\psi}{dx^2} - \kappa^2\psi = 0, \quad \kappa = \left[\frac{2m}{\hbar^2}(U_0 - \varepsilon) \right]^{1/2}. \quad (3.11)$$

Как легко проверить, общим решением этого уравнения является

$$\psi = A_1 \exp(\kappa x) + A_2 \exp(-\kappa x), \quad 0 < x < a. \quad (3.12)$$

Мы не выписали вид волновой функции при $x < 0$ и при $x > a$, но сказанное выше убеждает нас, что для ее отыскания нужно определить четыре константы: A_1 , A_2 , амплитуды прошедшей и отраженной волн. Для этого служат четыре граничные условия. Зная волновую функцию, нетрудно вычислить коэффициенты прохождения и отражения. Коэффициент прохождения D (вероятность обнаружить частицу за потенциальным барьером) равен

$$D = \frac{4k^2\kappa^2}{(k^2 + \kappa^2)^2 \operatorname{sh}^2(a\kappa) + 4k^2\kappa^2}, \quad k = \left(\frac{2m\varepsilon}{\hbar^2} \right)^{1/2}, \quad (3.13)$$

$$R = 1 - D,$$

где R — коэффициент отражения.

Формула (3.13) дает возможность обсудить два вопроса, полезные для понимания квантовой механики.

1. Если в полученной с помощью аппарата квантовой механики формуле постоянную Планка \hbar положить равной нулю, то, как правило, должна получиться формула, которую можно было бы вывести,

используя уравнения классической механики ¹⁾. Это требование так называемого *принципа соответствия*. При построении квантовой механики принцип соответствия сыграл важную эвристическую роль.

Вспомните формулу (3.13) и в определения величин k и κ . Видно, что фактически постоянную Планка можно сократить везде, кроме аргумента гиперболического синуса. Поэтому, как и следовало ожидать, с уменьшением постоянной Планка коэффициент прохождения стремится к нулю. Стремится в данном случае очень быстро — экспоненциально.

Постоянная Планка — мировая постоянная. Неловко как-то принимать ее равной нулю. Поступим иначе: оценим коэффициент прохождения для случая, когда налетающая частица имеет массу m , равную грамму; высота барьера и энергия частицы — порядка работы, которую надо затратить, чтобы тот же грамм поднять на высоту $h = 1$ см над землей (то есть $\varepsilon - U_0 = mgh$, где g — ускорение силы тяжести); ширина барьера $a = 1$ см. Тогда $a\kappa = a[(2m/\hbar^2)(U_0 - \varepsilon)]^{1/2} \cong (am/\hbar)(2gh)^{1/2} \cong 10^{29}$, а $D < \exp(-10^{29})$. Думаю, такую незначительную вероятность можно считать равной нулю, будучи уверенным, что столь невероятное событие никогда не произойдет.

Ясно, что описание движения макроскопических тел не требует уточнений за счет квантовых поправок.

2. Формула (3.13) выведена с использованием уравнения (3.11). Те, кто знаком с классической электродинамикой, понимают, что оно совершенно аналогично волновому уравнению — следствию уравнений Максвелла. Минус — знак коэффициента перед функцией ψ — соответствует отрицательной диэлектрической проницаемости, то есть чисто мнимому показателю преломления. В среде с мнимым показателем преломления световая волна проникает, но при этом экспоненциально затухает. Об этом свидетельствуют формула (3.12) и гиперболический синус вместо обычного.

При $\varepsilon > U_0$ классическая частица вовсе не замечает потенциального барьера, а квантовая — частично проходит через промежуток $(0, a)$, частично отражается. Действует аналогия с электромагнитной волной, но с волной, распространяющейся в среде с неоднородным показателем преломления.

Становится очевидным, что туннельный эффект — результат существования у микрочастиц волновых свойств.

До сих пор рассматривалось инфинитное движение частицы. Как мы уже говорили, в этом случае спектр энергий непрерывен. Следующий пример — пример финитного движения.

¹⁾ Встретившись с ситуацией, когда это правило, казалось бы, нарушается, всегда можно обнаружить, что при выводе квантовомеханической формулы использовалось условие, из-за которого переходить к пределу $\hbar = 0$ нельзя.

Частица в потенциальной яме. Пусть $U = \infty$ при $|x| \geq a/2$ и $U = 0$ при $|x| < a/2$, а волновая функция при $|x| < a/2$ удовлетворяет следующему простому уравнению:

$$\frac{d^2\Psi}{dx^2} + k^2\Psi = 0, \quad k^2 = \frac{2m\varepsilon}{\hbar^2}. \quad (3.14)$$

При $|x| > a/2$ волновая функция тождественно равна нулю. Кроме того, можно показать, что $\Psi(a/2) = \Psi(-a/2) = 0$. Эти два равенства служат граничными условиями.

Тригонометрические функции (четная — косинус и нечетная — синус) удовлетворяют уравнению (3.14):

$$\Psi = A \cos(kx) \text{ или } \Psi = B \sin(kx).$$

Граничные условия будут выполнены, если в случае косинуса

$$k = k_n = (2n + 1)\frac{\pi}{a}, \quad n = 0, 1, 2, \dots,$$

а в случае синуса

$$k = k_n = 2n\frac{\pi}{a}, \quad n = 1, 2, 3, \dots$$

Из приведенных формул видно, что энергия частицы в потенциальной яме может иметь только *дискретные значения*:

$$\varepsilon_n = \frac{\hbar^2 k_N^2}{2m}, \quad k_N = \frac{N\pi}{a}, \quad (3.15)$$

где $N = 1, 2, 3, \dots$ — произвольное целое число.

Нормировочные постоянные A и B можно выбрать так, чтобы полная вероятность обнаружить электрон в яме была равна единице. Это означает, что интеграл от $|\Psi|^2$ от $-a/2$ до $+a/2$ должен быть равен единице, то есть $A_n = B_n = (2/a)^{1/2}$.

Что мы выяснили? Что энергия принимает дискретные значения. Но не только это.

Чем уже яма, тем больше энергия основного (наинизшего) состояния, равная $\hbar^2\pi^2/2a^2m$; частица не может упасть на дно, поскольку среди разрешенных значений энергии нет нулевого. Это естественные следствия соотношений неопределенности. Из-за ограничения положения частицы размером ямы она имеет неопределенность импульса. Воспользовавшись формулами Эйлера¹⁾, легко убедиться, что при измерении импульса в одном из найденных состояний мы с вероятностью, равной $1/2$, получим значения $+\hbar k_N$ и $-\hbar k_N$. Средний импульс равен нулю, но неопределенность импульса равна $\hbar k_N$. Чем уже яма, тем эта неопределенность больше (с указанными нами последствиями).

¹⁾ $\cos \alpha = (1/2)(e^{i\alpha} + e^{-i\alpha}), \quad \sin \alpha = (1/2i)(e^{i\alpha} - e^{-i\alpha}).$

Как по формуле (3.15) проверить, выполняется ли принцип соответствия? Для этого нужно убедиться, что при $\hbar = 0$ и минимальное значение энергии, и расстояния между уровнями равны нулю. Равенство нулю наименьшего значения энергии при $\hbar = 0$ очевидно. Остается зафиксировать энергию, а затем с помощью выражения (3.15) проанализировать отношение $\Delta\varepsilon/\varepsilon$, где $\Delta\varepsilon = \varepsilon_{N+1} - \varepsilon_N$. Легко убедиться в том, что если $\varepsilon \gg \varepsilon_1$, то $\Delta\varepsilon/\varepsilon \approx 2(\varepsilon_1/\varepsilon)^{1/2}$, $\varepsilon_1 = \hbar^2\pi^2/2a^2m$. Принцип соответствия, естественно, не нарушается. Более того, оценка, аналогичная сделанной нами для коэффициента прохождения, показывает, что при макроскопических значениях параметров квантовые поправки столь малы, что при решении большинства задач макрофизики нет никакой необходимости их учитывать.

В заключение перепишем формулу (3.15) в терминах волн де-Бройля. Вспоминая, что $2\pi/\lambda = k = \hbar/p$, а $p = (2m\varepsilon)^{1/2}$, убеждаемся, что на ширине потенциальной ямы укладывается целое число волн де-Бройля: $\lambda N = 2a$.

Атом водорода. Первым успехом волновой механики явилась последовательная теория атома водорода, основанная на решении уравнения Шредингера с потенциальной энергией, равной $-e^2/r$. Как ни удивительно, Шредингер знал ответ. Дело в том, что Нильс Бор, исходя из законов классической механики и навязав ей, казалось бы, незаконные требования, нашел дискретные электронные энергетические уровни в атоме водорода, а предположив, что излучение и поглощение световых квантов есть результат перехода электрона с уровня на уровень, получил правильную картину спектра. Не придерживаясь исторической последовательности событий, заметим: как оказалось в дальнейшем, подход Бора совпадает с *квазиклассическим приближением*, справедливым в случае, когда действие велико по сравнению с \hbar . (Действие — механическая характеристика движения той же размерности, что и постоянная Планка [эрг · с]). Несомненной удачей и Бора, и Шредингера было то, что задача об атоме водорода принадлежит к редкому классу задач, в которых решение, полученное в квазиклассическом приближении, совпадает с точным (по крайней мере для уровней энергии электрона).

Воспользуемся эклектическим приемом. Будем считать, что электрон вращается вокруг «бесконечно тяжелого» протона по классической траектории — окружности радиуса a . Энергия электрона, как показано в предыдущей главе, равна $\varepsilon = -e^2/2a$. Потребуем, чтобы длина окружности равнялась целому числу длин волн де-Бройля ($N\lambda = 2\pi a$). Поскольку из равенства центробежной силы и силы притяжения $a = me^2/p^2$, а $p = 2\pi\hbar/\lambda$, имеем

$$\varepsilon = -\frac{e^4 m}{2\hbar^2} \frac{1}{N^2}, \quad N = 1, 2, 3, \dots \quad (3.16)$$

Мы видим, что разрешенные значения энергии представляют собой систему дискретных уровней, сгущающихся с ростом энергии — при приближении к нулю. Положительной энергии соответствует инфинитное движение электрона (аналог движения по параболе или гиперболе). При $\varepsilon > 0$ спектр непрерывен.

Все состояния с $N > 1$ метастабильны: излучив квант света с энергией $\hbar\omega$, равной разности разрешенных значений энергии, электрон может опуститься на уровень с меньшей энергией. Состояние с $N = 1$ является *основным, устойчивым состоянием* электрона. Именно существованием таких состояний квантовая механика объясняет устойчивость атомов.

Для того чтобы ионизовать атом водорода, если он находится в основном состоянии, нужно увеличить энергию электрона на величину $e^4 m / 2\hbar^2 = 13,60$ эВ.

Эта величина используется как единица энергии в атомной физике и носит название Ридберг (в честь шведского физика Иоганесса Роберта Ридберга, 1854–1919, занимавшегося систематизацией атомных спектров).

Заключительные замечания. Не знаю, удалось ли мне показать, как квантовая механика позволяет понять устройство мира атомов. Хотелось поделиться не только сведениями о ряде правил и приемов, принятых при описании движения микрочастиц, но и своим восхищением перед математическим аппаратом квантовой механики, приспособленным к решению разнообразных задач, связанных с поведением атомных и субатомных частиц.

Несмотря на впечатляющие успехи квантовой механики, ее структура и логика привлекали и привлекают усиленное внимание.

Все работы, не останавливаясь на деталях, можно подразделить на два класса. К первому отнесем решение конкретных задач атомной физики: объяснение экспериментальных результатов и предсказание новых, специфических квантовых эффектов. Ко второму — выяснение непротиворечивости и логической строгости квантовой механики — науки, качественно новой по сравнению с классической механикой.

Благодаря работам первого класса понимание строения материи необычайно углубилось, огромное количество свойств и явлений получило объяснение, такие дисциплины, как ядерная физика, физика твердого тела, физика магнетизма, физика квантовых жидкостей и другие, из описательных превратились в логически стройные, основанные на квантовой механике области науки.

Всесторонняя проверка логики квантовой механики была проведена в работах, принадлежащих второму классу. Она убедила в безукоризненности основ квантовой механики.

То, что к этому вопросу возвращаются, показывает, насколько трудно привыкнуть к способу описания действительности, необходимому в микромире.

Как правило, встретившись с чем-то совсем необычным, человечество создает новое понятие и постепенно привыкает к нему. Для утверждения о том, что электроны, позитроны и другие представители микромира обладают свойствами и частиц, и волн, создан специальный термин — *корпускулярно-волновой дуализм*. Им пользуются. Но, по-видимому, из-за несочетаемости свойств корпускул (частиц) и волн смысл этого термина плохо укладывается в нашем сознании и чувство неудовлетворенности полностью не исчезает. Хочется задать вопрос: «Что же такое электрон? *И* корпускула, *и* волна? А возможно, *или* корпускула, *или* волна?» Очень трудно найти доступные воображению модели того, что невозможно себе представить. Электрон *и не* волна, *и не* классическая частица. Электрон есть электрон. Однако если попытаться все же отыскать наглядные модели в арсенале классической физики, то это и «*и ... , и ...*», и «*или ... , или ...*». Даже из тех примеров, которые мы рассмотрели, видно, что волновые свойства электрона, несомненно, надо учитывать. В его корпускулярных свойствах, наверное, никто не сомневается.

Мы уже знаем, что размер атома, ограничивающий область пространства, в которой, благодаря притяжению к ядру, сосредоточены электроны, определяется волновыми свойствами последних (см. выше). Более наглядную иллюстрацию волновых свойств частицы трудно придумать. Итак, *и ... , и ...*? Но в гл. 13 мы убедимся, что как ни оценивай размер электрона, размер атома во много раз его превышает.

Когда мы осуществляем какой-либо эксперимент с электронами, то в зависимости от постановки эксперимента электрон проявляет себя *либо* как частица, *либо* как волна. *Или ... , или!*

В фотоэффекте, или засвечивая зерно фотоэмульсии, электрон — частица; демонстрируя дифракцию на кристаллической решетке — волна. И никогда при анализе постановки и результатов экспериментов, ни реальных, ни мысленных, не были обнаружены противоречия с основными положениями квантовой механики. В частности, не удалось выйти за пределы соотношений неопределенности.

Наверное, следует еще раз напомнить, как много усилий приложил Альберт Эйнштейн, пытаясь доказать *неполноту* квантовой механики. Его дискуссии с Нильсом Бором, а также строгий анализ разнообразных ситуаций, проведенный многими физиками, похоже, позволили окончательно убедиться: *квантовая механика — логически безукоризненная, завершенная наука.*

Глава 4. СИСТЕМЫ КООРДИНАТ. РЕЛЯТИВИСТСКАЯ РЕВОЛЮЦИЯ

В главе, посвященной классической механике, мы обошли молчанием вопрос о том, относительно чего движется тело, подчиняясь законам Ньютона. Часто при решении практических задач сама постановка вопроса подсказывает необходимость введения конкретной системы координат. Когда автомашина движется по шоссе, совершенно естественно именно шоссе считать телом отсчета, а пройденный путь измерять по расстоянию вдоль пути следования. Для описания полета воздушного шара необходимо учитывать движение Земли, а при расчете полетов космических аппаратов только «неподвижные» звезды пригодны для роли тела отсчета.

Введение разумной системы координат, как будет ясно, предполагает абстрагирование.

Давайте не будем торопиться и расскажем несколько подробнее о том, как удастся выбрать систему координат, пригодную для описания любых движений.

Как было сказано в первом абзаце, казалось бы, наиболее естественный способ введения системы координат — указать тело, к которому можно, пусть мысленно, прикрепить координатные оси. Во времена Ньютона так и поступали, принимая за тело отсчета «неподвижные» звезды. Прилагательное «неподвижные» и здесь, и ранее взято в кавычки, поскольку оказалось, что неподвижных звезд нет.

Возникает скептический вопрос: не кажется ли Вам, что до выбора системы отсчета утверждение, что нечто неподвижно, вообще несколько нелогично? Ведь мы еще не умеем определять положение!

Наверное, лучше не искать реальное тело (или тела), которое может служить системой отсчета, а задуматься, каким свойствам она должна удовлетворять.

Принято использовать *инерциальную систему отсчета*.

О п р е д е л е н и е. Система отсчета называется инерциальной, если в ней тело, на которое не действуют никакие силы, покоится или движется с постоянной скоростью.

Введение инерциальной системы отсчета — результат абстрагирования. Заэкранировать полностью какое-либо тело от всех воздействий невозможно. Но если воздействие мало, от него можно абстрагироваться. Конечно, до тех пор пока мы не захотим исследовать влияние той

самой малой силы, которой мы пренебрегли при выборе инерциальной системы отсчета.

Выбор инерциальной системы неоднозначен. Система отсчета, движущаяся с постоянной скоростью относительно выбранной нами инерциальной системы, также инерциальна. В классической механике переход из одной инерциальной системы в другую осуществляется с помощью преобразования Галилея:

$$\mathbf{r}' = \mathbf{r} + \mathbf{V}t,$$

где \mathbf{r}' и \mathbf{r} — радиусы-векторы в разных системах, \mathbf{V} — скорость движения одной системы относительно другой, t — время. Согласно классической механике и в полном соответствии с нашими ощущениями *время во всех системах течет одинаково*. Если бы мы избрали системой отсчета тело, движущееся с ускорением, то и тогда оно не изменило бы своего течения. В доэйнштейновской физике время *абсолютно*.

Хотя инерциальная система отсчета — абстракция, мы с легкостью находим системы отсчета, прекрасно выполняющие роль инерциальных. Часто такой системой служит Земля. Правда до тех пор, пока мы передвигаемся по ней или около нее на небольшие расстояния. А в личных целях роль инерциальной системы прекрасно выполняют парожитная каюта, железнодорожный вагон, самолет, даже велосипед. Иногда (и не так уж редко) средства передвижения бунтуют: резкий поворот, неожиданная встряска, и мы с трудом удерживаемся на ногах или вдавливаемся в кресло, ощущая перегрузку. Это проявляют себя силы инерции, обязанные своим существованием ускорению выбранной нами системы отсчета — ее отличию от инерциальной.

Все разговоры о необходимости абстракции при введении инерциальной системы не могут поколебать ощущения, что способ описания геометрии пространства и течения времени, принятый в ньютоновской физике, соответствует нашим житейским наблюдениям. Высказывания типа «при ожидании время тянется страшно медленно» и «счастливые часов не наблюдают», очевидно, являются просто метафорами.

Казалось бы, ни у кого не вызывает сомнений, что перенос предмета (например, стержня) из одной системы в другую — вне зависимости от того, инерциальна система или нет — не изменяет его размеров. Правда, утверждая это, надо проявить осторожность: значительные силы инерции могут в определенных условиях деформировать тело. Поэтому лучше сказать более осторожно: в любой инерциальной системе тело имеет одни и те же размеры.

Почти в любой достаточно подробной книге, посвященной изложению теории относительности, перечисляются причины, заставившие А. Эйнштейна подвергнуть строгому логическому анализу принятые

представления о геометрии пространства и о течении времени. Во главу угла, наряду с эквивалентностью всех инерциальных систем, был поставлен *принцип независимости скорости света от скорости системы отсчета*.

Согласно теории относительности скорость света c — абсолютная величина.

Как уже упоминалось (см. гл. 2), до сих пор продолжает обсуждаться вопрос, знал ли Эйнштейн результат опыта Майкельсона (независимость скорости света от скорости Земли). Процитированная ранее фраза: «Именно это негативное открытие [Майкельсона] способствовало признанию правильности специальной теории относительности», по нашему мнению, доказывает, что открытие Майкельсона произошло *после* создания теории относительности. У меня сложилось впечатление: *Эйнштейн заранее знал, что так должно быть*.

В 1938-м году А. Эйнштейн и Л. Инфельд опубликовали для широкого круга читателей книгу по истории физики «Эволюция физики. Развитие идей от первоначальных понятий до теории относительности и квантов». В связи с работой над текстом «Бесед об абстракции в физике» я неоднократно обращался к этому источнику. Кроме познавательной пользы, каждый, кого интересуют проблемы научного описания Мира, получит огромное удовольствие от чтения «Эволюции физики», так как он сумеет проникнуть в мир идей одного из величайших мыслителей.

Здесь не место излагать содержание книги Эйнштейна и Инфельда. Однако мне кажется, что следует отметить одно обстоятельство, которому авторы уделяют много внимания.

Как уже говорилось, к началу 20-го века большинство физиков было уверено, что классическая механика и теория Максвелла являются прочным фундаментом всей физики. Похоже, один Эйнштейн ощущал: теория Максвелла и механика Ньютона не могут составлять единство, они несовместимы. Понимание того, что распространение электромагнитных волн не требует эфира, то есть среды-переносчика, а без признания абсолютности скорости света невозможна эквивалентность всех инерциальных систем отсчета, потребовало отказа от, казалось бы, очевидного преобразования Галилея. В «Эволюции физики» авторы ссылаются на точно установленный к 30-м годам факт независимости скорости света от скорости Земли. Однако, как показывает описание структуры уравнений Максвелла, Эйнштейн руководствовался не столько экспериментальными фактами, сколько абстрактными соображениями, основанными на полевом характере уравнений электродинамики. Невозможность приспособить теорию Максвелла к механике Ньютона заставила, отказавшись от преобразования Галилея, заменить их преобразованиями Лоренца, соответствующими уравнениям Максвелла.

Последствия столь кардинальных изменений не заставили себя ждать. Поскольку выводы противоречили интуиции, необходимо было тщательно проанализировать логику всей схемы. Как мы упоминали, пришлось пересмотреть такое очевидное понятие, как одновременность. Логика теории относительности оказалась безупречной, а результаты — удивительными.

Выяснилось, что время не абсолютно. Длина промежутка времени между двумя событиями зависит от того, в какой из инерциальных систем производить измерение. Пространственный размер предмета также меняется при переходе из одной системы в другую. Однако существует величина, не зависящая от выбора инерциальной системы отсчета. Она носит название *интервала*. Пусть x_1, y_1, z_1, t_1 и x_2, y_2, z_2, t_2 — координаты и время двух событий (естественно, по отношению к некоторой определенной системе отсчета). Величина

$$S_{12} = [c^2(t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2]^{1/2}$$

и есть интервал (точнее, интервал между двумя событиями). Итак, интервал является инвариантом, то есть его значение не зависит от выбора инерциальной системы отсчета.

В классической, доэйнштейновской физике считалось, что существует два инварианта: одномерный — временной и трехмерный — пространственный. Оказалось, что только их комбинация — четырехмерный интервал между двумя событиями — величина, сохраняющаяся при переходе из одной системы координат в другую. Это утверждение — основа построенной в 1905 году А. Эйнштейном специальной теории относительности. Ее часто называют *релятивистской теорией*. Инвариантность интервала означает объединение пространства и времени в единое 4-пространство (Г. Минковский). Отдельно, независимо друг от друга рассматривать трехмерное пространство и время неправомерно.

Как же быть с нашими ощущениями? Дело в том, что заметить неточность классической механики, «обнаружить релятивистские эффекты», непросто: особые черты теории относительности проявляются только при движении со скоростью, близкой к скорости света.

Пусть два события, происшедшие в одном месте (скажем, на перроне), разделены временным интервалом, равным Δt . Какой результат ($\Delta t'$) мы получим, если будем фиксировать события из окна поезда, проходящего мимо перрона со скоростью V , и измерять время по своим наручным часам? Теория относительности дает следующий ответ:

$$\Delta t' = \Delta t \left[1 - \frac{V^2}{c^2} \right]^{-1/2}.$$

Очевидно, что в реальной ситуации (при измерении в поезде и на перроне) мы не заметим отличия двух промежутков времени, а вот при расчете линейного ускорителя электронов пренебрежение релятивистским сжатием длин элементов ускорителя в системе координат электрона, движущегося с околосветовой скоростью, приведет к грубым ошибкам.

Огромное впечатление производит следующий факт: время жизни любой нестабильной частицы существенно зависит от того, покоится она или движется относительно системы отсчета, в которой мы это время жизни измеряем. Движущаяся частица живет дольше (в соответствии с приведенной формулой).

Общая теория относительности была создана А. Эйнштейном в 1915-м году. Она названа *общей*, поскольку релятивистской теории Эйнштейн дал название *частной теории относительности*. Термин «частная» подчеркивает, что последняя позволяет использовать для описания физических явлений только инерциальные системы отсчета. Анализируя описание физических явлений, происходящих в неинерциальной системе отсчета, Эйнштейн пришел к выводу, что логика требует возможности описывать физическую реальность, используя произвольно (даже с ускорением) движущуюся систему координат. Обобщение — переход от частной к общей теории относительности — базируется на тождественности инертной и гравитационной масс (к этому мы еще вернемся).

Хочется подчеркнуть, что до появления общей теории относительности не были известны какие-либо экспериментальные факты, не укладывавшиеся в классическую теорию тяготения. Стимулом, который побуждал Эйнштейна к созданию новой теории, являлась внутренняя неудовлетворенность, ощущение того, что релятивистские представления должны быть основой последовательной, логически непротиворечивой картины любых физических явлений в Мире. Закон всемирного тяготения, несомненно, противоречил этим представлениям. В Мире, в котором действуют релятивистские законы, нет места силе, зависящей от расстояния между предметами. Ведь существование подобной силы означало бы, что сигнал из одной точки пространства может достигнуть другой, удаленной от нее точки мгновенно, вопреки тому, что предельная скорость распространения сигнала равна скорости света. Таким образом, общая теория относительности по существу есть *релятивистская теория гравитации*. Решение, казалось бы, конкретной задачи привело к фундаментальным изменениям наших представлений о пространстве–времени.

Представим себе, что мы избрали Землю в качестве тела отсчета, и рассматриваем лишь движение по ее поверхности. Пренебрежем не только ускорениями, вызванными вращением Земли вокруг своей оси и движением по эллиптической орбите, но и отличием ее формы от сферы. Наконец, забудем про существование на нашей планете гор

и впадин ¹⁾. Таким образом, наше пространство — поверхность сферы. Его геометрия является сферической. Она отличается от геометрии на плоскости, которую принято называть евклидовой. Мерой отличия сферической геометрии от эвклидовой служит радиус сферы R . Чем он больше, тем больше должен быть рисунок (чертеж) на поверхности сферы, по которому можно заметить, что пространство не плоское.

А какова же геометрия нашего истинного пространства — вместилища всего? То, что оно четырехмерно, не снимает вопроса. Является ли пространство евклидовым или имеет кривизну? Подобный вопрос впервые задали не физики, а математики. Они даже пытались экспериментальным путем получить ответ, однако для этого не хватило точности измерений.

Причиной возникновения сомнений послужило открытие того, что евклидова геометрия — не единственная логически безупречная. Понято, что ответ на вопрос, какова «настоящая» геометрия, должен быть получен экспериментально или на основании физической теории в случае ее существования.

Принципиальный ответ дала общая теория относительности. Точнее, она выяснила, *как* получить этот ответ. Но главное, с созданием общей теории относительности кардинально изменилось наше восприятие пространства–времени как независимой сущности. Исчезло самостоятельное пространство–время, существующее вне зависимости от того, что в нем находится и движется. Перестало существовать столь привычное пространство, будто бы специально *приготовленное* к размещению в нем и движению по нему любых тел.

Изменение произошло, конечно, лишь в нашем сознании. Свойства пространства–времени не зависят от уровня наших знаний о Природе. Кривизна пространства–времени, согласно теории Эйнштейна, зависит от материи, которая в нем находится. Вблизи больших масс (звезд, скоплений звезд) пространство искривлено больше, вдали — меньше.

¹⁾ Может показаться, что пренебрежение рельефом — наиболее грубая идеализация. Действительно, когда рассматриваешь *реальное* движение *реальных* предметов по Земле, пренебрежение высотами разных участков земной поверхности над уровнем моря невозможно. Однако, во-первых, большую часть поверхности занимают океаны и моря, а во-вторых, горы высоки и впадины глубоки только по нашим ощущениям. Максимальная разность уровней на Земле, обязанная своим существованием рельефу (сумма высоты Эвереста и глубины Марианской впадины в Тихом океане), составляет приблизительно 20 км, то есть имеет тот же порядок, что и сплюснутость Земли (отличие геоида от сферы). Радиус же Земли ~ 6400 км. Итак, безразмерная характеристика неровности земной поверхности $20/6400 = 0,003125$. Параметр достаточно мал для идеализации. Поверхность земного шара меньше отличается от идеальной сферы, чем поверхность бильярдного шара (мне кажется, я не ошибаюсь).

Открытие зависимости свойств пространства–времени от материи произвело огромное впечатление. Это была истинная революция, хотя она и не затронула экономической или социальной сторон жизни. Но по тому, какую популярность приобрел Эйнштейн, можно судить, сколько людей почувствовали, пережили происшедшее. По-видимому, у всех людей, даже далеких от физики, глубоко укоренились представления о пространстве и времени. Их изменение воспринимается как важнейшее, фундаментальное событие, от которого невозможно отмахнуться, даже если нельзя его полностью понять. Особенно остро человечество отреагировало, когда наблюдения подтвердили предсказания общей теории относительности.

Математический аппарат общей теории относительности сложнее, чем математический аппарат классической механики, используемый, скажем, при решении задач теории движения космических тел. Сложности эйнштейновских теорий по сравнению с классическими посвящена следующая эпиграмма:

Был этот мир глубокой тьмой окутан.
«Да будет свет!» – и вот явился Ньютон.
Но сатана не долго ждал реванша:
Пришел Эйнштейн – и стало все как раньше.

В действительности, общая теория относительности не усложнила наши представления о Мире, а упростила их. Сначала сошлюсь на Эйнштейна. Простите за длинную цитату:

«... трудности, возникающие в процессе развития науки, вынуждают нашу теорию становиться все более и более абстрактной [...] наша постоянная конечная цель — все лучшее и лучшее понимание реальности. К логической цепи, связывающей теорию и наблюдение, прибавляются новые звенья. Чтобы очистить путь, ведущий от теории к эксперименту, *от ненужных и искусственных допущений* (выделено нами), чтобы охватить все более обширную область физики, мы должны делать цепь все длиннее и длиннее. Чем проще и фундаментальнее становятся наши допущения, тем сложнее математическое орудие нашего рассуждения; путь от теории к наблюдению становится длиннее, тоньше и сложнее. Хотя это и звучит парадоксально, но мы можем сказать: современная физика проще, чем старая физика, и поэтому она кажется более трудной и запутанной. Чем проще наша картина внешнего мира и чем больше фактов она охватывает, тем сильнее отражает она в наших умах гармонию Вселенной» («Эволюция физики», собрание научных трудов А. Эйнштейна. Т. 4. С. 492–493).

Не имея возможности подробно прокомментировать и снабдить примерами все приведенные утверждения, остановимся только на словах, выделенных нами курсивом. Одно из несомненных упрощений — уменьшение числа не следующих ниоткуда утверждений. Вспомните знаменитый опыт Галилея. Бросая с Пизанской башни предметы разной массы, Галилей обнаружил, что все они достигают земли за одно

и то же время, равное $(2h/g)^{1/2}$, где g — ускорение силы тяжести, а h — высота башни. Опыт позволяет определить величину ускорения силы тяжести: $g = 981,56 \text{ см/с}^2$ (это значение соответствует результатам современного эксперимента; точное значение g зависит от точки на земной поверхности; мы привели значение g на широте Москвы и на уровне моря).

Почему все тела имеют одинаковое ускорение силы тяжести? Для ответа на этот вопрос вычислим g . Сила притяжения тела массы m к Земле равна GmM/R^2 , где M — масса Земли, R — ее радиус, а $G = 6,67 \cdot 10^{-8} \text{ см}^3/\text{г} \cdot \text{с}^2$ — гравитационная постоянная. Воспользовавшись уравнением движения Ньютона и приравняв силу притяжения произведению массы тела m на ускорение g , находим

$$g = \frac{GM}{R^2}.$$

Формула не содержит массы тела. Значит, мы доказали, что ускорение силы тяжести не зависит от массы тела m ? Не совсем. Естественен вопрос: «Почему инертная масса (входящая в произведение массы на ускорение) и масса, которая входит в закон всемирного тяготения, то есть в выражение для силы притяжения, тождественны?»

Конечно, есть возможность уйти от ответа, сказав, что закон всемирного тяготения — первичный закон; он не нуждается в выводе; он утверждает, что сила притяжения пропорциональна произведению масс — тех самых, которые определяют инерцию тела. Это опасный путь, ведущий к застою.

Иногда приходится ограничиваться подобными ответами. Чаще всего тогда, когда наше понимание сути явления не позволяет получить реальный ответ на вопрос: почему? Хороший пример — спектр масс элементарных частиц. Пока (надеюсь!) приходится принимать его как данность.

Общая теория относительности показала, что у тела нет двух масс. Есть только одна. Она и определяет геометрию пространства вокруг тела. А отличие этой геометрии от евклидовой дает возможность описать притяжение тел друг к другу — *вывести* закон всемирного тяготения. Когда скорости и гравитационные силы не слишком велики, уравнения общей теории относительности переходят в знакомые уравнения классической механики. Вот тогда — за пределами общей теории относительности — и возникает вопрос: «Равны ли инертная и гравитационная массы?»

Нет и не может быть теорий, свободных от экспериментальной проверки. Эксперименты по проверке выводов общей теории относительности проводились с предельно возможной точностью. Результат таков: «... ускорение тела, свободно падающего в гравитационном поле, не зависит ни от его массы, ни от свойств вещества, из которого тело состоит» (Физическая энциклопедия, 1992. Т. 3. С. 51). Это означает,

что инертная и гравитационная массы равны, а точность последнего утверждения тем больше, чем больше масса притягивающего тела. Согласно экспериментам и наблюдениям вблизи Земли разность инертной и гравитационной масс не может превосходить 10^{-8} массы тела, вблизи Солнца — 10^{-12} массы тела. Если бы различие было больше, то эксперимент бы это обнаружил.

Убедились ли Вы, что общая теория относительности упростила наши представления о мире? Во всяком случае, вопроса о равенстве инерционной и гравитационной масс больше не существует.

Общая теория относительности служит примером совершенной физической теории. Она не только объяснила наблюдавшиеся несоответствия с механикой Ньютона: добилась согласия с теорией измеренных параметров движения Меркурия; выяснила, что искривление луча света вблизи большой массы вдвое больше, чем предсказывает классическая физика, причем специально проведенные при солнечном затмении наблюдения подтвердили предсказания теории. Общая теория относительности также дала возможность на опыте обнаружить предсказанные только ею эффекты.

Пожалуй, самое сильное впечатление (на меня уж точно!) произвел эксперимент, позволивший зафиксировать зависимость хода времени от гравитационного поля. Не знаю, что поражает больше — само предсказание теории или появившаяся благодаря открытию Мессбауэра ¹⁾ возможность наблюдать и измерять чудовищно малое изменение хода времени, обязанное своим существованием разности сил притяжения к Земле на разных высотах. Расположив источник и приемник электромагнитных колебаний на разных этажах башни маяка, физикам удалось обнаружить предсказанное теорией различие испущенной и поглощенной частот — несомненное свидетельство различия хода времени.

¹⁾ Об эффекте Мессбауэра немного сказано в заключительных замечаниях.

Глава 5. ЭЛЕМЕНТАРНЫЕ ЧАСТИЦЫ

Научное и бытовое словоупотребления часто не совпадают (мы это уже подчеркивали). Словарь русского языка под редакцией С. И. Ожегова (М.: Русский язык, 1975) предлагает пять разных смысловых значений для прилагательного *элементарный*. Строго говоря, ни одно из них не соответствует тому значению, которое несет на себе слово *элементарные* в названии главы. Для контраста сравните два выражения: *элементарная математика* и *элементарные частицы*.

Между прочим, в физике элементарных частиц используется весьма сложная, современная, отнюдь не элементарная математика.

Частицы, из которых состоит все, что нас окружает: весь вещный мир, все предметы, биологические объекты, а также другие планеты Солнечной системы и далекие звезды, это *молекулы, атомы, электроны, протоны, нейтроны* . . .

Многоточие поставлено здесь не случайно; оно подчеркивает: мы перечислили далеко не все, что имеет право претендовать на звание структурной единицы материи. Например, нами не упомянуты *кварки*.

Обилие «кирпичей мироздания» несколько настораживает. Возникают естественные вопросы: как считать, из чего состоит Мир? Из чего состоим мы сами? Химия отвечает: из молекул и атомов. Физика идет глубже: из протонов, нейтронов и электронов. А современная физика элементарных частиц выяснила, что протоны и нейтроны состоят из кварков. Значит, химия и «вчерашняя» физика ошибались? А физика элементарных частиц, отменив устаревшие ответы, дает новый, единственно верный ответ? Конечно, нет! В каком-то смысле правы все.

В политике одним из руководящих принципов является: «Разделяй и властвуй!». В науке на всем пути ее развития важную роль играл принцип, лишь частично совпадающий с принципом политиков: «Разделяй и познавай!». Такой подход можно назвать *методом осколков*. Разделил на составные части, исследовал то, на что разделил, и узнал, из чего состоит . . . Однако характер осколков существенно зависит от затраченных на их получение усилий.

Интересно, что кому вспомнится: разбитая посуда или конструктор?

Таким образом, понятие *составная часть*, в какой-то мере совпадающее с понятием *элементарная частица* (но ему не тождественное!), зависит от нас, от того, сколько усилий мы готовы потратить на разборку.

При испарении вещества самостоятельно разлагаются на молекулы. Химия знает множество способов разлагать молекулы на атомы. Наименование *атом* (от греч. — неделимый) зафиксировало античное неумение разлагать атомы на составные части. Атомная физика позволила исследовать структуру атома. Выяснилось, что атом — сложная система с массой, сосредоточенной в ее ядре, вокруг которого вращаются электроны.

Ядерная физика не только установила, что и ядра — сложные системы, состоящие из протонов и нейтронов, но и научилась их расщеплять; таким образом, осуществилась мечта алхимиков о превращении одного элемента в другой.

Расщеплять атомные ядра — непростая задача. Ее решение потребовало строительства гигантских ускорителей. Вместе с тем, некоторые ядра разделяются сравнительно легко: например, под действием медленных нейтронов; подобная реакция энергетически выгодна; она легла в основу действия атомной бомбы и атомного реактора.

Совсем сложной задачей является проникновение в глубь нуклонов (протона, нейтрона). Эта задача была решена лишь во второй половине XX-го века. Изучая структуру нуклонов, физики натолкнулись на загадочную, не имеющую прецедентов ситуацию: нуклон имеет структуру, но не делится. Кварки, из которых состоят нуклоны, не наблюдаются в свободном состоянии. Похоже, нет общей точки зрения на то, причислять ли кварки к элементарным частицам.

Последние два абзаца могут привести к мысли, что выбор *структурных единиц* и *элементарных частиц* целиком определяется научным содружеством. Большая доля истины в этом есть. Однако введение термина всегда фиксирует *объективные* свойства того, что мы изучаем, в данном случае — свойства микроскопических частиц и их конгломератов.

Важно понять, что термины *структурная единица* и *элементарная частица* не совпадают.

Последнее издание «Физической энциклопедии» перечисляет несколько сот частиц, которым присвоен титул *элементарная*. К ним относятся электроны, позитроны, нейтрино, мезоны, нуклоны и т. д. Там же детально описана кварковая структура многих из этих частиц, что, казалось бы, противоречит понятию элементарности.

Еще полвека назад было известно гораздо меньше элементарных частиц. Их нынешнее обилие породило потребность попытаться проникнуть в глубь элементарных частиц. Именно таким образом были открыты кварки. Бесконечный это процесс, или в конце концов будут найдены истинно элементарные частицы, думаю, никто сказать не может. Признаться, мне хочется верить, что процесс дробления имеет

естественный предел. Возможно, невылет кварков — первое свидетельство того, что предел существует.

Ответ на вопрос: «Из чего состоит данное тело?» — определяется выбором *структурной единицы* макроскопической системы. Выбор в большой мере зависит от того, какую задачу ставит перед собой исследователь.

Для исследователя естественно желание *объяснить* наблюдаемое явление. Однако в разное время под *объяснением* понимали разное. Сейчас, если речь идет об объяснении (понимании) какого-либо свойства макроскопического тела, оно обычно строится следующим образом: прежде всего выясняют, из каких микроскопических частиц состоит тело и как они движутся, а затем устанавливают, какое конкретное движение частиц соответствует интересующему нас свойству. Не должен вызвать удивления тот факт, что для объяснения разных свойств тел приходится по-разному углубляться в строение вещества. Свойство диктует необходимую глубину проникновения и тем самым определяет, какие частицы могут быть приняты за *структурные единицы тела*. Этим определяется выбор тех частиц, которые называют в ответ на вопрос: «Из чего состоит данное тело?»

Из сказанного очевидно, что одну и ту же частицу иногда следует считать структурной единицей, иногда нет.

Приведем один пример. Исследуется поликристалл. Для выяснения многих его свойств (прочности, пластичности, электро- и теплопроводности) достаточно знать, что он состоит из кристаллитов, знать, как они расположены друг относительно друга (есть текстура или ее нет), что из себя представляют кристаллические прослойки. Кристаллит при таком подходе — структурная единица поликристалла.

Ответ на вопрос: «Какова природа наблюдаемой электропроводности?» — требует понимания, из чего состоит кристаллит, из нейтральных атомов (тогда, скорее всего, это полупроводник) или из ионов и электронов (металл). Выяснив, скажем, что кристаллит — металл, мы, как правило, не должны углубляться в структуру ионов. Можно просто считать, что атом «потерял» свои Z валентных электронов, а ион с зарядом $+Ze$ служит для электронов источником поля сил, в котором они движутся. Этого достаточно для расчета зонной структуры (см. гл. 17), а «разрешив» ионам колебаться, мы можем вычислить электро- и теплопроводность. Итак, в данном случае структурные единицы — ионы с зарядом $+Ze$ и электроны, покинувшие атомы.

Способность металла отражать электромагнитные волны — результат наличия в нем свободных электронов. Но если мы хотим исследовать поглощение света металлом (или фотоэффект), то нам придется задуматься о строении ионов, о том, в каких состояниях находятся

электроны в составе иона. Роль структурных единиц начинают играть все электроны (не только валентные), а также ядра атомов.

Наконец, существуют ядерные эффекты, специфические для твердых тел. Например, эффект Мессбауэра — резонансное излучение и поглощение γ -квантов без отдачи ядра; распад ядра урана на осколки, которые, разлетаясь, взаимодействуют и с электронами, и с ионами. Рассматривая подобные эффекты, нельзя считать ядра структурными единицами. В таких явлениях структурные единицы — нуклоны.

В этой небольшой главе мы ни разу не употребили слова *абстракция*. Она целиком абстрактна. Конкретизация, в той мере, в которой она есть, присутствует здесь только в виде примеров.

Глава 6. ТОЖДЕСТВЕННОСТЬ И НЕРАЗЛИЧИМОСТЬ

Если речь идет о макроскопических структурах или даже небольших совокупностях атомов, таких как кристаллит в поликристалле (см. предыдущую главу), то каждая из них обладает своей особой неповторимостью. Как бы ни старались, к примеру, сделать все пули одинаковыми, опытный следователь сможет определить, какой из них был произведен смертельный выстрел. А вот элементарные частицы тождественны. Ни один электрон ничем не отличается от другого; то же можно сказать о протоне и нейтроне. Сложнее дело обстоит с атомами и молекулами. И те, и другие обладают определенной, причем жесткой, структурой. Заэкранированные от воздействия, они не будут изменяться. Если бы кто-то решил заменить «хранящуюся» молекулу на какую-то другую того же сорта (например, одну молекулу воды на другую), мы не сумели бы определить, осуществил ли он замену. То же самое с атомами. Ни один атом не отличим от своего собрата. Так, атом азота невозможно отличить от любого другого атома азота.

Правда (и именно поэтому дело обстоит несколько сложнее), и атом, и молекулу можно пометить. Например, возбудить электронную оболочку того сложного образования, которое в зависимости от состава именуется атомом или молекулой. Тем самым можно выделить какую-нибудь из этих частиц. Но не нужно думать, что метка (возбуждение) прикреплена к атому или к молекуле, как царапина на пуле, позволяющая ее идентифицировать. Если волею судьбы неподалеку от возбужденного атома окажется невозбужденный, а потом их траектории разойдутся, то мы принципиально не сможем сказать, остался ли возбужденным тот же атом или возбуждение «перебралось» на другой.

Пометить атом или молекулу можно и более вычурным способом. Многие ядра существуют в разных модификациях, именуемых *изотопами*. Они отличаются друг от друга числом нейтронов при том же числе протонов. Число протонов задает число электронов в электронной оболочке атома и тем самым большинство его химических (атомных) свойств. Переместиться нейтрону из ядра в ядро непросто, поэтому нейтроны устойчиво метят атом. Последнее позволяет с помощью изотопов следить за перемещением отдельных атомов и молекул. Так и говорят: *метод меченых атомов*. Этот метод широко применяется в разных сферах — от материаловедения до медицины.

Среди микрочастиц особенно важное и почетное место занимает электрон — необходимая составная часть любого атома. Электрон —

первая открытая (Дж. Дж. Томсоном, 1897) и наиболее изученная элементарная частица. В настоящей главе мы будем говорить об электронах, хотя многое из того, о чем будет рассказано, относится и к другим частицам.

Как и другие элементарные частицы, все электроны тождественны.

Тождественность как бы специально предназначена для создания *понятия электрон*. Используя последнее, мы не должны задумываться, идет ли речь о конкретном электроне или об электроне как абстрактном понятии. Это одно и то же. У электрона нет таких черт, которые можно отбросить, не превратив его в нечто иное.

Остановимся чуть подробнее на тождественности электронов. Начнем с позиции классической физики. Вне зависимости от того, как мы описываем электроны, с помощью формул классической физики или квантовой, они не теряют своей тождественности. Электроны, прилетевшие из космического пространства в составе космических лучей, и электроны в лучевой трубке телевизора не отличаются друг от друга. Тождественность электронов и вообще элементарных частиц редко подчеркивается в классической физике. Не потому, что это неважный факт, а потому, что воспринимается он как самоочевидный.

Ситуации, когда тождественность частиц необходимо учитывать, встречаются в классической физике нередко. Особенно это касается статистической физики. При построении теории идеальных газов ¹⁾ необходимо уметь подсчитывать физически *отличающиеся* состояния совокупности частиц газа. Вот цитата из 5-го тома курса Л. Ландау и Е. Лифшица «Статистическая физика»:

«... получим всего [столько-то] возможных распределений, среди которых, однако, есть *тождественные*, отличающиеся лишь перестановкой частиц (частицы все *одинаковы*). Число перестановок N частиц есть $N!$ ».

Некоторые слова выделены нами курсивом. Упоминание об одинаковости частиц для подчеркивания очевидности сказанного помещено авторами в скобки.

Великий Джеймс Клерк Максвелл, живший в 19-м веке, 1831–1879 гг. (как мало он прожил!), знаменит главным образом формулировкой уравнений, носящих его имя. Уравнения Максвелла создали современную электродинамику, объединили электричество, магнетизм и оптику. Но кроме того, Максвелл был одним из создателей *кинетической теории газов*. Естественно, он столкнулся с необходимостью учитывать тождественность частиц газа, и задолго до открытия структуры атома задумался: «Собственно говоря, почему все атомы одного элемента тождественны?» Действительно, все окружающие нас

¹⁾ *Идеальный газ* — типичный пример абстракции. Это газ, при описании свойств которого можно абстрагироваться от взаимодействия его частиц друг с другом (ниже мы еще будем иметь дело с этим примером абстракции).

предметы хоть незначительно, но отличаются друг от друга, а атомы почему-то тождественны. Я нашел рассуждения на эту тему в опубликованных лет 30 назад переводах его лекций для сравнительно широкой аудитории (к сожалению, у меня нет их под рукой, когда я пишу этот текст). Поражает пронизательность Максвелла. Он понял, что причина неразличимости атомов — существование *структуры*. Атомы тождественны потому, что одинаково, по какому-то неизвестному в то время закону построены. А ведь, вспомните, атом тогда считался последней структурной единицей вещества. Само слово «*атом*» означает «неделимый».

Хочу поделиться мыслью, которая, правду сказать, не имеет прямого отношения к теме рассказа.

Скорее всего, Максвелл был верующим человеком. В те времена неверующих существовало немного. Казалось бы, ему было естественно подумать: «Бог создал все атомы водорода или кислорода одинаковыми». И все... Однако такой ответ его не устраивает. Факт тождественности, как любое явление, требует объяснения. Возможно, Максвелл рассуждал примерно следующим образом: «Всякое созидание начинается с выработки плана постройки. Совокупность планов составляет законы природы. Поняв их, мы сможем постичь, как устроен и функционирует окружающий нас мир. В частности, ответить на вопрос, почему атомы одного элемента одинаковы». Однако, боюсь, что я приписываю Максвеллу свои мысли.

Хочется еще раз подчеркнуть, что квантовая механика *объяснила*, почему все атомы одинаковы. Правда, почему тождественны электроны, протоны, нейтроны, мне представляется, остается необъясненным фактом — одним из тех фактов, которые лежат в основе наших попыток (притом весьма успешных) объяснить устройство окружающего нас Мира.

Вернемся к классическим частицам. После довольно подробных разговоров об их тождественности, наверное, несколько странно прозвучит заявление о том, что классические частицы нельзя считать принципиально неразличимыми. Более того, понятия *тождественность* и *неразличимость* не следует смешивать.

Поясним последнее утверждение.

Классическая частица движется по определенной траектории. Зафиксировав ее в какой-то момент времени, мы можем непрерывно следить за ее судьбой. Из-за этого нельзя наблюдаемую классическую частицу спутать с какой-нибудь другой частицей, а уж тем более заменить одну частицу на другую. Классические частицы именно поэтому не следует считать принципиально неразличимыми, хотя они тождественны.

А вот квантовые частицы *принципиально неразличимы*. Их неразличимость — следствие отсутствия траектории, наличия у квантовой

частицы волновых свойств, существования принципа неопределенности.

Неразличимость — типично квантовое свойство.

С принципиальных позиций микромир, мир элементарных частиц, атомов и молекул проще окружающего нас мира вещей хотя бы тем, что мир вещей построен из элементов микромира. Кроме того, все микрочастицы одинаковы, тождественны. Ранее мы убедились, что создание любого, даже самого простого, понятия, например понятия «стол», требует выделения общих для всех столов свойств и пренебрежения менее существенными. Эти выделение и пренебрежение — один из типов абстрагирования (см. введение). Создавая понятие «электрон», нам не пришлось опускать какие-либо свойства, имеющиеся у электронов, чтобы все их объединить одним понятием. Оно включает в себя все свойства электрона, а каждый электрон несет на себе свойства всех электронов. Это рассуждение, конечно, не означает, что понятие «электрон» является простым или самоочевидным. Само понятие простоты достаточно сложно. Нас серьезно обучали, что «электрон так же неисчерпаем, как атом». Что это точно означает, думаю, не знает никто. И думаю, что атом сложнее электрона, а любая элементарная частица проще, чем макроскопическая конструкция из элементарных частиц, в том числе и тем, что любая элементарная частица принципиально неотличима от себе подобных.

Может показаться, что неразличимость элементарных частиц (в частности, электронов) несет лишь философскую нагрузку. Это не так. Неразличимость следует учитывать при расчетах разных величин, необходимых для описания физических свойств реально существующих объектов и результатов экспериментов над ними. Рассматривая математический аппарат квантовой механики, мы не могли обратить внимание на этот факт, поскольку ограничились одной частицей.

Рассмотрим несколько явлений, в которых принимает участие более одного электрона.

Начнем с рассеяния электрона на атоме. В атоме есть свои электроны. При расчете вероятности процесса рассеяния необходимо учесть возможность обмена между рассеивающимся электроном и электронами атома. Такие процессы называют *процессами обмена*. Важно подчеркнуть, что отличить процесс рассеяния с обменом от рассеяния без обмена принципиально невозможно. Существует единый процесс рассеяния. Зафиксировав рассеянный электрон, нельзя выяснить, он ли столкнулся с атомом или это один из электронов атома. Подобный вопрос столь же лишен смысла, как вопрос, какова координата электрона, если известен его импульс.

Вместе с тем, и при принципиальной неразличимости электронов процессы обмена необходимо учитывать.

Рассмотрим другой процесс рассеяния, напоминающий рассеяние электрона на атоме, — рассеяние мюона на атоме. Мюон очень похож на электрон; он имеет тот же заряд, но его масса приблизительно в 207 раз больше массы электрона. При рассеянии мюона на атоме отлична от нуля вероятность того, что атом захватит мюон, отдав один из своих электронов. Произойдет обмен электрона на мюон. В противоположность предыдущему случаю, это реальный процесс, который можно отличить от рассеяния мюона без его захвата атомом. Более того, можно выделить атомы, у которых один из электронов заменен мюоном, и исследовать их свойства ¹⁾.

Неразличимость электронов проявляется не только в процессах рассеяния, но и во всех процессах, в которых участвуют два или более электрона. При взаимодействии атомов друг с другом процессы обмена электронами приводят к существованию дополнительной энергии взаимодействия — *обменной энергии*, не имеющей классического аналога.

Без учета неразличимости и следствий из нее (в частности, обменной энергии) нельзя объяснить существование ферромагнетизма, антиферромагнетизма и других более сложных магнитных явлений, являющихся обменными.

Хотя формально (в уме, на бумаге) электроны можно менять местами, как бы далеко атомы ни были расположены друг от друга, необходимо помнить, что обменная энергия очень быстро спадает с расстоянием между атомами. Достаточно разнести их на расстояние, превышающее размеры атома в два-три раза, и обменная энергия практически обращается в нуль. Таким образом, она убывает значительно быстрее, чем, например, энергия электростатического притяжения или отталкивания между ионами. Последнее означает, что вычисляя энергию взаимодействия между далеко разнесенными ионами, неразличимостью электронов попросту можно пренебречь.

Неразличимость, как уже было сказано, не является специфическим свойством электронов. Все элементарные частицы (протоны, нейтроны, мюоны, все мезоны и кварки), каждая в своем классе, неразличимы: протоны не отличимы друг от друга, а один нейтрон нельзя отличить от другого. Следовательно, введение понятий «протон», «нейтрон» и т. д., как и введение понятия «электрон», не требует разделения свойств каждой отдельной элементарной частицы на учитываемые и на те, которыми пренебрегают, чтобы не вступать в противоречие с вводимым понятием. Каждый протон имеет свойства всех протонов, каждый нейтрон — всех нейтронов и т. д.

¹⁾ Атом водорода, в котором электрон заменен мюоном, получил специальное название *мюоний*. Исследование свойств мюония — интересная область физики, принадлежащая физике твердого тела, ядерной физике, а также физике элементарных частиц.

При этом в физике элементарных частиц, как и в любой области физики, примеры использования абстракции при введении новых понятий подобрать совсем нетрудно. Вот один простой пример. Протоны и нейтроны обладают многими общими свойствами, не зависящими от того, имеет частица заряд или нет. Абстрагируясь от наличия заряда у протона и от отсутствия заряда у нейтрона, вводят понятие *нуклон*, объединяющее эти две частицы.

Пожалуй, никогда в обыденной жизненной практике, объединяя предметы новым понятием, мы не знаем столь точно, какими именно свойствами жертвуем, как в случае введения понятия «нуклон». А жертва отнюдь не мала — электрический заряд. Жертва ли это? В каком-то смысле, да. Конечно, протон не лишается способности притягивать электрон. Но приходится изъясняться: «В одном из состояний нуклон обладает положительным зарядом и, следовательно, притягивает электрон». Понятие «нуклон» не всегда полезно.

Уточним смысл слова «полезно».

Число *нуклонов* в ядре определяет атомный вес, а число *протонов* — атомный номер. Обе характеристики очень важны: атомный вес — в ядерной физике, а атомный номер — в химии.

Понятию «элементарные частицы» посвящена небольшая, но, как нам кажется, важная предыдущая глава. Этот термин — хороший пример абстракции. Когда частицу называют элементарной, то, как правило, тем самым подчеркивают нежелание (а скорее, отсутствие необходимости на данный момент!) заниматься ее внутренней структурой. Мы знаем, что нуклоны состоят из кварков, но в подавляющем большинстве случаев об этом можно не думать и считать нуклоны элементарными частицами.

Последний пример еще раз подчеркивает: процесс выработки понятий, сопровождаемый абстрагированием, является творческим процессом.

Отсюда следует вывод: абстрагирование — часто вполне осознанный прием, который используется при решении определенных задач.

А вот следующее несколько шокирующее утверждение приведено только, чтобы заинтересовать тех, кто впервые сталкивается с описываемыми понятиями: *неразличимость бывает разной*¹⁾.

¹⁾ Мой соавтор посетовал, что я играю словами: если частицы неразличимы, значит неразличимы, неразличимость не может быть разной. Из дальнейшего станет ясно, что понимается под *различием неразличимостей* в данном тексте.

Глава 7. СПИН. ФЕРМИОНЫ И БОЗОНЫ

Как и принципиальная неразличимость, *отличие* неразличимостей не имеет аналога в классической физике. Оно тесно связано со *спином*. Нам предстоит познакомиться с этим понятием.

Какой образ прежде всего приходит в голову, когда произносится слово «частица»? Не знаю, как вам, а мне — маленький шарик, нечто маленькое и твердое. Конечно, это неправильно. Квантовая частица имеет и корпускулярные, и волновые черты. Как мы увидим (см. гл. 13), частице нельзя приписать определенный размер. Если же реальную частицу наделять свойствами классических, то, скорее, надо было бы представлять ее в виде материальной точки. . . Не укладывается в голове: частица вещества — точка? Нечто, вовсе не имеющее размеров? Нет, скорее уж, шарик! А как движется шарик? Помимо поступательного движения, он может вращаться, как волчок.

С вращением связан определенный момент импульса. Его величина имеет размерность $г \cdot см^2/с$. Для многих, наверное, будет неожиданным, что такую же размерность имеет знаменитая постоянная Планка, без участия которой не обходится, как мы знаем, ни одна формула квантовой физики.

Обязанный своим существованием вращению момент импульса классического твердого тела, естественно, может иметь любую величину и быть куда угодно направленным. Момент импульса или просто момент — вектор ¹⁾. Вращающийся шарик имеет момент, направленный по оси вращения. Направление оси вращения может быть произвольным. Классический шарик может вращаться вокруг любой оси, вокруг которой ему придали вращение. Если шарик не вращается, то его момент равен нулю.

Вернемся к элементарным частицам. Выяснилось, что электрон, протон и нейтрон имеют *собственный момент импульса*. Здесь очень важным является слово *собственный*. Оно означает, что вращение присуще электронам и другим элементарным частицам как их внутреннее неотъемлемое свойство. Вращение электронов невозможно прекратить, как невозможно отобрать у электрона или у протона их

¹⁾ Часто можно услышать, что момент импульса — *псевдовектор*. Это правильное уточнение. Различие между вектором и псевдовектором нетрудно проиллюстрировать с помощью зеркала и катушки ниток: направление момента — псевдовектора — определяется направлением вращения катушки, а оно не изменяется при отражении (рис. 7.1).

заряды. И величину собственного момента импульса частицы изменить нельзя. Собственный момент импульса частицы принято называть ее *спином*.

Итак, спин — одна из неотъемлемых характеристик частицы. Составляя визитные карточки элементарных частиц (электрона, протона, нейтрона, мезонов, нейтрино и т. д.), перечисляя их характеристики, мы должны назвать не только массу, заряд, но и спин.

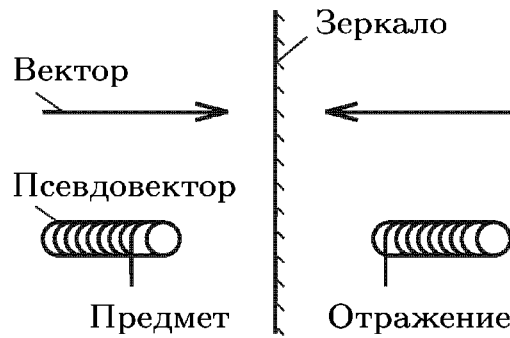


Рис. 7.1. Вектор и псевдовектор

Прежде чем привести значения спинов различных элементарных частиц, познакомим читателя с необычным свойством любого квантового момента импульса.

Так же как квантовая частица не может одновременно иметь определенные значения скорости и координаты, ее *момент импульса не может иметь определенные значения всех трех своих проекций*. Это еще одно проявление принципа неопределенности.

Квантовая механика допускает, чтобы у момента импульса были определенными модуль, то есть длина, и одна из проекций на какую-либо произвольно направленную ось.

Не знаю, как происходило знакомство с квантовой механикой у других. Меня поразило очень многое. Пожалуй, больше всего — удивительная логичность и непротиворечивость квантовой механики, казалось бы, нарушающей привычную логику. С интересом читал я статьи, касающиеся дискуссии между Эйнштейном и Бором. Кое-какими мыслями по этому поводу я поделился в главе, посвященной квантовой механике.

Одним из конкретных фактов, произведших на меня сильное впечатление, является *пространственное квантование*, о котором предстоит рассказать ¹⁾.

Чтобы рассказать о пространственном квантовании, совершим небольшое тактическое отступление. Пусть необходимо изобразить классический вектор момента импульса. Как это сделать? Сложность, пожалуй, заключается в том, что существует множество способов.

¹⁾ Более или менее серьезно изучая квантовую механику, о туннельном эффекте я был уже слышан. Поэтому этот наиболее поразительный квантовый феномен не произвел того впечатления, которого заслуживает.

Поступим следующим образом. Построим сферу с центром в точке, выбранной в качестве начала координат (она соответствует нулевому вектору), и радиусом, равным величине момента импульса. Вектор импульса, проведенный из начала координат, естественно, упирается в сферу. Чтобы зафиксировать его направление, сферы недостаточно: надо выбрать оси координат. Удобно взять три взаимно ортогональные орта и к ним привязать оси X , Y и Z . Куда их направлять, безразлично. Это наше дело. После выбора координатных осей направление вектора импульса можно задать двумя углами. Кроме того, можно просто задать проекции этого вектора на оси. Сумма их квадратов и равна квадрату величины импульса.

Весь длинный предыдущий абзац иллюстрируется рис. 7.2. К каким изменениям приведет квантовая механика? Перечислим их по порядку.

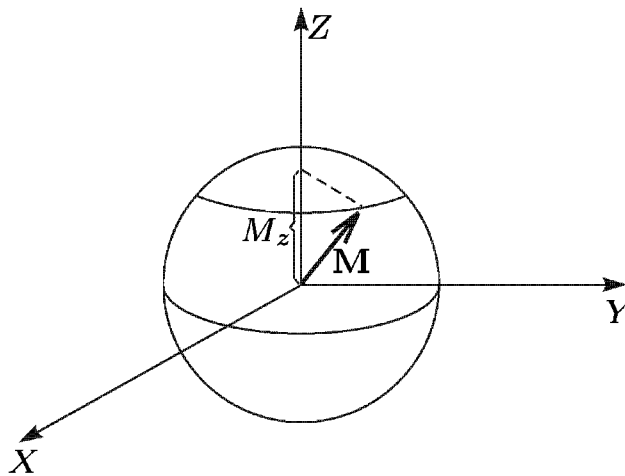


Рис. 7.2. Вектор момента количества движения M

Как мы уже говорили, определенные значения одновременно имеют длина вектора импульса и одна из его проекций, например M_z . Кроме того, длина этого вектора и его проекция могут принимать отнюдь не все значения, а только дискретные.

Проекция момента на ось Z принимает следующие значения:

$$-s\hbar, -(s-1)\hbar, \dots, (s-1)\hbar, s\hbar$$

(всего $2s + 1$ значений). Чтобы $2s + 1$ было целым числом, s должно быть либо целым, либо полуцелым. В первом случае число проекций нечетно, а во втором — четно.

Перечисленные утверждения относятся к моменту импульса любой природы. Когда речь идет об орбитальном моменте, букву s часто заменяют буквой l .

Нас главным образом будут интересовать собственные моменты импульса — спины. Прежде всего, спин электрона.

Величиной спина принято считать s , т. е. величину максимальной проекции спина в единицах \hbar . Про частицу, у которой максимальная проекция на избранную ось равна $s\hbar$, говорят, что ее спин равен s .

Если классический вектор имеет максимально возможную проекцию на какую-либо ось, то, значит, он направлен вдоль данной оси. Его длина, естественно, равна этой проекции. С квантовым спином сложнее. Когда его максимальная проекция равна $s\hbar$, квадрат длины вектора спина равен $s(s+1)\hbar^2$. Он больше квадрата максимальной проекции. Причину этого понять легко. У спина, кроме максимальной, существуют две другие проекции. Поскольку все три не могут одновременно иметь определенные значения, то, хотя проекция на избранную ось максимальна, остальные две проекции не равны нулю — они вообще не имеют определенного значения.

Хорошее наглядное представление спина — вектор, стрелочка. Он крутится вокруг оси (см. рис. 7.2). Квантовая механика запрещает спину совпадать с осью: всегда существует отличный от нуля угол его отклонения от любой избранной оси (минимальное значение последнего дается выписанными выше формулами). Проекция спина вдоль избранной оси имеет определенное значение, а проекция на плоскость всегда «крутится».

Необходимо подчеркнуть, что если изотропия ничем не нарушена, выбор направления оси (ее называют осью квантования) совершенно произволен. А вот если частицу поместить в магнитное поле, то за ось квантования естественно принять именно направление вдоль магнитного поля.

Частицы с полуцелым спином называют *фермионами* в честь великого итало-американского физика Энрико Ферми. Частицы с нулевым или целым спином называют *бозонами* в честь индийского физика Шатьендраната Бозе.

Электроны, протоны и нейтроны являются фермионами. У каждой из этих частиц спин равен $1/2$ ($s = 1/2$). Последнее означает, что спин как электрона, так и любого из нуклонов относительно произвольной оси может ориентироваться лишь двумя способами: либо по оси (проекция равна $\hbar/2$), либо против нее (проекция равна $-\hbar/2$).

У фотона спин равен 1 ($s = 1$). Фотон — бозон. Некоторые мезоны также являются бозонами. Среди них есть и частицы с нулевым спином.

Список того, что сейчас принято называть элементарными частицами, включает 350 наименований. Среди них есть и фермионы, и бозоны. Спины элементарных частиц в большинстве случаев не очень велики. Однако просматривая список элементарных частиц (см. 5-й том Физической энциклопедии, М., 1998. С. 599–601), среди фермионов я обнаружил две группы частиц со спином $11/2$, а среди бозонов — группы частиц со спином 4 . Все кварки — фермионы со спином $1/2$.

Момент импульса, обязанный своим существованием движению в пространстве, называют орбитальным, хотя, как мы знаем, квантовая механика не допускает движения по орбите. Орбитальный момент

всегда имеет целочисленное значение. Величину его максимальной проекции обозначают, как мы уже говорили, буквой l (а иногда, L).

Чем больше квантовый момент импульса, тем он ближе к классическому: с ростом s и l (или L) растет число возможных проекций на ось квантования, а длина момента приближается к величине максимальной проекции. Это проявление общего принципа соответствия, согласно которому формулы и выводы квантовой механики переходят в формулы и выводы классической ньютоновской механики, если характеристики движения соответствуют условиям применимости последней. В формулировку таких условий обязательно должна входить постоянная Планка. Условие того, что момент импульса можно описывать формулами классической механики, выглядит особенно просто: величина последнего должна во много раз превосходить \hbar .

Познакомившись с понятием спин, узнав, что электроны являются фермионами, продолжим выяснение того, как проявляет себя неразличимость.

Проще и естественнее всего изучать неразличимость на примере двух частиц. Для определенности — на примере двух электронов. Чтобы электроны не разлетелись кто куда, поместим их в поле ядра, имеющего положительный заряд, равный по величине заряду двух электронов. Легко видеть, что получился атом гелия. Поскольку ядро атома в тысячи раз тяжелее электронов, можно считать его неподвижным. Заметьте: неподвижное ядро — абстракция типа идеализации. Однако это настолько элементарный случай, что «неприлично» привлекать к нему внимание. Не учитывая движения ядра, можно вывести приближенные, но достаточно точные формулы, описывающие движение электронов (как при классическом, так и при квантовом описании). Правда, аналитически точно решить задачу о движении двух электронов в поле ядра непросто. В дальнейшем нетрудно учесть движение ядра и вычислить поправки, обязанные ему своим существованием. Подобные поправки малы, что подтверждает допустимость приближения.

При классическом описании состояние системы двух электронов определяется траекториями, по которым они движутся, при квантовом — волновой функцией, заданной в конфигурационном пространстве (если состояние стационарно, как движение по определенной замкнутой траектории, то зависимость пси-функции от времени можно исключить).

От какого числа переменных зависит пси-функция двух электронов в стационарном состоянии? В предыдущей главе говорилось, что от шести; каждый электрон обладает тремя степенями свободы, а $2 \cdot 3 = 6$. Однако мы не учитывали спина. К трем пространственным координатам каждого электрона нужно добавить значение проекции его спина. С учетом спина у каждого электрона не 3, а 4 степени

свободы! Следовательно, волновая функция двух электронов зависит от 8 переменных: $2 \cdot 4 = 8$ ¹⁾).

Вернемся к неразличимости. Переставим мысленно местами два электрона. Формально это означает, что в волновой функции, зависящей от восьми переменных, четыре переменные, относящиеся к одному электрону, надо поменять местами с четырьмя переменными, относящимися к другому электрону. Что произойдет при такой операции с волновой функцией? Оказывается, она при этом изменит знак. Смена знака — следствие того, что электроны — фермионы. Все частицы с полужелым спином ведут себя аналогично. А вот если бы мы переставляли бозоны (частицы с целым или нулевым спином), волновая функция вовсе не изменилась бы.

Естественно возникает вопрос: «Что это за неразличимость, если при перестановке частиц местами волновая функция меняет знак?» Как уже говорилось в главе, посвященной квантовой механике, процедура вычислений с помощью волновой функции физических величин, описывающих результаты экспериментов, устроена таким образом, что смена ее знака не влияет на результат. Следовательно, смена знака у пси-функции не противоречит неразличимости.

Все вышеизложенное должно пояснить утверждение о *различии неразличимостей*. Фермионы и бозоны обладают разными неразличимостями и, когда речь идет не об одной частице, разительно не похожи друг на друга.

Фермионы — индивидуалисты, а бозоны — коллективисты.

Поясним, что это значит.

Фермионы названы индивидуалистами потому, что смена знака у пси-функции при перестановке двух частиц приводит к принципу запрета, согласно которому в каждом состоянии может находиться лишь один фермион.

Бозоны — коллективисты, так как в любом состоянии может скапливаться произвольное число бозонов.

Принцип запрета, которому подчиняются фермионы, — один из важнейших принципов квантовой физики. По имени сформулировавшего его в 1924–25 гг. физика В. Паули принцип запрета называют *принципом Паули*.

Принцип Паули лежит в основе объяснения структуры атомов и периодического закона Менделеева. Действительно, представьте себе, что ничто не мешает произвольному числу электронов находиться в одном состоянии. В любом атоме все электроны оказались бы на самом нижнем энергетическом уровне. Атом, который имеет сложную структуру именно потому, что электроны занимают разные энергетические уровни, являлся бы бесформенным сгустком электронов, прижатых

¹⁾ Вас не должно смущать, что одна переменная (проекция спина) принимает только два значения: $+1/2$ и $-1/2$.

к ядру (нижний энергетический уровень «расположен» вблизи ядра). Считая таковым устройство атома, невозможно объяснить химические свойства различных атомов. А ведь они прекрасно объясняются, если при заполнении уровней учитывать требования принципа Паули. Без соблюдения последнего добавление протона к ядру, означающее переход в соседнюю клеточку таблицы Менделеева, почти ничего не изменит в строении атома. А мы знаем, сколь разительно могут отличаться атомы из соседних клеточек.

Итак, причина различия свойств атомов с разным числом электронов заключается в положении электронов в атоме. Для понимания свойств атомов надо, во-первых, выяснить, что собой представляют состояния электрона, движущегося в поле, создаваемом ядром и совокупным действием остальных электронов, а во-вторых, знать, как заполняют электроны атомные уровни.

Энергия электрона в атоме (вспомним ранее рассмотренный пример простейшего атома — атома водорода) принимает дискретные значения (электрон занимает энергетические уровни). Состояние электрона характеризуется четырьмя числами. Три из них описывают его орбитальное движение, а четвертое — проекцию спина. От проекции спина энергия электрона как правило практически не зависит. Поэтому различием состояний с противоположными спинами в хорошем приближении можно пренебречь, то есть от роли спина абстрагироваться. Найдя расположение энергетических уровней, приступим к вопросу их заполнения электронами.

Согласно принципу Паули в каждом состоянии может находиться один электрон; если считать, что от проекции спина энергия не зависит, то на каждом уровне может находиться не более двух электронов; в последнем случае электроны имеют противоположные направления спинов.

Электроны заполняют атомные уровни, руководствуясь требованием минимума энергии, но так, чтобы при этом не нарушался принцип Паули. Атом нейтрален. Значит, число электронов совпадает с числом протонов в ядре атома. Нильс Бор показал, что расположение электронов на атомных уровнях с соблюдением принципа Паули объясняет химические свойства атомов.

Принцип Паули столь важен, что хотелось бы показать, как он выводится из антисимметрии волновой функции. Пусть греческие буквы α и β обозначают состояния, а арабские числа 1 и 2 — набор четырех величин (трех координат и проекции спина) каждая. Тогда антисимметричное (по перестановке частиц) состояние двух электронов, в котором один находится в α -состоянии, а другой — в β -состоянии, имеет следующую структуру:

$$\Psi(1, 2) = \psi_\alpha(1)\psi_\beta(2) - \psi_\beta(1)\psi_\alpha(2).$$

Ясно, что если состояния совпадают ($\alpha = \beta$), то $\Psi(1, 2) \equiv 0$, то есть в одном состоянии оба электрона находиться не могут.

Теперь спросим себя: «Позволяет ли неразличимость электронов наблюдать и исследовать один электрон?»

Почти всегда отдельный электрон — *идеализация*. Ставя эксперимент, мы имеем дело с огромным коллективом электронов. Их число чаще всего не поддается подсчету. Для того чтобы изучить свойства электрона или исследовать с его помощью другой физический объект, мы пытаемся выделить отдельный электрон, прибегая не столько к экспериментальным ухищрениям, сколько к помощи теоретических представлений, основанных на принципе неразличимости, поскольку уверены, что все электроны ведут себя одинаково.

Пометить электрон нельзя. Однако можно попытаться изолировать его от других электронов, например локализовать в определенной области пространства. Тогда правомерно говорить «этот электрон». Существует метод измерения магнитного момента электрона, основанный на резонансном отклике отдельного электрона на высокочастотный сигнал. Для проведения эксперимента в приборе локализуется один (!) электрон. Говорят, что при этом лаборант должен следить, чтобы электрон «не сбежал», поскольку процесс локализации очень труден. Пройдет время, неизбежно возникнет взаимодействие с другими электронами, и данный электрон потеряет свою индивидуальность — превратится в просто электрон.

Иными словами, некоторое время электрон может быть похож на обычный предмет. Но есть и кардинальное отличие. Любой предмет (тот же стол) также может потерять свою индивидуальность, скажем, рассыпаться на щепки, превратиться в труху. Однако при этом он перестанет быть столом, а электрон, теряющий свою индивидуальность, остается электроном.

Кажется, предпринимались попытки построить квантовую электродинамику, в которой существует лишь один электрон, многократно (бесконечнократно) повторяющий себя.

Воистину абстрактная теория! К сожалению, не знаю подробностей.

Неразличимость квантовых частиц делает понятия «электрон», «протон» и т. д. удивительно похожими на математические. Число 5 — всегда 5. Можно выяснить, что именно в конкретном случае подсчитывается, однако само число 5 не зависит от того, к чему его относят. Электрон — всегда просто электрон, где бы он ни находился, в чем бы ни участвовал. Как мы уже подчеркивали, к элементарной частице неприменим вопрос: «Какая из . . .?». У нее, как и у многих математических понятий, отсутствует индивидуальность.

Объединяясь, бозоны создают более сложную частицу, и частица эта тоже бозон. А вот из фермионов могут получиться либо бозоны, либо фермионы. Если в сложную частицу объединяется четное число

фермионов, получается бозон, если нечетное — фермион. Причина проста: у частицы, состоящей из четного числа фермионов, всегда целый или равный нулю спин, а у частицы, состоящей из нечетного числа фермионов, спин полуцелый.

Прекрасный пример — два изотопа гелия.

У тяжелого, наиболее распространенного изотопа He^4 в ядре содержится 2 протона и 2 нейтрона; у легкого, весьма редкого изотопа He^3 — 2 протона и 1 нейтрон. В каждом из атомов обоих изотопов 2 электрона. Тяжелый изотоп — бозон, легкий — фермион. О макроскопических свойствах изотопов гелия достаточно подробно будет рассказано в гл. 19.

Глава 8. ФЕНОМЕНОЛОГИЯ И АБСТРАКЦИЯ

Слово «феноменология» в словарях разъяснено как философский термин. В то же время в дискуссиях на семинарах, в журнальных статьях, монографиях и даже учебниках по теоретической физике оно часто встречается, как непосредственно, так и в словосочетаниях (*феноменологическая теория, феноменологическое понятие, феноменологическая величина*). Попробуем познакомить читателя с тем, какой смысл вкладывают физики в слово «феноменология» и в производные от него. Связь феноменологии и абстракции будет очевидна. Ее нет необходимости подчеркивать.

Пожалуй, словосочетание «феноменологическая теория» встречается чаще, чем «феноменологическое понятие», «феноменологическая величина», но мы начнем именно с них.

Феноменологические понятия и величины. Каждое научное исследование (в том числе физическое) использует определенный набор понятий и представлений. Одни из них требуют объяснений, определений, другие воспринимаются как очевидные. Эти очевидные для большинства физиков понятия, именующие физические сущности, свойства которых и сам факт существования, по мнению большинства, находятся вне данного исследования, а в некоторых случаях и вообще не могут быть выведены из более глубокого уровня познания, и есть *феноменологические понятия*.

По-видимому, «феноменологическое понятие» — не всеми принятый термин. В какой-то мере некоторые из феноменологических понятий по смыслу близки к априорным (Кант). Думаю, что используя данный термин, удастся избежать выяснения их происхождения.

В настоящей главе будет приведено несколько примеров. В последующих главах мы также встретимся с феноменологическими понятиями.

Примером феноменологического понятия *докварковой эры* могут служить *элементарные частицы*, о которых мы кратко рассказали в гл. 5. Последняя фраза несет на себе дополнительную нагрузку, демонстрируя следующий факт: феноменологические понятия не составляют застывшее множество. Перемещение границы познания превращает их в объяснимые, в выводимые. Вспомним молекулу, атом, атомное ядро. Величины, характеризующие элементарные частицы (их заряды, массы, спины), также являются примерами феноменологических величин.

Примером феноменологического понятия в дорелятивистской физике является абсолютное трехмерное *пространство*. Позднее его сменило *4-пространство-время Минковского*.

Пространство и время, а также их объединение — пространство-время — могут помочь при разъяснении характерных черт феноменологических понятий.

Феноменологические понятия, особенно фундаментальные, наделяются свойствами, которые кажутся не требующими вывода или проверки, хотя, если вдуматься, они отнюдь не столь очевидны.

Пространство и время — наиболее характерные примеры последнего. Каждый из нас замечал, что время течет по-разному. То «не успеваешь оглянуться», то каких-нибудь полчаса воспринимаешь как вечность. Однако все мы *знаем* (или, точнее, знали в доэйнштейновскую эру), что время однородно и *одно на всех*. Физик, записывающий уравнение движения в своей рабочей тетради, знает, что он исследует события на той же временной оси, на которой расположены исторические события, а считать начальные условия заданными при $t = 0$ — условность: начало отсчета времени выбрано из соображений удобства.

А пространство? Опыт убеждает, что пространство неоднородно и анизотропно. Стоит повернуться, чтобы убедиться: то, что мы видим слева, совершенно не похоже на то, что мы видим справа. Переместившись на несколько шагов, мы попадаем в совершенно другую обстановку. Однако *известно*, что дело не в пространстве: наши впечатления определяются тем, что *находится в пространстве, а само пространство однородно и изотропно*. Мы столь в этом уверены, что легко соглашаемся, когда данное представление о пространстве распространяется туда, где его справедливость можно проверить только косвенно.

Переосмысление феноменологических понятий воспринимается очень остро; новые представления с трудом входят даже в сознание тех, кому, грубо говоря, «до этого нет дела». Ощущение того, что столь фундаментальные понятия, как пространство и время, — совсем не то, что мы о них думали, воспринимается как катастрофа. Многими забывается, а большинство просто не знает, что пересмотр «содержания» фундаментальных понятий не затрагивает тех областей использования, где они себя оправдали. (Этот вопрос обсуждался в гл. 4 именно на примере пространства и времени.)

Приведем еще несколько примеров фундаментальных феноменологических понятий.

Причинность. Причинность — философская категория. И при серьезном обдумывании отнюдь не простая. Наверное, все согласны с тем, что причина предшествует следствию. Вот как это сформулировано в статье «Принцип причинности» в Физической энциклопедии (Т. 4, Научное издательство «Большая Российская энциклопедия»,

М., 1994): «событие-причина предшествует событию-следствию». Проще говоря, *будущее не влияет на прошлое*. Из данной, казалось бы, совершенно абстрактной констатации, оказывается, следуют весьма нетривиальные и вполне конкретные свойства разнообразных явлений (о них еще будет говориться в этой главе).

Обратимость времени и необратимость макроскопических процессов. Подзаголовок объединяет два противоположные понятия.

Обратимость времени означает, что уравнения механики (как классической, так и квантовой) не изменяются, если произвести инверсию времени, то есть заменить t на $-t$. Правда, следует оговорить: в классические уравнения движения не должны входить силы, имеющие макроскопический характер, такие как силы трения. Кроме того, если в уравнения входит магнитное поле, то при инверсии времени нужно направить его в обратную сторону, поскольку источником магнитного поля служат токи, а они при замене t на $-t$ изменяют свое направление на противоположное.

Необратимый макроскопический процесс — это такой процесс, который может самопроизвольно протекать только в одном определенном направлении. Все неравновесные процессы необратимы. Горячее тело, окруженное холодной средой, всегда остывает. Необратимость как физический закон сформулирована в виде *второго закона термодинамики* или *закона возрастания энтропии*.

Сделаем небольшое отступление. Слово «энтропия» из физики и теории информации, где оно имеет четкий смысл, попало в журналистский обиход и приобрело некий мистический оттенок. В термодинамике энтропия — одна из многих величин, характеризующих тепловое состояние тела, строгими соотношениями связанная с другими термодинамическими величинами (например, с теплоемкостью). Энтропия положительна. Ее микроскопический смысл выясняется в статистической физике. Энтропия — мера внутренней неупорядоченности системы. При всех процессах, происходящих в замкнутой системе, она либо возрастает, либо остается постоянной. Поскольку в результате релаксационных процессов тела стремятся к равновесию, естественно, что в равновесии энтропия максимальна.

Стоит задуматься, почему при заданных внешних условиях равновесное состояние физической системы характеризуется максимальным беспорядком, а для того чтобы вывести систему из равновесного состояния, то есть «навязать порядок», необходимо приложить усилия извне.

Инвариантность уравнений классической и квантовой механики по отношению к замене t на $-t$ осложняет вывод закона возрастания энтропии. Полностью затруднения не преодолены до сих пор. Однако это не мешает физикам быть уверенными в том, что реальные макроскопические движения необратимы, а философам утверждать, что необратимость — фундаментальное свойство времени.

Такие фундаментальные понятия, как пространство, время, причинность, обратимость и необратимость, неравновесность и им подобные не исчерпывают круг феноменологических понятий.

Любая физическая работа, посвященная новому исследованию, начинается с постановки задачи. Никогда не бывает так, чтобы каждое используемое понятие, каждый термин, каждая величина требовали бы объяснения, а все утверждения — вывода. Если произвести разделение всех используемых в работе терминов, то, как правило, большинство из них предполагается известными и автору, и читателям. Все они составляют набор феноменологических понятий данной работы, а их численные характеристики суть феноменологические величины. Ясно, что в разных работах эти наборы существенно различаются: то, что в одной работе является феноменологическим понятием, в другой может быть предметом теоретического и/или экспериментального исследования и вывода.

Проиллюстрируем последний абзац двумя примерами из одной области — из *физики твердого тела*. И первое понятие, которое мы обсудим, именно *твердое тело*.

Твердое тело — одно из привычайших физических понятий. Его изучению посвящен большой раздел физики. Однако даже специалисты, пытаясь дать *определение* понятию «твердое тело», становятся в тупик. Твердые тела обладают столь различными свойствами, что теряешься, пытаясь выбрать те из них, которые следует считать *определяющими*. Определение твердого тела, принятое в механике, было приведено в гл. 1.

Если не добиваться строгости, то, наверное, можно считать твердое тело символом неизменности. Мы привыкли, что твердые тела сохраняют свою форму, не меняют ее самопроизвольно (см., однако, сказанное в гл. 1).

Твердое тело — сплошная среда. Это утверждение — не тавтология. Оно подчеркивает, что существует круг явлений, исследуя которые, нет необходимости учитывать атомное строение твердого тела. Приведем один пример: распространение тепла — теплопроводность.

Если тело достаточно велико, то его внутреннюю область можно считать бесконечным *пространством*, в котором происходят исследуемые явления. Эпитет «бесконечное» в данном случае означает лишь то, что внутри тела законы, которым подчиняются эти явления, не зависят от его границ.

Называя внутреннюю часть твердого тела *пространством*, мы не переносим на него свойства настоящего пространства. Это пространство может быть изотропным (поликристалл, аморфное тело), а может быть и анизотропным (монокристалл, текстурированный поликристалл). Важно, что его геометрия определяет геометрические свойства исследуемых явлений: в изотропном теле распространение тепла изотропно, в анизотропном — анизотропно.

Кристаллических твердых тел значительно больше, чем аморфных. Более того, истинно аморфные тела (пластмассы, стекла) в какой-то мере являются не твердыми телами, а застывшими жидкостями. Все они находятся не в стабильном, а в метастабильном состоянии и в подходящих условиях могут кристаллизоваться.

Для людей, далеких от физики, кристалл — образец, имеющий огранку. На первый взгляд, утверждение, что среди твердых тел кристаллических больше, чем аморфных, не соответствует нашему опыту. Это происходит потому, что мы чаще встречаемся с поликристаллами, чем с монокристаллами, а поликристаллы «притворяются» аморфными телами. Для физика кристалл — периодическая структура, то есть многократное тождественное повторение определенным образом расположенной группы атомов, составляющих элементарную ячейку (кристаллическая решетка построена из ячеек как из «кирпичиков»). Поскольку тождественное повторение — простейший способ построения макроскопического тела из микрочастиц, кристалл представляет собой простейший пример макроскопической структуры.

Круг явлений, при рассмотрении которых необходимо учитывать атомную, дискретную структуру твердого тела, очень широк (рассеяние рентгеновских лучей, рассеяние электронов и многие другие). Для тех из них, которые происходят не в непосредственной близости от границ кристалла, кристаллическая решетка является бесконечным пространством. Ее геометрия (и, прежде всего, периодическая структура) определяет многие черты наблюдаемых явлений (см. главы 15, 16).

Приведем еще один пример, мысль о котором навеяна тем, что твердому телу и его кристаллической решетке «присвоено» наименование пространства. В пространстве происходят события. В них принимают участие электроны, ионы, атомы примесей, меченные атомы. . . Когда интересны не столько они, сколько происходящие с ними явления, все участники событий воспринимаются как феноменологические понятия, а их характеристики — как феноменологические величины.

Большинство магнитных свойств твердых тел определяется электронами, каждый из которых является микроскопическим магнитиком. Величина его магнитного момента носит название *магнетона Бора* и как правило обозначается буквой μ . Магнетон Бора измерен и равен приблизительно $0,93 \cdot 10^{-20}$ эрг/Гс. В физике магнитных явлений твердых тел магнитный момент электрона, как и его заряд e в теории электропроводности, есть феноменологическое понятие.

В релятивистской квантовой механике магнитный момент электрона находит объяснение: $\mu = e\hbar/2mc$ (Дирак). Подстановка численных значений входящих в это выражение величин (e , m , c , \hbar) приводит к выписанному выше результату. Современная теория (квантовая электродинамика) предсказывает существование аномального магнитного момента и вычисляет $\Delta\mu$ (небольшую поправку к приведенной

формуле). Результат расчета с фантастической точностью подтвержден экспериментом.

Как в релятивистской квантовой механике, так и в квантовой электродинамике невозможно обойтись без целого набора феноменологических понятий и величин. Все величины, позволившие вычислить магнетон Бора, являются таковыми. И, как мне представляется, пока нет оснований ожидать, что скоро будет создана глубокая и всеобъемлющая теория, которая позволит понять природу мировых постоянных — e , m , c , \hbar и еще некоторых величин (например, гравитационной постоянной G , входящей в закон всемирного тяготения и равной приблизительно $6,67 \cdot 10^{-8} \text{ см}^3/\text{г} \cdot \text{с}^2$). В настоящее время ни одна физическая теория не обходится без феноменологических понятий и величин.

Феноменологические теории. Когда думаешь о большинстве феноменологических физических теорий, понимаешь, что к ним лучше подходит термин феноменологическое *описание*. Особенно это справедливо для нашего первого примера, в котором речь пойдет о теории электропроводности.

Закон Ома. Взяв за пример закон Ома, постараемся не только выявить черты феноменологических теорий, но и сформулировать границы применимости этого закона, указать, чем пришлось пожертвовать для того, чтобы иметь право выписать соотношения, составляющие его содержание.

Для многих проводников (металлов, полупроводников, электролитов и т. д.) характерна линейная связь между током J , идущим по проводнику, и приложенной к нему разностью потенциалов V . Эта линейная связь и называется законом Ома (сформулирован в 1827-м году). Коэффициент пропорциональности между V и J называют сопротивлением и часто обозначают буквой R .

Линейная зависимость тока от разности потенциалов предполагает отсутствие зависимости сопротивления от тока. Справедливо ли подобное предположение? Нет! При малом токе это так, однако с его ростом, в одних проводниках раньше, в других позже, зависимость $R(J)$ проявит себя. Величать линейную связь между током и разностью потенциалов *законом* — несомненное преувеличение.

Сопротивление, как показывают опыт и простые качественные соображения, пропорционально длине проводника L и обратно пропорционально площади его сечения ($R = \rho L/S$). Коэффициент пропорциональности ρ называют *удельным сопротивлением*. Удельное сопротивление — одна из важнейших характеристик проводника. Его значение, зависимость от температуры, давления, магнитного поля и состояния образца тщательно измеряют, сообщают в статьях, приводят в справочниках.

Казалось бы, вполне безобидное утверждение «удельное сопротивление — характеристика проводника» не всегда точно. Дело в том, что

оно подразумевает не конкретную проволочку или пластинку, а *вещество* (медь, олово, железо, ртуть, кремний или германий) со вполне определенными примесями. . . *Удельное* сопротивление введено именно для того, чтобы величина, характеризующая явление прохождения тока по проводнику, не зависела от конкретного образца, а определяла вещество проводника, давая возможность сравнивать один проводник с другим.

Выбор удельного сопротивления основной характеристикой прохождения тока по проводнику, в частности, означает утверждение о его независимости от размеров образца. Однако последнее предполагает пренебрежение поверхностным сопротивлением. Об этом не следует забывать. Особенно сейчас — в век миниатюрных схем и приборов. Следствием того, что в сопротивлении образца участвуют поверхность и близкие к ней слои вещества, является неустраняемая зависимость удельного сопротивления от размеров. Правда, если проводник достаточно велик, вклад поверхностного сопротивления оказывается столь мал, что им можно (и нужно!) пренебречь, абстрагироваться от него.

С помощью удельного сопротивления удастся переформулировать закон Ома, записать его, введя вместо тока J плотность тока \mathbf{j} , а вместо разности потенциалов V — напряженность электрического поля \mathbf{E} ¹⁾:

$$\mathbf{E} = \rho \mathbf{j}. \quad (8.1)$$

Обратим внимание на то, что величины, независимое измерение которых дает возможность найти удельное сопротивление проводника (напряженность электрического поля \mathbf{E} и плотность тока \mathbf{j}), — векторы. Их выбор и, как следствие, установление характеристики, определяющей изучаемое свойство (в данном случае, удельного сопротивления ρ), — заслуга *феноменологической теории проводимости*, одним из положений которой является закон Ома. (Напомним: удельная проводимость σ — величина, обратная удельному сопротивлению ($\sigma = 1/\rho$).)

Прохождение тока по проводнику принадлежит к тем явлениям, при феноменологическом описании которых можно считать проводник сплошной средой. А если вспомнить, что, как сказано выше, мы «решили» пренебречь поверхностным сопротивлением, то проводник следует считать бесконечным. Таким образом, соотношение (8.1) описывает линейную связь между векторами в пространстве проводника. Если проводник изотропен (например, поликристалл или аморфное тело), то выписанное соотношение строго соответствует геометрии пространства: изотропия предполагает отсутствие причин для несовпадения

¹⁾ Ток J и разность потенциалов V связаны с \mathbf{j} и \mathbf{E} следующим образом: $\mathbf{j} = J/S$, $\mathbf{E} = V/L$.

направлений векторов \mathbf{E} и \mathbf{j} . А что, если проводник анизотропен (монокристалл)?

Следствием анизотропии, на первый взгляд, должна быть зависимость удельного сопротивления ρ от направления векторов \mathbf{E} и \mathbf{j} . Однако это не так! Геометрический анализ показывает: анизотропия среды приводит к тому, что при произвольном направлении вектора \mathbf{j} вектор \mathbf{E} оказывается ему непараллельным. Связь между \mathbf{j} и \mathbf{E} остается линейной, но записывается значительно сложнее, чем в случае изотропного проводника:

$$E_i = \sum \rho_{ik} j_k. \quad (8.2)$$

Таблица из девяти величин (ρ_{xx} , ρ_{xy} и т. д.) носит название *тензора сопротивлений*. Последний является тензором второго ранга (по числу индексов) и обозначается так, как записано в последней формуле. Только знак суммы опускают. Договорились, что по дважды встречающимся индексам (в формуле (8.2) — по k от 1 до 3) подразумевается суммирование. Иногда вовсе опускают индексы, а над буквой, обозначающей тензор, ставят шляпку:

$$\mathbf{E} = \hat{\rho} \mathbf{j}. \quad (8.3)$$

Свойства тензоров так же хорошо изучены, как и свойства векторов.

Тензорная связь между линейно зависящими друг от друга векторами — следствие абстрактной геометрической теоремы, никак не связанной с природой векторных величин. Любые векторы будут определяться друг через друга подобным образом, если между ними существует линейная связь.

Феноменологическая теория использует не только геометрические факты и теоремы, но и законы природы, часто выраженные в виде феноменологических положений.

Обратимость уравнений движения относительно замены течения времени на обратное, несомненно, является одним из основных законов природы (см. выше).

Для тензора сопротивлений обратимость, дополненная некоторыми положениями статистической физики, приводит к неожиданному результату: тензор удельных сопротивлений обязан быть симметричным, т. е. должно выполняться равенство $\rho_{ik} = \rho_{ki}$. Это утверждение — частный случай общего *принципа симметрии кинетических коэффициентов*¹⁾ (Л. Онсагер).

Чтобы почувствовать нетривиальность утверждения о симметрии тензора сопротивлений, отметим, что перестановка индексов означает

¹⁾ К сожалению, я не умею *объяснить* на том уровне, который используется в данном изложении, какова связь между обратимостью времени и принципом симметрии кинетических коэффициентов.

изменение условий эксперимента. Действительно, ρ_{yx} определяет напряженность электрического поля вдоль оси y , в то время как ток течет вдоль оси x , а ρ_{xy} — напряженность электрического поля вдоль оси x , когда ток течет вдоль оси y . При этом направление осей произвольно.

Симметрия тензора второго ранга позволяет выбрать оси координат таким образом, чтобы тензор сопротивлений принял простейший вид:

$$\rho = \begin{pmatrix} \rho_1 & 0 & 0 \\ 0 & \rho_2 & 0 \\ 0 & 0 & \rho_3 \end{pmatrix}. \quad (8.4)$$

Величины ρ_1, ρ_2, ρ_3 называются *главными значениями* тензора второго ранга, а оси 1, 2, 3 — его осями. Оси 1, 2 и 3 взаимно перпендикулярны. В симметричном кристалле главные значения могут совпадать — либо все три (например, в кубическом кристалле), либо два из них (как в гексагональном кристалле).

Не хочется обойти вниманием несомненно нетривиальное следствие симметрии тензора сопротивлений: сопротивление кубического кристалла изотропно. При этом кристалл, естественно, фактически анизотропен. Только для обнаружения его анизотропии надо выбрать какой-либо другой метод — не сопротивление. Поскольку наше утверждение базируется лишь на симметрии тензора второго ранга, очевидно: изотропно будет любое свойство кубического кристалла, описываемое симметричным тензором второго ранга (его теплопроводность, диэлектрическая и магнитная проницаемости, а как частное следствие, и оптические свойства кубических кристаллов).

Нетривиальны и свойства предельно анизотропных кристаллов: хотя их элементарная ячейка — косоугольный параллелепипед, главные оси тензоров второго ранга все равно взаимно перпендикулярны.

Кроме того, на компоненты тензора сопротивлений накладывает ограничения *закон возрастания энтропии*, то есть необратимость процесса прохождения тока по проводнику. Согласно ему, главные значения тензора сопротивлений должны быть положительны. Дело в том, что возрастание энтропии, умноженное на абсолютную температуру, есть количество тепла, выделяемое в 1 с в проводнике при прохождении тока. Оно равно $\rho_1 j_1^2 + \rho_2 j_2^2 + \rho_3 j_3^2$, а эта величина *всегда* положительна, если положительны главные значения тензора сопротивления.

В случае тензора сопротивлений симметрия кинетических коэффициентов связывает между собой однородные по смыслу величины — разные компоненты одного и того же тензора. Если проводник помещен в магнитное поле (находится между полюсами магнита), то сравнивая компоненты тензора сопротивлений, необходимо переключить

полюса магнита. Соотношения симметрии Онсагера модернизируются:

$$\rho_{ik}(-\mathbf{H}) = \rho_{ki}(\mathbf{H}). \quad (8.5)$$

И наконец, нельзя не упомянуть, что наибольшую славу принципу симметрии кинетических коэффициентов принесли термоэлектрические явления, при описании которых он требует равенства величин, относящихся к разным свойствам проводника. Равенства эти были установлены У. Томсоном (лордом Кельвиным) в 1854 г., задолго до вывода обобщенного принципа симметрии кинетических коэффициентов, и получили название соотношений Томсона (ему принадлежит введение понятия абсолютной температуры — шкалы Кельвина, 1848 г.).

Подчеркнем: все сказанное о законе Ома и о сопротивлении — результат феноменологической теории проводимости — следствие геометрических теорем и ограничений (предположения о линейности, неучета роли границ образца; можно еще добавить постоянство тока). Также пришлось воспользоваться некоторыми общими положениями, или, точнее, следствиями из них. Имеются в виду принцип симметрии кинетических коэффициентов и закон возрастания энтропии. К чему нам не пришлось обращаться, так это к конкретизации механизма прохождения тока. Каков бы ни был механизм переноса заряда, если он сопровождается выделением тепла в проводнике, все сказанное будет соответствовать действительности. Оговорка «если» добавлена из-за сверхпроводников, в которых перенос заряда не сопровождается выделением тепла.

У феноменологического описания любого явления есть важная функция — выбор величин, которые могут служить характеристикой явления или свойства. В случае электропроводности роль таких величин играют удельное сопротивление ρ или обратная ему величина: $\sigma = 1/\rho$ — удельная электропроводность (в монокристалле и ρ , и σ — тензоры второго ранга). Микроскопическая теория, основанная на понимании механизма явления, должна позволить вычислить характеристику, которая была введена феноменологической теорией. Можно сформулировать так: феноменологическое описание ставит задачу для микроскопической теории.

Разобранный пример закона Ома дает возможность схематически описать структуру почти любой феноменологической теории.

Феноменологическая теория начинается с выбора величин для описания явления или свойства. После этого декларируют некоторые соотношения между выбранными величинами. Подчеркнем: именно декларируют, а не выводят. Часто аргументом для декларации служат эксперимент или следствия из него. Соотношения чаще всего просты. Иначе от них мало пользы, да и трудно их сформулировать. Нередко используют линейные соотношения между избранными величинами.

Численные значения коэффициентов в декларированных соотношениях могут и должны быть определены из эксперимента или из

более глубокой, например микроскопической, теории. Как правило, декларированные соотношения не абсолютно точны и справедливы во вполне определенной области значений параметров.

Феноменологическая теория практически всегда использует идеализацию: при ее формулировке приходится пренебрегать многими деталями явления.

Очень важная черта любой феноменологической теории — возможность привлечь общие законы природы (закон сохранения энергии, закон возрастания энтропии, изотропию и анизотропию «пространства», обратимость времени и другие).

Следующий пример феноменологической теории сильно отличается от предыдущего. Он не столько связан с выбором и введением характеристики явления, сколько демонстрирует, как действуют абстрактные принципы, например однородность времени и принцип причинности.

Дисперсионные соотношения. Если на систему действует малое по величине переменное возмущение, скажем, частоты ω , то ее реакцию называют *линейным откликом*. Коэффициент, связывающий возмущение с результатом его действия, называют *обобщенной восприимчивостью*.

Ощущается и подчеркивается абстрактность подхода: не называют отклик чего на что — отклик и все, важно, что он *линейный*, т. е. результат линейно зависит от возмущения. Эпитет *обобщенная* тоже подчеркивает отсутствие конкретности.

Скажем несколько слов о *комплексных числах*¹⁾. Физика имеет дело с реально существующими объектами. Для описания их свойств и происходящих с ними явлений физики используют весь разнообразный арсенал *математической реальности*.

Математическая реальность существенно отличается от физической. Можно, конечно, спросить: «*Действительно* ли существуют комплексные числа?». Действительного числа, квадрат которого равен -1 , естественно, нет. Язык — прекрасный свидетель. Прилагательное *действительное* имеет тот же корень, что наречие *действительно*. Но после того как введены комплексные числа и определены правила действия над ними, комплексные числа *существуют*. Они превратились в математическую реальность. Непротиворечивость — надежное свидетельство существования комплексных чисел. Если пользоваться введенными для них правилами, то производя любые необходимые действия, нельзя прийти к противоречию, как и при операциях с действительными числами.

¹⁾ См. главы 6 и 8 математической части книги. Советую не только посмотреть, но и внимательно перечитать. Тогда легче будет восприниматься содержание настоящей главы.

Без комплексных чисел было бы весьма неуютно и математикам, и физикам.

Приведем один математический пример.

Без комплексных чисел не существовало бы основной теоремы алгебры, утверждающей, что число корней алгебраического уравнения равно его степени: у квадратного уравнения два корня, у кубического — три и т. д.

И все же. . . Разве может масса частицы выражаться комплексным числом? Или ее координата? Скорость? Нет, конечно. Как же комплексные числа попали в физику? Главным образом, из-за желания упростить вычисления и добиться наглядности. Так, решая линейное дифференциальное уравнение, значительно легче использовать комплексные функции, чем ограничиваться действительными. Однако ответ (физическая величина) должен выражаться действительным числом. Теория должна включать в себя получение действительного ответа, даже если в процессе решения приходилось применять комплексные функции и числа.

Это общепринятый метод. Перечитайте гл. 3 физической части книги. Волновая функция комплексна, оператор импульса содержит мнимую единицу. Но ответы (значения физических величин и соответствующие им вероятности), конечно, являются действительными числами.

Вернемся к дисперсионным соотношениям.

Опишем несколько подробнее, как вводится обобщенная восприимчивость. Пусть две физические величины, обозначим их A и B , зависят от времени. Известно (из теоретических соображений или из эксперимента), что между ними существует линейная связь. Что это значит? Оказывается, если одну из этих величин, например A , умножить на некоторый постоянный множитель q , то и B превратится в qB .

Исследования связи между физическими величинами A и B удобно начать с рассмотрения простейшего случая, предположив, что возмущение зависит от времени периодически:

$$A = A_0 \cos(\omega t + \varphi). \quad (8.6)$$

Возмущение характеризуется тремя величинами: амплитудой A_0 , частотой ω и фазой φ . Все они — действительные числа. Если и существуют для них какие-то ограничения, то только связанные с предположением о линейности (скорее всего, амплитуда A_0 должна быть мала).

Следующий шаг также преследует соображения удобства. Удобно считать, что A — реальная часть комплексной функции $A_k = A_0 \exp[i(\omega t + \varphi)]$, рассматривать, какое возмущение вызывает A_k , то есть «вычислять» значение B_k , а потом брать реальную часть от B_k : $B = \text{Re } B_k$ — реакция на возмущение A . Глагол «вычислять»

взяты в кавычки, поскольку буквально вычислить B можно только тогда, когда теория явления, описываемого интересующим нас соотношением, полностью построена. В этом случае нет необходимости в феноменологической теории. Важно то, что вне зависимости от конкретных свойств явления и самой физической системы, к которой оно относится, опираясь на абстрактные принципы, можно сделать нетривиальные выводы.

Например, только *из однородности времени* следует выражение

$$B_k = \alpha(\omega) A_0 \exp[i(\omega t + \varphi)], \quad (8.7)$$

где $\alpha(\omega)$ и есть обобщенная восприимчивость. Обобщенная восприимчивость — комплексная функция частоты ω . Подчеркнем, что и A , и B колеблются с одинаковой частотой (следствие однородности времени). Существование у $\alpha(\omega)$ мнимой части ($\text{Im } \alpha(\omega) \neq 0$) означает, что фазы колебаний A и B не совпадают. Не конкретизируя механизм действия возмущения, обычно удается показать, что поглощенная энергия пропорциональна $\text{Im } \alpha(\omega) |A_0|^2$, откуда $\text{Im } \alpha(\omega) > 0$.

Таким образом, действительная часть обобщенной восприимчивости, $\text{Re } \alpha$, отвечает за величину реакции, а $\text{Im } \alpha$ — за сдвиг фазы колебаний и поглощенную энергию.

До сих пор мы не столько проводили абстрактные рассуждения, сколько вводили удобные обозначения. Но когда необходимые величины введены, а их физический смысл понят, начинается самое интересное. Чтобы продвинуться дальше, надо сделать еще один шаг, на первый взгляд только усложняющий всю картину. Будем считать, что $\omega = \text{Re } z$, где z — комплексная переменная. Таким образом, обобщенная восприимчивость превратилась в комплексную функцию комплексной переменной. После прочтения гл. 8 математической части Вас не должно удивить, что $\alpha(z)$ — аналитическая функция. Более того, если принять во внимание *принцип причинности* в его наиболее естественной формулировке («будущее не влияет на прошлое»), то в зависимость $B(t)$, выраженную в виде функционала от A , должны входить значения A в моменты времени, предшествующие моменту t или совпадающие с ним. Тогда выясняется, что аналитическая функция $\alpha(z)$ обладает следующим свойством: в верхней полуплоскости ($\text{Im } z > 0$) она не имеет особенностей. Из этого математического результата выводятся неожиданные следствия: действительная и мнимая части функции $\alpha(\omega) = \text{Re } \alpha + i \text{Im } \alpha$ связаны между собой интегральными соотношениями. Обратите внимание на то, что обе функции ($\text{Re } \alpha$ и $\text{Im } \alpha$) — функции настоящей частоты.

Указанные интегральные соотношения называются *дисперсионными соотношениями* или *соотношениями Крамерса–Кронига* (по имени двух физиков-теоретиков, независимо установивших их в 1926-м году для диэлектрической проницаемости).

Дисперсионные соотношения связывают между собой две реальные физические величины, описывающие разные стороны явления. Например, если обобщенная восприимчивость α является диэлектрической проницаемостью, то ее действительная часть — $\text{Re } \alpha$ — есть квадрат показателя преломления, а мнимая часть — $\text{Im } \alpha$ — пропорциональна коэффициенту затухания электромагнитной волны.

Представьте себе, что мы умеем *измерять* $\text{Im } \alpha(\omega)$, то есть поглощенную энергию (правда, при всех частотах). Тогда $\text{Re } \alpha(\omega)$ можно не измерять, а *вычислить*. Верно и обратное: знание функции $\text{Re } \alpha(\omega)$ позволяет вычислить $\text{Im } \alpha(\omega)$.

Как мы уже отметили, вычисление обобщенной восприимчивости требует знания всех процессов, происходящих в системе. Часто такое вычисление затруднительно или даже невозможно. Дисперсионные соотношения позволяют сделать некоторые важные выводы, не опираясь на точное значение обобщенной восприимчивости. Например, именно они дают возможность *доказать*, что статическая диэлектрическая проницаемость любого тела больше единицы. Это означает, что в природе нет электрических аналогов диамагнетиков.

Дисперсионные соотношения характерны не только для обобщенных восприимчивостей, описывающих свойства макроскопических систем. Они играют важную роль в ядерной физике и в физике элементарных частиц. Дисперсионные соотношения, по существу, встречаются практически в любом разделе физики.

Долгое время физики надеялись построить на дисперсионных соотношениях теорию взаимодействия между любыми элементарными частицами, не прибегая к конкретизации сил, действующих между ними. Хотя такая всеобъемлющая программа не реализовалась, дисперсионные соотношения, несомненно, являются одним из триумфов феноменологической, абстрактной (!) теории.

У феноменологических теорий есть существенный недостаток (наверное, внимательный читатель его заметил): ответы, которые получены без детального рассмотрения механизма взаимодействия частиц изучаемой системы, не полны.

Результат феноменологической теории проводимости сводится к тому, что главные значения тензора сопротивлений положительны, но в рамках этой теории мы не можем определить ни величину сопротивления, ни его температурную зависимость. Установив строго выполняющиеся дисперсионные соотношения, нельзя ответить на вопрос, в каком диапазоне частот исследуемое воздействие особенно эффективно.

Враг абстракции, а значит, и феноменологии скажет: «Не нужна мне такая теория! Я построю *микроскопическую* теорию, и из нее автоматически будут следовать все результаты феноменологической!». На это трудно возразить. Особенно, если утверждение немного изменить: «... из нее должны автоматически следовать».

Не всегда удастся построить достаточно полную микроскопическую теорию. При построении практически каждой теории используются приближения, чем-то пренебрегают, что-то опускают. Выводы феноменологической теории могут служить надежным контролем. Если результат им не удовлетворяет, ищи ошибку! Каждый опытный физик может рассказать, как сравнение полученных им результатов (в расчете или на опыте) с выводами феноменологической теории заставило его пересмотреть эти результаты. Чрезвычайно редко возникает ситуация, заставляющая пересмотреть фундаментальные принципы, на основе которых получены выводы феноменологической теории. Тот, кто знает о подобных фактах, скорее всего, знает их не по собственному опыту, а из книг по истории физики, поскольку каждый из них знаменует крупное открытие.

Глава 9. ПАРАМЕТР ПОРЯДКА — АБСТРАКТНОЕ ПОНЯТИЕ

Несмотря на то что слова «фазовый переход» в названии главы отсутствуют, она посвящена именно фазовым переходам. Теория фазовых переходов — один из наиболее впечатляющих примеров феноменологической теории.

Попытаемся продемонстрировать некоторые важные черты теории фазовых переходов. Как и в других главах, нередко придется ограничиваться лишь констатацией фактов с надеждой, что читатель поверит нам на слово. Все же мы можем обещать, что постараемся разъяснить большинство встречающихся терминов.

Начнем. Первый термин, который требует разъяснения — *фаза*, или, чтобы не путать с фазой волны, — *термодинамическая фаза*.

Естественной характеристикой любой физической системы, состоящей из макроскопического числа частиц, служит ее состав. Физические системы способны *при неизменном составе* изменять свое состояние. Всякое состояние, *качественно* отличающееся от других состояний своими физическими свойствами, называется фазой.

Важно уточнить, что понимается под термином «качественно».

Очевидные примеры качественно отличных состояний (различных фаз) — вода и пар, вода и лед.

Менее наглядный пример. Практически в любом кристалле при изменении температуры ячейка кристалла несколько изменяет свои размеры, благодаря чему изменяется удельный объем тела. Если при этом кристалл не изменяет своей симметрии, то изменение объема не воспринимается как качественное, а кристалл при разных температурах не считается находящимся в разных фазах. Описанное явление называется тепловым расширением, хотя есть тела, которые при увеличении температуры сжимаются. Однако бывает, что при определенной температуре (назовем ее *критической*) кристалл изменяет свою *симметрию*.

Изменение симметрии всегда трактуется как качественное изменение, а модификации кристалла различной симметрии — как разные фазы. Симметрия изменяется скачком. Иногда фазы с различной симметрией в пределе (при критической температуре) отличаются друг от друга, а иногда совпадают. В первом случае говорят, что произошел *фазовый переход 1-го рода*, а во втором случае — *фазовый переход 2-го рода* (см. ниже).

Для того чтобы не возникло ощущения, что разные фазы вещества обязательно имеют разную симметрию, напомним, что у пара и у жидкости симметрии одинаковые. С кристаллами тоже иногда происходят переходы, при которых симметрия не меняется. В кристаллах такие переходы называются *гомеоморфными*. Например, церий (Ce) при температуре приблизительно 300 К скачком на 16 % изменяет свой удельный объем. По обе стороны от этой температуры фазы церия различны. Проявляется это не только в различии удельного объема и, следовательно, размеров элементарной ячейки, но и в различии электронных свойств церия, хотя и выше, и ниже перехода Ce — металл.

Приведем еще несколько примеров, из которых видно, что главное отличие фаз одного вещества может состоять не только, и даже не столько в расположении его атомов (как в выше приведенных примерах), сколько в существенном изменении иных свойств.

Многие металлы при понижении температуры ниже критической переходят в сверхпроводящее состояние, то есть приобретают способность проводить ток при нулевой разности потенциалов. Состояние металла при высокой температуре — его *нормальная* фаза, а при низкой температуре — *сверхпроводящая*. Незначительное изменение кристаллической решетки в сверхпроводящей фазе по сравнению с нормальной — вторичный и не слишком важный эффект.

Вещества, в которых структурные единицы (атомы, молекулы, ионы) обладают магнитными моментами (являются микроскопическими магнетиками), при высоких температурах — *парамагнетики*, то есть их магнитные моменты направлены в пространстве случайным образом, а их тепловое движение — случайное изменение ориентации магнетиков. Как следствие, макроскопический магнитный момент тела — суммарный магнитный момент структурных единиц — равен нулю. Конечно, воздействуя магнитным полем, микромагнетики можно «заставить» выстроиться, в результате чего у тела появляется суммарный (как правило, небольшой) магнитный момент.

При температуре ниже критической магнетики спонтанно выстраиваются, повинувшись силам, действующим между ними.

Точнее, при критической температуре магнетики *начинают* выстраиваться (упорядочиваться). Чем ниже температура, тем больше порядка в расположении микроскопических магнетиков.

В зависимости от того, как выстроены магнетики при нуле температуры, низкотемпературная фаза (она существует при температуре ниже критической) называется *ферромагнитной*, *антиферромагнитной* или *ферримагнитной*. В ферромагнитной фазе все микроскопические магнетики при $T = 0$ «смотрят» в одну сторону, в антиферромагнитной фазе их направления чередуются так, что суммарный магнитный момент равен нулю. В ферримагнетиках магнетики тоже чередуются, но суммарный магнитный момент отличен от нуля.

Переход между различными фазами одного вещества, происходящий при определенных значениях внешних параметров (температуры, давления, магнитного поля и т. д.), называется *фазовым переходом*. Значение параметров, при которых происходит фазовый переход, называется *точкой фазового перехода*, или, как мы говорили выше, *критической точкой*. Фазовый переход следует отличать от постепенного превращения одного состояния в другое.

Приведем два примера изменения состояния вещества не путем фазового перехода.

С ростом температуры любой молекулярный газ постепенно ионизируется и превращается в плазму.

Увеличивая магнитное поле, можно предельно намагнитить парамагнетик, превратив его в искусственный ферромагнетик.

Приведенные ранее примеры показывают, что фазовые переходы — нередкое явление. Некоторые из них хорошо известны даже тем, кто далек от физики: превращение жидкости в газ (кипение), таяние льда (плавление), замерзание воды (кристаллизация). Их знают все.

О других фазовых переходах наслышаны.

В последнее время часто упоминается переход металлов из нормального состояния в сверхпроводящее. Упоминание перехода в сверхпроводящую фазу имеет причину: всех, причастных к технике, волновали поиски и, в конечном итоге, обнаружение высокотемпературных сверхпроводников.

Многие, по-видимому, знают о попытках создать приемлемые условия для превращения (фазового перехода) дешевого графита в драгоценный алмаз.

Переход «графит–алмаз» — один из примеров *структурных* переходов. Структурные переходы встречаются нередко, но редко привлекают внимание неспециалистов. Чтобы привлечь внимание к структурным фазовым переходам, в порядке шутки можно вспомнить, что популярный американский писатель Курт Воннегут в центр сюжета своего романа «Колыбель для кошки» поставил переход воды в особую модификацию льда — *лед-девять*.

Перечисление фазовых переходов можно продолжать и продолжать...

Несмотря на все разнообразие, фазовые переходы допускают простую классификацию. Она игнорирует конкретные физические особенности фаз вещества до и после перехода, а сосредотачивает внимание на характере самого перехода (в этом смысле классификация носит абстрактный характер). Как мы упоминали, различают фазовые переходы 1-го и 2-го рода.

Наименование двух типов фазовых переходов переходами 1-го и 2-го рода принадлежит П. Эренфесту (голландскому физику, несколько лет работавшему в Санкт-Петербурге). Он связал различие переходов

с тем, какая производная (первая или вторая) *термодинамического потенциала* терпит скачок при фазовом переходе.

Слово «термодинамика» и его производные уже несколько раз встречались в нашем тексте (и в этой главе, и в предыдущей). Рассказ о параметре порядка, и вообще о фазовых переходах, невозможно продолжить, не познакомив читателя, хотя бы в общих чертах, с термодинамикой. Кроме того, Вы увидите, что беседы об абстракции в физике были бы явно не полны, если бы в них не упоминалась термодинамика.

Термодинамикой называется наука о наиболее общих тепловых свойствах макроскопических тел. Принятый в ней подход не требует привлечения упрощенных моделей рассматриваемых явлений. Поэтому все выводы термодинамики имеют поистине универсальный характер. Самый серьезный приговор неправильной работе звучит так: «Противоречит термодинамике».

Для описания поведения макроскопических систем термодинамика выработала специальную терминологию, ввела величины, отличающиеся от тех, которыми пользуются и которые надежно служат для описания движения отдельных частиц. Часто применяемые термодинамические понятия, такие как температура, энтропия, давление, плотность, удельный объем, теплоемкость, не встречаются при описании механических движений.

Знание строения макроскопических тел позволяет выяснить, что собой представляют все термодинамические величины с точки зрения молекулярно-кинетической теории. Несколько примеров мы приведем ниже. Надо, однако, иметь в виду, что термодинамика может быть сформулирована как замкнутая наука. Нет нужды привлекать микроскопическую расшифровку используемых в ней понятий. Достаточно понимания их роли в описании свойств *макроскопических* систем.

Наиболее характерный пример — *теплота*. Было установлено, что теплота — одна из форм энергии, в которую, не нарушая закона сохранения энергии, может превращаться механическая и любая другая форма энергии. Введение теплоты как формы энергии дало возможность сформулировать один из основных принципов термодинамики.

В рамках нашего рассказа термодинамику, наверное, можно считать самым значительным успехом абстракции в физике. При этом она содержит огромное количество конкретных результатов, нашедших свое применение во всех областях техники. Например, строгое введение коэффициента полезного действия и оценка его предельного значения — заслуга термодинамики.

Термодинамика основана на трех главных принципах. Принято называть их *тремя началами термодинамики*. Три начала термодинамики фактически являются краткой формулировкой, суммирующей результаты многочисленных и не имеющих исключений экспериментов.

Начала термодинамики можно сформулировать в виде запретов.

Первое начало утверждает невозможность построения вечного двигателя 1-го рода — воображаемой машины, совершающей работу, не заимствуя энергию извне. Такая машина нарушала бы закон сохранения энергии.

Второе начало термодинамики запрещает создание вечного двигателя 2-го рода — воображаемой машины, способной уменьшить энергию теплового резервуара и целиком превратить ее в работу без каких-либо других изменений в окружающей среде. Работа такой машины приводила бы к уменьшению энтропии замкнутой системы.

Третье начало термодинамики утверждает недостижимость абсолютного нуля температуры. Оно устанавливает начало отсчета энтропии: энтропия любой равновесной системы равна нулю при $T = 0$ К. Обращается в нуль и теплоемкость любого тела при стремлении температуры к нулю. В этом причина недостижимости абсолютного нуля температуры.

Многие понятия, играющие важную роль в термодинамике, представляют собой *идеализацию*.

Приведем несколько примеров, которые позволяют понять, что идеализация использует реально существующие свойства макроскопических тел. *Замкнутая система* — система, связью которой с «внешним миром» можно пренебречь. Даже понятие *тело* — макроскопическая система, заключенная в определенный объем, — пренебрегает возможностью испарения частиц с его поверхности и адсорбции на поверхность тела. Существует и такое понятие, как *обратимость*. Термодинамические процессы *необратимы*, но если энтропия за счет изменения параметров изменяется незначительно, то процесс можно считать обратимым. Как показывает анализ, для обратимости необходимо, чтобы внешние параметры изменялись бесконечно медленно.

Замкнутая система по истечении времени всегда приходит в *равновесное состояние*. Равновесным называется такое состояние макроскопической системы, в котором отсутствуют потоки ¹⁾ массы, заряда, импульса и т. д. Процесс прихода в равновесное состояние (*релаксация*) сопровождается ростом энтропии. В равновесии энтропия максимальна.

Выводом и анализом соотношений между различными характеристиками, описывающими свойства тел именно в равновесии, и занимается термодинамика. Поэтому равновесные свойства и равновесные характеристики часто называют *термодинамическими свойствами* и *термодинамическими характеристиками*.

Ограничиться только описанием *состояний* нельзя. В термодинамике важное место принадлежит исследованию изменений

¹⁾ Это утверждение не относится к сверхпроводникам и сверхтекучим жидкостям, в которых равновесие достигается при фиксированных потоках заряда и массы (см. гл. 19).

состояний — термодинамических *процессов*. Само название «термодинамика» содержит указание на изменение, движение.

Мы понимаем, что изложение несколько декларативно. Возможно, использованные понятия станут яснее, когда мы приведем их статистический смысл. Хочется еще раз подчеркнуть, что важнейшая черта термодинамики — строгая феноменологичность, опора на фундаментальные законы природы. Именно это делает ее выводы особенно убедительными.

Разработан совершенный математический аппарат, позволяющий из основных принципов термодинамики получать большое количество конкретных и нетривиальных соотношений и утверждений. Важную роль при анализе и расчетах играют так называемые *термодинамические потенциалы*. Все термодинамические потенциалы суть внутренняя энергия тела, однако они отличаются друг от друга: каждый из них — функция своих переменных (*параметров состояния*). Примерами потенциалов являются *энергия* — функция энтропии и объема, *свободная энергия* — функция температуры и объема, *термодинамический потенциал* (в узком смысле слова) — функция давления и температуры, *энтальпия* — функция энтропии и давления.

Термодинамические величины бывают двух сортов. Одни из них (*экстенсивные*) пропорциональны объему или массе системы. К ним относятся все потенциалы и, например, энтропия. Другие (*интенсивные*) от объема и массы не зависят (температура, давление). Иногда в качестве параметров состояния удобно вводить число частиц в системе. Число частиц, естественно, экстенсивная переменная. Ей соответствует интенсивная переменная — химический потенциал (см. гл. 10). Если тело находится под воздействием внешних полей (электрического и/или магнитного), то поля — интенсивные параметры, а поляризуемости (электрическая и/или магнитная) — экстенсивные.

Мы уже отмечали, что энтропия в равновесии максимальна. Кроме того, в равновесии однородны температура и давление. Эти утверждения справедливы и тогда, когда система состоит из разных соприкасающихся друг с другом частей.

Максимальность энтропии приводит к тому, что в равновесии минимальны термодинамические потенциалы ¹⁾. В частности, если система может находиться, скажем, в двух фазах, то стабильна та из фаз,

¹⁾ Условия минимума термодинамических потенциалов требуют сравнительно тонкого анализа. Не имея возможности его привести, заметим только, что потенциалы всегда имеют минимум по отношению к микроскопическим переменным, которые характеризуют состояние. Например, небольшая деформация ячейки кристалла, несомненно, приведет к возрастанию соответствующего термодинамического потенциала. Используя термодинамические переменные, надо исходить из условия минимума энергии как функции энтропии и объема.

чей термодинамический потенциал меньше. Фаза, имеющая больший потенциал, *метастабильна*.

Термодинамика получила существенное развитие и в своем классическом (неквантовом) варианте была завершена в XIX-м—начале XX-го веков. В ее создании принимали участие многие замечательные ученые; среди них мы упомянем Л. Больцмана (1844–1906) и Дж. У. Гиббса (1839–1903), в работах которых все основные термодинамические соотношения были получены в рамках *статистической физики*. Понимание атомно-молекулярного строения вещества дало возможность выяснить природу макроскопических величин — термодинамика получила логическое завершение.

Статистическая физика позволяет придать выражению для свободной энергии компактный вид:

$$F = -T \ln Z, \quad (9.1)$$

где $Z = \sum_n \exp(-E_n/T)$. Сумму Z называют *статистической суммой*; входящая в нее величина E_n — уровни энергии тела, а T — температура тела в энергетических единицах (это тот не слишком редкий случай, когда квантовая формула *выглядит* проще, чем классическая). Знание свободной энергии позволяет вычислить остальные термодинамические потенциалы, а также другие термодинамические величины (энтропию, давление и т. д.). Надо, конечно, иметь в виду, что конкретное вычисление статистической суммы возможно только тогда, когда известны уровни энергии тела, то есть когда решена *механическая задача* о взаимодействии частиц тела. Роль формулы (9.1) — сведение термодинамической формулы к механической. Кроме того, многие результаты, описывающие связи между макроскопическими величинами, могут быть получены без знания уровней энергии, а качественные выводы — сделаны лишь на основании структуры формулы (9.1) и физического смысла входящих в нее величин.

Согласно термодинамическому определению энтропия есть взятая с обратным знаком производная по температуре от свободной энергии. При статистическом выводе формулы (9.1) было использовано понятие энтропии как меры неупорядоченности частиц, из которых состоит тело. По смыслу введения энтропия тем больше, чем больше число микроскопических состояний, осуществляющих данное макроскопическое состояние. Отсюда ясно, что для того, чтобы состояние было равновесным (в равновесном состоянии энтропия максимальна), оно должно осуществляться максимальным числом микросостояний, или, другими словами, быть максимально неупорядоченным (при фиксированных условиях).

Наиболее простой объект статистической физики — газ. Рассмотрение поведения газа при разных температурах дает возможность «ощутить», что собой представляют термодинамические величины.

Температура есть мера средней энергии, приходящейся на одну степень свободы. Давление можно вычислить, рассматривая отражение молекул газа от стенок сосуда, в котором газ находится, и вычисляя переданный стенке импульс. На примере газа удобно продемонстрировать влияние квантовых свойств частиц на поведение макроскопической системы (см. гл. 10).

Макроскопические величины — результат усреднения. Практически всегда они с большой точностью равны своим средним значениям. Однако как ни малы отклонения, они все же существуют. Все физические величины *флуктуируют*. Теория флуктуаций физических величин является главой термодинамики. Оценка величины флуктуаций может быть произведена термодинамическим путем, без ссылки на микроскопические процессы, происходящие в теле.

Особую роль играют флуктуации основных физических величин. Приведем одну формулу. Средний квадрат отклонения (флуктуации) температуры T обратно пропорционален теплоемкости тела C :

$$\langle(\Delta T)^2\rangle = \frac{T^2}{C}. \quad (9.2)$$

Теплоемкость C — экстенсивная величина; она пропорциональна числу частиц в теле N . Нетрудно показать, что $|\Delta T| \cong T/N^{1/2}$. Следовательно, отклонение температуры от своего среднего значения для тела размером в 1 см^3 не превосходит $10^{-12}T$. Казалось бы, физические величины равны своим средним значениям с такой баснословной точностью, что и говорить не о чем.

Не будем торопиться с выводами.

На замечании о флуктуациях заканчивается сделанное в общих чертах отступление о термодинамике. В начале главы мы сказали, что она будет посвящена фазовым переходам. Физика фазовых переходов — часть термодинамики.

Мы уже знаем, что фазовые переходы бывают разными.

Фазовые переходы причисляют к переходам 1-го рода, если переход из фазы в фазу сопровождается выделением (или поглощением) тепла, скачком удельного объема и других величин, которые выражаются через первые производные одного из термодинамических потенциалов.

Опишем фазовый переход 1-го рода несколько подробнее.

Пусть физическая система может существовать в виде двух различных фаз. Обозначим термодинамические потенциалы (в узком смысле слова) обеих фаз через Φ_A и Φ_B . При фиксированном давлении (для определенности равном атмосферному) термодинамические потенциалы обеих фаз — функции температуры. Предположим, что при $T = T_{\text{крит}}$ значения термодинамических потенциалов совпадают, при $T < T_{\text{крит}}$ устойчива A -фаза, а при $T > T_{\text{крит}}$ — B -фаза. Это означает, что $\Phi_A < \Phi_B$ при $T < T_{\text{крит}}$ и $\Phi_A > \Phi_B$ при $T > T_{\text{крит}}$ (рис. 9.1).

Очень важно, что по обе стороны от перехода могут существовать обе фазы. Равенство термодинамических потенциалов не означает равенства первых производных. Отсюда следует существование скачков (см. выше) и скрытой теплоты перехода (она выражается через скачок энтропии).

Для обеих фаз точка фазового перехода — обычная точка: в ней термодинамический потенциал каждой из фаз (Φ_A и Φ_B) не имеет никакой особенности как функция величин, от которых он зависит.

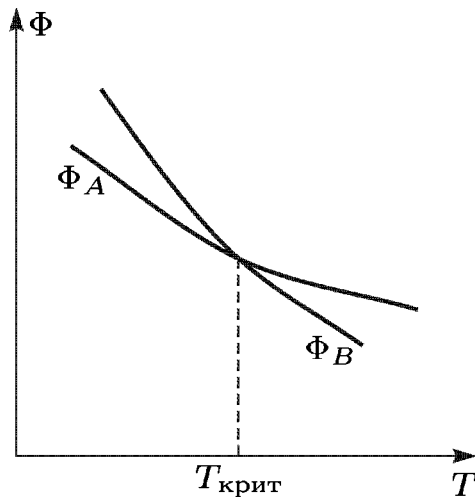


Рис. 9.1. Фазовый переход 1-го рода; Φ_A и Φ_B — термодинамические потенциалы разных фаз

Причиной перехода служит только то, что с одной стороны от $T_{\text{крит}}$ энергетически выгодней одна фаза, а с другой — другая. Фактически обе фазы никак не связаны друг с другом (лишь тем, что соответствуют разным модификациям одного и того же вещества). Поскольку каждая из фаз отделена от другой потенциальным барьером, фазовый переход может несколько задержаться. Возможны и практически обязательны перегрев низкотемпературной фазы и переохлаждение высокотемпературной. Поэтому у большинства фазовых переходов 1-го рода существуют петли гистерезиса.

Совершенно иная природа фазовых *переходов 2-го рода*. При фазовом переходе 2-го рода первые производные от термодинамического потенциала непрерывны, а *вторые* испытывают аномалии, иногда — просто скачки. Тепло при фазовом переходе 2-го рода не поглощается и не выделяется, удельный объем не меняется, но теплоемкость тела и коэффициент теплового расширения обнаруживают аномалии.

Итак, от того, начиная с какой производной термодинамического потенциала проявляется фазовый переход, зависит его наименование. Теоретически исследован фазовый переход 3-го рода (см. гл. 10). Формально даже можно ввести переходы с дробным родом. Например, И. М. Лифшиц предсказал существование в металле своеобразной аномалии, которую, используя терминологию Эренфеста, можно трактовать как переход 2,5-го рода. Подобные аномалии были обнаружены в некоторых металлах при воздействии давления.

Вернемся к фазовым переходам 2-го рода.

Одна из характерных черт фазовых переходов 2-го рода состоит в том, что и выше, и ниже точки перехода устойчива только одна из фаз. Естественно, по обе стороны от точки перехода устойчивы разные фазы. Поэтому переохлаждение и перегрев невозможны, петли гистерезиса отсутствуют. Важно, что на зависимости термодинамического потенциала от внешних параметров (температуры, давления и т. п.) фазовый переход 2-го рода проявляется как *особая точка*.

Общее свойство всех переходов 2-го рода — появление (как правило, в низкотемпературной фазе) нового свойства, причем не скачком, а плавно. Ландау обратил внимание на то, что зарождение нового свойства при фазовом переходе 2-го рода приводит к спонтанному (самопроизвольному) изменению симметрии тела. Обязательно происходит ликвидация (разрушение) одного из элементов симметрии.

Приведем несколько примеров.

При переходе в ферромагнитное состояние появляется макроскопический магнитный момент — за счет упорядочения магнитных моментов атомов. Выше температуры перехода они «смотрят» в разные стороны, а суммарный магнитный момент равен нулю. Направление, полного магнитного момента нарушает симметрию кристалла. В точке перехода магнитный момент еще равен нулю, возникла только *возможность* нарушения симметрии, но с понижением температуры он растет и достигает своего максимального значения при температуре, равной абсолютному нулю. Магнитные моменты атомов при абсолютном нуле температуры полностью упорядочены. Полный порядок!

Аналогичное явление происходит при появлении сверхпроводимости: симметрию разрушает сверхпроводящий ток. Одновременно он упорядочивает движение электронов.

Весьма наглядно изменяется симметрия в упорядочивающихся сплавах. Рассмотрим сплав CuZn, имеющий объемно-центрированную кубическую решетку. При высоких температурах сплав беспорядочен: атомы Cu и Zn с одинаковой вероятностью занимают любую позицию. Начиная с некоторой температуры, узлы решетки приобретают индивидуальность: в одних (в центрах кубов) с большей вероятностью располагаются атомы Cu, а в других (в вершинах кубов) — атомы Zn. С понижением температуры различие вероятностей возрастает, а при нуле температуры все атомы Cu располагаются в центрах кубов, а все атомы Zn — в их вершинах. Сплав полностью упорядочен.

Мерой отличия упорядоченной фазы Ландау избрал *параметр порядка*, описывающий в каждом конкретном случае то свойство, которое нарушает симметрию, имеющую место при высокой температуре. Это магнитный момент в ферромагнетике, плотность сверхпроводящих электронов в сверхпроводнике, определенным образом отнормированная вероятность в упорядочивающихся сплавах и т. д.

Параметр порядка всегда можно выбрать так, чтобы он изменялся от нуля (в точке фазового перехода 2-го рода) до единицы (при абсолютном нуле температуры) (рис. 9.2). В высокотемпературной фазе параметр порядка *тождественно* равен нулю.

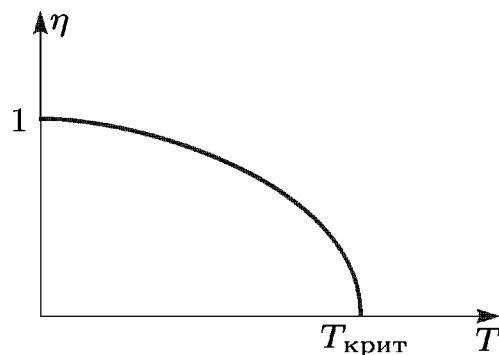


Рис. 9.2. Температурная зависимость параметра порядка

Исходя из того соображения, что вблизи температуры перехода параметр порядка мал, усиленного *предположением*, что термодинамический потенциал разложим по целым степеням малого параметра порядка, Ландау построил качественно справедливую теорию фазовых переходов 2-го рода. Во время ее создания (1937 г.) предположение о разложимости казалось очень убедительным. Однако вскоре обнаружилось, что не все фазовые переходы 2-го рода хорошо описываются теорией Ландау. Построение непротиворечивой теории многие годы было дразнящей задачей для физиков-теоретиков. Ландау ушел из научной жизни (автомобильная катастрофа в 1962 г.), не зная «ответ», который был получен в конце 60-х—начале 70-х годов усилиями многих физиков-теоретиков, хотя Нобелевская премия за создание теории критических явлений (1982 г.) присуждена лишь *одному* из них — американскому физику Кеннету Вильсону. Рассказ о получении «ответа» — в русле нашего дальнейшего повествования.

Теория Ландау, сыгравшая фундаментальную роль в понимании природы фазовых переходов 2-го рода и вообще критических явлений, служит примером *абстрактной феноменологической теории*, в которой главное внимание сознательно уделено не конкретному описанию механизма фазового превращения, а появлению и роли абстрактной характеристики — параметра порядка, описывающего симметрию системы.

Не все фазовые переходы 2-го рода хорошо описываются теорией Ландау. Чтобы пояснить, в чем дело, сравним кривые зависимости от температуры теплоемкости C сверхпроводника и гелия (рис. 9.3, *а*, *б* соответственно). На обоих рисунках температурный интервал выбран так, что он содержит точку фазового перехода. При этом, если переход в сверхпроводящее состояние теория Ландау описывает правильно, то переход гелия в сверхтекучее состояние — неправильно. Она не объясняет отчетливо видной аномалии вблизи точки перехода. Термин *аномалия* означает здесь признание того, что поведение, согласующееся с теорией Ландау, является нормальным. Строго говоря, как мы увидим, поведение гелия при сверхтекучем переходе более нормально, чем поведение сверхпроводников.

Теория Ландау фазовых переходов 2-го рода построена без учета флуктуаций. Казалось бы, малость термодинамических флуктуаций оправдывает подобное упрощение, однако это далеко не всегда так. Если мы утверждаем, что некоторая величина мала, то обязательно должны уточнить, по сравнению с чем. Это относится и к параметру порядка. Его флуктуациями можно пренебречь только в том случае, когда они малы по сравнению с самим параметром порядка. В точке перехода параметр порядка обращается в нуль. Следовательно, всегда есть такая область температур вблизи точки перехода, где нельзя обойтись без учета флуктуаций.

Как же быть с теорией Ландау? Из параметров, характеризующих физическую систему вблизи перехода, можно построить безразмерную величину. Ее часто обозначают буквой G и называют числом Гинзбурга ¹⁾. У различных веществ числа Гинзбурга бывают существенно разными. Причина последнего принципиально ясна: величина G зависит от сил, действующих между частицами, принимающими участие в фазовом переходе.

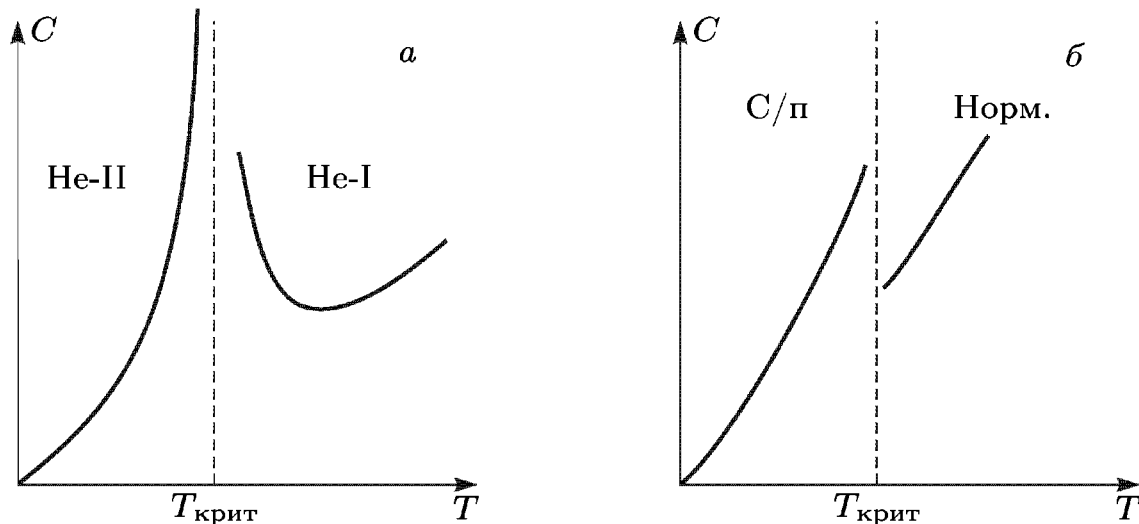


Рис. 9.3. Температурная зависимость теплоемкости вблизи а) сверхтекучего и б) сверхпроводящего переходов

Если число Гинзбурга значительно меньше единицы, для теории Ландау есть место: она справедлива в интервале температур, определяемом неравенством

$$G \ll \frac{|T - T_{\text{крит}}|}{T_{\text{крит}}} \ll 1. \quad (9.3)$$

Если G относительно велико, для теории Ландау места нет, и теорию фазового перехода необходимо строить с учетом флуктуаций. Отметим, что в непосредственной близости от температуры перехода $T_{\text{крит}}$ теория Ландау всегда отказывается.

Температурная область, в которой флуктуации играют важную роль, обязательно существует. Ее так и называют — *флуктуационной*. Правда в сверхпроводниках, сегнетоэлектриках, при многих ориентационных магнитных переходах и т. д. флуктуационная область очень узка; теория Ландау хорошо описывает подобные фазовые переходы. При сверхпроводящем переходе флуктуационная область столь узка, что ее практически невозможно обнаружить.

Анализ флуктуаций параметра порядка позволил не только обнаружить факт существования флуктуационной области, но и выяснить,

¹⁾ А.А. Абрикосов и В.Л. Гинзбург в 2003 г. получили Нобелевскую премию за работы по сверхпроводимости.

как должны вести себя физические величины при приближении к точке фазового перехода.

Во флуктуационной области термодинамические характеристики системы ведут себя на удивление однотипно. Поведение термодинамических характеристик разных систем описывается степенными законами, причем соответствующие степени — не целые числа. Их называют *критическими показателями*.

Критические показатели обладают свойством универсальности: они не зависят от физической природы вещества, и даже от физической природы фазового перехода. Критические показатели определяются только тем, как переход изменяет симметрию системы. По самой своей сути *критические индексы — абстрактные характеристики фазового перехода*.

Для того чтобы подчеркнуть роль абстракции в теории фазовых переходов 2-го рода, мысленно проследим цепочку абстрактных понятий:

параметр порядка \rightarrow *симметрия* \rightarrow *критические показатели*.

Вычисление критических показателей завершает термодинамическую теорию фазовых переходов 2-го рода. Их расчет требует использования весьма абстрактных рассуждений и методов и является очень сложным. Вот если бы мы жили в четырехмерном мире, осуществить вычисления было бы гораздо проще. Физики-теоретики пошли по весьма нестандартному пути, который получил название *эпсилон-разложения* (ε -разложения), где, как всегда, ε — символ малого параметра. Что же в данном случае играет его роль? Приведем цитату из Физической энциклопедии: «Эпсилон-разложение — метод приближенного вычисления *критических показателей* в статистической физике с помощью разложения . . . по степеням малого параметра $\varepsilon = 4 - d$ », где $d = 3$ — истинная размерность нашего пространства. Курсив в цитате означает, что в Энциклопедии есть статья «Критические показатели». В небольшой статье (в ней всего две таблицы) собраны основные критические показатели. Многие физики считают, что их вычисление завершило теорию фазовых переходов.

Подчеркнем еще раз: весь подход, особенно его последний этап, основан на использовании истинной абстракции. Действительно, разложение по $\varepsilon = 4 - 3$ означает, во-первых, «материализацию» четырехмерного пространства, а во-вторых, и всех промежуточных пространств с дробными размерностями между $d = 4$ и $d = 3$.

Глава 10. ЯВЛЯЕТСЯ ЛИ ИДЕАЛИЗАЦИЯ АБСТРАКЦИЕЙ?

Когда я по настойчивому призыву своего соавтора впервые задумался об абстрактных понятиях в физике, на ум, естественно, приходили многие известные мне примеры теорий и подходов. Однако почти каждый раз, подумав, я понимал, что речь идет об идеализации. Формально, конечно, идеализация — один из видов абстракции. Вместе с тем, идеализация используется в физике (да и в любой другой науке) столь часто, что перечислить примеры невозможно, так как нет возможности остановиться, прекратить перечисление. Один пример все же приведем. Он характерен.

Идеальный газ, естественно, является идеализацией. Его частицы не взаимодействуют друг с другом. Именно поэтому его именуют *идеальным*. Конечно, между частицами реальных газов существует взаимодействие; они сталкиваются, длина свободного пробега каждой частицы не равна бесконечности, как предполагают при описании свойств идеального газа. Вместе с тем, при описании многих свойств реальных газов (например, их термодинамических, равновесных свойств) взаимодействием между частицами можно пренебречь. Поправки, которые возникнут, если мы учтем взаимодействие, будут чаще всего очень малы. Их можно опустить и использовать абстрактную идеализацию — идеальный газ.

Вместе с тем, полного пренебрежения взаимодействием идеализация не допускает. Расчет термодинамических характеристик идеального газа включает в себя выяснение того, как зависят теплоемкость и другие характеристики от температуры. Состояние системы частиц характеризуется общей для всего газа температурой, только если газ находится в состоянии равновесия. Обычно молчаливо предполагают, что газ находится в равновесии. Однако как он может прийти в состояние равновесия, если его частицы не взаимодействуют между собой? Это абсолютно невозможно. В отсутствие взаимодействия газ никогда не пришел бы в состояние равновесия. Чтобы определить время, необходимое для установления равновесия, необходимо учесть взаимодействие между частицами газа. Обычно не только рассчитывая, но и на практике, например при измерении термодинамических характеристик газов, предполагают, что время установления прошло и газ успел прийти в состояние равновесия.

Ощутили ли Вы, что проблема, упомянутая в предыдущем абзаце, имеет вполне практическое значение? Действительно, производя изме-

рения, приходится изменять внешние условия. Иначе вместо кривой получишь одну точку. Меняя внешние условия, необходимо помнить, что делать это надо сравнительно медленно, чтобы газ успел каждый раз прийти в равновесие. Обычно время установления равновесия очень мало. Однако иногда оно столь велико, что требование проводить измерение в равновесных условиях делает эксперимент весьма утомительным, растягивая его на многие часы.

Нередко газ находится в условиях, в которых внешние причины поддерживают его неравновесное состояние (например, когда стенки сосуда находятся при разных температурах и по газу распространяются потоки тепла, или на концах трубы, по которой течет газ, устанавливают разное давление). Мерой теплопроводности служит коэффициент теплопроводности, а движением газа по трубе «руководит» его вязкость. Обе характеристики (коэффициенты теплопроводности и вязкости) определяются длиной свободного пробега частиц газа. Они могут быть вычислены только при учете взаимодействия между частицами.

Если пренебречь взаимодействием между частицами нельзя, но оно мало, часто используют компромиссный термин — *почти идеальный газ*. Когда можно, взаимодействием пренебрегают, когда пренебречь нельзя, о нем вспоминают.

Манипулирование взаимодействием (учитываем—не учитываем), конечно, формально превращает исследуемый объект (реальный газ) в абстрактный: многое в него привнесено в попытке получить ясный и достаточно точный ответ на поставленный вопрос. Однако ведь это обычное дело. Измеряя некоторую величину, например размер комнаты, и используя тот или иной метод измерения, а главное, помня, для чего мы это делаем, мы ограничиваем себя разной точностью. Тем не менее не становится же от этого измеряемый объект абстракцией?!

Лучше не быть формалистом и отличать идеализацию от абстракции.

С другой стороны, став на такую точку зрения, можно задуматься: а есть ли в теоретической физике по-настоящему, истинно абстрактные понятия? Здесь слова «по-настоящему, истинно» взяты не из философского, а из житейского словаря. Надеемся, нам удастся показать, что в физике нередко и с пользой применяют то, чего нет в природе, что нельзя именовать иначе, чем абстракцией. Усилим: истинной абстракцией.

Оказывается, тот самый идеальный газ, который, как мы только что убедились, служит прекрасным примером идеализации, в определенных условиях может стать хорошим примером истинной абстракции. Дело в том, что в некоторых условиях обычный идеальный газ превращается в квантовый, однако подобных условий в природе не существует. Их нужно либо искусственно создавать, либо искать, причем там, где, казалось бы, исследуемый объект совсем не похож на газ.

В параграфе, посвященном фермионам и бозонам, мы обещали рассказать, что собой представляют их макроскопические совокупности — идеальные газы фермионов и бозонов. Пришло время выполнить обещание.

Газов в природе много. Все они существуют при не слишком низкой температуре. Большинство газов конденсируется в жидкость при достаточно высокой температуре. Например, пары воды — при $100\text{ }^\circ\text{C}$, т. е. $\sim 400\text{ K}$. У легких, а тем более инертных газов температура конденсации существенно ниже. Рекордсмен среди них — гелий. Он превращается в жидкость при температуре около 4 K . Однако для нас и эта температура — не слишком низкая.

Понятия «фермионы» и «бозоны» фиксируют различие квантовых свойств частиц (см. гл. 7). Подчеркивая, что один из интересующих нас газов — газ фермионов, а другой — бозонов, мы должны быть готовыми к тому, что различия в их поведении будут проявляться тогда, когда смогут проявиться квантовые свойства газов.

Все частицы — либо фермионы, либо бозоны. В этом смысле, любой газ — либо газ фермионов, либо газ бозонов, либо смесь фермионов и бозонов. Однако заметно ли отличаются газы фермионов и бозонов друг от друга? Оказывается, при высоких температурах практически не отличаются. Их свойства не зависят от того, фермионные они или бозонные¹⁾. На каждую степень свободы приходится $(1/2)kT$ энергии газа, так что каждая его частица имеет среднюю энергию, равную $(3/2)kT$. Следовательно, среднее значение квадрата импульса $\langle p^2 \rangle = 3mkT$, где m — масса частицы. Если число частиц в 1 см^3 равно n , то среднее расстояние между ними составляет $n^{-1/3}$. Приведенные нами формулы помогут понять, когда, при какой температуре T , должны проявляться квантовые свойства идеального газа, т. е. когда для описания его свойств заведомо нельзя использовать формулы классической физики.

Рассмотрим отдельную частицу газа. Вокруг нее находятся другие частицы. На долю одной частицы приходится область пустого пространства с линейными размерами порядка $n^{-1/3}$. Можно сказать иначе: каждая частица локализована в области с размерами порядка $n^{-1/3}$. Значит, $\Delta x \approx n^{-1/3}$. Следовательно, на основании соотношений неопределенности, $\Delta p \geq \hbar n^{1/3}$. Теперь задумаемся. Если неопределенность импульса превышает его среднее значение, можно ли пользоваться классическими представлениями? Ответ очевиден: конечно, нет!

¹⁾ Мы интересуемся теми свойствами, относительно которых можно считать частицы газов элементарными (см. гл. 5 второй части книги). Строго говоря, это означает, что частицы газа только перемещаются (у каждой из них три степени свободы плюс спиновая; от ориентации спина энергия не зависит).

Комбинируя выписанные формулы, нетрудно убедиться, что при температуре порядка, и тем более ниже температуры $T_{\text{кван}} \approx \hbar^2 n^{2/3} / mk$ классическими формулами пользоваться нельзя. Однако $T_{\text{кван}}$ столь низка, что ни один газ при ней не останется газом, а все газы, кроме гелия, давно (при значительно более высокой температуре) превратятся в твердые тела — кристаллы. Именно поэтому мы утверждали, что газы фермионов и бозонов могут служить примером истинной абстракции. Нужно, правда, сказать, что сейчас делаются весьма успешные попытки искусственно создать газ бозонов (так называемый бозе-газ). Исследованием свойств искусственно созданного бозе-газа занимаются в нескольких лабораториях ¹⁾.

Что касается ферми-газа, то создавать его нет необходимости. Каждый металлический предмет содержит в себе естественный ферми-газ. Это электроны проводимости. Подставив вместо m в формулу для $T_{\text{кван}}$ массу электрона, а вместо n — Z/a^3 (где Z — валентность металла, a — размер его кристаллической ячейки), мы убедимся, что для электронов металла $T_{\text{кван}} \approx 10^5$ К, т. е. если можно рассматривать их как газ, то только как квантовый. Уверен, что последние слова вызовут скептическое замечание:

— Какой же это газ? Да еще идеальный?! Ведь это электроны. Они заряжены, отталкиваются друг от друга, притягиваются к ионам! А Вы говорите — газ!

Не следует торопиться. Пока только скажем, что рассматривая электроны проводимости как квантовый газ фермионов, удалось понять многие свойства реальных металлов.

Перед нами встает трудная задача. Авторам предстоит описать, а читателям понять, чем отличаются газы фермионов и бозонов. Несколько последующих абзацев представляют собой попытку познакомиться с выводами квантовой статистической физики — науки о больших (макроскопических) коллективах квантовых частиц.

Представим себе пространство импульсов. Это несложно. Построим три обычные взаимно перпендикулярные оси координат, только отложим на них не x , y и z , а компоненты импульса — p_x , p_y , p_z . Точка в таком пространстве обозначает импульс частицы. Теперь разобьем все импульсное пространство на бесконечно малые ячейки, объем которых составляет $dp_x dp_y dp_z$. Будем (мысленно, конечно) измерять импульс у частиц газа, заполняющего единичный объем (1 см^3), и в зависимости от полученного результата ставить точку в одну из ячеек импульсного пространства. Покоящиеся частицы попадут в начало координат. Всего в кубическом сантиметре n частиц. Как они будут распределены по ячейкам импульсного пространства? Если газ находится в равновесии,

¹⁾ За экспериментальное исследование свойств бозе-газа группа американских физиков получила Нобелевскую премию по физике за 2001-й год.

то распределение по импульсному пространству зависит только от энергии $\varepsilon = p^2/2m$ частиц и описывается одной из двух функций:

для газа бозонов — функцией Бозе–Эйнштейна

$$F_{\text{БЭ}} = \frac{1}{\exp(\varepsilon - \zeta)/kT - 1},$$

для газа фермионов — функцией Ферми–Дирака

$$F_{\text{ФД}} = \frac{1}{\exp(\varepsilon - \zeta)/kT + 1}.$$

Как мы видим, в функции Бозе–Эйнштейна и Ферми–Дирака параметрами входят две величины: температура газа T и химический потенциал ζ . Последний уже упоминался при рассказе о термодинамике. Выясним физический смысл БЭ- или ФД-функции. Зная их, можно записать, сколько частиц находится в ячейке импульсного пространства:

$$dn = \frac{g}{(2\pi\hbar)^3} F_{\text{БЭ(ФД)}} dp_x dp_y dp_z,$$

где $g = 2s + 1$ — число проекций спина частицы на избранную ось (для электрона $g = 2$). Количество частиц во всех ячейках импульсного пространства ¹⁾ всегда должно составлять n . При изменении температуры число частиц не меняется. Именно это и обеспечивает химический потенциал ζ . Он служит *нормировочной величиной*. Естественно, химический потенциал зависит от числа частиц газа в 1 см^3 и от температуры.

Нетрудно показать, что при высокой температуре химический потенциал отрицателен, а по модулю превышает kT . Приглядитесь к формулам, описывающим ФД- и БЭ-функции. Если $|\zeta| \gg kT$ и $\zeta < 0$, экспонента значительно больше единицы. Следовательно, с хорошей точностью единицей вовсе можно пренебречь. Теперь ФД- и БЭ-функции не отличаются друг от друга, что естественно, так как при высоких температурах любой газ описывается формулами классической физики, которая не отличает фермионы от бозонов:

$$F_{\text{ФД}} \approx F_{\text{БЭ}} \approx F_{\text{М}}, \quad F_{\text{М}} \approx \exp\left[-\frac{\varepsilon - \zeta}{kT}\right].$$

Буква М в индексе высокотемпературной, классической функции написана в честь Максвелла, который не только сформулировал уравнения электродинамики, но и построил кинетическую теорию газов.

Осталось сказать, какая температура является достаточно высокой для того, чтобы газ можно (нужно!) было считать классическим.

¹⁾ Суммирование в данном случае означает интегрирование по всему импульсному пространству.

Думаю, вы не удивитесь, узнав, что высокой следует считать температуру, существенно превышающую $T_{\text{кван}}$.

Теперь можно сравнить ферми- и бозе-газы в условиях, когда их различие максимально — при $T \ll T_{\text{кван}}$.

Начнем с ферми-газа. На рис. 10.1 показана $F_{\text{ФД}}$ при $T \ll T_{\text{кван}}$ и $T = 0$. Она абсолютно не похожа на максвелловскую, изображенную на рис. 10.2. Такое впечатление, что нечто мешает частицам занять низкоэнергетические состояния. Действительно, мешает *принцип запрета Паули*. Как мы уже отмечали, он действует сильнее, чем стремление системы иметь при абсолютном нуле, т. е. в основном состоянии, наименьшую энергию. Функцию Ферми–Дирака при $T = 0$ называют фермиевской ступенькой. Все состояния с энергией, меньше некоторой энергии ε_F , заняты, а с более высокой — свободны. Граничную энергию называют энергией Ферми. По порядку величины она равна $kT_{\text{кван}}$.

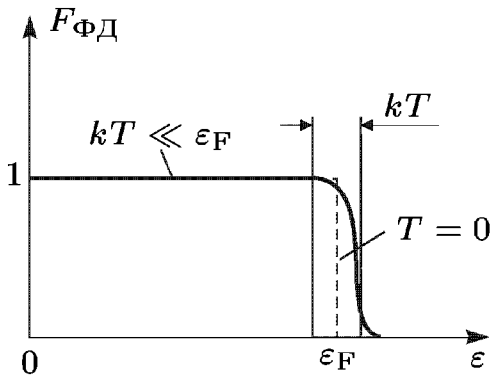


Рис. 10.1. Функция распределения Ферми–Дирака

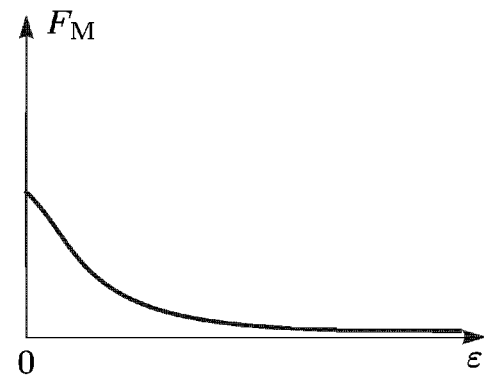


Рис. 10.2. Функция распределения Максвелла

С понижением температуры фермионы постепенно перераспределяются по энергетическим состояниям. Перераспределение не сопровождается никакими катаклизмами. В частности, химический потенциал — монотонно возрастающая функция температуры. При $T \approx T_{\text{кван}}$ химический потенциал меняет знак. При $T = 0$ он равен энергии Ферми.

Бозе-газ по своим свойствам разительно отличается от ферми-газа (конечно, при $T \leq T_{\text{кван}}$). Легко убедиться, что при температуре порядка $T_{\text{кван}}$ химический потенциал бозе-газа, как и у ферми-газа, обращается в нуль. Однако изменить знак он не может¹⁾. Возникает странная ситуация. Складывается впечатление, что по мере понижения температуры частицы исчезают. Куда они деваются? Парадокс разрешил Эйнштейн. Он обратил внимание на то, что пересчитывая частицы, т. е. вычисляя нормировочный интеграл, мы учитываем только движущиеся частицы — те, у которых $\varepsilon \neq 0$. Вместе с тем, ниже некоторой температуры (порядка $T_{\text{кван}}$) конечное число частиц «кон-

¹⁾ Формальная причина такова: при $\zeta > 0$ нормировочный интеграл расходится (обращается в бесконечность).

денсироваться» в состояние с $\varepsilon = 0$. Слово «конденсируется» мы взяли в кавычки, поскольку с конденсацией газа в жидкость это явление не имеет ничего общего. И все же его называют *Бозе–Эйнштейновской конденсацией*. Бозе–Эйнштейновская конденсация — фазовый переход, причем очень мягкий. По классификации Эренфеста это переход 3-го рода. Когда он происходит, не только не выделяется и не поглощается тепло, но и теплоемкость непрерывна. Скачок имеет производная по температуре от теплоемкости.

При абсолютном нуле температуры все частицы имеют энергию, равную нулю. Поскольку у частиц определен (нулевой) импульс, соотношения неопределенности требуют их полной делокализации. Координаты частиц не имеют никакого определенного значения.

Сейчас исследования Бозе–Эйнштейновской конденсации — одна из модных задач. Недавно экспериментаторы достигли существенного успеха. К сожалению, рассказать об этих исследованиях в рамках настоящей книги невозможно. Однако причину, по которой идеальный газ фермионов может служить хорошей моделью для описания электронных свойств металлов, кажется, удастся объяснить.

Мы высказали предположение о том, что считать электроны в металле идеальным газом мешает электростатическое (кулоновское) взаимодействие. Действительно, между электронами, а также между электронами и ионами действуют силы, обратно пропорциональные квадрату расстояния между ними. Энергия взаимодействия между частицами обратно пропорциональна расстоянию r между ними. Обозначим ее $U_{\text{взаим}}$: $U_{\text{взаим}} = e^2/r$. Мы уже знаем, что среднее расстояние между частицами — порядка $n^{-1/3}$. Поэтому среднее значение энергии взаимодействия в расчете на одну частицу приблизительно равно $e^2 n^{1/3}$. Среднее значение кинетической энергии нетрудно вычислить, воспользовавшись фермиевской ступенькой. В расчете на одну частицу $E_{\text{кин}} \approx E_{\text{кван}} \approx kT_{\text{кван}}$, т.е. $E_{\text{кин}} \approx \hbar^2 n^{2/3}/m$. Мы видим, что в газе фермионов при низких температурах и потенциальная, и кинетическая энергии растут с ростом числа частиц газа n в кубическом сантиметре (в классическом газе средняя кинетическая энергия частицы вовсе не зависит от n). Вспомните формулы! Кинетическая энергия растет с n быстрее, чем энергия взаимодействия. Это означает, что в достаточно плотном газе (при $n \gg (me^2/\hbar^2)^3$) энергия взаимодействия значительно меньше кинетической. Следовательно, в хорошем приближении энергией взаимодействия можно пренебречь и считать газ идеальным (очередной пример идеализации). Поскольку $\hbar^2/me^2 \approx 10^{-8}$ см, нет ничего удивительного в том, что идеальный газ фермионов служит хорошей моделью для описания электронных свойств металлов.

Глава 11. В ПОИСКАХ ИСТИННОЙ АБСТРАКЦИИ. ВСЕГДА ЛИ ПРОСТРАНСТВО ТРЕХМЕРНО?

Использование истинной абстракции, то есть того, чего попросту не существует в природе, оказывается, может приносить пользу, особенно в тех областях теоретической физики, в которых понимание затруднено сложностью реально существующих систем. Часто исследуют заведомо нереальную систему — более простую, чем реальная, и позволяющую разобраться в вопросе, на который ищут ответ. Подчеркнем, что речь не идет об идеализации: нереальная система не является приближением к реальной.

Анализ не существующих реально систем и ситуаций оправдан только тем, что учит исследователя. Он либо способствует развитию аппарата исследования, который может пригодиться для решения реальной задачи, либо полученный результат подсказывает, каков может быть результат при решении реальной задачи. В последнем случае перенесение выводов от нереальной к реальной системе требует не только большой осторожности, но и настоящего искусства. Им редко кто владеет в совершенстве. Использование такого приема нельзя назвать повсеместным. Однако про физиков-теоретиков часто говорят, что они ищут кошелек не там, где потеряли, а там, где светлее. Эту дразнилку, правда, обычно применяют в том случае, когда критика не устраивает используемая абстракция.

Использование истинной абстракции, требующее, как мы сказали, большого теор-физического искусства, всегда внушает естественное опасение, но в случае успеха, т.е. признания коллегами, особенно ценится.

Конечно, необходимы примеры. Их очень непросто приводить, так как они требуют детализации и подробностей, выходящих за пределы возможностей настоящей книги. Все же попробуем привести хотя бы один пример.

Исследование свойств систем, описываемых модельными гамилтонианами. Так принято называть исследования макроскопических систем частиц, реальное, истинное взаимодействие между которыми заменено удобным взаимодействием, позволяющим провести исчерпывающий анализ.

Например, предполагается, что сила взаимодействия между частицами не убывает с удалением частиц друг от друга, а вовсе не зависит

от расстояния между ними. Оказывается, в ряде случаев удается не только получить ответы на вопросы, на которые для реальных частиц ответы неизвестны, но и перенести полученные результаты из модели (абстракции) в реальность.

D-мерное пространство. Открывая журнал, в котором печатаются статьи по теоретической физике, обязательно увидишь большую букву D (реже — маленькую d), не всегда даже поясняемую в тексте: читатель привык, что D (или d) — размерность пространства. Конечно, и автор, и читатель знают, что мы живем в трехмерном мире ($D = 3$), но иногда удобно, нужно, необходимо рассматривать объекты, размерность которых отлична от трех ($D \neq 3$). Бывает, что ничего, кроме идеализации, в этом нет. Например, если электроны прижаты электрическим полем к поверхности кристалла или жидкого гелия, а температура очень низка, фактически ниже одного абсолютного градуса, то свойства электронов описываются формулами, справедливыми для двухмерного электронного газа.

Экспериментальные исследования двухмерного электронного газа оказались столь богатыми открытиями, что заслуживают сравнительно подробного рассказа.

Сначала сделаем одно общее замечание.

Может показаться, что экспериментальная физика на протяжении всего своего существования занималась реально имеющимися в природе объектами. Галилей сбрасывал с Пизанской башни различные предметы, и в его рассуждениях важную роль играли связанные вместе и летящие поодиночке кирпичи; Ньютону открыть закон всемирного тяготения, по легенде, помогло свалившееся на голову яблоко.

Однако...

Мир пронизан радиоволнами, которые не только названием, а в большой степени и происхождением обязаны радио, изобретенному Поповым и Маркони. Наверное, именно с радиоволн главным предметом исследования физиков стали рукотворные объекты.

В отличие от геологов, физики в большинстве случаев исследовали не природные кристаллы, а искусственно созданные образцы, на приготовление которых часто затрачивалось больше усилий, чем непосредственно на измерения.

Гелий существует на Земле в газообразном состоянии, а наиболее интересные его свойства были обнаружены у жидкого гелия, полученного лишь в 1908-м году голландским физиком Каммерлинг-Оннесом.

Энергичные протоны и электроны прилетают на Землю в составе космических лучей, но планомерное и принесшее замечательные результаты исследование энергичных элементарных частиц началось и успешно продолжается с помощью искусственно созданных ускорителей заряженных частиц.

Приведенный неполный перечень примеров призван показать, что физика активно использует искусственно созданные объекты исследо-

вания или искусственно созданные условия для существующих объектов. В конце XX-го века стали появляться новые объекты, значительно дальше отстоящие от естественных, имеющих в природе. Фактически это детали будущих приборов. Во всяком случае, они специально создаются для нужд техники. К подобным объектам принадлежат, например, элементы деталей компьютеров — печатные схемы.

Изучение физических свойств самых миниатюрных из деталей приборов и устройств породило новый раздел физики конденсированного состояния. Его назвали *мезоскопикой*. Этот раздел занимает промежуточное место между макро- и микроскопической физикой.

Двухмерный электронный газ. Исследование качества и свойств полевого транзистора ¹⁾ сделало двухмерный электронный газ одним из самых модных объектов изучения. Важный элемент полевого транзистора — электрическое поле E , создающее на границе с полупроводником треугольную потенциальную яму (рис. 11.1).

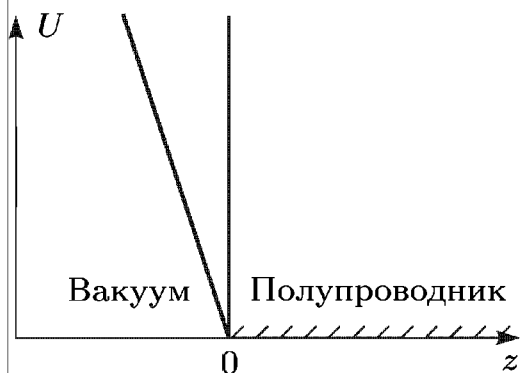


Рис. 11.1 Треугольная потенциальная яма ($E < 0$)

В этой яме скапливаются электроны, которые *могут создать* двухмерный газ. Яма, конечно, условное наименование, поскольку в плоскости поверхности образца электроны движутся как свободные частицы. Правильнее было бы сказать «траншея», «овраг».

Неопределенность утверждения (могут создать) вызвана следующим. Будет ли газ двухмерным, квазидвухмерным или попросту трехмерным, зависит от соотношения нескольких параметров.

Как и в любой потенциальной, в треугольной яме энергия движения электрона перпендикулярно поверхности принимает дискретные значения. Используем идеализацию: будем считать, что температура равна абсолютному нулю. Если число электронов, приходящихся на единицу поверхности, таково, что все они (с учетом принципа Паули, конечно) помещаются ниже второго уровня, то электронный газ над поверхностью полупроводника окажется *двухмерным*. Если электроны заполняют несколько уровней, газ *квазидвухмерен*, а если на многих уровнях есть электроны, то это обыкновенный *трехмерный газ*.

Согласно принципу Паули двухмерный, как и трехмерный, электронный газ заполняет при нуле температуры все энергетические уровни, расположенные ниже энергии Ферми. Таким образом, условие существования двухмерного газа приобретает форму неравенства:

¹⁾ Транзистор — полупроводниковый прибор, способный усиливать электрические сигналы. Транзисторный эффект был открыт в 1946 г. американскими физиками Дж. Бардиным, У. Браттейном и У. Шокли.

энергия Ферми должна быть меньше, чем расстояние между двумя наиболее низкими уровнями.

Не знаю, стал ли бы двухмерный газ столь модным объектом исследования, если бы эксперименты с полевым транзистором не привели к открытию в 1982-м году немецким физиком фон-Клитцингом удивительно интересного явления — квантового эффекта Холла.

Эффект Холла — один из гальваномангнитных эффектов. Гальваномангнитные эффекты (явления) представляют собой совокупность эффектов (явлений), наблюдаемых, когда проводник, по которому течет ток, помещен в магнитное поле. Для наблюдения эффекта Холла проводник с током помещают в перпендикулярное току магнитное поле, в результате чего появляется электрическое поле (поле Холла), перпендикулярное и току, и магнитному полю. Эффект Холла был открыт еще в 1879 г. американским физиком Эдвином Гербертом Холлом. Он обусловлен силой Лоренца, действующей на движущийся электрон. Во многих случаях коэффициент между полем Холла и плотностью тока линейно зависит от магнитного поля. Он носит название холловского сопротивления и обозначается R_H . Часто с хорошей точностью справедливо следующее равенство: $R_H = H/nes$, где n — число электронов проводимости в единице объема проводника.

В полевом транзисторе (т. е., по сути, в двухмерном электронном газе) тоже наблюдается эффект Холла. Для этого необходимо направить магнитное поле перпендикулярно плоскости поверхности образца. При относительно высоких температурах эффект Холла в двухмерном случае мало отличается от обычного эффекта Холла. Однако в области температур ~ 1 К зависимость холловского сопротивления от магнитного поля теряет линейный характер; возникает ряд характерных ступенек, причем величина R_H на них с высокой точностью выражается комбинацией мировых констант:

$$R_H = \frac{1}{N} \frac{2\pi\hbar}{e^2c},$$

где $N = 1, 2, 3, \dots$ — целые числа.

Выписанное соотношение выполняется с такой удивительной точностью, что с его помощью можно уточнить многие атомные константы, а величину R_H при $N = 1$ воспринимать как эталон сопротивления. Появление чередующихся плато у холловского сопротивления получило название *квантового эффекта Холла*, а когда выяснилось, что плато существует и при некоторых дробных значениях N , в него добавили прилагательное *целочисленный*, чтобы отличать от *дробного квантового эффекта Холла*.

Естественен вопрос, какое отношение квантовый эффект Холла имеет к абстракции. Во-первых, речь о нем зашла в связи с исследованием низкоразмерных систем. Во-вторых, что кажется нам более важным, квантование обладает удивительной способностью создавать

тождественные, неотличимые объекты. Значения RH не зависят от существенных (казалось бы) условий эксперимента: размеров образца, влияния границ, степени совершенства образца и т. п.

Тождественность значений RH заставляет вспомнить тождественность атомов и молекул.

Тождественность элементарных частиц (в частности, электронов) воспринимается как естественное, присущее им свойство. Она является элементом феноменологической констатации, на основе которой создается сегодняшняя научная картина физического мира. Однако тождественность атомов и молекул, имеющих сложную внутреннюю структуру, есть результат квантования. Ее природу мы понимаем. Закрадывается мысль: когда мы глубже проникнем в природу элементарных частиц, мы поймем природу и их тождественности. Одновременно, возможно, выяснится, почему заряд у всех элементарных частиц одинаков, а массы столь различны. Сегодня, похоже, рассуждения на тему о природе тождественности элементарных частиц — просто спекуляция.

Напомним, что мы занялись пространствами с размерностью D , отличающейся от трех, в поисках истинной абстракции. Обратившись к 2-мерному пространству, мы отвлеклись и стали рассматривать свойства вполне реального 2-мерного газа электронов. Тем самым мы не вышли за пределы идеализации.

Продолжим поиски. Даже просто проглядывая современные работы по теоретической физике, нетрудно заметить, что размерность D — не всегда целое число. Возникает естественная реакция: «Это заведомая абстракция! Как иначе воспринимать дробную размерность?» Однако приглядевшись внимательнее, мы поймем, что дробной размерностью описывают реально существующие в природе объекты. Не обошлось, конечно, и без идеализации.

Фракталы и фрактальная размерность. Начнем с определения, взятого из Физической энциклопедии (М., 1998. Т. 5. С. 371): «Фракталы — множества с крайне нерегулярной разветвленной или изрезанной структурой. Термин «фракталы» введен Б. Мандельбротом, хотя подобные объекты изучались в математике с конца 19-го века». Пропустим несколько абзацев, в которых приведены примеры математического построения фракталов, и заключим цитату фразами, непосредственно относящимися к нашей теме: «Примерами естественных фракталов являются береговая линия материков и островов, снежинки, броуновские кривые и т. д. . . Большой интерес к фракталам в физической литературе связан с тем, что фракталы возникают в реальных физических задачах, причем в типичных, а не экзотических ситуациях.»

Когда речь идет о естественных фракталах, всегда можно в принципе определить размерность геометрической структуры, которая претендует на «право» быть фракталом. Самая вычурная снежинка имеет

своей границей 2-мерную поверхность, а береговая черта на географической карте — 1-мерная линия, даже если она изображает Норвегию с ее беспорядочно расположенными фиордами разных размеров. Однако при этом фрактальный характер поверхностей и линий только затрудняет исследование свойств содержащих фракталы физических объектов. При рассмотрении усредненных характеристик объектов, содержащих фракталы или представляющих собой фракталы, эти объекты можно характеризовать эффективной размерностью. Ее так и называют — *фрактальной размерностью*. Математики создали аппарат, позволяющий по свойствам фракталов находить число, равное фрактальной размерности. Оно характеризует степень разветвленности и изрезанности. Фрактальная размерность всегда больше истинной размерности фрактала.

Приведем пример нахождения фрактальной размерности.

Рассмотрим кристаллическую решетку, ребра которой, связывающие узлы, могут быть либо проводниками, либо изоляторами. Распределение проводников и изоляторов случайно. Вероятность проводящей связи равна x . Узлы считаются связанными друг с другом, если между ними по цепочке проводящих ребер может пройти электрический ток. В этом случае проводящие ребра и связанные ими узлы образуют кластер.

Если вероятность x мала, то проводящие связи, как правило, далеки друг от друга, а превалирующие кластеры имеют небольшое количество узлов и ребер. Однако с ростом x размеры кластеров увеличиваются, и начиная с некоторого критического значения $x = x_{\text{крит}}$, кластер пронизывает всю решетку. Его естественно назвать критическим кластером. Теория позволяет вычислить значение $x_{\text{крит}}$. Для плоской решетки, состоящей из шестиугольников, $x_{\text{крит}} = 0,6527$; из квадратов — $x_{\text{крит}} = 0,5$; из треугольников — $x_{\text{крит}} = 0,3473$. В случае трехмерных решеток: для простой кубической решетки $x_{\text{крит}} = 0,25$; для объемно центрированной — $x_{\text{крит}} = 0,18$; для гранецентрированной — $x_{\text{крит}} = 0,12$.

Существенная информация о критическом кластере заключена в его фрактальной размерности. Значение последней зависит от истинной размерности решетки. Критический кластер в двухмерной решетке имеет фрактальную размерность, равную 1,9, а в трехмерной — 2,5. Два последние числа содержат важные данные о критических кластерах. Кроме того, фрактальная размерность дает возможность сравнить разные фракталы и тем самым ощутить, что собой представляет каждый из них.

Рассмотрим *последний пример, относящийся к совершенно другой области*. Наверное, я не вправе его приводить, так как предмет знаю только понаслышке. Речь пойдет о космогонии — сценарии возникновения Вселенной. Слышал, что есть теории, согласно которым наш мир

родился многомерным ($D \gg 4$; $4 = 3 + 1$, где 1 — время), а в процессе развития большая часть размерностей «вымерзла». Создан даже специальный термин — компактизация. Он означает, что только вдоль 4-х из всех координат могут существовать большие интервалы между двумя точками на них. Остальные координаты замыкаются. Демонстрационную модель придумать легко. Сверните лист бумаги в трубку и считайте ее бесконечно длинной. По одной координате (вдоль оси трубки) можно двигаться, как в обычном пространстве, а по другой не существует расстояний, больших $2\pi R$, где R — радиус трубки.

Откровенно говоря, для меня это такая же абстракция, как и для Вас, дорогой читатель. Может быть, даже большая, чем для Вас. Многие из того, что Вам кажется абстракцией, мною воспринимается как прием. Однако в данном случае голова кругом идет. Вместе с тем, хочется верить, что если наука поставила перед собой задачу понять, как развивалась Вселенная от Большого взрыва до сегодняшних дней, она ее решит. Возможно, многомерный этап на ранней стадии существования Вселенной — один из реальных этапов ее развития. Если это так, пройдет время, теория будет общепризнана, и компактизованные координаты не станут восприниматься абстрактными. «В процессе познания чудо превращается в тривиальность», — утверждал Эйнштейн.

Физика существенно отличается от математики не только объектом исследования. В этих науках пользуются различными способами описания, по-разному задают вопросы и записывают ответы на них. Физики и математики говорят на разных языках, как правило, неплохо понимая друг друга, и уж во всяком случае уверенно используя достижения друг друга.

Обе науки не могут обойтись без чисел, причем в математике число имеет более абстрактный характер, чем в физике. Геометрию по традиции относят к математике, хотя фактически она должна относиться к физике. В геометрии числа имеют размерность. Нельзя сложить объем тела, выраженный числом, с числом, обозначающим длину тела. Для математики это исключение, а физик практически всегда имеет дело с размерными числами. Они, как правило, являются целью и вычислений, и измерений. Чем именно вызван интерес к размерности в рассказе об абстракции и ее применениях в физике, надеемся, скоро станет понятным.

Говорят, исключения подтверждают правила. Встречаются ситуации, в которых наибольший интерес вызывают не столько сами размерные физические величины, полученные в эксперименте или в расчете, сколько безразмерные числа, извлеченные из непосредственных результатов.

Приведем две такие ситуации.

Пусть изучается какое-либо явление. Исследователи пытаются понять, от чего оно зависит и убеждаются, что от некоторого воздействия явление вовсе не зависит. Результат, возможно, весьма важен, а числом не выражается. Правда, и это утверждение тоже можно выразить числом. Обозначим характеристику явления буквой a , а воздействие — буквой b . Независимость a от b означает, что $da/db \equiv 0$. Появилось число — нуль. Конечно, оно не добавило ничего нового. Вместе с тем, употребление вместо слов числа наводит на вопрос: «Какова точность полученного результата?» Аккуратные исследователи обязательно сообщают, за какую точность они отвечают.

Ограничимся случаями, когда каждый полученный физический результат можно выразить числом. Конечно, из чисел можно при необходимости построить кривую, наборы кривых, поверхность. Иногда важно не столько значение измеренной величины, сколько вид получившейся кривой. Например, измеряя некоторую величину, экспериментатор заносил результаты в тетрадь (или они автоматически

заносились в память компьютера). По полученным данным была построена кривая, разглядывая которую, экспериментатор обнаружил, что она очень напоминает трилистник. Он радостно сообщает (руководителю, товарищу по работе, научной общественности — статьей): в исследуемом объекте есть ось симметрии третьего порядка. Таким образом, результатом эксперимента становится не первичное измерение, а число 3 — число лепестков у кривой.

Перечитывая последний абзац, я живо представил себе экспериментатора, о котором в нем говорится. Скорее всего, строя кривую по результатам измерений, он использовал декартову систему координат, а получив кривую с тремя максимумами, перестроил ее в полярных координатах, понимая, что в такой системе координат легче увидеть трилистник. Однако трилистник ли это изображен? Один лепесток может несколько отличаться от двух других, а некоторые точки выпадать. Строя кривую, экспериментатор оставит их чуть в стороне. Неужели он пытается подправить данные? Нет! Наоборот: прежде чем продемонстрировать свою кривую, он тщательно продумает природу ошибок и возможную их величину. Каждая точка снабжается крестиком, изображающим величину ошибок; построенная кривая не может пройти не по крестикам. Зачем же экспериментатор так стремится построить трилистник? Для того чтобы «сделать по виду несвязанные скучные числа согласованными, осветить их теорией»¹⁾. Если теории нет, ее заменяет интуиция — предвестник теории.

Обратите внимание на то, какую важную роль в описанном, правда, гипотетическом, но, мне думается, реалистическом эпизоде играет идеализация — один из типов абстракции.

Вспомним эпизод, не созданный нашим воображением, а взятый из истории физики XX-го века. Исследуя прохождение пучка атомов Ag через неоднородное магнитное поле, немецкие физики Отто Штерн и Вальтер Герлах обнаружили на чувствительной пластинке две полосы, два следа осажденных атомов. Расстояние между полосками позволяет определить величину магнитного момента атома серебра. Однако не менее важно само число 2 (количество полосок). Именно оно — главный и несомненно фундаментальный результат, который позволил убедиться в том, что спин электрона равен $1/2$. Кроме того, этот опыт дал возможность непосредственно увидеть пространственное квантование (см. гл. 7). Описанный эксперимент вошел в историю физики как опыт Штерна–Герлаха. Статья с его описанием опубликована в 1922 г. (см. приложение к настоящей главе).

¹⁾ А. Эйнштейн, Л. Инфельд; в связи с работой Н. Бора о расположении спектральных линий атома водорода — работой, показавшей, что «новая механика» может объяснить экспериментальные данные (*Эйнштейн А. Эволюция физики. Собрание научных трудов.* — М., 1967. Т. IV. С. 525.).

Два опыта (гипотетический и действительно имевший место) были описаны для того, чтобы показать — наиболее существенным результатом эксперимента может оказаться безразмерное число. Правда, безразмерные числа (3 — в первом случае, 2 — во втором) возникли, строго говоря, не непосредственно в эксперименте, а в процессе обдумывания полученных данных. Непосредственные результаты, зафиксированные приборами, несомненно, имели размерность.

Пора заняться тем, чему посвящена настоящая глава, то есть размерностями физических величин.

Следующая фраза может показаться несколько неожиданной. Несмотря на обилие физических величин и принятых единиц их измерения, существует всего три независимые размерности. Основные (фундаментальные) размерности можно выбирать по-разному. Нам привычны: *сантиметр* (см) — *размерность длины*, *секунда* (с) — *размерность времени*, *грамм* (г) — *размерность массы*.

Приведем рассуждение, напоминающее доказательство того, что можно ограничиться тремя фундаментальными размерностями. Любое физическое явление, в конечном счете, проявляется в движении материальных объектов. Для описания движения (для кинематики) достаточно уметь описать зависимость пройденного пути от времени. Для этого нужны единица длины (см) и единица времени (с). Причиной движения (точнее, изменения движения) служит сила. Динамика требует единицы силы. Ее можно взять как одну из основных. Однако удобнее в качестве третьей фундаментальной размерности выбрать размерность массы (г). Тогда размерность силы нетрудно определить, воспользовавшись уравнением движения Ньютона (см. гл. 1). Если сила обозначается буквой f , то ее размерность будет обозначаться как $[f]$. Итак, $[f] = \text{г} \cdot \text{см}/\text{с}^2$ (квадратные скобки мы и в дальнейшем будем использовать для обозначения размерности). Размерность энергии в этих единицах имеет вид $\text{г} \cdot \text{см}^2/\text{с}^2$.

Несколько примеров помогут понять, как можно размерность фактически любой физической величины свести к трем фундаментальным.

Температуру принято измерять в градусах, но только из соображений удобства и по традиции. Существует, как известно, несколько температурных шкал. В большинстве стран используют шкалу Цельсия, в которой нулю градусов соответствует точка таяния льда при нормальном давлении, однако в США до сих пор пользуются шкалой Фаренгейта, в которой 0°C соответствует 32 градуса. В чем причина устойчивого консерватизма в выборе системы единиц в США, мне неизвестно. Знаю лишь, что новым жителям этой страны непросто привыкнуть не только к температурной шкале Фаренгейта, но и к фунтам, галлонам, унциям, милям, футам, дюймам и т. д. Наиболее «физической» является температурная шкала Кельвина. Температуру,

выраженную по шкале Кельвина, принято называть абсолютной. Ее обозначение не сопровождается значком градуса; после числа градусов ставят заглавную букву К. Температуру 300 К называют комнатной, хотя $300 \text{ К} = 26,84 \text{ }^\circ\text{С}$ — жарковато для жилой комнаты ¹⁾.

Абсолютная температура является мерой энергии теплового движения атомных частиц в теле (см. гл. 9). Поэтому существует механический эквивалент тепла, и температуру можно выразить в энергетических единицах. Другими словами, температура имеет размерность энергии. Переводным множителем из градусов в эрги ²⁾ служит постоянная Больцмана ($k = 1,410^{-16}$ эрг/К).

Задумывались ли Вы, почему постоянная Больцмана так мала? Оказывается, kT по порядку величины есть средняя энергия одной (!) атомной частицы тела при температуре T .

При вычислении средней скорости молекулы идеального газа можно пользоваться формулой

$$\frac{1}{2}mv^2 = \frac{3}{2}kT.$$

Если подставить сюда температуру в абсолютных градусах, а массу в граммах, то скорость $v = (3kT/m)^{1/2}$ будет измеряться в см/с.

Размерность электромагнитных величин. Основными электромагнитными характеристиками являются напряженности электрического (**E**) и магнитного (**H**) полей.

Начнем с напряженности электрического поля. Принято указывать **E** в вольтах на метр (В/м). Нас интересует $[E]$, выраженная через фундаментальные размерности. Воспользуемся тем, что энергия самого электрического поля, накопленная в объеме V , равна $VE^2/8\pi$, то есть

$$\text{см}^3 \cdot [E^2] = \text{г} \cdot (\text{см}/\text{с})^2, \quad [E] = \text{г}^{1/2}/\text{с} \cdot \text{см}^{1/2}.$$

Знание размерности напряженности электрического поля позволяет определить размерность заряда. Обозначим заряд буквой Q . Сила, действующая на него со стороны электрического поля, равна QE , поэтому $[Q] \cdot [E] = \text{г} \cdot \text{см}/\text{с}^2$, откуда $[Q] = \text{г}^{1/2} \cdot \text{см}^{3/2}/\text{с}$.

Размерность заряда можно определить и другим способом. Согласно закону Кулона сила, действующая между двумя одинаковыми зарядами, расположенными на расстоянии R друг от друга, равна Q^2/R^2 . Следовательно,

$$[Q]^2 \cdot [R]^{-2} = \text{г} \cdot \text{см}/\text{с}^2,$$

¹⁾ Много-много лет назад от своего коллеги Я. С. Кана я впервые услышал, что с появлением шкалы Кельвина появилась возможность отмечать не только *на сколько* одна температура больше или меньше другой, но и *во сколько* раз одна температура больше или меньше другой. Это утверждение произвело на меня впечатление и запомнилось.

²⁾ Эрг — единица макроскопической энергии; $1 \text{ эВ} = 1,60219 \cdot 10^{-12}$ эрг.

или $[Q] = \text{г}^{1/2} \cdot \text{см}^{3/2} / \text{с}$.

Мы видим, что полученные двумя разными способами результаты совпадают.

Размерность напряженности магнитного поля не отличается от размерности напряженности электрического поля ($[H] = [E]$), поскольку энергия магнитного поля в объеме V определяется такой же формулой, что и энергия электрического поля. Нетрудно проверить, что $[\mu] \cdot [H] = [\text{энергия}]$ (см. приложение).

Используемые на практике размерности напряженностей E и H различны.

Размерность оптических характеристик удобно выяснять, считая свет то электромагнитными волнами, то фотонами (используя корпускулярно-волновой дуализм). Частота ω волны имеет размерность $1/\text{с}$. Размерность длины волны — см. Скорость света c , как всякая скорость, имеет размерность см/с. Энергия фотона $\hbar\omega$, естественно, имеет ту же размерность, что и любая энергия, а его импульс (равный $\hbar\omega/c$) — размерность импульса. При выяснении размерности физических величин, в определение которых входит постоянная Планка h , удобно помнить следующие соотношения:

$$[\hbar] = [\text{координата}] \cdot [\text{импульс}], \quad [\hbar] = [\text{энергия}] \cdot [\text{время}].$$

Поскольку световые волны являются электромагнитными, их амплитуда имеет размерность $[E] = [H] = \text{г}^{1/2} / \text{с} \cdot \text{см}^{1/2}$.

Оптика как техническая дисциплина, а не как часть физики, использует единицы измерения освещенности (люкс), светового потока (люмен) и силы света (кандела). Все они связаны друг с другом, а поскольку кандела «равна силе света в заданном направлении источника, испускающего монохроматическое излучение частотой $540 \cdot 10^{12}$ Гц, энергетическая сила света которого в этом направлении $1/683$ Вт/ср»¹⁾, все три величины могут быть выражены через три фундаментальные (см, с, г).

Познакомившись с размерностями и поняв, что фундаментальных среди них всего лишь три, можно объяснить, как использовать размерности физических величин в абстрактном приеме, который так и именуют — *метод размерностей*.

Один из важных этапов понимания сути явления (или свойства) — вывод описывающей его формулы. Нередко мы знаем, от каких величин должна зависеть характеристика явления (свойства), но строгий математический вывод затруднителен, иногда невозможен. В подобных случаях имеет смысл воспользоваться соображениями размерности величин, которые, по нашему мнению, должны войти в формулу.

¹⁾ Физическая энциклопедия. — М., 1990. Т. 2. С. 236. Герц (Гц) — единица частоты; Ватт (Вт) — единица мощности (Мощность = Энергия/с); стерадиан (ср) — мера телесного угла (угловые характеристики безразмерны).

Проиллюстрируем возможности такого подхода (несомненно, весьма абстрактного) на примере самой простой квантовой системы — атома водорода (электрона, движущегося в поле ядра — протона). Зададим два вопроса. Каков размер атома водорода a ? Каков порядок величины средней скорости электрона v ? Ответ должен зависеть от трех размерных величин: заряда электрона e , его массы m и постоянной Планка \hbar . Из этих величин нетрудно составить комбинации размерности длины (см) и скорости (см/с). Они и будут определять порядки интересующих нас величин:

$$a = \frac{\hbar^2}{me^2}, \quad v = \frac{e^2}{\hbar}.$$

Обратите внимание на то, что используя классическую механику, т. е. без постоянной Планка \hbar , нельзя составить из e и m комбинаций размерности длины и скорости. Классические частицы — электроны (здесь лучше сказать «тела», «планеты») могут вращаться вокруг протона («солнца») по орбите любого размера. Скорость такой «планеты», естественно, будет функцией расстояния от нее до «солнца»: $v = e(ma)^{-1/2}$.

Подчеркнем еще раз: тождественность атомов — непосредственное следствие квантовой механики.

Полученные соотношения, естественно, не исчерпывают решения задачи. Таким путем нельзя выяснить, каковы уровни энергии электрона в поле ядра, что собой представляют волновые функции и многое другое, без чего атомная физика неполна. Однако, согласитесь, возможность оценить столь простым способом важные характеристики атома в его основном состоянии весьма полезна. При этом оценка оказывается достаточно точной. Последнее, конечно, является следствием простоты системы. В сложных случаях, при описании свойств, строения и поведения макроскопических систем, численный множитель, отличающий точный результат от оценки, полученной сравнением размерностей, может быть как значительно больше, так и значительно меньше единицы. Вместе с тем, как это ни поразительно, он нередко оказывается порядка единицы.

Приложение

У нас имеется возможность подробно рассмотреть, как планировался эксперимент по измерению магнитного момента атома.

В 1920-м году, за два года до того, как Штерн и Герлах опубликовали результаты своего исторического опыта, два молодых ученых, будущих нобелевских лауреата — П.Л. Капица (1894–1984) и Н.Н. Се-

менов (1896–1986) ¹⁾ — разработали метод измерения магнитных моментов атомов путем пропускания атомного пучка через неоднородное магнитное поле. Непосредственным результатом эксперимента должна была стать форма пятна на пластинке. О пространственном квантовании Капица и Семенов тогда еще не знали. Свои соображения они опубликовали в Журнале Русского физико-химического общества (1922. Т. 50. С. 159–160). Форма пятна, точнее, его ширина, естественно, выражалась в сантиметрах. Формула, описывающая ожидаемые результаты (ширину пятна), связывала параметры установки, предполагаемое значение величины магнитного момента и градиент магнитного поля. Все величины выражались в принятых единицах (размеры — в см, скорости — в см/с, градиент магнитного поля — в Гс/см, а магнитный момент атома — в единицах СГСМ). Как и всякая правильная физическая формула, полученное авторами выражение, демонстрирует, в частности, правильную размерность: левая и правая стороны равенства имеют одинаковые размерности.

У внимательного читателя может возникнуть подозрение, что мы в своем рассказе допустили ошибку, и Штерн и Герлах, обнаружив в эксперименте две полоски, определили спин не электрона, а атома серебра. Расскажем об их опыте чуть подробнее.

Пусть атом Ag имеет магнитный момент μ . Энергия его взаимодействия с магнитным полем напряженности \mathbf{H} равна $\mu\mathbf{H}$. Если магнитное поле, направленное вдоль пучка, неоднородно в перпендикулярном пучку направлении y , то на атом действует сила $f_y = \mu dH/dy$.

Из-за пространственного квантования под действием силы f_y пучок разделится на столько пучков, сколько проекций имеет вектор μ . В электронной оболочке Ag содержится 47 электронов. Почему же магнитный момент атома Ag имеет всего две проекции, как спин электрона? При чем здесь вообще спин — собственный момент количества движения?

Оказывается, 46 из 47 электронов атома серебра в результате действия принципа запрета Паули организованы так, что их суммарный магнитный момент равен нулю. Поэтому магнитный момент атома Ag определяет один электрон. При вращении любое заряженное тело превращается в магнит. Это относится и к микроскопическим частицам. Правда, существует (точнее, может существовать) количественное различие. Величина магнитного момента пропорциональна моменту импульса. Если последний является орбитальным, то коэффициент пропорциональности (он называется *гиромангнитным отношением*

¹⁾ П.Л. Капица получил в 1978 г. Нобелевскую премию за открытия в области физики низких температур, а Н.Н. Семенов — в 1956 г. Нобелевскую премию по химии за открытие цепных реакций. Осуществить задуманный эксперимент по измерению магнитных моментов атомов они не сумели из-за разрухи в стране в первые послереволюционные годы.

и часто обозначается буквой g) равен $Q/2Mc$, где Q — заряд тела, M — его масса, а c — скорость света. Если момент импульса — спиновый, то гиромагнитное отношение вдвое больше. Орбитальный момент 47-го электрона в атоме Ag равен нулю. Следовательно, магнитный момент атома Ag составляет $e\hbar/2mc$. Надеюсь, Вы помните, что собственный (спиновый) момент импульса электрона — $(1/2)\hbar$. Число проекций магнитного момента атома серебра (в данном случае, на направление магнитного поля \mathbf{H}) равняется двум и совпадает с числом проекций спина электрона.

Глава 13. БЕЗРАЗМЕРНЫЕ ОТНОШЕНИЯ

Современное «здание» физики огромно. Мы уже использовали этот образ. Хотя, работая на одном из «этажей», можно годами не спускаться и не подниматься на соседние, все «строители» ощущают единство конструкции. Проявляется оно не только в том, что все заняты постижением единой природы, но и в том, что результаты, полученные на одном «этаже», практически всегда служат исходным «строительным материалом» на других. Приведем последовательность нескольких «этажей», не пытаясь перечислить все и не вдаваясь в подробности:

*Элементарные частицы, электроны → Ядра атомов → Атомы →
Молекулы → Газы, жидкости, плазма, твердые тела.*

Стрелки указывают направление, откуда «подвозят» на соответствующий «этаж» «строительный материал».

Строение самых различных тел в настоящее время известно довольно подробно, описание их свойств покоится на весьма общих законах физики, а современная вычислительная техника и современный математический аппарат теоретической физики обладают необозримыми возможностями. Поэтому можно осмелиться сформулировать вопрос: из каких величин должны в *конечном итоге* состоять все физические характеристики относительно сложных объектов (от ядер атомов и дальше по направлению стрелок на схеме)?

Одной из задач, которые ставит перед собой теоретическая физика, является вывод формул, описывающих то или иное явление или свойство (мы уже об этом упоминали). Итогом теории служит формула, связывающая физические величины разной природы. Физические величины имеют размерность. Уже из этого ясно, что формулы не могут содержать только безразмерные числа. Среди входящих в них размерных величин, по-видимому, должны присутствовать численные характеристики элементарных частиц и электронов — их массы, заряды разной природы, спины. Опыт последнего столетия убедительно показывает, что необходимо присутствие по крайней мере еще двух величин — постоянной Планка \hbar и скорости света c . Постоянная Планка — мера корпускулярно-волнового дуализма микрочастиц. Скорость света — максимальная скорость распространения сигнала; без нее не обходится ни одна формула релятивистской механики (см. гл. 4).

Попытки осуществить амбициозную программу вычисления характеристик атомных и макроскопических систем через характеристики

электронов и других элементарных частиц носят название *расчетов из первых принципов*. В полном объеме такая программа не осуществлена. Однако для того чтобы поверить в возможность ее полной реализации, есть достаточно оснований: разделенная на этапы, она демонстрирует значительные успехи и . . . ставит новые вопросы.

Не хотелось бы лишать себя права задать вопрос о генезисе характеристик элементарных частиц. Почему элементарные частицы имеют именно такие характеристики? С помощью каких величин они должны (могут) быть выражены?

Похоже, что на пути углубления наших знаний о первичных сущностях есть успехи. Однако мы не имеем возможности даже назвать их. Это увело бы нас в *terra incognita* — в незнакомую (нам!) область. Остережемся.

Думаю, можно смело утверждать: будущая теория не станет ставить перед собой задачи конкретного вычисления массы электрона в граммах ($\sim 10^{-27}$ г) или его заряда в электростатических единицах ($\sim 4,8 \cdot 10^{-10}$ CGSE). Скорее всего, она сможет ответить на вопрос, почему отношение масс протона и электрона близко к 1838 вне зависимости от выбранных единиц или почему $e^2/\hbar c \approx 1/137$. Комбинацию $e^2/\hbar c$ принято считать *безразмерным зарядом*. Составленная из столь разнородных величин, она действительно безразмерна¹⁾. Когда утверждают, что заряд мал, опираются именно на это соотношение. Отношение $e^2/\hbar c$ часто так и называют — «1/137-я», и все знают, о чем идет речь.

Сравнение разнородных величин одной размерности иногда приводит к возникновению непредставимо больших или непредставимо малых чисел. Два протона отталкиваются друг от друга, поскольку одноименно заряжены, и притягиваются друг к другу согласно закону всемирного тяготения. Обе силы обратно пропорциональны квадрату расстояния между протонами, а их отношение (по модулю) составляет

$$\frac{e^2}{G(m_p)^2},$$

где m_p и e — масса и заряд протона; гравитационная постоянная G равна

$$G = 6,67 \cdot 10^{-8} \text{ см}^3/\text{г} \cdot \text{с}^2.$$

Подставив численные значения, убедимся: электростатическое отталкивание в 10^{36} раз больше гравитационного притяжения. Сможет ли когда-нибудь теория вывести это число? Думаю, что да. Не напрасно

¹⁾ Интересные соотношения, связывающие между собой далекие друг от друга величины, есть и в математике. Например, соотношение между $e \equiv \lim_{n \rightarrow \infty} (1 + 1/n)^n$ и $\pi: e^{i\pi} = -1$.

же говорят об успехах на пути к созданию Великого объединения всех сил, существующих в природе?!

Далее в настоящей главе мы сосредоточим внимание на 1/137-й. Однако сначала скажем несколько слов о фантастической малости гравитационных сил по сравнению с электромагнитными. Написал «фантастическая малость» и стало не по себе. Идя в гору, особенно с тяжелым рюкзаком на спине, трудно представить, что гравитационные силы так малы. Задумавшись, конечно, мы понимаем, что дело тут в матушке-Земле, которая всей своей массой ($M_3 \approx 6 \cdot 10^{24}$ кг, а ее радиус $R \approx 6400$ км) притягивает нас к себе. Чтобы оценить гравитацию Земли, вспомним формулу, определяющую ускорение силы тяжести:

$$g = \frac{GM_3}{R^2}.$$

Подстановка численных значений входящих в нее величин приводит к привычному значению $\sim 10^3$ см/с². Эта формула дает также и возможность представить, что будет чувствовать астронавт на других планетах.

Число с 36-ю нулями представить невозможно. Одно очевидно: область учета сил гравитации бесконечно далека от атомной и субатомной физики. Следует признать, что вся картина Мира оказалась бы иной, если бы гравитационные и электрические силы были бы близки друг к другу.

Не устаю удивляться устройству Мира и восхищаться им.

Вдумайтесь! Несмотря на все разнообразие, безграничный Мир (в том числе и мы сами) построен из трех видов частиц: электронов, протонов и нейтронов в разных сочетаниях! Разнообразие — результат качественного отличия структур нашего Мира. Почему они столь разительно отличаются друг от друга? Ответ на этот вопрос в каждом конкретном случае требует сравнения разнообразных физических величин. Сравнение допускают только величины одной размерности. Поэтому можно сказать иначе: различие структур обусловлено тем, какую величину имеют определенные *безразмерные комбинации физических величин*.

Когда речь идет об устройстве Мира, простейшей структурой, долгое время считавшейся элементарнейшей, несомненно, является атом. Зная его строение (то, что он состоит из ядра и электронной оболочки, то есть вращающихся вокруг ядра электронов), можно заметить, что атомы демонстрируют две принципиально различные совокупности субатомных частиц: нуклоны в ядре напоминают конденсированную среду, а атомные оболочки — газ. В ядре частицы практически соприкасаются друг с другом, а в электронной оболочке находятся далеко друг от друга. Последнее утверждение можно подтвердить оценкой.

Я уверен, что у каждого, кто более или менее подробно знакомится с устройством нашего Мира, есть среди его черт свои «любимчики» —

черты конструкции, которые ему особенно по душе. Мне очень нравится безразмерный заряд электрона — отношение $e^2/\hbar c$. Нравится потому, что оно неожиданно возникает в разных ситуациях и часто проявляет свою малость неожиданным образом.

Начнем с атома водорода. Мы уже знаем, что электрон в нем находится на расстоянии порядка ангстрема от ядра — протона ($1 \text{ \AA} = 10^{-8} \text{ см}$). Далеко это или близко? По сравнению с «человеческими масштабами», конечно, очень близко. Однако размер человека и размеры, доступные человеческим чувствам, тут не при чем ¹⁾. По сравнению с ядром это расстояние огромно, так как размер протона по порядку величины (10^{-14} см) в миллион раз меньше расстояния от протона до электрона. Что же будет, если проводить сравнение с «размером» самого электрона? Слово «размер» взято здесь в кавычки, поскольку относиться к тем размерам, которыми мы будем оперировать, надо с большой осторожностью. Из характеризующих электрон величин (e, m, c, \hbar ; см. выше) можно составить две комбинации размерности длины:

$$r_c = \frac{e^2}{mc^2}, \quad \lambda_K = \frac{\hbar}{mc}.$$

Первая называется электростатическим радиусом электрона. Формально значение r_c можно считать радиусом сферы, вне которой электростатическая энергия электрона по порядку величины совпадает с его энергией покоя (mc^2).

Вторая комбинация размерности длины — комптоновская длина волны. Она названа в честь американского физика Артура Комптона, открывшего эффект, непосредственно продемонстрировавший корпускулярную природу фотона. Эффект Комптона заключается в изменении длины волны рентгеновского излучения при рассеянии его электронами. Рассмотрение законов сохранения показывает, что энергия и импульс электрона и фотона при этом перераспределяются. Изменение энергии фотона означает изменение его частоты ω . Поскольку частота фотона и его длина волны λ жестко связаны соотношением $2\pi/\lambda = \omega/c$, изменение ω влечет за собой изменение λ . Величина λ_K как раз и есть мера изменения длины волны рентгеновского излучения. Комптоновскую длину можно отождествить с «размером» электрона, так как при попытке локализовать электрон в области с линейными размерами, меньшими λ_K , согласно соотношениям неопределенности Гейзенберга, неопределенность его импульса достигнет mc , а энергии — mc^2 , что достаточно для рождения пар *частица–античастица*

¹⁾ Это не совсем так. Существуют интересные соображения о возможных размерах высших организмов в условиях, имеющих место на Земле. В рассуждениях, естественно, учитываются размеры атомов и молекул, из которых состоит любой живой организм.

с массой m (например, электрон–позитрон; см. гл. 14). Из-за этого локализация электрона в области с размерами λ_K или меньше приводит к возникновению сгустка вещества, состоящего из таких пар и занимающего внутренность сферы радиуса λ_K . Подобный электрон–сгусток полусерьезно, полушутливо называют электроном, «одетым шубой». «Одетый шубой» электрон имеет линейные размеры порядка λ_K . Сравним теперь все три размера (r_c , λ_K и a):

$$\frac{r_c}{\lambda_K} = \frac{e^2}{\hbar c}, \quad \frac{\lambda_K}{a} = \frac{e^2}{\hbar c}.$$

Роль $1/137$ -й отчетливо видна. Кроме того, мы узнали, что атом фактически пуст. Даже «одетый шубой» электрон не заполняет атома водорода, а занимает лишь незначительную часть его объема ($\sim 1/137^3$).

Атом водорода был рассмотрен для простоты. В любом атоме, даже если электронов в нем значительно больше одного, так что можно говорить, пусть условно, о газе электронов, окружающих ядро, газ электронов разрежен — расстояния между ними существенно (в 137 раз!) превышают «размеры» электрона. Повторим, и опять без разъяснения: многонуклонное ядро — плотный сгусток, расстояние между нуклонами в нем того же порядка, что размер нуклона.

Безразмерное отношение массы нуклона к массе электрона определяет распределение ролей между ядрами и электронами в динамических свойствах макроскопических тел.

Немного арифметики. Протонов в ядрах столько же, сколько электронов в атоме. В легких ядрах нейтронов столько же, сколько протонов, а в тяжелых — больше, чем протонов. Один нуклон в 1838 раз тяжелее электрона. Следовательно, практически вся масса твердого тела сосредоточена в ядрах атомов. Электроны в тысячи раз легче.

Отрицательно заряженные электроны отталкиваются друг от друга. Отталкиваются друг от друга и положительно заряженные ядра. Тела не могли бы существовать, если бы протоны ядер не компенсировали заряд электронов, а электроны — заряд ядер. Нейтральность тел обеспечивается и электронами, и ядрами.

Все микрочастицы, из которых состоит тело, накрепко связаны друг с другом. Однако при отличной от нуля температуре они движутся. Это и есть тепловое движение в твердом теле¹⁾. Один из основных типов движения частиц — их колебания вокруг положений равновесия. Хороший наглядный образ таков: твердое тело — набор осцилляторов с разными частотами. Частоты колебаний заполняют целые полосы, которых тем больше, чем больше атомов в ячейке кристалла (см. гл. 16). Чем атомы тяжелее, тем частоты меньше.

¹⁾ Не следует забывать, что и при $T = 0$ К движение атомных частиц не прекращается. Электроны движутся вокруг ядер, атомы и молекулы совершают нулевые колебания (см. гл. 16).

Вспомните формулу для определения частоты колебаний маятника: она обратно пропорциональна квадратному корню из массы маятника. Частоты колебаний атомов в твердом теле обратно пропорциональны квадратному корню из масс атомов, а массы атомов определяются массами ядер.

Положение атомов в твердом теле не зафиксировано навсегда. Если тело не полностью однородно, то его атомы путем диффузии пытаются занять такое положение, при котором в среднем тело окажется однородным. Диффузия — случайное перемещение атома на значительное расстояние. При каждом перемещении он должен преодолеть энергетический барьер между положениями равновесия.

Чем создаются барьеры? Атомные барьеры — результат взаимодействия электронной оболочки блуждающего атома с электронными оболочками атомов из его ближайшего окружения. Скорость перемещения зависит от массы атома: чем он тяжелее, тем вероятность его перемещения меньше.

При рассмотрении движения электронов в твердом теле ядра в хорошем приближении можно считать неподвижными. Повторим: роль ядер сводится главным образом к компенсации заряда электронов. Однако не только. Основой наших представлений о движении электронов служит анализ их движения под действием периодической силы, созданной ядрами (см. гл. 17). Строго говоря, ядра представляют собой периодическую структуру, если, конечно, абстрагироваться от их теплового и диффузионного движения (в частности, от их колебаний).

Одна из важных задач теории твердого тела — расчет равновесных структур, выяснение того, как энергетически выгодно расположиться атомам. При расчете структуры тела, казалось бы, надо учесть все силы взаимодействия между частицами, из которых оно состоит. Между заряженными частицами действуют электростатические силы, обусловленные наличием у электронов и ядер зарядов ¹⁾. Вместе с тем, существованием у электронов (тем более, у ядер) магнитных моментов можно пренебречь. Почему? Ведь между магнетиками существует специфическое взаимодействие. Как мы увидим, магнитным взаимодействием при расчете структуры тела можно пренебречь потому, что *безразмерный заряд электрона значительно меньше единицы*.

Если заряды расположены на расстоянии r друг от друга, то действующая между ними электростатическая сила равна $f_{эл} = e^2/r^2$. Кроме того, между электронами, благодаря наличию у них магнитного момента (электрон — магнитный диполь — магнетик), существует

¹⁾ При расчете электростатического взаимодействия между микрочастицами надо учесть электронейтральность, отталкивание одноименно заряженных частиц, положение электронов в атоме и поляризующее влияние атомов друг на друга при конечном расстоянии между ними.

так называемое магнитно-дипольное взаимодействие. Сила магнитно-дипольного взаимодействия устроена сложно, но по порядку величины равна $f_m \sim \mu^2/r^4$. Таким образом, она спадает с увеличением расстояния быстрее, чем кулоновская. Сравним эти силы друг с другом для случая, когда расстояние между электронами составляет порядка боровского радиуса a (как в твердом теле):

$$\frac{f_m}{f_{эл}} = \left(\frac{e^2}{\hbar c} \right)^2 = \left(\frac{1}{137} \right)^2.$$

Парадоксальная ситуация: электростатическая сила взаимодействия значительно больше магнитно-дипольной из-за того, что безразмерный заряд мал!

Можно было бы привести еще несколько примеров фундаментальной роли малости безразмерного заряда электрона. Например, магнитная восприимчивость твердого диамагнетика всегда, когда диамагнетизм не обязан своим существованием переходу металла в сверхпроводящее состояние, значительно меньше единицы¹⁾. Причина — малость все той же постоянной тонкой структуры. Можно показать, что диамагнитная восприимчивость по порядку величины просто равна $(e^2/\hbar c)^2$. Опять!

Думается, роль $1/137$ -й продемонстрирована. Некоторые дополнительные соображения по этому поводу будут высказаны в заключении.

¹⁾ Мы используем систему единиц, в которой магнитная восприимчивость — безразмерная величина, то есть магнитный момент и магнитное поле имеют одинаковую размерность (см. гл. 12).

Глава 14. АБСТРАКЦИЯ ИЛИ РЕАЛЬНОСТЬ?

В современной физике встречаются термины, которые, несомненно, претендуют на эмоциональное их восприятие: узкое горло, шарм, ливень, цвет, вакансия, коллапс, дырка, хаос, античастица и многие другие. Мы привели термины в художественном беспорядке и среди них поместили заранее выбранные — те, которые нам понадобятся в этой главе: вакансия, дырка, античастица. Добавим к ним еще позитрон и позитроний, имеющие привычную латинскую этимологию: *positivus* — положительный.

Начнем с *вакансии*. Если бы можно было создать объемное изображение реального кристалла, мы обнаружили бы, что совершенный на первый взгляд кристалл обладает дефектами. Эти дефекты различны. Одни устроены весьма непросто, другие можно легко описать. Наверняка, мы заметили бы, что некоторые узлы кристаллической решетки не заняты атомами или ионами — теми частицами, из которых построен кристалл. Незанятый атомом или ионом узел и есть вакансия. Из-за того что он пуст, немного нарушен баланс сил, а соседние атомы чуть сдвинуты со своих привычных мест. Всякое нарушение строгого порядка сопровождается подобными сдвигами атомов. Такова реальность.

При конечной температуре атомы не неподвижны. В основном они совершают малые колебания вокруг положений равновесия. Однако изредка возможна флуктуация. В 9-й главе мы говорили о термодинамических флуктуациях и о том, каков масштаб флуктуаций в среднем. В среднем флуктуации малы. Вместе с тем, иногда происходят катастрофические на микроуровне события. Например, в результате флуктуации атом, расположенный рядом с вакансией, может приобрести достаточную энергию и перепрыгнуть на свободное место, покинув свое. И это реальность.

С другой стороны, это же событие можно описать иначе: *переместилась вакансия*. Оказывается, вакансия может двигаться! Более того, движение атомов кристалла можно трактовать как перемещение вакансий: откуда уходят вакансии, туда приходят атомы. Думаю, ясно, что в диффузионных процессах, происходящих в твердых телах, движение вакансий играет важную, по сути определяющую роль. Описывая диффузию, удобно считать, что вакансии представляют собой газ, растворенный в твердом теле. Частицы газа служат «извозчиками», перевозя атомы с места на место. Эта сюрреалистическая картина

может быть детально проверена тщательным изучением диффузии и самодиффузии¹⁾).

Образ *вакансия-частица* особенно впечатляет, когда, благодаря волновым свойствам атомных частиц, вакансия теряет свою локализацию и превращается в квантовую частицу, состояние которой характеризуется определенным импульсом. Эта квантовая частица даже получила специальное наименование — *вакансион*. Делокализация вакансий — превращение классической вакансии в квантовую частицу — требует определенных и довольно жестких условий: атомы кристалла должны быть легки, а силы связи между ними — слабы. (В гл. 16 мы убедимся, что делокализованную вакансию лучше называть *квазичастицей*.)

Условия делокализации вакансии осуществляются в так называемых *квантовых кристаллах*, лучше всего в твердом гелии — в гелии под давлением. (Без давления гелий не затвердевает.) Физика квантовых кристаллов — новая, интересная и активно развивающаяся область низкотемпературной физики твердого тела.

Дырка. Я сомневался, не следует ли взять слово «дырка» в кавычки. В последнее время в физике полупроводников, превратившей это обиходное слово в термин, кавычки не ставят. Что же такое дырка?

Квантовомеханическое рассмотрение движения электрона в периодическом поле сил показало, что его энергия в этом случае может принимать только разрешенные значения, занимающие определенные полосы — зоны (на рис. 14.1 они заштрихованы). В свободном пространстве электрону разрешено иметь произвольные значения энергии, электрону в атоме — дискретные. Движение под действием периодической силы — промежуточный случай (подробнее об электронах в периодическом поле сил будет рассказано в гл. 17).

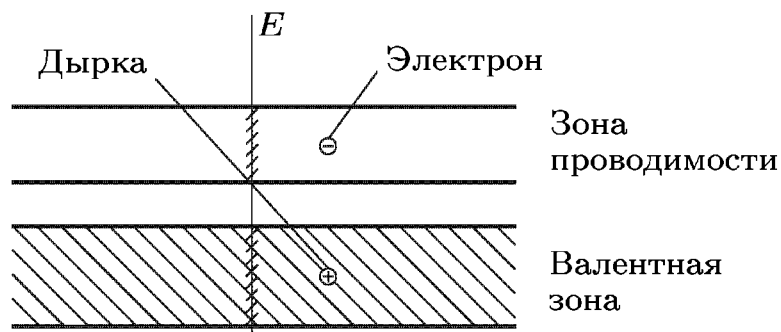


Рис. 14.1 Зонная структура для электрона в периодическом поле

Вспомним, что принцип запрета Паули не позволяет электронам собраться на нижнем энергетическом уровне даже при абсолютном нуле температуры. Электроны заполняют зоны так, чтобы принцип Паули не нарушался.

¹⁾ Самодиффузия — случайное блуждание собственных атомов тела. Она изучается с помощью радиоактивных меток.

Если все электроны твердого тела, заняв свои места, оставляют одну или несколько зон частично заполненными, то твердое тело называется металлом (проводником). Твердое тело является диэлектриком (изолятором), если в нем имеются только целиком заполненные и полностью пустые зоны, а между ними располагается запрещенная полоса значений энергии.

Более подробно о классификации твердых тел будет рассказано в гл. 17. Упомянем только, что электроны заполненной зоны под действием не слишком большого электрического поля не перемещаются в пространстве — полный ток электронов заполненной зоны равен нулю. Если же поле столь велико, что наступает электрический пробой, то в каком-то смысле безразлично, является ли тело изолятором или проводником.

Если не все места в зоне заняты, то есть имеются незаполненные состояния (дырки), то подсоединив тело к электрической цепи, мы убедимся, что по нему потечет ток, причем такой, будто заряд переносят дырки. Существует способ установить, какой знак имеют перемещающиеся частицы. Оказывается, положительный! Следует помнить: достоверно установлено, что ионы практически не принимают участия в переносе заряда.

Конечно, если вдуматься, все становится ясным: дырка есть отсутствие отрицательного заряда. Она воспринимается как положительный заряд. Поэтому дырку *удобно считать* положительно заряженной ... частицей. Именно удобно! Можно провести вычисления, не вводя дырок. Они «сами появятся».

Если у диэлектрика запрещенная зона энергий сравнительно мала, так что при комнатных температурах заметная доля электронов в результате теплового возбуждения переходит в пустую зону, а в заполненной зоне появляются дырки, то диэлектрик называют *полупроводником*. Таковы Ge, Si, Se, InSb, InAs. Именно их использование в бесконечно разнообразных приборах и устройствах совершенно преобразило нашу жизнь за вторую половину XX-го века.

О роли полупроводников в технике нельзя говорить походя. В данном контексте нас интересует лишь то, что дырки, наряду с электронами, являются равноправными частицами. Как и электроны, дырки — фермионы.

Рассматривая явления, происходящие в полупроводниках, необходимо помнить, что полупроводник обычно содержит два газа заряженных частиц — газ отрицательно заряженных электронов и газ положительно заряженных дырок. Иногда, правда, в полупроводнике присутствуют только электроны или только дырки. Этот факт специально отмечают: p-Ge — германий, содержащий дырки; n-Ge — германий, содержащий электроны.

Конечно, электроны есть в каждом теле, в каждом атоме. Подчеркивая наличие электронов в полупроводнике, утверждают, что в ис-

ходно пустой зоне имеются электроны, способные принять участие в проводимости. Их так и называют — *электронами проводимости*, а зону, в которой они находятся, — *зоной проводимости*. Когда по полупроводнику течет ток дырок, говорят о *дырочной проводимости*; зона, в которой расположены дырки, называется *валентной зоной*.

Позитрон — первая античастица. Физический вакуум. Когда Дирак (Поль Адриан Морис Дирак, 1902–1984 гг., Нобелевская премия 1933 г.) сформулировал свое знаменитое уравнение (1930 г.), положившее начало релятивистской квантовой механике, возникло странное противоречие. Оказалось, что решения уравнения (состояния) могут описывать не только реальные электроны с положительной энергией ε , зависящей от импульса p так, как предписывает релятивистская механика ($\varepsilon = (m^2c^4 + c^2p^2)^{1/2}$). Они описывают и состояния с отрицательной энергией ($\varepsilon = -(m^2c^4 + c^2p^2)^{1/2}$), которые, казалось бы, ничему реальному не соответствуют. Отмахнуться от них не было никакой возможности, хотя бы потому, что электрон с положительной энергией должен «упасть» в состояние с отрицательной энергией, излучив лишнюю энергию в виде световых квантов. Последнее всегда происходит, если атомный электрон оказывается в состоянии с высокой энергией и имеются свободные состояния с энергией ниже.

Аналогия с атомом подсказала Дираку нетривиальное решение проблемы. Он предположил, что все состояния с отрицательной энергией заняты, а наблюдать электроны в этих состояниях нельзя, так как они не принимают участия ни в каких физических процессах.

Идея Дирака заставила пересмотреть представление о вакууме как о состоянии, в котором нет частиц. Все уровни с $\varepsilon < 0$ заполнены электронами. Непосредственно вакуум ненаблюдаем, хотя его энергия, заряд и т. п. равны бесконечности. Бесконечные значения этих величин служат началом отсчета. Наблюдаемое конечное значение физической величины A есть разность

$$A(\text{электронов}) = A(\text{вакуума} + \text{электронов}) - A(\text{вакуума}).$$

Если извлечь электрон из вакуума (то есть из состояния с $\varepsilon < 0$), то образуется дырка. Нарушая однородность вакуума, она наблюдаема и ведет себя как частица, тождественная электрону, но обладающая положительным зарядом. Ее назвали *позитроном*. Когда в 1932 г. Андерсон обнаружил в космических лучах положительно заряженный электрон, идея Дирака получила замечательное подтверждение.

Позитрон является *античастицей* по отношению к частице — электрону. Термин «античастица» подчеркивает не только тождественность всех, кроме знака заряда, свойств электрона и позитрона. Электрон и позитрон могут аннигилировать, излучая освободившуюся энергию в виде γ -квантов.

Позитрон оказался первенцем в семействе античастиц. Каждый фермион (нейтрино, мюон, протон и т. д.) имеет свою античастицу.

Это стало столь привычным, что античастицам не присваивают специальных названий, просто добавляют приставку *анти-*: антипротон, антинейтрино и т. д.

Изложенная концепция весьма усложнила представление о вакууме. Каждый тип фермионов имеет свой вакуум, а реальный физический вакуум — конгломерат всех возможных вакуумов. Вероятно, многих мы еще не знаем.

Аппарат современной релятивистской квантовой теории модифицирован так, что вакуум, заполненный фермионами с отрицательной энергией, остается «за кадром». Однако фактически применяемые методы расчета основаны на идее Дирака.

Продумав, о чем я хочу рассказать в настоящей главе, начал искать первую фразу и остановился на такой: «Откроем словарь». Признаться, я не ожидал, что Словарь (речь идет о Словаре русского языка С.И. Ожегова. — М.: Русский язык, 1975) даст столько материала для размышления. Слово «точка» используется в семи различных смыслах, даже если опустить выделенное «точка, см. точить». Из семи словоупотреблений нам интересны следующие.

1. «След от прикосновения, укола чем-либо острым (кончиком карандаша, пера, иглы и т. п.), маленькое круглое пятнышко».

3. «Основное понятие геометрии, а также механики, физики — место, не имеющее измерения, граница отрезка линии». В этом пункте приведены такие примеры: точка пересечения прямых; точка приложения сил; точка опоры; точка отсчета.

Геометрический пример — точка пересечения прямых — вполне строго соответствует определению. Действительно, прямая линия в геометрии не имеет толщины — поперечного размера. Следовательно, не имеет размера и пересечение прямых. Однако любая точка опоры, точка приложения сил, если речь не идет о силах типа гравитационной или кулоновской, с необходимостью имеет какой-то, пусть небольшой размер и, скорее, напоминает точку в первом определении, чем в третьем. Просто (?) мы привыкли опускать уточняющие слова. Например, такие: «Силу можно считать приложенной в точке, если размер области, на которую действует сила, значительно меньше, чем размер предмета, на который она действует».

«Маленькому круглому пятнышку» в Словаре соответствует пример: «ситец в красных точках». Вооружись мы микроскопом, пятнышко предстанет перед нами огромной поверхностью, заполняющей все поле зрения. Не вызовет удивления, если в каких-то расчетах его радиус будет принят равным бесконечности. Для оправдания использования подобной абстракции, если таковое потребуются, можно сказать, что краска покрывает число нитей, значительно превышающее единицу, — обстоятельство, позволяющее при необходимости считать радиус маленького круглого пятнышка-точки равным бесконечности.

Многие физические величины являются функциями координат, или, как иногда говорят, функциями точки. Пусть речь идет о волновой функции элементарной частицы, зависящей от координаты \mathbf{r} (см. гл. 3 второй части книги). Несомненно, под \mathbf{r} следует понимать вектор, проведенный из точки, выбранной нами в качестве начала

координат (с нулевыми координатами), в точку, в которой нас интересует значение ψ -функции. Компоненты вектора \mathbf{r} принято обозначать латинскими буквами x , y и z . Какие значения они могут принимать? «Любые», — ответит неискушенный читатель. «Откуда Вы знаете?» — последует вопрос. Более ста лет математики выясняют, как устроена числовая ось. Об этом рассказано в математической части книги. Числовая ось — создание человеческого ума. Вместе с тем, пространство мы воспринимаем как реальность, как то, *что есть*. Его устройство должно открыться в результате исследования. Если бы мы априори знали, какова структура пространства, исследовать было бы нечего.

В прошлом (наверное, до 20-го века) как правило считали, что вопроса об устройстве (структуре) пространства не существует: оно рассматривалось лишь как вместительное всего. Правда, слова *вместительное всего* недостаточно определены. Помещая нечто в пространство, надо уметь определять его положение. Для этого можно, точнее, необходимо ввести *систему координат*. Выбор за нами. Нужно ясно представлять себе: пока мы не выбрали систему координат, ничего вразумительного о движении тел в пространстве сказать невозможно (или очень трудно)¹⁾.

Повторим вопрос: «Какие значения могут иметь координаты?». Вне зависимости от модификации наших взглядов на пространство–время, ответ кажется очевидным: «Любые ...». Наученные осторожности, подумаем, что означает подобный ответ? По-видимому, признание непрерывности пространства и времени. Сказав, понимаешь, что не так уж это и очевидно. Часы отмеряют время дискретными интервалами, даже такие непрерывнодействующие, как песочные: нельзя измерить песочными часами временной интервал, который меньше времени уменьшения количества песка в верхней чашке на одну песчинку. С началом использования для измерения времени атомных и ядерных процессов дискретные интервалы стали очень малы: для атомных часов $\sim 10^{-16}$ с. Однако возможно, дискретность не является результатом недостатка конкретных технических устройств, а лежит в природе вещей? Тогда из-за единства пространства–времени дискретность времени требует и дискретности пространства.

Дискретность пространства не так трудно себе представить. Возможно, оно напоминает кристаллическую решетку, а точки в нем могут иметь только дискретные координаты — целочисленные значения в длинах элементарной ячейки? Какой величины могла бы быть ячейка? По-видимому, она должна быть меньше 10^{-20} см. До этих расстояний пространство исследовано и дискретность не обнаружена.

¹⁾ Последние слова приведены из осторожности: можно было бы рассматривать только системы тел и изучать их движение относительно друг друга. Однако такой подход очень усложнил бы построение механики. О выборе системы координат и переходе из одной системы в другую рассказано в гл. 4.

Если бы удалось построить теорию дискретного, или, как принято говорить, квантованного пространства, то его ячейка действительно оказалась бы истинно элементарной. Однако пока построить такую теорию никому не удалось, хотя усилий талантливые физики-теоретики потратили и тратят очень много. Почему же они стремятся проквантовать пространство–время? Что побуждает тратить усилия? Казалось бы, нет экспериментальных фактов, которые указывали бы на подобную квантованность. Вместе с тем, имеются разнообразные абстрактные соображения в пользу квантованного пространства. Например, если бы удалось проквантовать пространство, удалось бы избавиться от обращения в бесконечность многих физических величин. Бесконечности мучают физиков-теоретиков уже десятки лет. С ними научились справляться, но методы ликвидации бесконечностей воспринимаются как искусственные приемы и не всех удовлетворяют.

Оставим релятивистскую физику с ее глубокими проблемами. Есть надежда, что в ближайшие годы нас ждут в этой науке удивительные открытия. Значит, возникнут и новые проблемы. . .

Обратимся к макроскопической физике. Большой ее раздел занимается макроскопическими телами, практически игнорируя их атомную дискретную структуру. Конечно, здесь мы снова имеем дело с абстракцией! Называют эту часть макроскопической физики *физикой сплошной среды*. Гидродинамика, теория упругости, макроскопическая электродинамика — примеры наук, рассматривающих физические тела как сплошные. В подавляющем большинстве случаев свойства тел неоднородны — характеристики макроскопических тел являются функциями координат (функциями точки). Например, удельное сопротивление ρ неоднородного проводника зависит от координаты \mathbf{r} , т. е. $\rho = \rho(\mathbf{r})$. Что в данном случае нужно понимать под \mathbf{r} ? Конечно, необходимо быть последовательным: если считаешь тело сплошной средой, считай! Следовательно, \mathbf{r} — просто радиус-вектор с произвольным значением проекций (координат) x , y и z . Однако последовательность требует большего. Надо принять, что элемент объема $dV = dx dy dz$ включает в себя достаточное количество атомов тела. Это позволяет считать dV элементом объема макросистемы¹⁾. Кроме того, чтобы не вступить в противоречие, следует рассматривать только функции, которые почти не меняются на межатомных расстояниях. Так, распространение видимого света, длина волны которого в десятки тысяч раз превышает атомные размеры, можно исследовать в рамках электродинамики сплошных сред, а для рассмотрения рассеяния рентгеновских

¹⁾ При предельном переходе от дискретной среды к сплошной не должно нарушаться условие $dV \gg a^3$, где a — межатомное расстояние. Предполагается, что предельный переход осуществляется в два этапа: сначала устремляют к нулю межатомные расстояния, а потом переходят от конечных приращений к дифференциалам.

лучей, длина волны которых порядка размеров атома, необходимо учитывать атомную структуру тела.

Бесконечно-малый элемент объема dV , превышающий во много раз a^3 , где a — расстояние между атомами, является наглядным и в то же время, конечно, абстрактным образом. Подобные абстрактные образы приходится применять весьма часто. Надо добавить, с большой пользой.

Понятие *квазичастица* — одно из популярнейших в современной физике конденсированного состояния. Можно сказать, что *квазичастица* служит символом квантовой теории твердых тел и квантовых жидкостей.

Приставка «квази» (от лат. *quasy*) означает *якобы, почти* и в обычной речи часто встречается в словах, имеющих уничижительный оттенок (например, квазиученый). В научной терминологии подобный смысл никак не проявляется. Говоря, что процесс *квазилинейный*, предполагают, что он описывается почти линейной зависимостью. И только. С понятием «*квазиклассическое приближение*» мы уже встречались, когда рассматривали переход от квантовой механики к классической.

Впервые термин *квазичастица* возник не в физике твердого тела. То, что свет имеет не только волновую, но и корпускулярную природу, было зафиксировано введением термина *квант света* (А. Эйнштейн, 1905 г.), а в 1929-м году Г. Льюисом¹⁾ был введен термин *фотон*. Думаю, фотон начали в начале 30-х годов именовать *квазичастицей*, подчеркивая ощущение, что основой описания света служат электромагнитные волны, а световые кванты — фотоны — лишь напоминают частицы, они *якобы-частицы, почти-частицы — квазичастицы*. Способствовало подобному ощущению отсутствие у фотона массы. Открытие безмассовых частиц — нейтрино — помогло «превращению» фотона в настоящую частицу. Термин «*квазичастица*» оказался свободным.

Внедрение квантовой механики в физику конденсированного состояния привело к введению представлений о квазичастицах.

Как мы убедимся, различных типов квазичастиц много. Это означает, что термин *квазичастица* — *абстрактное понятие*. С его помощью объединяют кванты элементарных возбуждений разного типа.

*Квазичастица — квант элементарного возбуждения
конденсированного тела.*

Мы хотим разъяснить выделенное утверждение и показать, что введение квазичастиц полезно. В частности, оно придает квантовой теории конденсированного состояния наглядный характер.

¹⁾ Г. Льюис — американский физико-химик (1875–1946 гг.).

Особо важную роль квазичастицы играют в физике кристаллов и в физике квантовых жидкостей. Квантовым жидкостям и квазичастицам в них будет посвящена отдельная глава.

Начнем с кристаллов. Напомним, что главное их свойство — периодичность. Идеальный кристалл бесконечен и построен из тождественных «кирпичиков», из кристаллических ячеек, каждая из которых содержит несколько атомов, ионов или молекул.

Заключительные главы, по нашему мнению, играют весьма важную роль. Предупреждаем, что они сложнее других. Наверное, они потребуют более сосредоточенного чтения, чем предыдущие. Поскольку материала много, мы решили эту главу и главу 19 разделить на сравнительно независимые параграфы, снабдив каждый названием соответствующей квазичастицы.

§ 16.1. Экситон Френкеля

Элементарное возбуждение в изолированном атоме означает переход одного из электронов в свободное возбужденное состояние. При этом, конечно, среди занятых состояний появляется свободное место (дырка). В молекулах картина чуть сложнее: кроме описанных, возможны элементарные возбуждения без перестройки электронных оболочек атомов, входящих в их состав. Атомы в молекуле не закреплены намертво в своих положениях равновесия; они могут колебаться. Подобные колебания осуществляют элементарные возбуждения, которых не возникает в атомах.

Наглядную картину элементарных возбуждений атомов и молекул можно конкретизировать, добавив численные характеристики их энергий: *энергия электронного возбуждения* $\varepsilon_{эл}$ — порядка эВ; *энергия колебательного возбуждения* $\varepsilon_{кол} = \hbar\omega$, где ω — частота колебания, которая имеет порядок $10^{12 \div 13} \text{ с}^{-1}$, т. е. $\varepsilon_{кол} \sim 10^{-2} \div 10^{-3}$ эВ.

Первой квазичастицей, о которой мы постараемся рассказать, будет *экситон*. Введен экситон в физику твердого тела Я.И. Френкелем в 1931-м году. Он же ввел в употребление слово *экситон*, образовав его от английского *excitation* — возбуждение. Слово *экситон* построено по тому же принципу, что и названия частиц: электрон, протон, нейтрон и т. д.

Что же такое экситон?

Если изолированные атом или молекула находятся в возбужденном состоянии, то перейти в основное они могут лишь одним способом — излучив излишнюю энергию в виде кванта света. Из-за малости заряда электрона (см. гл. 13) процесс излучения света атомом, обусловленный взаимодействием электронов с электромагнитным полем, сравнительно

но медленный ¹⁾). Поэтому при расчете атомных энергетических уровней от излучения можно вовсе абстрагироваться и считать *стационарными* не только основное, но и возбужденные его состояния.

Какова же судьба электронного возбуждения в кристалле? От излучения света мы по-прежнему будем абстрагироваться. Вспомним о неразличимости атомов. Что может помешать некоторому атому передать возбуждение соседнему? Ничто. Более того, такая передача обязательно происходит, делая возбужденное состояние выделенного атома *нестационарным*. Анализ показывает, что *стационарным является делокализованное состояние*. Возбужденное состояние не высвечивается, а перемещается по кристаллу от одного атома к другому. *Перемещающееся по кристаллу возбужденное состояние атома или молекулы и есть экситон*.

Чуть подробнее. Возбужденное состояние атома (молекулы) означает, что в электронной оболочке атомной частицы, наряду с электроном в возбужденном состоянии, появилась дырка — состояние, не заполненное электроном. Как мы уже знаем, дырка ведет себя как положительно заряженная частица. Значит, возбужденное состояние атома (молекулы) является комбинацией дырки и возбужденного электрона. О других электронах можно просто не думать.

Таким образом, экситон есть движение связанных между собой электрона и дырки. Экситон нейтрален. Его перемещение не сопровождается переносом заряда. Чтобы подчеркнуть структуру экситона, в котором расстояние между электроном и дыркой не больше размера атома или молекулы, его часто называют *экситоном малого радиуса* или *экситоном Френкеля* (по имени ученого, предсказавшего его существование). Об экситонах большого радиуса будет рассказано ниже.

Движение экситона по кристаллу очень напоминает движение свободной квантовой частицы в пустом пространстве. Стационарным состоянием свободной частицы является состояние с определенным импульсом. Ее пси-функция — плоская волна. Геометрия пространства, в котором движется частица, проявляется в динамике частиц. В частности, именно однородность пространства имеет своим следствием тот факт, что *импульс* любой величины и любого направления (импульс — вектор) есть сохраняющаяся величина. Это позволяет заданием импульса характеризовать стационарные состояния свободной частицы.

Состояние экситона в кристалле определяется вектором, очень похожим по своим свойствам на импульс. Он называется *квазиимпульсом*.

Наличие в кристалле экситона не предполагает движения атомов. От узла к узлу перемещается только возбуждение, а не сам атом.

¹⁾ Говоря на языке классической физики, возбужденный электрон успевает многократно облететь вокруг ядра прежде, чем излучит квант света и перейдет в основное состояние.

Последнее дает возможность считать кристалл пространством с целочисленными координатами, в котором движется экситон. Пространство, в котором *существует* экситон, не однородно, а периодически. Мы нарочно употребили слово *существует*, чтобы подчеркнуть: на все, что происходит с экситоном, накладывает отпечаток периодичность этого пространства.

Как же проявляется периодичность пространства?

Давайте мысленно построим пространство квазиимпульсов, т. е. систему координат, на осях которой отложены компоненты квазиимпульса. Мы уже строили пространство импульсов (см. гл. 10). В отличие от него, *пространство квазиимпульсов периодически*, как и сам кристалл. По своим размерам ячейка пространства квазиимпульсов обратно пропорциональна размерам ячейки кристалла. Поэтому пространство квазиимпульсов часто называют *обратным пространством*.

Из факта периодичности обратного пространства (забегая вперед, отметим, что у любой квазичастицы обратное пространство периодически) следуют важные выводы. Вот один из них. Значения энергии экситона заполняют некоторую полосу. Ее называют *зоной*.

Доказательство последнего утверждения столь просто, что его можно изложить. Энергия экситона зависит от квазиимпульса периодически. Где-то она имеет минимум, где-то — максимум. Между минимальным и максимальным располагаются все остальные значения энергии. Они и составляют зону.

Можно рассуждать чуть конкретнее. В кристалле содержится N атомов. Какой из них возбужден, безразлично. В таких случаях принято говорить о вырождении. Состояние кристалла с одним возбужденным атомом N -кратно вырождено. Правда, здесь не учтено взаимодействие атомов. Из-за взаимодействия энергия возбужденного состояния (теперь уже надо говорить не о состоянии отдельного атома, а о состоянии всего кристалла) зависит от того, какова энергия перемещения возбуждения по кристаллу. В результате энергетический уровень атома (молекулы) расширяется в полосу-зону. Обычно ширина экситонной зоны заметно меньше, чем энергия электронного возбуждения атома или молекулы. Проведенное рассмотрение подсказывает оценку ширины зоны. Она имеет порядок энергии взаимодействия между соседними атомами или молекулами.

Сходство квазиимпульса с импульсом столь велико, что скорость экситона v и скорость свободной частицы выражаются одной и той же формулой:

$$v = \frac{d\varepsilon}{dp}, \quad (16.1)$$

только в случае экситона p — квазиимпульс, а энергия экситона $\varepsilon \neq p^2/2m$. Как мы уже знаем, $\varepsilon = \varepsilon(p)$ — периодическая функция квазиимпульса p . В минимуме энергии скорость экситона равна нулю.

Если квазиимпульс отсчитывать от точки, в которой энергия минимальна, то разложение последней по степеням квазиимпульса начинается с квадратичного члена:

$$\varepsilon = \varepsilon_{\min} + \left. \frac{d^2 \varepsilon(\mathbf{p})}{d(\mathbf{p})^2} \right|_{p=0} \left(\frac{\mathbf{p}^2}{2} \right). \quad (16.2)$$

Разложение (16.2) дает возможность ввести понятие *эффективной массы*, позволяющее придать зависимости энергии экситона от квазиимпульса привычный вид:

$$\varepsilon = \frac{\mathbf{p}^2}{2m_{\text{эф}}}, \quad \frac{1}{m_{\text{эф}}} = \left. \frac{d^2 \varepsilon(\mathbf{p})}{d(\mathbf{p})^2} \right|_{p=0}. \quad (16.3)$$

Энергию мы отсчитываем от ее минимального значения ε_{\min} . Необходимо, конечно, помнить, что формула (16.3) справедлива только для тех экситонов, энергия которых лишь чуть больше минимальной.

Нетрудно сообразить, что у экситонов с энергией, близкой к максимальной, эффективная масса отрицательна (!).

Как правило, энергия электронного возбуждения атома (молекулы) столь велика, что появление экситонов за счет теплового возбуждения маловероятно и им можно пренебречь. Таким образом, в тепловых свойствах кристаллов экситоны роли не играют. Их роль велика в оптических свойствах кристаллов. Более того, экситоны были «созданы», чтобы объяснить оптические свойства некоторых кристаллов, поглощение света в которых не сопровождается появлением фотопроводимости: энергия фотона тратится не на ионизацию, а на создание нейтрального экситона.

Прежде чем рассказывать о других квазичастицах, надо обратить внимание на следующий факт. Среди отмеченных свойств экситонов есть такие, которыми должны обладать все квазичастицы в кристаллах. Перечислим их:

- 1) состояние любой квазичастицы определяется ее квазиимпульсом;
- 2) пространство квазиимпульсов периодически;
- 3) энергия — периодическая функция квазиимпульса; разрешенные значения энергии заполняют зону конечной ширины;
- 4) скорость любой квазичастицы определяется формулой (16.1)¹).

§ 16.2. Фононы

Фононы — важнейшие из квазичастиц. К этому утверждению надо отнестись, как к любой рекламе, чуть иронически. Однако обратить

¹) Формулы (16.2), (16.3) могут быть применены не ко всем квазичастицам (см. ниже).

на него внимание имеет смысл. Важность фононов заставляет нас рассказать о них подробнее, чем о других квазичастицах.

Придется начать издалека, с описания тепловых свойств твердых тел.

Любое твердое тело, казалось бы, может служить символом неподвижности. Однако мы понимаем, что все тела состоят из атомов, ионов, молекул. Знание строения вещества заставляет задуматься, как движутся частицы, из которых состоят тела, что представляет собой тепловое движение частиц в теле. Концепция квазичастиц призвана ответить на эти вопросы.

Тепловое движение часто называют *хаотическим*. По-моему, естественное понимание словосочетания *хаотическое движение* таково: как угодно, беспорядочно.

Давайте буквально на несколько минут «оторвемся» от твердого тела и подумаем, что означает выражение «хаотическое движение» для газа. Значит ли оно, что частицы газа, принимающие участие в тепловом движении, могут двигаться по окружностям или колебаться вокруг каких-либо точек в сосуде? «Да нет, — отмахнется любой. — Это означает, что частицы между столкновениями летят по разным случайным направлениям, так что в среднем газ никуда не перемещается». А с какими скоростями движутся частицы?

Не будем продолжать воображаемую беседу, просто вспомним то, о чем мы уже говорили в 10-й главе. При высоких температурах любые частицы (и фермионы, и бозоны) движутся одинаково: среднее значение квадрата их скорости $\langle v^2 \rangle$ определяется температурой газа T ; однако при понижении температуры фермионы и бозоны ведут себя различно.

Воспользуемся примером газа для того, чтобы сделать два важных для дальнейшего вывода. Один из них очевиден: частицы, которые осуществляют тепловое движение, двигаются так, как *могут* двигаться (только в газе, где частицы свободны, они перемещаются с постоянной скоростью). Второй вывод менее очевиден: квантовые особенности теплового движения проявляются при понижении температуры.

Запомним сказанное и, вернувшись к твердому телу, зададим себе вопрос: «Можно ли на опыте определить, как движутся микрочастицы, из которых построено твердое тело?».

В настоящее время имеется множество методов исследования движения микрочастиц. Мы же поставим себя в условия, когда физики лишь подозревали о существовании атомов и молекул, а объектами их исследования были *макроскопические* тела. Можно ли узнать, чему равна внутренняя тепловая энергия тела? Оказывается, не только можно, но и принципиально несложно. Правда, не саму тепловую энергию W , а ее производную по температуре T — *теплоемкость* $C = dW/dT$. Выписанная формула является определением: теплоемкость есть отношение dW к dT , где dW — поглощаемое телом

бесконечно малое количество тепла при бесконечно малом повышении температуры на dT . Следовало бы уточнить, в каких условиях происходит повышение температуры: от этого зависит количество поглощенной теплоты. Однако для твердых тел теплоемкость слабо зависит от условий эксперимента. Поэтому мы не будем их оговаривать.

Возникла специальная дисциплина, разрабатывающая экспериментальные методы измерения тепловых характеристик тел, прежде всего теплоемкости. Эта дисциплина называется *калориметрия* (от лат. *calor* — тепло и греч. *metreo* — измеряю; такое вот греко-латинское слово). Появилась калориметрия в XIX веке, но методы измерений непрерывно совершенствуются; в настоящее время теплоемкость можно измерять с большой точностью.

К каким же результатам приводит измерение теплоемкости твердых тел? Если проводить его при высоких температурах и ограничиться сравнительно простыми по составу телами, то результат можно сформулировать в виде *закона Дюлонга и Пти*, согласно которому *молярная* теплоемкость при постоянном объеме для всех простых твердых тел одинакова и составляет приблизительно 25 Дж/моль · К. Это число нашло объяснение. Будем исходить из простейшей модели твердого тела, считая, что каждый атом совершает малые колебания вокруг положения равновесия. При этом он принимает участие в трех независимых колебаниях. Если T — температура тела, то средняя энергия колебаний атома равна $3k_B T$ (где k_B — постоянная Больцмана). В моле вещества N атомов, где N — число Авогадро. Оказывается, что $3Nk_B = 24,9$ Дж/моль · К.

Для дальнейшего следует обратить внимание не столько на согласие вышеизложенного с результатами экспериментов, сколько на имеющие место регулярные расхождения.

То, что теплоемкость простых твердых тел подчиняется закону Дюлонга и Пти по сути означает только правильность наших представлений о том, что атомы совершают малые колебания вокруг положений равновесия. Но если так, то молярная теплоемкость *всегда* должна быть равна $3Nk_B$. Однако с понижением температуры теплоемкость всех твердых тел всегда уменьшается.

Со сложными кристаллами ситуация несколько сложнее. Не всегда легко показать, что их теплоемкость удовлетворяет закону Дюлонга и Пти. Причина заключается в том, что температура, при которой производят измерения, низка. Существует уверенность, что если бы температура была более высокой, то согласие с законом Дюлонга и Пти было бы. Не всегда, правда, удается в этом убедиться, так как тело раньше плавится, чем достигается нужная температура.

Альберт Эйнштейн уже в 1907-м году понял: причина несогласий в том, что при низких температурах нельзя пользоваться классическими выражениями; необходимо учитывать квантование энергии малых колебаний.

За несколько лет до этого для объяснения законов излучения Макс Планку пришлось предположить, что энергия микроскопического осциллятора может принимать лишь дискретные значения, равные $n\hbar\omega$, где n — целые положительные числа, ω — частота колебаний, а \hbar — постоянная Планка, деленная на 2π (см. гл. 3). В дальнейшем выяснилось, что энергетические уровни осциллятора суть $(n + 1/2)\hbar\omega$.

С принципиальной точки зрения слагаемое $(1/2)\hbar\omega$ очень существенно. Его отсутствие означало бы, что энергия основного состояния осциллятора равна нулю. Однако тогда и импульс, и координата колеблющейся частицы *одновременно* равнялись бы нулю, что противоречит соотношениям неопределенности. Движение осциллятора в основном состоянии, при котором его энергия равна $(1/2)\hbar\omega$, называют *нулевым колебанием*, подчеркивая, что при этом $n = 0$.

Введение в физику постоянной Планка, как мы знаем, было событием огромного масштаба: родилась новая, *квантовая* физика. Работа Эйнштейна, названная им «Теория излучения Планка и теория удельной теплоемкости», по праву знаменует рождение *квантовой теории твердого тела*.

Эйнштейн ограничился простейшим предположением, считая, что все атомы твердого тела колеблются с одинаковой частотой. Будем производить расчет средней энергии $\langle E \rangle$ осциллятора, исходя из идеи Планка. Зная, что вероятность состояния осциллятора с энергией $n\hbar\omega$ пропорциональна $\exp(-n\hbar\omega/k_B T)$, нетрудно вычислить его среднюю энергию:

$$\langle E \rangle = \frac{\hbar\omega}{\exp(\hbar\omega/k_B T) - 1}. \quad (16.4)$$

Видно, что при высоких температурах ($\hbar\omega \ll k_B T$) средняя энергия осциллятора, как ей и положено по законам классической физики, равна $k_B T$, а с понижением температуры уменьшается. Вместе со средней энергией уменьшается и теплоемкость. При стремлении температуры к абсолютному нулю ($\hbar\omega \gg k_B T$) теория Эйнштейна предсказывает экспоненциальное падение теплоемкости до нуля:

$$C = 3Nk_B \left(\frac{\hbar\omega}{k_B T} \right)^2 \exp(-\hbar\omega/k_B T). \quad (16.5)$$

Эксперимент определенно указывает, что теплоемкость твердых тел стремится к нулю значительно медленнее, чем предсказывает теория Эйнштейна.

Более реалистическая модель принадлежит Питеру Дебаю, справедливо считающемуся одним из основоположников квантовой теории твердого тела. Работа, о которой пойдет речь, вышла из печати в 1912-м году. Основываясь на идее Эйнштейна о необходимости учитывать квантование энергии осцилляторов, к их введению Дебай подошел совсем иначе, чем Эйнштейн. Его модель как бы игнорировала атомное

строение кристалла. Твердое тело он рассматривал с позиции механики сплошных сред, считая его *упругим континуумом*. Согласно модели упругого континуума тепловое движение может происходить только в виде упругих колебаний, а последние в виде звуковых волн распространяются по твердому телу.

Звуковую волну, как и любую другую, можно характеризовать волновым вектором \mathbf{k} и частотой ω , причем между частотой и волновым вектором существует простая линейная связь: $\omega = sk$, где s — скорость звука. В упругом теле, в отличие от жидкостей и газов, звуковые волны бывают трех типов. Одна — продольная, а две — поперечные. Скорости волн с помощью уравнений теории упругости выражаются через упругие характеристики тела. Продольная и поперечные скорости несколько различаются по величине: скорость продольного звука $s_{\text{прод}}$ больше скоростей поперечного звука $s_{\text{поп}}$, которые в изотропном упругом теле равны между собой (эффектов, обязанных своим существованием анизотропии, мы рассматривать не будем). Итак, звуковые волны в твердом теле бывают

$$\begin{aligned} \text{продольные} \quad \mathbf{u}_{\text{прод}} &= \mathbf{u}_0^{\text{прод}} \exp i(\mathbf{k}\mathbf{r} - \omega_{\text{прод}}t), & \omega_{\text{прод}} &= s_{\text{прод}}k, \\ \text{поперечные (их две)} \quad \mathbf{u}_{\text{поп}} &= \mathbf{u}_0^{\text{поп}} \exp i(\mathbf{k}\mathbf{r} - \omega_{\text{поп}}t), & \omega_{\text{поп}} &= s_{\text{поп}}k. \end{aligned} \quad (16.6)$$

Амплитуда продольной волны $\mathbf{u}_0^{\text{прод}}$ параллельна волновому вектору \mathbf{k} , а амплитуды поперечных волн $\mathbf{u}_0^{\text{поп}}$ перпендикулярны \mathbf{k} . Тип волны называют ее *поляризацией*.

Заставить колебаться отдельный атом (или выделенный элемент объема тела) практически невозможно. Из-за взаимодействия между атомами локализованное колебание немедленно превратится в волну. Звуковые волны тоже взаимодействуют между собой, но во много раз слабее, чем отдельные атомы тела друг с другом. В хорошем приближении их можно считать вообще невзаимодействующими. Такой подход — пример разумной идеализации.

Теперь, опустив индексы, указывающие тип волны, выделим временную зависимость:

$$q(t) = q(0) \exp(-i\omega t). \quad (16.7)$$

Уравнение, которому подчиняется функция $q(t)$, есть уравнение малых колебаний:

$$\frac{d^2 q(t)}{dt^2} + \omega^2 q(t) = 0. \quad (16.8)$$

Какова при классическом подходе энергия каждой звуковой волны, то есть волны определенной поляризации и с определенным волновым вектором \mathbf{k} ? Все зависит от амплитуды, которая описывается функцией $q(t)$ и, естественно, может быть любой. Энергия определяется величиной $q^2(0)$.

При квантовом подходе энергия звуковой волны определяется допустимыми значениями энергии малых колебаний, которые, как мы знаем, принимают дискретные значения, равные $n\hbar\omega$, где $n = 0, 1, 2, \dots$ — любые целые числа. Итак, энергия звуковой волны E_k^j в теле принимает квантованные значения, а именно $n\hbar\omega^j(\mathbf{k})$ с $n = 0, 1, 2, \dots$. Индекс j принимает три значения и обозначает поляризацию волны.

Используя модель упругой непрерывной среды, Дебай, конечно, понимал, что она применима только до тех пор, пока длина звуковой волны ($\lambda = 2\pi/k$) значительно превосходит межатомные расстояния. В случае коротких волн необходим микроскопический подход, основанный на исследовании колебаний атомов кристаллической решетки. В дальнейшем колебания молекул и атомов кристаллических решеток были тщательно изучены. Дебай, пытаясь предельно упростить задачу, выдвинул изящную идею. Он предположил, что линейная зависимость частоты колебаний от волнового вектора не нарушается, но величина волнового вектора не может быть больше некоторого значения, которое естественно обозначить k_D . Как же выбрать значение предельного волнового вектора? Ответ прост. И в его простоте — успех модели. Закон Дюлонга и Пти свидетельствует о том, что при высоких температурах все имеющиеся в теле осцилляторы дают одинаковый по величине вклад во внутреннюю (тепловую) энергию тела. При этом вклад каждого осциллятора — его средняя энергия — вовсе не зависит от частоты. Следовательно, правильное значение теплоемкости при высоких температурах получится, если полное число осцилляторов приравнять утроенному числу атомов в теле. Отсюда

$$k_D = \left(\frac{6\pi^2 N}{V} \right)^{1/3}, \quad (16.9)$$

где N — число атомов в кристалле, V — объем кристалла. В случае простой кубической решетки с одним атомом на ячейку $k_D = (6\pi^2)^{1/3}/a$, где a — линейный размер ячейки кристалла.

Повторим: по Дебаю, тепловое движение в твердом теле осуществляют звуковые волны. Если абстрагироваться от взаимодействия между волнами, то тепловая энергия тела окажется равной сумме энергий отдельных звуковых волн. Каждой отдельной волне можно поставить в соответствие осциллятор с частотой $\omega^j(\mathbf{k})$. Следовательно, частоты колебаний осцилляторов одного сорта, соответствующего фиксированной поляризации волны, занимают полосу: $0 < \omega < \omega_D^j = s_j k_D$.

Энергия каждого осциллятора может принимать целочисленные, квантованные значения: $n\hbar\omega^j(\mathbf{k})$, где $n = 0, 1, 2, \dots$ — любые целые числа. Среднее значение энергии осциллятора по-прежнему определяется формулой Планка (16.4), в которой, конечно, надо учесть зависимость частоты от волнового вектора и от поляризации.

Для вычисления колебательной части тепловой энергии тела следует найти сумму средних энергий всех осцилляторов, то есть произвести интегрирование по k внутри сферы радиуса k_D (см. (16.9)). Результат дает возможность рассчитать зависимость теплоемкости C от температуры T , которая хорошо описывает теплоемкость простых твердых тел в широком диапазоне температур. При высоких температурах мы, естественно, получим закон Дюлонга и Пти, а при стремлении температуры к нулю теплоемкость стремится к нулю пропорционально T^3 :

$$C(T) = (2\pi^2/5)Nk_B(T/\theta_D)^3, \quad (16.10)$$

где θ_D — температура Дебая. Это характеристическая температура в описываемой модели. Здесь

$$\theta_D = \hbar\omega_D/k_B; \quad \omega_D = \langle s \rangle k_D; \quad 3/\langle s \rangle^3 = 2/s_{\text{поп}}^3 + 1/s_{\text{прод}}^3. \quad (16.11)$$

Среднее значение скорости звука $\langle s \rangle$ выбрано с учетом суммарного вклада волн всех поляризаций в тепловую энергию тела.

Температура Дебая *отделяет* область низких температур, требующую квантового подхода, от высоких температур, при которых классическое рассмотрение не вносит заметной ошибки.

Следует отметить, что в таком виде определение температуры Дебая не связано напрямую с моделью Дебая. В более реалистической модели тоже можно ввести температуру, выше которой использование классического подхода вполне оправдано.

Обычно температура Дебая составляет приблизительно 100 ÷ 200 К, так что комнатная температура оказывается высокой, а теплоемкость при ней удовлетворяет закону Дюлонга и Пти. Однако для разных веществ температуры Дебая довольно значительно различаются. Вот несколько примеров значений температур Дебая:

$$\begin{aligned} \text{Pb} &— 90 \text{ К}, & \text{Ag} &— 210 \text{ К}, \\ \text{KBr} &— 180 \text{ К}, & \text{NaCl} &— 280 \text{ К}, & \text{C (алмаз)} &\sim 2000 \text{ К}. \end{aligned}$$

Простота и успешность модели Дебая дали возможность с ее помощью сформулировать интерполяционную модель, которая удовлетворительно описывает температурную зависимость многих твердых тел. Подчеркнем, что в двух предельных случаях, вблизи абсолютного нуля и при температурах, существенно превышающих θ_D , модель Дебая дает точные значения.

Существование двух разных моделей колебаний атомных частиц твердого тела (Эйнштейна и Дебая) позволяет показать возможность и пользу сочетания моделей.

Если в каждой элементарной ячейке кристалла находится двух-атомная молекула, атомы в которой сильно связаны друг с другом, то в звуковых колебаниях она принимает участие как нечто целое.

Модель Дебая хорошо описывает эти колебания и их вклад в теплоемкость с тем очевидным изменением, что под N следует понимать не число атомов, а число молекул, совпадающее с числом элементарных ячеек кристалла.

Колебания атомов в каждой из молекул хорошо описываются моделью Эйнштейна, которую можно несколько усовершенствовать. Поскольку каждая из молекул не изолирована от своих соседей, к молекулярным колебаниям естественно применить рассуждения, изложенные при описании экситона Френкеля. Колебания, распространяющиеся по кристаллу, часто так и называют — экситонами, а иногда им присваивают особое наименование — *оптические колебания*. Некоторые (но не все!) из них проявляют себя в оптических свойствах кристаллов. Этому все, кроме описанных выше акустических, колебания и обязаны своим названием.

Частота молекулярных колебаний ω_0 определяет характеристическую температуру в модели Эйнштейна:

$$\theta_E = \hbar\omega_0/k_B. \quad (16.12)$$

Ширина зон оптических колебаний обычно заметно меньше $\hbar\omega_0$. Если это так, то формулы (16.4) и (16.5) теории Эйнштейна непосредственно применимы к оптическим колебаниям кристалла.

Теперь можно продемонстрировать, как объединение двух моделей описывает тепловые свойства кристалла. При низких температурах, когда вклад оптических колебаний экспоненциально мал, тепловые свойства определяются акустическими колебаниями, причем в формуле (16.10) под буквой N надо понимать число молекул, совпадающее с числом ячеек кристалла. При самых высоких температурах ($T \gg \theta_E, \theta_D$) выполняется закон Дюлонга и Пти, в котором N надо заменить на νN , где ν — число атомов в элементарной ячейке (в описываемом примере $\nu = 2$). В промежуточной области температур вклад в теплоемкость твердого тела дают и акустические, и оптические колебания.

Конспективно, но не прибегая к упрощающим моделям, опишем, что собой представляют колебания атомных частиц, из которых построена кристаллическая решетка твердого тела.

Если в элементарной ячейке кристалла содержится ν атомов, то существует 3ν ветвей колебаний. Они отличаются поляризацией. Частоты каждой из ветвей заполняют полосу (зону). Частота — периодическая функция квазиволнового вектора \mathbf{k} . Как и пространство квазиимпульсов, пространство квазиволновых векторов периодически. Три из 3ν ветвей колебаний называются *акустическими*, а остальные $3\nu - 3$ — оптическими. Минимальные частоты акустических ветвей равны нулю. Если длина k квазиволнового вектора мала и удовлетворяет условию $ak \ll 1$, то зависимость частот от квазиволнового вектора линейна (см. (16.6)).

Мы видим, что объединение моделей Эйнштейна и Дебая хорошо описывает истинное положение вещей.

Следующий шаг — введение квазичастиц-фононов.

Элементарным возбуждением следует считать волну, энергия которой равна $\hbar\omega$. То обстоятельство, что энергия каждой звуковой волны есть сумма энергий квантов ($\hbar\omega$), позволяет считать, что звуковая энергия распространяется по твердому телу в виде квантов — квазичастиц-фононов.

Как и всякая квазичастица, фонон имеет квазиимпульс $\mathbf{p} = \hbar\mathbf{k}$, а его энергия $\varepsilon = \hbar\omega_j(\mathbf{k})$, где индекс j обозначает сорт фонона — поляризацию звуковой волны. Вводя фононы, степень возбуждения осциллятора n следует воспринимать как число фононов данного сорта. Оно может быть любым ($n = 0, 1, 2, \dots$). Значит, *фононы* являются *бозонами* (см. гл. 7).

Формула Планка для среднего (равновесного) числа фононов имеет вид

$$\langle n \rangle = [\exp(\hbar\omega_j/kT) - 1]^{-1}. \quad (16.13)$$

Мы сознательно подчеркнули, что она описывает равновесное распределение фононов. В теле, которое не находится в термодинамическом равновесии, распределение фононов может весьма существенно отличаться от планковского (см. гл. 10).

Совокупность фононов можно рассматривать как *газ бозонов*, поскольку формула Планка совпадает с функцией Бозе–Эйнштейна. Правда, его химический потенциал $\zeta \equiv 0$. Так и должно быть. Ведь химический потенциал представляет собой нормировочную постоянную, задача которой — обеспечить постоянное значение полного числа частиц газа. Однако газ фононов — особый газ, число квазичастиц в котором зависит от температуры: при низких температурах их мало, с ростом T их число растет.

Используя формулу (16.13), можно найти суммарное число фононов N_Φ в кристалле. При низких температурах оно уменьшается с T по степенному закону (пропорционально T^3). При этом все фононы — акустические. Оптические вымерзают. При высоких температурах ($T \gg \theta_D$) в теле присутствуют фононы всех сортов; их число линейно растет с температурой и существенно превышает число атомов тела ($N_\Phi \sim N(T/\theta_D)$). Последнее не должно вызывать удивления, поскольку степень возбуждения каждого осциллятора высока ($\langle n \rangle \gg 1$).

Подчеркнем, что возрастание числа квазичастиц с ростом температуры, а значит, и равенство нулю химического потенциала — общее свойство всех газов квазичастиц. При $T = 0$ (в *основном* состоянии) квазичастиц в теле вообще нет, поскольку они являются квантами элементарных *возбуждений*.

К фононам, как и к любым квазичастицам, применимы формулы (16.1)–(16.3). Правда, к акустическим фононам при низких частотах формулы (16.2), (16.3) не применимы; для них нельзя ввести эффективную массу, поскольку частоты акустических волн (энергии акустических фононов) обращаются в нуль по линейному закону. Вне зависимости от принятого метода расчета колебаний атомных частиц кристалла, акустические фононы с длиной волны, существенно превышающей постоянную решетки, представляют собой обыкновенные звуковые волны (см. (16.6)).

Назвав фононы важнейшими квазичастицами, мы хотели подчеркнуть, что именно они — главные носители тепла в твердом теле, по крайней мере при высоких температурах. Кроме того, фононы присутствуют в твердых телах любого типа. Другие квазичастицы значительно более специфичны. Описывая некоторые из них, мы укажем, в телах какого типа они существуют.

Хочется, чтобы у читателя сложилось следующее представление. Если в твердом теле микроскопические частицы могут каким-то образом двигаться, то каждому типу движения можно сопоставить определенный тип квазичастиц.

§ 16.3. Магноны

В настоящем параграфе речь пойдет о магнетиках. К сожалению, размеры книги не позволяют подробно остановиться на природе магнитных явлений и на свойствах магнетиков. Мы ограничимся описанием структуры основного состояния магнетиков и соответствующих специфических квазичастиц.

Отличительная черта атомов, из которых построен кристалл магнетика, состоит в наличии у них магнитных моментов. Обозначим магнитный момент атома буквой μ (он является вектором). При высокой температуре магнитные моменты направлены в пространстве случайным образом, так что суммарный магнитный момент тела равен нулю.

Тепловое движение атомных частиц магнетика не исчерпывается колебаниями. Их магнитные моменты могут спонтанно (не под воздействием внешней силы) менять свое направление. Изменение направления магнитных моментов — специфическое движение микроскопических магнетиков. Оно, естественно, отсутствует у атомных частиц без магнитных моментов.

Внешнее магнитное поле \mathbf{H} намагничивает тело. При высоких температурах и небольшом по величине магнитном поле тело приобретает магнитный момент \mathbf{M} , пропорциональный \mathbf{H} . Если магнитный момент отнести к единице объема, то коэффициент пропорциональности между \mathbf{M} и \mathbf{H} безразмерен:

$$\mathbf{M} = \kappa \mathbf{H}. \quad (16.14)$$

Коэффициент пропорциональности κ называют магнитной восприимчивостью. Все тела, состоящие из частиц, имеющих магнитные моменты, являются *парамагнетиками*. Это утверждение означает, что $\kappa > 0$.

Если в ячейке кристалла содержится несколько атомов, но только один из них имеет магнитный момент, то кристалл все равно является парамагнетиком, хотя атомы, не обладающие магнитными моментами, намагничиваются против магнитного поля — они *диамагнитны*. По модулю диамагнитная восприимчивость $\kappa_{\text{диам}}$ как правило значительно меньше, чем парамагнитная. Это следствие малости безразмерного заряда электрона (см. гл. 13).

Поскольку энергия взаимодействия магнитного момента μ с магнитным полем \mathbf{H} равна $-\mu\mathbf{H}$, атомным магнитным моментам выгоднее быть направленными по магнитному полю, а не против него. В этом и состоит причина парамагнетизма.

Мы пока не упоминали о силах взаимодействия между магнитными моментами. При высоких температурах они не очень важны, хотя их существование несколько изменяет температурную зависимость магнитной восприимчивости. При понижении температуры силы взаимодействия начинают играть определяющую роль. Что же происходит?

При некоторой температуре происходит фазовый переход 2-го рода. Ниже нее атомные магнитные моменты начинают упорядочиваться (см. гл. 9). В результате при абсолютном нуле температуры все микроскопические магнетики оказываются выстроены в строгом порядке.

Силы, заставляющие магнитные моменты выстраиваться, имеют квантовую природу (в классической физике им нет аналога) и называются *обменными* (мы уже упоминали их в 6-й главе, когда рассказывали о принципиальной неразличимости квантовых частиц). В зависимости от кристаллической структуры магнетика и от обменных сил атомные магнитные моменты могут в разных телах выстраиваться по-разному. Остановимся на двух предельных случаях: в *ферромагнетиках* при $T = 0$ все атомные магнетики параллельны друг другу, а в *антиферромагнетиках* каждый магнетик окружен антипараллельными соседями.

Для простоты будем считать, что спин каждого магнитного атома равен $1/2$. Значит, ориентироваться он может лишь двумя способами. Если перевернуть один из магнетиков, то энергия магнетика повысится по атомным масштабам довольно значительно. Обозначим ее изменение буквой J . Оценить J можно, заметив, что температура фазового перехода из парамагнитного состояния в магнито-упорядоченное по порядку величины определяется значением J . Температуру фазового перехода в ферромагнитное состояние T_C называют температурой Кюри, а температуру T_N перехода в антиферромагнитное состояние — температурой Нееля:

$$\frac{J}{k_B} \sim T_C \text{ или } T_N. \quad (16.15)$$

Задумаемся над судьбой перевернутого атомного магнитного момента. Из-за взаимодействия с соседями он не сможет закрепиться на каком-нибудь определенном атоме. Как и в случае экситона, или возбужденного колебательного состояния молекулы, стационарно не локализованное состояние неправильно направленного магнитного момента, а волна — перевернутый магнитный момент создает эстафету переворачиваний.

Волна переворачиваний называется *спиновой*. Спиновая волна — специфическое элементарное возбуждение магнетика. Соответствующая ей специфическая для магнетиков квазичастица — *магнон*.

Существование в ферромагнетиках спиновых волн как элементарных возмущений впервые было предсказано Феликсом Блохом в 1930-м году.

С помощью спиновой волны возмущение, обязанное своим существованием неправильному положению атомного магнетика, «размазывается» по всему кристаллу. В среднем направления магнитных моментов соседних атомов тем ближе друг к другу, чем больше длина спиновой волны. Поэтому с ростом длины волны энергия магнона уменьшается. Правда, уменьшается различно в ферромагнетиках и в антиферромагнетиках. Ниже мы приведем две формулы, а пока повторим: энергия магнона, как любой квазичастицы, является периодической функцией квазиимпульса.

Напомним, что квазиимпульс пропорционален квазиволновому вектору ($\mathbf{p} = \hbar\mathbf{k}$), а квазиволновой вектор (как и волновой вектор) обратно пропорционален длине волны ($k = 2\pi/\lambda$). Таким образом, длинноволновым спиновым волнам соответствуют магноны с малыми квазиимпульсами.

При $p \rightarrow 0$ магноны в ферромагнетиках и в антиферромагнетиках очень различны. В ферромагнетиках они напоминают квантовые частицы:

$$\varepsilon = \frac{p^2}{2m_{\text{эф}}}, \quad \frac{1}{m_{\text{эф}}} = \frac{2Ja^2}{\hbar^2}, \quad (16.16)$$

где a — постоянная решетки (расстояние между соседними атомами). Обратите внимание: чем больше J , тем меньше эффективная масса магнона. Это естественно, так как J определяется энергией связи соседних магнитных моментов. Чем она больше, тем легче магнону передвигаться по кристаллу. Скорость магнона $\mathbf{v} = \mathbf{p}/m_{\text{эф}}$ (см. (16.1)). Может показаться, что зависимость от расстояния между атомами противоречит последнему утверждению. Неужели перемещение магнона тем легче, чем больше расстояние между атомами? Конечно, это не так. Дело в том, что обменное взаимодействие, как мы уже отмечали, очень быстро спадает при увеличении расстояния между

атомами. Поэтому произведение Ja^2 с ростом постоянной решетки a уменьшается.

В антиферромагнетиках при $p \rightarrow 0$ магноны напоминают фононы:

$$\varepsilon = up, \quad u = J(a/\hbar). \quad (16.17)$$

Мы видим, что и здесь скорость магнона u тем больше, чем больше J .

Магноны являются бозонами. Вывод базируется на теоретико-физическом анализе, но в случае ферромагнетиков можно привести рассуждение, позволяющее понять, на чем он основан. Спиновая волна — волна отклонения спина (в данном случае удобнее говорить о спине, а не о магнитном моменте). Если по кристаллу распространяется одна спиновая волна, то каково бы ни было значение спина атома, суммарный спин всех атомов тела на единицу отличается от его спина в основном состоянии, в котором все спины имеют максимальную проекцию. Это позволяет приписать магнону спин, равный 1. Отсюда следует, что магнон — бозон.

Одно из отличий магнетиков от других твердых тел состоит в том, что при $T \neq 0$ в них присутствует газ магнонов. Магноны особенно успешно описывают свойства магнетиков при низких температурах ($T \ll T_C, T_N$), когда их сравнительно мало и они редко сталкиваются друг с другом. Тогда газ магнонов можно считать идеальным, а расчеты и все представления особенно просты. Число магнонов $N_{\text{маг}}$ в ферро- и антиферромагнетиках различно:

$$\begin{aligned} \text{для ферромагнетиков} \quad N_{\text{маг}} &\sim N(T/T_C)^{3/2}, \\ \text{для антиферромагнетиков} \quad N_{\text{маг}} &\sim N(T/T_C)^3. \end{aligned}$$

Приведенная оценка позволяет понять, как зависят от температуры некоторые характеристики магнетиков.

У ферромагнетиков отклонение величины магнитного момента $M(T)$ от его максимального значения, достигаемого при $T = 0$, обязательно своим существованием магнонам. Поэтому при $T \ll T_C$

$$M(T) = M(0)[1 - \alpha(T/T_C)^{3/2}], \quad (16.18)$$

где α — численный множитель, различный для разных магнетиков. Формулу (16.18) называют *законом Блоха*. Магноны также вносят вклад в теплоемкость. При низких температурах он пропорционален $(T/T_C)^{3/2}$.

У антиферромагнетиков при $T \ll T_N$ вклад магнонов в теплоемкость пропорционален третьей степени температуры. Вклады магнонов и фононов можно разделить, поскольку магнонная теплоемкость зависит от магнитного поля (мы на этой зависимости не останавливались).

§ 16.4. Экситон Ванье–Мотта

Все квазичастицы, которые мы до сих пор рассматривали, нейтральны. Они не несут на себе заряда. Экситон Ванье–Мотта, как и экситон Френкеля, тоже нейтрален, но построен из заряженных частиц — электрона проводимости и дырки. Как Вы понимаете, речь пойдет о полупроводниках (см. гл. 14).

В *основном состоянии полупроводника* в верхней заполненной электронами зоне (ее называют *валентной*, так как она заполняется валентными электронами атомов, из которых состоит полупроводник) свободных мест нет — она полна. Зона проводимости при этом совсем пуста.

Элементарное возбуждение электронов полупроводника происходит следующим образом: один электрон из валентной зоны приобретает энергию, превышающую ширину запрещенной зоны, и попадает в зону проводимости.

Дальнейшая судьба электрона в зоне проводимости и дырки в валентной зоне неоднозначна. Они могут вместе с себе подобными оставаться свободными и осуществлять проводимость полупроводника. Однако электрон и дырка притягиваются друг к другу (дырка заряжена положительно). Поэтому они могут образовать водородоподобный атом — атом, состоящий из электрона и дырки. Это и есть *экситон Ванье–Мотта*.

Энергетические уровни экситона Ванье–Мотта напоминают уровни энергий в атоме водорода. Они расположены в запрещенной для свободных электронов и дырок энергетической полосе, отделяющей зону проводимости от валентной зоны. Как и уровни энергии в атоме водорода, энергетические уровни (правда, отсчитанные от дна зоны проводимости) обратно пропорциональны n^2 , где $n = 1, 2, 3, \dots$ — целые числа. Однако коэффициент пропорциональности отличается от константы Ридберга (см. (3.16)). Вместо массы электрона в выражение для аналога константы Ридберга входит эффективная масса (см. гл. 1)

$$m_{\text{эф}} = \frac{m_n m_p}{m_n + m_p},$$

где m_n и m_p — массы электрона и дырки.

Масса m_n электрона в полупроводнике как правило не равна эффективной массе дырки, m_p . Обе они вычисляются по формуле (16.3). Энергии электрона проводимости и дырки, как и энергия любой квазичастицы, являются периодическими функциями квазиимпульса. В экситон Ванье–Мотта объединяются электрон, имеющий энергию, близкую к минимуму в зоне проводимости, и дырка с энергией, близкой к максимуму в валентной зоне. Обе они масштаба массы свободного электрона, но могут и существенно отличаться (например, быть в $10 \div 100$ раз меньше). Итак, формула, описывающая энергетические

уровни экситона Ванье–Мотта, имеет вид (за нуль принято минимальное значение энергии в зоне проводимости)

$$E = -\frac{m_{\text{эф}}e^4}{2\hbar^2} \left(\frac{1}{n^2}\right), \quad n = 1, 2, 3, \dots; \quad m_{\text{эф}} = \frac{m_n m_p}{m_n + m_p}. \quad (16.19)$$

Экситон Ванье–Мотта напоминает *позитроний* — водородоподобный атом, состоящий из электрона и позитрона. Правда, в позитронии эффективная масса, входящая в выражение для значений энергетических уровней, равна $m_e/2$, так как электрон и позитрон имеют одинаковые массы.

Существует еще один атом, о котором стоит вспомнить. Это мюоний (см. гл. 6). Эффективная масса в формуле, аналогичной (16.19), заметно отличается от массы мюона, так как отличие массы протона от массы мюона меньше, чем от массы электрона.

Как и всякая квазичастица, экситон Ванье–Мотта в полупроводнике не локализован. Он может перемещаться. Состояние поступательного движения экситона Ванье–Мотта определяется квазиимпульсом и описывается, конечно, не $m_{\text{эф}}$. Эффективная масса поступательного движения экситона обязана периодической зависимости его энергии от квазиимпульса (см. формулы (16.2) и (16.3)).

Обнаруживается экситон Ванье–Мотта по спектрам поглощения света полупроводниками.

§ 16.5. Квазичастицы-бозоны и кванты макроскопических волн

Мы перечислили отнюдь не все квазичастицы, являющиеся бозонами. По сути все их и невозможно перечислить. Хотя бы потому, что неизвестно, сколько их. Некоторые квазичастицы-бозоны еще будут упоминаться в следующей главе, посвященной квазичастицам-фермионам. Обратите внимание на то, что экситонам в каком-то смысле место в главе о фермионах: и экситон Френкеля, и экситон Ванье–Мотта образованы из фермионов.

Забудем на некоторое время не только о квазичастицах, но даже об атомарном строении твердого тела. Как мы уже отмечали, многие явления, происходящие в твердых телах, можно рассматривать, считая твердое тело непрерывной средой. В частности, к таким явлениям относятся распространяющиеся по телу колебания, длина волны которых превосходит расстояние между атомами. На данном этапе можно абстрагироваться даже от природы волны. Не важно, что это за волна. Важно то, что она является *классической*, то есть ее можно описывать уравнениями классической физики, а квантовые поправки малы. Что для этого необходимо? Энергия волны должна быть велика по сравнению с величиной кванта энергии ($\hbar\omega$). Кроме того, волна должна

быть слабозатухающей. Тогда можно абстрагироваться от затухания и считать, что волна осуществляет одно из *возможных* движений тела.

Если речь идет о когерентной плоской волне, то она имеет вполне определенные значения волнового вектора \mathbf{k} и частоты ω . Теперь перейдем на язык квантовой механики. Существование классической волны означает, что в теле *много квазичастиц с квазиимпульсом* $\mathbf{p} = \hbar\mathbf{k}$ и с энергией $\varepsilon = \hbar\omega$. Таким образом, квазичастицы, соответствующие движению в виде классической волны, являются бозонами. Фермионов в каждом состоянии не может быть больше одного.

В зависимости от строения кристалла в нем могут распространяться различные волны: акустические и оптические колебания, спиновые волны, волны поляризации и многие, многие другие. Их квантовые аналоги — квазичастицы-бозоны.

Ранее мы неоднократно подчеркивали фундаментальное отличие между фермионами и бозонами. Принадлежность одному из двух классов именуется статистикой (*ферми-статистика* и *бозе-статистика*). Формально статистика проявляется при любом числе частиц и/или квазичастиц, лишь бы их было больше единицы. Подчеркнем, что в квантовой физике взаимодействуют невзаимодействующие частицы или квазичастицы. Придав утверждению нарочито парадоксальную формулировку, мы хотим обратить на него внимание читателей.

Газ частиц чувствителен к статистике, если он достаточно плотен или при фиксированной плотности находится при достаточно низкой температуре. Как в этом смысле ведут себя газы квазичастиц? Забегая несколько вперед (квазичастицам-фермионам посвящена следующая глава), запишем выражения для среднего равновесного значения количества квазичастиц при температуре T :

$$\begin{aligned} \text{для бозонов} \quad \langle n \rangle &= [\exp(\hbar\omega/kT) - 1]^{-1}; \\ \text{для фермионов} \quad \langle n \rangle &= [\exp(\hbar\omega/kT) + 1]^{-1}. \end{aligned}$$

Отличие, как видно из выписанных формул, заключается только в знаке перед единицей. Следовательно, если единицей можно пренебречь, то квазичастицы не проявляют своей статистики. Чтобы можно было пренебречь единицей, необходимо сильное неравенство: $\hbar\omega \gg kT$. Причина безразличия к статистике ясна: квазичастиц, энергия которых велика по сравнению с kT , очень мало и они не могут проявить свои статистические, коллективные свойства. Сказанное не означает, что подобная квазичастица теряет свою квантовую сущность. Дискретность энергии как суммы квантов — основной признак квантования — сохраняется, а существование неравенства не позволяет устремить постоянную Планка к нулю, без чего невозможен предельный переход к классической физике.

Последнее замечание особенно существенно для квазичастиц, введенных с помощью квантования классических волн (прекрасный пример — фононы). Допускает ли выражение $\langle n \rangle = [\exp(\hbar\omega/kT) - 1]^{-1}$ переход к классике? Да, но для этого необходимо выполнение неравенства $\hbar\omega \ll kT$. Раскладывая экспоненту, получаем $\langle n \rangle \sim kT/\omega$. Отсюда следует классическое значение для средней энергии осциллятора с частотой ω (сравните со словами, сказанными после формулы (16.4)).

Глава 17. ЭЛЕКТРОНЫ ПРОВОДИМОСТИ. КВАЗИЧАСТИЦЫ-ФЕРМИОНЫ

Все квазичастицы, которые мы рассматривали (фононы, магноны, экситон Френкеля), являются квантами *коллективных* движений в твердом теле. Даже экситон Ванье–Мотта — не исключение, поскольку движение дырки фактически есть движение *всех* электронов валентной зоны кроме одного — того, который перешел в зону проводимости.

Каждая введенная квазичастица имеет свои атрибуты. Квазичастица очень похожа на частицу. Состояние частицы характеризуется импульсом, а квазичастицы — квазиимпульсом. От значения импульса зависят энергия и скорость частицы, а от значения квазиимпульса — энергия и скорость квазичастицы. Можно сказать, что *квазичастица* — *одночастичное возбуждение тела*, то есть возбуждение, очень похожее на частицу.

Структурными единицами *массы* тела служат частицы, из которых оно состоит. Квазичастицы — структурные единицы *энергии* тела, то есть энергия тела, отсчитанная от энергии его основного состояния, равна сумме энергий квазичастиц. Это равенство выполняется с большой точностью и позволяет описывать возбужденное состояние твердого тела как состояние газа квазичастиц.

Для того чтобы сформулировать свойства *отдельной* квазичастицы, необходимо рассматривать движение *всех* атомных или субатомных частиц, из которых состоит тело.

Некоторые квазичастицы, однако, имеют атрибуты, совпадающие с атрибутами отдельных частиц. Из рассмотренных уже квазичастиц примером может служить магнон, спин которого строго равен отклонению из положения равновесия *спина отдельного атома*.

Настоящая глава посвящена *электронам проводимости*. Каждый электрон в любом теле имеет заряд, равный заряду свободного электрона. Даже дырке, оказывается, можно не приписывать положительный заряд, если учесть, что она представляет собой *незанятое электроном состояние с отрицательной эффективной массой* (см. (16.3)).

Тот факт, что металлы проводят электрический ток, известен давно. И то, что ток переносят электроны, было известно, как мне представляется, уже к концу XIX-го века. Более того, в 1916-м году было непосредственно измерено отношение e/m (заряда к массе) электронов проводимости. Это отношение — основная характеристика заряженной частицы. Оно оказалось таким же, как у свободных электронов.

Остановимся кратко на том, *как* было измерено отношение e/m . Экспериментаторы (Толмен и Стюарт) использовали инерционные эффекты. Их несколько, и все они носят название по имени обнаруживших их исследователей (эффекты Толмена–Стюарта). Один из методов в принципе столь прост, что о нем можно рассказать на языке вполне доступных формул. Если проводник движется с ускорением a , то на каждый электрон действует сила инерции ma . Если такой проводник включить в цепь, то действие этой силы возбудит в ней ток. В разомкнутом проводнике возникает электрическое поле E , компенсирующее действие силы инерции. Электрическое поле и ускорение можно измерить, а условие компенсации

$$eE = ma \quad (17.1)$$

дает возможность определить e/m .

Откуда в металле берутся свободные электроны, ведь атомы, из которых он построен, нейтральны? Атом — система связанных между собой электронов и ядер. Часть электронов в атомах металлов сравнительно слабо связана с ядром. Это *валентные* электроны. При объединении отдельных атомов в металлический образец валентные электроны покидают атомы. Последние ионизируются. Любой металлический образец представляет собой кристаллическую решетку из ионов. Между ними движутся бывшие валентные электроны. Если каждый ион имеет заряд, равный $+Ze$, то всего в единице объема металла $n = Zn_j$ электронов, где n_j — число ионов в единице объема.

В целом металл, конечно, нейтрален. Пока проводник не включен в электрическую цепь, электроны не покидают образец. Причина заключается в том, что ионы создают потенциальную яму, из которой электроны не могут выбраться, не приобретя необходимой энергии (ее называют *работой выхода*). Можно пронаблюдать высвобождение электронов из образца. Существует несколько эффектов, демонстрирующих эмиссию электронов из металла и отличающихся методом передачи им необходимой энергии: фотоэффект, термоэлектронная эмиссия, холодная эмиссия, электрон–электронная эмиссия и другие. Мы сознательно не разъясняем упомянутые термины, поскольку не собираемся рассказывать об эмиссионных явлениях. Несколько типов эмиссий перечислено только для того, чтобы убедить читателя в наличии *в металле свободных электронов*. Вылетевшие из металлов электроны, конечно, ничем не отличаются от любых других, даже прилетевших из космоса в составе космических лучей.

Если читатель осознал, что в металле есть свободные электроны, то у него может возникнуть вопрос: «При чем же здесь *квазичастицы*? Электроны-то уж заведомо частицы!». Хочется думать, что недоумение скоро пройдет.

На примере электронов проводимости постараемся показать, как происходило и происходит усложнение модели, позволяющее все более

адекватно описывать свойства макроскопических тел (в данном случае — электронных проводников).

Последовательно рассказывая о постепенно усложняющихся моделях, мы не будем придерживаться аргументации их авторов. Смена воззрений, как правило, лишает многие соображения убедительности. Особенно когда известно, как поступательное движение научной мысли уже модифицировало первоначальную теорию.

Первая подробно разработанная модель металла, позволившая описать его электро- и теплопроводность, была создана к 1900-у году. По имени авторов она получила название *модели Друде–Лоренца*.

Согласно модели Друде–Лоренца электроны проводимости представляют собой газ совершенно свободных электронов. Предполагается, что пока к проводнику не приложена извне разность потенциалов, электрическое поле в металле попросту равно нулю. В каком-то смысле это справедливо, если обращать внимание только на макроскопическое поле и пренебрегать микроскопическими, созданными отдельными заряженными частицами. Подобное предположение не слишком необычно. Например, вводя простое и естественное понятие *плотности вещества*, мы фактически пренебрегаем тем, что атомы расположены лишь в отдельных точках внутри тела. Все средние, макроскопические величины отличаются от истинных значений, поскольку атомно-молекулярная «рябь» не принимается во внимание.

К газу электронов Друде и Лоренц применили молекулярно-кинетическую теорию *классического* газа. Подчеркивая, что газ классический, мы допускаем анахронизм: в то время в сознании физиков не существовало никаких других газов.

Введение длины свободного пробега дало возможность выразить основные характеристики металлов (электропроводность σ и теплопроводность κ) через параметры электронного газа: число электронов в единице объема n , заряд электрона e , его массу m , среднюю скорость $\langle v \rangle$, длину свободного пробега l или время свободного пробега τ . В принципе, каждый используемый параметр допускает независимое измерение. Оценка длины свободного пробега показала, что $l \gg a$, где a — межатомное расстояние. Тем самым получила подтверждение гипотеза о свободе электронов проводимости и возникла уверенность в их делокализации. Мысль о делокализации электронов в металле сохранила свою ценность до настоящего времени.

Важным достижением теории Друде–Лоренца принято считать вывод *закона Видемана–Франца*.

Закон Видемана–Франца — эмпирическое соотношение, открытое в 1853-м году. Согласно ему отношение κ/σ не зависит от сорта металла. В 1882-м году Л. Лоренц дополнил утверждение, заметив, что отношение $\kappa/(\sigma T)$ равно универсальной постоянной L , получившей название *числа Лоренца* (см. ниже). В настоящее время закон

Видемана–Франца формулируется следующим образом:

$$\kappa/(\sigma T) = L, \quad L = (\pi^2/3)(k_B/e)^2 = 2,45 \cdot 10^{-8} \text{ Вт} \cdot \text{ Ом}/\text{К}^{-2}. \quad (17.2)$$

Вывод формулы (17.2), правда, с несколько отличающимся значением числа Лоренца, следовал из выражений для σ и κ , а также из справедливого утверждения о том, что тепло в металле в основном переносят электроны.

Как уже говорилось (см. гл. 10), свойства газа электронов проводимости практически при любой температуре надо описывать формулами *квантовой статистики*. Эту же мысль можно выразить иначе: *электронный газ в металле вырожден*.

Привлечение квантовой статистики к теоретическому исследованию свойств металлов — заслуга Зоммерфельда. С 1928-го года электронная теория металлов, исходящая из предположения о свободе электронов проводимости, называется теорией Друде–Лоренца–Зоммерфельда.

Переход в формулах от классического (невырожденного) газа к квантовому (вырожденному), естественно, изменил температурные зависимости многих физических величин. Как правило, выведенные формулы по меньшей мере качественно, а часто и количественно соответствуют экспериментальным данным. В частности, было найдено правильное значение числа Лоренца.

Особенно показательно разрешение *парадокса теплоемкости*. Под этим названием фигурирует вопрос, почему металлы подчиняются закону Дюлонга и Пти. Если в них есть свободные электроны, то законы классической статистики требуют, чтобы средняя энергия каждого электрона равнялась $(3/2)k_B T$. Отсюда, казалось бы, обязательно, следует, что теплоемкость металла должна на величину $(3/2)k_B n$ превосходить теплоемкость изолятора, в котором свободных электронов нет. Однако последнего не наблюдается.

Вырождение электронного газа ликвидирует парадокс.

При $T \ll T_{\text{кв}} \approx \hbar^2 n^{2/3} / (mk_B)$ лишь небольшая доля электронов проводимости, порядка $n_{\text{эф}} = n(T/T_{\text{кв}})$, принимает участие в тепловом движении. Остальные как бы «заморожены» в глубине фермиевского распределения (см. рис. 10.1). Поскольку для электронного газа большинства металлов выписанное неравенство всегда справедливо, электронная теплоемкость в широком диапазоне температур линейно зависит от T и мала по сравнению с классическим значением теплоемкости атомов (или ионов) кристаллической решетки. Правда, при низкой температуре ($T \ll \theta_D$, где θ_D — температура Дебая; см. гл. 16) решеточная теплоемкость стремится к нулю пропорционально T^3 , то есть быстрее, чем электронная. Это дает возможность выделить вклад электронов в теплоемкость металла. Калориметрические исследования при низких температурах — надежный метод изучения электронов проводимости.

Надо признать, что теория Друде–Лоренца–Зоммерфельда необычайно продуктивна. С ее помощью удалось объяснить многие свойства металлов. Однако у нее есть серьезный недостаток: для теории Друде–Лоренца–Зоммерфельда все металлы на одно лицо. Это как бы теория металла вообще. Вместе с тем, металлы весьма существенно различаются. Описать подобное различие теория Друде–Лоренца–Зоммерфельда не может. Кроме того, все металлы — кристаллы. Каждый обладает анизотропией. Анизотропия остается за пределами теории Друде–Лоренца–Зоммерфельда: газ свободных электронов изотропен.

Имеется и еще одна немаловажная проблема. Одни вещества являются металлами, то есть проводниками, другие — диэлектриками (изоляторами). Можно ли указать критерий металла?

Кроме того, непонятно, что собой представляют вещества, свойства которых занимают промежуточное положение между свойствами проводников и изоляторов, то есть полупроводники и полуметаллы? Ранее об этом сказано явно недостаточно.

Прежде чем описать более сложную, но значительно более адекватную модель металла, покажем, что модель Друде–Лоренца–Зоммерфельда позволяет выявить некоторые важные черты элементарных возбуждений электронной подсистемы металла, черты, не стирающиеся при усложнении модели.

Попробуем представить себе основное состояние электронов. Как мы уже знаем, при $T = 0$ их движение не прекращается. Все состояния с энергией, меньшей энергии Ферми (обозначим ее ε_F), заняты, а с энергией, большей ε_F , свободны (см. гл. 10). Это нетрудно изобразить графически, если нарисовать импульсное пространство: при $T = 0$ сфера радиуса $p_F = (2m\varepsilon_F)^{1/2}$ однородно заполнена электронами, а вне ее электронов вообще нет (рис. 17.1). (Всем величинам, имеющим в своих обозначениях индекс «F», присвоено имя Ферми: ε_F — энергия Ферми; p_F — фермиевский импульс; сфера радиуса p_F — сфера Ферми.) Каково элементарное возбуждение газа электронов?

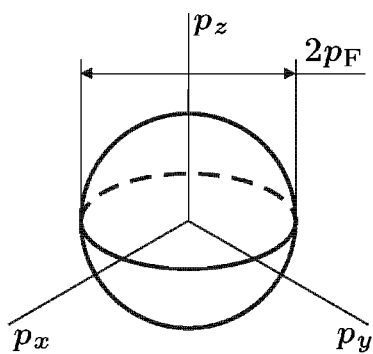


Рис. 17.1 Сфера Ферми

Для того чтобы энергия электронов увеличилась, необходимо, чтобы какие-то электроны приобрели импульс, превышающий p_F . Элементарное возбуждение состоит в том, что *один* электрон получает импульс $p > p_F$ и освобождается *одно* состояние с импульсом $p < p_F$. В таком случае говорят, что появилась дырка.

Важной чертой элементарных возбуждений — фермионов — является то, что они рождаются парами. Одновременно появляются

частица ¹⁾ с $p > p_F$ и дырка с $p < p_F$. Вычислим, как изменится энергия электронного газа E за счет рождения подобной пары:

$$\begin{aligned} E &= E_0 + p_1^2/2m - p_2^2/2m = \\ &= E_0 + p_1^2/2m - p_F^2/2m + p_F^2/2m - p_2^2/2m; \quad p_1 > p_F, \quad p_2 < p_F. \end{aligned} \quad (17.3)$$

Здесь E_0 — энергия основного состояния. Энергию элементарных возбуждений (частицы и дырки) принято отсчитывать от энергии Ферми. Как правило, нас интересуют элементарные возбуждения с энергией, значительно меньшей ε_F . Тогда, обозначив энергию частицы ε_e , а энергию дырки ε_h , имеем

$$\varepsilon_e = v_F(p - p_F); \quad \varepsilon_h = v_F(p_F - p). \quad (17.4)$$

Здесь $v_F = p_F/m$ — фермиевская скорость, то есть скорость электрона на поверхности Ферми. В первой формуле $p > p_F$, во второй $p < p_F$. Фермиевская скорость достаточно велика: $v_F \sim 10^8$ см/с. Например, скорость звука в металлах приблизительно раз в сто меньше.

Формулы (17.4) — лишь результат разложения. Однако сделав это преобразование, частице (электрону) и дырке можно, по-видимому, присвоить наименование *квазичастиц*. Во-первых, как бы изменилась зависимость энергии от импульса, а во-вторых, число частиц и дырок не остается неизменным (!). Чем больше энергия газа электронов, чем выше температура, тем больше квазичастиц (частиц и дырок). Правда, полное число частиц равно числу дырок.

Как мы увидим в дальнейшем, для того чтобы считать электроны и дырки квазичастицами, есть и более убедительные основания, чем те, которые мы пока привели. Итак, *квазичастицы-фермионы это электроны (частицы) и дырки*.

Современная теория металлов основана на *зонной теории*. В создании последней принимали участие многие физики-теоретики. Мы не можем рассказать о вкладе каждого из них. Наверное, не будет ошибкой сказать, что основоположником зонной теории следует считать Феликса Блоха (он же первый понял, что представляет собой спиновая волна; см. гл. 16).

Попытаемся сформулировать основные положения зонной теории.

Электроны в твердом теле движутся в поле сил, действующих на них со стороны ионов кристаллической решетки. От того, каковы конкретно эти силы, сначала следует абстрагироваться. Важны соотношения симметрии и то, что электрон движется в *периодическом* поле сил. Блох выяснил структуру ψ -функции электрона. Она оказалась

¹⁾ Частицами мы здесь считаем только те из электронов, которые «находятся» вне сферы Ферми. Дырки, естественно, «находятся» внутри нее.

модулированной плоской волной (зависимость от времени мы опустили):

$$\psi_{\mathbf{k}}^s(\mathbf{r}) = \exp(i\mathbf{k}\mathbf{r})u_{\mathbf{k}}^s(\mathbf{r}), \quad (17.5)$$

где $u_{\mathbf{k}}(\mathbf{r})$ — периодическая функция, период которой совпадает с периодом кристаллической решетки; \mathbf{k} — квазиволновой вектор ($\mathbf{p} = \hbar\mathbf{k}$ — квазиимпульс); все характеризующие состояние электрона и зависящие от квазиимпульса величины (и ψ -функция, и энергия) — периодические функции \mathbf{p} ; $s = 1, 2, 3, \dots$ — номер зоны. Дело в том, что фиксация значения квазиимпульса не определяет однозначно состояние. Состояний при заданном \mathbf{p} бесконечно много. Если зафиксировать номер состояния s и перебрать все значения \mathbf{p} внутри ячейки \mathbf{p} -пространства, то значения энергии различных состояний заполнят полосу конечной ширины — зону. Именно поэтому число s , нумерующее состояние, названо номером зоны.

Согласно принципу Паули в каждом состоянии с определенными значениями \mathbf{p} и s может находиться не более двух электронов, причем при наличии двух электронов они должны обладать антипараллельными спинами. Следующее утверждение особенно важно для дальнейшего: размер ячейки \mathbf{p} -пространства таков, что в ней может поместиться $2N$ электронов, где N — число ячеек кристалла¹⁾. Таким образом, $2N$ электронов полностью заполняют энергетическую зону.

Надеюсь, Вы понимаете, к какой мысли я пытаюсь Вас подвести. Для металла-проводника обязательно наличие частично заполненной электронами зоны или нескольких зон. Почему, будет рассказано в конце этой главы.

Начнем заполнять зоны электронами. Пусть на каждую ячейку кристалла приходится Z электронов, то есть суммарный заряд расположенных в ней атомных ядер равен $+Ze$. Значит, в кристалле ZN электронов. Если Z — четное число, то электроны могут заполнить $Z/2$ зон, а кристалл окажется диэлектриком-изолятором. Если Z — нечетное, одна зона останется полупустой.

Неужели справедлив столь простой критерий: нечетное Z — металл, четное — диэлектрик? Не верится! Проверим себя на первых четырех элементах таблицы Менделеева: водороде, гелии, литии и бериллии. Кристалл водорода состоит из молекул H_2 , в каждой из которых содержится два электрона ($Z = 2$). Следовательно, он должен оказаться диэлектриком. Так оно и есть! Гелий, как мы уже упоминали, кристаллизуется только под давлением. Сейчас это для нас несущественно. В каждой ячейке кристалла гелия один атом, то есть два электрона. Снова диэлектрик! А вот литий, у которого на ячейку

¹⁾ Каждая ячейка \mathbf{p} -пространства содержит все состояния с данным номером зоны. Квазиимпульсы, отличающиеся на период \mathbf{p} -пространства, описывают тождественные состояния.

приходится три электрона, является металлом. «Подводит» бериллий. У него $Z = 4$, а он — металл.

Причина проста: энергетические зоны перекрываются (рис. 17.2). Из-за этого металлов среди элементов значительно больше половины, как было бы, если бы заполнение зон определялось четностью Z .

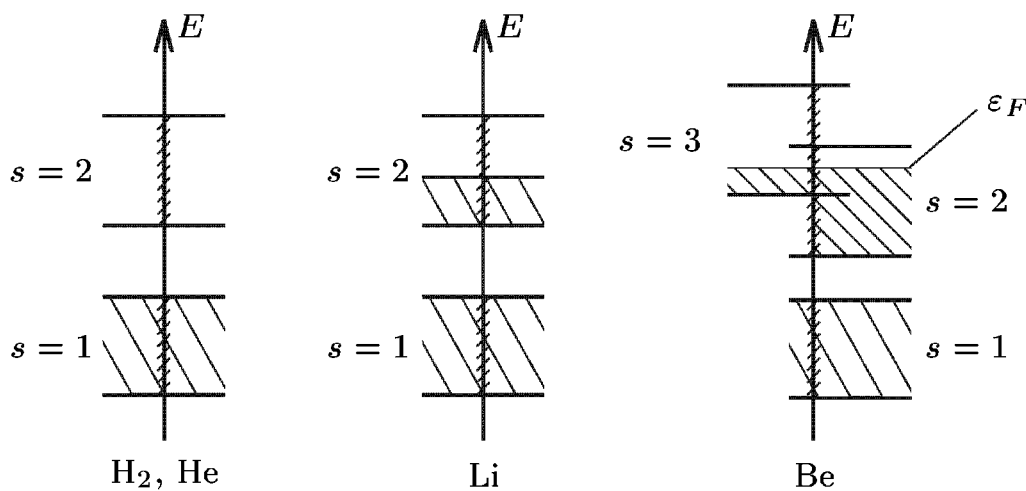


Рис. 17.2. Зонная структура первых четырех элементов таблицы Менделеева. Разрешенные состояния показаны волновой линией, а заполненные — косой штриховкой. У Be вторая и третья зоны перекрываются

В схеме, связывающей электрические свойства тела с заполнением зон, находят себе место и полупроводники, и полуметаллы. Полупроводники — диэлектрики с узкой запрещенной энергетической зоной. Ее ширина такова, что при невысокой температуре значительное количество электронов из заполненной при $T = 0$ энергетической зоны попадает в пустую при $T = 0$ зону. Полуметаллы — металлы, но только потому, что зоны слегка перекрываются: в частично заполненных зонах очень мало электронов или свободных мест.

Хотя причина принципиального различия твердых тел теперь ясна, определить, какое из них является металлом, а какое диэлектриком, можно, только производя расчет зонной структуры. Лишь когда в результате подобного расчета выяснится, что имеет место один из перечисленных случаев, можно будет отнести твердое тело к одному из указанных типов. Надо сказать, что в последнее десятилетие расчет зонной структуры твердых тел удается производить с большой точностью.

Ниже мы сосредоточим свое внимание на металлах.

В модели Друде–Лоренца–Зоммерфельда основное состояние электронов проводимости — заполненная ими ферми-сфера. В зонной теории поверхности равной энергии — сложные периодические функции квазиимпульса. Поверхность Ферми, отделяющая занятые электронами состояния от свободных, как правило, достаточно вычурна. Иногда удобно изображать только ее кусок, помещающийся в одной ячейке \mathbf{p} -пространства, а иногда, особенно если она непрерывно проходит через

границы ячеек, ее изображают так, чтобы была ясна топология всей ферми-поверхности.

Оказалось, что многие свойства металлов чувствительны к форме ферми-поверхности. Возникла своеобразная спектроскопия, позволяющая восстановить форму поверхности Ферми по опытным данным (преимущественно в магнитном поле). В настоящее время известны поверхности Ферми практически всех одноатомных металлов, а также многих металлических соединений. В качестве примера на рис. 17.3 показана поверхность Ферми свинца.

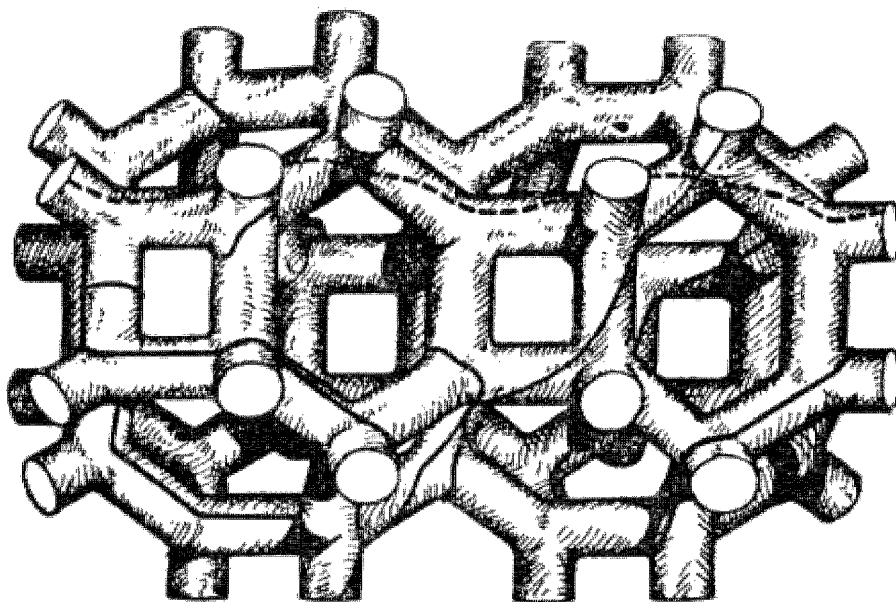


Рис. 17.3 Поверхность Ферми свинца

Элементарные возбуждения электронной системы металла рождаются, как мы уже говорили, парами: электрон (частица) в свободной части \mathbf{p} -пространства и свободное место (дырка) среди занятых при $T = 0$ состояний. Если, как и раньше, рассматривать квазичастицы, энергия которых близка к энергии Ферми, то энергию квазичастиц, отсчитанную от ε_F , можно разложить по степеням отклонения от поверхности Ферми (по $\mathbf{p} - \mathbf{p}_F$):

$$\begin{aligned} \text{для электрона} \quad \varepsilon_e &= v_F(\mathbf{p} - \mathbf{p}_F); \\ \text{для дырки} \quad \varepsilon_{\text{дыр}} &= v_F(\mathbf{p}_F - \mathbf{p}). \end{aligned} \quad (17.6)$$

Обратите внимание, что анизотропия поверхности Ферми приводит к анизотропии энергии электрона и дырки: фермиевская скорость v_F зависит от того, в какой точке \mathbf{p}_F на поверхности Ферми находится электрон.

Как видно из формул (17.6), для того чтобы найти энергетический спектр электронов и дырок, кроме формы поверхности Ферми, надо знать скорости фермиевских электронов. Поскольку $v_F = d\varepsilon/dp$ при $\mathbf{p} = \mathbf{p}_F$, из геометрических соображений ясно, что вектор скорости v_F является вектором нормали к поверхности Ферми.

Если для изображения основного состояния электронов проводимости достаточно знать форму ферми-поверхности, то для нахождения энергии квазичастиц, то есть электронов и дырок, необходимо нанести на поверхность Ферми векторы скорости. Давайте представим себе, что из этого получится.

Сначала ограничимся моделью Друде–Лоренца–Зоммерфельда. В ее случае поверхность Ферми с векторами скорости на ней напоминает свернувшегося ежика.

Когда в 50-х годах прошлого века выяснилось, сколь сложны поверхности Ферми многих металлов, удивление исследователей выразилось в том, что некоторые из них называли *монстрами*. Покрыв их векторами скорости, мы получим *ощетинившегося монстра*. Особенно впечатляюще должен выглядеть воображаемый рисунок, если свободная от электронов часть p -пространства расположена внутри поверхности Ферми. Чтобы представить себе чудовище со щетиной, направленной внутрь, надо обладать развитым воображением (например, таким, как у польского писателя-фантаста Станислава Лема).

При всех геометрических сложностях, схематически описанных нами, некоторые физические свойства зонных электронов очень напоминают свойства электронов в пустом пространстве. Зонные электроны — электроны в поле сил, обусловленных ионами кристаллической решетки (иногда мы будем именовать их, как и раньше, электронами проводимости). Сходство между зонными и свободными электронами — следствие *периодичности* поля сил. Именно периодичность сил позволяет ввести *квазиимпульс* (см. выше). Сходство особенно отчетливо проявляется тогда, когда, описывая движение зонного электрона по проводнику, можно пользоваться классическими представлениями.

Вся рассмотренная ситуация — пример успешного использования абстракции на разных этапах. Вывод зависимости энергии зонного электрона от квазиимпульса — квантовомеханическая задача. Анализ движения под действием внешних сил показывает, что в некоторых весьма распространенных случаях можно пренебречь отличием квазиимпульса от импульса, а движение описывать классически. При этом от квантовомеханического подхода сохраняется сложная зависимость энергии от импульса. О тех случаях, когда при описании движения зонной частицы достаточно использовать законы классической механики, и пойдет речь ниже.

Классические уравнения движения зонного электрона очень напоминают уравнения Ньютона:

$$d\mathbf{p}/dt = \mathbf{F}; \quad d\mathbf{r}/dt = \mathbf{v}; \quad \mathbf{v} = d\boldsymbol{\varepsilon}/d\mathbf{p}. \quad (17.7)$$

Хотя по виду они не отличимы от более привычных уравнений ньютоновской механики, отличия, конечно, все же есть. Основных — два. Во-первых, действующая на зонный электрон сила (мы обозначили ее, как обычно, буквой \mathbf{F}) есть не полная сила, а только внешняя по

отношению к кристаллу, например сила, обусловленная приложенной к проводнику разностью потенциалов. Во-вторых, из-за сложной зависимости энергии от импульса (в данном контексте приставку «квази» можно опустить) скорость электрона $v \neq p/m$; она является сложной функцией p .

Надо сказать, что последствия указанных отличий очень впечатляют. Приведем два примера.

1. *Электрон проводимости в постоянном и однородном магнитном поле \mathbf{H} .*

На такой электрон действует сила Лоренца ($\mathbf{F} = (e/c)[\mathbf{v}\mathbf{H}]$).

Рассматривать движение электрона под воздействием магнитного поля с помощью уравнений классической механики можно в случае, когда поле не слишком велико (условие опускаем). Траектория каждого электрона в вакууме — навитая на магнитные силовые линии спираль. В кристалле некоторые из зонных электронов совершают инфинитное движение во всех трех измерениях. Кроме того, хотя у электронов проводимости, как и у всех электронов, заряд отрицательный, некоторые из них (из тех, которые движутся по спирали) в плоскости, перпендикулярной магнитному полю, вращаются так, будто имеют положительный заряд.

2. *Электрон проводимости под действием постоянного и однородного электрического поля \mathbf{E} ($\mathbf{F} = e\mathbf{E}$).*

Электрическое поле может либо ускорять, либо замедлять (тормозить) электрон. В свободном пространстве ускорение соответствует увеличению энергии, а замедление — уменьшению. С электронами проводимости дело обстоит не так просто: среди них встречаются такие, которые, замедляясь, увеличивают свою энергию, а ускоряясь, уменьшают.

Все особенности динамики зонных электронов объясняются сложной зависимостью энергии от квазиимпульса и могут быть связаны с геометрией поверхностей равной энергии. При изучении движения электронов под действием внешних полей удобно использовать зависимость

$$\varepsilon = \varepsilon(\mathbf{p}), \quad (17.8)$$

а не разложение (17.6).

Нетривиальность динамики зонных электронов способствует тому, чтобы именовать их квазичастицами, а не частицами. Особенно следует подчеркнуть тот факт, что зависимость (17.8) является результатом действия периодического поля сил, источником которых служат не только ионы, но и все электроны (конечно, кроме того, движение которого мы рассматриваем). До сих пор о роли коллектива электронов в движении отдельного электрона мы умалчивали (абстрагировались от коллектива).

Строго учесть самосогласованное движение электронов позволила теория ферми-жидкости Ландау. О ней мы расскажем в гл. 19.

Пока электроны проводимости для нас представляют собой ферми-газ. Правда, несколько более сложный, чем тот, о котором мы рассказывали в гл. 10. Последнее — следствие сложности зависимости энергии от квазиимпульса. Однако многие свойства газа электронов проводимости очень напоминают свойства обычного ферми-газа¹⁾. Например, их теплоемкость при низких температурах линейно зависит от T ; парамагнитная восприимчивость, обязанная своим существованием электронам металла, не зависит от температуры. Обычный ферми-газ имеет те же температурные зависимости. Правда, численные значения характеристик несколько отличаются, хотя, как правило, усложнение зависимости (17.8) по сравнению с обычной ($\varepsilon = p^2/2m$) не слишком изменяет порядок величины (естественно, плотности обоих газов предполагаются одинаковыми).

Вместе с тем, переход от теории Друде–Лоренца–Зоммерфельда к зонной теории — весьма существенный шаг вперед. Прежде всего, потому, что выяснилось: электрон в периодическом поле движется как свободная частица.

Наверное, именно сейчас надо объяснить, почему кристалл, у которого есть частично заполненная электронами зона, является металлом, то есть проводит электрический ток. Металл — проводник.

Если электродвижущая сила отсутствует, или, проще, в металлическом образце нет разности потенциалов и он не включен в цепь, то ток по металлу не течет, хотя электроны, как Вы понимаете, движутся. Однако движутся они хаотически, направленное движение отсутствует. Правда, в этом хаосе присутствует своеобразный порядок, особенно ощутимый при $T = 0$: все электроны распределены по энергетическим уровням, не превышающим энергию Ферми.

Если на электрон проводимости действует сила $\mathbf{F} = e\mathbf{E}$ (результат наличия разности потенциалов), то эта сила сможет изменить его состояние только в том случае, если у электрона есть возможность переместиться в \mathbf{p} -пространстве, не нарушая принципа Паули. Поскольку за бесконечно малый интервал времени dt квазиимпульс тоже изменяется на бесконечно малую величину $d\mathbf{p}$, подобное перемещение в \mathbf{p} -пространстве (по крайней мере, при $T = 0$) доступно лишь электронам, имеющим энергию, равную ε_F . Вспомнив, что газ электронов проводимости в металле всегда сильно вырожден ($k_B T \ll \varepsilon_F$), понимаем: только тогда, когда в теле есть частично заполненная зона, оно может проводить ток при низких температурах (быть металлом).

¹⁾ Задумаемся над употреблением эпитета «обычный». На самом деле обычных ферми-газов в природе попросту нет, а любой кусок металла — хранитель необычного газа электронов проводимости.

Надо признаться, что мы проявили некоторую поспешность. Если электрон может свободно двигаться под действием электрического поля, то, казалось бы, он должен напоминать электрон в свободном пространстве между катодом и анодом. Приобретая энергию от электрического поля в свободном пространстве, электрон отдает ее при столкновении с анодом. В результате, если ток электронов велик, то анод может просто расплавиться. Не очень похоже на происходящее в проводнике. Когда по проводнику течет ток, весь проводник нагревается более или менее равномерно по всей длине. Дело в том, что объяснив в предыдущем абзаце *проводимость*, то есть свободу перемещения под действием электрического поля, мы проигнорировали *сопротивление* — способность электронов отдавать приобретенную от электрического поля энергию.

Свобода электрона ограничена столкновениями. Он не сталкивается с правильно расположенными в пространстве ионами и в периодическом поле сил движется как блоховская волна (см. (17.5)). Однако каждое нарушение строгой периодичности электрон воспринимает как препятствие — сталкивается с ним. Мерой свободы служит длина свободного пробега l — среднее расстояние, пролетаемое электроном без столкновений. Как мы уже отмечали, $l \gg a$, где a — размер ячейки кристалла. Именно последнее неравенство — мера несвободы (или свободы) электрона. В следующей главе мы расскажем о длине свободного пробега несколько подробнее, а здесь приведем выражение для удельной электропроводности σ металла, кристаллическая решетка которого имеет кубическую симметрию (вывод мы опускаем):

$$\sigma = S_F e^2 l / 12 \pi^3 \hbar^3, \quad (17.9)$$

где S_F — площадь поверхности Ферми. Подчеркнем: сколь бы ни была вычурна и анизотропна поверхность Ферми, у металла с кубической симметрией она также будет обладать кубической симметрией. Важнее другое. По требованию симметрии (см. гл. 8) тензор удельных электропроводностей вырождается в скаляр — кубический кристалл не может обладать анизотропией электропроводности.

Пропорциональность проводимости не объему поверхности Ферми, а ее площади — следствие принципа Паули: только электроны, расположенные на поверхности Ферми в непосредственной близости от свободной части \mathbf{p} -пространства, могут изменять свой квазиимпульс под воздействием постоянной силы (см. выше).

Глава 18. КВАЗИЧАСТИЦЫ НЕ ВПОЛНЕ СВОБОДНЫ

Любое твердое тело состоит из огромного числа частиц. Можно считать, что из бесконечного. Введение квазичастиц начинается с попытки выяснить, нельзя ли степени свободы, описывающие движение частиц твердого тела, разделить на группы (их называют подсистемами), которые слабо связаны между собой. Если вы вернетесь к квазичастицам, рассмотренным в предыдущих главах, то убедитесь, что это действительно так. Движение спинов считается независимым от колебаний атомов, а рассматривая движение электронов, мы попросту «запрещаем» ионам двигаться. Хотя это не обсуждалось, каждый раз были основания для абстрагирования от взаимодействия (связи) между подсистемами.

Однако и это не все.

Исследуя движение отдельной квазичастицы, мы абстрагировались от всего, что мешает ей «чувствовать себя свободной». Предполагалось, что каждая из рассмотренных квазичастиц движется так, будто она *полностью* свободна, ни с чем не взаимодействует. Только в конце предыдущей главы мы отступили от этого правила. Настоящая глава целиком посвящена причинам, мешающим квазичастицам быть полностью свободными. Причинам, которые определяют *длину свободного пробега* квазичастиц.

В такой общей формулировке тема слишком широка. Она охватывает целые главы физической кинетики — науки о неравновесных свойствах физических систем. Нам придется ограничиться лишь несколькими примерами: мы обсудим некоторые процессы, определяющие природу длины свободного пробега фононов и электронов.

Начнем с *фононов*.

Рассмотрение фононов как невзаимодействующих квазичастиц основано на *гармоническом приближении*. В гармоническом приближении энергия тела представляет собой квадратичную функцию отклонений атомов от положений равновесия. То, что это является *приближением*, очевидно. Воспринимается оно как разумное приближение: отклонения малы и пренебречь более высокими членами разложения, казалось бы, можно.

Необходимо, однако, подчеркнуть, что из вполне законного приближения следует важный *качественный* вывод — фононы не взаимодействуют друг с другом. Справедлив ли он? Если не только не учитывать

взаимодействие между фононами, но и не обращать внимание на нарушения строгой периодичности кристаллической решетки, то длина свободного пробега фононов окажется равной бесконечности. Вот это уж действительно абстракция!

Конечно, длина свободного пробега фонона (как и любой квазичастицы) ограничена размером кристалла. Но к тому, что кристалл можно считать имеющим бесконечные размеры, думаю, Вы уже привыкли, и подобную идеализацию не воспринимаете как серьезную абстракцию. Однако возникает справедливое впечатление, что даже в строго периодическом кристалле должны существовать внутренние причины, ограничивающие длину свободного пробега фононов.

Такие причины действительно существуют. Это *ангармонизмы* — отброшенные, более высокие, чем квадратичные, члены разложения энергии по степеням отклонений. Как правило, для расчета длины свободного пробега фононов достаточно ограничиться членами третьего и четвертого порядка по отклонениям. Физическая кинетика разработала строгую и удобную процедуру вычисления не только длины пробега, но и характеристик (кинетических коэффициентов), определяющих кинетические свойства твердых тел (например, коэффициента теплопроводности или коэффициента поглощения звуковой энергии). К сожалению, мы не сможем остановиться ни на вычислении кинетических коэффициентов, ни на описании самих макроскопических кинетических процессов, не будем обсуждать ни перенос тепла по твердому телу, ни поглощение звука.

Постараемся описать (причем довольно схематично) взаимодействие фононов друг с другом. За это взаимодействие ответственны ангармонизмы.

Ангармонизмы малы по сравнению с гармоническими (квадратичными) слагаемыми. Это позволяет входящие в ангармонизмы вектора смещений разложить по функциям, которые описывают невозмущенное ангармонизмами движение. Фигурально можно сказать: «разложить по фононам». В кубический ангармонизм будет входить три фонона, а в пропорциональные четвертой степени отклонений слагаемые — четыре. Записанные таким образом ангармоничные слагаемые позволяют вычислить вероятность разных событий, происходящих с фононами:

— кубические члены описывают распад одного фонона на два и столкновение двух фононов, приводящее к рождению нового фонона (столкнувшиеся фононы при этом гибнут);

— члены четвертого порядка описывают столкновение двух фононов с сохранением их числа, хотя, возможно, сорт фононов изменяется.

При непосредственном вычислении вероятности определенного процесса (столкновения или распада) можно убедиться в *выполнении законов сохранения энергии и квазиимпульса*.

Закон сохранения энергии квазичастиц вполне обычен: их энергия до столкновения должна равняться энергии квазичастиц после столкновения. С законом сохранения квазиимпульса дело обстоит сложнее. Периодичность p -пространства позволяет изменение квазиимпульса. Он может измениться на величину периода. Столкновения квазичастиц, при которых квазиимпульс сохраняется, называются *нормальными*, а столкновения, сопровождающиеся несохранением квазиимпульса, — *процессами переброса*.

Процессы переброса ввел в рассмотрение Рудольф Пайерлс. Хотя вероятность подобных процессов обычно меньше вероятности нормальных, их роль велика. Постараемся объяснить, почему.

Для объяснения придется несколько отступить от темы. Рассмотрим течение газа по трубе. Его частицы, конечно, сталкиваются друг с другом. Законы сохранения приводят к тому, что ни импульс, ни энергия частиц не теряются, а лишь перераспределяются между ними. Что же тормозит газ? Оказывается, только столкновения его частиц со стенками трубы.

Вернемся к фононам. Если бы не было процессов переброса, то не было бы *внутренних* причин торможения. А ведь мы хотим найти именно их. Не учитывая процессы переброса, нельзя вычислить такую значительную характеристику кристалла, как коэффициент теплопроводности. Сравнивая результаты измерения коэффициента теплопроводности совершенных кристаллов больших размеров с теорией, по его температурной зависимости можно убедиться: процессы переброса играют важную роль.

Однако и нормальные процессы игнорировать нельзя: именно они ответственны за установление равновесия в газе фононов.

Наверное, Вы заметили, что в рассказе о столкновениях фононов мы не упоминали о дефектах кристаллической решетки. Конечно, фононы «чувствуют» отклонения от строгой периодичности. Однако они не слишком «чувствительны». Чем больше длина волны фонона, тем меньше он отклоняется случайным препятствием. Находясь на берегу моря или озера, приглядитесь, как волна проходит небольшое препятствие, почти не замечая.

В заключение хочется рассмотреть конкретный процесс распада одного фонона на два. Для простоты воспользуемся приближением Дебая и рассмотрим распад продольного фонона на два поперечных. Рассмотрение должно начинаться с анализа законов сохранения, чем мы и ограничимся.

Если пренебречь процессами переброса, то законы сохранения энергии и квазиимпульса будут выглядеть следующим образом (см. (17.6)):

$$\varepsilon_l(\mathbf{p}) = \varepsilon_t(\mathbf{p}_1) + \varepsilon_t(\mathbf{p}_2); \quad \mathbf{p} = \mathbf{p}_1 + \mathbf{p}_2,$$

или

$$s_l[(\mathbf{p}_1 + \mathbf{p}_2)^2]^{1/2} = s_t p_1 + s_t p_2.$$

Возведем последнее равенство в квадрат. Пусть θ — угол между векторами \mathbf{p}_1 и \mathbf{p}_2 . Тогда

$$\sin^2(\theta/2) = \frac{(1 - s_t^2/s_l^2)(p_1 + p_2)^2}{4p_1p_2}.$$

Эта формула, необходимая для вычисления вероятности распада, показывает, что он возможен, но только в случае, когда более быстрый фотон распадается на два более медленных.

Анализ законов сохранения необходим при отборе процессов, которые могут произойти.

Движение электронов проводимости металлов очень чувствительно к тому, каково состояние кристаллической решетки. Настолько, что длина свободного пробега электронов служит критерием качества кристалла. Для того чтобы оценить качество кристалла, сравнивают его удельное сопротивление при комнатной температуре ($T = 300$ К) с удельным сопротивлением при температуре жидкого гелия (порядка нескольких градусов Кельвина). Чем их отношение больше, тем кристалл чище. Удастся получить образцы металлов, у которых это отношение достигает сотен тысяч.

Сказанное требует пояснений.

Посмотрите на формулу (17.9). Для каждого определенного металла коэффициент, стоящий перед длиной свободного пробега l , содержит площадь поверхности Ферми (S_F), мировые постоянные и численные множители. Следовательно, выражению для удельного сопротивления ($\rho = 1/\sigma$) можно придать вид $A \cdot (1/l)$. Множитель A не меняется при изменении температуры и качества образца. Он представляет собой характеристику металла (меди, золота, свинца и т. д.), а свойства конкретного образца зависят от величины $1/l$.

Физический смысл величины $1/l$ прост. Это средняя вероятность рассеяния электрона на единице длины. Отсюда непосредственно следует: если в металле есть несколько *независимых* механизмов рассеяния, то

$$\frac{1}{l} = \sum_a \frac{1}{l_a}, \quad (18.1)$$

где индекс a , по которому происходит суммирование, обозначает механизм рассеяния.

Механизмы рассеяния можно разделить на два класса:

1) механизмы, присущие идеальному, бездефектному кристаллу. Вероятность каждого такого рассеяния зависит от температуры, а при $T = 0$ обращается в нуль. Обозначим суммарную длину пробега, определяемую этими механизмами рассеяния, через $l_{ид}$, а удельное сопротивление — через $\rho_{ид}$;

2) механизмы рассеяния на всевозможных дефектах кристаллической решетки. Обозначим соответствующую характерную длину свободного пробега через $l_{ос}$, а сопротивление $\rho_{ос}$ назовем *остаточным*,

так как оно *остаётся* при $T = 0$. Остаточное сопротивление от температуры зависит очень слабо.

Таким образом, сопротивление металла является суммой идеального и остаточного сопротивлений. При высоких температурах (начиная с комнатной $T_{\text{ком}}$) идеальное сопротивление как правило заметно превышает остаточное, а при низких (в несколько градусов Кельвина) остаточное сопротивление становится основной составляющей. Именно поэтому отношение $\rho(T_{\text{ком}})/\rho_{\text{ос}}$ может служить мерой качества кристалла.

Что определяет длину пробега электронов в идеальном, бездефектном кристалле? Электроны в нем сталкиваются с фононами, друг с другом, с иными квазичастицами (правда, для этого в металле должны присутствовать какие-либо другие квазичастицы, кроме фононов и электронов; например, если металл — магнетик, то электроны могут сталкиваться с магнонами).

Столкновения с фононами обычно являются главным, определяющим механизмом температурно-зависящей части сопротивления. Поэтому мы расскажем только о них. Предварительно напомним, что от температуры в выражении для удельного сопротивления зависит только вероятность столкновения.

Пометим индексом «ф» вероятность столкновения с фононами и приведем без вывода результат расчета:

$$\begin{aligned} 1/l_{\text{ид}}^{\text{ф}} &\approx (1/l_{\text{D}})(T/T_{\text{D}}) && \text{при } T \gg T_{\text{D}}; \\ 1/l_{\text{ид}}^{\text{ф}} &\approx (1/l_{\text{D}})(T/T_{\text{D}})^5 && \text{при } T \ll T_{\text{D}}. \end{aligned} \quad (18.2)$$

Здесь T_{D} — температура Дебая, а l_{D} — длина свободного пробега электронов при $T = T_{\text{D}}$. По порядку величины $l_{\text{D}} \approx 10^{-6}$ см. Описывая фононы, мы отмечали, что их полное число в теле при высоких температурах пропорционально T , а при низких — T^3 . Первая из формул (18.2) просто означает, что при высоких температурах вероятность рассеяния на фононах пропорциональна их числу. Почему же вероятность рассеяния при низких температурах не пропорциональна T^3 ? Чтобы ответить на этот вопрос, рассмотрим, что собой представляют «столкновения» электронов с фононами. (Скоро будет понятно, почему слово «столкновение» взято в кавычки.)

Пример учета ангармонизмов при описании взаимодействия фононов друг с другом показывает, что для описания взаимодействия электронов с фононами надо принять во внимание не учтенные ранее члены, содержащие и электронные, и фононные переменные. Вернемся к зонной теории и исследуем движение электрона в кристалле с не закрепленными в положениях равновесия ионами. Координату каждого иона представим в виде суммы координаты равновесного положения и вектора смещения. Поскольку вектор смещения мал, воспользуемся разложением. Нулевой (по смещениям) член разложения составляет

основу зонной теории, а первый — линейный по смещениям — именно тот, который мы ищем. Повторим: он *линеен* по смещениям. Согласно квантовой теории это означает, что учет подобных слагаемых позволит описать *рождение* и *уничтожение* фонона. Может ли при рождении или гибели фонона родиться или погибнуть электрон? Нет! Хотя бы потому, что заряд электрона не может ни исчезнуть, ни появиться. Простейший процесс с участием электрона — рассеяние. Рассеиваясь, он либо порождает фонон, либо поглощает его. Когда мы запишем законы сохранения, это будет вполне ясно.

Сделаю шутливое замечание о терминологии. Рассеяние, связанное с рождением или поглощением электроном фонона, называют их столкновением. С тем же основанием о волке, который догнал и съел зайца, можно сказать, что произошло «столкновение волка с зайцем». Правда, волк не может зайца родить. . .

Рассмотрим законы сохранения энергии и квазиимпульса при поглощении фонона электроном. Чтобы не возникло путаницы, будем записывать квазиимпульс и энергию фонона следующим образом: $\hbar\mathbf{k}$ и $\hbar\omega$. Если столкновение нормальное, то

$$\mathbf{p} + \hbar\mathbf{k} = \mathbf{p}'; \quad \varepsilon(\mathbf{p}) + \hbar\omega = \varepsilon(\mathbf{p}'). \quad (18.3)$$

В конце предыдущей главы мы подчеркивали, что проводимость осуществляется электронами, энергия которых равна энергии Ферми. Их квазиимпульс велик, поскольку у металлов электроны как правило заполняют большую часть ячейки обратного пространства ($p \sim \hbar/a$). Большинство фононов при низких температурах (то есть при $T/T_D \ll 1$) имеет малый квазиимпульс ($\hbar k \approx (\hbar/a)(T/T_D)$). Именно этот случай нас в основном и интересует. Если квазиимпульс фонона мал, то в уравнении (18.3) можно произвести разложение по степеням $\hbar\mathbf{k}$ и сохранить лишь первый его член. Поскольку $d\varepsilon(\mathbf{p})/d\mathbf{p} = \mathbf{v}$, система (18.3) сводится к одному уравнению:

$$k\mathbf{v}_F = \omega, \quad \text{или} \quad \cos \varphi = \omega/kv_F, \quad (18.4)$$

где φ — угол между квазиимпульсом фонона и вектором скорости электрона. Необходимое условие рождения фонона таково: $s/v_F \leq 1$, где $s = \omega/k$ — фазовая скорость волны, квант энергии которой и есть фонон. Как уже упоминалось, $s \ll v_F$. Поэтому поглощение (так же, как и рождение¹⁾) фононов электронами может происходить и происходит. Однако при низких температурах квазиимпульс фонона с большой вероятностью мал, поэтому и изменение квазиимпульса электро-

¹⁾ Для рождения фонона необходимо выполнение того же условия, что и для поглощения. Более того, если поглощаются главным образом фононы с малым квазиимпульсом, то и рождаются такие же. Это одно из нетривиальных следствий абстрактного (несомненно!) принципа независимости законов механики от направления течения времени.

на очень мало. Сопротивление определяется вероятностью заметного изменения Δp квазиимпульса p ($\Delta p \sim p$). Для этого электрону надо многократно «столкнуться» с фононом (как показывает расчет, порядка $(T_D/T)^2$ раз). Длина свободного пробега в соответствующее число раз увеличивается, что и зафиксировано второй из формул (18.2). При высоких температурах для заметного отклонения достаточно одного «столкновения».

Прежде чем закончить настоящую главу, обратимся еще раз к формуле (18.4). Не удивляет ли Вас, что в ней нет постоянной Планка \hbar ? Действительно, это классическая формула — условие испускания волны частицей, движущейся со скоростью, превышающей фазовую скорость волны. Формула (18.4) служит обоснованием возможности существования самых разных физических явлений. Вот неполный перечень некоторых из них (приведенный только для того, чтобы подчеркнуть их разнообразие):

- поглощение коротковолнового звука металлом;
- затухание электромагнитных волн в плазме (так называемое затухание Ландау);
- излучение электроном, летящим со сверхсветовой скоростью, световых волн (черенковское излучение ¹⁾);
- щелчок бича или громкий звук, долетевший от пронесшегося мимо самолета, представляющий собой свидетельство того, что превышена скорость звука в воздухе (отношение скорости тела к скорости звука называют числом Маха).

С равенством (18.4) мы еще встретимся при описании сверхтекучести (см. гл. 19)

Подведем итоги. Квазичастицы полностью свободны только приближенно. При описании их движения столкновениями можно пренебречь, абстрагироваться от них. Однако существует широкий круг явлений (так называемые *кинетические явления*), которые невозможно описать, не учитывая столкновений квазичастиц. Одна из задач физической кинетики — обнаружение главных, наиболее существенных столкновений и вычисление величин, характеризующих их интенсивность. Оценкой свободы квазичастиц (законности абстракции) служат неравенства

$$\varepsilon \gg \hbar/\tau; \quad p \gg \hbar/l,$$

где $\tau = l/\langle v \rangle$ — время свободного пробега квазичастицы, $\langle v \rangle$ — ее средняя скорость; остальные обозначения известны.

¹⁾ Явление названо в честь открывателя — П.А. Черенкова. Объяснили его И.Е. Тамм и И.М. Франк. Нобелевскую премию за 1958 г. получили все три советских физика.

Глава 19. КВАНТОВЫЕ ЖИДКОСТИ

Начнем с пояснения названия главы. Законы течения обычных жидкостей — законы гидродинамики — являются следствиями законов классической механики. *Квантовыми* называют такие жидкости, макроскопическое движение которых (например, течение по трубе) происходит не так, как того требуют законы классической гидродинамики. Для описания свойств квантовых жидкостей придется привлекать квантовую механику.

Много ли в природе квантовых жидкостей? По-видимому, много. По современным представлениям, внутренность многих звезд — квантовые жидкости из ядерных частиц. Однако на Земле квантовых жидкостей в готовом виде совсем не существует. Их приходится искусственно создавать с помощью низких температур. Материал для создания квантовых жидкостей — гелий (He). Он назван гелием (солнечным) потому, что впервые был открыт с помощью спектрального анализа на Солнце, а лишь потом его источники обнаружили на Земле. Природный гелий состоит из двух стабильных изотопов — He^4 (99,999862 %) и He^3 , в ядре которого не хватает одного нейтрона. Атом He^4 — бозон, а He^3 — фермион. Гелий остается газообразным вплоть до 4,22 К. Жидкий гелий впервые получен в 1908-м году Камерлинг-Оннесом.

Попытки кристаллизовать гелий закончились неудачей: при атмосферном давлении он остается жидким вплоть до абсолютного нуля температуры. Затвердевает He при давлении выше 25 атм. Уже то, что гелий остается жидкостью при абсолютном нуле, выдает роль квантовых эффектов. Классические жидкости с обязательностью кристаллизуются при низких температурах; квантовой жидкостью становится потому, что если она превратится в кристалл, нулевые колебания ее атомов будут столь велики, что разрушат кристаллический порядок. Большой величине амплитуды нулевых колебаний атомов гелия способствует слабое взаимодействие между ними (атомы He инертны); к тому же у них относительно малая масса. Мэру квантовости можно оценить. Соответствующий параметр носит название *параметра де-Бура*:

$$\Lambda = \frac{\hbar/a}{(mE)^{1/2}}, \quad (19.1)$$

где a — среднее расстояние между атомами, m — масса атома, а E — энергия взаимодействия между атомами. Гелий имеет наибольший

параметр де-Бура по сравнению со всеми твердыми телами. Поэтому при $T = 0$ К он представляет собой не кристалл, а жидкость.

Хотя в природном газе содержится очень мало легкого изотопа гелия, удается получить макроскопические количества этой уникальной жидкости. Помогает ядерная энергетика: среди ее отходов присутствуют атомы He^3 . В руках экспериментаторов оказался целый набор квантовых жидкостей, будто специально предназначенный для исследования их свойств: бозе-жидкость (He^4), ферми-жидкость (He^3) и растворы бозе- и ферми-жидкостей ($\text{He}^4 + \text{He}^3$). Мы остановимся только на свойствах жидких He^4 и He^3 .

§ 19.1. Бозе-жидкость (He^4)

Наиболее интересное квантовое свойство жидкого He^4 — *сверхтекучесть* — способность течь без трения. Остроумными экспериментами П.Л. Капица показал, что в гелии возможны два типа движения: *нормальное*, очень похожее на течение обычных жидкостей, и *сверхтекучее*, не сопровождающееся трением о стенки капилляра, по которому протекает жидкость.

Сверхтекучесть жидкого гелия наблюдается при температуре ниже $T_\lambda = 2,17$ К. Температура T_λ называется *λ -точкой*; это точка фазового перехода 2-го рода (см. гл. 9). Она получила свое название по виду кривой (см. рис. 9.3, а), изображающей температурную зависимость теплоемкости вблизи точки перехода. При $T < T_\lambda$ жидкий гелий обозначают как HeII, а при $T > T_\lambda$ — как HeI.

Переход гелия в сверхтекучее состояние можно описать как разделение жидкости на две компоненты: нормальную с плотностью ρ_n и сверхтекучую с плотностью ρ_s . Естественно, $\rho_n + \rho_s = \rho$, где ρ — плотность гелия. Необходимо отчетливо понимать, что никакого деления атомов He на два сорта не происходит. Смысл вводимых плотностей заключается в том, что $\rho_n v_n$ есть переносимый единицей объема гелия импульс при нормальном движении со скоростью v_n ; $\rho_s v_s$ — переносимый единицей объема гелия импульс при сверхтекучем движении со скоростью v_s . Отношение ρ_s/ρ служит параметром порядка. Оно изменяется от нуля при $T = T_\lambda$ до единицы при $T = 0$ (см. рис. 9.2).

Теорию сверхтекучести построил Ландау, установив, каков энергетический спектр элементарных возбуждений в гелии.

При $T = 0$ гелий, оставаясь жидким, находится в основном состоянии. Элементарные возбуждения в жидком гелии представляют собой макроскопические звуковые волны. Если длина волны велика, то, как всегда в звуковых волнах, частота ω пропорциональна волновому вектору \mathbf{k} ; коэффициентом пропорциональности является скорость звука ($u = 235$ м/с). С ростом волнового вектора линейность нарушается и на кривой $\omega = \omega(k)$ появляется сравнительно глубокий минимум. При дальнейшем росте волнового вектора кривая $\omega = \omega(k)$

обрывается. Более коротких незатухающих волн в гелии не существует. Как и при введении фононов в кристаллах, в случае гелия можно перейти от зависимости $\omega = \omega(k)$ к зависимости

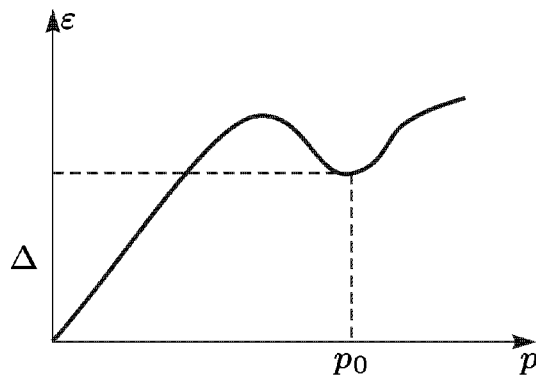


Рис. 19.1. Зависимость энергии фононов-ротонов от импульса

энергии $\varepsilon = \hbar\omega$ от импульса $p = \hbar k$. На рис. 19.1 приведена зависимость $\varepsilon = \varepsilon(p)$. Квазичастицы в HeII называют фононами, если речь идет о малых импульсах, и ротонами при $p \sim p_0$. Когда Ландау, воспользовавшись сформулированным им характером энергетического спектра гелия, объяснил природу сверхтекучести (см. ниже), кривая, изображенная на рис. 19.1, была гипотезой, правда, подтвержденной температурной зависимостью термодинамических характеристик HeII. В настоящее время зависимость $\varepsilon = \varepsilon(p)$ не вызывает сомнений, в частности, потому, что она непосредственно измерена с большой точностью (с помощью неупругого рассеяния нейтронов).

Стоит отметить, что в настоящее время гелий — самая изученная жидкость. Его многочисленные свойства исследованы и экспериментально, и теоретически. Проблема не считается решенной до тех пор, пока нет *количественного* согласия между теорией и экспериментом. Вместе с тем, несмотря на более чем полувековое изучение жидкого гелия, он до сих пор преподносит исследователям интересные открытия.

Во всех жидкостях могут распространяться звуковые волны, но только в HeII фононы-ротоны исчерпывают возможные движения ¹⁾. В нем отсутствуют характерные для обычных жидкостей индивидуальные движения атомов. В низкотемпературной фазе гелий представляет собой сильно скоррелированную систему. Именно в возникновении корреляции между атомами He⁴ заключена суть перехода в сверхтекучее состояние.

При $T = 0$ фононов-ротонов в гелии нет вовсе. При $T > 0$ они, как и фононы в кристалле, представляют собой бозе-газ, число частиц (квазичастиц!) в котором растет с ростом температуры. Средой,

¹⁾ В HeII возможны и возбуждения другого типа — квантованные вихри (микроскопические смерчи). Их образование сопровождается разрушением сверхтекучести по оси вихря.

в которой существуют фононы-ротоны, является гелий в основном состоянии. Он существенно отличается от кристаллической решетки — вместителища фононов в твердых телах. Во-первых, гелий однороден и изотропен. Поэтому \mathbf{p} здесь — настоящий импульс, а не квазиимпульс. Во-вторых, гелий — жидкость; он может перетекать, не сохраняет формы и заполняет сосуд, в котором находится.

Теперь можно разъяснить, что собой представляют два типа движения, о которых говорилось выше.

Нормальное движение — это движение газа фононов-ротонов. Квазичастицы сталкиваются со стенками сосудов, отдают им часть своего импульса, тормозятся, испытывают трение — все, как в обычной жидкости. Воспользовавшись тем, что квазичастицы представляют собой бозе-газ, можно вычислить, чему равен импульс единицы объема гелия, если его скорость равна v_n . Это позволяет узнать, какую массу переносит единица объема газа фононов-ротонов. Так определяется величина и температурная зависимость ρ_n ; разность $\rho - \rho_n$ есть ρ_c — плотность сверхтекучей компоненты.

Сверхтекучее движение — движение невозбужденного гелия, при котором газ фононов-ротонов в среднем покоится. При сверхтекучем движении единицы объема со скоростью v_c переносится масса ρ_c .

К сожалению, мы не можем подробно рассказать о том, как представление о двух типах движения дает возможность описать поистине парадоксальные свойства гелия. Мы только постараемся объяснить, почему сверхтекучее движение не сопровождается трением, приводящим к торможению течения.

Для простоты рассмотрим движение через гелий при $T = 0$ макроскопического предмета массы M (это стенки капилляра, по которому течет гелий; мы перешли в систему координат, движущуюся вместе с гелием). В результате трения энергия движения тела уменьшается за счет того, что часть ее переходит в тепло. Тепловое движение в He осуществляют фононы-ротоны. Повышению температуры соответствует увеличение числа квазичастиц. Вот этой простой мыслью мы и воспользуемся.

Выясним, разрешают ли законы сохранения импульса и энергии рождение фонона-ротона. Если такое событие разрешено, то должны выполняться следующие равенства:

$$\mathbf{P}' = \mathbf{P} + \mathbf{p}; \quad (P')^2/2M = P^2/2M + \varepsilon(p),$$

где \mathbf{P}' и \mathbf{P} — импульсы тела до и после рождения квазичастицы. Подставим значение \mathbf{P}' из первого уравнения во второе. Слагаемое $p^2/2M$ пренебрежимо мало. Заметив, что $\mathbf{P}/M = \mathbf{V}$ — скорость тела, имеем

$$Vp = \varepsilon(p), \quad \text{или} \quad \cos \theta = \varepsilon(p)/Vp,$$

где θ — угол, определяющий направление импульса фонона-ротона относительно скорости тела. Необходимым условием рождения фонона-ротона является неравенство

$$\frac{\varepsilon(p)/p}{V} < 1.$$

Однако если $V < \min[\varepsilon(p)/p]$, то рождение фононов-ротонов невозможно. Фононно-ротонный спектр гелия таков, что

$$\min[\varepsilon(p)/p] \neq 0,$$

то есть если $V < V_{\text{крит}} = \min[\varepsilon(p)/p]$, то движение тела не сопровождается трением. Это утверждение, естественно, переносится на скорость течения жидкости по капилляру. Таким образом, выведенные формулы являются условием существования сверхтекучести. Их называют *условием Ландау*.

В гл. 10 мы рассказали о бозе-газе частиц. Может показаться, что свойства HeII очень напоминают свойства бозе-газа при температуре ниже температуры бозе-эйнштейновской конденсации. Сходство, конечно, есть, но спектр элементарных возбуждений бозе-газа не удовлетворяет условию Ландау: идеальный бозе-газ не обладает сверхтекучестью. А вот неидеальный бозе-газ, как показал Н.Н. Боголюбов, обладает свойством сверхтекучести и может служить моделью сверхтекучей квантовой жидкости. Надо, правда, иметь в виду, что HeII — жидкость, а не газ.

§ 19.2. Ферми-жидкость (He^3)

В книге «Физики шутят» есть такая острота: «Господь Бог, видя, что физики-теоретики продвинулись в понимании устройства Мира, говорит: «Подбросим им нелинейность...». Насколько проще было бы рассказывать о свойствах He^3 , если бы не была открыта его сверхтекучесть! Правда, температура перехода в сверхтекучее состояние составляет порядка тысячной градуса Кельвина, а до нее жидкий He^3 можно считать *нормальной ферми-жидкостью*. Это означает, что свойства He^3 хорошо описываются теорией ферми-жидкости Ландау. О сверхтекучести ферми-жидкости мы поговорим позже, а пока сформулируем некоторые положения теории Ландау.

Ферми-жидкость очень напоминает ферми-газ:

1) все атомы He^3 при $T = 0$ заполняют ферми-сферу, радиус которой определяется той же формулой, что и в случае ферми-газа: $p_F = (3\pi^2 n)^{1/3} \hbar$, где n — плотность жидкости;

2) элементарные возбуждения рождаются парами: частица вне сферы Ферми и дырка внутри нее. Вблизи ферми-сферы зависимость энергии частицы и дырки от импульса похожа на приведенную нами

в гл. 17:

$$\begin{aligned} \text{для частицы} \quad \varepsilon &= v_F(p - p_F), & p > p_F; \\ \text{для дырки} \quad \varepsilon &= v_F(p_F - p), & p_F > p. \end{aligned}$$

Однако здесь $v_F = p_F/m_{\text{эф}}$, а $m_{\text{эф}} \neq m$, где m — масса атома He^3 . Эффективную массу можно вычислить по известному взаимодействию атомов гелия друг с другом. Теория Ландау объяснила многие свойства жидкого He^3 и предсказала ряд свойств, которые в дальнейшем были открыты.

Теория ферми-жидкости естественным образом переносится на электроны проводимости. Правда, состояние частиц (электронов) и дырок определяет квазиимпульс, а не импульс, а ферми-сферу заменяет ферми-поверхность. При этом многие из формул, выведенные согласно теории ферми-жидкости Ландау, не отличаются от формул, которые получены в предположении, что электроны проводимости — ферми-газ. Последнее относится и к температурной зависимости теплоемкости, и к значению константы в законе Видеманца–Франца, и к зависимости сопротивления от температуры.

Современная электронная теория металлов — прекрасный пример активного использования абстракции. Удастся не только обосновать применение упрощенной теории, но и выяснить границы ее применимости.

Важная особенность электронной жидкости в металлах — неустраняемое взаимодействие электронов с фононами. Его проявление мы рассматривали на примере зависимости сопротивления от температуры (см. гл. 18). Однако столкновения не исчерпывают эффект подобного взаимодействия.

На примере He^3 мы видели, что взаимодействие атомов жидкости друг с другом не сводится к рассеянию. Конспективно излагая теорию ферми-жидкости Ландау, о рассеянии мы вообще не упоминали. Мы говорили о *перенормировке* массы, приведшей к *количественному* изменению зависимости энергии частиц и дырок от импульса. Всегда ли перенормировка ограничивается количественными изменениями? Оказывается, нет. Далее мы приведем один, но очень впечатляющий пример.

§ 19.3. Сверхпроводимость

В 1911 г., изучая в Лейдене (Голландия) сопротивление ртути при температуре жидкого гелия, Г. Камерлинг-Оннес обнаружил, что при 4 К оно скачком обращается в нуль. Так была открыта *сверхпроводимость* — одно из интереснейших макроскопических квантовых явлений. То, что сверхпроводимость — квантовое явление, естественно, было понято не сразу. Долгое время она оставалась дразнящей загадкой физики металлов. Лишь в 1957-м году, через 46 лет после открытия,

американскими физиками Дж. Бардиным, Л. Купером и Дж. Шриффером была построена микроскопическая теория сверхпроводимости.

После 1911-го года было обнаружено множество сверхпроводников. Их поведение тщательно исследовано. Накоплен огромный экспериментальный материал. Макроскопические свойства сверхпроводников подробно описаны. Характер фазового перехода из нормального в сверхпроводящее состояние строго установлен: в магнитном поле это переход I-го рода, а в его отсутствии — переход II-го рода. Выяснилось, что потеря сопротивления сопровождается изменением магнитных свойств металла. Он превращается в идеальный диамагнетик, то есть его магнитная восприимчивость составляет $-1/4\pi$, а среднее магнитное поле в толще сверхпроводника равно нулю (эффект Мейсснера).

Нормальное состояние с отличным от нуля сопротивлением может быть восстановлено, если либо пропустить по образцу достаточно большой ток, либо поместить его в достаточно сильное магнитное поле.

Не имея возможности даже конспективно рассказать о свойствах сверхпроводников, отметим только одно из них, получившее название *изотопического эффекта*. В 1950-м году впервые было обнаружено, что температура T_K сверхпроводящего перехода зависит от изотопического состава сверхпроводника, причем приближенно выполняется равенство $T_K M^{1/2} = \text{const}$.

В 1955-м году в Оксфорде вышла монография Р. Пайерлса «Квантовая теория твердых тел». Она заканчивалась словами: «Имея такой ключ, как изотопический эффект, и применяя теорию Фрелиха и Бардина в качестве отправной точки, мы можем ожидать дальнейшего прогресса в этой области»¹⁾. Это было сказано за два года до появления в печати теории сверхпроводимости. Упомянутая теория Фрелиха–Бардина является теорией взаимодействия электронов с фононами, не ограниченной рассмотрением только рассеяния квазичастиц. Изотопический эффект указывает на то, что с большой вероятностью сверхпроводимость связана со взаимодействием электронов с колебаниями атомов кристаллической решетки (или, на квантовом языке, с фононами). Напомним, что все частоты таких колебаний обратно пропорциональны $M^{1/2}$.

Прежде чем очень кратко (по необходимости) рассказать о том, почему металл становится сверхпроводником, сделаем небольшое отступление.

Мы неоднократно подчеркивали, что характер идеализации зависит от задачи, которую ставит перед собой теория. При рассмотрении сопротивления металлов нам было *достаточно* использовать первое приближение теории возмущений при учете электрон–фононного вза-

¹⁾ Цитирую по русскому переводу А.А. Абрикосова (М.: Иностранная литература, 1956).

имодействия. Для понимания природы сверхпроводимости этого недостаточно; нужно более высокое приближение.

Любопытно отметить, что роль электрон–фононного взаимодействия можно менее определенно, чем по изотопическому эффекту, но все же усмотреть и при сравнении температурных зависимостей металлов, один из которых ниже точки перехода становится сверхпроводником, а другой нет. При $T > T_K$ сопротивление «будущего» сверхпроводника выше (рис. 19.2).

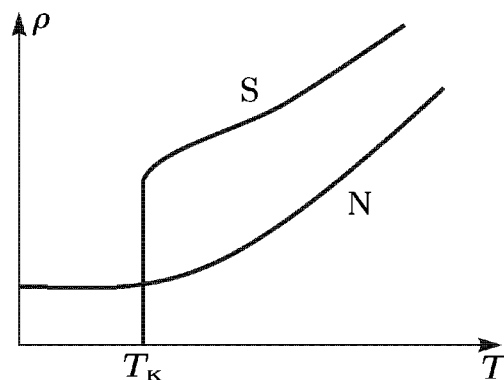


Рис. 19.2. Температурные зависимости сопротивления нормального (N) и сверхпроводящего (S) металлов

Как уже говорилось, первое приближение теории возмущения при учете электрон–фононного взаимодействия описывает излучение и поглощение фонона электроном. Второе приближение можно трактовать как взаимодействие электронов друг с другом с помощью фононов.

В квантовой физике у каждого взаимодействия есть специфический агент-переносчик. Электростатическое кулоновское взаимодействие осуществляется через фотоны, за ядерные силы ответственны π -мезоны.

Если один электрон в металле излучает фонон, а другой его поглощает, то в результате электроны притягиваются друг к другу. Притяжение невелико, но во многих случаях способно преодолеть кулоновское (электростатическое) отталкивание. Последнему благоприятствует тот факт, что притягиваются относительно далеко находящиеся друг от друга электроны, отталкивание между которыми уменьшено процессом, получившим название *экранирования*. Экранирование представляет собой небольшой сдвиг зарядов в целом нейтральной системы, несколько компенсирующий заряды электронов и ионов.

Небольшое притяжение способно привести к образованию своеобразных квазиатомов или квазимолекул, состоящих из двух электронов. Они получили название куперовских пар (по имени предсказавшего их существование Л. Купера). Заряд куперовской пары равен $2e$.

Образование куперовских пар приводит к перенормировке энергетического спектра элементарных возбуждений. Теперь для увеличения энергии электронов необходимо разорвать куперовскую пару, преодолев энергию связи между ними. Существование куперовских

пар приводит к тому, что энергия электронов и дырок не обращается в нуль. Энергетическая зависимость для сверхпроводника имеет щель Δ (рис. 19.3):

$$\varepsilon = [\Delta^2 + v_F^2(p - p_F)^2]^{1/2},$$

где $p > p_F$ соответствует электронам, а $p < p_F$ — дыркам.

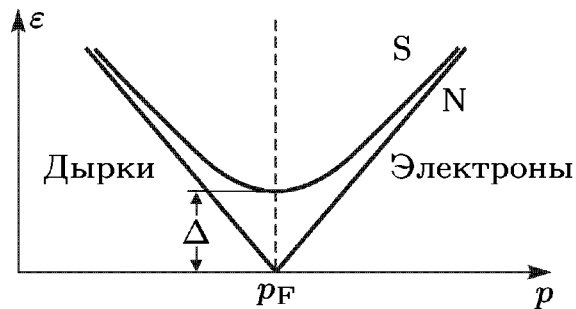


Рис. 19.3. Энергетический спектр электронов и дырок в нормальном и сверхпроводящем металлах

Согласно теории БКШ (такой аббревиатурой обычно заменяют фамилии авторов теории сверхпроводимости — Бардина, Купера, Шриффера) ширина щели при $T = 0$ по порядку величины близка к $k_B T_K$, а при $T = T_K$ обращается в нуль. Заметим, что величина $\Delta = \Delta(T)$ может служить параметром порядка при описании сверхпроводимости.

Подведем итоги.

Электрон–фононное взаимодействие приводит к перенормировке энергетического спектра элементарных возбуждений. Она вызвана тем, что электроны объединяются в пары. Перенормировка охватывает только электроны, имеющие энергию, равную или близкую энергии Ферми. Последнее является следствием того, что энергия фонона значительно меньше фермиевской энергии. Однако надо помнить, что перенормируются «самые главные» электроны. Именно они ответственны за металлические свойства — электро-, теплопроводность и линейный ход температурной зависимости теплоемкости. Заметим, что электроны, составляющие пару, имеют противоположные импульсы, поэтому их суммарный импульс равен нулю. Размер куперовской пары Δr_K определяется величиной щели: $\Delta r_K = v_F \hbar / \Delta$. Думаю очевидно, что для области, в которой локализованы электроны в паре (размер пары), справедливо соотношение неопределенности. Численное значение Δr_K по атомным масштабам очень велико: $\Delta r_K \sim 10^{-6} \div 10^{-5}$ см. Наглядно представить себе, что такое электронная жидкость в сверхпроводящем металле, боюсь, невозможно, но два обстоятельства несомненны: после объединения электронов в пары электронная жидкость состоит именно из куперовских пар, а куперовские пары — бозоны (по-видимому, спин пары как правило равен нулю). Таким образом, в сверхпроводящем состоянии электронная жидкость представляет

собой скорее бозе-, чем ферми-жидкость. Правда, элементарные возбуждения — электроны и дырки — фермионы. При этом зависимость их энергии от импульса удовлетворяет условию Ландау. Вывод: электроны в сверхпроводнике — сверхтекучая жидкость.

Эксперимент подтвердил справедливость нарисованной выше картины. Был измерен заряд куперовской пары. Он оказался равным $2e$. Существование щели Δ также подтверждено многочисленными экспериментами.

Вернемся к гелию. Природа сверхтекучести He^3 близка природе сверхпроводимости: и здесь также образуются пары (их существование, как и переход в сверхтекучее состояние He^3 , предсказал Л. Питаевский)...

Хотел написать: «Образуются пары со всеми вытекающими отсюда последствиями». В действительности последствия значительно интересней, чем можно было ожидать. Оказывается, у He^3 две сверхтекучие фазы, своеобразное поведение в магнитном поле, более сложная, чем у БКШ-сверхпроводников, структура параметра порядка. В нескольких словах об этом не расскажешь.

Глава 20. ВОЗМОЖНА ЛИ ОТРИЦАТЕЛЬНАЯ ТЕМПЕРАТУРА?

Конечно, Вы понимаете, что речь пойдет об абсолютной температуре. Если пользоваться одной из обычных шкал (Цельсия, Реомюра или Фаренгейта), то отрицательная температура вполне возможна. Правда, по шкале Фаренгейта нуль — очень низкая температура (по этой шкале температура таяния льда составляет $+32^{\circ}\text{F}$).

Нуль по абсолютной шкале температуры означает, что теплового движения в теле вообще нет. Как же можно еще понизить температуру? Да и абсолютный нуль недостижим. Последнему утверждению, как мы отмечали, даже присвоили титул третьего начала термодинамики. Понизить температуру ниже абсолютного нуля, похоже, действительно нельзя. Однако, возможно, для получения отрицательной температуры нужно не опускать ее, а поднимать? Подождем пока отвечать на этот вопрос.

Температура относится к столь обычным, домашним понятиям, что ее строгий смысл забывается, а многим, скорее всего, и неизвестен. Напомним: *температура характеризует термодинамически равновесное состояние тела*. Состояние термодинамического равновесия предполагает не только однородность температуры тела, но и то, что тело находится в равновесии со всем своим окружением. Если температуры двух соприкасающихся тел различаются, то обязательно идет процесс их выравнивания, а равновесие наступает лишь тогда, когда температуры сравняются.

Выравнивание температур — сравнительно медленный процесс. Следя за изменением температуры каждого из тел, можно забыть о том, что оба они, строго говоря, не находятся в состоянии равновесия, а процесс приближения к состоянию равновесия характеризовать зависимостью от времени разности их температур. Это, между прочим, типичный пример абстракции — идеализация.

При произнесении слова «идеализация» возникает желание проверить, имеем ли мы на нее право, то есть подкреплена ли она какими-либо неравенствами.

Поясню, о чем идет речь.

В неравновесном состоянии микроскопические частицы, из которых состоит тело, движутся совсем не так, как в равновесном. Рассмотрим два контактирующие между собой тела *A* и *B*. Выведем тело *B* из состояния равновесия и предоставим его своей судьбе. Судьба тел *A* и *B*,

если они действительно ни с чем кроме друг друга не взаимодействуют, предопределена: постепенно оба они должны перейти в состояние равновесия¹⁾).

Переход тела в состояние равновесия называют *процессом релаксации*. Часто процесс релаксации естественным образом происходит в два этапа. Сначала весьма быстро (как правило, за миллионные доли секунды) выведенное из равновесия тело B приходит в состояние внутреннего равновесия, то есть равновесия, не связанного с существованием тела A . В частности, температура тела B , отвечающая средней энергии его частиц, не будет совпадать с температурой тела A . Потом значительно медленнее (в течение минут, иногда часов) выравниваются температуры обоих тел. Рассмотренный сценарий предполагает, что взаимодействие между телами A и B существенно слабее, чем взаимодействие частиц внутри каждого из них.

Право на идеализацию дается неравенством, которое мы запишем, не используя математических символов:

*время первого этапа релаксации значительно меньше времени
выравнивания температур.*

Подчеркнем: это очень сильное неравенство! Время выравнивания температур может в миллионы раз превосходить время первого этапа релаксации. Состояние тела после окончания первого этапа релаксации часто называют *квазиравновесным*.

Приведенное неравенство объясняет, почему в большинстве случаев мы можем не задумываться о том, в равновесном или квазиравновесном состоянии находится тело, температуру которого мы измеряем. Как правило, это квазиравновесное состояние, однако неравновесность предельно мала. Мерой малости служит отношение времен релаксации.

Углубимся мысленно в структуру тела. Для определенности рассмотрим два объекта:

- 1) полупроводник n -типа, в котором достаточно много электронов;
- 2) диэлектрик с парамагнитными примесями.

Первый объект. Из-за того что электроны в тысячи раз легче атомов, обмен энергией между ними и атомами или ионами кристаллической решетки полупроводника весьма затруднен. Поэтому все, что мы говорили о релаксации в системе из двух тел, применимо к электронам полупроводника и к его кристаллической решетке, хотя и кристаллическая решетка, и электроны принадлежат одному телу.

¹⁾ Стоит подчеркнуть, что тела теплокровных животных, к которым относится человек, никогда не находятся в равновесии с окружающей их средой.

Правда, различие времен релаксации в этом случае не столь велико. Однако поскольку обмен энергией внутри электронной подсистемы и внутри подсистемы атомов/ионов, из которых состоит кристаллическая решетка, идет быстрее, чем между электронами и тяжелыми частицами, можно сделать следующий вывод:

*в одном теле могут сосуществовать две подсистемы
с отличающимися температурами.*

Второй объект. Пусть частицы примеси имеют спин, равный одной второй. Тогда из-за пространственного квантования каждая из них может находиться только в двух спиновых состояниях. Любому состоянию соответствует свой энергетический уровень. Количество атомов примеси, находящихся в каждом из состояний, зависит от температуры. При очень низкой температуре почти все частицы примеси находятся на нижнем уровне. С ростом T число частиц на верхнем уровне увеличивается, однако в условиях равновесия их всегда меньше, чем на нижнем уровне. Лишь при бесконечно большой температуре (абстракция) заселенности обоих уровней сравниваются.

Что же произойдет, если мы искусственно заселим верхний уровень бóльшим числом частиц, чем нижний? Нередки условия, при которых взаимодействие между спиновыми степенями свободы сильнее, чем между спинами и кристаллической решеткой. Заметим, что перейти с верхнего уровня на нижний без участия атомов тела невозможно. Пренебрегая взаимодействием спинов с решеткой, мы обрекаем спины примесных частиц на в некотором смысле изолированное существование. Они могут успеть прийти в квазиравновесное состояние до того, как придут в состояние равновесия со всем твердым телом. Какая температура окажется у спинов, если в состоянии с бóльшей энергией будет больше частиц, чем в состоянии с меньшей? Как ни парадоксально, *отрицательная!*

Сделанное в предыдущем абзаце утверждение подтверждается законом Аррениуса, согласно которому

$$N_2/N_1 = \exp(-\Delta E/kT), \quad \Delta E = E_2 - E_1 > 0,$$

где индексы «1» и «2» относятся к нижнему и верхнему уровню соответственно; T — абсолютная температура спиновой подсистемы; N_1 и N_2 — заселенности уровней. Видно, что при $N_2/N_1 > 1$ с необходимостью $T < 0$.

Описанная выше заселенность уровней называется *инверсной*. Реализовать ее непросто, но, как говорится, игра стоит свеч. Инверсная заселенность позволяет черпать из тела энергию в удобной для использования форме, заставляя частицы возвращаться в нормальное состояние заселенности (по этому принципу функционирует ряд современных приборов).

Итак, мы ответили на вопрос, вынесенный в название главы. Однако при чем же здесь абстракция? Дело в том, что и первый, и второй рассмотренные объекты по предположению — тела в неравновесном состоянии. Абстрагируясь от детальных черт их состояний, мы описываем эти тела сравнительно просто, считая, что имеется частичное равновесие внутри естественно выделенных подсистем. Это удобное описание, позволяющее рассмотреть многие явления наглядно, не прибегая к громоздким вычислениям.

Потеря подробностей полностью оправдана: абстрагирование, если оно законно, всегда приносит пользу.

Завершить беседы об абстракции в физике мы решили конкретным примером и выбрали для этого случай введения отрицательной температуры.

Заключительные замечания. Природа учит абстрагировать

В ненаучном обиходе, как мы неоднократно повторяли, глагол «абстрагировать» чаще всего употребляется с негативным оттенком. «Ты абстрагируешься от. . .» — далее следует перечень того, что именно собеседник не учел, забыл упомянуть, чем пренебрег. Если разговор идет на повышенных тонах, то забывается, что прийти к общему мнению можно как правило только тогда, когда собеседники смогут от чего-то абстрагироваться. Спор же фактически возникает из-за того, что они не могут договориться, от чего конкретно можно абстрагироваться.

Одной из загадок науки Эйнштейн считал возможность исследования природы с помощью математики. Разделяя мнение великого ученого, мы хотим отметить, что применять математику к изучению явлений природы было бы невозможно, если бы не абстракция. Часто подчеркивают единство природы. Все связано со всем. К счастью, это не конструктивная формула. Явления, тела, разнообразные типы движения естественным образом вычленяются из общей массы. Это свойство мира, в котором мы существуем, представляет собой необходимое условие возможности его изучения. Невозможно было бы исследовать явления природы, если бы каждое из них зависело от всех остальных. Само понятие *явление* предполагает выделенность из всего происходящего.

Некоторые физики-теоретики заняты исследованием соотношений между различными величинами, обеспечивающими своеобразие нашего мира. По их мнению, если бы значения отдельных величин отличались от истинных, то в звездах не могли бы протекать термоядерные реакции, Солнце не давало бы Земле тепла, на Земле не было бы жизни и т. д., и т. п. В своем желании понять устройство мира они утверждают, что мир таков, как он есть, *потому что* иначе не могла бы возникнуть жизнь, а в результате эволюции — познающий субъект — человек. Подобное рассуждение мне представляется вненаучным.

Уверен, что *познаваемым* делает мир абстрагирование. Возможность абстрагирования — одно из свойств нашего Мира.

Одно из достоинств абстрагирования — возможность задавать вопросы и получать на них ответы *последовательно*.

Абстрагировавшись от взаимодействия между молекулами газа, позволив себе пренебречь их размерами, ученые поняли, почему давление P , объем V и абсолютная температура T связаны между собой знаменитым уравнением Бойля–Мариотта–Гей-Люссака:

$$PV = RT,$$

и сумели определить газовую постоянную R (если V — объем одного моля газа, то $R = 1,9872$ кал/моль·К). После учета размеров молекул исправленное уравнение Бойля–Мариотта–Гей-Люссака оказалось способным описать фазовый переход газ–жидкость и существование критической точки.

Электромагнитные волны пронизывают космическое пространство. С одной стороны, они, несомненно, ведут независимое существование, а с другой — приносят сведения о весьма удаленных объектах Вселенной, свет от которых доходит до нас за миллионы лет. Космос не пуст, но взаимодействие с веществом не уничтожает информацию, несомую светом. Причиной этого является малость заряда, позволяющая абстрагироваться от взаимодействия электромагнитных волн с веществом космоса.

Изучая свойства атомов, молекул, конденсированных систем, мы можем не думать о структуре атомных ядер, абстрагироваться от нее. Правда, часто неожиданно, обнаруживаются явления, как бы принадлежащие двум областям физики. Особенно яркий пример — ядерный гамма-резонанс без отдачи ядра (*эффект Мессбауэра*). Причина отсутствия отдачи заключается в том, что ядро встроено в твердое тело. Эффект Мессбауэра дал в руки экспериментаторам прекрасный метод исследования плохо доступных свойств твердых тел. Однако в плане нашего рассказа новый эффект не заставил пересмотреть *объяснение* ни свойств твердых тел, ни ядерных явлений. Каждая область продолжает свое независимое существование и при этом, не без успеха используя достижения своего соседа, легко абстрагируется от его существования по соседству.

Двадцатый век заслуженно считается веком двух революций в физике. Так окрестили создание теории относительности и квантовой механики. Замечательная особенность революционизировавших естествознание теорий — их бережное отношение к своим предшественницам. И из теории относительности, и из квантовой механики выводятся уравнения классической ньютоновской механики. Подобный вывод позволяет выяснить пределы применимости классической механики и даже определить, какова точность классических уравнений. В естественных науках нередко можно оценить меру использованной абстракции.

Изучая природу, физики создают определенную картину Мира. Как художник не может изобразить на полотне сразу все и вынужден

делать выбор, так и физик разделяет природу, отделяет слабо связанные части, создает вместо одной картины целый атлас — множество картин, которые все вместе изображают постепенно постигаемый Мир. Каждая страница атласа — результат абстракции, пренебрежения тем, чем можно пренебречь, что можно перенести на другую страницу.

Огромную роль в отборе материала для «карт атласа» играют руководящие идеи, которые всегда представляют собой абстракцию. Именно они упорядочивают результаты экспериментов и наблюдений, определяют их место «в атласе» и «на карте». Без идей, без теории природа безымянна и неисчислима. Число привнесено в природу теорией. Эксперимент служит проверкой созданной теории. Теория предсказывает определенные числа и соотношения между ними. Опыт дает возможность проверить ее выводы. Сравнение чисел, полученных на основании теории, с числами, добытыми экспериментальным путем, есть основной этап, без которого невозможно выяснить, понимаем ли мы природу. Совпадение — надежный критерий справедливости используемой абстракции. Точность — мера абстракции.

Природа учит абстрагировать и дает возможность проверить законность использования абстракции.

ОГЛАВЛЕНИЕ

Введение	3
--------------------	---

I. Эволюция абстрактных понятий в математике

Введение	16
--------------------	----

Глава 1. Простейшие структуры	19
---	----

§ 1.1. Целые числа сами по себе	19
---	----

§ 1.2. Целые числа со сложением и умножением	20
--	----

Глава 2. Теория чисел	21
---------------------------------	----

§ 2.1. Две составляющие теории чисел	21
--	----

§ 2.2. Великая теорема Ферма	22
--	----

Глава 3. Развитие понятия числа	24
---	----

§ 3.1. Рациональные числа	24
-------------------------------------	----

Глава 4. Иррациональные числа	26
---	----

§ 4.1. Иррациональные числа и греческая математика	27
--	----

§ 4.2. Первое признание иррациональных чисел	28
--	----

§ 4.3. Новое действие — переход к пределу	29
---	----

§ 4.4. Итерационный метод и теория пределов	30
---	----

§ 4.5. Рациональные числа и рациональные точки	32
--	----

§ 4.6. Разрезы Дедекинда	32
------------------------------------	----

§ 4.7. Теория Дедекинда	34
-----------------------------------	----

§ 4.8. Основная теорема теории пределов	35
---	----

§ 4.9. Бесконечные десятичные дроби	37
---	----

§ 4.10. Обратная задача теории пределов	37
---	----

Глава 5. Векторы	40
----------------------------	----

§ 5.1. Сложение векторов и умножение на числа	40
---	----

§ 5.2. Скалярное произведение векторов	41
--	----

§ 5.3. Векторное произведение	42
---	----

§ 5.4. Физические величины как векторы	43
--	----

Глава 6. Комплексные числа	46
§ 6.1. Появление комплексных чисел	46
§ 6.2. Комплексные числа	47
§ 6.3. Три «лика» комплексных чисел	48
§ 6.4. Пример: аналитическая геометрия	48
§ 6.5. Три лика комплексных чисел (продолжение)	50
§ 6.6. Комплексная плоскость	52
§ 6.7. Формула Муавра	54
§ 6.8. Метод комплексификации	56
§ 6.9. Алгебраические уравнения	58
Глава 7. Функции и алгоритмы	60
§ 7.1. Эволюция понятия функции	60
§ 7.2. «Хорошие» и «парадоксальные» функции	63
Глава 8. О теории аналитических функций	67
§ 8.1. Аналитические функции	67
§ 8.2. Первое удивительное свойство аналитических функций	68
§ 8.3. Второе свойство. Ряды Тейлора	69
§ 8.4. Контурные интегралы	71
§ 8.5. Аналитические функции встречаются на каждом шагу	72
§ 8.6. Теорема Лиувилля	73
Глава 9. Конечные и бесконечные множества	75
§ 9.1. Первые шаги теории	75
§ 9.2. Теория множеств и логика	76
§ 9.3. Сравнение множеств по Кантору	77
§ 9.4. Мощность множества	79
Глава 10. Похвальное слово элементарной алгебре	82
Глава 11. Математика и физика. Немного истории	84
Глава 12. Аксиоматический метод Гильберта	88
§ 12.1. Сущность метода	88
§ 12.2. Непротиворечивость системы аксиом	90
§ 12.3. Полнота системы аксиом	94
§ 12.4. Проблема континуума	95

§ 12.5. Теория Геделя	95
§ 12.6. Краткие выводы	97
Глава 13. Пространства, операторы и функциональный анализ	98
§ 13.1. Основные типы пространств	100
Глава 14. Линейные пространства	101
§ 14.1. Сложение и умножение как абстрактные операции	101
§ 14.2. Линейные пространства	102
Глава 15. Метрические пространства	104
§ 15.1. Два примера метрических пространств	104
§ 15.2. «Хорошие» (полные) и «плохие» (неполные) пространства	105
§ 15.3. Непрерывные функции и компакты	109
Глава 16. Гильбертовы и евклидовы пространства	112
§ 16.1. Скалярное произведение	112
§ 16.2. Простейшие теоремы и новые понятия	114
§ 16.3. Абстрактное описание гильбертова пространства	118
Глава 17. Банаховы пространства и нормированные кольца	120
§ 17.1. Банаховы пространства	120
§ 17.2. Нормированные кольца	121
Глава 18. Приложения теории гильбертовых и евклидовых пространств	123
§ 18.1. Неравенство Коши	123
§ 18.2. Приложения к теории аппроксимации	124
§ 18.3. Обработка результатов ядерного эксперимента	126
Глава 19. Решение задач и теоремы существования	129
§ 19.1. О роли компьютеров	129
§ 19.2. Два лика задачи о предсказании будущего	130
§ 19.3. Почему математики считают важным доказательство теорем существования?	132
§ 19.4. Пример: между Сциллой и Харибдой	134
§ 19.5. Операторы и уравнения	135

§ 19.6. Рассказ о принципе Дирихле	137
§ 19.7. Как поймать льва в пустыне?	138
§ 19.8. Принцип неподвижной точки	139
§ 19.9. Ограниченные операторы	141
§ 19.10. Сжимающие операторы	142
Глава 20. Спектры в физике и математике	144
§ 20.1. Немного истории	144
§ 20.2. Спектр оператора и физические спектры	147
§ 20.3. Эрмитовы операторы	149
§ 20.4. Эрмитовы операторы в конечномерном пространстве . . .	151
§ 20.5. Интегральные операторы Фредгольма в абстрактной форме	153
§ 20.6. Спектр вполне непрерывного эрмитового оператора не пуст	155
§ 20.7. Спектр вполне непрерывного эрмитового оператора	156
Глава 21. Спектры и теория Фурье	159
§ 21.1. Теория Фурье	159
§ 21.2. Обобщенные ряды Фурье	161

II. Абстракция в физике

Введение	166
Глава 1. Классическая механика	168
Глава 2. Причина революций	174
Глава 3. Квантовая механика	180
Глава 4. Системы координат. Релятивистская революция	195
Глава 5. Элементарные частицы	204
Глава 6. Тождественность и неразличимость	208
Глава 7. Спин. Фермионы и бозоны	214

Глава 8. Феноменология и абстракция	223
Глава 9. Параметр порядка — абстрактное понятие . . .	238
Глава 10. Является ли идеализация абстракцией?	251
Глава 11. В поисках истинной абстракции. Всегда ли пространство трехмерно?	258
Глава 12. Размерные величины	265
Глава 13. Безразмерные отношения	273
Глава 14. Абстракция или реальность?	280
Глава 15. Что такое точка?	285
Глава 16. Квазичастицы-бозоны	289
§ 16.1. Экситон Френкеля	290
§ 16.2. Фононы	293
§ 16.3. Магноны	302
§ 16.4. Экситон Ванье–Мотта	306
§ 16.5. Квазичастицы-бозоны и кванты макроскопических волн.	307
Глава 17. Электроны проводимости. Квазичастицы-фермионы	310
Глава 18. Квазичастицы не вполне свободны	323
Глава 19. Квантовые жидкости	330
§ 19.1. Бозе-жидкость (He^4)	331
§ 19.2. Ферми-жидкость (He^3)	334
§ 19.3. Сверхпроводимость	335
Глава 20. Возможна ли отрицательная температура? . .	340
Заключительные замечания. Природа учит абстрагировать	344

КАГАНОВ Моисей Исаакович
ЛЮБАРСКИЙ Григорий Яковлевич

АБСТРАКЦИЯ В МАТЕМАТИКЕ И ФИЗИКЕ

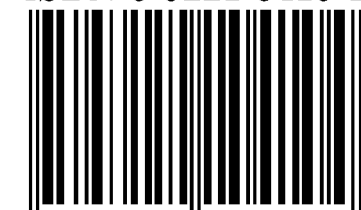
Редактор *О.А. Пенина*
Оригинал-макет: *М.В. Башевой*
Оформление переплета: *А.Ю. Алехина*

ЛР № 071930 от 06.07.99. Подписано в печать 18.08.04.
Формат 60×90/16. Бумага офсетная. Печать офсетная.
Усл. печ. л. 22. Уч.-изд. л. 24. Заказ №

Издательская фирма «Физико-математическая литература»
МАИК «Наука/Интерпериодика»
117997, Москва, ул. Профсоюзная, 90
E-mail: fizmat@maik.ru, fmlsale@maik.ru; <http://www.fml.ru>

Отпечатано с готовых диапозитивов
в ФГУП «Ивановская областная типография»
153008, г. Иваново, ул. Типографская, 6.
E-mail: 091-018@adminet.ivanovo.ru

ISBN 5-9221-0410-1



9 785922 104104