

КУРС ОБЩЕЙ ФИЗИКИ

А.С. Кингсеп, Г.Р. Локшин, О.А. Ольхов

ОСНОВЫ ФИЗИКИ

Том I

МЕХАНИКА.

ЭЛЕКТРИЧЕСТВО И МАГНЕТИЗМ.

КОЛЕБАНИЯ И ВОЛНЫ.

ВОЛНОВАЯ ОПТИКА

Издание второе,
исправленное

Под редакцией А.С. Кингсепа

*Рекомендовано Министерством образования
Российской Федерации в качестве учебника
для студентов высших учебных заведений*



МОСКВА
ФИЗМАТЛИТ®
2007

УДК 530.1
ББК 22.3
К 41

~ *Технический университет* ~

Кингсеп А. С., Локшин Г. Р., Ольхов О. А. **Курс общей физики. Основы физики.** Учеб. пособие: для вузов. В 2 т. Т. I. **Механика. Электричество и магнетизм. Колебания и волны. Волновая оптика** / Под ред. А.С. Кингсепа. — 2-е изд., испр. — М.: ФИЗМАТЛИТ, 2007. — 704 с. — ISBN 978-5-9221-0753-2.

Курс общей физики подготовлен в соответствии с программой бакалавриата по техническим специальностям и является победителем конкурса Министерства образования Российской Федерации. Курс адресован студентам технических университетов с углубленным изучением физики, а также студентам физико-математических факультетов классических университетов. Изложение ведется на современном уровне при достаточно высокой степени формализации, но математической подготовки, выходящей за рамки технического университета, у читателя не требуется — все необходимые дополнительные сведения включены непосредственно в данный курс.

Рекомендовано Министерством образования Российской Федерации в качестве учебника для студентов высших учебных заведений.

Учебное издание

КИНГСЕП Александр Сергеевич
ЛОКШИН Геннадий Рафаилович
ОЛЬХОВ Олег Алексеевич

КУРС ОБЩЕЙ ФИЗИКИ. ОСНОВЫ ФИЗИКИ

Том I

**МЕХАНИКА. ЭЛЕКТРИЧЕСТВО И МАГНЕТИЗМ.
КОЛЕБАНИЯ И ВОЛНЫ. ВОЛНОВАЯ ОПТИКА**

Редакторы *Артоболевская Е.С., Миртова Д.А.*

Оригинал-макет: *Затекин В.В.*

Оформление переплета: *Гришина Н.В.*

Подписано в печать 17.04.07. Формат 60х90/16. Бумага офсетная. Печать офсетная.

Усл. печ. л. 44,0. Уч.-изд. л. 48,8. Тираж 2000 экз. 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 ОАО «Чебоксарская типография № 1»

428019, г. Чебоксары, пр. И. Яковлева, 15

ISBN 978-5-9221-0753-2



9 7859 2 107532

ISBN 978-5-9221-0753-2

© ФИЗМАТЛИТ, 2001, 2007

ОГЛАВЛЕНИЕ

Предисловие	10
-----------------------	----

Раздел первый

МЕХАНИКА

Введение	13
Глава 1. Пространство. Время. Движение	15
1.1. Механика и математика. Научный метод познания	15
1.2. Пространство и время	16
1.3. Системы отсчета. Радиус-вектор движущейся точки	21
1.4. Частицы и поля. Классическая механика Ньютона	24
Вопросы	27
Глава 2. Введение в кинематику	28
2.1. Кинематика материальной точки	28
2.2. Равномерное движение по окружности	31
2.3. Приближение абсолютно твердого тела и приближение материальной точки	35
2.4. Преобразования Галилея и закон сложения скоростей	38
Вопросы и задачи	40
Глава 3. Законы Ньютона	41
3.1. Закон инерции	41
3.2. Второй закон Ньютона	43
3.3. Выбор единиц измерения и системы единиц. Размерности физических величин	46
3.4. Понятие импульса. Третий закон Ньютона	49
Вопросы	53
Глава 4. Примеры приложений законов Ньютона	54
4.1. Исследование закона движения материальной точки	54
4.2. Движение материальной точки под действием постоянной силы	57
4.3. Реактивное движение	62
4.4. Колебательное движение: гармонические колебания, резонанс	65
Вопросы и задачи	71
Глава 5. Работа и энергия. Закон сохранения энергии при движении материальной точки во внешнем силовом поле	73
5.1. Работа и кинетическая энергия	73
5.2. Потенциальная энергия материальной точки во внешнем силовом поле. Закон сохранения энергии	78
5.3. О законе сохранения энергии и непотенциальных силах	84

5.4. Простые примеры	88
5.5. Равновесие и устойчивость	91
Вопросы и задачи	94
Глава 6. Замкнутая система тел. Энергия взаимодействия и внутренняя энергия	95
6.1. Особенности движения замкнутой системы из двух взаимодействующих материальных точек. Приведенная масса	95
6.2. Центр масс системы материальных точек.	98
6.3. Потенциальная энергия взаимодействия. Закон сохранения энергии для замкнутой системы материальных точек.	101
6.4. Закон всемирного тяготения	106
6.5. Упругие и неупругие соударения	114
Вопросы и задачи	118
Глава 7. Уравнение моментов. Динамика твердого тела . . .	120
7.1. Момент импульса и момент силы	120
7.2. Законы Кеплера	123
7.3. Вращение абсолютно твердого тела вокруг неподвижной оси	126
7.4. Следствия уравнения моментов	133
7.5. Трехмерное движение твердого тела. Гироскопы	136
7.6. Плоское движение твердого тела.	140
Вопросы и задачи	142
Глава 8. Элементы механики сплошных сред	144
8.1. Упругие деформации. Закон Гука.	144
8.2. Сдвиг и кручение	148
8.3. Течение идеальной жидкости. Уравнение непрерывности . .	150
8.4. Архимедова сила. Уравнение Бернулли	153
8.5. Вязкость. Течение Пуазейля	157
8.6. Турбулентность	161
Вопросы и задачи	163
Глава 9. Законы механики в неинерциальных системах отсчета	165
9.1. Принцип относительности Галилея	165
9.2. Законы механики в неинерциальных системах отсчета. Силы инерции.	167
Вопросы и задачи	175
Глава 10. Введение в релятивистскую механику	177
10.1. Постоянство скорости света для всех систем отсчета. Принцип относительности Эйнштейна и преобразования Лоренца	177
10.2. Следствия из преобразований Лоренца. Сокращение длины и замедление времени	182
10.3. Импульс и энергия в релятивистской механике	186
Вопросы и задачи	189

Раздел второй

ЭЛЕКТРИЧЕСТВО И МАГНЕТИЗМ

Введение	191
Глава 1. Электрическое поле в вакууме	195
1.1. Электрический заряд	195
1.2. Электрическое поле	197
1.3. Теорема Гаусса	199
1.4. Потенциал. Понятие электрической емкости	202
1.5. Уравнения Лапласа и Пуассона	205
1.6. Электрический диполь	207
Задачи	209
Глава 2. Электрическое поле в веществе	211
2.1. Проводники в электрическом поле	211
2.2. Поляризация диэлектриков. Понятие электрической индукции	213
2.3. Граничные условия в электростатике	217
2.4. Диэлектрики с квазиупругими и жесткими диполями	219
2.5. Пьезоэлектрики и сегнетоэлектрики	221
2.6. Энергия электрического поля в вакууме и в веществе	224
Задачи	229
Глава 3. Постоянный ток	230
3.1. Ток как движение зарядов	230
3.2. Закон сохранения заряда. Постоянный ток	233
3.3. Закон Ома. Закон Джоуля–Ленца	235
3.4. Электродвижущая сила. Правила Кирхгофа	239
Задачи	242
Глава 4. Магнитное поле тока	244
4.1. Взаимодействие проводников с током. Понятие о магнитном поле	244
4.2. Магнитные поля простейших токовых конфигураций. Векторные свойства магнитного поля	248
4.3. Вектор-потенциал	253
4.4. Магнитный диполь. Понятие о магнитном моменте	255
Задачи	260
Глава 5. Магнитное поле в веществе	262
5.1. Микро- и макроскопическое описание поля в веществе	262
5.2. Диа- и парамагнетизм	265
5.3. Ферромагнетизм	268
5.4. Граничные условия на поверхности раздела	272
5.5. Магнитные цепи	275
5.6. Эффект Холла	277
Задачи	278

Глава 6. Электромагнитная индукция	280
6.1. Индуктивность и взаимная индукция	280
6.2. Принцип Ленца. Закон Фарадея	282
6.3. Интегральная и локальная форма закона электромагнитной индукции	285
Задачи	286
Глава 7. Энергия и силы в магнитном поле	288
7.1. Проводники в магнитном поле. Магнитное давление	288
7.2. Диполь в магнитном поле	290
7.3. Энергия магнитного поля	292
7.4. Подъемная сила электромагнита	298
Задачи	300
Глава 8. Квазистационарное электромагнитное поле.	302
8.1. Условие квазистационарности электрической цепи. Релаксационные процессы.	302
8.2. Колебательный контур.	305
8.3. Вынужденные колебания. Переменный ток	309
8.4. Скин-эффект.	315
Задачи	318
Глава 9. Уравнения Максвелла	321
9.1. Ток смещения. Обобщение теоремы о циркуляции	321
9.2. Уравнения электромагнитного поля	323
9.3. Теорема Пойнтинга	325
9.4. Импульс электромагнитного поля	329
Задачи	331
Глава 10. Электромагнитные волны	333
10.1. Вопросы волновой динамики. Волновое уравнение	333
10.2. Плоская монохроматическая волна	336
10.3. Электромагнитные волны в вакууме. Скорость света	338
10.4. Энергия и импульс электромагнитной волны	341
10.5. Излучение электромагнитных волн	345
Задачи	349
Глава 11. Электромагнитные волны в прозрачных средах.	351
11.1. Распространение волн в сплошной среде. Показатель преломления	351
11.2. Отражение и преломление на плоской границе	353
11.3. Формулы Френеля.	356
11.4. Поляризационные эффекты. Поток энергии через границу	359
11.5. Электромагнитные волны и теория относительности	361
Задачи	364
Глава 12. Элементы физики плазмы.	366
12.1. Газовый разряд	366
12.2. Понятие плазмы	371

12.3. Ленгмюровские колебания и дебаевское экранирование	375
12.4. Электромагнитные волны в плазме	377
Задачи	381

Раздел третий

ФИЗИКА КОЛЕБАНИЙ И ВОЛН.

ВОЛНОВАЯ ОПТИКА

Введение.	383
Глава 1. Кинематика колебаний	384
1.1. Гармонические колебания	384
1.2. Векторная интерпретация и комплексное представление гармонических колебаний. Фазовая плоскость	386
1.3. Некоторые важные задачи сложения гармонических колебаний.	390
1.4. Модулированные колебания.	395
1.5. Спектральное разложение	399
1.6. Векторные колебания. Фигуры Лиссажу	411
Задачи	414
Глава 2. Колебания в линейных системах	416
2.1. Примеры простейших колебательных систем. Общность уравнений, описывающих колебания различной физической природы	416
2.2. Свободные колебания гармонического осциллятора	420
2.3. Превращения энергии при свободных колебаниях гармонического осциллятора	421
2.4. Затухающие колебания гармонического осциллятора	425
2.5. Связанные осцилляторы	434
2.6. Вынужденные колебания гармонического осциллятора (гармоническая внешняя сила)	438
2.7. Гармонические колебания в линейных стационарных системах.	445
2.8. Спектральный анализ линейных колебательных систем	448
Задачи	457
Глава 3. Параметрические колебания. Ангармонический осциллятор. Автоколебания	458
3.1. Параметрические колебания	458
3.2. Ангармонический осциллятор	461
3.3. Автоколебания	464
3.4. Автогенератор Ван-дер-Поля	466
Глава 4. Кинематика волн	474
4.1. Основные понятия и определения. Простейшие типы волн. Волновое уравнение	474

4.2. Монохроматические волны. Комплексная амплитуда. Уравнение Гельмгольца	482
4.3. Векторные волны	484
Задачи	486
Глава 5. Упругие волны	487
5.1. Продольные упругие волны в твердом теле	487
5.2. Упругие волны в жидкостях и газах	490
5.3. Плотность и поток энергии в упругой волне. Вектор Умова	492
5.4. Стержень, закрепленный на концах. Собственные моды колебаний	496
5.5. Поведение звука на границе раздела двух сред	499
5.6. Поперечные волны в струне	500
5.7. Общие выводы	502
5.8. Эффект Доплера	504
Задачи	506
Глава 6. Электромагнитные волны	508
6.1. Уравнения Максвелла и волновое уравнение	508
6.2. Поляризация электромагнитных волн	513
6.3. Энергетические характеристики электромагнитных волн. Вектор Пойнтинга	514
6.4. Стоячая электромагнитная волна	517
6.5. Излучение колеблющегося диполя	518
6.6. Отражение электромагнитной волны от идеального проводника	522
6.7. Волноводы	524
6.8. Электромагнитная волна на границе раздела двух диэлектриков	527
Глава 7. Интерференция волн	529
7.1. Принцип суперпозиции волн	529
7.2. Интерференция монохроматических волн	530
7.3. Квазимонохроматические волны	535
7.4. Функция когерентности	541
7.5. Интерференция квазимонохроматических волн	544
7.6. Связь функции когерентности и распределения энергии по спектру. Соотношение неопределенностей	549
7.7. Пространственная когерентность и интерференционные явления при использовании протяженных источников	553
7.8. Интерферометры и интерферометрия	559
Вопросы и задачи	562
Глава 8. Дифракция	563
8.1. Граничные условия Кирхгофа	565
8.2. Спектральный метод решения задачи дифракции (метод Рэлея)	569

8.3. Дифракция на периодических структурах (эффект Талбота)	576
8.4. Область геометрической оптики.	578
8.5. Дифракция Фраунгофера.	580
8.6. Принцип Гюйгенса–Френеля	583
8.7. Дифракция Френеля. Дифракционные задачи с осевой симметрией.	585
8.8. Дифракция Френеля на одномерных структурах	592
8.9. Дифракция Фраунгофера.	595
8.10. Разрешающая способность спектральных приборов	603
8.11. Оптическое изображение и пространственная фильтрация. Разрешающая способность оптических систем	611
8.12. Принципы голографии.	623
8.13. Разрешающая способность голограммы	629
8.14. Схема с наклонным опорным пучком.	630
8.15. Объемная голограмма	631
8.16. Чем отличается голографическое изображение от фотографии.	635
Вопросы и задачи	636
Глава 9. Дисперсия волн.	638
9.1. Фазовая и групповая скорость. Формула Рэлея	638
9.2. Дисперсия электромагнитных волн	644
Вопросы и задачи	651
Глава 10. Волны в анизотропных средах. Элементы кристаллооптики.	653
10.1. Модель анизотропной среды	653
10.2. Волны в одноосных кристаллах	655
10.3. Преломление на границе анизотропной среды, двойное лучепреломление	660
10.4. Дихроизм. Поляризаторы.	663
10.5. Электрооптические и магнитооптические эффекты.	664
Вопросы и задачи	670
Глава 11. Элементы нелинейной оптики.	672
11.1. Модель нелинейной среды.	672
11.2. Эффект удвоения частоты. Оптическое выпрямление	673
11.3. Генерация третьей гармоники. Самофокусировка и самоканализация	680
11.4. Параметрические процессы в нелинейных средах.	684
11.5. Комбинационное рассеяние света.	687
11.6. Вынужденное комбинационное рассеяние	689
11.7. Вынужденное рассеяние Мандельштама–Бриллюэна (ВРМБ)	691
11.8. Обращение волнового фронта	693
Именной указатель	697
Предметный указатель	698

Предисловие

То что мы сейчас понимаем под словом «физика», сформировалось как наука совсем недавно, примерно 200 лет тому назад. До этого все ученые назывались «естествоиспытателями», а в древние времена просто «философами». По-гречески «фило» — значит «любить», а «софия» — мудрость. Поэтому философами называли людей, любивших мыслить. Конечно, человек всегда задумывался над тем, как произошел мир, что лежит «в основе вещей», почему планеты движутся по небосклону, камень скатывается с горы в ущелье и т. п. Инстинктивно, начиная с глубокой древности, наши предки пытались обобщить свои наблюдения над окружающим миром таким образом, чтобы придать им предсказательную силу, пытаясь определить заранее, когда наступит засуха, произойдет затмение Солнца или наводнение. Со временем багаж познаний древних естествоиспытателей пополнялся, одновременно расширялся и круг вопросов. Сегодня под естественными науками мы понимаем те области знания, в которых может быть проведен эксперимент, т. е. проверка тех предположений, моделей, которые следуют из вновь выдвинутых или хорошо известных теорий, из анализа проведенных опытов. Тем самым, естественные науки существенно расширили свою методологию по сравнению с философией, превратившись из чисто созерцательных в экспериментальные.

Физика — одна из естественно-научных дисциплин, точное определение которой дать очень трудно, а может быть, и невозможно. Физика тесно переплетена со многими другими дисциплинами, которые зачастую выделялись в самостоятельные области исследований. Не всегда легко провести грань между собственно физикой и механикой, между атомной физикой и химией; существуют также и пограничные науки (например, биофизика, химическая физика), активно использующие физические методы для решения задач другой области знания либо технических приложений. Впрочем, одну отличительную черту физики можно указать точно: физика — это наука об основополагающих закономерностях, определяющих процессы и явления в природе. Ее основная задача — выявлять фундаментальные черты нашего мира, предоставляя другим наукам основы знаний, открывая новые возможности инженерных решений, прорыв в новые технологии. Физика тесно связана с развитием нашего понимания природы, фактически она определяет современное естественно-научное мировоззрение.

Как и все естественные науки, физика — наука экспериментальная, но, как наука фундаментальная, она более всех формализована, в ней чрезвычайно важную роль играет теория. Теория в физике — не только обобщение массива экспериментальной информации, но и метод исследования, связанный с разработкой предположений и постановкой экспериментальных задач. Но надо четко представлять себе, что теоретические выводы признаются в физике фактом лишь после их полного экспериментального подтверждения, оставаясь до того лишь гипотезами.

Трудно переоценить и ту огромную роль в развитии человеческой цивилизации, которую играли и играют, казалось бы, чисто научные успехи физики в понимании законов природы. Достаточно вспомнить, что развитие термодинамики оказало существенное влияние на создание и совершенствование тепловых машин, создание электродинамики уже напрямую предопределило появление электродвигателя, электрического освещения, телефона и телеграфа, радиосвязи и телевидения. Понимание процессов, происходящих в полупроводниках, обусловило грандиозные успехи электроники и, в частности, вычислительной техники (а значит, в немалой степени, современной информатики). Создание лазеров революционизировало всю оптику и все большее влияние оказывает на электронику. Достижения ядерной физики XX в. послужили основой для ядерной энергетики; следует отметить, что во многих европейских странах более половины энергии производится на АЭС.

Помимо физиков-профессионалов с физикой соприкасаются так или иначе специалисты самых различных профессий. Химик, использующий эффект Мёссбауэра; биолог, изучающий строение вируса с помощью электронного микроскопа; археолог, применяющий метод радиоактивного датирования для определения возраста найденных им предметов быта древнего человека; металлург, определяющий по спектру излучения температуру расплавленного металла; инженер, проектирующий современные здания либо бытовую технику, — все они, быть может, непосредственно не ощущают необходимости понять, что такое волновая функция электрона или энтропия, однако их работа базируется на достижениях естественных наук, и прежде всего физики. Поэтому физика является неотъемлемой частью профессионального образования в большинстве областей человеческой деятельности.

Традиционно физика делится на, казалось бы, очень разные, независимые друг от друга разделы — механику, термодинамику, оптику, электричество, атомную физику. Такой подход позволяет, с одной стороны, выделить из всего многообразия явлений природы тот круг вопросов, который можно описать с единой точки зрения, на основе общих для них законов. С другой стороны, все окружающие нас явления взаимосвязаны, и успехи науки во многом обязаны именно пониманию единых закономерностей природы. Природа ведь не знает, на какие главы и параграфы мы разделили ее изучение. Эту диалектику в изучении явлений природы мы старались по мере возможности воспроизвести в нашей книге.

Учебник написан в соответствии с программой бакалавриата по физике, утвержденной Министерством общего и профессионального образования РФ для технических направлений и специальностей высших учебных заведений.

Учебник назван нами «Основы физики», потому что авторы старались изложить основополагающие современные экспериментальные факты и их теоретические обоснования, идеи и принципы, которые прежде всего должны помочь молодому человеку понять окружающий мир и мощь человеческого познания природы. Кроме того, и может быть, это главное, эти знания должны послужить основой для дальнейшего изучения либо прикладных дисциплин, либо курсов, конкретизирующих общие положения физической науки. Фундамент

физических знаний должен помочь будущему инженеру разбираться в постоянно возникающих новых технологиях, материалах, приборах и методах измерений.

Современный курс физики никоим образом не может сводиться только к учебнику и лекционному курсу. Не меньшая роль принадлежит лабораторным работам и семинарским занятиям, а также заданиям для самостоятельной работы. Чтобы помочь студенту в самостоятельной работе, мы дополнили практически каждую главу вопросами и задачами, часть задач приводится с решениями, в некоторых даны только указания к решению, а в остальных приводятся только ответы.

Мы старались написать этот учебник так, чтобы при его чтении как можно реже нужно было обращаться к пособиям по математике — математические понятия и приемы, которые выходят за рамки стандартного курса, объясняются в тех местах, где изложение материала требует обращения к высшей математике. Конечно, мы предполагаем у нашего студента знание школьного курса математики, он должен также уметь дифференцировать и интегрировать простейшие математические выражения, владеть экспоненциальной формой записи комплексных чисел и уметь решать простые дифференциальные уравнения.

Любой учебник предназначен не только для студентов, но и для преподавателей. Мы надеемся, что преподаватели найдут в этом учебнике много полезного для своей работы, выделяя те моменты, которые принципиально важны, либо, напротив, опустив несущественные для конкретного вуза, конкретной специальности.

Учебник написан профессорами кафедры общей физики Московского физико-технического института и состоит из двух томов. В первый том, общим редактором которого является А.С. Кингсеп, вошли следующие разделы: «Механика» (автор О.А. Ольхов при участии В.Е. Белонучкина и А.С. Кингсепа), «Электричество и магнетизм» (автор А.С. Кингсеп) и «Колебания и волны» (автор Г.Р. Локшин). Во второй том, редактором которого является Ю.М. Ципенюк, включены «Квантовая физика атома, ядра и элементарных частиц» (авторы Д.А. Заикин и Ю.М. Ципенюк) и «Статистическая физика и термодинамика» (авторы В.Е. Белонучкин и Ю.М. Ципенюк).

Мы глубоко признательны нашим редакторам Е.С. Артоболевской, О.А. Пениной, Н.С. Рытовой, О.В. Савельяновой и Д.А. Миртовой, доброжелательные критические замечания которых и полезные советы способствовали существенному улучшению содержания учебника.

Мы далеки от мысли, что данный учебник лишен недостатков, и заранее благодарны нашим коллегам и читателям за любые замечания.

Авторы

Раздел первый

МЕХАНИКА

ВВЕДЕНИЕ

Открывая курс общей физики, полезно еще раз отметить то обстоятельство, что физика — наука естественная, а следовательно — экспериментальная. Среди естественных наук физика — в силу фундаментальности объектов исследования и их свойств — наиболее формализована. Поэтому ее первичное изучение нередко порождает иллюзию «выводимости» или аксиоматичности физических законов. На самом деле все основные сведения в естественных науках следуют из эксперимента, и им же проверяются в конечном счете любые теоретические модели.

Механика может рассматриваться как наиболее фундаментальная естественная наука. С нее обычно и начинается изучение всего курса общей физики. Предметом механики являются законы движения и равновесия тел при известном и/или заданном взаимодействии между ними. (Подчеркнем еще раз, что аксиоматичность есть свойство математики, но не естественных наук, поэтому наши начальные определения с неизбежностью будут не полны, полуинтуитивны). Механика изучает перемещение тел в пространстве; а любое изменение, развитие, взаимодействие предполагает и перемещение чего-либо, будь то галактики, звезды и планеты, машины и механизмы, атомы и молекулы, клетки живых организмов и сами живые организмы, электромагнитные и звуковые сигналы. По этой причине механические представления о движении пронизывают все естественные науки, в том числе и физику. Кроме того, решая задачи повседневной жизни и наблюдая все то, что нас окружает, мы с первых шагов нашей жизни оказываемся непрестанно вовлеченными в мир бесчисленных перемещений и механических воздействий.

Но есть и другая, не менее важная причина, почему механика является стартовым курсом при изучении физики. Дело в том, что история становления механики как науки в современном понимании — это есть одновременно и история развития и становления того, что принято называть «современным научным подходом». Сейчас трудно представить, что первые (не слишком успешные) попытки четко сформулировать правила движения тел при различных условиях предпринимались уже более двух с половиной тысяч лет тому назад в Греции, в знаменитой

школе-лицее, руководимой выдающимся мыслителем древности Аристотелем. Но как отличить ошибочное правило от истинного, и что вообще понимать под истинными законами движения или каких-либо других явлений природы? Чтобы найти ответы на эти естественные вопросы, потребовалось более двух тысячелетий напряженной работы бесчисленной армии исследователей в различных областях знания, пока не были выработаны общие принципы установления, формулировки и проверки законов, описывающих наблюдаемые явления природы, и именно эти принципы лежат в основе того, что называется современным научным мировоззрением. Именно при изучении законов механики из-за ее непростой и долгой истории и блестящих результатов можно одновременно осваивать и основные элементы современного метода познания любых явлений природы, понимать приближенный характер наших знаний о природе, представлять место и взаимосвязь теории и эксперимента и, наконец, даже грамотно вести спор на профессиональную тему. Все это не менее важно, чем знание законов движения и умение решать задачи из задачника, так как понимание логики научного мышления окажется неоценимым подспорьем и при изучении других наук, и при овладении любой новой профессией, да и при решении многих проблем повседневной жизни.

В соответствии с международным стандартом, мы отдаем в нашем курсе предпочтение Международной системе единиц (СИ); впрочем, в механике, в отличие от электродинамики или атомной физики, никакого различия в формализме из предпочтения той или иной традиционной системы единиц не следует. Необходимые сведения из курса математики, существенно выходящие за пределы школьного курса, будут даны по мере надобности, но за доказательствами соответствующих теорем нам придется отсылать любознательного читателя к специальной литературе.

Глава 1

ПРОСТРАНСТВО. ВРЕМЯ. ДВИЖЕНИЕ

1.1. Механика и математика. Научный метод познания

Как уже было отмечено, первые попытки установления объективных законов, управляющих движением, предпринимались еще в античные времена. Но, благодаря создателю механики Исааку Ньютону (1643–1727 гг.), анализ опытных данных, введение новых физических понятий и установление соответствия между ними были впервые поставлены на строгое математическое основание, и математика стала единым языком науки и могучим средством, помогающим использовать ее достижения для практических целей человеческого сообщества.

Как язык науки математика играет двоякую роль. Во-первых, точность математических формулировок позволяет определять физические величины универсальным способом, когда задание конкретного значения этой величины любым специалистом, находящимся в любом месте, будет правильно и однозначно понято любым другим специалистом. Например, если через научный журнал сообщается, что измеренная на опыте скорость какого-либо объекта была равна одному метру в секунду, то эта информация будет одинаково и однозначно понята всеми заинтересованными читателями только потому, что существует строгое математическое определение скорости $v = dr/dt$, точно указывающее, что именно означают слова «один метр в секунду». И мы далее увидим, что именно проблема строгого, объективного определения понятия скорости движения тела дала толчок к развитию высшей математики.

Во-вторых, формулировка наблюдаемых взаимосвязей каких-либо явлений с помощью математических соотношений исключает их неверное, неоднозначное истолкование, которое, однако, неизбежно, если ту же взаимосвязь явлений пытаться выразить, не прибегая к математическому языку, а просто «на словах». Например, Аристотель так формулировал на словах основной закон движения: *для того чтобы поддерживать движение тела с постоянной скоростью, на него необходимо воздействовать некоторой силой*. С точки зрения повседневного опыта, это утверждение кажется почти очевидным: действительно, чтобы передвигать, например, какой-либо предмет по горизонтальной

поверхности, необходимо прилагать определенное усилие. Тем не менее, с точки зрения современной механики, это утверждение ошибочно, и эта ошибка, обусловленная невозможностью в те времена строго математически определить понятие «сила», продержалась в умах философов и ученых более двух тысяч лет (в частности, благодаря огромному авторитету Аристотеля среди ученых древности). Эта ошибка была исправлена только после того, как Ньютон дал современное, знакомое нам по школе, определение силы: $\mathbf{F} = m\mathbf{a}$. Из него, в частности, следует, что для движения тела с постоянной скоростью (когда ускорение \mathbf{a} равно нулю) внешней силы не требуется, а повседневный опыт свидетельствует в этом случае о равновесии сил, приложенных к телу.

В сравнении с другими естественными науками, физика и, в частности, механика в наибольшей степени допускают количественное описание и наиболее формализованы. Но тем более важно подчеркнуть, что, как и во всех естественных науках, базовая информация при этом следует исключительно из эксперимента. Равным образом не нужно забывать, что основные понятия в естественных науках, включая механику, отнюдь не аксиоматичны — при любом уровне формализации — но являются следствием человеческого опыта и отражают достигнутое на сегодняшний день понимание материального мира. Практически количественные соотношения между ними определяются по имеющимся экспериментальным данным.

Таким образом, математика дает лишь один из инструментов исследования природы, но не может подменять собою естественные науки. (Непонимание этого обстоятельства является источником многих «школьных» ошибок — и не только у школьников). Но есть еще одно принципиальное свойство естественных наук, которое выходит за рамки математических моделей — *принципиальная неточность, приближительность любого количественного описания*. Ни одна формула в механике, физике, химии и т. д. не может считаться абсолютно точной, она справедлива лишь в пределах границ ее применимости, указанных или подразумеваемых. И ни одна экспериментально определенная величина не имеет смысла, если не указано, в каких условиях и с какой точностью она измерена.

1.2. Пространство и время

Попытаемся определить, что мы понимаем под словом «движение». Хотя это понятие и представляется совершенно очевидным, нам необходимо сформулировать такое его определение, которое содержало бы в себе указание, во-первых, на способ из-

мерения различных свойств возможных движений и, во-вторых, давало бы возможность выразить эти измеряемые свойства на общепринятом научном языке, то есть с помощью математических знаков и формул. Достаточно популярным является следующее определение: движение — это изменение относительного положения тел в пространстве с течением времени. В таком утверждении как бы заранее подразумевается, что понятия «пространство» и «время» являются совершенно естественными и не нуждающимися в каком-либо специальном, формальном определении.

В действительности оказывается, что сами эти понятия могут быть определены лишь через материальные предметы и события, с ними происходящие. Можно сказать, что движение есть перемещение рассматриваемого тела относительно других тел. Нет других тел — нет и движения. Таким образом, пространство задается расположением материальных объектов. Время, в свою очередь, воспринимается через последовательность событий. Если ничего не происходит, нет и течения времени. Зададимся вопросом: что было за секунду до Большого взрыва, с которого, по современным понятиям, началась эволюция нашего мира? Ответ: если наши представления об эволюции Вселенной правильны, то этот вопрос не имеет смысла, поскольку не было самого этого *до*.

В отличие от пространства, которое для нас представимо хотя бы какими-то зрительными ассоциациями, время — понятие более абстрактное, основанное более всего на нашем жизненном опыте, который, в частности, свидетельствует о таком его фундаментальном свойстве, как однонаправленность. На нем базируются понятия причинно-следственных связей.

Эта парадоксальная ситуация не приводит, однако, к каким-либо практическим затруднениям. Дело в том, что хотя мы и не можем сформулировать логически безупречного определения пространства или времени, мы, тем не менее, можем сделать самое главное, что необходимо для описания и изучения любого явления, а именно, мы можем указать способ измерения его свойств и способ представления результатов измерений на математическом языке. Для этой цели введем следующее определение: *измерение какой-либо величины — это ее сравнение с однородной величиной, условно принятой за эталонную единицу измерения.*

Тогда, если мы, например, представляем себе пространство как нечто, отделяющее одно тело от другого, то мы можем измерить это «нечто» с помощью твердого однородного прямолинейного стержня, который договоримся считать эталонной мерой расстояния между двумя любыми телами в пространстве. Измерив, сколько раз уложится наш эталон (или доли этого эталона) на кратчайшем расстоянии между нашими телами, мы

выразим это расстояние, или, как еще говорят, длину, в виде определенного числа. Это число и будет выражать интересующее нас расстояние в «единицах длины», определяемых выбором эталонного стержня.

В течение долгого времени эталоном длины, который получил название *метр*, служило определенное расстояние между двумя штрихами, нанесенными на стержне особой формы, изготовленном из сплава платины и иридия и находившемся в Международном бюро мер и весов во Франции. Точные копии этого эталона имелись и в других странах. Сейчас, однако, отказались от такого эталона, так как он слишком чувствителен к изменению внешних условий и уже не удовлетворяет возросшим требованиям современной науки и техники. В настоящее время единица длины метр определяется как длина пути, проходимого светом в вакууме за время $1/299\,792\,458$ с.

Имея в своем распоряжении способ измерения пространственных интервалов между телами, можно на этой основе перейти к обсуждению тех свойств пространства, которые могут быть исследованы с помощью указанных измерений. Таких свойств два: *трехмерность* и *евклидовость*. Начнем по порядку с того свойства пространства, которое на словах выражается утверждением, что «наше пространство трехмерно». В повседневной жизни мы связываем это свойство с тем очевидным и привычным фактом, что всякий предмет (а следовательно, и занимаемый предметом элемент пространства) можно с той или иной точностью охарактеризовать тремя параметрами: «высотой», «длиной» и «шириной». Более строго, трехмерность пространства определяется тем, что для однозначного определения положения одной точки пространства A относительно другой точки B необходимо в общем случае задать три пространственных

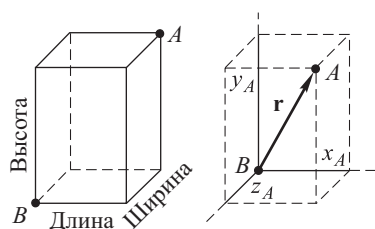


Рис. 1.1

интервала (рис. 1.1). В частном случае, когда с точкой B совпадает начало прямоугольной системы координат, вышеупомянутые три пространственных интервала совпадают с соответствующими координатами x_A , y_A , z_A точки A .

Теперь — о втором свойстве пространства, которое также определяется опытным путем и

формулируется в виде утверждения о том, что «наше пространство является евклидовым пространством», то есть в нем выполняются все теоремы известной нам евклидовой геометрии, такие, например, как теорема Пифагора или теорема о том, что сумма углов треугольника равна 180° . Как и трехмерность, это свойство кажется очевидным и не требующим

для его подтверждения никаких измерений. Однако оно станет не столь очевидным, если мы будем рассматривать возможные геометрические свойства не трехмерного пространства, а более простого для восприятия двумерного пространства, примером которого является, например, поверхность стола, где положение любого предмета определяется всего двумя координатами (рис. 1.2). На плоской поверхности стола выполняются все теоремы евклидовой геометрии, и поэтому такое двумерное пространство является евклидовым двумерным пространством.



Рис. 1.2

Но дело в том, что двумерное пространство может быть не только плоским, но и «кривым», как, например, поверхность шара. На поверхности шара теоремы евклидовой геометрии перестают быть справедливыми. Это свойство «неевклидовости» может быть установлено путем прямых измерений.

Но если двумерное пространство может быть кривым (неевклидовым), то откуда мы знаем, что наше трехмерное пространство не является таковым? Кривизну двумерного пространства, кривизну поверхности шара мы воспринимаем, находясь в пространстве трех измерений. Точно так же искривление трехмерного пространства реализуется в четвертом измерении, которое мы, трехмерные существа, представить не в состоянии. Однако астрономические исследования свидетельствуют о том, что на доступных исследованию расстояниях (а это миллиарды свето-

вых лет) кривизна мирового пространства в среднем отсутствует. Имеются лишь указания на ничтожное локальное искривление пространства в непосредственной близости от звезд. Поэтому в рамках излагаемой в этой книге классической механики пространство считается евклидовым (не «искривленным»), то есть таким, в котором всюду строго выполняются все теоремы привычной для нас евклидовой геометрии.

Итак, наше пространство *трехмерно* и *евклидово*. Характерные пространственные масштабы, с которыми приходится сталкиваться при исследовании природы, приведены в табл. 1.

Таблица 1.1

Пространственно-временные масштабы окружающего мира

Пространство (масштаб)	Время
Между звездами 10^{16} м (световой год)	Номо sapiens 10^{12} с (100000 лет)
Солнечная система 10^{11} м	Оборот Земли вокруг Солнца 10^7 с (один год)
Земля 10^6 м	Оборот Земли вокруг оси 10^4 с (одни сутки)
Человек 1 м	Один удар сердца 1 с
Живая клетка 10^{-6} м	Прохождение электромагнитного импульса по нервному волокну 10^{-1} с
Атом 10^{-10} м	Время одной операции ЭВМ 10^{-9} с
Атомное ядро 10^{-15} м	Время жизни некоторых элементарных частиц 10^{-23} с

Движение, как уже говорилось, происходит не только в пространстве, но и во времени. На примере с определением пространства мы уже поняли, что более важным для описания свойств какого-либо первичного понятия является не возможность его безупречного формального определения, а указание на возможность его *измерения* с помощью какого-то единого эталона. С доисторических времен в качестве меры времени было принято рассматривать какое-либо повторяющееся явление природы, так как число циклов, число повторов этого явления есть удобный и естественный способ измерения времени. В течение многих столетий с этой целью использовались хорошо известные песочные часы.

Как любопытный исторический факт, можно отметить, что Галилей при изучении законов движения тел использовал в каче-

стве часов биение собственного пульса. Развитие науки и практические нужды общества потребовали единого эталона времени, по которому можно было бы сверять ход любых часов; до недавнего времени таковым являлся период обращения Земли вокруг своей оси. Точнее, за эталонную единицу времени, называемую *секундой*, принималась $1/86\,400$ часть средних суток. Однако сейчас физики научились использовать периодические процессы, которые происходят в микромире и которые могут служить гораздо более точными эталонами времени, чем средний период вращения Земли вокруг своей оси. Конкретно, за эталон времени сейчас принят период колебаний электромагнитного поля, испускаемого атомами химического элемента цезия при определенном изменении их внутренней энергии. В течение одной секунды совершается $9\,192\,631\,770$ вышеуказанных эталонных колебаний.

Из табл. 1 можно получить представление о характерных временных интервалах, с которыми приходится иметь дело в житейской и научной практике.

Можно в принципе поставить вопрос о том, как связан ход часов с их собственным движением. На сегодняшний день установлено, что в движущихся друг относительно друга часах время течет по-разному, но это отличие можно обнаружить лишь при относительных скоростях, близких к скорости света, $c \approx 300\,000$ км/с. С такими скоростями физикам приходится иметь дело при изучении движения и превращений элементарных частиц, прилетающих к нам из космоса или разгоняемых до таких скоростей в специальных ускорителях. При $v \ll c$ указанное изменение хода часов ничтожно, что позволяет считать время *абсолютным*.

1.3. Системы отсчета. Радиус-вектор движущейся точки

Принципиально важным оказывается свойство *относительности* всякого движения. В самом деле, находящийся в нашей ладони и покоящийся относительно нее предмет может совершать при этом сложное движение вместе с рукой относительно нашего тела. А если мы не стоим на месте, этот же предмет будет совершать еще более запутанное движение относительно каких-либо окружающих нас вещей. И, наконец, вместе с нами тот же предмет будет участвовать в очень сложном движении, обусловленном вращением Земли вокруг своей оси и вокруг Солнца.

Любые попытки установить какие-либо закономерности движения естественно начинать с выбора той совокупности тел,

относительно которых представляется важным или удобным рассматривать движение интересующего нас объекта, и такую совокупность тел принято называть *системой отсчета*. В качестве системы отсчета можно себе представить, например, три взаимно перпендикулярных недеформируемых прямолинейных стержня, пересекающихся в одной точке и жестко связанных в этой точке друг с другом. Главное свойство этой системы отсчета состоит в том, что в каком бы месте она ни находилась, положение любой точки интересующего нас тела однозначно определяется в ней тремя расстояниями вдоль стержней-линий, отсчитанными от точки их пересечения. Эти расстояния принято обозначать буквами x , y , z , и они образуют хорошо известную *декартову систему координат* (см. рис. 1.1). Существуют и другие системы координат, выбор которых определяется соображениями целесообразности и удобства: полярные, сферические, параболические и т. д. Например, для определения положения объектов у поверхности Земли часто используется «географическая» система, где вместо параметров x , y , z используются широта, долгота и высота над уровнем моря. Число независимых координат равно размерности пространства, т. е. в общем случае трем, но нередко физические явления удается смоделировать задачей меньшей размерности. Подчеркнем особо, что система координат — понятие математическое, тогда как система отсчета — физическая категория.

Итак, мы видим, что положение любой точки тела относительно выбранной системы отсчета однозначно определяется тройкой чисел x_A , y_A и z_A . С другой стороны, легко видеть, что вместо этой тройки чисел положение той же точки тела можно определить, задав одно число — длину отрезка, соединяющего точку с началом координат, и указав направление этого отрезка из начала координат в выбранную точку тела. Такая величина, количественная характеристика которой имеет не только числовое значение, но и направление, называется *вектором*. Строго говоря, не всякую величину, характеризующуюся модулем и направлением, можно назвать вектором. Принципиально важным оказывается правило сложения векторов, которое должно удовлетворять известным условиям:

$$\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}; \quad (\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c}); \quad \lambda(\mathbf{a} + \mathbf{b}) = \lambda\mathbf{a} + \lambda\mathbf{b}.$$

В трехмерном пространстве правила сложения векторов имеют вид, представленный на рис. 1.3. Из них, как мы вскоре увидим, прямо вытекает фундаментальный закон классической механики — преобразования Галилея. Алгебраическим операциям с векторами однозначно соответствуют точно такие же операции с их компонентами (проекциями):

$$(\mathbf{a} + \mathbf{b})_x = a_x + b_x; \quad (\mathbf{a} - \mathbf{b})_y = a_y - b_y \text{ и т. д.}$$

Существенность правил сложения иллюстрируется таким примером: величина, имеющая то же направление, что и вектор \mathbf{b} , но по модулю равная \sqrt{b} , этим правилам не удовлетворяет, а значит, и вектором именоваться не может. Истинный вектор должен к тому же вполне определенным образом вести себя при преобразовании системы координат. Наконец, для векторов определены известные операции

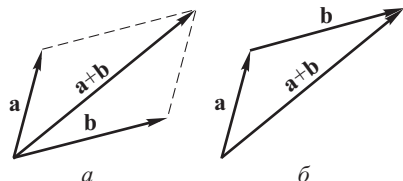


Рис. 1.3

скалярного произведения $(\mathbf{a} \cdot \mathbf{b})$ (или $\mathbf{a} \cdot \mathbf{b}$), векторного произведения $[\mathbf{a} \mathbf{b}]$ (или $\mathbf{a} \times \mathbf{b}$), смешанного произведения $\mathbf{c}[\mathbf{a} \mathbf{b}]$. Таким образом, утверждение о том, что некоторая величина есть вектор, будет достаточно сильным, и достаточно много говорящим о ее физических свойствах. Вектором (векторной величиной) определяется не только положение какой-либо точки в пространстве, но и многие другие важные величины, используемые для описания явлений природы, такие как, например, скорость и сила, напряженность электрического поля и т. д. Для сравнения заметим, что такие физические величины, как длина, масса, энергия, температура, вполне определяются заданием числового значения величины, и их свойства не связываются в нашем сознании с каким-либо направлением в пространстве (в отличие от векторных величин их называют скалярными величинами или *скалярами*). Что касается вектора, задающего положение любой точки движущегося тела в заранее выбранной системе координат, то такой вектор принято называть *радиусом-вектором* рассматриваемой точки, а декартовы координаты точки x_A, y_A, z_A являются проекциями этого вектора на соответствующие координатные оси, однозначно определяющими как длину, так и направление этого радиуса-вектора. Обычно радиус-вектор обозначают буквой \mathbf{r} , (либо той же буквой со стрелкой сверху \vec{r}). Длину любого вектора называют его *модулем* и обозначают обыкновенной, не жирной буквой (без стрелки сверху), так что модуль радиуса-вектора обозначается, как r . Модуль любого вектора — скаляр, по определению всегда положительный: $r \equiv \sqrt{(\mathbf{r} \mathbf{r})}$.

При движении тела его различные точки изменяют с течением времени свое положение в пространстве и, соответственно, изменяются со временем их радиусы-векторы, то есть радиус-вектор \mathbf{r} в общем случае является функцией времени t , $\mathbf{r} = \mathbf{r}(t)$, которая полностью определяется зависимостью от времени трех его проекций $x(t)$, $y(t)$ и $z(t)$ в заданной системе отсчета. Мы сделали первый важный шаг на пути привлечения математики для описания движения тел, показав, что положение любой

точки тела в пространстве определяется векторной величиной — радиусом-вектором $\mathbf{r}(t)$, который изменяется по величине и направлению вместе с изменением со временем положений точек движущегося тела. Это сразу позволяет нам объединить словесное, зачастую недостаточно четкое описание механических явлений со строгим, единым для всех языком математики. В частности, утверждение, что «движение тела это есть его перемещение в пространстве относительно других тел с течением времени», является, как мы видим, утверждением, что «движение тела — это изменение со временем радиусов-векторов всех точек рассматриваемого тела в выбранной системе отсчета».

Итак, мы можем сказать, что *основная задача механики состоит в нахождении математических соотношений, определяющих, как радиусы-векторы различных точек движущихся тел меняются со временем.*

Умение находить эту временную зависимость решает одну из главных практических задач — предсказать, как будут двигаться тела в зависимости от возможных начальных условий. Графически изменение со временем радиуса-вектора любой точки можно изобразить в виде линии, вдоль которой перемещается конец радиуса-вектора. Эта линия и является *траекторией* движения соответствующей точки.

1.4. Частицы и поля. Классическая механика Ньютона

Принципиальным для классической механики является понятие *материальной точки*. По сути вся механика строится на основании законов движения материальной точки, постепенно усложняясь и переходя к рассмотрению все более сложных объектов — и так вплоть до механики жидкости и газа.

Первоначальное понятие материальной точки базируется на системах координат, которые обсуждались в предыдущем параграфе. Это означает, что физическое тело рассматривается в качестве материальной точки, если для описания его движения достаточно временной зависимости трех координат или, что то же, одного радиуса-вектора $\mathbf{r}(t)$. Прежде всего, это соответствует телу пренебрежимо малых размеров — отсюда и термин. Но так же можно описывать перемещение тел совершенно произвольных размеров и формы, если они совершают поступательное движение, т.е. траектории всех точек тела идентичны и могут быть совмещены параллельным переносом. Наконец, и при более сложных движениях, когда существенны ориентационные координаты, движение центра масс тела, как мы увидим в последующих главах, подчиняется законам механики материальной точки.

Механика Ньютона называется классической потому, что ее становление произошло задолго до утверждения современных атомистических представлений о строении вещества. Как уже говорилось, любое количественное соотношение (а значит — и любая область физики, представляемая некоторой совокупностью количественных соотношений) имеет смысл лишь при указании границ применимости. Таким образом, в современном понимании термин «классическая механика» должен быть определен количественно, а именно: должны быть сформулированы условия, при которых ее можно использовать с заданной точностью и/или оценена максимальная возможная ошибка.

Во-первых, сложение движений в рамках классической механики сводится к сложению радиусов-векторов. Если материальная точка в некоторой системе отсчета перемещается по закону $\mathbf{r}_1(t)$, а сама система отсчета движется так, что ее начало координат в некоторой новой системе перемещается по закону $\mathbf{r}_2(t)$, то в этой новой системе отсчета материальная точка движется по закону $\mathbf{r}_1(t) + \mathbf{r}_2(t)$. Это эквивалентно преобразованиям Галилея, о которых мы будем говорить в следующем параграфе. Сколько естественно выглядит это правило, уже в XX в., почти через 200 лет после смерти Ньютона было установлено (прежде всего здесь следует назвать имя выдающегося физика-теоретика Альберта Эйнштейна (1879–1955)), что оно представляет собой лишь частный случай, приближение, отвечающее пределу нерелятивистских скоростей $v \ll c \approx 3 \cdot 10^8$ м/с. Механика тел, движущихся с относительными скоростями, сравнимыми со скоростью света c , выглядит заметно сложнее и в гораздо меньшей степени согласуется с нашим здравым смыслом, поскольку опыта с такими движениями человек в повседневной практике не имеет. Эта механика называется *релятивистской*; иногда ее по традиции называют *специальной теорией относительности*. Мы с ней познакомимся в одной из последующих глав данного раздела.

Во-вторых, существенной ревизии подвергается механика Ньютона при переходе к рассмотрению физики молекул, атомов и элементарных частиц. И это обстоятельство было понято в XX в. усилиями датского физика-теоретика Нильса Бора (1885–1962), немецкого физика-теоретика Вернера Гейзенберга (1901–1976) и многих других замечательных ученых. Механика микромира — квантовая механика — не может использовать понятие материальной точки, поскольку все частицы, как оказывается, обладают волновыми свойствами. Пренебречь этим обстоятельством позволительно лишь до тех пор, пока реальные размеры тел, рассматриваемых нами как материальные точки, или расстояния между телами, другими словами, пространственные

масштабы задачи, существенно превышают т. н. длину волны де Бройля (1892–1987):

$$\lambda_{\text{Б}} = \frac{h}{mv},$$

где m — масса тела, v — его скорость, а $h \approx 6,6 \cdot 10^{-34}$ Дж·с — постоянная Планка. Квантовая механика необходима уже для ответа на вопрос, что такое валентность с точки зрения физики или, например, что такое твердое тело и как оно устроено. С этой наукой наше знакомство состоится в заключительных разделах нашего курса.

Можно сказать, что механика Ньютона корректна в случае предельного перехода $h \rightarrow 0$ и одновременно $v/c \rightarrow 0$. Но этот предельный случай вмещает такое многообразие предметов и явлений и столь важен с практической точки зрения, что классическая механика и на сегодняшний день остается не только предметом утилитарного применения, но и живой, развивающейся областью естествознания.

Подавляющая часть объектов окружающего мира, в особенности доступных непосредственному человеческому восприятию (с помощью органов чувств), состоит из мельчайших частичек вещества — атомов. В дальнейшем мы будем называть их *частицами*, поскольку структура атомов и законы их «внутренней жизни» не могут быть предметом рассмотрения классической механики. Частицами называют также обычно и входящие в состав атомов электроны, атомные ядра, протоны, нейтроны и множество других, так называемых, элементарных частиц, о которых будет идти речь в соответствующем разделе курса общей физики. Наконец, физики используют понятие макрочастиц, состоящих из большого числа атомов. В рамках классической механики частицей можно называть любой физический объект, который с достаточной точностью описывается приближением материальной точки.

Но окружающий нас мир состоит не только из атомов. Так например, иначе устроены электромагнитные волны, пронизывающие окружающее нас пространство и почти непрестанно воздействующие на наши органы чувств либо как видимый свет, либо как радиоволны или телевизионные сигналы. (Более того, по сути все наши органы чувств, а равно и все наши измерительные приборы воспринимают окружающий мир через электромагнитное взаимодействие). Электромагнитные волны представляют собой электрические и магнитные *поля* — особый вид объектов природы, отличных от частиц. Опыт показывает, что все, что нас окружает, есть либо частицы, либо поля, либо структуры, объединяющие частицы и поля, и все они могут совершать движение друг относительно друга. Заметим, однако, что в релятивистской квантовой физике принципиальная разница между частицей и

полем исчезает (иногда говорят, что частица есть квант соответствующего поля).

В классической и даже релятивистской (но не квантовой) физике понятия частицы и поля разграничены. Движение частиц или их ансамблей — макроскопических тел — оказывается главным предметом рассмотрения, тогда как полем обусловлено взаимодействие частиц. В том случае, когда источники поля остаются за рамками механической задачи, рассматривается движение в заданном поле, если же исследуется движение замкнутой системы как результат действия полей, ею же порожденных, то задача называется самосогласованной.

Поля, известные физике на сегодняшний день, бывают четырех видов и, соответственно, представляют четыре типа взаимодействия. Мы с детства знакомы с силой тяжести и в какой-то мере умеем измерять ее. Самое естественное понятие о массе следует из измерения веса тел. Это дает нам понятие о гравитационном взаимодействии. Очень многие хрестоматийные задачи механики рассматривают поведение частицы или системы частиц в гравитационном поле.

Столь же универсальным в окружающем нас мире оказывается электромагнитное взаимодействие; как уже отмечалось, вся наша связь и взаимодействие с внешним миром именно этим взаимодействием обусловлены. Это, правда, лишь в XX в. было понято должным образом. Подобно гравитационному, электромагнитное взаимодействие не ограничено расстоянием, что и отражает, в сущности, поэтический штамп «свет далекой звезды».

Еще два типа полей и взаимодействий — сильное и слабое — напротив, ограничены радиусом действия, и притом столь малым, что рассматривать их можно только в рамках квантовой физики. В нашем курсе мы рассмотрим их во втором томе применительно к физике ядра и элементарных частиц. Классическая механика с этими полями дела иметь не может.

Теперь, наконец, мы можем определить содержание данного раздела нашей книги. В ней рассказывается о *законах движения макроскопических тел* при скоростях $v \ll c$ и пространственных масштабах, примерно соответствующих нашему повседневному опыту. Наука, которая изучает относительное движение таких тел, называется *классической механикой*.

Вопросы

1. Каковы основные физические свойства нашего пространства?
2. Чему равна сумма углов треугольника, изображенного на рис. 1.2, включающего в себя четверть верхней полусферической поверхности?
3. Что такое часы?
4. Является ли проекция векторной величины на одну из координатных осей (x , y , z) скалярной величиной?

Глава 2

ВВЕДЕНИЕ В КИНЕМАТИКУ

2.1. Кинематика материальной точки

Кинематикой называется раздел механики, в котором изучаются самые общие характеристики движения, не зависящие от характера взаимодействия между телами. Существует следующее простое определение кинематики: геометрия движения.

В классической и релятивистской механике кинематика оказывается различной. Кинематика движения тела конечных размеров сложнее, чем кинематика точки, хотя в классической механике она и сводится к последней. Темой данного параграфа будет нерелятивистская кинематика материальной точки.

Фундаментальные кинематические понятия — скорость и ускорение. Из школьного курса физики читателю знакомы средняя скорость и среднее ускорение. Пусть материальная точка совершает одномерное перемещение, и пусть она сместилась на расстояние Δx за промежуток времени Δt . Тогда, по определению, средняя скорость есть

$$\langle v \rangle = \frac{\Delta x}{\Delta t}. \quad (2.1)$$

Пусть, далее, средние скорости на двух последовательных участках пути, отстоящих друг от друга на Δt , отличаются на величину $\Delta \langle v \rangle$. Среднее ускорение может быть определено как

$$\langle a \rangle = \frac{\Delta \langle v \rangle}{\Delta t}. \quad (2.2)$$

Естественным обобщением этих первичных определений оказываются понятия мгновенной скорости и ускорения:

$$v = \lim_{\Delta t \rightarrow 0} \frac{\Delta x}{\Delta t}, \quad (2.3)$$

$$a = \lim_{\Delta t \rightarrow 0} \frac{\Delta v}{\Delta t}. \quad (2.4)$$

Заметим, что именно необходимость адекватного кинематического описания послужила Ньютону и Г. Лейбницу (1646–1716) стимулом для создания основ дифференциального и интегрального исчисления. Следующим естественным обобщением определений (2.3), (2.4) будет переход к неодномерному движению, которое, естественно, уже нельзя будет описать одной координатой.

натой x . В тех случаях, когда по условиям задачи движение тела можно заменить на движение материальной точки, его положение в пространстве определяется радиусом-вектором $\mathbf{r}(t)$ этой точки, который при движении меняется со временем, и нахождение закономерностей этого изменения и является основной задачей механики.

Пусть тело движется по криволинейной траектории (рис. 2.1 а). И пусть в момент времени t оно оказалось в точке A , положение которой относительно выбранной системы координат определяется радиусом-вектором $\mathbf{r}_A(t)$. Через промежуток времени Δt тело оказалось в точке B с радиусом-вектором $\mathbf{r}_B(t + \Delta t)$. Расстояние вдоль траектории между точками (например, между точками A и B) называется *путем* Δs_{AB} , пройденным телом, а разность векторов $\mathbf{r}_B - \mathbf{r}_A = \Delta \mathbf{r}_{AB}$ называется вектором *перемещения* тела из точки A в точку B . Если нам известен закон движения тела по траектории, т.е. известна функция времени $\mathbf{r}(t)$, то мы всегда сможем найти и длину пути, скажем, из точки A в точку B , и вектор перемещения $\Delta \mathbf{r}_{AB}$ (построив, например, траекторию графически). А чему же при этом равняется скорость тела, допустим, в точке A ? Идея

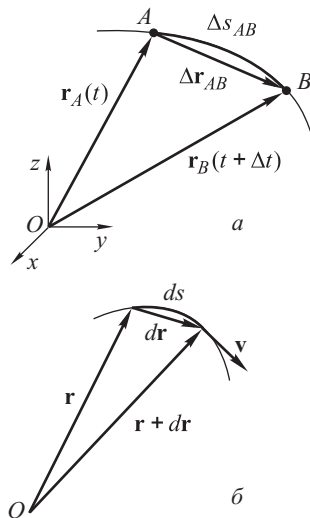


Рис. 2.1

определения скорости в точке A в момент времени t состоит в том, чтобы выразить ее, как частное от деления перемещения из точки A в соседнюю, близкую к ней точку траектории, на промежуток времени, которое затрачивает тело на это перемещение.

Таким образом, в трехмерном случае, то есть при движении в пространстве, естественно, изменяется не только величина, но и направление скорости. Поэтому скорость должна быть определена как предел отношения приращения вектора $\Delta \mathbf{r}$ к промежутку времени Δt

$$\mathbf{v}_A = \lim_{\Delta t \rightarrow 0} \frac{\Delta \mathbf{r}}{\Delta t} = \frac{d\mathbf{r}}{dt}. \quad (2.5)$$

Хотя определение скорости и содержит предел $\Delta t \rightarrow 0$, но мы должны понимать, что в реальном опыте предел $\Delta t = 0$ все равно никогда не достигим, так как любое измерение осуществляется всегда с некоторой ошибкой, как говорят, с конечной точностью. Более того, *экспериментальное значение любой физической величины имеет смысл лишь постольку, поскольку указано, с какой точностью эта величина измерена*. Поэтому для опыт-

ного измерения скорости тела в какой-то точке его траектории необходимо с максимальной точностью измерить приращение $\Delta \mathbf{r}$ за минимально возможный при заданной точности промежуток времени Δt . Частное от деления этих величин и определяет экспериментальное значение скорости, а используемые при этом приборы и метод измерения определяют, с какой конкретно точностью известна скорость интересующего нас тела.

Определение скорости (2.5) можно выразить следующими словами: *скорость материальной точки является производной ее радиуса-вектора по времени*. Правила вычисления производных различных функций изучаются в соответствующих разделах математики. Для нас важно понимание того, что если нам известен закон движения тела по траектории (известна функция времени $\mathbf{r}(t)$), то мы по известным и простым математическим правилам сможем вычислить $\mathbf{v}(t)$ — значение скорости тела в любой точке этой траектории.

Какие свойства скорости как физического понятия особенно важны? Прежде всего, *скорость является векторной величиной* и, следовательно, однозначно определяется, например, тремя своими проекциями на оси какой-либо декартовой системы координат. В такой системе координат указанные проекции v_x , v_y , v_z определяются в соответствии с (2.5) как

$$v_x = \frac{dx}{dt}, \quad v_y = \frac{dy}{dt}, \quad v_z = \frac{dz}{dt}, \quad (2.6)$$

где $x(t)$, $y(t)$, $z(t)$ — проекции радиуса-вектора $\mathbf{r}(t)$ движущейся материальной точки на соответствующие оси координат.

Любой вектор можно определить, не только указав три его координаты в какой-либо системе отсчета, но и задав его направление и его модуль (длину). В пределе $\Delta t \rightarrow 0$ бесконечно малое перемещение $d\mathbf{r}$ совпадает с бесконечно малым элементом траектории, точнее, в этом пределе модуль $d\mathbf{r}$ совпадает с бесконечно малым изменением пути ds , *направление $d\mathbf{r}$ совпадает с направлением движения, которое определяется касательной к траектории в заданной точке* (рис. 2.1 б), а *модуль скорости равен производной пути по времени*

$$v = \frac{ds}{dt}.$$

С учетом (2.6) модуль скорости можно представить также в виде

$$v = \sqrt{v_x^2 + v_y^2 + v_z^2}.$$

Когда при движении тела модуль скорости остается постоянным (равным, например, v_0), то такое движение называют *равномерным*. Если при этом весь пройденный телом путь s представить как сумму элементарных перемещений ds , то из

(2.5) мы получим

$$s = \sum ds = \sum v_0 dt = v_0 \sum dt = v_0 t,$$

где мы учли, что сумма элементарных промежутков времени dt равна времени t , за которое тело прошло путь s . Таким образом, для равномерного движения мы получаем

$$v_0 = \frac{s}{t},$$

то есть *модуль скорости равномерного движения равен пути, проходимому телом за единицу времени*. Именно таково понимание скорости в обыденной жизни.

Аналогичным образом определяется при неоднородном движении и вектор ускорения материальной точки:

$$\mathbf{a}(t) = \frac{d\mathbf{v}}{dt} = \lim_{\Delta t \rightarrow 0} \frac{\Delta \mathbf{v}}{\Delta t}. \quad (2.7)$$

Итак, *ускорение — это производная скорости по времени*.

Так же, как и радиус-вектор $\mathbf{r}(t)$ и скорость $\mathbf{v}(t)$, ускорение $\mathbf{a}(t)$ является векторной величиной. Модуль и направление $\mathbf{a}(t)$ можно всегда вычислить, если известен закон движения тела по траектории, то есть известна зависимость радиуса-вектора от времени.

2.2. Равномерное движение по окружности

В качестве примера таких вычислений рассмотрим, как определяется модуль и направление ускорения при равномерном движении материальной точки по окружности радиуса R со скоростью, модуль которой равен v . С таким типом движения приходится сталкиваться очень часто: от вращения на карусели до движения спутников и космических кораблей.

Важным шагом при решении любой механической задачи является выбор системы отсчета. В нашем случае мы поместим начало декартовой системы координат в центр окружности, по которой движется тело, а оси x и y расположим в плоскости движения (рис. 2.2). Пусть в начальный момент времени тело находилось в наивысшей точке окружности A , то есть начальные значения проекций радиуса-вектора равны $x(0) = 0$, $y(0) = R$. Проекция $z(t)$ всегда равна нулю, так как движение происходит в плоскости xy . Обозначим меняющийся со временем угол между $\mathbf{r}(t)$ и Oy через $\varphi(t)$.

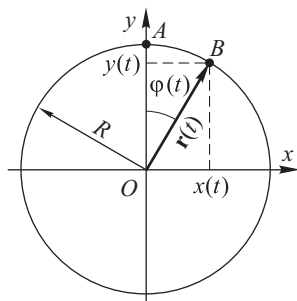


Рис. 2.2

Тогда траектория движения будет описываться зависимостью от времени проекций радиуса-вектора $\mathbf{r}(t)$, а именно:

$$\begin{aligned} x(t) &= R \sin \varphi(t), \\ y(t) &= R \cos \varphi(t). \end{aligned} \quad (2.8)$$

Отношение пройденной телом дуги окружности AB к радиусу R определяет угол $\varphi(t)$, измеренный в радианах. При равномерном движении тела по окружности длина дуги AB равна пройденному пути $s(t) = vt$, и поэтому

$$\varphi(t) = \frac{s(t)}{R} = \frac{v}{R} t \equiv \omega t, \quad (2.9)$$

где $\omega = v/R$ — *угловая скорость* движения тела по окружности.

Определим сначала проекции скорости v_x и v_y . Используя правила для вычисления производных от тригонометрических функций, получаем с помощью (2.8), (2.9):

$$v_x = \frac{dx}{dt} = R \frac{d}{dt} (\sin \omega t) = R\omega \cos \omega t, \quad v_y = \frac{dy}{dt} = -R\omega \sin \omega t.$$

Как следствие, получаем известную формулу, связывающую линейную и угловую скорости при равномерном движении по окружности:

$$v = \sqrt{v_x^2 + v_y^2} = R\omega.$$

Теперь можно получить и значения проекций ускорения на оси декартовых координат:

$$a_x = \frac{dv_x}{dt} = -R\omega^2 \sin \omega t, \quad a_y = \frac{dv_y}{dt} = -R\omega^2 \cos \omega t.$$

Отсюда находим окончательное выражение для модуля вектора \mathbf{a} :

$$a = \sqrt{a_x^2 + a_y^2} = R\omega^2 = \frac{v^2}{R}. \quad (2.10)$$

Нетрудно усмотреть, что a_x и a_y меняются со временем синхронно с проекциями радиуса-вектора $x(t)$ и $y(t)$ по законам $\sin \omega t$ и $\cos \omega t$, то есть векторы $\mathbf{a}(t)$ и $\mathbf{r}(t)$ связаны соотношением

$$\mathbf{a}(t) = -\omega^2 \mathbf{r}(t). \quad (2.11)$$

Знак «минус» означает, что вектор \mathbf{a} направлен противоположно вектору \mathbf{r} , т.е. *ускорение материальной точки при ее равномерном движении по окружности направлено к центру*. По этой причине ускорение, модуль которого определяется формулой (2.10), а направление совпадает с нормалью к траектории, называют обычно *центростремительным ускорением*.

Если же сама скорость неоднородного движения меняется по величине, этому соответствует ненулевая компонента ускорения по направлению касательной. А вот по направлению так

называемой бинормали — вектора, ортогонального плоскости, содержащей векторы нормали и касательной, проекции как скорости, так и ускорения всегда равны нулю. Понятие угловой скорости можно обобщить и на этот случай, подобно тому, как это делается при рассмотрении прямолинейного неравномерного движения для линейной скорости. Определим *угловую скорость* при неравномерном движении по окружности как производную по времени естественной — угловой — координаты:

$$\omega(t) = \frac{d\varphi(t)}{dt}. \quad (2.12)$$

Более сложные траектории, а также основные законы неравномерного движения по криволинейной траектории будут рассмотрены в последующих главах нашего курса.

В механике, как и в любом последующем разделе физики, особую ценность количественным результатам придает их инвариантность по отношению к выбору системы координат. В частности, векторная алгебра, а в дальнейшем и векторный анализ, тем и привлекательны в качестве языка, на котором формулируются законы физики. Формулы (2.11), (2.12) представляют собой первый шаг в данном направлении, но и они оставляют без должной определенности вопрос о плоскости, в которой происходит движение тела по окружности. Иными словами, мы полностью определим движение тела в пространстве, если дополним наше рассмотрение уравнением плоскости (см. рис. 2.2).

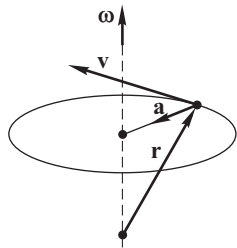


Рис. 2.3

Эту проблему мы можем с успехом разрешить, введя *вектор угловой скорости* (рис. 2.3). Определяем вектор ω как перпендикулярный плоскости нашей окружности и направленный так, чтобы при наблюдении со стороны его конца движение точки по окружности происходило *против* часовой стрелки. Иначе говоря, вектор как бы «завинчивается» в пространство по закону правого винта. Тогда скорость и ускорение при равномерном движении по окружности могут быть представлены в виде векторных произведений:

$$\frac{d\mathbf{r}}{dt} = [\omega \mathbf{r}]; \quad \frac{d\mathbf{v}}{dt} = [\omega \mathbf{v}]. \quad (2.13)$$

Напомним, что по определению векторного произведения вектор $\mathbf{c} = [\mathbf{a} \mathbf{b}]$ ортогонален плоскости $(\mathbf{a} \mathbf{b})$, образует с этими векторами правую тройку: $(\mathbf{a}, \mathbf{b}, \mathbf{c}) \sim (x, y, z)$ и равен по модулю $c = ab \sin \angle(\mathbf{a}, \mathbf{b})$.

Остановимся кратко на инвариантности формул (2.13), то есть на независимости этих соотношений от выбора системы

координат. Первое из этих соотношений позволяет выбирать начало координат в любой точке оси вращения, в данном случае — прямой, нормальной к плоскости нашей окружности и проходящей через ее центр. Эта инвариантность прямо следует из свойства векторного произведения: если $\mathbf{a} \parallel \mathbf{b}$, то $[\mathbf{a} \mathbf{b}] = 0$. Второе из равенств (2.13) справедливо всегда, если центр окружности неподвижен. Заметим, что кинематически оно эквивалентно первому. Действительно, оба вектора, \mathbf{r} и \mathbf{v} , при указанных выше условиях равномерно вращаются, не изменяя модуля, с одной и той же угловой скоростью $\boldsymbol{\omega}$. Таким образом, можно сформулировать и более общее утверждение. Если некоторая векторная величина \mathbf{b} изменяется во времени так, что это изменение можно представить как вращение с угловой скоростью $\boldsymbol{\omega}$, то имеет место соотношение

$$\frac{d\mathbf{b}}{dt} = [\boldsymbol{\omega} \mathbf{b}]. \quad (2.14)$$

Далее, в гл. 7 будет рассмотрено вращательное движение твердого тела, и тогда первое из соотношений (2.13) можно будет трактовать уже как *векторное поле* скоростей, т.е. заданную векторную функцию в пространстве (и может быть, во времени). Скорость \mathbf{v} привязывается в этом случае не к материальной точке как элементу твердого тела, а к точке пространства, и с этой скоростью движется элемент тела в тот момент, когда он «посещает» данную точку пространства. (В дальнейшем такое представление переносится на течение жидкости.)

Следует, отметить, что $\boldsymbol{\omega}$ — не совсем обычный вектор. При поворотах системы координат он, подобно истинным векторам, например, \mathbf{r} и \mathbf{v} , сохраняет величину и направление. Но если подвергнуть систему координат операции отражения $\{x, y, z \Rightarrow -x, -y, -z\}$, то мы должны поменять знак $\boldsymbol{\omega}$. Это следует формально из того, что формулы (2.13) должны определять \mathbf{r} и \mathbf{v} инвариантным образом. Возможна и простая физическая (или геометрическая) интерпретация этого свойства: мы определили $\boldsymbol{\omega}$ через правило винта, но при перемене знака всех трех координат правый винт трансформируется в левый и наоборот.

Такие величины называются *псевдовекторами*. Псевдовектором оказывается векторное произведение любых двух истинных векторов. (Соответственно, смешанное произведение $\mathbf{a}[\mathbf{b} \mathbf{c}]$ называется *псевдоскаляром*.) Таким образом, мы видим, что объявлять векторной любую величину, имеющую направление и представимую в трех проекциях, было бы неразумно. Повышенные требования к объекту, претендующему на звание вектора, придают истинно векторным законам дополнительную доказательную силу.

Но пока и поскольку мы не оперируем отражением системы координат, $\boldsymbol{\omega}$ может рассматриваться именно как вектор, из чего

проистекают вполне практические следствия. При решении задач на динамику твердого тела конечных размеров часто бывает удобно представлять движение тела как сумму двух или более вращений,

$$\omega = \omega_1 + \omega_2,$$

либо производить обратную операцию. Что же касается переноса начала координат либо перехода в систему отсчета, движущуюся относительно данной с некоторой скоростью, то к этим операциям в контексте формул (2.13) мы еще вернемся в последующих главах.

2.3. Приближение абсолютно твердого тела и приближение материальной точки

Не только история науки, но и повседневный человеческий опыт показывают, что приступая к изучению какого-либо нового явления, разумно не пытаться сразу же искать объяснения всех его зачастую очень сложных сторон и деталей, а двигаться шаг за шагом от понимания простых закономерностей к пониманию все более полной картины. Так и нам при изучении механики следует начать с рассмотрения простейших типов движения, на примере которых мы освоим основные идеи и понятия этой науки, после чего круг изучаемых движений можно будет расширить. В самом деле, трудно рассчитывать на то, что мы с первых же шагов сможем, например, разобраться в законах птичьего полета. Только лишь для строгого определения положения птицы относительно выбранной системы координат в некоторый момент времени необходимо знать радиусы-векторы различных точек крыльев, корпуса, хвостового оперения, головы (рис. 2.4), так что математическое описание полета неизбежно будет содержать соотношения между большим числом зависящих от времени величин. Получить из этих соотношений траекторию полета наверняка окажется необычайно сложной задачей.

Сформулируем теперь наше первое важное упрощение. Оно состоит в том, что мы будем на первых этапах изучать движение только *абсолютно твердых тел*. Этим термином принято обозначать тела, *при движении которых можно пренебречь их деформациями, то есть смещениями одних частей тела относительно других*. Под влиянием внешних воздействий все тела испытывают те или иные деформации. Стальной стержень

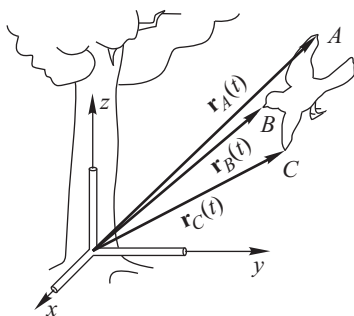


Рис. 2.4

даже при обычных изменениях атмосферного давления и температуры меняет свои размеры и форму, хотя эти изменения можно заметить только с помощью весьма чувствительных приборов. Но если этот стержень является деталью достаточно грубого механизма, то такими деформациями можно пренебречь и его движение можно будет рассчитывать при конструировании механизма по законам движения абсолютно твердого тела, о которых пойдет речь в дальнейшем. Однако если тот же стержень является деталью сложного электронного прибора, то учет его деформаций может оказаться главной задачей конструктора, и тогда приближение абсолютно твердого тела будет неприменимо. Поэтому представление об абсолютно твердом теле как о теле, не подверженном вообще никаким деформациям, является всего лишь определенным приближением к реальной действительности, удобной идеализацией (или, как говорят физики, моделью), применимость или неприменимость которой при рассмотрении движения тела зависит от конкретных условий.

Относительная простота законов движения абсолютно твердого тела обусловлена уже тем, что для строгого математического описания его положения в какой-либо системе координат

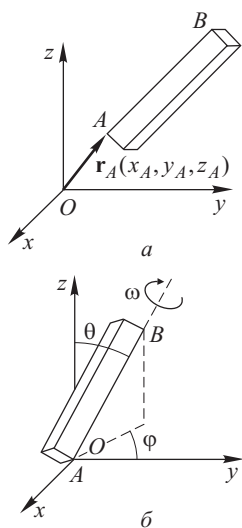


Рис. 2.5

требуется задавать гораздо меньше параметров (координат), чем для деформируемого тела. Число независимых параметров, однозначно определяющих положение тела в пространстве, называют *числом степеней свободы*. Посмотрим, чему равно число степеней свободы абсолютно твердого тела. Представим для большей наглядности, что наше тело имеет форму параллелепипеда, как изображено на рис. 2.5, и пусть линия AB является пересечением каких-либо двух его граней. Полное число степеней свободы такого тела можно определить, например, как сумму двух слагаемых: одно — это число степеней свободы точки A и другое — это число независимых параметров, которые однозначно определяют положение тела при уже фиксированном в пространстве положении точки A . Число степеней свободы точки A равно трем — это, например, три декартовы координаты x_A, y_A, z_A , определяющие радиус-вектор \mathbf{r}_A точки A в заранее выбранной системе координат (рис. 2.5 а). При фиксированном положении одной точки A положение всего тела еще может быть различным, и это положение можно окончательно зафиксировать, если задать три угла поворота θ, φ и ω (рис. 2.5 б). Угол θ — это угол

наклона линии AB к оси z исходной системы координат. Угол φ — это угол поворота линии AB вокруг оси z при фиксированном угле наклона θ . И, наконец, угол ω — это угол поворота всего тела вокруг линии AB . Итак, *число степеней свободы абсолютно твердого тела равно шести*: три координаты точки $A — x_A, y_A, z_A$, и три угла поворота — θ, φ, ω .

Таким образом, для описания произвольного движения абсолютно твердого тела необходимо знать, как меняются со временем всего шесть параметров, т.е. необходимо знание шести функций от времени: $x_A(t), y_A(t), z_A(t), \theta(t), \varphi(t), \omega(t)$. Хотя мы пока еще не знаем законов движения твердого тела, т.е. соотношений или уравнений, позволяющих определить эти функции и, тем самым, предсказать траекторию движения тела, можно заранее предположить, что эти законы будут в общем случае очень сложными — даже из школьной математики мы знаем, что чем больше число неизвестных, тем сложнее решение задачи. Поэтому рассмотрим сначала движения, описание которых не требует столь большого количества зависящих от времени величин.

Заметим для этого, что любое движение абсолютно твердого тела можно разложить на два основных вида движения — *поступательное* и *вращательное*. Поступательное движение — это такое движение, при котором любая прямая, связанная с движущимся телом, остается параллельной самой себе (рис. 2.6). При вращательном движении все точки тела движутся по окружностям, центры которых лежат на одной и той же прямой, называемой *осью вращения* (рис. 2.7). При этом происходит изменение ориентации тела. Отметим, что при конечных перемещениях поворота тела относительно только одной оси, вообще говоря, недостаточно, чтобы перевести его в новое положение. Сама ось вращения может находиться вне тела. Например, перемещение тела из положения a в положение b на рис. 2.8 можно представить как поступательное движение из положения a в положение b и последующего поворота тела на 180° вокруг оси AB .

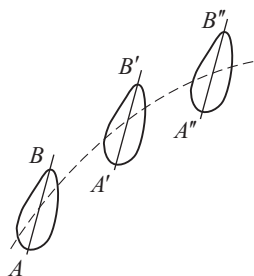


Рис. 2.6

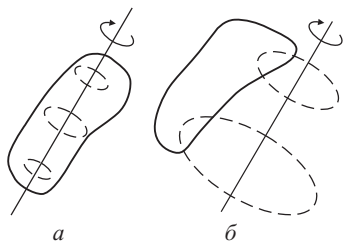


Рис. 2.7

Главным для нас является сейчас тот факт, что если абсолютно твердое тело совершает только поступательное движение, то все его положения в пространстве однозначно определяются положением всего одной какой-либо его точки, например, точ-

ки A на рис. 2.8. Следовательно, при математическом описании поступательного движения твердого тела его размеры не играют роли, и все это тело можно заменить точкой, положение которой в пространстве определяется в общем случае всего тремя степенями свободы — например, декартовыми координатами $x(t)$, $y(t)$, $z(t)$, определяющими радиус-вектор этой точки. Это и есть приближение материальной точки.

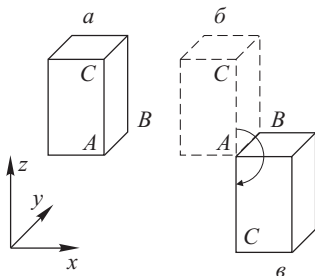


Рис. 2.8

Рассмотренный выше пример с поступательным движением твердого тела — не единственный, когда при описании движения «большого» тела его

можно заменить материальной точкой. Замена «реального» тела на материальную точку может быть оправдана и во всех тех случаях, когда в рамках поставленной задачи влиянием размеров тела можно пренебречь. Например, при вычислении траектории движения Земли вокруг Солнца можно пренебречь ее вращением вокруг своей оси, а также ее размерами, т. к. они малы по сравнению с расстоянием до Солнца. В результате положение Земли в пространстве будет определяться одной точкой и описание ее движения значительно упростится.

Имеется еще одна очень важная причина, почему приближение материальной точки играет в механике исключительную роль. Дело в том, что тело любого размера и любой формы можно представить как совокупность очень малых его частей, взаимодействующих друг с другом. Каждую из таких частей-частиц можно рассматривать как материальную точку и, следовательно, задачу о движении любого тела можно свести к задаче о движении совокупности материальных точек. И мы увидим в дальнейшем, что именно такой прием используется, например, при нахождении законов вращательного движения твердого тела и законов движения жидкости. Из всего вышесказанного следует, что приближение материальной точки (или, как говорят иногда, модель материальной точки) лежит в основе классической механики.

2.4. Преобразования Галилея и закон сложения скоростей

Остановимся теперь на вопросе о том, как изменяется скорость и ускорение материальной точки для наблюдателя, который переходит из одной системы отсчета в другую, движущуюся относительно первой. Так, например, если космическую ракету

запускают с поверхности Земли, то ее скорость относительно Солнца определится правилом, по которому следует сложить скорость ракеты относительно Земли и скорость самой Земли относительно Солнца.

Принципиальный ответ на наш вопрос следует из того факта (являющегося обобщением огромного массива опытных данных), что положение точки в пространстве и, соответственно, ее скорость и ускорение представляются векторными величинами. Тогда правило сложения векторов дает

$$\mathbf{r} = \mathbf{r}_1 + \mathbf{r}_2 \Rightarrow \frac{d\mathbf{r}}{dt} = \frac{d\mathbf{r}_1}{dt} + \frac{d\mathbf{r}_2}{dt} \Rightarrow \frac{d^2\mathbf{r}}{dt^2} = \frac{d^2\mathbf{r}_1}{dt^2} + \frac{d^2\mathbf{r}_2}{dt^2},$$

или, что то же,

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2; \quad \mathbf{a} = \mathbf{a}_1 + \mathbf{a}_2.$$

Рассмотрим этот вопрос подробнее. Пусть имеются две системы отсчета, движущиеся друг относительно друга с постоянной скоростью \mathbf{v}_0 (рис. 2.9). Систему, обозначенную на рисунке

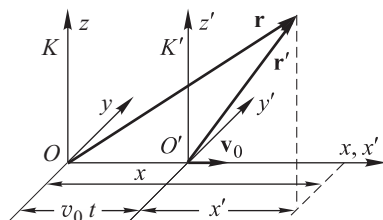


Рис. 2.9

буквой K , будем условно считать неподвижной. Тогда вторая система K' будет двигаться равномерно и прямолинейно. Координатные оси обеих систем выберем так, чтобы оси Ox и $O'x'$ совпадали, а оси Oy и $O'y'$ и оси Oz и $O'z'$ были бы параллельны друг другу. Пусть координаты некоторой движущейся материальной точки M в системе K равны x, y, z , а в системе $K' — x', y', z'$. Время начнем отсчитывать с того момента, когда начала координат обеих систем совпали. Установим сначала связь между координатами нашей движущейся точки в системах K и K' , а потом определим связь между скоростями в разных системах. Как видно из рисунка 2.9, $x = x' + v_0 t$. Очевидно также, что $y = y'$ и $z = z'$. Одно из основных положений классической механики, обобщающее как повседневный опыт, так и огромный массив экспериментальной информации, утверждает, что при $v \ll c$ время абсолютно, т. е. в обеих наших системах время течет одинаково: $t = t'$. В результате получаем четыре соотношения:

$$x = x' + v_0 t', \quad y = y', \quad z = z', \quad t = t', \quad (2.15)$$

называемых *преобразованиями Галилея* (по имени итальянского физика и астронома Галилео Галилея (1564–1642)).

Помня, что скорость — это производная радиуса-вектора по времени, найдем связь между скоростями точки M в обеих

системах, продифференцировав соотношения (2.15) по времени:

$$\frac{dx}{dt} = \frac{dx'}{dt} + v_0, \quad \frac{dy}{dt} = \frac{dy'}{dt}, \quad \frac{dz}{dt} = \frac{dz'}{dt},$$

то есть соотношения между проекциями скоростей имеют вид

$$v_x = v'_x + v_0, \quad v_y = v'_y, \quad v_z = v'_z.$$

Последние формулы эквивалентны следующему соотношению между вектором скорости \mathbf{v} по отношению к системе K и вектором скорости \mathbf{v}' по отношению к системе K' :

$$\mathbf{v} = \mathbf{v}' + \mathbf{v}_0. \quad (2.16)$$

Продифференцируем по времени равенство (2.16). Учитывая, что \mathbf{v}_0 постоянна, получим

$$\frac{d\mathbf{v}}{dt} = \frac{d\mathbf{v}'}{dt}, \quad \text{или} \quad \mathbf{a} = \mathbf{a}'. \quad (2.17)$$

Другими словами, *ускорение какого-либо тела во всех системах отсчета, движущихся друг относительно друга равномерно и прямолинейно, одинаково.*

Вопросы и задачи

1. Сколько степеней свободы у материальной точки? Сколько степеней свободы у велосипеда?

2. Придумайте задачу, когда при рассмотрении движения велосипеда его можно считать материальной точкой.

3. Вычислите ускорение Земли при ее движении вокруг Солнца, считая ее орбиту круговой, и сравните это ускорение с ускорением свободного падения на поверхности Земли.

Ответ: $a/g \approx 6 \cdot 10^{-4}$.

4. Начертите приближенно траекторию движения относительно Солнца какого-либо тела, покоящегося на экваторе Земли.

5. На рис. 2.10 дан график зависимости скорости некоторого тела от времени. Постройте графики ускорения этого тела и пройденного им пути.

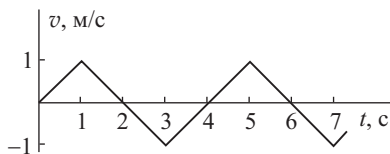


Рис. 2.10

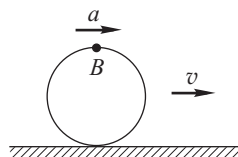


Рис. 2.11

6. Шар радиуса R катится по горизонтальной плоскости без проскальзывания с ускорением a . Определите ускорение точки B (рис. 2.11) в момент времени, когда скорость тела равна v .

Ответ: $a_B = \sqrt{a^2 + (v^2/R)^2}$.

Глава 3

ЗАКОНЫ НЬЮТОНА

Как известно, основные законы механики в ее классическом варианте были сформулированы Ньютоном. И хотя их формулировка была не вполне адекватна современному пониманию, они, тем не менее, с успехом послужили основанием для построения динамики — науки о движении тел при заданном взаимодействии между ними. Благодаря глубокому пониманию физики дела и сильной математической интуиции, Ньютон ввел именно столько законов, сколько было нужно для этой цели — три. «Школьная» их формулировка читателю хорошо известна:

$$\mathbf{v} = \text{const}, \quad \text{если} \quad \mathbf{F} = 0; \quad (3.1)$$

$$\mathbf{F} = m\mathbf{a}; \quad (3.2)$$

$$\mathbf{F}_{12} = -\mathbf{F}_{21}. \quad (3.3)$$

Мы опускаем подробные пояснения, поскольку все обозначения — общепринятые.

На сегодняшний день такая трактовка едва ли может нас удовлетворить. В самом деле, почему первый закон Ньютона (3.1) не рассматривается как тривиальное следствие второго? Еще сложнее обстоит дело с зависимостью (3.2). Эта формула содержит сразу две новых величины — силу и массу, которые следовало бы определять независимо. Поэтому нам следует дать согласованную и, по возможности, непротиворечивую формулировку главных законов динамики (3.1)–(3.3).

3.1. Закон инерции

Ньютон при формулировке законов механики придавал должное значение системе отсчета. Он полагал, однако, что не только время, но и пространство абсолютно, иными словами, постулировал существование выделенной системы отсчета, связанной с пространством как таковым. Хорошим приближением к такой системе он считал систему координат, привязанную к «неподвижным» звездам. Именно в такой системе отсчета формулировались законы Ньютона. Как следствие сильных неравенств, обуславливающих применимость классической механики, это приближение оказалось не так уж и плохо.

Но в действительности, как мы теперь знаем, звезды не неподвижны. Более того, развитие физики заставило научное сообщество отказаться от идеи абсолютного пространства. По современным представлениям, никаких аналогов абсолютного пространства (вроде, например, «мирового эфира») также не существует. Тем не менее, законы механики, включая релятивистскую и квантовую (но не общую теорию относительности!), принято формулировать изначально в особой системе отсчета, которая базируется на понятии свободного физического тела. Тело должно быть свободно от взаимодействия с другими телами, а это возможно лишь для тела, абсолютно уединенного. В самом деле, сколь ни слабы были бы взаимодействия между двумя, тремя и т. д. телами, все же поля — их носители — на конечном расстоянии никогда в точности нулевыми не будут. Значит, совершенная свобода предполагает бесконечную удаленность тела, являющегося предметом рассмотрения, от любых других тел.

Реальная физика, конечно, может трактовать это свойство исключительно в смысле некоторого предельного перехода, а реальные ситуации рассматривать как приближения к описанному идеальному случаю с той или иной точностью.

«Привязав» некоторую систему отсчета к такому телу, мы обнаружим, что в ней тело, ни с чем не взаимодействующее, *покоится*. Но тогда и в любых других системах отсчета, движущихся относительно нашей выделенной системы равномерно и прямолинейно, ускорение свободного тела должно быть равно нулю. В нерелятивистской механике это прямо следует из соотношений (2.17), в релятивистской же — это свойство придется постулировать особо. Таких систем — бесконечно много (континуум), и все они равноправны; в общем случае произвольное свободное тело в произвольной системе отсчета такого рода должно двигаться с некоторой постоянной скоростью \mathbf{v} .

Система отсчета, в которой абсолютно свободное тело должно пребывать в состоянии покоя или равномерного прямолинейного движения, называется инерциальной. Именно в инерциальных системах принято формулировать физические законы, в частности законы механики, дабы придать им по возможности универсальную силу. Поскольку абсолютно свободное тело может мыслиться лишь как идеализация, предельный случай, существование таких систем должно постулироваться на основе опыта обращения с физическими законами. *Существование хотя бы одной, а следовательно, бесконечного числа инерциальных систем отсчета, и есть первый закон Ньютона в современном его понимании.*

3.2. Второй закон Ньютона

Попытаемся привести закон (3.2) к корректному виду. Для этого нам необходимо некоторое обобщение совокупности экспериментальных данных *применительно к приближению материальной точки*.

Во-первых, опыт говорит о том, что для тел пренебрежимо малых размеров или в случаях, когда по условиям задачи ориентационные степени свободы несущественны, *взаимодействием тел определяется вектор ускорения каждого тела*.

Во-вторых, если с телом, ускорение которого определяется в заданном силовом поле, жестко связать еще одно, точно такое же тело, то ускорение падает в два раза (при том, однако, что оба тела могут рассматриваться порознь и вместе как материальные точки). Дальнейшее обобщение этого факта может быть выражено так: при ускорении в формуле (3.2) стоит некоторый *аддитивный коэффициент t* . Это значит, что при жестком соединении любого числа произвольных тел закон (3.2) останется в силе, причем масса совокупного тела будет равна сумме масс тел, его составляющих. Мера коэффициента t произвольна и должна эталонироваться.

В третьих, взаимодействие материальных точек друг с другом, а равно и их поведение во внешнем поле, задается некоторой векторной величиной \mathbf{F} — со всеми типичными свойствами векторов, включая правило векторного сложения (см. гл. 1). Ускорение каждого тела пропорционально \mathbf{F} .

Эта величина связана с законом Гука (см. гл. 9), а также с законом всемирного тяготения (см. ниже), и в итоге механика представляет собой сложную замкнутую систему взаимосвязанных физических соотношений. То обстоятельство, что закон (3.2) подразумевает, как мы уже отмечали, введение сразу двух новых понятий, создает известные логические трудности. (Они обусловлены возможным произволом в определении массы и силы.) Разумеется, система всех физических понятий в целом свободна от этого парадокса, и в принципе, современная физика и рассматривает их в совокупности, опираясь на известные экспериментальные данные.

Обсудим более подробно вопрос о массе тела. Являясь фундаментальной физической величиной, масса определяет и инерционные, и гравитационные свойства тел, т.е. определяет различные физические явления. В ньютоновской теории гравитации масса служит источником силы всемирного тяготения, притягивающей все тела друг к другу. Как следствие, ускорение тела, свободно падающего в гравитационном поле, не зависит ни от его массы, ни от свойств вещества, из которого оно состоит. Эта закономерность проверена в поле Солнца с точностью

порядка 10^{-12} . Обычно ее называют равенством инертной и гравитационной масс.

В нерелятивистском приближении, т.е. когда мы имеем дело со скоростями, много меньшими скорости света, масса тела служит мерой содержащегося в теле вещества и имеет место закон сохранения и аддитивности массы: масса изолированной системы тел не меняется со временем и равна сумме масс тел, составляющих эту систему. Поэтому в нерелятивистской механике масса тела пропорциональна его объему. В то же время в релятивистской механике масса тела не является аддитивной величиной. Правильно говорить лишь об энергии покоя тела. Эта энергия не равна сумме энергий покоя ее составляющих, причем необходимо учитывать еще энергию связи между частицами. Именно поэтому, например, масса ядра не равна сумме масс составляющих его частиц — нейтронов и протонов. Однако когда мы переходим к макроскопическим телам, величина энергии связи и поверхностная энергия оказываются столь малыми, что их можно не учитывать. Так, например, свободная поверхность воды дает вклад в массу порядка $7 \cdot 10^{-19}$ кг/м².

В настоящее время за единицу массы принят *килограмм — масса платино-иридиевого тела, хранящегося в Международном бюро мер и весов во Франции*.

Определиться с соотношением между силой, массой и ускорением помогает измерение веса тел в однородном поле тяжести. В земных условиях оно, правда, не так уж и однородно, но того, с чем реально приходится иметь дело в повседневной жизни, вполне достаточно для введения первичного понятия. Опираясь на эксперименты с пружинными весами, — при малом относительном удлинении пружины (чтобы не сказалась нелинейность деформации) — можем ввести понятие веса как величины, пропорциональной удлинению пружины. Вес оказывается пропорционален массе тела: $F_g \propto m$. Затем, убедившись, что все тела в пустоте падают с одинаковым ускорением g , приходим к соотношению $F_g = mg$. Оно явно выглядит как частный случай формулы (3.2).

Уравнение (3.2) при исследовании движения тел предпочтительнее использовать в виде

$$\mathbf{F} = m \frac{d^2 \mathbf{r}}{dt^2}. \quad (3.4)$$

Тем самым исключаются возможные недоразумения в случае неравноускоренного или непрямолинейного движения. Второй закон Ньютона, согласно опыту, описывает произвольное движение материальной точки.

Сила тяжести представляет собой гравитационное взаимодействие, или гравитационное поле. Как мы уже отмечали в гл. 1,

помимо гравитационного, лишь электромагнитное взаимодействие может быть предметом рассмотрения классической механики. Но последнее проявляется часто не в «чистом» виде типа закона Кулона, а в форме приближенных, усредненных сил, таких, например, как *упругие силы* и *силы трения*. Вывод их из основных законов электродинамики далеко выходит за рамки классической механики, которая пользуется *феноменологическим* описанием таких сил, базирующемся на обширном экспериментальном материале.

Под действием приложенных сил всякое реальное тело деформируется, то есть изменяет свои размеры и форму. При этом меняют свое расположение и входящие в состав молекул тела элементарные электрические заряды. Под действием электромагнитных сил они стремятся восстановить свое первоначальное положение, а вместе с ними и все тело стремится вернуться в исходное положение. Если после прекращения действия внешних сил тело принимает в точности первоначальные размеры и форму, деформация называется *упругой*. Упругие деформации наблюдаются в том случае, когда сила, обусловившая деформацию, не превосходит некоторого определенного для каждого конкретного тела предела (предела упругости). Возникающая в результате упругой деформации сила со стороны тела, обусловленная стремлением внутренних электрических зарядов восстановить свою первоначальную конфигурацию, называется *упругой силой*.

Возьмем пружину, имеющую в недеформированном состоянии длину l_0 , и приложим к ее концам равные по величине, но противоположно направленные силы \mathbf{F}_1 и \mathbf{F}_2 (рис. 3.1). Под действием этих сил пружина растянется на некоторую величину Δl , после чего наступит равновесие. Это означает, в соответствии со вто-

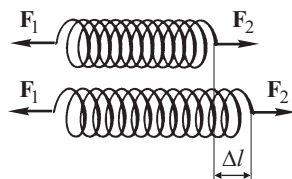


Рис. 3.1

рым законом Ньютона, что равна нулю векторная сумма сил, приложенных к телу. В точке, где приложена внешняя сила, равновесие обусловлено равенством этой силы и упругой силы. В то же время в каждой внутренней точке растянутой пружины равновесие обусловлено упругими силами, возникающими в пружине. Опыт показывает, что при небольших деформациях удлинение пружины Δl оказывается пропорциональным растягивающей силе: $\Delta l \sim F$ ($F = F_1 = F_2$). Соответственно, упругая сила оказывается пропорциональной удлинению пружины:

$$F = k\Delta l. \quad (3.5)$$

Коэффициент пропорциональности k называется *жесткостью* или *коэффициентом жесткости* пружины. Утверждение о

пропорциональности между упругой силой и деформацией носит название *закона Гука*, названного по имени известного английского физика Р. Гука (1635–1703) — современника Ньютона. Это утверждение относится не только и не столько к деформации пружины, но прежде всего к малым деформациям простых образцов, выполненных из кристаллических веществ, а также — с большими оговорками — аморфных и высокополимерных материалов.

Электромагнитное взаимодействие лежит также в основе *сил трения*, которые играют очень важную роль при движении окружающих нас тел. Мы еще остановимся подробнее на этом явлении, здесь же отметим, что эффект трения обусловлен взаимодействием электронных оболочек атомов (молекул), прилежащих к поверхности тела.

Из закона (3.4) можно определить размерность силы через базовые (эталонные) величины и, соответственно, единицу силы — 1 *ньютон* (1 Н):

$$[F] = [m] \frac{[l]}{[t^2]} \implies 1 \text{ Н} = 1 \text{ кг} \cdot \text{м}/\text{с}^2. \quad (3.6)$$

3.3. Выбор единиц измерения и системы единиц. Размерности физических величин

Остановимся подробнее на весьма важном вопросе о единицах физических величин. (Поскольку физика занимается самыми фундаментальными законами природы, она порождает единицы измерения, которые используются также в химии, биологии и технических науках.)

Законы физики устанавливают соотношения между физическими величинами, доступными измерению. Измерить какую-либо физическую величину (например, время) означает сравнить ее с величиной того же вида (в нашем случае — с определенным промежутком времени), принятой за единицу измерения этой величины. В принципе, для каждой физической величины, независимо от других, можно выбрать свою единицу измерения. Например, за единицу ускорения можно было бы выбрать ускорение свободного падения g и «присвоить» ей какое-либо специальное название. Оказывается, однако, что можно ограничиться выбором единиц всего для нескольких величин, принятых за основные. При этом единицы измерения для всех остальных величин будут уже определяться автоматически через эти основные единицы с помощью физических законов, связывающих выбранные величины друг с другом.

Именно таким образом в настоящее время установлена *Международная система единиц — СИ* (от начальных букв английского названия *System International*). Основными единицами

СИ в механике являются единицы измерения для трех фундаментальных свойств окружающего нас мира: пространства, времени и массы. Конкретно, такими единицами в системе СИ являются: единица длины — *метр* (сокращенное обозначение — м), единица времени — *секунда* (с) и единица массы — *килограмм* (кг). Кроме этих трех единиц, СИ принимает в качестве основных единицу силы тока — *ампер* (А), единицу температуры — *кельвин* (К), единицу силы света — *канделу* (кд) и единицу количества вещества — *моль* (моль). Об этих единицах речь пойдет в соответствующих разделах курса.

Выбранные основные единицы соответствующим образом связаны с эталонными величинами, о которых мы говорили ранее. Так, метр равен длине пути, проходимого светом в вакууме за $1/299\,792\,458$ с (приблизительно метр равен $1/40\,000\,000$ доле длины земного экватора). Секунда равна сумме $9\,192\,631\,770$ периодов электромагнитных волн определенного типа, испускаемых атомом цезия (приблизительно она равна $1/86\,400$ средних солнечных суток). Наконец, килограмм совпадает с эталоном массы, то есть равен массе платино-иридиевого стержня, хранящегося в Международном бюро мер и весов во Франции. Эта масса близка к массе $1\,000\text{ см}^3$ чистой воды. На практике применяются также кратные единицы: километр ($1\text{ км} = 10^3\text{ м}$), сантиметр ($1\text{ см} = 10^{-2}\text{ м}$), миллиметр ($1\text{ мм} = 10^{-3}\text{ м}$), микрометр ($1\text{ мкм} = 10^{-6}\text{ м}$); минута, равная 60 секундам; один грамм, равный $1/1\,000$ килограмма и т. д. В исследовательской практике иногда пользуются внесистемными единицами, удобными для каких-либо конкретных задач, например: 1 ангстрем (\AA) $= 10^{-10}\text{ м}$; 1 электронвольт (эВ) $\approx 1,6 \cdot 10^{-19}\text{ Дж}$; 1 парсек (пк) $\approx 3,1 \cdot 10^{16}\text{ м}$ и т. д. В зависимости от порядка величины производных единиц для их наименования используются приставки: милли- $= 10^{-3}$, микро- $= 10^{-6}$, нано- $= 10^{-9}$, пико- $= 10^{-12}$, фемто- $= 10^{-15}$; кило- $= 10^3$, мега- $= 10^6$, гига- $= 10^9$, тера- $= 10^{12}$ и т. д., а для обозначения — соответственно: м, мк, н, п, ф; к, М, Г, Т.

Единицы измерения всех *системных* физических величин являются производными от основных единиц. Так, за единицу скорости в СИ принимается скорость равномерно движущегося тела, проходящего в единицу времени (секунду) путь, равный единице длины (метр). Обозначается эта единица как м/с. За единицу ускорения принимается ускорение равномерно-переменного движения, при котором скорость тела за единицу времени (секунду) меняется на единицу скорости (на 1 м/с). Эта единица в СИ обозначается как м/с². Единица силы в СИ определяется соотношением (3.6): 1 Н равен силе, под действием которой тело с массой 1 кг получает ускорение 1 м/с^2 .

В тех случаях, когда единице измерения присвоено имя собственное, его сокращение принято писать с большой буквы, и не только как таковое, но и в составных сокращениях. Например: 1 ньютон = 1 Н, 1 милливольт = 1 мВ, 1 мегаджоуль = = 1 МДж и т. д.

Кроме системы единиц измерения СИ, в науке и технике используются иногда и другие системы единиц. В научной практике часто применяется так называемая СГС-система. Основными единицами этой системы являются сантиметр, грамм и секунда. Единица силы в СГС-системе называется *диной* (дин). Одна дина равна силе, под действием которой тело с массой 1 г получает ускорение 1 см/с^2 . Между ньютоном и диной имеется следующее соотношение:

$$1 \text{ Н} = 1 \text{ кг} \cdot 1 \text{ м/с}^2 = 10^3 \text{ г} \cdot 10^2 \text{ см/с}^2 = 10^5 \text{ дин.}$$

Нетрудно видеть, что изменение основных единиц влечет за собой изменение производных единиц. Для того чтобы охарактеризовать эту связь производных и основных единиц измерения, в физике вводится понятие *размерности физической величины*, которая показывает, как изменяется значение единицы измерения этой величины при изменении основных единиц. Для обозначения размерности произвольной величины используется ее буквенное обозначение, взятое в квадратные скобки. Так, например, символ $[v]$ означает размерность скорости. Размерность основных величин обозначается специальным образом: размерность длины — L , времени — T , массы — M . Таким образом, обозначив длину буквой l , время буквой t , массу буквой m , можно написать:

$$[l] = L; \quad [m] = M; \quad [t] = rmT.$$

Какова, например, размерность скорости? Модуль скорости определяется соотношением $v = \Delta s / \Delta t$ (для сколь угодно малых Δt). Так как физические определения и законы не могут зависеть от выбора единиц измерения фигурирующих в них величин, размерности обеих частей уравнений, выражающих эти законы, должны быть одинаковы. Размерность Δs равна L , размерность Δt равна T .

Следовательно, размерность скорости равна

$$[v] = LT^{-1}.$$

Последняя запись означает, что при увеличении единицы длины в n_1 раз единица измерения скорости увеличится в n_1 раз, а соответствующее число, которым выражается скорость в этих единицах, уменьшится в n_1 раз. А при увеличении единицы времени в n_2 раза единица измерения скорости уменьшится в n_2 раза, а выражающее скорость число увеличится в n_2 раза. На-

пример, пусть задано значение скорости $v = 10$ м/с, а мы хотим представить ее в единицах км/ч. В этом случае $n_1 = 1\,000$, а $n_2 = 3\,600$. В результате в новых единицах измерения значение скорости станет равным

$$v = 10 \cdot (3\,600/1\,000) \text{ км/ч} = 36 \text{ км/ч}.$$

Аналогично скорости можно установить размерность ускорения:

$$a = \frac{[\Delta v]}{[\Delta t]} = \text{LT}^{-1}/\text{T} = \text{LT}^{-2}.$$

Размерность силы (см. (3.6)):

$$[F] = [m][a] = \text{MLT}^{-2}.$$

Аналогичным образом устанавливаются размерности всех прочих величин. В каждом конкретном случае «инструментом» для введения новой единицы служит физический закон, в котором впервые появляется соответствующая величина. Обратим внимание еще раз на методологическую проблему, которую в традиционном изложении порождает формула (3.2).

Отметим, что контроль размерности физических формул является мощным инструментом проверки правильности проведенных вычислений. Более того, в современной физике (а прежде всего — как раз в механике) на этой идее основаны некоторые теоретические методы получения новой информации (точнее, они базируются на законах подобия или скейлинге — родственных понятиях более высокого уровня). По этой причине в физических формулах обычно не принято доводить сокращение до максимально тривиального вида с точки зрения алгебры, вместо этого предпочтительно формировать сомножители с наиболее выразительной размерностью или безразмерные отношения.

3.4. Понятие импульса. Третий закон Ньютона

Закон (3.4) описывает не только одномерное, но сколь угодно сложное движение материальной точки со всеми особенностями траектории, обсуждавшимися в предыдущей главе. Он, конечно же, информативнее «школьной» формулы (3.2), потому что прямо дает в наши руки дифференциальное уравнение, решая которое при надлежащих начальных или граничных условиях, мы можем в принципе определить траекторию $\mathbf{r}(t)$.

Но оказывается, что с физической точки зрения предпочтительна не математически совершенная форма второго закона Ньютона (3.4), а несколько иная:

$$\frac{d\mathbf{p}}{dt} = \mathbf{F}, \quad \text{где} \quad \mathbf{p} \equiv m\mathbf{v}. \quad (3.7)$$

Величина \mathbf{p} называется *импульсом* материальной точки (старое название, до сих пор употребительное в теоретической механике — *количество движения*). Если записать (3.7) в приближенной форме конечных малых приращений,

$$\Delta \mathbf{p} \approx \mathbf{F} \Delta t,$$

то величина, стоящая в правой части, именуется *импульсом силы* за время Δt . В настоящее время этот термин употребляется редко.

Представим теперь любое макроскопическое тело (твердое, жидкое, сыпучее) как совокупность материальных точек и присвоим каждой из них индекс суммирования, обозначив его в общем случае буквой i . Тогда для тела в целом из (3.7) следует:

$$\frac{d\mathbf{p}}{dt} = \sum_i \mathbf{F}_i, \quad \text{где} \quad \mathbf{p} \equiv \sum_i m_i \mathbf{v}_i. \quad (3.8)$$

Опыт (включая прецизионные эксперименты) показывает, что соотношения (3.7), (3.8) справедливы не только тогда, когда они тривиальным образом следуют из (3.4), но и в тех случаях, когда приходится иметь дело с телами переменной массы. Более того, (3.7), (3.8) могут быть перенесены в релятивистскую механику, где закон (3.4) уже не имеет места. А в квантовой механике, где в принципе отсутствует понятие скорости, понятие импульса сохраняется и играет роль одного из самых фундаментальных. Наряду с энергией, которую мы введем в гл. 4, импульс принадлежит универсальному языку физики, который может быть использован безотносительно к приближению, в котором мы работаем. Именно на этом языке формулируется обобщение третьего закона Ньютона (3.3) на случай любой, а не только ньютоновской, механики.

Вводя понятие инерциальной системы отсчета, мы опирались на идеализированную модель уединенного физического тела. Мы предполагали при этом, что означенное тело можно либо смоделировать материальной точкой, либо представить его в виде жестко связанной совокупности таковых. Это, конечно, модель достаточно грубая — по крайней мере, атомы, из которых состоит любое макроскопическое тело, отнюдь не являются статическими системами, так как движение электронов в атоме или нуклонов в ядре — свойства неустранимые. Рассмотрим систему более свободную. Пусть материальные точки не будут, вообще говоря, жестко связаны друг с другом. Они в общем случае все взаимодействуют друг с другом, но иногда этим взаимодействием можно с заданной точностью пренебречь, а иногда, напротив, образуют какие-то связанные ансамбли — систему макроскопических тел. Таким образом мы переходим от идеи абсолютно изолированного тела к идее *замкнутой системы физических тел* или, что то же, *изолированной системы материальных точек*.

Фундаментальным законом природы является закон сохранения импульса: *векторная сумма импульсов всех тел замкнутой системы в инерциальной системе отсчета не изменяется со временем*, хотя каждый из импульсов по отдельности может с течением времени испытывать существенные изменения. Если замкнутая система состоит из N тел, то закон сохранения импульса можно сформулировать в виде соотношения

$$\sum_{i=1}^N \mathbf{p}_i(t) = m_1 \mathbf{v}_1(t) + m_2 \mathbf{v}_2(t) + \dots + m_N \mathbf{v}_N(t) = \text{const.} \quad (3.9)$$

Значение входящей в закон сохранения (3.9) константы (инварианта) может быть различным, и оно определяется начальными условиями, то есть значениями импульсов всех тел в момент образования замкнутой системы тел.

Покажем теперь, что закон сохранения импульса, во-первых, дает возможность измерить массу любого тела путем сравнения ее с массой выбранного эталонного тела, и, во-вторых, позволяет выяснить физический смысл массы как особого свойства всех частиц нашего мира. Для этого рассмотрим простейший случай, когда замкнутая система состоит всего из двух взаимодействующих тел. Тогда закон сохранения импульса (3.9) принимает вид

$$m_1 \mathbf{v}_1(t) + m_2 \mathbf{v}_2(t) = \text{const.} \quad (3.10)$$

Характер взаимодействия между телами не играет роли — это может быть звезда, вокруг которой вращается единственная планета, или два различных сталкивающихся резиновых мячика, воздействием на которые со стороны других тел можно по каким-то причинам пренебречь. Интересно отметить, что векторное равенство (3.10) может быть изображено в виде треугольника (см. рис. 1.3б), т. е. в любой системе отсчета векторы скоростей частиц в любой момент времени лежат в одной плоскости.

Рассмотрим два момента времени: один — непосредственно перед столкновением, а другой — сразу после столкновения. Так как суммарный импульс при столкновении сохраняется, то мы получаем из (3.10)

$$m_1 \mathbf{v}_{1(\text{до})} + m_2 \mathbf{v}_{2(\text{до})} = m_1 \mathbf{v}_{1(\text{после})} + m_2 \mathbf{v}_{2(\text{после})},$$

где слева стоят значения скоростей до столкновения, а справа — непосредственно после столкновения.

Последнее равенство удобнее представить как

$$-m_1[\mathbf{v}_{1(\text{после})} - \mathbf{v}_{1(\text{до})}] = m_2[\mathbf{v}_{2(\text{после})} - \mathbf{v}_{2(\text{до})}],$$

или в более компактной форме:

$$-m_1 \Delta \mathbf{v}_1 = m_2 \Delta \mathbf{v}_2. \quad (3.11)$$

В последнем соотношении $\Delta \mathbf{v}_1$ и $\Delta \mathbf{v}_2$ означают изменения скоростей соответствующих тел в результате их столкновения. Эти изменения скоростей являются, как и сами скорости, векторными величинами, и направлены, как следует из (3.11), по одной прямой в противоположные стороны. Для модулей этих изменений скоростей Δv_1 и Δv_2 получаем из (3.11) равенство

$$m_1 \Delta v_1 = m_2 \Delta v_2. \quad (3.12)$$

Важно отметить, что величины Δv_1 и Δv_2 могут быть измерены опытным путем.

А теперь представим себе, что массу одного из сталкивающихся тел (например, m_2) мы условились считать эталоном массы ($m_2 = m_\varepsilon$). Проводя тогда наблюдения за столкновением произвольного тела с эталонным телом, мы всегда сможем выразить с помощью соотношения (3.12) его неизвестную до этого массу m_x через эталонную массу m_ε и через измеренные на опыте изменения скоростей:

$$m_x = m_\varepsilon \frac{\Delta v_\varepsilon}{\Delta v_x}.$$

На практике массу тела чаще измеряют, используя другие соотношения. Важная роль закона сохранения импульса для определения массы тела состоит в том, что при этом определении не требуется вводить никаких других величин, которые были бы нам до этого неизвестны. Необходимо лишь условиться об эталонной единице массы.

Обсудим еще раз вопрос о том, какое физическое свойство тел характеризует их масса — свойство, которое бы находило отражение в наших повседневных контактах с различными предметами. Из соотношения (3.12) следует, что чем больше масса тела, тем меньше оно изменяет свою скорость в процессе взаимодействия. Хорошо всем знакомое свойство тел оказывать сопротивление при попытках привести его в движение или изменить величину или направление его скорости называется «инертностью». И мы видим, что чем больше масса тела, тем сильнее оно «сопротивляется» изменению его скорости. Таким образом, масса характеризует «инерционную способность» тела, и можно, следовательно, сказать, что *масса тела — это мера инертности тела, в принципе она может быть измерена на основе закона сохранения импульса*. Во всяком случае, это единственный способ «взвешивания» элементарных частиц.

Последним по порядку, но не по значению, будет вывод из закона сохранения импульса классической формы третьего закона Ньютона (3.3). Рассмотрим замкнутую систему двух тел 1 и 2. Продифференцируем по времени условие постоянства их полного

импульса $\mathbf{p}_1(t) + \mathbf{p}_2(t) = \text{const}$:

$$\dot{\mathbf{p}}_1(t) + \dot{\mathbf{p}}_2(t) = 0,$$

а далее, учитывая для каждого из них закон (3.7), приходим к окончательному результату:

$$\mathbf{F}_{12} = -\mathbf{F}_{21},$$

где \mathbf{F}_{12} — сила, действующая на тело 1 со стороны тела 2, и наоборот. Мы смогли указать точный «адрес» каждой силы, поскольку тел в нашей системе всего два. Поэтому то, что нами получено, еще не полностью эквивалентно (3.3). Доказательство в случае произвольного числа тел не столь математически очевидно; его проще оказывается провести «в обратную сторону»:

$$\frac{d}{dt} \sum_i \mathbf{p}_i = \sum_i \dot{\mathbf{p}}_i = \sum_i \sum_{j \neq i} \mathbf{F}_{ij} = 0, \quad \text{если} \quad \mathbf{F}_{ij} = -\mathbf{F}_{ji}.$$

(Условие второго суммирования $j \neq i$ подразумевает, что тело само себя ни тянуть, ни толкать в инерциальной системе отсчета не может, т. е. $\mathbf{F}_{ii} = 0$).

Хотя закон сохранения импульса, как мы уже отметили, более универсален с точки зрения современной физики, традиционный третий закон Ньютона (3.3) в классической механике также занимает достойное место, особенно в статике — науке о равновесии физических тел и систем тел.

Вопросы

1. Сформулируйте закон сохранения импульса.
2. Что такое масса и что такое сила?
3. Какой тип фундаментальных взаимодействий лежит в основе сил трения?
4. Что такое вес тела?
5. Какова размерность импульса?
6. Можно ли вывести закон инерции (3.1) из второго закона Ньютона (3.2)?
7. Является ли масса скалярной величиной?
8. Что общего между абсолютно свободным телом и замкнутой системой тел?

Глава 4

ПРИМЕРЫ ПРИЛОЖЕНИЙ ЗАКОНОВ НЬЮТОНА

4.1. Исследование закона движения материальной точки

Исследование закона движения означает получение временной зависимости всех координат, определяющих положение тела в пространстве. В частности, для материальной точки это три координаты, определяющие зависимость $\mathbf{r}(t)$. Как мы уже знаем, второй закон Ньютона в формулировке (3.4) представляет собой уравнение, содержащее не сам радиус-вектор, а его вторую производную по времени. И если зависимость силы от положения материальной точки известна, то есть известна зависимость \mathbf{F} от \mathbf{r} , то соотношение (3.4) представляет собой *дифференциальное уравнение*, где подлежащая определению функция $\mathbf{r}(t)$ входит под знаком производной. Если нам известна зависимость действующей на тело силы от его положения в пространстве и если мы знаем положение тела и его скорость в какой-то момент времени, то мы можем с помощью уравнения (3.4) вычислить его положение и скорость в любой из последующих моментов времени. Следовательно, второй закон Ньютона дает возможность предсказать всю траекторию движения материальной точки в зависимости от времени, если известны начальные условия: начальное значение радиуса-вектора $\mathbf{r}(0)$ и начальное значение скорости $\mathbf{v}(0)$.

Заметим, что бросая камень или мяч в цель, человек интуитивно использует приобретенную из опыта информацию о связи начального положения руки и ее начальной скорости с последующей траекторией полета брошенного предмета. Это приобретенное на опыте знание человечество использовало с первых шагов своей эволюции, но только после формулировки второго закона Ньютоном в XVII столетии у человека появилась возможность перейти от интуитивного предсказания характера движения тел к точному расчету, и это послужило одной из причин резкого ускорения технического прогресса на нашей планете.

А теперь покажем, как второй закон «работает» в качестве инструмента для вычисления траекторий движения. Пусть перед нами стоит задача вычислить в зависимости от времени

в некоторой инерциальной системе траекторию движения тела массы m , которое можно рассматривать как материальную точку (рис. 4.1). Величина и направление силы, действующей на тело в каждой точке пространства, предполагаются известными, то

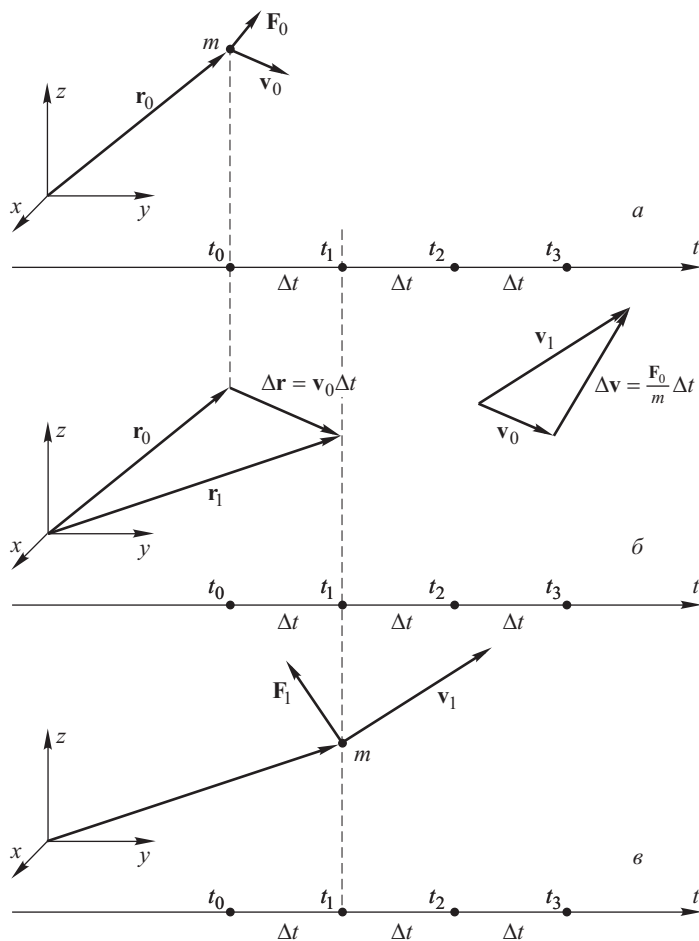


Рис. 4.1

есть известна $\mathbf{F}(\mathbf{r})$ — зависимость этой силы от радиуса-вектора тела. Такая ситуация встречается очень часто. Один из самых простых примеров — движение вблизи поверхности Земли под действием силы тяжести, которая везде равна по величине mg и направлена к поверхности.

Начальные условия для движения рассматриваемого тела заключаются в том, что в некоторый начальный момент времени

t_0 нам, во-первых, известно положение тела в пространстве (то есть его начальный радиус-вектор \mathbf{r}_0) и, во-вторых, известна его начальная скорость \mathbf{v}_0 . Эта начальная ситуация изображена на рис. 4.1 а, где также представлена шкала времени, которое мы разбили на малые интервалы Δt , разделяющие следующие друг за другом моменты t_1, t_2, t_3, \dots и так далее. Будем подразумевать, что интервал Δt будет в конечном итоге выбираться сколь угодно малым, то есть устремлен к нулю. Это позволит нам использовать физические величины (скорость, ускорение), выраженные через производные по времени от соответствующих функций. Но для наглядности будем пока представлять Δt просто как очень малый временной интервал.

Теперь покажем сначала, что, зная $\mathbf{r}_0, \mathbf{v}_0$ и значение действующей на тело в точке \mathbf{r}_0 силы \mathbf{F}_0 , мы можем с помощью второго закона Ньютона определить \mathbf{r}_1 и \mathbf{v}_1 — *новое положение тела* и его *новую скорость* через интервал времени Δt . Учитывая, что мы подразумеваем под Δt сколь угодно малый интервал, перепишем определение скорости в виде

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \frac{\Delta \mathbf{r}}{\Delta t}. \quad (4.1)$$

Если применить это соотношение в точке \mathbf{r}_0 , то из него мы сразу получаем $\Delta \mathbf{r}_0$ — изменение начального радиуса-вектора тела за интервал Δt , то есть

$$\Delta \mathbf{r}_0 = \mathbf{v}_0 \Delta t.$$

А зная это изменение начального положения тела, мы можем найти его *новое положение*, то есть новое значение радиуса-вектора в момент t_1 (см. рис. 4.1 б):

$$\mathbf{r}_1 = \mathbf{r}_0 + \Delta \mathbf{r}_0 = \mathbf{r}_0 + \mathbf{v}_0 \Delta t. \quad (4.2)$$

Найдем теперь с помощью второго закона новое значение скорости \mathbf{v}_1 в момент t_1 . Используем соотношения

$$m \frac{d\mathbf{v}}{dt} = \mathbf{F}, \quad \mathbf{v} = \frac{d\mathbf{r}}{dt} \quad (4.3)$$

в точке \mathbf{r}_0 в момент t_0 и определим из них приращение скорости:

$$\Delta \mathbf{v}_0 = \frac{\mathbf{F}_0}{m} \Delta t. \quad (4.4)$$

Зная $\Delta \mathbf{v}_0$, получаем для *нового значения скорости* в точке \mathbf{r}_1 в момент t_1 :

$$\mathbf{v}_1 = \mathbf{v}_0 + \Delta \mathbf{v}_0 = \mathbf{v}_0 + \frac{\mathbf{F}_0}{m} \Delta t. \quad (4.5)$$

Итак, мы сумели, стартуя с начального положения тела, найти его новое положение на траектории \mathbf{r}_1 и новую скорость \mathbf{v}_1 в момент t_1 . Но теперь мы пришли к ситуации, полностью аналогичной начальной (см. рис. 4.1 в). И так как значение

силы в новой точке \mathbf{r}_1 нам известно, то, продолжая вычисления точно тем же способом, мы шаг за шагом найдем положение и скорость тела во все последующие моменты времени t_2, t_3, \dots и так далее. Последнее и означает, что, зная начальное положение и начальную скорость материальной точки, можно с помощью второго закона Ньютона вычислить, то есть предсказать, всю последующую траекторию ее движения. На практике вычисление траектории осуществляют одним из двух способов. Во-первых, можно использовать компьютер, «поручив» ему проделывать всю ту процедуру последовательного нахождения точек траектории, о которой мы говорили выше. Для этого компьютеру надо только задать величину «временного шага» Δt и ввести в компьютер программу вычислений, определяемых соотношениями (4.1)–(4.5). Во многих случаях, однако, можно обойтись без компьютера. Мы говорили уже, что второй закон Ньютона является дифференциальным уравнением, конкретный вид которого зависит от физических условий задачи (фактически, от характера зависимости действующей на тело силы от его положения в пространстве — $\mathbf{F}(\mathbf{r})$). Решение такого уравнения часто удается найти, используя хорошо известные правила интегрирования и дифференцирования, а в некоторых случаях (составляющих большую часть задач в студенческих задачниках) даже с помощью элементарной алгебры. Некоторые примеры таких решений мы рассмотрим в последующих параграфах.

4.2. Движение материальной точки под действием постоянной силы

Перейдем сейчас от обсуждения общих свойств второго закона Ньютона как средства для вычисления траекторий к рассмотрению конкретных задач и конкретных движений, которые можно детально исследовать с помощью этого основного уравнения механики. Как и при изучении любого явления, разумно начать с рассмотрения таких движений, которые, с одной стороны, допускали бы относительно простое математическое решение задачи, а, с другой, имели бы достаточно большое практическое значение.

Закон движения тела, то есть зависимость его радиуса-вектора \mathbf{r} от t , определяется характером действующей на него силы и является, согласно второму закону Ньютона, решением уравнения (4.3).

В случае, когда на тело никакие силы не действуют, то есть $\mathbf{F} = 0$, решением уравнения (4.3) является движение с постоянной скоростью $\mathbf{v} = \text{const}$. Это простейший тип движения — свободное движение. Следующий по сложности тип движения —

это такое движение, при котором действующая на тело сила хотя и не равна нулю, но не меняется по величине и направлению при перемещении тела. Это означает, что правая часть первого из уравнений (4.3) является константой. Любое уравнение решается проще, когда входящие в него коэффициенты остаются постоянными, не зависящими от времени и координат, так что этот тип движения наверняка окажется наименее сложным с точки зрения его математического исследования. С другой стороны, движение под действием постоянной силы является очень важным с практической точки зрения и часто встречается в повседневной жизни.

Прежде всего, к такому типу движения относится при определенных условиях движение под действием силы тяжести. Сила тяжести, как и любая сила, является векторной величиной. Примем упрощающее предположение, что ее модуль постоянен и равен mg . Но так как эта сила направлена к центру Земли, то ее направление в разных точках земной поверхности различно. Однако при исследовании движений тел, перемещающихся на расстояния, которые намного меньше радиуса Земли ($R_3 \approx 6\,000$ км), можно пренебречь кривизной земной поверхности и с хорошей точностью считать, что сила тяжести не меняет своего направления, оставаясь перпендикулярной этой поверхности. В этих условиях сила тяжести может рассматриваться постоянной как по модулю, так и по направлению. Помимо силы тяжести, с постоянными силами приходится часто сталкиваться при рассмотрении работы различных технических устройств, когда их различные детали испытывают действие постоянных сил со стороны других деталей.

Мы подробно разберем случай движения тела под действием постоянной силы на конкретном примере определения закона движения брошенного камня или снаряда, вылетевшего из ствола орудия. Какой вид имеет траектория камня? От чего зависит дальность полета? Древнегреческий философ и ученый Аристотель (384–322 до н. э.) утверждал, например, что на начальном участке траектория брошенного под углом к вертикали тела является прямой линией, и это, вроде бы, подтверждалось непосредственными наблюдениями. Потребовалось почти два тысячелетия, чтобы понять, что траектория на самом деле является криволинейной на всех участках полета.

Изучение движения брошенного под углом к вертикали тела включает в себя несколько этапов, характерных для решения большинства задач механики.

Первый этап — определение типа движения. В нашем случае это означает выяснение вопроса о том, можем ли мы пренебречь влиянием размеров тела при решении задачи, то есть, можно ли рассматривать его движение, как движение материаль-

ной точки, и, следовательно, применять второй закон Ньютона (4.3). В общем случае на брошенное тело, помимо силы тяжести, действует сила, порожденная обтеканием воздуха, которое зависит от размеров и формы тела. Эта зависимость может сильно усложнить решение задачи и на практике может приводить к весьма замысловатым траекториям полета тела. Вспомним хотя бы о движении изобретенного австралийскими аборигенами бумеранга, который после броска возвращается к охотнику. Однако во многих случаях эффектами обтекания можно с достаточной точностью пренебречь. Именно такой случай мы и будем рассматривать, то есть мы воспользуемся для решения основным уравнением движения материальной точки (4.3).

Второй этап — физическая формулировка задачи: выбор системы отсчета, определение действующих сил и начальных условий. Рассмотрим случай, когда тело начинает движение в некоторой точке на поверхности Земли. В качестве системы координат выберем декартову систему, начало которой естественно поместить в эту точку. Оси Ox , Oy , Oz расположим так, чтобы ось Oz была перпендикулярна поверхности, а вектор начальной скорости \mathbf{v}_0 лежал, как на рис. 4.2а, в плоскости zOx . Начальный модуль скорости v_0 и ее начальное направление, то есть угол α между \mathbf{v}_0 и осью Ox , определяются теми условиями и причинами, которые приводят камень или снаряд в движение. Это, например, расположение и скорость движения руки при бросании камня или наклон орудия и мощность порохового заряда.

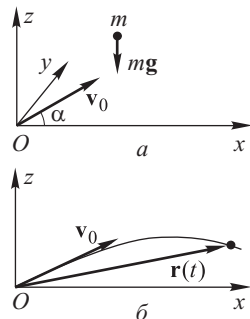


Рис. 4.2

Система отсчета, связанная с поверхностью Земли, не является, строго говоря, инерциальной, так как любая точка поверхности движется с ускорением, обусловленным вращением Земли вокруг своей оси и вокруг Солнца. Но для многих практических задач этот эффект «неинерциальности» является несущественным, и мы будем полагать, что и в нашей задаче этим эффектом можно пренебречь и считать выбранную систему отсчета инерциальной. В инерциальной системе отсчета справедлив второй закон Ньютона (4.3), где теперь под \mathbf{F} подразумевается постоянная сила тяжести. Мы изобразили эту силу на рис. 4.2а для некоторого произвольного момента времени после начала движения, поместив тело массы m в некоторой произвольной точке над поверхностью. Истинное положение тела в различные моменты времени, то есть траекторию его движения, мы сможем определить только после окончательного решения задачи.

Третий этап — математическая формулировка задачи: запись уравнений, соответствующих физической формулировке. Уравнение (4.3) содержит в качестве неизвестных векторные величины $\mathbf{r}(t)$ и $\mathbf{v}(t)$. Поэтому оно фактически представляет собой совокупность трех уравнений для трех проекций вышеупомянутых величин. Для проекций радиуса-вектора тела введем обозначения: $r_x = x$, $r_y = y$, $r_z = z$. Взяв проекции на оси координат от левой и правой частей уравнения (4.3), мы получаем три уравнения:

$$m \frac{dv_x}{dt} = 0, \quad v_x = \frac{dx}{dt}; \quad x(0) = 0, \quad v_x(0) = v_0 \cos \alpha, \quad (4.6)$$

$$m \frac{dv_y}{dt} = 0, \quad v_y = \frac{dy}{dt}; \quad y(0) = 0, \quad v_y(0) = 0, \quad (4.7)$$

$$m \frac{dv_z}{dt} = -mg, \quad v_z = \frac{dz}{dt}; \quad z(0) = 0, \quad v_z(0) = v_0 \sin \alpha. \quad (4.8)$$

Справа от каждого из уравнений записаны начальные условия, являющиеся неотъемлемыми элементами физической и математической формулировки задачи. Знак «минус» перед mg в последнем уравнении отражает тот факт, что сила тяжести направлена в отрицательном направлении оси Oz .

Четвертый этап — математическое решение задачи. Теперь предстоит чисто «техническая» работа — решение сформулированных ранее уравнений с помощью известных в математике приемов и методов. В произвольном случае уравнения могут оказаться столь сложными, что их решение невозможно без привлечения компьютерной техники. В нашем же, относительно простом случае, решение уравнений (4.3), удовлетворяющее начальным условиям, может быть найдено на основе элементарных знаний о правилах вычисления производных от простейших функций.

Рассмотрим, например, уравнение (4.8). Составляющая скорости v_z имеет вид: $v_z = \text{const} - gt$. Константу определяем из условия $v_z(0) = v_0 \sin \alpha$. Интегрируем еще раз: $z(t) = \text{const} + v_0 t \sin \alpha - gt^2/2$. Новую константу определяем из условия $z(0) = 0$. Аналогично находятся решения уравнений (4.6), (4.7), так что окончательно решение выглядит следующим образом:

$$x(t) = v_0 t \cos \alpha, \quad y(t) = 0, \quad z(t) = v_0 t \sin \alpha - \frac{gt^2}{2}. \quad (4.9)$$

Найденные выражения определяют зависимость от времени всех трех проекций радиуса-вектора $\mathbf{r}(t)$ тела, движущегося под действием силы тяжести. Тем самым задача о нахождении траектории движения решена. Для того чтобы ответить на вопрос, к какому типу геометрических кривых (гипербола, парабола и т. п.) относится найденная траектория, достаточно выразить t через x в первом из равенств (4.9) и подставить результат в вы-

ражение для $z(t)$. Это дает следующее уравнение траектории в плоскости zOx :

$$z = \operatorname{tg} \alpha \cdot x - \frac{g}{2v_0^2 \cos^2 \alpha} x^2. \quad (4.10)$$

Из геометрии известно, что это соотношение представляет собой уравнение параболической кривой, и следовательно, ни на одном из участков полета тела его траектория не является прямой линией (см. рис. 4.2 б).

Из соотношения (4.10) можно определить, например, дальность полета тела x_{Π} . При падении на поверхность $z = 0$, и из этого условия находим с помощью (4.10)

$$x_{\Pi} = \frac{v_0^2}{g} \sin 2\alpha. \quad (4.11)$$

Пятый этап — проверка полученного решения. Этот, последний, этап является неотъемлемым и очень важным элементом решения всякой физической задачи. Самым очевидным способом проверки является поэтапное повторение всего процесса решения. Но помимо этого существуют специальные приемы, позволяющие быстро проверить полученный ответ на наличие в нем грубых ошибок. Наиболее известны два таких приема.

Прием первый — проверка ответа по размерности. Речь здесь идет о том, что полученный результат должен удовлетворять правилам размерности, о которой мы говорили в предыдущей главе. Если результатом вычисления является, например, скорость, то и ответ должен иметь размерность скорости. Если вычислялась сила, то ответ должен иметь размерность силы и т. п. В нашем случае размерность полученного ответа правильна:

$$[x_{\Pi}] = [v_0^2]/[g] = (L^2 T^{-2})/(L T^{-2}) = L.$$

Прием второй — проверка ответа по заранее очевидным результатам. Часто бывает так, что при некоторых значениях входящих в задачу физических величин ответ можно «предсказать» заранее, так как он является очевидным. Как правило, результат очевиден для предельных значений каких-либо величин, например, при нулевых или бесконечных значениях времени или расстояния, при нулевой или бесконечной массе и так далее. Проверим этим способом тот же результат (4.11) для дальности полета тела. Очевидно, например, что при нулевой начальной скорости v_0 тело останется в исходном положении и дальность полета x_{Π} должна равняться нулю. Мы видим, что найденное нами решение удовлетворяет этому условию. Очевидно также, что дальность должна равняться нулю, если начальная скорость направлена строго вертикально, и мы видим в (4.11), что действительно, $x_{\Pi} = 0$ при $\alpha = \pi/2$. Наконец, очевидно, что в отсутствие силы тяжести (то есть при $g = 0$) дальность полета должна

быть бесконечной, и мы видим, что наш ответ удовлетворяет и этому условию. Как и в случае с размерностью, проведенная выше проверка не дает стопроцентной гарантии отсутствия ошибки (хотя бы, например, в численных коэффициентах). Но она, во всяком случае, страхует от совершения грубых ошибок.

4.3. Реактивное движение

Во многих случаях траекторию и временные характеристики движения материальной точки можно установить, вообще не обращаясь ко второму закону Ньютона. Примером такой ситуации может быть *реактивное движение*, которое осуществляется за счет выброса части общей массы тела в каком-либо направлении. Именно так движутся, например, космические корабли и обычные осветительные ракеты.

Рассмотрим задачу о реактивном движении для наиболее простого случая, когда ракета движется в дальнем космосе, так что воздействием на нее Земли и других планет можно пренебречь. Решение всякой задачи о движении начинается, как мы уже знаем, с выбора системы отсчета. В качестве таковой мы выберем некоторую инерциальную систему, конкретная привязка которой для данной задачи несущественна — это может быть система, связанная, например, с Солнцем или с какой-либо другой

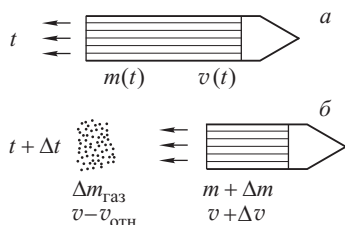


Рис. 4.3

звездой. На рис. 4.3 *а* схематически изображена летящая ракета в некоторый момент времени t , когда ее полная масса (масса корпуса и масса находящегося в ней топлива) равна $m(t)$, а скорость равна $v(t)$ и направлена вдоль положительного направления оси x декартовой системы координат. Предполагается, что топливо (в виде раскаленных газов) выбрасывается из задней части ракеты таким образом, что ракета и выброшенное топливо движутся по одной прямой, совпадающей с осью x .

Скорость выброса топлива относительно корпуса ракеты $v_{\text{отн}}$ считается известной — она определяется конструктивными особенностями и зависит от типа топлива, формы сопла, температуры сгорания топлива в сопле и т. д. Будем далее предполагать, что эта скорость $v_{\text{отн}}$ остается во время полета постоянной. Наша цель состоит в том, чтобы определить, по какому временному закону будет перемещаться ракета и как будут меняться со временем ее скорость и полная масса. После выбора системы отсчета и задания физических условий задачи необходимо выяснить, можем ли мы рассматривать движение ракеты как дви-

жение материальной точки, то есть считать его не зависящим от размеров ракеты. Только в этом случае мы сможем использовать, например, закон сохранения импульса. В нашем случае это приближение будет справедливо при условии, что ракета или реактивный снаряд не будет «кувыркаться» в полете. Не останавливаясь на формальном описании этого свойства (обусловленного динамикой твердого тела конечных размеров), ограничимся ссылкой на то обстоятельство, что аппарат, сконструированный должным образом, в полете устойчив.

Чтобы понять, какие именно законы движения нам следует использовать, остановимся коротко на принципе реактивного движения. Этот принцип очень прост. Ракета воздействует определенным образом на выбрасываемое из сопла вещество (газы). Выбрасываемое вещество, в свою очередь, воздействует на ракету и увеличивает ее скорость в противоположном направлении. Если воздействием всех других тел можно, как в нашем случае, пренебречь, то ракета вместе с выброшенным веществом является замкнутой системой. Суммарный импульс такой системы не меняется во времени, и именно закон сохранения импульса лежит в основе решения нашей задачи.

Предположим, что за малый промежуток времени Δt масса и скорость ракеты получают приращения Δm и Δv (величина Δm отрицательна). На рис. 4.3б изображена ситуация к моменту времени $t + \Delta t$: ракета имеет к этому моменту полную массу $m + \Delta m$ и скорость $v + \Delta v$, в то время, как выброшенная за время Δt масса газа $\Delta m_{\text{газ}}$ летит относительно выбранной системы отсчета со скоростью $v_{\text{газ}} = v - v_{\text{отн}}$. Такая величина получается по правилу сложения скоростей, если учесть, что скорость ракеты направлена в положительном направлении оси x , а скорость выброшенного топлива — в отрицательном.

Из закона сохранения импульса следует, что суммарный импульс ракеты и топлива имеет одно и то же значение в моменты t и $t + \Delta t$, то есть

$$mv = (m + \Delta m)(v + \Delta v) + \Delta m_{\text{газ}} v_{\text{газ}}. \quad (4.12)$$

Если устремить промежуток времени Δt , а с ним и приращения Δm и Δv к нулю, то в (4.12) можно отбросить произведение $\Delta m \cdot \Delta v$ как величину малую по сравнению со слагаемыми, содержащими малые приращения в первой степени (как бесконечно малую высшего порядка). Ввиду сохранения массы $\Delta m + \Delta m_{\text{газ}} = 0$. Пользуясь этим, получаем из (4.12)

$$dv = -v_{\text{отн}} \frac{dm}{m}, \quad (4.13)$$

где мы при переходе к бесконечно малым пределам заменили, как

принято в математике, конечные приращения на дифференциалы соответствующих величин.

Равенство (4.13) — это соотношение между бесконечно малыми величинами. Для того чтобы получить соотношение между конечными, доступными измерению, значениями скорости ракеты и ее массы, следует просуммировать левую и правую части равенства (4.13). Такое суммирование бесконечно малых выражается, как известно, операцией интегрирования:

$$\int dv = -v_{\text{отн}} \int \frac{dm}{m},$$

где мы вынесли постоянное значение скорости газов $v_{\text{отн}}$ за знак интеграла. Воспользовавшись известными правилами интегрирования, получаем

$$v = -v_{\text{отн}} \ln m + C.$$

Значение постоянной интегрирования C определяется начальными условиями. Допустим, что в начальный момент времени скорость ракеты равна нулю, а ее масса равна m_0 . Тогда предыдущее уравнение дает

$$0 = -v_{\text{отн}} \ln m_0 + C,$$

откуда $C = v_{\text{отн}} \ln m_0$. Следовательно,

$$v = v_{\text{отн}} \ln \frac{m_0}{m}, \text{ или } \frac{m_0}{m} = \exp \frac{v}{v_{\text{отн}}}. \quad (4.14)$$

Соотношение (4.14) называют *формулой Циолковского* по имени К.Э. Циолковского (1857–1935).

Формула Циолковского позволяет рассчитать запас топлива, необходимый для сообщения ракете определенной скорости v . Допустим, например, что ракете надо сообщить «первую космическую скорость», то есть такую скорость, чтобы она начала двигаться вокруг Земли по окружности. Эта скорость приблизительно равна $v = 8 \text{ км/с}$ (в дальнейшем мы покажем, почему первая космическая скорость равна именно такой величине). В современных ракетах скорость газовой струи достигает нескольких километров в секунду. Если считать скорость газовой струи равной $v_{\text{отн}} = 2 \text{ км/с}$, то из формулы Циолковского получается, что для достижения $v = 8 \text{ км/с}$ необходимо, чтобы отношение конечной массы ракеты к ее начальной массе было равно $m_0/m = 54,6$. Это означает, что почти 98% массы ракеты приходится на топливо.

Не представляет особых трудностей обобщение проведенного рассмотрения на случай переменной во времени скорости истечения газов. Для этого уравнение (4.13) надо просто разделить на dt . Учтем также, что все импульсы и их приращения, входящие в (4.12), суть истинные векторы, и придадим окончательному

уравнению векторную форму:

$$m \frac{d\mathbf{v}}{dt} = \mathbf{u} \frac{dm}{dt}, \quad (4.15)$$

где \mathbf{u} — скорость истечения газа из сопла (в векторной форме нет нужды приписывать ей знак «—», как мы это делали с $v_{\text{отн}}$). Формула (4.15) богаче по содержанию, чем (4.13) или (4.14). Она включает случай непараллельных \mathbf{v} и \mathbf{u} , что необходимо обеспечить при повороте ракеты. Более того, поскольку (4.15) есть не что иное как модифицированный второй закон Ньютона, в него можно добавить любую внешнюю силу, и например, описывать таким образом подъем ракеты в поле тяжести Земли:

$$m \frac{d\mathbf{v}}{dt} = \mathbf{u} \frac{dm}{dt} + \mathbf{F}. \quad (4.16)$$

Уравнение (4.16) называют *уравнением Мещерского* — по имени русского механика И.В. Мещерского (1859–1935).

Как мы видим, прямое использование второго закона Ньютона предоставляет большие возможности, нежели законы сохранения, но и решение при этом оказывается более трудным. Отметим в заключение этого параграфа, что уравнения вида (4.16) в физике принято называть *динамическими уравнениями*, а решения этих уравнений (типа соотношения (4.14)) — *интегралами этих уравнений*.

Кроме того, во всякой механической системе существуют функции, которые остаются постоянными при движении. Эти функции называются *интегралами движения*. Таковыми являются, например, энергия и импульс замкнутой системы.

4.4. Колебательное движение: гармонические колебания, резонанс

Рассмотрим еще один пример того, как второй закон Ньютона позволяет исследовать особенности движения, широко распространенного в природе и технике. Речь идет о колебаниях, когда тело в той или иной мере повторяет траекторию своего движения. Часто такое движение возникает под действием упругих сил — как, например, колебания пружин и элементов строительных конструкций. Колебания могут возникать и в результате действия силы тяжести — таковы колебания всевозможных маятников.

Изучение колебаний важно еще потому, что колебательные процессы встречаются не только среди механических движений, но они свойственны самым разнообразным явлениям природы. Так, например, на колебательных процессах основана вся радиотехника. Атомы в твердых телах непрерывно совершают колебательные движения, и характер этого движения определяет

важнейшие свойства окружающих нас тел, таких как прочность или способность проводить тепло.

Колебательное движение может иметь очень сложный вид. Мы рассмотрим простейший тип — *гармонические колебания*, то есть такие колебания, при которых колеблющаяся величина (например, отклонение маятника) изменяется со временем по закону синуса или косинуса. Этот вид колебаний, несмотря на свою простоту, особенно важен по двум причинам: во-первых, многие колебания в природе и технике очень близки к гармоническим, и, во-вторых, периодические процессы с любой другой зависимостью от времени могут быть представлены, как наложение (суперпозиция) истинно гармонических колебаний.

Основные особенности гармонических колебаний мы рассмотрим на примере колебаний соединенного с пружиной тела массы m , находящегося на горизонтальной поверхности, трением о

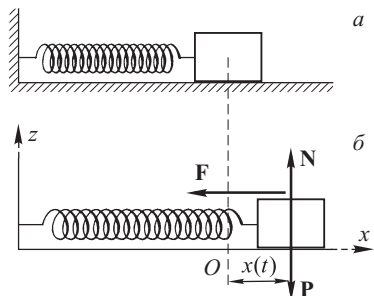


Рис. 4.4

которую можно пренебречь (рис. 4.4 а). Наша цель — определить траекторию движения тела; решение этой задачи, как и в предыдущем параграфе, можно разбить на несколько уже известных нам этапов.

Первый этап — определение того, с каким типом движения мы имеем дело. Мы рассмотрим случай, когда тело можно рассматривать, как материальную точку, и использовать для нахождения его

траектории второй закон Ньютона. Как и в предыдущем параграфе, постановку вопроса об адекватности такого приближения и возможных отклонениях от него нам придется отложить до знакомства с динамикой твердого тела конечных размеров.

Следующий этап — выбор системы координат, определение действующих на тело сил и начальных условий. В нашей задаче естественно выбрать декартову систему координат, у которой, например, ось Ox направлена вдоль пружины, а начало координат соответствует «равновесному» положению тела, когда пружина не растянута и не сжата.

Мы должны также определиться с упругими свойствами пружины: это не может быть гибкая пружина, работающая только на растяжение, но провисающая при сжатии. Мы предположим, что пружина симметричным образом реагирует на нагрузку любого знака:

$$F_y = kx(t),$$

где k — коэффициент жесткости пружины. При $x(t) > 0$ пружина растянута и упругая сила направлена в сторону, про-

типоволожную положительному направлению оси Ox (как на рис. 4.4 б). При $x(t) < 0$ пружина сжата, и упругая сила будет направлена в положительном направлении.

Кроме упругой силы, на тело со стороны Земли действует сила тяжести \mathbf{P} и со стороны поверхности — упругая сила \mathbf{N} (реакция опоры). Эти силы направлены по вертикали, вдоль оси Oz . Так как в этом направлении тело при горизонтальных колебаниях не испытывает ускорения, то эти силы должны быть по модулю равны.

Теперь остается выбрать начальные условия, если они не заданы заранее условиями задачи. Мы рассмотрим случай, когда в начальный момент времени пружина была растянута на некоторую величину x_0 , а скорость тела в этот момент времени была равна нулю. Взяв проекцию на ось Ox от левой и правой части второго закона Ньютона (3.4), получаем окончательно следующую математическую формулировку задачи о движении колеблющегося тела:

$$m\ddot{x} = -kx; \quad x(0) = x_0, \quad \dot{x}(0) = 0, \quad (4.17)$$

где мы, как это часто принято, с помощью одной или двух точек над x обозначили первую и вторую производную по времени. Знак «минус» в правой части уравнения отражает тот факт, что при положительных x (пружина растянута) упругая сила направлена в отрицательном направлении оси Ox , а при отрицательных x (пружина сжата) эта сила направлена в положительном направлении оси Ox .

Теперь можно перейти к следующему этапу — нахождению решения уравнения (4.17), удовлетворяющего начальным условиям. Эта задача немного сложнее той, которую мы решали, определяя траекторию движения тела под действием силы тяжести. Теперь сила зависит от координаты, причем эта зависимость описывается линейным законом: сила пропорциональна координате. Чтобы найти функцию $x(t)$, удовлетворяющую уравнению (4.17), мы сначала перепишем это уравнение в более компактном виде

$$\ddot{x} = -\omega_0^2 x, \quad \omega_0 \equiv \sqrt{\frac{k}{m}}. \quad (4.18)$$

В математике существуют специальные правила решения уравнений типа (4.18). Но мы поступим проще. Уравнение (4.18) означает, что искомая величина $x(t)$ удовлетворяет условию: вторая производная от нее по времени равна ей же самой со знаком «минус», умноженной на постоянную величину ω_0^2 . Вспомним теперь, что производная по времени от $\sin \omega_0 t$ равна $\omega_0 \cos \omega_0 t$, а производная от $\omega_0 \cos \omega_0 t$ равна $-\omega_0^2 \sin \omega_0 t$. Другими словами, взяв вторую производную от $\sin \omega_0 t$, мы получим $-\omega_0^2 \sin \omega_0 t$, то есть функция $\sin \omega_0 t$ является решением нашего уравнения.

Легко теперь убедиться, что решением нашего уравнения будет и более общая зависимость x от t по закону синуса:

$$x(t) = A \sin(\omega_0 t + \varphi), \quad (4.19)$$

где A и φ две произвольных пока константы, для определения которых как раз будет достаточно двух начальных условий нашей задачи.

Из начальных условий (4.17) с учетом (4.19) имеем два соотношения:

$$x(0) = x_0 = A \sin \varphi, \quad \dot{x}(0) = 0 = A \omega_0 \cos \varphi.$$

Из второго из этих соотношений следует: $\varphi = \pi/2$, а тогда из первого соотношения имеем $A = x_0$. Таким образом, удовлетворяющее начальным условиям решение нашей задачи имеет вид

$$x(t) = x_0 \sin \left(\omega_0 t + \frac{\pi}{2} \right). \quad (4.20)$$

Итак, отклонение тела x от положения равновесия меняется со временем по синусоидальному закону, то есть тело совершает *гармонические колебания*. На рис. 4.5а графически

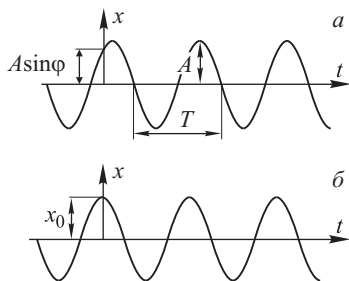


Рис. 4.5

представлена зависимость x от t (4.19) в случае произвольных значений A и φ , а на рис. 4.5б графически представлено наше решение (4.20) в случае выбора конкретных начальных условий. Так как синус меняется в пределах от -1 до $+1$, то константа A определяет максимальное отклонение тела в обе стороны от равновесного положения $x = 0$. Это максимальное отклонение от положения равновесия в процессе гармонических колебаний называется *амплитудой колебаний*. Амплитуда A — постоянная положительная величина. В нашей задаче амплитуда равна x_0 — начальному отклонению тела от равновесного положения. Аргумент синуса $(\omega_0 t + \varphi)$ называется *фазой колебания*, а значение фазы при $t = 0$, то есть константа φ , называется *начальной фазой колебания*. В нашей задаче начальная фаза равна $\pi/2$.

Поскольку синус — периодическая функция с периодом 2π , различные положения тела, совершающего гармонические колебания, повторяются через такой промежуток времени T , за который фаза колебаний получает приращение 2π (см. рис. 4.5а). Этот промежуток времени T называется *периодом колебаний*.

Его можно определить из условия

$$\omega_0(t + T) + \varphi = (\omega_0 t + \varphi) + 2\pi,$$

откуда

$$T = \frac{2\pi}{\omega_0}. \quad (4.21)$$

Число колебаний в единицу времени называется *частотой колебаний* ν . Чтобы найти связь между ν и продолжительностью одного колебания T , следует, очевидно, поделить единицу времени на T :

$$\nu = \frac{1}{T}. \quad (4.22)$$

За единицу частоты принимается частота таких колебаний, период которых равен одной секунде (одно колебание в секунду). Эту единицу называют *герцем* (Гц). Частота в тысячу герц называется *килогерцем* (кГц), в миллион герц — *мегагерцем* (МГц). Из (4.21) следует, что

$$\omega_0 = \frac{2\pi}{T}. \quad (4.23)$$

Таким образом, ω_0 равно числу колебаний за 2π секунд. Величину ω_0 называют *круговой* или *циклической* частотой. Она связана с обычной частотой ν соотношением

$$\omega_0 = 2\pi\nu. \quad (4.24)$$

Для нашей конкретной задачи о колебаниях тела на пружине по гладкой поверхности значение ω_0 дается соотношением (4.18), то есть оно определяется массой тела и коэффициентом жесткости пружины. Для других возможных случаев гармонических колебаний их частота, как мы увидим позже, определяется другими физическими величинами. Это может быть, например, масса и размеры маятника, ускорение свободного падения. Для колебаний электрического тока в радиотехнических устройствах циклическая частота зависит от электрических характеристик, таких как индуктивность и емкость.

Колебательное движение обладает одним интересным и практически очень важным свойством, заключающимся в его особой «чувствительности» к дополнительному воздействию на колеблющееся тело со стороны внешней силы, которая сама меняется со временем по гармоническому закону. Чтобы понять основные особенности этого явления, мы рассмотрим то же самое тело на пружине, движение которого мы только что изучали, но только теперь на тело вдоль пружины действует еще внешняя, как принято говорить, «вынуждающая» сила $\mathbf{F}_B(t)$ (рис. 4.6). Пусть ее горизонтальная проекция изменяется со временем по закону:

$$F_B^x(t) = F_0 \sin \omega t. \quad (4.25)$$

Как изменится движение нашего тела в этом случае? Уравнение движения с учетом дополнительной силы примет, очевидно, вид

$$m \frac{d^2 x}{dt^2} = -kx + F_0 \sin \omega t. \quad (4.26)$$

Поделим обе части этого уравнения на m , после чего с учетом (4.18) получаем

$$\frac{d^2 x}{dt^2} = -\omega_0^2 x + \frac{F_0}{m} \sin \omega t. \quad (4.27)$$

Даже без привлечения специальной математики представляется естественным ожидать, что по прошествии определенного

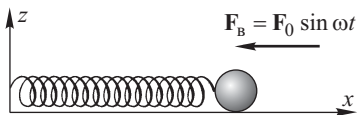


Рис. 4.6

времени движение нашего тела будет в определенной степени следовать за изменением вынуждающей силы, как это, например, происходит при раскачивании качелей. Самое простое предположение — это то, что через определенный промежуток

времени тело будет совершать колебания с той же частотой, с какой колеблется вынуждающая сила, то есть решение уравнения движения (4.27) примет вид

$$x(t) = A \sin \omega t. \quad (4.28)$$

Подставив это выражение в (4.27), легко убедиться, что мы нашли правильное решение, если амплитуда A удовлетворяет соотношению

$$-\omega^2 A = -\omega_0^2 A + \frac{F_0}{m}.$$

Определив отсюда A , получаем окончательно искомое решение уравнения движения (4.27):

$$x(t) = \frac{F_0}{m(\omega_0^2 - \omega^2)} \sin \omega t. \quad (4.29)$$

Итак, под действием вынуждающей силы тело совершает гармонические колебания с той же частотой ω , с которой меняется сама эта сила. И чем ближе оказывается частота вынуждающей силы ω к частоте свободных колебаний тела ω_0 (*собственной частоте*), тем больше амплитуда результирующих колебаний тела, а при $\omega = \omega_0$ амплитуда вообще формально становится бесконечной. Для реальных систем амплитуда, разумеется, всегда остается конечной; в нашем случае бесконечность при $\omega = \omega_0$ обусловлена тем, что мы ограничились линейным приближением $\ddot{x} \propto -x$ и не учли влияния силы трения. При учете трения амплитуда, хотя и резко возрастает, но остается конечной величиной.

Явление резкого увеличения амплитуды колебательной системы при совпадении частоты действующей на нее внешней переменной силы с собственной частотой этой системы (частотой свободных колебаний) называется резонансом. Оно играет исключительно важную роль в науке и технике. Заметим, прежде всего, что детали многих машин и механизмов изготавливаются из упругих материалов и, следовательно, могут совершать гармонические колебания с определенными собственными частотами. Эти детали связаны с другими элементами конструкции или машины, которые могут оказывать на них меняющееся периодически со временем воздействие, играющее роль вынуждающей силы. И если частота этого воздействия станет близкой к собственной частоте, то из-за резонанса могут возникнуть колебательные движения с большой амплитудой, которые отрицательно скажутся на работе механизма или даже приведут к его разрушению.

Но, пожалуй, с полезными следствиями резонансных явлений в технике и науке приходится сталкиваться чаще. Например, в радиотехнике широко используется тот факт, что амплитуда колебаний электрического тока в электрических цепях может резко возрастать, когда частота подводимого к цепи внешнего электромагнитного поля (играющего роль вынуждающей силы) совпадает с собственной частотой колебаний тока в цепи. Настраивая свой радиоприемник на определенную волну, мы как раз и добиваемся совпадения собственной частоты колебаний тока в приемнике с одной из частот приходящих издалека слабых электромагнитных волн. В научных исследованиях явление резонанса широко используется как основа так называемого спектрального метода. Он состоит в экспериментальном определении возможных собственных частот исследуемого объекта (спектра собственных частот) путем воздействия на него переменной вынуждающей силой и измерения тех ее частот, при которых наблюдается резонанс. С другой стороны, собственные частоты определенным образом связаны с внутренними свойствами объекта исследования. Поэтому, измеряя спектр собственных частот объекта, можно получить информацию о многих его важных внутренних свойствах, которые другими способами получить трудно или вообще невозможно.

Вопросы и задачи

1. Какие начальные условия должны быть известны, чтобы можно было однозначно рассчитать траекторию движения материальной точки под действием заданной силы?
2. Что такое гармонические колебания?
3. Что такое фаза гармонического колебания?
4. Что такое резонанс?

5. Под каким углом к поверхности следует бросить камень, чтобы он улетел на максимальное расстояние? Сопротивление воздуха не учитывать.

Ответ: $\pi/4$.

6. Лодка под парусом развила скорость v_0 . Найти зависимость скорости лодки после спуска паруса от пройденного лодкой расстояния x по стоячей воде, если силу сопротивления воды движению лодки можно считать пропорциональной скорости?

Ответ: $v = v_0 - \alpha x/m$, где m — масса лодки, α — коэффициент сопротивления воды.

7. Чему равно F_r/F_e — отношение силы гравитационного взаимодействия электрона и протона в атоме водорода к силе их электрического (кулоновского) взаимодействия?

Ответ: $F_r/F_e \sim 10^{-40}$ (!).

8. Ракета стартует без начальной скорости в пространстве, свободном от силовых полей. Масса ракеты меняется по закону $m = m_0 \exp(-\lambda t)$, где постоянная $\lambda = 10^{-2} \text{ с}^{-1}$; скорость истечения газов из сопла $u = 1000 \text{ м/с}$. Какое расстояние L_T пройдет ракета к моменту времени T , когда ее масса уменьшится в $n = 100$ раз?

Решение. Подставляя m и, соответственно, $dm = -\lambda m_0 \exp(-\lambda t) dt$ в закон сохранения импульса $-u dm + m dv = 0$, получаем $dv = u \lambda dt$. С учетом начальных условий имеем $v = dL/dt = u \lambda dt$, откуда $L = u \lambda t^2/2$. Из условия $m_0/n = m_0 \exp(-\lambda T)$ следует, что $T = \ln(n/\lambda)$. Отсюда окончательно получаем $L_T = u(\ln n)^2/2\lambda \approx 1000 \text{ км}$.

Глава 5

РАБОТА И ЭНЕРГИЯ. ЗАКОН СОХРАНЕНИЯ ЭНЕРГИИ ПРИ ДВИЖЕНИИ МАТЕРИАЛЬНОЙ ТОЧКИ ВО ВНЕШНЕМ СИЛОВОМ ПОЛЕ

5.1. Работа и кинетическая энергия

В повседневной жизни из понятий механики нам приходится чаще всего сталкиваться с понятием работы. Каждый человек непрерывно работает с первых шагов своей жизни, перемещая с места на место различные предметы (в том числе, и свое тело), изучая науки в школе и институте и совершая потом определенные действия на рабочем месте. Вокруг нас работают тысячи машин и механизмов.

Столь же часто, как понятие «работа», встречается нам и понятие «энергия»; это объясняется тем, что энергия, как мы увидим далее, отражает способность отдельных тел или сложных систем совершать работу. Поиск источников энергии всегда являлся одной из самых главных задач человеческого общества. Важнейшим из них для нас является Солнце, электромагнитное излучение которого приносит на Землю энергию, без которой само существование жизни было бы невозможным.

Перейдем теперь к строгому определению этих понятий. В повседневной жизни понятие работы чаще всего связано с определенными усилиями, которые приходится прилагать, чтобы переместить какое-либо тело из одного положения в другое. А посему строгое определение работы объединяет величину перемещения тела с силой, которая действовала на тело в каждой точке этого перемещения. Пусть тело перемещается из точки 1 в точку 2 под действием некоторой силы $\mathbf{F}(\mathbf{r})$, направление и модуль которой могут меняться вдоль траектории произвольным образом (рис. 5.1). Разобьем весь путь из 1 в 2 на столь малые отрезки, что движение на этих отрезках можно заменить на прямолинейное, а действующую на тело силу можно считать постоянной. Другими словами, представим всю траекторию как совокупность малых векторов перемещений $\Delta\mathbf{r}$. Проекцию си-

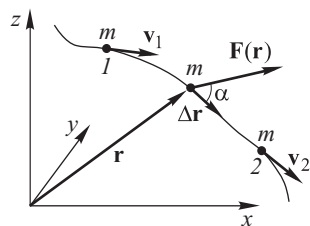


Рис. 5.1

лы \mathbf{F} на каждый из этих векторов перемещений обозначим через F_{\parallel} . Тогда по определению величина

$$\Delta A = F_{\parallel} \Delta r = F \Delta r \cos \alpha \quad (5.1)$$

называется *работой*, совершаемой силой \mathbf{F} на пути Δr (Δr — модуль вектора перемещения $\Delta \mathbf{r}$). Здесь угол α — это угол между векторами \mathbf{F} и $\Delta \mathbf{r}$.

Определение работы (5.1) можно записать более компактно, как скалярное произведение векторов \mathbf{F} и $\Delta \mathbf{r}$, а именно:

$$\Delta A = \mathbf{F} \Delta \mathbf{r}. \quad (5.2)$$

Полная работа, совершенная силой \mathbf{F} на пути из 1 в 2, определяется как сумма всех работ ΔA :

$$A = \sum_{\text{от } 1}^{\text{до } 2} \Delta A = \sum_{\text{от } 1}^{\text{до } 2} \mathbf{F} \Delta \mathbf{r}. \quad (5.3)$$

В частном случае, когда траектория является прямой линией и сила \mathbf{F} постоянна по модулю и направлена вдоль перемещения тела из точки 1 в точку 2 ($\cos \alpha = 1$), для полной работы из (5.3) получаем

$$A = \sum_{\text{от } 1}^{\text{до } 2} F \Delta r = F \sum_{\text{от } 1}^{\text{до } 2} \Delta r = FR, \quad (5.4)$$

где R — расстояние между точками 1 и 2. Работа может быть как положительной, так и отрицательной — все зависит от угла между силой и перемещением. Если в этом же примере сила была бы направлена не вдоль, а против перемещения тела ($\cos \alpha = -1$), то работа оказалась бы отрицательной ($A = -FR$). В этом случае говорят «над телом совершена работа».

Чтобы вычислить работу для произвольной траектории и произвольной силы, следует в определении (5.3) перейти к пределу $\Delta \mathbf{r} \rightarrow 0$. Переход к пределу избавляет от произвола в выборе малых перемещений $\Delta \mathbf{r}$ точно так же, как это было при определении скорости неравномерного движения через производную пути по времени. В нашем случае предельное значение $\Delta \mathbf{r}$ обозначается как $d\mathbf{r}$. При таком предельном переходе сумма в (5.4) преобразуется в интеграл, так что определение работы (5.4) вдоль траектории от точки 1 до точки 2 окончательно можно записать в виде

$$A_{12} = \int_1^2 \mathbf{F} d\mathbf{r}. \quad (5.5)$$

Величина работы, как видно из ее определения, не зависит от того, как долго продолжалось движение, но лишь от пройденного

пути и действующей силы. Во многих случаях, однако, особенно в технике, оказывается важным и то, с какой скоростью совершается работа. Поэтому в механике вводится понятие мощности. *Мощность* — это работа, совершаемая в единицу времени, и ее строгое определение имеет вид

$$W = \frac{\delta A}{dt}. \quad (5.6)$$

Единицы работы и мощности. В системе СИ единицей работы является *джоуль* (Дж), который равен работе, совершаемой силой в 1 Н на пути в 1 м. Единицей мощности в СИ является *ватт* (Вт). Ватт — это такая мощность, при которой за одну секунду совершается работа в один джоуль (Дж/с). Часто используют кратные единицы: *киловатт* (1 кВт = 10^3 Вт), *мегаватт* (1 МВт = 10^6 Вт) и другие. Иногда (сейчас все реже и реже) можно встретиться с единицей мощности, которая называется *лошадиной силой* (л. с.). Соотношение между ваттом и лошадиной силой: 1 л. с. = 736 Вт.

На заре цивилизации полезную работу получали, используя мускульные усилия человека или домашних животных. Затем человек стал осваивать природные источники для совершения работы — такими были, например, водяные мельницы, где полезная работа совершалась за счет движения водных потоков. И наконец, на рубеже XVII и XVIII веков была изобретена паровая машина, и это изобретение ознаменовало начало первой промышленной революции, кардинально изменившей условия человеческого существования. Вслед за этим и родилось понятие об «энергии», как о некой субстанции, содержащейся в горячем паре, которая может превращаться в работу. С тех пор поиск все новых и новых источников энергии стал одной из основных научно-технических задач человеческого общества.

Заметим сразу, что используемое в механике понятие энергии не содержит в себе каких-либо новых свойств пространства, времени и движения по сравнению с теми, которые нашли свое отражение, скажем, во втором законе Ньютона. Как мы сейчас увидим, эта физическая величина довольно просто выражается через известные уже нам механические величины, такие как масса, скорость и т. п. Важная роль энергии в механике определяется двумя обстоятельствами. Во-первых, величина энергии тела или системы тел определяет возможность совершать работу. Во-вторых, мы покажем, исходя из второго закона Ньютона, что энергия замкнутой системы тел сохраняется, и этот закон сохранения, как и закон сохранения импульса, является полезным инструментом при решении многих механических задач.

Рассмотрим сначала простейший случай движения одного тела, которое можно рассматривать как материальную точку. Его

движение подчиняется второму закону Ньютона:

$$m \frac{d\mathbf{v}}{dt} = \mathbf{F},$$

где \mathbf{F} — сила, действующая на тело.

В механике вводятся два понятия: кинетическая энергия и потенциальная энергия. Мы начнем с обсуждения кинетической энергии. *Кинетической энергией T материальной точки* называется величина

$$T = \frac{mv^2}{2}. \quad (5.7)$$

Сначала покажем формально, что кинетическая энергия тела характеризует его способность совершать работу, а потом проиллюстрируем указанное свойство на конкретном примере. Начнем с того, что представим второй закон Ньютона в несколько иной форме. Для этого умножим скалярно левую и правую части уравнения (3.4) на вектор $d\mathbf{r}$ — вектор бесконечно малого перемещения тела вдоль траектории. В результате получаем

$$m \frac{d\mathbf{v}}{dt} d\mathbf{r} = \mathbf{F} d\mathbf{r}. \quad (5.8)$$

Перегруппировав бесконечно малые множители, запишем левую часть (5.8) в виде

$$m \frac{d\mathbf{v}}{dt} d\mathbf{r} = m d\mathbf{v} \frac{d\mathbf{r}}{dt} = m \mathbf{v} d\mathbf{v}.$$

Скалярное произведение векторов по определению равно произведению их модулей на косинус угла между ними, то есть

$$\mathbf{v} d\mathbf{v} = v |d\mathbf{v}| \cos \beta = |v| |d\mathbf{v}| \cos \beta,$$

где v и $|d\mathbf{v}|$ — модули вектора скорости \mathbf{v} и вектора бесконечно малого приращения скорости $d\mathbf{v}$, β — угол между \mathbf{v} и $d\mathbf{v}$. В последнем соотношении мы учли, что $|d\mathbf{v}| \cos \beta$ есть составляющая бесконечно малого приращения скорости, которая направлена

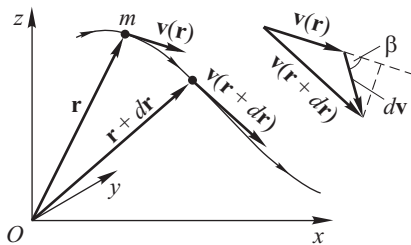


Рис. 5.2

вдоль вектора скорости \mathbf{v} и поэтому равна приращению длины (модуля) этого вектора dv (рис. 5.2). Это означает, что мы можем записать левую часть соотношения (5.8) в виде

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

С учетом этого соотношения получаем, вместо (5.8), выражение

$$d\left(\frac{mv^2}{2}\right) = \mathbf{F} d\mathbf{r}, \quad \text{или} \quad dT = \delta A. \quad (5.9)$$

Тем самым доказано, что изменение кинетической энергии материальной точки при ее перемещении на бесконечно малое расстояние $d\mathbf{r}$ равно работе, совершенной на этом расстоянии действующей на материальную точку силой. Обозначение малого приращения работы δA , вместо дифференциала, является общепринятым. Так обозначаются в физике приращения величин, не являющихся в пространстве функциями точки. Особенно важно будет отличать такие приращения от так называемых полных дифференциалов в термодинамике.

Если тело совершает конечное перемещение вдоль некоторой траектории из точки 1 в точку 2 под действием силы $\mathbf{F}(\mathbf{r})$ (см. рис. 5.1), то всю эту траекторию можно разбить на элементарные бесконечно малые перемещения $d\mathbf{r}$, по соотношению (5.9) на каждом из этих перемещений определить изменение кинетической энергии, а затем просуммировать все эти изменения. Сумма бесконечно малых величин вычисляется с помощью интегрирования, так что для перемещения тела из точки 1 в точку 2 справедливо равенство

$$\int_1^2 dT = \int_1^2 \mathbf{F}(\mathbf{r}) d\mathbf{r} = \int_1^2 \delta A \quad (5.10)$$

или, что то же,

$$T_2 - T_1 = A(1 \rightarrow 2). \quad (5.11)$$

Как мы уже неоднократно подчеркивали, сила есть истинный вектор, поэтому правая часть (3.4) в общем случае представляет собой векторную сумму всех действующих на тело сил, как говорят, результирующую силу. Итак, *приращение кинетической энергии материальной точки на любом отрезке траектории равно работе на этом отрезке результирующей всех действующих на нее сил*. Отметим, что приращение кинетической энергии может быть положительным или отрицательным в зависимости от знака суммарной совершенной силой работы (сила может либо ускорять, либо тормозить движение тела). Еще раз обратим внимание на то, что в правой части (5.11) стоит величина, зависящая от формы траектории $1 \rightarrow 2$, что и выражается обозначением δA .

Из (5.11) следует, что энергия имеет такую же размерность, как и работа, и, следовательно, измеряется в тех же единицах.

Полученное соотношение между работой и кинетической энергией указывает на способ получения полезной работы путем использования кинетической энергии каких-либо тел. Это можно показать на следующем примере. Представим себе горизонтальный цилиндр с поршнем (рис. 5.3). Пусть на поршень слева налетает какая-либо частица. В процессе столкновения с поршнем ча-

стица будет действовать на поршень с некоторой упругой силой, так что за время столкновения поршень приобретет некоторую

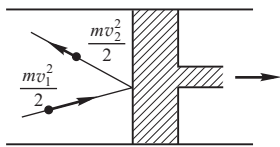


Рис. 5.3

начальную скорость и, следовательно, некоторую начальную кинетическую энергию T_0 , из-за чего скорость частицы по модулю уменьшится, а, следовательно, уменьшится и ее кинетическая энергия на некоторую величину ΔT . Конкретные значения T_0 и ΔT зависят от деталей столкновения, и для наших рассуждений они сейчас не играют роли. Важно, что приобретенную поршнем кинетическую энергию можно теперь превратить в полезную работу. Если на поршень действует только сила трения, то эта энергия будет «истрачена» на преодоление этой силы, то есть на перемещение поршня, как полезного груза, на определенное расстояние L . Если, например, сила трения $F_{\text{тр}}$ при перемещении остается постоянной, то величину L можно вычислить по формуле (5.11), то есть из соотношения $T_0 = F_{\text{тр}} L$. Несмотря на несколько абстрактный характер нашего примера, рассмотренный способ превращения «чужой» кинетической энергии в полезную работу послужил в свое время основой для создания первых паровых машин.

5.2. Потенциальная энергия материальной точки во внешнем силовом поле. Закон сохранения энергии

Наряду с кинетической энергией важную роль в механике играет понятие потенциальной энергии, которое также тесно связано с понятием работы. Но если кинетическая энергия материальной точки определяется, наряду с массой, модулем ее скорости, то потенциальная энергия, как мы увидим, зависит от положения тела в пространстве.

Как и при обсуждении свойств кинетической энергии, мы будем пока рассматривать движение одной материальной точки под действием заданной силы. Если задана векторная функция точки пространства $\mathbf{F}(\mathbf{r})$, то ее называют *силовым полем*. Функция $\mathbf{F}(\mathbf{r})$, во-первых, не должна зависеть от скорости тела (ни от величины, ни от направления), во-вторых, она должна быть постоянна во времени, и в-третьих, *работа в данном силовом поле по любому замкнутому пути должна обращаться в нуль* (рис. 5.4), что математически записывается следующим образом:

$$\oint_C \mathbf{F}(\mathbf{r}) d\mathbf{r} = 0. \quad (5.12)$$

Такие силы (силовые поля) называют *консервативными*; для них и только для них может быть введено понятие потенциальной энергии.

Примем условно какое-либо произвольное положение материальной точки за начальное (радиус-вектор этого положения в выбранной системе координат обозначим через \mathbf{r}_0). Тогда *потенциальной энергией $U(\mathbf{r})$ материальной точки, положение которой по отношению к выбранной системе координат определяется радиусом-вектором \mathbf{r} , называется работа, совершаемая действующей на эту материальную точку результирующей силой, при перемещении материальной точки из положения \mathbf{r} в начальное положение:*

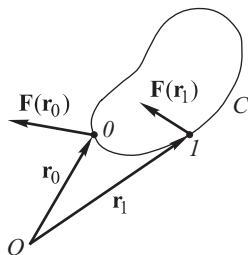


Рис. 5.4

$$U(\mathbf{r}) = A(\mathbf{r} \rightarrow \mathbf{r}_0). \quad (5.13)$$

Корректность такого определения следует из условия консервативности (5.12). В самом деле (см. рис. 5.4), если $\oint_C \mathbf{F}(\mathbf{r}) d\mathbf{r} = 0$, то $\int_0^1 \mathbf{F}(\mathbf{r}) d\mathbf{r} = - \int_1^0 \mathbf{F}(\mathbf{r}) d\mathbf{r}$ при следовании в одном и том же направлении вдоль контура C . Заменим направление следования (т.е. движения) в любом из этих направлений на противоположное. Как следствие, получаем, что работа при перемещении между точками O и I не зависит от выбора пути, по которому оно происходит:

$$\int_1^0 \mathbf{F}(\mathbf{r}) d\mathbf{r} = \text{inv.}$$

Но это как раз и означает, что потенциальная энергия (5.13) определена однозначным образом как функция координат.

Проиллюстрируем наше определение важным примером. Опыт показывает, что силы взаимодействия двух изолированных материальных точек зачастую направлены вдоль соединяющей их прямой и по модулю зависят только от расстояния между ними. Такие силы называются *центральными*. Примером может служить сила гравитационного взаимодействия Солнца с планетой, если их рассматривать как точечные массы, или сила электростатического взаимодействия двух точечных зарядов. Рассмотрим движение материальной точки под действием центральной силы, направленной, например, от одного и того же центра O и по модулю зависящей только от расстояния до него (рис. 5.5). Покажем, что работа при перемещении материальной точки из положения I в положение 2 по двум произвольным путям a и b одинакова, то есть от формы пути не зависит.

Для доказательства проведем из центра O концентрические сферические поверхности, радиусы которых отличаются на бесконечно малую величину dR . Эти поверхности пересекут обе

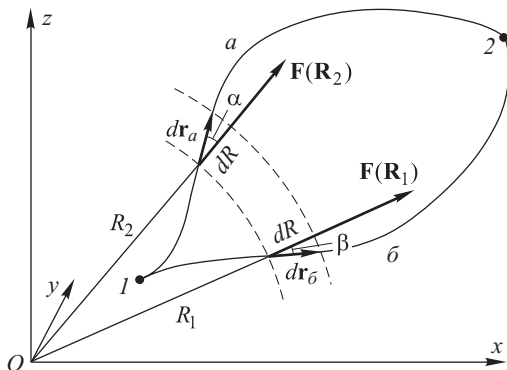


Рис. 5.5

наши траектории a и b на бесконечно малые отрезки пути. Бесконечно малые отрезки криволинейной траектории можно заменить на бесконечно малые прямолинейные векторы перемещений: на рисунке эти векторы перемещений между сферой с радиусом R и сферой с радиусом $R + dR$ по траекториям a и b обозначены через dr_a и dr_6 соответственно. Докажем сначала, что работа δA_a вдоль перемещения dr_a равна работе δA_6 вдоль перемещения dr_6 . В соответствии с определением (5.2) для δA_a получаем

$$\delta A_a = F(R) dr_a \cos \alpha = F(R) dR. \quad (5.14)$$

Аналогично получаем для δA_6 :

$$\delta A_6 = F(R) dr_6 \cos \beta = F(R) dR. \quad (5.15)$$

В написанных выше формулах угол α обозначает, как показано на рисунке, угол между силой $\mathbf{F}(\mathbf{R}_2)$ и перемещением dr_a , а угол β — угол между $\mathbf{F}(\mathbf{R}_1)$ и dr_6 . Кроме того, мы использовали равенства

$$dr_a \cos \alpha = dr_6 \cos \beta = dR.$$

Сравнивая (5.14) и (5.15), мы видим, что действительно $\delta A_a = \delta A_6$. Повторяя вышеприведенные рассуждения для всех отрезков пути a и пути b между соседними сферическими поверхностями, мы получим, что элементарные работы δA на всех этих отрезках равны, а, следовательно, равны и суммарные работы на путях a и b . Так как пути a и b были выбраны произвольно, то это означает, что *работа, совершаемая при перемещении материальной точки центральной силой, не зависит от формы*

пути, а зависит только от начального и конечного положения этой материальной точки.

Представим себе теперь, что на заданную материальную точку действуют силы со стороны N других материальных точек, причем все эти силы $\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_N$ являются центральными. Результирующая сила, действующая на заданную материальную точку уже не является центральной — в разных точках пространства она не будет направлена к какому либо одному и тому же центру (для наглядности на рис. 5.6 изображен случай, когда на материальную точку массы m действуют две центральные силы гравитационного притяжения со стороны двух материальных точек с массами m_1 и m_2).

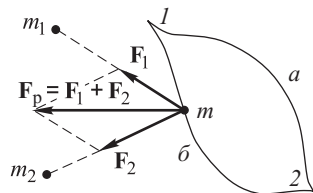


Рис. 5.6

Для дальнейших рассуждений воспользуемся тем фактом, что результирующая сила \mathbf{F}_p равна векторной сумме составляющих ее сил $\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_N$. Это заранее не очевидное свойство сил получило название *принципа суперпозиции сил*. Следовательно справедливо соотношение:

$$\mathbf{F}_p = \sum_{i=1}^N \mathbf{F}_i, \quad (5.16)$$

и тогда элементарная работа при перемещении материальной точки на величину $d\mathbf{r}$ при действии результирующей силы \mathbf{F}_p может быть записана согласно определению (5.2) в виде

$$\delta A_p = \mathbf{F}_p d\mathbf{r} = \sum_{i=1}^N \mathbf{F}_i d\mathbf{r} = \sum_{i=1}^N \delta A_i, \quad (5.17)$$

где δA_i — элементарная работа, совершаемая центральной силой \mathbf{F}_i . Отсюда для полной работы A_p результирующей силы при перемещении из некоторой точки 1 в точку 2 получаем

$$A_p = \int_1^2 dA_p = \sum_{i=1}^N \int_1^2 dA_i = \sum_{i=1}^N A_i = A_1 + A_2 + \dots + A_N. \quad (5.18)$$

Каждая из работ A_1, A_2, \dots, A_N совершается центральной силой и поэтому не зависит от пути. Следовательно и работа A_p не зависит от пути.

Вернемся к определению потенциальной энергии (5.13):

$$U(\mathbf{r}) = A(\mathbf{r} \rightarrow \mathbf{r}_0).$$

Так как работа не зависит от пути, то после выбора какого-либо нулевого положения \mathbf{r}_0 потенциальная энергия однозначно зави-

сит только от положения материальной точки в пространстве. Что касается выбора нулевого положения, то этот выбор определяется соображениями удобства решения конкретной механической задачи.

Произвол в выборе нулевого положения приводит к тому, что *потенциальная энергия определена с точностью до произвольной константы*. В самом деле, вычтем из потенциальной энергии при одном выборе начального положения $U(\mathbf{r}) = A(\mathbf{r} \rightarrow \mathbf{r}_0)$ потенциальную энергию при другом выборе начального положения $U'(\mathbf{r}) = A(\mathbf{r} \rightarrow \mathbf{r}'_0)$:

$$U(\mathbf{r}) - U'(\mathbf{r}) = A(\mathbf{r} \rightarrow \mathbf{r}_0) - A(\mathbf{r} \rightarrow \mathbf{r}'_0). \quad (5.19)$$

Так как работа не зависит от пути, то справедливо соотношение

$$A(\mathbf{r} \rightarrow \mathbf{r}_0) = A(\mathbf{r} \rightarrow \mathbf{r}'_0) + A(\mathbf{r}'_0 \rightarrow \mathbf{r}_0).$$

Подставляя это равенство в (5.19), получаем

$$U(\mathbf{r}) = U'(\mathbf{r}) + A(\mathbf{r}'_0 \rightarrow \mathbf{r}_0),$$

то есть изменение нулевого положения действительно приводит к изменению потенциальной энергии на величину $A(\mathbf{r}'_0 \rightarrow \mathbf{r}_0)$, не зависящую от \mathbf{r} — положения материальной точки в пространстве.

Сумма кинетической и потенциальной энергий называется полной энергией материальной точки и обозначается обычно буквой E :

$$E(\mathbf{r}) = T(\mathbf{r}) + U(\mathbf{r}). \quad (5.20)$$

Покажем, что при движении материальной точки в поле консервативных сил или, что то же, в *потенциальном поле* ее полная энергия не меняется. Для этого воспользуемся соотношением, связывающим бесконечно малое изменение кинетической энергии с работой:

$$dT = \mathbf{F} d\mathbf{r} = \delta A. \quad (5.21)$$

Это равенство справедливо для движения под действием любой силы, а не только потенциальной. Но лишь в этом последнем случае его правую часть можно выразить через изменение потенциальной энергии:

$$dU = U(\mathbf{r} + d\mathbf{r}) - U(\mathbf{r}) = A(\mathbf{r} + d\mathbf{r} \rightarrow \mathbf{r}_0) - A(\mathbf{r} \rightarrow \mathbf{r}_0). \quad (5.22)$$

Так как работа в случае потенциальной силы не зависит от пути, то

$$A(\mathbf{r} \rightarrow \mathbf{r}_0) = A(\mathbf{r} \rightarrow \mathbf{r} + d\mathbf{r}) + A(\mathbf{r} + d\mathbf{r} \rightarrow \mathbf{r}_0).$$

С учетом этого равенства соотношение (5.22) можно записать в виде

$$dU = -A(\mathbf{r} \rightarrow \mathbf{r} + d\mathbf{r}) = -\delta A.$$

Подставляя полученный результат в (5.21), приходим к равенству

$$dT = -dU, \quad d(T + U) = 0, \quad dE = 0.$$

Дифференциал (бесконечно малое приращение) любой функции равен нулю, если эта функция является константой, то есть мы получили

$$E(\mathbf{r}) = T(\mathbf{r}) + U(\mathbf{r}) = \text{const.} \quad (5.23)$$

Это соотношение выражает *закон сохранения энергии*.

Укажем еще одно весьма полезное формальное следствие равенства (5.22). Предположим вначале ради упрощения выкладок, что либо потенциальная энергия зависит только от одной координаты, либо движение происходит лишь вдоль одной координаты. В обоих этих случаях можем положить $U \equiv U(x)$. Соответственно, в элементе работы δA существенна лишь компонента силы F_x (с учетом ее знака). Положив $\delta A = F_x dx$, получаем

$$F_x = -\frac{dU}{dx}. \quad (5.24)$$

В качестве примера приведем выражение для потенциальной энергии деформированной пружины. При малых деформациях связь между силой и смещением, как известно, линейна: $F = -kx$, что соответствует квадратичной зависимости для потенциальной энергии: $U = -kx^2/2$.

В принципе ту же операцию мы можем провести и в общем трехмерном случае. Рассматривая малые перемещения вдоль осей Ox , Oy , Oz , мы получим три дополняющих друг друга соотношения:

$$F_x = -\frac{dU}{dx} \Big|_{y,z=\text{const}}, \quad F_y = -\frac{dU}{dy} \Big|_{z,x=\text{const}}, \quad F_z = -\frac{dU}{dz} \Big|_{x,y=\text{const}}.$$

Входящие в них производные по одной координате при фиксированных остальных называются в математике *частными* производными и обозначаются через $\partial/\partial x$ и т.д. Таким образом, вектор силы \mathbf{F} в компонентах может быть представлен следующим образом:

$$F_x = -\frac{\partial U}{\partial x}, \quad F_y = -\frac{\partial U}{\partial y}, \quad F_z = -\frac{\partial U}{\partial z}. \quad (5.25)$$

Эти соотношения имеют более общий характер, нежели сама ньютонова механика. В практическом отношении они удобны для вычисления силы, если известна потенциальная энергия как функция координат.

Соотношения (5.25) можно представить в компактной форме, введя формально вектор

$$\nabla \equiv \mathbf{e}_1 \frac{\partial}{\partial x} + \mathbf{e}_2 \frac{\partial}{\partial y} + \mathbf{e}_3 \frac{\partial}{\partial z},$$

который в математике называется *градиентом*. Мы еще неоднократно встретимся с такими операторами в разделе «Электричество и магнетизм». Такой формализм позволяет переписать (5.25) в предельно коротком виде:

$$\mathbf{F} = -\nabla U.$$

5.3. О законе сохранения энергии и непотенциальных силах

В элементарных курсах физики закон сохранения энергии (5.23) зачастую толкуется упрощенно. Мы попытаемся в данном параграфе дать о нем более адекватное представление.

Прежде всего заметим, что, сколь бы ни естественным казалось предположение о центральном характере сил взаимодействия между двумя материальными точками, оно отнюдь не универсально. Например, взаимодействие между двумя электронами, движущимися с непараллельными скоростями, к потенциальным силам не сводится. Поэтому и примеры наши с центральными силами (а они всегда потенциальны) — это именно примеры, но не доказательство.

Эксперимент — основа любой естественной науки — показывает, что все многообразие силовых полей в рамках механики подразделяется на три класса. Помимо *консервативных* сил, нам придется рассмотреть еще *гироскопические* и *диссипативные*.

Гироскопическими называются силы, работа которых всегда в точности равна нулю. Очевидно, $\mathbf{F} d\mathbf{r} \equiv 0$ при $\mathbf{F} \neq 0$, если $\mathbf{F} \perp d\mathbf{r}$, а поскольку $d\mathbf{r} = \mathbf{v} dt$, то это означает $\mathbf{F} \perp \mathbf{v}$. Примеров таких сил можно привести немного, наиболее очевидным является случай силы Лоренца, которая действует на частицу с зарядом q в магнитном поле \mathbf{B} :

$$\mathbf{F} = q \cdot [\mathbf{v} \mathbf{B}] \implies \mathbf{F} \perp \mathbf{v}.$$

Эти силы, как следует из самого определения, ни к потенциальной энергии, ни к закону сохранения энергии отношения не имеют.

Диссипативными называются силы, работа которых понижает энергию $E = T + U$ (диссипация — затухание, рассеяние энергии). Это силы трения, знакомые нам из школьного курса; к диссипативным относятся также силы вязкого сопротивления жидкости или газа (они-то и не дают парашютисту падать с ускорением g), сила радиационного торможения при ускорении заряженных частиц и многие другие. Этим классом сил обусловлено то обстоятельство, что закон сохранения энергии (5.23) в рамках одной только механики оказывается верным лишь приблизительно, с большей или меньшей точностью.

В принципе, полная энергия E механической системы может даже расти во времени. Это просто означает, что наша система не достаточно хорошо изолирована от других тел, и над ней совершается работа извне. По этой причине точность закона (5.23) не может быть выше той точности, с которой мы можем считать законы движения не зависящими от внешнего мира.

Но как бы тщательно мы не изолировали механическую систему, сколь бы скрупулезно ни учитывали количественно влияние внешней среды, мы никогда не сможем выполнить условие (5.23) с любой наперед заданной точностью, причем изменение полной энергии будет происходить именно в сторону потерь. Это означает, что вступает в игру диссипация, и предельная точность закона (5.23) определяется именно уровнем диссипации. Обычно при этом говорят: «Энергия сохраняется, но переходит в другие виды», но что такое эти «другие виды», нам еще предстоит разбираться на протяжении всего нашего курса.

Мы рассмотрим пример сил трения между твердыми телами как наиболее хрестоматийный. Популярный образ зацепляющихся неровностей на поверхности соприкасающихся тел не дает правильного понимания природы явления, ибо сам механизм зацепления остается как бы «за скобками». В действительности в основе сил трения лежит электромагнитное взаимодействие электронных оболочек атомов и молекул. Феноменологические законы (т. е. законы, обобщающие напрямую опытные данные) представляют собой усредненные и огрубленные количественные соотношения. Сформулируем их для случая *сухого трения скольжения*. Этим термином называют трение между поверхностями двух твердых тел, скользящих друг по другу при отсутствии между ними какой-либо прослойки, например смазки. В случае сухого трения сила трения возникает не только при скольжении одной поверхности по другой, но и при попытках вызвать такое скольжение у покоящегося вначале тела. В последнем случае она называется *силой трения покоя*. Рассмотрим два соприкасающихся тела 1 и 2, причем тело 2 закреплено неподвижно (рис. 5.7). Пусть тело 1 прижимается к телу 2 с силой F_n , направленной по нормали к поверхности соприкосновения тел. Она называется *силой нормального давления*. Она может быть обусловлена притяжением первого тела к Земле (силой тяжести) или другими причинами. Если тело 1 (вместе с телом 2) в инерциальной системе отсчета не перемещается в вертикальном направлении и его ускорение в этом направлении равно нулю, то это означает согласно второму закону Ньютона, что на него со стороны второго тела

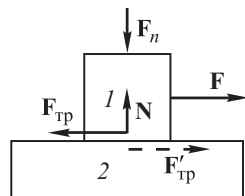


Рис. 5.7

действует уравнивающая сила, равная по модулю F_n и направленная вверх. Ее называют *силой реакции опоры* и обычно обозначают через \mathbf{N} . По своему происхождению она аналогична упругой силе и также обусловлена электромагнитным взаимодействием молекул в поверхностных областях тела 1 и тела 2.

Если попытаться переместить первое тело, подействовав на него внешней горизонтальной силой \mathbf{F} , то обнаружится, что для каждой конкретной пары тел и каждого значения силы нормального давления имеется определенное минимальное по модулю значение F_0 силы \mathbf{F} , при котором тело 1 удастся сдвинуть с места. При значениях внешней силы, заключенных в пределах $0 < F < F_0$, тело остается в покое. По второму закону Ньютона это возможно только в том случае, когда сила \mathbf{F} уравнивается равной ей и противоположно направленной силой, которая и есть сила трения покоя $\mathbf{F}_{\text{тр}}$ (см. рис. 5.7). Эта сила, действующая на тело 1 со стороны тела 2, по модулю равна величине внешней силы F до тех пор, пока последняя не превосходит наибольшего значения силы трения покоя F_0 . Из опыта следует, что во многих практически интересных ситуациях сила F_0 оказывается пропорциональна нормальному давлению F_n : $F_0 = \alpha_0 F_n$.

Когда внешняя сила \mathbf{F} превзойдет по модулю F_0 , тело начнет скользить, причем до некоторого значения скорости проскальзывания движение будет равномерным. При этом, как следует из второго закона Ньютона, сила трения должна быть равна по модулю составляющей приложенной силы, параллельной соприкасающимся поверхностям, и полностью ее уравнивать. Отсутствие у тела ускорения означает, что скорость проскальзывания зависит от приложенной силы, и притом нередко весьма сложным образом.

Увеличивая приложенную силу F , мы в конце концов превзойдем *предельную силу трения скольжения*, и далее движение будет происходить уже с ускорением

$$m\mathbf{a} = \mathbf{F} + \mathbf{F}_{\text{тр}}; \quad F_{\text{тр}} = F_{\text{max}} = \alpha F_n, \quad (5.26)$$

причем при специальной обработке соприкасающихся поверхностей сила трения скольжения может оказаться практически не зависящей от скорости движения тела и, что особенно интересно — от площади соприкосновения трущихся тел. Разумеется, эти свойства не являются ни вполне точными, ни вполне универсальными, но все же правильно описывают большой массив экспериментальных данных. Безразмерный коэффициент пропорциональности называется *коэффициентом трения* (соответственно покоя α_0 или скольжения α). Он зависит от природы и состояния трущихся поверхностей, в частности от их шероховатости. Коэффициенты трения покоя и скольжения практически никогда не совпадают друг с другом.

Попробуем проанализировать эту ситуацию с точки зрения закона сохранения энергии. Из (5.26) легко усмотреть, что работа силы F уже не равна приращению кинетической энергии тела $mv^2/2$. Часть ее, равная работе силы трения $\int F_{\text{тр}} dx$, как бы «исчезает» — это и есть диссипация.

Еще выразительнее проявляются диссипативные эффекты при $F_{\text{тр}} < F_{\text{max}}$ — работа совершается, но кинетическая энергия не изменяется.

Энергия, соответствующая работе сил трения, переходит к атомам поверхности — увеличивается их кинетическая энергия. Это увеличение кинетической энергии атомов находит свое отражение в повышении температуры трущихся поверхностей. (Подробнее о связи хаотического движения атомов вещества с его тепловыми свойствами будет рассказано в одном из последующих разделов курса — термодинамике). Поэтому говорят, что «потерянная» из-за трения кинетическая энергия переходит в тепло. Она может переходить также в акустические возбуждения, электромагнитные волны и даже радиоактивные излучения, но сути дела это не меняет.

Получим формальное соотношение, определяющее изменение полной энергии материальной точки из-за действия на нее неконсервативных сил. Для этого воспользуемся соотношением (5.9), связывающим бесконечно малое изменение кинетической энергии с работой:

$$dT = \mathbf{F} d\mathbf{r} = (\mathbf{F}_k + \mathbf{F}_n) d\mathbf{r} = \delta A_k + dA_n,$$

где \mathbf{F}_k обозначает результирующую всех консервативных сил, а \mathbf{F}_n — результирующую всех неконсервативных сил. Пусть тело совершает конечное перемещение по некоторой траектории из точки 1 в точку 2. Тогда бесконечно малые изменения в полученных соотношениях следует просуммировать вдоль этого участка траектории:

$$\int_1^2 dE = \int_1^2 \delta A_n, \quad \text{или} \quad E_2 - E_1 = A_n(1 \rightarrow 2).$$

Итак: при действии на материальную точку неконсервативных сил изменение ее полной энергии на любом участке траектории равно работе на этом участке результирующей неконсервативных сил. В частности, работа сил трения скольжения всегда отрицательна, так как они всегда направлены против перемещения тела, поэтому силы трения всегда приводят к уменьшению полной механической энергии движущегося тела.

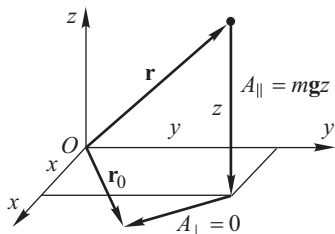
Таким образом, когда мы говорим, что в изолированной системе энергия сохраняется, это означает, что энергия может переходить из одной формы в другую, но ее количество остается

постоянным, оно не зависит от времени. Закон сохранения энергии является строгим законом природы, справедливым для всех известных взаимодействий.

5.4. Простые примеры

Закон сохранения энергии — не только фундаментальный закон природы, но и эффективный метод решения задач. В этом отношении полезны любые законы сохранения (как принято называть их в механике — *интегралы движения*), и принцип их использования совершенно универсален — не решая уравнений движения (например, не вычисляя траекторию материальной точки), сразу же связать начальное и конечное состояния системы. Именно по такому принципу мы выводили формулу Циолковского из закона сохранения импульса.

Начнем с движения материальной точки под действием силы тяжести. Чтобы использовать закон сохранения энергии для исследования движения, необходимо сначала определить конкретное выражение для потенциальной энергии, то есть найти ее зависимость от положения тела в пространстве. Для этого необходимо выбрать в пространстве некоторую произвольную точку с радиусом-вектором \mathbf{r}_0 и вычислить работу, совершаемую силой тяжести при перемещении материальной точки с массой m из заданного положения \mathbf{r} в положение \mathbf{r}_0 .



$$U = A(\mathbf{r} \rightarrow \mathbf{r}_0) = A_{\parallel} + A_{\perp} = mgz$$

Рис. 5.8

Чтобы сделать изложенную выше процедуру более наглядной, воспользуемся рис. 5.8. На этом рисунке положение материальной точки с массой m определяется ее радиусом-вектором \mathbf{r} в декартовой системе координат xOy , у которой плоскость xOy совпадает с поверхностью

Земли. Для определения потенциальной энергии выберем точку с радиусом-вектором \mathbf{r}_0 в любом месте на поверхности Земли. Работу силы тяжести при перемещении из точки \mathbf{r} в точку \mathbf{r}_0 , которая, как мы знаем, не зависит от пути, вычислим вдоль пути, состоящего из двух отрезков: вдоль вертикали до поверхности и далее вдоль произвольной траектории по поверхности до точки с радиусом-вектором \mathbf{r}_0 . На первом участке работа равна произведению модуля силы тяжести mg на величину пути до поверхности Земли, т.е. на координату z нашей материальной точки. А на втором участке работа равна нулю, так как здесь сила всюду перпендикулярна перемещению. Таким образом, работа силы тяжести при перемещении материальной

точки из положения с радиусом-вектором \mathbf{r} в положение с радиусом-вектором \mathbf{r}_0 , то есть потенциальная энергия, $U(\mathbf{r}) = mgz$.

Следовательно, закон сохранения энергии при движении материальной точки под действием только силы тяжести имеет вид следующего соотношения, справедливого для любой точки траектории:

$$E = \frac{mv^2(\mathbf{r})}{2} + mgz = \text{const.} \quad (5.27)$$

Если, например, тело начинает падать с некоторой высоты z_0 с нулевой начальной скоростью, то из закона сохранения энергии (5.27) легко получить закон нарастания скорости по мере приближения тела к поверхности:

$$mgz_0 = mgz + \frac{mv^2(z)}{2} \implies v(z) = \sqrt{2g(z_0 - z)}.$$

Определим выражение для потенциальной энергии и для второго простейшего типа движения материальной точки, рассмотренного в предыдущей главе, — для гармонических колебаний груза по поверхности под действием упругой силы со стороны пружины (см. рисунки 4.5, 4.6). В качестве произвольного начального положения с радиусом-вектором \mathbf{r}_0 в определении потенциальной энергии используем равновесное положение груза при $x = 0$. Тогда потенциальная энергия в положении x равна работе упругой силы $F_y = -kx$ при перемещении материальной точки из положения с координатой x в положение с координатой $x = 0$: $U(x) = kx^2/2$. Таким образом, закон сохранения полной энергии при гармонических колебаниях материальной точки по поверхности записывается в виде

$$\frac{mv^2(x)}{2} + \frac{kx^2}{2} = \text{const.} \quad (5.28)$$

С помощью этого соотношения можно решать многие задачи о гармонических колебаниях, не прибегая к интегрированию уравнений движения. Так например, если нам задана начальная скорость груза v_0 в положении равновесия $x = 0$, то из (5.28) легко найти амплитуду x_0 возникающих после этого гармонических колебаний. При максимальном отклонении скорость груза, а следовательно и его кинетическая энергия равны нулю. Применяя (5.28) к начальному моменту времени и к моменту, когда достигается максимальное отклонение, получаем

$$\frac{mv_0^2}{2} = \frac{kx_0^2}{2}, \quad x_0 = v_0 \sqrt{\frac{m}{k}}.$$

Особо подчеркнем, что закон сохранения энергии при движении во внешнем силовом поле справедлив только при движении

материальной точки под действием сил, постоянных во времени. (В противном случае условие (5.12) не было бы инвариантно по отношению к выбору траектории $\mathbf{r}(t)$). О таком движении

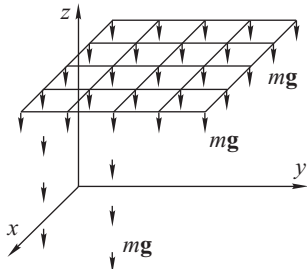


Рис. 5.9

говорят иногда, как о движении в *постоянном силовом поле*, понимая под словом «поле» пространство, каждой точке которого соответствует вектор действующей в пространстве силы. (На рис. 5.9 в качестве примера изображено постоянное силовое поле силы тяжести). Если условие постоянства силового поля не выполняется, полная энергия материальной точки, вообще говоря, не сохраняется. С одним из таких

примеров мы имели дело в предыдущей главе, когда рассматривали явление резонанса, возникающего при колебаниях тела, на которое помимо упругой силы действует еще внешняя вынуждающая сила. Полная энергия в этом случае растет с ростом амплитуды вынужденных колебаний.

Полезно рассмотреть также пример, когда для решения задачи используется не сохранение энергии, а ее полная диссипация. Это также нередко дает возможность получить простой ответ, не решая уравнений движения.

Уменьшение энергии, обусловленное действием сил трения, можно определить с помощью соотношения между изменением кинетической энергии и работой (5.11), которое часто оказывается полезным при решении задач. В качестве наглядного примера рассмотрим такую задачу (рис. 5.10). На горизонтальной поверхности лежит тело массы m , которому в начальный момент сообщают скорость поступательного движения v_0 . Из-за трения о поверхность тело останавливается на некотором расстоянии l от исходного положения. Чему равно это расстояние, если коэффициент трения равен α ? Ответ можно, конечно, найти путем решения соответствующего уравнения движения (второго закона Ньютона). Но решение можно проще и короче получить с помощью соотношения (5.11), из которого сразу получаем

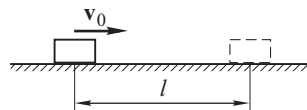


Рис. 5.10

$$\frac{mv_0^2}{2} = F_{\text{тр}}l = \alpha mgl, \quad l = \frac{v_0^2}{2\alpha g}.$$

Конечно, результат этот приближителен, а иногда может быть и просто неверен, поскольку мы воспользовались приближением постоянной силы трения. Но принцип решения подобных задач наш пример демонстрирует должным образом.

5.5. Равновесие и устойчивость

Механика — наука о движении (состояние покоя есть частный случай движения). Наиболее «рафинированное» состояние покоя (или движения с постоянной скоростью) присуще уединенному телу, но этот случай представляет собой идеализированную модель и никогда не реализуется. Наблюдаемое в принципе состояние покоя в системе отсчета, которую мы с необходимой точностью можем считать инерциальной, всегда сводится к одному из двух частных случаев.

1. Силы, действующие на тело, столь малы, и/или характерное время задачи столь коротко, что в пределах точности эксперимента либо точности решения теоретической задачи мы вправе пренебречь ускорением. Если Δx — характерный пространственный масштаб, а Δt — временной, F — масштаб действующей силы, а M — массы, то мы можем рассматривать тело как покоящееся при условии

$$\Delta x \gg \frac{F}{M} (\Delta t)^2.$$

2. Другой возможный случай, в меньшей степени обусловленный малыми характерными временами, — случай *равновесия*. Как было определено в гл. 3, сила есть вектор, а второй закон Ньютона (3.4) — закон векторный. Мы уже отмечали, что это обстоятельство — обобщение огромного массива экспериментальной информации, и что вектор — величина, не просто определяемая модулем и направлением, но и обладающая некоторыми специальными свойствами, в частности, она подчиняется *принципу суперпозиции*. Как следствие, если к телу приложены две, три и т. д. силы, то ускорение определяется их векторной суммой:

$$\frac{d^2 \mathbf{r}}{dt^2} = \frac{1}{m} \sum_i \mathbf{F}_i.$$

Равновесие материальной точки означает, что векторная сумма сил, к ней приложенных, равна нулю:

$$\sum_i \mathbf{F}_i = 0. \quad (5.29)$$

Если же речь идет о равновесии макроскопического тела, одного условия (5.29) недостаточно. К особенностям движения и равновесия тела конечных размеров мы обратимся в гл. 7, здесь же ограничимся бесспорным обобщением формулы (5.29): тело пребывает в *равновесии*, если для любой из материальных точек («физически бесконечно малых объемов»), каковые из этого тела можно выделить, выполнено условие (5.29), включающее как внешние силы, приложенные непосредственно к данному

элементу, так и силы взаимодействия с другими элементами данного тела. Разумеется, ни для макроскопического тела, ни даже для такого, которое можно представить как материальную точку, равенство (5.29) никогда не выполняется с совершенной точностью, однако условие малости, позволяющее пренебречь ускорением, налагается уже не на каждую из действующих сил, но лишь на их векторную сумму.

Условия равновесия тел рассматриваются обычно в разделе теоретической механики, именуемом *статикой*. Мы не будем в нашем курсе изучать ее детально, но остановимся на одном из важнейших вопросов как в механике, так и вообще в физике — проблеме *устойчивости равновесия*.

Помимо точности, с которой выполнено само условие равновесия (5.29), существует еще одна причина, способная выводить из равновесия тело или систему тел. Она заключается в хаотических внешних воздействиях, которые невозможно учесть в рамках (5.29). Они могут иметь, например, тепловую природу (поскольку тепловые эффекты обусловлены движением на молекулярном уровне), динамикой Вселенной («свет далекой звезды» вполне может нарушить равновесие идеальной системы), но даже и «замороженная» в мысленном эксперименте Вселенная не будет свободна от квантовых флуктуаций (это предмет следующего тома нашего курса).

Таким образом, равновесие лишь постольку реализуемо, поскольку механическая система *устойчива* по отношению хотя бы к достаточно малым внешним воздействиям. Наука об устойчивости сложна и обширна, но мы введем некоторые основные понятия на примере одномерной устойчивости равновесия материальной точки — см. рис. 5.11.

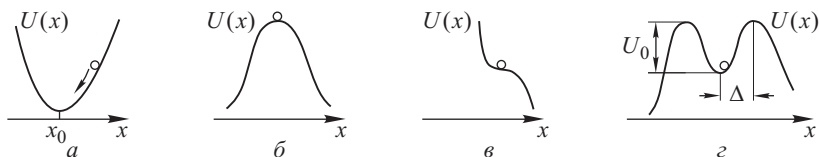


Рис. 5.11

1. Если при малом нарушении равновесия возникает отклик в виде возвращающей силы, материальная точка не может уйти из точки равновесия x_0 слишком далеко (рис. 5.11 а). Само воздействие может заключаться либо в перемещении точечной массы в некоторую точку $x \neq x_0$, либо в придании ей некоторой начальной скорости. Она, однако, останется в некоторой окрестности точки x_0 , если выполнено условие

$$\frac{dU}{dx} > 0 \quad \text{при} \quad x > 0; \quad \frac{dU}{dx} < 0 \quad \text{при} \quad x < 0,$$

что сразу следует из (5.24). Иными словами, устойчивому равновесию отвечает минимум потенциальной энергии, что в одномерной задаче эквивалентно условиям:

$$\left. \frac{dU}{dx} \right|_{x=x_0} = 0, \quad \left. \frac{d^2U}{dx^2} \right|_{x=x_0} > 0. \quad (5.30)$$

Как мы уже знаем из гл. 4, эффект возвращающей силы сводится к некоторым колебаниям вблизи положения равновесия. Если функцию $dU(x)/dx$ можно при малых отклонениях линеаризовать, то это будут просто гармонические колебания — например, (4.18). При учете трения либо иных механизмов диссипации (вязкость, излучение акустических или электромагнитных волн, химические реакции и пр.) колебания должны быть затухающими, и частица со временем — по крайней мере, асимптотически — возвращается в положение равновесия.

2. Предположим, что функция $U(x)$ не имеет минимума в точке равновесия (рис. 5.11 б, в). Тогда нет и устойчивости. В случае максимума $U(x)$ — рис. 5.11 б — смещение частицы в любую сторону приводит к тому, что и ускорение оказывается направлено в сторону смещения, так что вернуться в точку x_0 частица не может. В ситуации, изображенной на рис. 5.11 в, частица может вернуться в x_0 , если либо смещение, либо начальная скорость были отрицательны — но вернется она только один раз, а далее воспроизводится сценарий, соответствующий рис. 5.11 б. Иногда выделяют особо случай безразличного равновесия, когда $F(x) \equiv 0$, т. е. $U(x) = \text{const}$. Но он отличается от вариантов рис. 5.11 б, в лишь функционально, возвращения частицы в положение равновесия не происходит. Истинное устойчивое равновесие может быть обусловлено лишь минимумом потенциальной энергии (5.30).

В современной механике и физике различают *линейную* и *нелинейную* устойчивость. Проблема иллюстрируется рис. 5.11 г. При малых отклонениях от равновесия частица в него возвращается. Но если она выходит за границу $|x - x_0| > \Delta$ или ей сообщают кинетическую энергию, превышающую U_0 , частица в равновесие уже не вернется. Величины U_0 и Δ определяют *порог* нелинейной неустойчивости. Легко сообразить, что при достаточной амплитуде воздействия любая система оказывается нелинейно неустойчивой — хотя бы на уровне разрушения. Но это уже вовсе не обязательно связано со случайными самопроизвольными отклонениями от равновесия.

Все примеры данного параграфа достаточно естественным образом переносятся и на неоднородный случай, и на случай макроскопических тел, а равно и сложных систем, когда, быть может, не все параметры системы ξ_i , определяющие устойчи-

вость, сводятся к пространственным координатам. Правильный язык для такого «перевода» — формализм потенциальной энергии, а главная проблема — отыскание функции $U(\xi_1, \xi_2, \dots)$.

Вопросы и задачи

1. Что такое потенциальная энергия материальной точки?
2. Что такое принцип суперпозиции сил?
3. Может ли потенциальная энергия быть отрицательной?
4. Почему затухают колебания пружины?
5. Почему нагреваются трущиеся поверхности?
6. Имеет ли значение знак потенциальной энергии для устойчивости равновесия?

7. Тело, размерами которого можно пренебречь, начинает скользить с нулевой начальной скоростью по гладкой поверхности, образованной четвертью сферической поверхности радиуса R из ее верхней точки. Достигнув нижней точки, тело продолжает двигаться по шероховатой поверхности, где коэффициент трения скольжения равен α . Чему равно расстояние l , которое тело пройдет по шероховатой поверхности до остановки? Переход от криволинейной поверхности к плоской горизонтали считать сглаженным так, что удара в этой точке не происходит.

Ответ: $l = R/\alpha$.

8. Какова минимальная работа, которую надо затратить, чтобы втащить волокном тело массы m на горку длины L и высоты H ? Коэффициент трения равен α .

Ответ: $A_{\min} = mg(H + \alpha L)$.

9. Математический маятник (груз малых размеров на легком подвесе длины l) находится в положении равновесия. Определите, какую скорость v надо сообщить грузу, чтобы он мог совершить полный оборот, для двух случаев: груз подвешен а) на жестком стержне и б) на нити.

Ответ: а) $v^2 \geq 4gl$; б) $v^2 \geq 5gl$.

Указание. Во втором случае скорость в верхней точке должна быть достаточно велика, чтобы сила тяжести не превышала необходимой центростремительной силы.

10. На столе лежит доска массы $M = 1$ кг, а на доске — груз массы $m = 2$ кг. Какую силу F надо приложить к доске, чтобы она выскользнула из-под груза? Коэффициент трения между грузом и доской равен $\alpha_1 = 0,25$, а между доской и столом — $\alpha_2 = 0,5$.

Решение. Сумма сил, действующих на доску, равна $F_1 = F - [\alpha_1(M + m)g + \alpha_2 mg]$, на груз — $F_2 = \alpha_2 mg$. Чтобы доска могла выскользнуть из-под груза, ее ускорение должно быть больше ускорения груза, т.е. $F_1/M \geq F_2/m$. Отсюда получаем $F \geq (\alpha_1 + \alpha_2)(M + m)g \approx 22,5$ Н.

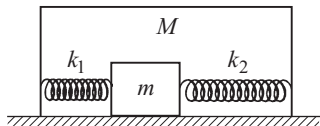


Рис. 5.12

11. Тело массы m колеблется без трения, двигаясь по дну коробки массы M , лежащей на горизонтальной поверхности стола, под действием двух пружин жесткости k_1 и k_2 (см. рис. 5.12). Коэффициент трения коробки о стол равен α . При какой амплитуде колебаний тела Δ коробка начнет двигаться по поверхности стола?

Ответ: $\Delta_{\min} = [\alpha(M + m)g]/(k_1 + k_2)$.

Глава 6

ЗАМКНУТАЯ СИСТЕМА ТЕЛ. ЭНЕРГИЯ ВЗАИМОДЕЙСТВИЯ И ВНУТРЕННЯЯ ЭНЕРГИЯ

6.1. Особенности движения замкнутой системы из двух взаимодействующих материальных точек. Приведенная масса

До сих пор мы изучали наиболее простые движения — движения, при которых размеры тела не играют роли, т.е. использовали приближение (модель) материальной точки. При этом мы рассматривали лишь случай, когда сила, действующая на материальную точку, известна (движение в поле тяжести, гармонические колебания, закон сохранения энергии при движении в постоянном силовом поле).

Теперь, руководствуясь правилом «от простого — к сложному», мы сделаем следующий шаг — рассмотрим особенности движения нескольких взаимодействующих материальных точек. В этом случае действующую на каждую из них силу со стороны остальных тел нельзя уже считать заданной величиной, так как она зависит от меняющегося положения этих остальных тел, траектории движения которых заранее не известны. Так, например, в одном узле сложного механизма часто оказываются соединенными несколько деталей, оказывающих воздействие друг на друга и совершающих друг относительно друга сложные движения. Один из простейших примеров такого взаимодействия изображен на рис. 6.1: две материальные точки с массами m_1 и m_2 соединены пружиной и совершают по горизонтальной плоскости колебания друг относительно друга. Ситуация здесь отличается от той, которая рассматривалась в гл. 4, где мы изучали гармонические колебания одного тела, соединенного пружиной со стенкой.

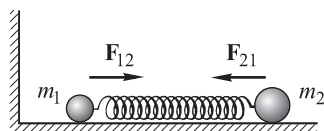


Рис. 6.1

Изучение движения нескольких взаимодействующих тел мы начнем с простейшего случая — движения всего двух материальных точек, при движении которых можно пренебречь всеми действующими на них силами, кроме сил, с которыми они действуют друг на друга. Другими словами, эти две материальные точки образуют замкнутую систему тел. Массы этих тел обозначим

через m_1 и m_2 (рис. 6.2). Силу, действующую на тело 1 со стороны тела 2, обозначим \mathbf{F}_{12} . Она, в соответствии с третьим законом Ньютона, равна по величине и противоположна по направлению силе \mathbf{F}_{21} , действующей на тело 2 со стороны тела 1. Положение наших тел в некоторой произвольно выбранной инерциальной системе отсчета определяется в момент времени t соответствующими радиусами-векторами этих тел $\mathbf{r}_1(t)$ и $\mathbf{r}_2(t)$.

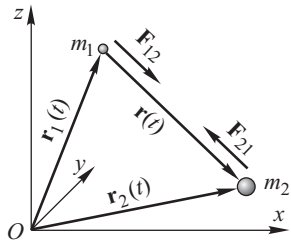


Рис. 6.2

Перейдем к изучению особенностей движения этих двух тел с помощью второго закона Ньютона. Записав этот закон для каждой из материальных точек,

получим следующую систему уравнений относительно интересующих нас радиусов-векторов \mathbf{r}_1 и \mathbf{r}_2 :

$$\begin{aligned} m_1 \frac{d^2 \mathbf{r}_1}{dt^2} &= \mathbf{F}_{12}(|\mathbf{r}_2 - \mathbf{r}_1|), \\ m_2 \frac{d^2 \mathbf{r}_2}{dt^2} &= \mathbf{F}_{21}(|\mathbf{r}_2 - \mathbf{r}_1|). \end{aligned} \quad (6.1)$$

Прежде всего мы рассмотрим задачу об *относительном движении* двух тел. На языке математики относительное движение определяется зависимостью от времени радиуса-вектора $\mathbf{r}(t) = \mathbf{r}_2 - \mathbf{r}_1$, соединяющего две рассматриваемые материальные точки. Уравнение для $\mathbf{r}(t)$ легко получить из нашей системы двух уравнений (6.1) следующим способом. Поделим обе части первого из уравнений на m_1 , а обе части второго уравнения — на m_2 . После этого вычтем из левой и правой частей второго уравнения левую и правую части первого соответственно. В результате получаем

$$\frac{d^2(\mathbf{r}_2 - \mathbf{r}_1)}{dt^2} = \left(\frac{1}{m_2} + \frac{1}{m_1} \right) \mathbf{F}_{21}(|\mathbf{r}_2 - \mathbf{r}_1|),$$

где мы учли, что $\mathbf{F}_{12} = -\mathbf{F}_{21}$. Последнее уравнение можно окончательно записать в виде

$$\mu \frac{d^2 \mathbf{r}}{dt^2} = \mathbf{F}_{21}(r), \quad (6.2)$$

где буквой μ мы обозначили величину, равную

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \quad (6.3)$$

и называемую *приведенной массой*.

Уравнение (6.2) отражает замечательное свойство задачи о движении двух взаимодействующих материальных точек: *задача об относительном движении двух материальных точек*

в замкнутой системе сводится к задаче о движении одной материальной точки с приведенной массой под влиянием той же самой силы взаимодействия.

Одним из примеров замкнутой системы двух тел является система «Солнце–Земля», если пренебречь влиянием других тел Солнечной системы, что в действительности можно сделать с хорошей точностью. Для вычисления траектории движения Земли относительно Солнца нет необходимости решать систему из двух уравнений (6.1), т. е. находить сначала $\mathbf{r}_1(t)$ и $\mathbf{r}_2(t)$ (траектории Солнца и Земли в некоторой инерциальной системе координат) и только после этого уже определять $\mathbf{r}(t) = \mathbf{r}_2(t) - \mathbf{r}_1(t)$ (траекторию относительного движения). Вместо этого можно сразу найти траекторию относительного движения, решая всего одно уравнение (6.2), что и сделал в свое время Ньютон, блестяще подтвердив расчетами знаменитые опытные законы Кеплера. Правда в этом случае в уравнение относительного движения войдет приведенная масса, которая для системы «Солнце–Земля» равна

$$\mu = \frac{M_c M_3}{M_c + M_3} = \frac{M_3}{1 + M_3/M_c}.$$

Масса Солнца примерно в двести тысяч раз больше массы Земли. Поэтому во многих практически важных задачах отношением масс в знаменателе последней формулы можно с хорошей точностью пренебречь, и считать, что приведенная масса для системы «Солнце–Земля» совпадает с массой Земли.

Что касается задачи об относительном движении двух тел, соединенных невесомой пружиной (рис. 6.1), то в этом случае на тела помимо упругой силы, играющей здесь роль силы \mathbf{F}_{12} , действуют еще сила тяжести и реакция опоры. Эти силы, однако, уравновешивают друг друга и не оказывают влияния на движение тел. Поэтому в отсутствие сил трения движение тел по горизонтальной поверхности подчиняется законам движения замкнутой системы, и, следовательно, задача об их относительном движении сводится к рассмотренной в гл. 4 задаче о движении одного тела, но с приведенной массой. Например, для случая колебаний двух тел с одинаковой массой m их приведенная масса μ , в соответствии с (6.3), равна $m/2$. Тогда для вычисления, скажем, частоты относительных колебаний таких двух тел можно воспользоваться уже известной нам формулой (4.18) для частоты колебаний одного тела, соединенного пружиной со стенкой, а именно,

$$\omega = \sqrt{\frac{k}{\mu}} = \sqrt{\frac{2k}{m}},$$

т. е. частота колебаний в рассматриваемом случае в $\sqrt{2}$ раз больше, чем в случае колебаний одного тела.

6.2. Центр масс системы материальных точек

Если бы мы не вычитали, а складывали уравнения (6.1), у нас получился бы просто закон сохранения импульса

$$\frac{d}{dt}(m_1 \mathbf{v}_1 + m_2 \mathbf{v}_2) = 0.$$

Его можно переписать чисто формально как закон постоянства во времени некоторой скорости \mathbf{V}_C :

$$\mathbf{V}_C \equiv \frac{m_1 \mathbf{v}_1 + m_2 \mathbf{v}_2}{m_1 + m_2} = \text{const.} \quad (6.4)$$

Перейдем в систему отсчета, движущуюся со скоростью (6.4). Скорости частиц 1 и 2 при этом преобразуются следующим образом:

$$\mathbf{v}'_1 = \mathbf{v}_1 - \mathbf{V}_C = m_2 \frac{\mathbf{v}_1 - \mathbf{v}_2}{m_1 + m_2}; \quad \mathbf{v}'_2 = \mathbf{v}_2 - \mathbf{V}_C = m_1 \frac{\mathbf{v}_2 - \mathbf{v}_1}{m_1 + m_2}, \quad (6.5)$$

т.е. в новой системе отсчета они выражаются через скорость относительного движения. Свяжем скорость \mathbf{V}_C с радиусом-вектором некоторой точки \mathbf{r}_C :

$$\mathbf{V}_C \equiv \frac{d}{dt} \mathbf{r}_C; \quad \mathbf{r}_C \equiv \frac{m_1 \mathbf{r}_1 + m_2 \mathbf{r}_2}{m_1 + m_2}. \quad (6.6)$$

Отметим, что определение (6.6) совпадает с известным из школьного курса физики понятием центра тяжести. Для доказательства перенесем начало координат в точку \mathbf{r}_C . Тогда, совершенно аналогично (6.5), получим

$$\mathbf{r}'_1 = \mathbf{r}_1 - \mathbf{r}_C = m_2 \frac{\mathbf{r}_1 - \mathbf{r}_2}{m_1 + m_2}; \quad \mathbf{r}'_2 = \mathbf{r}_2 - \mathbf{r}_C = m_1 \frac{\mathbf{r}_2 - \mathbf{r}_1}{m_1 + m_2}.$$

Таким образом, $m_1 \mathbf{r}'_1 = -m_2 \mathbf{r}'_2$ (центр тяжести определяется равенством произведений массы на «плечо»). Но определения (6.4) и (6.6) более корректны и более универсальны, поскольку без каких-либо проблем обобщаются на любое число материальных точек, а следовательно, и на макроскопические тела. Точку C в механике — и вообще в физике — принято называть *центром масс* или *центром инерции* системы материальных точек.

Пусть в некоторой инерциальной системе координат положения взаимодействующих материальных точек с массами m_1, m_2, \dots, m_N задаются в каждый момент времени t посредством радиусов-векторов $\mathbf{r}_1(t), \mathbf{r}_2(t), \dots, \mathbf{r}_N(t)$ (рис. 6.3а). Тогда *центром масс рассматриваемой системы материальных точек называется такая точка, радиус-вектор которой $\mathbf{R}_C(t)$ выражается через радиусы-векторы $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$ материальных точек по формуле*

$$\mathbf{R}_C(t) = \frac{m_1 \mathbf{r}_1(t) + m_2 \mathbf{r}_2(t) + \dots + m_N \mathbf{r}_N(t)}{m_1 + m_2 + \dots + m_N}. \quad (6.7)$$

Подчеркнем, что в общем случае положение центра масс не совпадает с положением какой-либо из материальных точек системы (рис. 6.3 б), хотя иногда такое может и случиться.

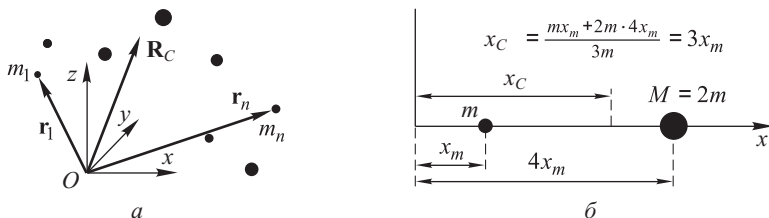


Рис. 6.3

Продифференцируем по времени левую и правую части равенства (6.7). Производная радиуса-вектора по времени есть по определению скорость, так что в результате мы получаем

$$\mathbf{V}_C = \frac{1}{m_1 + m_2 + \dots + m_N} \{m_1 \mathbf{v}_1(t) + m_2 \mathbf{v}_2(t) + \dots + m_N \mathbf{v}_N(t)\}, \quad (6.8)$$

где \mathbf{V}_C — скорость центра масс; $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N$ — скорости материальных точек. Величина $m_1 \mathbf{v}_1$ в (6.8) — импульс первой материальной точки, $m_2 \mathbf{v}_2$ — импульс второй точки и т. д. Таким образом, в фигурных скобках выражения (6.8) стоит сумма импульсов рассматриваемой системы материальных точек, т. е. импульс \mathbf{P} всей системы. Следовательно, равенство (6.8) можно переписать в виде

$$\mathbf{P} = \{m_1 + m_2 + \dots + m_N\} \mathbf{V}_C. \quad (6.9)$$

В системе отсчета, где центр масс покоится, $\mathbf{P} = 0$.

Если нас не интересует относительное движение материальных точек, а интересует движение системы как целого, то тогда всю систему можно рассматривать как одну материальную точку, движущуюся со скоростью \mathbf{V}_C и обладающую импульсом \mathbf{P} . Вспомним, что масса материальной точки есть, по определению, коэффициент пропорциональности между импульсом и скоростью. Поэтому стоящий в равенстве (6.9) коэффициент пропорциональности, заключенный в фигурные скобки, есть масса M рассматриваемой системы:

$$M = m_1 + m_2 + \dots + m_N, \quad (6.10)$$

т. е. *масса системы материальных точек равняется сумме масс этих точек*. Соотношение (6.10), согласно которому масса сложного тела равна сумме масс его частей, кажется нам привычным и очевидным. Однако, как мы еще убедимся, в релятивистской механике (т. е. в более общем случае) ситуация будет

совершенно иной. В предельном случае ньютоновой механики равенство (6.10) представляет собой частный случай определенного физического закона — закона сохранения массы.

В отсутствие внешних сил, т. е. для замкнутой системы, сумма импульсов всех тел системы не зависит от времени; тогда из (6.9) следует важное свойство движения центра масс замкнутой системы материальных точек:

$$\mathbf{V}_C = \text{const},$$

т. е. центр масс замкнутой системы материальных точек неподвижен или движется равномерно и прямолинейно, хотя каждая из материальных точек может совершать сложное движение. Приведенное выше утверждение называют иногда теоремой о движении центра масс.

Мы сейчас докажем следующее важное свойство кинетической энергии: кинетическая энергия T системы материальных точек равна сумме кинетической энергии всей массы системы, мысленно сосредоточенной в ее центре масс и движущейся вместе с ним, и кинетической энергии T' той же системы в ее относительном движении по отношению к системе отсчета, движущейся вместе с центром масс:

$$T = \frac{1}{2} M V_C^2 + T' = \frac{1}{2} M V_C^2 + \sum_{i=1}^N \frac{m_i v_i^2}{2}, \quad (6.11)$$

где $M = m_1 + m_2 + \dots + m_N$, \mathbf{V}_C — скорость центра масс в исходной системе отсчета, \mathbf{v}_i — скорость i -ой материальной точки относительно системы отсчета, движущейся вместе с точкой C . Такую систему обычно называют «системой центра масс», «системой центра инерции» или просто «ц-системой». (Систему отсчета, в которой поставлена задача, если эта система не совпадает с ц-системой, принято называть лабораторной системой отсчета или л-системой).

Для доказательства получим вначале более общее соотношение, связывающее кинетическую энергию в двух системах отсчета (рис. 6.4). Для координат и скоростей точек в старой системе \mathbf{R}_i , \mathbf{V}_i и в новой системе \mathbf{r}_i , \mathbf{v}_i запишем преобразования Галилея:

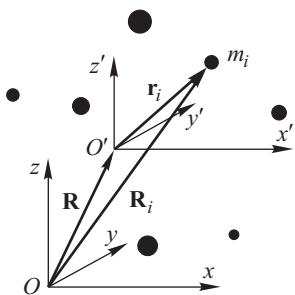


Рис. 6.4

$$\mathbf{R}_i = \mathbf{r}_i + \mathbf{R}; \quad \mathbf{V}_i = \mathbf{v}_i + \mathbf{V},$$

где \mathbf{R} — радиус-вектор перехода из старой системы в новую, а \mathbf{V} — соответственно скорость движения новой системы относительно старой. Тогда кинетическую энергию в старой системе

отсчета можно представить в виде

$$T = \sum_i \frac{m_i V_i^2}{2} \equiv \sum_i \frac{m_i}{2} (\mathbf{V}_i \mathbf{V}_i) = \sum_i \frac{m_i}{2} (\mathbf{v}_i + \mathbf{V})^2 = \\ = \frac{1}{2} \sum_i m_i (v_i^2 + V^2 + 2\mathbf{v}_i \mathbf{V}). \quad (6.12)$$

Правую часть (6.12) можно представить в виде трех сумм:

$$T = \sum_i \frac{m_i v_i^2}{2} + \frac{1}{2} V^2 \sum_i m_i + \mathbf{V} \sum_i m_i \mathbf{v}_i \equiv T' + \frac{1}{2} M V^2 + \mathbf{V} \mathbf{P}, \quad (6.13)$$

где \mathbf{P} — полный импульс системы материальных точек в новой системе отсчета. Соотношение (6.13) принято называть *теоремой Кёнига* по имени немецкого физика А. Кёнига (1856–1901). Если же новая система совпадает с ц-системой, то суммарный импульс в ней равен нулю, $\mathbf{V} = \mathbf{V}_C$, а значит, имеет место соотношение (6.11).

В заключение этого параграфа отметим два важных свойства, вытекающих из определения центра масс. Во-первых, частицы в (6.7) можно объединять в какие угодно группы, например:

$$\frac{m_1 \mathbf{r}_1 + m_2 \mathbf{r}_2 + \dots + m_N \mathbf{r}_N}{m_1 + m_2 + \dots + m_N} = \left[(m_1 + m_2) \frac{m_1 \mathbf{r}_1 + m_2 \mathbf{r}_2}{m_1 + m_2} + \right. \\ \left. + (m_3 + m_4) \frac{m_3 \mathbf{r}_3 + m_4 \mathbf{r}_4}{m_3 + m_4} + \dots \right] / (m_1 + m_2 + \dots + m_N).$$

Отсюда, как легко сообразить, следует, что центр масс любой системы макроскопических тел может быть найден как центр масс системы материальных точек, в предположении, что масса каждого тела сосредоточена в его собственном центре масс.

И во-вторых, от суммирования в (6.7) нетрудно перейти к интегрированию, если мы вычисляем положение центра масс тела с непрерывным распределением плотности вещества $\rho(\mathbf{r})$:

$$\mathbf{R}_C = \left(\int_V \mathbf{r} \rho(\mathbf{r}) dv \right) / \left(\int_V \rho(\mathbf{r}) dv \right).$$

6.3. Потенциальная энергия взаимодействия. Закон сохранения энергии для замкнутой системы материальных точек

В предыдущей главе мы познакомились с понятиями кинетической и потенциальной энергии для специального и относительно простого типа движения — движения одной материальной точки под действием заданной силы (движение в постоянном силовом поле). Теперь мы познакомимся с тем, как формулируется

закон сохранения энергии для более сложного типа движения — движения многих взаимодействующих тел, образующих замкнутую систему.

Начнем с рассмотрения простейшей системы тел — системы, состоящей всего из двух материальных точек, а силу взаимодействия между ними будем полагать потенциальной и зависящей лишь от $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$. Эта оговорка весьма существенна. Например, сила взаимодействия между заряженными частицами, вообще говоря, непотенциальна и зависит от их скоростей, сила взаимодействия между двумя диполями, даже будучи потенциальной, зависит от ориентации частиц в пространстве и т. д. Запишем уравнения движения рассматриваемых материальных точек в некоторой инерциальной системе координат:

$$\begin{aligned} m_1 \frac{d\mathbf{v}_1}{dt} &= \mathbf{F}_{12}(|\mathbf{r}_2 - \mathbf{r}_1|) + \mathbf{F}_1(\mathbf{r}_1), \\ m_2 \frac{d\mathbf{v}_2}{dt} &= \mathbf{F}_{21}(|\mathbf{r}_2 - \mathbf{r}_1|) + \mathbf{F}_2(\mathbf{r}_2), \end{aligned} \quad (6.14)$$

где по-прежнему \mathbf{r}_1 и \mathbf{r}_2 — радиусы-векторы рассматриваемых материальных точек; $\mathbf{v}_1, \mathbf{v}_2$ — их скорости; m_1 и m_2 — их массы, а $\mathbf{F}_{12}, \mathbf{F}_{21}$ — силы, с которыми они действуют друг на друга. Мы предполагаем теперь, что наряду с силами \mathbf{F}_{12} и \mathbf{F}_{21} на тела действуют еще внешние, как говорят, сторонние, силы \mathbf{F}_1 и \mathbf{F}_2 (то есть рассматриваемая система тел не является замкнутой).

Умножим обе части первого из уравнений (6.14) на бесконечно малое перемещение первого тела $d\mathbf{r}_1$ и обе части второго уравнения на $d\mathbf{r}_2$, а затем сложим левые и правые части полученных таким образом соотношений. С учетом того, что $\mathbf{F}_{12} = -\mathbf{F}_{21}$, получаем

$$m_1 \frac{d\mathbf{v}_1}{dt} d\mathbf{r}_1 + m_2 \frac{d\mathbf{v}_2}{dt} d\mathbf{r}_2 = \mathbf{F}_{21}(r) d\mathbf{r} + \mathbf{F}_1 d\mathbf{r}_1 + \mathbf{F}_2 d\mathbf{r}_2,$$

где $d\mathbf{r} = d\mathbf{r}_2 - d\mathbf{r}_1$, и, поскольку $d\mathbf{r}_1/dt = \mathbf{v}_1$, $d\mathbf{r}_2/dt = \mathbf{v}_2$, отсюда следует

$$d\left(\frac{m_1 v_1^2}{2}\right) + d\left(\frac{m_2 v_2^2}{2}\right) = \mathbf{F}_{21}(r) d\mathbf{r} + \mathbf{F}_1 d\mathbf{r}_1 + \mathbf{F}_2 d\mathbf{r}_2. \quad (6.15)$$

Определим понятие потенциальной энергии взаимодействия: *потенциальной энергией взаимодействия $u_{12}(r)$ между материальными точками 1 и 2, находящимися на расстоянии r друг от друга, называется работа, совершаемая силой взаимодействия при удалении одной из них от другой с расстояния r в бесконечность, т. е.*

$$u_{12}(r) = A_{B3}(r \rightarrow \infty). \quad (6.16)$$

Определение конкретного выражения для $u_{12}(r)$ сводится, как мы видим, к вычислению $A_{\text{вз}}(r \rightarrow \infty)$ для конкретной силы взаимодействия \mathbf{F}_{21} , и в следующем параграфе мы покажем в качестве примера, как находится потенциальная энергия гравитационного взаимодействия. Бесконечность в (6.16) принято вводить для определенности; разумеется, при вычислении сил либо работы на конечном участке пути потенциальная энергия, как и в постоянном силовом поле (см. гл. 5), определена с точностью до константы. Если силы взаимодействия между материальными точками являются потенциальными, то работа не зависит от пути. Это, в частности, справедливо для важного частного случая центральных сил:

$$\mathbf{F}_{12} = F_{12}(r) \frac{\mathbf{r}}{r}; \quad \mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2. \quad (6.17)$$

Здесь \mathbf{r}/r — единичный вектор, так называемый *опт* направления. Закон взаимодействия (6.17) никогда не бывает справедлив в точности, но очень во многих случаях используется как достаточно хорошее приближение к реальности.

Подчеркнем одно отличие потенциальной энергии взаимодействия между двумя материальными точками u_{12} от потенциальной энергии одной материальной точки в постоянном силовом поле $U(\mathbf{r})$, о которой шла речь в гл. 5. Потенциальная энергия взаимодействия двух материальных точек зависит только от расстояния между ними r — если, конечно, не вступают в игру ориентационные координаты, скажем, направление дипольных моментов частиц (о чем мы будем говорить в следующем разделе курса). Чтобы подчеркнуть это различие, потенциальную энергию взаимодействия двух материальных точек называют обычно просто *энергией взаимодействия*, и мы так и будем ее называть в дальнейшем.

Из соотношения (6.15) следует связь между работой и приращением кинетической энергии при бесконечно малых перемещениях:

$$dT = \delta A_{\text{вз}} + \delta A_{\text{вн}}. \quad (6.18)$$

Здесь T обозначает суммарную кинетическую энергию рассматриваемых материальных точек — $T = m_1 v^2(\mathbf{r}_1)/2 + m_2 v^2(\mathbf{r}_2)/2$, $\delta A_{\text{вз}} = \mathbf{F}_{21} d\mathbf{r}$ обозначает бесконечно малую работу силы \mathbf{F}_{21} , а $\delta A_{\text{вн}} = \mathbf{F}_1 d\mathbf{r}_1 + \mathbf{F}_2 d\mathbf{r}_2$ — бесконечно малую работу внешних сил.

Из определения энергии взаимодействия (6.16) получаем

$$du_{12} = u_{12}(r + dr) - u_{12}(r) = A_{\text{вз}}(r + dr \rightarrow \infty) - A_{\text{вз}}(r \rightarrow \infty). \quad (6.19)$$

Ввиду нашего предположения о потенциальности сил взаимодействия, работа их от пути не зависит. Поэтому работу

$A_{вз}(r \rightarrow \infty)$ при удалении тела с расстояния r на бесконечность можно совершить, удалив сначала тело по произвольным траекториям на расстояние $r + dr$, а затем удалив его из этого положения на бесконечность, т. е.

$$A_{вз}(r \rightarrow \infty) = A_{вз}(r \rightarrow r + dr) + A_{вз}(r + dr \rightarrow \infty).$$

Подставляя полученный результат в (6.19), получаем

$$du_{12} = -A_{вз}(r \rightarrow r + dr) = -\delta A_{вз}, \quad (6.20)$$

как следствие, соотношение (6.18) принимает вид

$$dT = -du_{12} + \delta A_{вн}.$$

После перегруппировки слагаемых последнее равенство можно представить как

$$d(T + u_{12}) = \delta A_{вн}. \quad (6.21)$$

Если внешние силы отсутствуют, то и $\delta A_{вн} = 0$. В этом случае из (6.21) следует, что для *замкнутой системы*

$$d(T + u_{12}) = 0, \quad \text{или} \quad dE = 0,$$

где E обозначает полную энергию замкнутой системы из двух материальных точек. В этом случае из (6.21) следует, что *полная энергия замкнутой системы двух тел остается при их движении постоянной*:

$$E = \frac{m_1 v_1^2}{2} + \frac{m_2 v_2^2}{2} + u_{12} = \text{const}. \quad (6.22)$$

Следует отметить, что утверждение, выраженное формулой (6.22), имеет характер намного более общий, нежели наш вывод, представляющий по сути частный пример. Значение константы (инварианта, интеграла движения) в (6.22) может быть различным, и оно определяется значениями кинетической энергии и энергии взаимодействия в некоторый (например, начальный) момент времени.

Сформулируем теперь закон сохранения энергии для общего случая, когда замкнутая система состоит не из двух, а из произвольного числа N взаимодействующих материальных точек. Формулировка является обобщением формулы (6.22), а именно: *при движении N материальных точек в замкнутой системе сохраняется ее полная энергия, равная сумме кинетических энергий всех материальных точек и всех попарных потенциальных энергий взаимодействия между ними*:

$$E = \sum_l \frac{m_l v_l^2}{2} + \sum_{l>m} u_{lm} = \text{const}, \quad (6.23)$$

где суммирование по индексам l и m проводится от 1 до N , и во второй сумме при этом содержатся только слагаемые с $l > m$, чтобы не учитывать дважды вклады от энергии взаимодействия каждой из пар материальных точек. Кроме того, для того чтобы сделать запись формулы более компактной, мы ввели обозначения: $v(\mathbf{r}_l) = v_l$ и $u(|\mathbf{r}_l - \mathbf{r}_m|) = u_{lm}$. Так для замкнутой системы из трех материальных точек (например, тройная звезда) закон сохранения (6.16) запишется в виде (рис. 6.5):

$$\frac{mv_1^2}{2} + \frac{mv_2^2}{2} + \frac{mv_3^2}{2} + u_{21} + u_{31} + u_{32} = \text{const.}$$

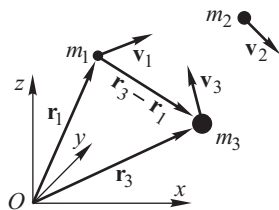


Рис. 6.5

Сколь ни естественно выглядит формула (6.23) в качестве обобщения закона (6.22), применимость ее ограничена важным дополнительным условием. Мы предположили, что имеет место *принцип суперпозиции сил*. В случае их потенциальности это означает *аддитивность потенциальной энергии*:

$$\mathbf{F}_i = \sum_{k \neq i} \mathbf{F}_{ik} \implies U = \sum_{l > m} u_{lm}.$$

Нарушения этих условий не столь уж и редки (например, взаимодействие нуклонов в ядре не есть сумма их попарных взаимодействий друг с другом). Но в классической механике, пока и поскольку мы не обращаемся к космическим масштабам либо нелинейным свойствам сплошной среды, закон сохранения энергии в форме (6.23) имеет достаточно широкую область применимости. Напомним, что мы показали на примере системы из двух материальных точек, что бесконечно малое изменение их полной энергии dE равно бесконечно малой работе внешних сил $\delta A_{\text{вн}}$.

Обобщая это на случай произвольного числа материальных точек, мы имеем, следовательно, равенство

$$dE = \delta A_{\text{вн}}.$$

Применительно к конечным изменениям энергии оно означает, что *изменение полной энергии системы материальных точек равно работе внешних сил*. При этом всю потенциальную энергию частиц, составляющих нашу систему, вместе с кинетической энергией в ц-системе, зачастую удобно интерпретировать как *внутреннюю энергию*, тем самым представляя систему частиц в качестве одного макроскопического тела.

С учетом теоремы Кёнига (6.13) энергию E системы N материальных точек в любой инерциальной системе координат можно

записать в виде

$$E = T + \sum_{l>m} u_{lm} = \frac{1}{2} M V_C^2 + \sum_{i=1}^N \frac{m_i v_i^2}{2} + \sum_{l>m} u_{lm} \equiv \frac{1}{2} M V_C^2 + U,$$

где внутренняя энергия U равна

$$U = \sum_{i=1}^N \frac{m_i v_i^2}{2} + \sum_{l>m} u_{lm}.$$

Так как \mathbf{v}_i — скорости относительно центра масс, то внутренняя энергия системы не зависит от скорости ее движения как целого и определяется только ее собственными, внутренними свойствами — *внутренними степенями свободы*. Мы рассмотрим данный вопрос на должном уровне в разделе, посвященном термодинамике.

Если система как целое покоится ($\mathbf{V}_C = 0$), то $E \equiv U$ — вся энергия системы равна ее внутренней энергии соответственно, $dU = \delta A_{\text{вн}}$.

6.4. Закон всемирного тяготения

Из астрономических наблюдений известно, по крайней мере, со времен Коперника, что Земля и прочие планеты Солнечной системы обращаются по замкнутым криволинейным орбитам вокруг Солнца. Следовательно, тела эти не свободны, на них постоянно действуют какие-то силы. Идентичность формы орбит дает основание считать, что взаимодействие с Солнцем доминирует. Из третьего закона Ньютона следует, что Земля действует на Солнце с той же силой, с которой Солнце притягивает Землю. Но при рассмотрении законов движения можно работать в системе отсчета, связанной с Солнцем, поскольку при $M_c \gg M_3$ приведенная масса (6.3) с хорошей точностью совпадает с массой планеты, а центр масс системы (см. (6.6)) практически совпадает с центром Солнца.

Представим себе, что мы задались целью определить, как зависит от расстояния сила, заставляющая Землю вращаться вокруг Солнца. Для этого можно определить, например, центростремительные ускорения в точке наибольшего удаления Земли от Солнца (в апогее) и в точке наименьшего удаления (в перигее). Во времена Ньютона данные об орбитах планет были уже хорошо известны, т. е. были известны расстояние от Солнца до Земли в апогее r_a и расстояние в перигее r_p , а также значение скорости Земли в различных точках ее траектории. Это позволило Ньютону вычислить ускорение Земли в апогее a_a и в перигее a_p (в этих точках ускорение Земли, а, следовательно, и сила

направлены вдоль одной прямой (вдоль большой полуоси), как показано на рис. 6.6). В результате оказалось, что отношение сил, действующих на Землю в этих точках, равно отношению квадратов соответствующих расстояний:

$$\frac{F_a}{F_n} = \frac{r_n^2}{r_a^2}. \quad (6.24)$$

Естественно предположить, что сила притяжения Земли к Солнцу, возможно, и при любых их взаимных расположениях зависит от расстояния таким же образом. Такое предположение оказалось в хорошем соответствии с законами Кеплера, которые будут рассмотрены в следующей главе. Имея в виду эти соображения, Ньютон, в конечном счете, высказал гениальную догадку, что не только между Солнцем и Землей, а между любыми телами, размерами которых можно пренебречь по сравнению с расстоянием между ними, действует сила взаимного притяжения, подчиняющаяся закону: *сила, с которой две материальные точки притягивают друг друга, пропорциональна массам этих точек и обратно пропорциональна квадрату расстояния между ними:*

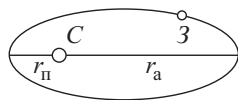


Рис. 6.6

$$F = \gamma \frac{m_1 m_2}{r^2}. \quad (6.25)$$

Здесь γ — коэффициент пропорциональности, называемый постоянной всемирного тяготения или гравитационной постоянной, числовое значение которой зависит от выбранных единиц измерения. В общепринятой сейчас системе СИ γ имеет значение

$$\gamma = 6,67 \cdot 10^{-11} \text{ м}^3/(\text{кг} \cdot \text{с}^2).$$

Сформулированный выше закон носит название *закона всемирного тяготения*, а взаимодействие, о котором идет речь в этом законе, называется *гравитационным взаимодействием*.

В отличие, скажем, от сил электростатического взаимодействия, гравитационная сила (6.25) всегда есть сила притяжения. В векторной форме она может быть представлена следующим образом:

$$\mathbf{F}_{12} = -\gamma \frac{m_1 m_2}{r_{12}^2} \frac{\mathbf{r}_{12}}{r_{12}}. \quad (6.26)$$

Выражения (6.25), (6.26) записаны для взаимодействия двух материальных точек, т.е. для тел, размерами которых можно пренебречь из-за того, например, что их размеры намного меньше расстояния между ними (такая, в частности, ситуация имеет место в случае Солнца и Земли). Из них следует, что силу притяжения, действующую со стороны Земли на всякое тело

вблизи ее поверхности, можно с хорошей точностью считать постоянной, не зависящей от удаления тела от земной поверхности. Допустим, что тело массы m находится от поверхности Земли на высоте h , намного меньшей радиуса Земли R_3 , $h \ll R_3$ (напомним, $R_3 \approx 6\,000$ км). Тогда сила, действующая на тело и направленная к центру Земли, равна

$$F_m = \gamma \frac{m M_3}{(R_3 + h)^2} \approx m \cdot \gamma \frac{M_3}{R_3^2} = mg,$$

где M_3 — масса Земли, а величина $g = \gamma M_3 / R_3^2$ и есть *ускорение свободного падения*, т. к. она равна тому постоянному (с точностью до h/R_3) ускорению, с которым движутся к поверхности Земли все тела под действием ее притяжения. Учитывая, что $M_3 = 5,98 \cdot 10^{24}$ кг, $R_3 = 6,38 \cdot 10^6$ м, получаем для ускорения свободного падения хорошо известное из школы значение $g = 9,8$ м/с², которое с высокой степенью точности подтверждается на опыте. Силу $F_m = mg$, с которой любое тело притягивается к Земле вблизи ее поверхности, называют *силой тяжести*.

Осталось внести ясность в понятие «вес тела». Чаще всего процесс взвешивания состоит в том, что тело помещают на поверхность определенного устройства — весов, и на шкале этого устройства появляется значение веса, выраженное в единицах измерения силы (например, в килограммах). Таким образом, *вес тела* — это сила, с которой

взвешиваемое тело действует на весы.

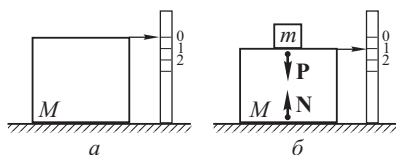


Рис. 6.7

На рис. 6.7 схематически изображены весы в виде твердого тела, у которого изменение уровня верхней поверхности может быть измерено с помощью

стрелки и соответствующей шкалы, и когда на весах ничего нет, стрелка показывает на нуль (рис. 6.7а). Когда на весах находится тело массы m , то из-за возникающих на поверхности соприкосновения упругих сил эта масса действует на весы с некоторой силой P , которая по определению и есть вес тела.

Допустим сначала, что весы вместе с телом покоятся относительно Земли, поверхность которой будем рассматривать, как инерциальную систему отсчета. Для вертикальных составляющих всех сил, действующих в системе «весы плюс взвешиваемое тело», можно согласно второму закону Ньютона записать следующие два равенства:

$$N - Mg - P = 0; \quad N - (m + M)g = 0,$$

где N — сила реакции опоры, на которой расположены весы.

Отсюда получаем: $P = mg$, т.е. когда весы покоятся, вес тела равен действующей на него силе тяжести.

Теперь рассмотрим случай, когда весы с телом движутся вертикально вверх или вниз с ускорением, равным по модулю a :

$$N - Mg - P = \pm Ma; \quad N - (m + M)g = \pm(m + M)a,$$

где знак «плюс» справа соответствует движению вверх, а знак «минус» — вниз. Отсюда получаем, что вес тела в этом случае равен: $P = m(g \pm a)$, т.е. при движении весов с телом с ускорением вверх вес тела увеличивается, а при движении вниз — уменьшается. В последнем случае, если $a = g$ (свободное падение), вес тела становится равным нулю. О такой ситуации говорят, как о *состоянии невесомости*.

Функциональная зависимость (6.25), (6.26) позволяет с очевидностью отнести гравитационную силу к консервативным. Вычислим потенциальную энергию взаимодействия двух материальных точек с массами m_1 и m_2 , находящихся на расстоянии r друг от друга. В соответствии с определением (6.16), получаем из (6.26)

$$u_{12}(r) = A_{\text{вз}}(r \rightarrow \infty) = \int_r^{\infty} \delta A_{\text{вз}} = \int_r^{\infty} \mathbf{F}_{12}(r) d\mathbf{r} = - \int_r^{\infty} \gamma \frac{m_1 m_2}{r^2} \frac{\mathbf{r} d\mathbf{r}}{r}, \quad (6.27)$$

Для удобства вычислений будем удалять тело 2 от тела 1 вдоль соединяющей их прямой линии (рис. 6.8). При этом способе разведения материальных точек вектор бесконечно малого перемещения тела 2 $d\mathbf{r}$ направлен вдоль вектора \mathbf{r} , который направлен от тела 1 к телу 2. Это означает, что $\mathbf{r} d\mathbf{r} = r dr$. Подставляя это соотношение в (6.27), получаем

$$u_{12}(r) = - \gamma m_1 m_2 \int_r^{\infty} \frac{dr}{r^2} = \gamma m_1 m_2 \frac{1}{r} \Big|_r^{\infty} = - \gamma \frac{m_1 m_2}{r}.$$

Итак, окончательно, потенциальная энергия гравитационного взаимодействия двух материальных точек с массами m_1 и m_2 , расположенных на расстоянии r_{12} друг от друга, равна

$$u_{12}^{\text{гр}} = u_{21}^{\text{гр}} = - \gamma \frac{m_1 m_2}{r_{12}}. \quad (6.28)$$

Рассмотрим вопрос о том, какие начальные скорости необходимо сообщить спутнику, чтобы он оставался на замкнутой орбите вблизи Земли, а также — какие скорости необходимо сообщить телу на Земле, чтобы оно преодолело силу земного притяжения и поки-

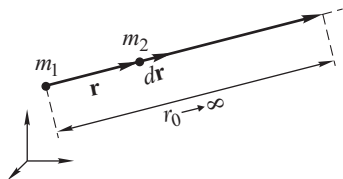


Рис. 6.8

нуло окрестность Земли, или преодолело притяжение Солнца и покинуло Солнечную систему. Эти скорости получили название «космических скоростей» (их называют, соответственно, первой, второй и третьей космической скоростью).

Определим сначала первую космическую скорость V_1 . Вблизи поверхности Земли, на высотах $h \ll R_3$, действующая на тело массы m сила притяжения с хорошей точностью равна силе тяжести mg . Для простоты рассмотрим движение спутника по круговой орбите. В этом случае уравнение движения (второй закон Ньютона) имеет вид

$$m \frac{V_1^2}{R_3 + h} = mg.$$

Пренебрегая h по сравнению с $R_3 \approx 6,4 \cdot 10^6$ м, получаем отсюда

$$V_1 = \sqrt{gR_3} \approx 8 \cdot 10^3 \text{ м/с} = 8 \text{ км/с}. \quad (6.29)$$

Это и есть выражение для *первой космической скорости*. Именно такую скорость необходимо сообщить телу, чтобы оно стало спутником Земли.

Вторую космическую скорость можно найти из закона сохранения энергии (6.28). Землю и космический корабль будем рассматривать, как замкнутую систему двух тел, так как влиянием Солнца и других планет можно с хорошей точностью пренебречь. Задача о движении двух взаимодействующих тел сводится к задаче о движении одного тела под действием той же силы, но с приведенной массой μ :

$$\mu = \frac{M_3 m}{M_3 + m},$$

где M_3 — масса Земли, m — масса корабля. Поскольку масса Земли намного больше, можно положить $\mu = m$. Это означает, что влиянием корабля на движение Земли можно пренебречь, т.е. считать Землю неподвижной и рассматривать движение спутника, как движение в постоянном гравитационном силовом поле Земли.

«Освободить» тело от земного притяжения означает дать ему возможность «уйти на бесконечность», где $u_{12}^{\text{rp}} = 0$. Минимальная кинетическая энергия на поверхности Земли, которую ему для этого надо сообщить, отвечает нулевой кинетической энергии в бесконечно удаленной точке:

$$\frac{mV_2^2}{2} - \gamma \frac{mM_3}{R_3} = T_\infty = 0, \quad (6.30)$$

откуда следует

$$V_2 = \sqrt{2\gamma \frac{mM_3}{R_3}} = \sqrt{2gR_3}. \quad (6.31)$$

Таково выражение для *второй космической скорости*. Из сравнения с (6.29) видно, что вторая космическая скорость в $\sqrt{2}$ раз больше первой: $V_2 \approx 11$ км/с.

Для того чтобы навсегда покинуть Солнечную систему, тело должно преодолеть, помимо земного притяжения, еще и силы притяжения к Солнцу. Необходимая для этого скорость V_3 существенно зависит от направления запуска, поскольку скорость Земли на орбите сравнима с искомой величиной: $V_3 \approx 30$ км/с. При запуске в направлении орбитального движения Земли эта скорость минимальна и составляет около 17 км/с. Она называется *третьей космической скоростью*.

Применяя к различным ситуациям закон всемирного тяготения (6.25), мы оставили открытым вопрос, до какой степени мы вправе это делать, если в реальной задаче фигурируют далеко не точечные массы. Ни в формуле (6.29), ни в левой части формулы (6.30) Земля отнюдь не может быть представлена как материальная точка. В действительности мы молчаливо подразумевали, что сферически симметричное распределение массы может рассматриваться («глядя извне») как материальная точка. Покажем, что это действительно так.

Итак, нашей задачей является определение модуля и направления силы гравитационного притяжения, действующей между Землей (с массой M) и материальной точкой (с массой m), которая находится на расстоянии r от центра Земли (рис. 6.9 а). Сделаем два упрощающих предположения. Во-первых, будем считать Землю строго сферическим телом, хотя она слегка сплюснута вдоль своей оси вращения. Во-вторых, будем предполагать, что плотность вещества Земли всюду постоянна. Оба эти предположения выполняются с хорошей точностью — достаточно сказать, что силы тяжести измеренные на полюсе и на экваторе отличаются на доли процента. Это означает, что наша задача сводится к определению силы гравитационного притяжения между материальной точкой и однородным шаром (рис. 6.9 б).

Представим нашу Землю в виде совокупности материальных точек — бесконечно малых объемов с массой dM , притом каждый из них взаимодействует с материальной точкой m с известной силой тяготения $d\mathbf{F}$, величина и направление которой определяются их взаимным расположением (рис. 6.9 б). Результирующая сила притяжения со стороны шара определится затем векторной суммой всех этих элементарных сил $d\mathbf{F}$.

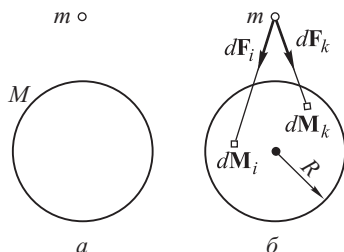


Рис. 6.9

При вычислении результирующей силы мы используем симметрию задачи. А именно, представим сплошной шар как совокупность тонких шаровых слоев, вложенных друг в друга

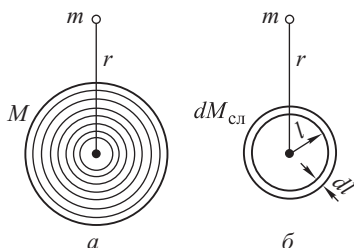


Рис. 6.10

и вычислим сначала силу притяжения между исходной материальной точкой и одним из шаровых слоев, радиус которого обозначим через l , толщину — через dl , а массу — через $dM_{\text{сл}}$ (рис. 6.10 б). Для нахождения силы взаимодействия между слоем и материальной точкой выделим сначала из слоя кольцо, края которого отвечают значениям угла θ и $\theta + d\theta$ (рис. 6.11 а). На рис. 6.11 б кольцо обозначено штриховыми линиями. Каждый бесконечно малый элемент кольца с массой dM_k находится на одном и том же расстоянии a от массы m и поэтому действует на эту массу с силой всемирного тяготения dF_k , которая по модулю одинакова для всех малых элементов кольца и равна

$$dF_k = \gamma \frac{dM_k m}{a^2}. \quad (6.32)$$

Из симметрии задачи ясно, что при вычислении суммарной силы со стороны всего кольца следует учесть только вклад от $dF_{\parallel} = dF_k \cos \varphi$ — составляющей силы dF_k вдоль прямой, соединяющей массу m с центром шарового слоя. Вклады от перпендикулярных к этому направлению составляющих dF_{\perp} взаимно уничтожаются при их суммировании от всех элементов кольца dM_k .

Итак, сила F_k , действующая на массу m со стороны одного кольца, направлена к центру шарового слоя и по модулю равна с учетом (6.32)

$$dF_{\text{кол}} = \sum_{\text{по кольцу}} dF_k \cos \varphi = \gamma \frac{m \cos \varphi}{a^2} \sum_{\text{по кольцу}} dM_k = \gamma \frac{m M_k}{a^2} \cos \varphi, \quad (6.33)$$

где M_k — масса кольца. Выражение для M_k мы найдем,

(рис. 6.10 а), и вычислим сначала силу притяжения между исходной материальной точкой и одним из шаровых слоев, радиус которого обозначим через l , толщину — через dl , а массу — через $dM_{\text{сл}}$ (рис. 6.10 б).

Для нахождения силы взаимодействия между слоем и материальной точкой выделим сначала из слоя кольцо, края которого отвеча-

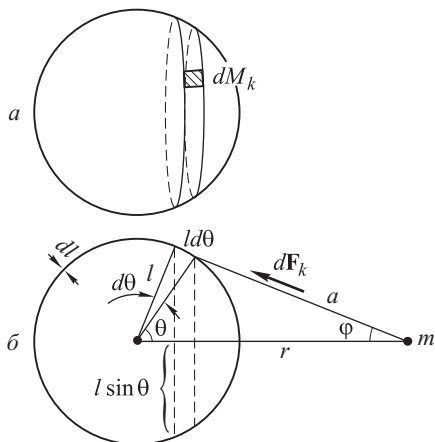


Рис. 6.11

если учтем, что радиус кольца равен $l \sin \theta$, а ширина — $l d\theta$ (рис. 6.11 б). Следовательно, площадь кольца определяется выражением $2\pi l^2 \sin \theta d\theta$. Если толщина шарового слоя dl , а плотность ρ , то масса кольца, равная произведению его объема на плотность, составляет

$$M_k = 2\pi l^2 \sin \theta d\theta \cdot dl \cdot \rho = 2\pi \rho l^2 dl \sin \theta d\theta.$$

Подставляя это выражение в (6.33), получаем

$$dF_{\text{кол}} = \gamma m \cdot 2\pi \rho l^2 dl \frac{\sin \theta \cos \varphi d\theta}{a^2}. \quad (6.34)$$

Весь рассматриваемый нами шаровой слой можно представить, как совокупность колец, положение которых однозначно задается углом θ (при заданных r и l величины a и φ однозначно зависят от θ). Поэтому суммарная сила $dF_{\text{сл}}$ со стороны всего слоя определяется суммой вкладов (6.34) от всех колец:

$$dF_{\text{сл}} = \sum_{\text{по всем кольцам}} dF_{\text{кол}} = \gamma m \cdot 2\pi \rho l^2 dl \int \frac{\sin \theta \cos \varphi d\theta}{a^2}. \quad (6.35)$$

В (6.35) удобно перейти от интегрирования по θ к интегрированию по всем возможным расстояниям a . Для этого воспользуемся известной из тригонометрии теоремой косинусов (см. рис. 6.11 б):

$$a^2 = l^2 + r^2 - 2lr \cos \theta, \quad l^2 = a^2 + r^2 - 2ar \cos \varphi. \quad (6.36)$$

Продифференцировав первое из этих соотношений (при постоянных r и l), получаем

$$2a da = 2lr \sin \theta d\theta. \quad (6.37)$$

Определив $\sin \theta d\theta$ из (6.37) и $\cos \varphi$ из (6.36), получим после подстановки результатов в (6.35)

$$dF_{\text{сл}} = \gamma \frac{m\pi\rho l dl}{r^2} \int_{r-l}^{r+l} \frac{a^2 + r^2 - l^2}{a^2} da, \quad (6.38)$$

где мы учли, что a может изменяться в пределах от $a = r - l$ до $a = r + l$.

Интеграл в (6.38) разбивается на сумму двух интегралов

$$\int_{r-l}^{r+l} da + (r^2 - l^2) \int_{r-l}^{r+l} \frac{da}{a^2}.$$

Нетрудно убедиться, что после интегрирования для обоих слагаемых получается одинаковый результат, а именно, $2l$. Поэтому получаем из (6.38)

$$dF_{\text{сл}} = \gamma \frac{m\rho \cdot 4\pi l^2 dl}{r^2}.$$

Выражение $4\pi l^2 dl$ дает объем слоя, а $\rho \cdot 4\pi l^2 dl$ — его массу $dM_{\text{сл}}$. Таким образом, находим окончательно, что шаровой слой с массой $dM_{\text{сл}}$ действует на находящуюся вне слоя материальную точку с массой m с силой, направленной к центру слоя и равной по модулю:

$$dF_{\text{сл}} = \gamma \frac{m dM_{\text{сл}}}{r^2}, \quad (6.39)$$

где r — расстояние от точечной массы m до центра слоя. Любопытно отметить, что если точечная масса находится внутри шарового слоя, то действующая на нее результирующая сила со стороны всего слоя равна нулю. В этом легко убедиться, изменив в (6.38) пределы интегрирования по a от значения $a = l - r$ до $a = l + r$, которые соответствуют нахождению массы m внутри слоя.

Просуммировав выражение (6.39) по всем слоям, получим силу, с которой шар притягивает массу m ,

$$F = \int dF_{\text{сл}} = \gamma \frac{m}{r^2} \int dM = \gamma \frac{mM}{r^2}.$$

Таким образом, *действие шара на материальную точку эквивалентно действию помещенной в центре шара материальной точки с массой, равной массе шара.*

6.5. Упругие и неупругие соударения

Обратимся к изучению круга явлений, которые, в зависимости от области физики, именуются *столкновениями*, *соударениями*, *рассеянием*. Преимущественно процессы такого рода важны в физике микромира (едва ли не главная задача в физике элементарных частиц — задача о рассеянии). Столкновения составляют один из важнейших предметов рассмотрения в физической кинетике, т. е. в молекулярной физике, физике плазмы, растворов и т. д. Но и в небесной механике, коль скоро речь идет не о регулярных планетных или звездных системах, но об астероидах, кометах, фрагментах, образовавшихся как следствие взрывных процессов, данная проблема занимает достойное место.

Главная особенность *взаимодействий*, которые могут быть квалифицированы как столкновения, состоит в следующем. Участвующие в них частицы (тела) как бы «приходят из бесконечности» и в конечном состоянии «уходят на бесконечность»,

где взаимодействием можно пренебречь. Сразу ясно, например, что взаимодействие Земли и Солнца не может быть отнесено к этой категории. А вот соударение бильiardных шаров — в принципе, может, хотя о бесконечностях в пределах бильiardного стола говорить и не принято. Но при достаточной и вполне разумной степени идеализации задачи (пренебрежение трением о сукно, тем более — обменом импульса через возмущение воздуха или гравитационным взаимодействием шаров) можно утверждать, что шары взаимодействуют *в процессе* удара, но не взаимодействуют *до* или *после* него. Значительно сложнее представить таким образом столкновение заряженных частиц, взаимодействующих по закону Кулона, поскольку сила и потенциальная энергия их взаимодействия не обращаются в нуль ни на каком конечном расстоянии. Отсюда и возникает бесконечность в корректном определении процесса столкновения, а в реальной ситуации мы всегда имеем дело с некоторым приближением к таковому.

Мы в рамках курса механики ограничимся достаточно простыми примерами, по преимуществу такими, когда при соударении тела приходят в непосредственный контакт друг с другом. Термин «соударение» как раз и относится обычно к классической механике макроскопических тел. В этом случае определение траекторий тел после соударения путем решения уравнений движения оказывается часто очень сложной, а иногда вообще невыполнимой задачей. Вот тут-то особенно полезными оказываются законы сохранения энергии и импульса, применение которых к задачам о соударениях мы сейчас рассмотрим.

При соударении макроскопические тела деформируются. При этом некоторая часть кинетической энергии, которой обладали тела перед ударом, переходит в потенциальную энергию упругой деформации, а некоторая часть кинетической энергии переходит во внутреннюю энергию образующих тела атомов и молекул. В зависимости от того, насколько меняется внутренняя энергия тел, при решении задач используют нередко одно из двух приближений: абсолютно упругий и абсолютно неупругий удар.

Абсолютно упругим ударом называют такое соударение тел, при котором переходом части их энергии во внутреннюю энергию тел можно пренебречь. Можно считать, что при таком ударе кинетическая энергия переходит полностью или частично в потенциальную энергию упругой деформации. Затем тела восстанавливают свою форму, отталкивая друг друга. В результате потенциальная энергия упругой деформации переходит обратно в кинетическую энергию, и тела разлетаются со скоростями, величина и направление которых определяется двумя условиями — законом сохранения полной энергии и законом сохранения полного импульса сталкивающихся тел. При *абсолютно неупругом*

ударе тела «слипаются», т. е. после удара они движутся с одинаковой скоростью либо покоятся. Кинетическая энергия тел полностью или частично превращается в их внутреннюю энергию. Кинетическая энергия тел до и после соударения имеет различное значение, и выполняется лишь закон сохранения импульса.

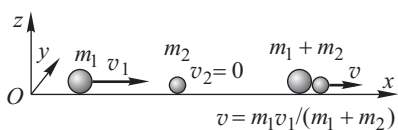


Рис. 6.12

Пример абсолютно неупругого удара показан на рис. 6.12.

Рассмотрим также пример абсолютно упругого удара, причем ограничимся случаем центрального удара двух однородных шаров, один из которых первоначально покоится. (Удар называется центральным, если шары до удара движутся вдоль прямой, проходящей через их центры.) Пусть вращение шаров отсутствует и пусть справедливо приближение, в котором два тела образуют как бы замкнутую систему, или, что то же, внешние силы, приложенные к шарам, уравниваются друг друга (как, например, при скольжении шаров без трения по горизонтальной поверхности). Обозначим

массы шаров через m_1 и m_2 . Пусть второй шар до удара неподвижен, а первый до удара двигался в положительном направлении оси x со скоростью v_{10} (рис. 6.13). Скорости шаров после удара обозначим через v_1 и v_2 . Эти неизвестные пока величины являются проекциями на ось x соответствующих векторов \mathbf{v}_1 и \mathbf{v}_2 , и, следовательно, их знак, полученный в результате решения, определит, в каком направлении оси x шары будут двигаться после удара.

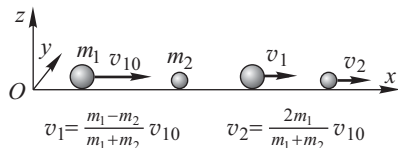


Рис. 6.13

Запишем условия сохранения энергии и импульса:

$$\frac{m_1 v_{10}^2}{2} = \frac{m_1 v_1^2}{2} + \frac{m_2 v_2^2}{2}, \quad (6.40)$$

$$m_1 v_{10} = m_1 v_1 + m_2 v_2. \quad (6.41)$$

Решение системы уравнений (6.40), (6.41) — задача элементарная. Искомые значения для v_1 и v_2 :

$$v_1 = \frac{m_1 - m_2}{m_1 + m_2} v_{10}, \quad v_2 = \frac{2m_1}{m_1 + m_2} v_{10}, \quad (6.42)$$

Отметим некоторые особенности движения шаров при их упругом центральном ударе, вытекающие из полученного решения (6.42). Если массы шаров равны, то из (6.42) следует, что $v_1 = 0$, $v_2 = v_{10}$. Это означает, что в таком случае первый шар после удара останавливается, а второй движется с той скоро-

стью, которая была у первого шара до удара. Именно по этой причине наилучшей защитой от быстрых нейтронов оказываются вещества, содержащие как можно больше водорода. При столкновении нейтрона с ядром атома водорода — протоном — последний перехватывает практически всю кинетическую энергию (поскольку $M_p \approx M_n$), а нейтрон останавливается.

Рассмотрим случай упругого столкновения тела со стенкой, т. е. с телом, массу которого можно считать бесконечно большой. Для этого случая из (6.42) получаем $v_1 = -v_{10}$, $v_2 = 0$, т. е. в результате упругого столкновения со стенкой первый шар меняет свою скорость на противоположную, отскакивая от стенки с той же по модулю скоростью, с которой он к стенке подлетал. Если же столкновение происходит с движущейся стенкой (которая в принципе моделирует любое массивное тело), то весь этот сценарий просто переносится в систему отсчета движущейся стенки. Возвращаясь в лабораторную систему, получаем

$$v_2 = v_{20} = \text{const}; \quad v_1 = -v_{10} + 2v_{20}, \quad (6.43)$$

в предположении, что v_{10} , v_{20} считаются однонаправленными (направления «вперед-назад» учитываются посредством знака \pm). Подчеркнем, что соударение, упругое в системе отсчета стенки, безусловно остается таковым и в л-системе: при корректном подсчете кинетической энергии уже нельзя пренебрегать поправками к v_1 , v_2 порядка m_1/m_2 в формуле (6.43).

Пусть теперь рассматривается упругое столкновение двух частиц, которое уже не может быть описано в рамках одномерной модели. Припишем индекс i начальным значениям всех физических величин, а индекс f — конечным. Закон сохранения импульса в произвольной системе отсчета будет существенно трехмерным:

$$\mathbf{p}_{1i} + \mathbf{p}_{2i} = \mathbf{p}_{1f} + \mathbf{p}_{2f} \equiv \mathbf{p}, \quad (6.44)$$

где \mathbf{p} — сохраняющийся суммарный импульс системы сталкивающихся частиц. В ц-системе $\mathbf{p} = 0$, и размерность уравнения (6.44) понижается. Действительно, векторы $\mathbf{p}_{1i} = -\mathbf{p}_{2i}$ определяют некоторую прямую, векторы $\mathbf{p}_{1f} = -\mathbf{p}_{2f}$ — еще одну прямую, а две пересекающиеся прямые задают некоторую плоскость (исключая тот вырожденный случай, когда $\mathbf{p}_{1i} \parallel \mathbf{p}_{1f}$, так что задача оказывается одномерной). Следовательно, в ц-системе процесс рассеяния всегда происходит в некоторой плоскости, поэтому в общем случае вместо двух переменных v_1 , v_2 в уравнениях (6.40), (6.41), мы будем иметь дело с четырьмя. Процесс рассеяния, ввиду его фактической двумерности, удобно представить графически, как это показано на рис. 6.14.

В ц-системе $|\mathbf{p}_{1i}| = |\mathbf{p}_{2i}| = p_c$, тогда кинетическая энергия в ц-системе (которая при упругом столкновении сохраняется)

равна

$$T = \text{const} = \frac{p_{Ci}^2}{2} \left(\frac{1}{m_1} + \frac{1}{m_2} \right) = \frac{p_{Cf}^2}{2} \left(\frac{1}{m_1} + \frac{1}{m_2} \right).$$

Отсюда следует, что и после соударения $|\mathbf{p}_{1f}| = |\mathbf{p}_{2f}| = p_C$. Таким образом, концы векторов $\mathbf{p}_{1i}, \dots, \mathbf{p}_{2f}$ в ц-системе можно представить лежащими на окружности, как это показано на рис. 6.14а: в результате рассеяния векторы $\mathbf{p}_{1,2}$ поворачиваются, как показано на рисунке, не изменяя своей длины (модуля). Похожая картинка получилась бы и для скоростей частиц,

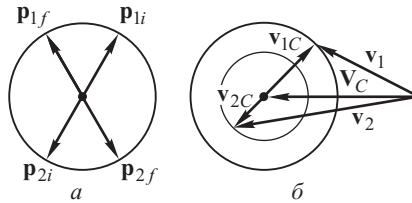


Рис. 6.14

только концы векторов $\mathbf{v}_{1,2C}$ располагались бы на концентрических окружностях с радиусами, соответственно равными p_C/m_1 , p_C/m_2 . На рис. 6.14б показано, как можно построить векторы скорости частиц в л-системе, складывая $\mathbf{v}_{1,2C}$ с вектором скорости центра масс. Такое построение можно провести и для начальных, и для конечных скоростей.

Если бы столкновение было неупругим, но мы бы знали, какая именно доля кинетической энергии в ц-системе утрачена, можно было бы провести аналогичное построение, но радиус соответствующих окружностей был бы разным для начального и конечного состояний. Эта проблема — учет неупругости с известными потерями — чрезвычайно важна в контексте молекулярной, атомной и ядерной физики, ибо моделирует реакции — от химических до ядерных.

Вопросы и задачи

1. Что такое приведенная масса?
2. Что такое энергия взаимодействия трех материальных точек?
3. Напишите выражение для энергии гравитационного взаимодействия двух материальных точек с массами m_1 и m_2 .
4. Как связана сила взаимодействия между двумя материальными точками с их энергией взаимодействия?
5. Что такое центр масс?
6. Чем отличаются упругие соударения от неупругих?
7. Что такое первая, вторая и третья космические скорости?
8. Определить долю энергии δ , теряемую частицей массы m_1 при упругом центральном столкновении с неподвижной частицей массы m_2 . При каком соотношении масс m_1/m_2 потеря энергии максимальна? Используя полученный

результат, объяснить, почему в ядерных реакторах для замедления нейтронов используется рассеяние их на ядрах атомов легких (дейтерий, углерод), а не тяжелых элементов.

Ответ: $\delta = 4m_1m_2/(m_1 + m_2)^2$. Отсюда видно, что при $m_1 = m_2$ теряется вся энергия ($\delta = 1$).

9. Лодка длиной L и массы M с находящимся на ее корме человеком массы m неподвижно стоит на спокойной воде носом у причала перпендикулярно ему. На сколько отодвинется лодка от причала, когда человек пройдет по лодке с кормы на нос?

Ответ: $x = mL/(m + M)$.

10. На дне маленькой запаянной пробирки, подвешенной над столом на нити, сидит муха, масса которой равна массе пробирки. Расстояние от дна пробирки до стола равно длине пробирки l . Нить пережигают, и за время падения муха перелетает со дна в верхний конец пробирки. Определите время падения пробирки.

Ответ: $t = \sqrt{l/g}$.

11. Два маятника в виде шариков разных масс m_1 и m_2 подвешены на нитях разной длины l_1 и l_2 так, что шарики соприкасаются. Первый шарик отводят в плоскости нитей на небольшой угол β и отпускают. На какие углы отклонятся маятники после упругого удара?

Ответ: $\beta_1 = \beta(m_1 - m_2)/(m_1 + m_2)$, $\beta_2 = 2\beta m_1(l_1/l_2)/(m_1 + m_2)$.

12. Тело, движущееся со скоростью v , догоняет стенку, движущуюся в том же направлении со скоростью V . Найдите скорость тела u после упругого соударения.

Ответ: $u = -(v - 2V)$.

13. Шайба массы m , скользя по льду, сталкивается с неподвижной шайбой массы $3m$. Считая удар центральным и упругим, определите, на каком расстоянии S друг от друга остановятся шайбы, если скорость первой шайбы непосредственно перед ударом была равна v , а коэффициент трения между шайбами и льдом равен α .

Ответ: $S = v^2/4\alpha g$.

Глава 7

УРАВНЕНИЕ МОМЕНТОВ. ДИНАМИКА ТВЕРДОГО ТЕЛА

7.1. Момент импульса и момент силы

Введем два понятия, чрезвычайно важных для описания движения (и как частный случай — равновесия) макроскопического твердого тела. Но вводить их целесообразно на основе закона движения материальной точки:

$$\frac{d\mathbf{p}}{dt} = \mathbf{F}. \quad (7.1)$$

В векторной форме уравнение (7.1) имеет универсальный вид во всех инерциальных системах отсчета и не зависит от выбора начала координат. Предположим, что мы выбрали в качестве такового некоторую точку O . Пусть \mathbf{r} — радиус-вектор нашей материальной точки в системе координат с началом в точке O . Проведем некоторую тождественную операцию — умножим все уравнение (7.1) слева векторно на \mathbf{r} :

$$\left[\mathbf{r} \frac{d\mathbf{p}}{dt} \right] = [\mathbf{r}\mathbf{F}].$$

Учтем, далее, что для материальной точки $d\mathbf{r}/dt = \mathbf{v}$, $\mathbf{v} \parallel \mathbf{p}$. Отсюда следует, в частности, что

$$\left[\mathbf{r} \frac{d\mathbf{p}}{dt} \right] \equiv \frac{d}{dt} [\mathbf{r}\mathbf{p}] - \left[\frac{d\mathbf{r}}{dt} \mathbf{p} \right] = \frac{d}{dt} [\mathbf{r}\mathbf{p}],$$

так что в целом второй закон Ньютона (7.1) приводится к эквивалентному виду

$$\frac{d\mathbf{L}}{dt} = \mathbf{M}, \quad (7.2)$$

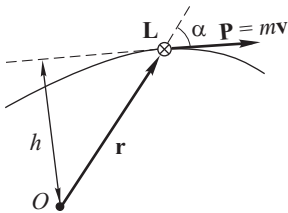


Рис. 7.1

где векторная величина $\mathbf{M} = [\mathbf{r}\mathbf{F}]$ именуется *моментом силы F относительно точки O* , а $\mathbf{L} = [\mathbf{r}\mathbf{p}]$, соответственно, *моментом импульса материальной точки относительно точки O* — (рис. 7.1).

В плоскости, заданной векторами \mathbf{r} и \mathbf{p} , модуль вектора \mathbf{L} может быть представлен, в соответствии с рис. 7.1, как $L = rp \sin \alpha = mvh$.

Пусть, например, материальная точка массы m движется по окружности радиуса R (рис. 7.2). Момент импульса этой

материальной точки относительно центра окружности O равен по модулю $L = mvR$. Вектор \mathbf{L} в этом случае перпендикулярен к плоскости окружности, причем направление движения частицы и вектор \mathbf{L} образуют, как принято говорить, правовинтовую систему. При постоянном радиусе траектории момент импульса может изменяться только за счет изменения модуля скорости. При равномерном движении материальной точки по окружности ее момент импульса остается постоянным и по модулю и по направлению.

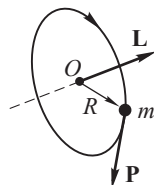


Рис. 7.2

Пока наши новые определения еще не наполнены должным содержанием — мы всего лишь переписали в новой форме второй закон Ньютона. Ситуация меняется, когда от единственной материальной точки мы переходим к системе материальных точек и, в качестве предельного случая — к макроскопическому телу. При движении друг относительно друга нескольких материальных точек момент импульса каждой материальной точки не будет оставаться постоянным. Но если эти тела образуют замкнутую систему, то оказывается, что будет оставаться неизменным их суммарный момент импульса относительно произвольного центра $\mathbf{L} = \mathbf{L}_1(t) + \mathbf{L}_2(t) + \dots + \mathbf{L}_N(t)$.

Для наглядности мы приведем доказательство для системы, состоящей всего из двух материальных точек, связанных центральной силой, т. е. $\mathbf{F}_{12} \parallel (\mathbf{r}_2 - \mathbf{r}_1)$, и рассмотрим случай, когда помимо сил взаимодействия на материальные точки действуют также внешние силы (рис. 7.3). Запишем уравнения для моментов импульсов обеих материальных точек относительно некоторого центра O . Согласно (7.2) имеем

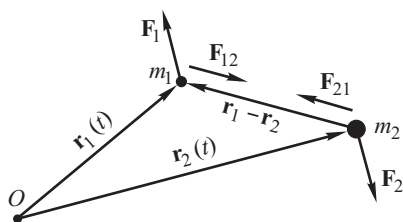


Рис. 7.3

тогда имеем

$$\begin{aligned}\frac{d\mathbf{L}_1}{dt} &= [\mathbf{r}_1 \mathbf{F}_{12}] + [\mathbf{r}_1 \mathbf{F}_1], \\ \frac{d\mathbf{L}_2}{dt} &= [\mathbf{r}_2 \mathbf{F}_{21}] + [\mathbf{r}_2 \mathbf{F}_2],\end{aligned}$$

где \mathbf{F}_{12} , \mathbf{F}_{21} — силы взаимодействия тел друг с другом,

\mathbf{F}_1 , \mathbf{F}_2 — некоторые результирующие внешние силы, действующие со стороны каких-то других тел на тела 1 и 2, соответственно. Сложив левые и правые части этих уравнений, получаем с учетом того, что $\mathbf{F}_{12} = -\mathbf{F}_{21}$

$$\frac{d(\mathbf{L}_1 + \mathbf{L}_2)}{dt} = [(\mathbf{r}_1 - \mathbf{r}_2) \mathbf{F}_{12}] + \mathbf{M}_{1\text{вн}} + \mathbf{M}_{2\text{вн}}, \quad (7.3)$$

где $\mathbf{M}_{1\text{вн}} = [\mathbf{r}_1 \mathbf{F}_1]$ и $\mathbf{M}_{2\text{вн}} = [\mathbf{r}_2 \mathbf{F}_2]$ — моменты внешних сил относительно центра O . Векторы $(\mathbf{r}_1 - \mathbf{r}_2)$ и \mathbf{F}_{12} параллельны,

поэтому их векторное произведение равно нулю. Следовательно, уравнение (7.3) для суммарного момента импульса $\mathbf{L} = \mathbf{L}_1 + \mathbf{L}_2$ системы двух материальных точек имеет вид

$$\frac{d\mathbf{L}}{dt} = \mathbf{M}_{1\text{вн}} + \mathbf{M}_{2\text{вн}} \equiv \mathbf{M}_{\text{вн}}. \quad (7.4)$$

Здесь $\mathbf{M}_{\text{вн}}$ естественно трактовать как полный момент внешних сил, приложенных к системе материальных точек. В случае замкнутой системы материальных точек внешний момент равен нулю, так что оказывается, что *момент импульса замкнутой системы материальных точек относительно произвольного центра остается постоянным.*

Таким образом, мы сформулировали *закон сохранения момента импульса.*

В силу принятого предположения о центральном характере сил взаимодействия между материальными точками этот закон все еще остается следствием второго закона Ньютона (7.1) и подобием закона сохранения импульса. Сделаем следующий шаг: откажемся от каких-либо предположений о характере взаимодействия внутри замкнутой системы. Тем самым теряет доказательную силу и наш вывод, но сам закон сохранения момента импульса остается уже как *экспериментальный* и *дополнительный* к законам движения материальной точки, сформулированным в гл. 3. (В аналитической механике, изучающей самые общие законы движения, закон сохранения энергии связан с однородностью времени, т.е. с инвариантностью физических законов относительно изменения начала отсчета времени, закон сохранения импульса замкнутой системы связывается с однородностью пространства, т.е. инвариантностью относительно пространственных сдвигов, тогда как закон сохранения момента импульса — с изотропностью пространства (инвариантностью относительно вращений пространства).)

Равным образом и уравнение (7.4) в общем случае не есть следствие (7.1), но представляет собой дополнительное уравнение движения. Соответствующий закон гласит, что скорость изменения со временем полного момента импульса замкнутой системы определяется только моментами внешних сил, а именно:

$$\frac{d}{dt} \sum_i \mathbf{L}_i = \sum_k \mathbf{M}_{\text{вн}k}. \quad (7.5)$$

Уравнение (7.5) принято называть уравнением моментов. Операции суммирования в левой и правой части (7.5) не эквивалентны: моменты импульса суммируются по всем частицам (или материальным точкам, или физически бесконечно малым объемам), а в правой части суммирование происходит по всем моментам внешних сил.

Подчеркнем еще раз, что момент силы в данном параграфе понимается именно по отношению к некоторой точке O , из которой проводится радиус-вектор \mathbf{r} в точку приложения силы. В действительности, как легко усмотреть из рис. 7.4а, с точки зрения уравнения моментов (7.5), имеет значение не точка, а линия приложения силы: перенося последнюю вдоль линии AA' , мы сохраняем плечо $h = r \sin \alpha$, а следовательно, не меняется и момент силы. Отметим также, что сумма моментов сил, приложенных к телу, именно вследствие этого свойства не равна моменту суммы сил. Наиболее выразительный пример — т. н. пара сил — представлен на рис. 7.4б. Здесь сумма сил, приложенных к телу, равна нулю (тем самым равно нулю и ускорение центра масс), но полный момент $M = h \cdot F \neq 0$. И кстати, полезно в качестве упражнения убедиться, что ни величина, ни направление момента пары сил не зависят от выбора точки O .

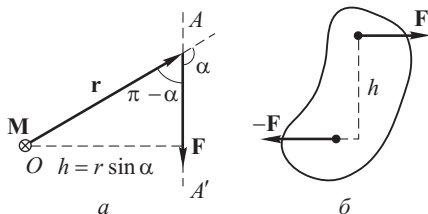


Рис. 7.4

7.2. Законы Кеплера

Еще задолго до рождения Ньютона немецкий ученый И. Кеплер (1571–1630) на основании обобщения астрономических наблюдений датского астронома Тихо Браге (1546–1601) установил три закона движения планет.

1. *Каждая планета движется по эллипсу, в одном из фокусов которого находится Солнце.*

2. *Радиус-вектор планеты, проведенный из Солнца как начала координат, описывает в равные промежутки времени равные площади.*

3. *Квадраты периодов обращения планет по эллипсам относятся как кубы больших полуосей.*

Разумеется, эти законы сформулированы с точностью до отношения радиусов Солнца и планет к характерным расстояниям между ними, но точность эта, как и математическая содержательность законов Кеплера, для начала XVII века просто поразительны.

Поскольку законы Кеплера оказались в свое время главным тестом для механики Ньютона, и поскольку они принципиально важны для описания динамики планет, мы посвятим настоящий параграф их последовательному выводу.

Напомним, что для частного случая центральных сил уравнение моментов не является дополнительным ко второму закону Ньютона, но оно прямо из него следует. В соответствии с законом всемирного тяготения (6.26) сила тяготения, связывающая Солнце и любую из планет, есть центральная сила, причем вектор \mathbf{r}_{12} в (6.26) — это просто радиус-вектор планеты в системе отсчета Солнца.

Мы будем пренебрегать взаимодействием планет друг с другом, так как массы их много меньше массы Солнца, хотя главной поправкой к нашим, т.е. кеплеровским, результатам будет как раз влияние планет-гигантов, особенно Юпитера. В силу того же обстоятельства приведенную массу в (6.2) можно просто приравнять массе планеты и пренебречь смещением Солнца относительно общего центра масс на фоне движения планет.

В приближении замкнутой системы «Солнце–планета» момент импульса с точностью до отношения масс равен

$$\mathbf{L} = m[\mathbf{r} \mathbf{v}] = \text{const}, \quad (7.6)$$

где m , \mathbf{r} , \mathbf{v} — соответственно масса, радиус-вектор и скорость планеты. Из того, что вектор \mathbf{L} перпендикулярен \mathbf{r} , \mathbf{v} , в частности, следует, что движение планеты — плоское, т.е. ее траектория лежит в некоторой плоскости, что с необходимостью предполагается законами Кеплера.

Далее, обратим внимание на то, что по определению векторного произведения (см. рис. 7.5),

$$|[\mathbf{r} \mathbf{v}]| dt = rv \sin \varphi \cdot dt \equiv 2 dS,$$

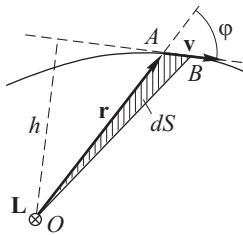


Рис. 7.5

где dS — малый элемент площади (площадь малого треугольника, ограниченного векторами \mathbf{r} и $\mathbf{v} dt$). Как следствие, из сохранения момента импульса (7.6) вытекает постоянство секториальной скорости:

$$\frac{dS}{dt} = \text{const}, \quad (7.7)$$

а это не что иное, как второй закон Кеплера.

В дальнейшем, помимо сохранения момента импульса, мы будем использовать закон сохранения энергии:

$$E = \text{const} = \frac{mv^2}{2} - \gamma \frac{Mm}{r}, \quad (7.8)$$

где M — масса Солнца.

Рассмотрение удобно проводить в полярных координатах в плоскости траектории планеты (рис. 7.6 а), соответственно разложив скорость \mathbf{v} на радиальную $v_r = \dot{r}$ и касательную $v_\tau = r\dot{\varphi}$

составляющие. Тогда $L = mv_\tau r = mr^2\dot{\varphi}$, и (7.8) можно переписать в виде

$$E = \frac{m}{2} v_r^2 + V(r); \quad (7.9)$$

где $V(r) \equiv -\gamma Mm/r + L^2/(2mr^2)$. График $V(r)$ представлен на рис. 7.6 б. Случай $L = 0$ (штриховая кривая) отвечает лобовому столкновению тела со звездой и никак не может соответствовать закону движения планет. При $L \neq 0$ сразу можно выделить два принципиально разных сценария. Если $E \geq 0$, тело, взаимодействующее со звездой, может уйти на бесконечность, так что и этот случай не имеет отношения к закону движения планет. Если же $E < 0$, то движение, как говорят, *финитно*, т. е. происходит в ограниченной области пространства. При этом радиальная скорость v_r обязательно должна быть знакопеременной — в противном случае планета либо столкнулась бы со звездой, либо ушла бы на бесконечность, впрочем, смена знака v_r прямо следует из зависимости (7.9). При условии $v_r = 0$ уравнение (7.9) имеет два корня r_1, r_2 — см. рис. 7.6 б — их называют точками поворота. Как показано на том же рисунке, в случае $E \geq 0$ траектория также имеет точку поворота r'_1 — но лишь одну, так что финитное движение невозможно. (Дело в том, что по самому смыслу полярных координат — рис. 7.6 а — физически значимый ответ для r может быть только положительным). Уравнение (7.9)

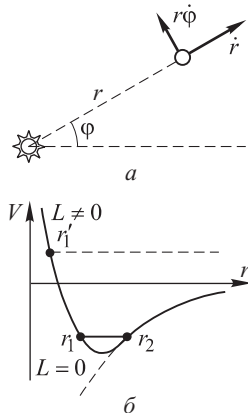


Рис. 7.6

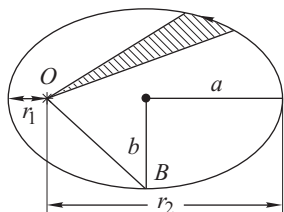


Рис. 7.7

описывает кривую второго порядка, но единственная кривая второго порядка, не уходящая на бесконечность, есть эллипс:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

Для полноты картины упомянем, что траектория пролетающего тела в поле тяготения звезды при $E > 0$ представляет собой гиперболу, а в вырожденном случае $E = 0$, соответствующем второй космической скорости, — параболу.

Рассмотрим подробнее эллиптическую траекторию (рис. 7.7). При $v_r = 0$ соотношение (7.9) вырождается в квадратное уравнение:

$$r^2 + \gamma \frac{Mm}{E} - \frac{L^2}{2mE} = 0.$$

Его корни, сумма которых, как видно из рисунка, равна длине большой оси эллипса, удовлетворяют соотношению:

$$2a = r_1 + r_2 = -\gamma \frac{Mm}{E} \implies E = -\gamma \frac{Mm}{2a}. \quad (7.10)$$

Постоянную секториальную скорость $\dot{S} = L/2m$ удобно привязать к точке B , где скорость v ортогональна малой оси эллипса: $\dot{S} = b \cdot v/2$. Принимая во внимание свойство фокуса эллипса $OB = a$, приходим к соотношению

$$\left. \frac{mv^2}{2} \right|_B = E - U = E + \gamma \frac{Mm}{a} = \gamma \frac{Mm}{2a}.$$

Таким образом, $v|_B \propto 1/\sqrt{a}$, а значит, $b = 2\dot{S}/v \propto \dot{S}\sqrt{a}$. Как известно, площадь эллипса равна $S = \pi ab$, с другой стороны, ее можно выразить через секториальную скорость: $S = \dot{S} \cdot T$, где T — период обращения. Окончательно,

$$a\dot{S}\sqrt{a} \propto \dot{S} \cdot T \implies T^2 \propto a^3. \quad (7.11)$$

Итак, все три закона Кеплера в рамках ньютоновой механики оказываются прямым следствием закона всемирного тяготения и уравнения моментов.

7.3. Вращение абсолютно твердого тела вокруг неподвижной оси

Переходя к рассмотрению движения макроскопических тел, определим сначала наиболее простой тип движения, зависящего от геометрии тела. И первый шаг на этом пути — исключение из рассмотрения движения деформируемых тел, так как заранее очевидно, что движения многих частей тела друг относительно друга (и, соответственно, их математическое описание) могут быть достаточно сложны. Это означает, что мы начнем с изучения движения *абсолютно твердых тел*. Но и среди всевозможных движений абсолютно твердого тела мы выберем, по-возможности, наиболее простое. Дело в том, что для однозначного определения положения абсолютно твердого тела в пространстве необходимо задать шесть независимых величин — *шесть степеней свободы*. Мы же на данный момент ограничимся простейшим, но практически важным случаем, когда существует лишь одна степень свободы, — *вращением абсолютно твердого тела вокруг неподвижной оси*. При этом изменение положения тела в пространстве однозначно определяется единственной координатой φ — углом поворота тела вокруг оси.

Вывод основного уравнения для вращения тела вокруг неподвижной оси мы для наглядности проведем на примере вращения тонкого однородного диска вокруг вертикальной оси, перпендикулярной его поверхности и проходящей через его центр (рис. 7.8). Слово «тонкий» означает, что толщина диска намного меньше его радиуса и что всеми эффектами, зависящими от толщины диска, можно пренебречь. Чтобы осуществить такое вращение, диск жестко соединен с твердым стержнем, направленным вдоль оси Oz и заключенным в подшипники. Опирающиеся на подшипники фланцы $\Phi\text{л}$, предотвращают перемещение диска в вертикальном направлении. Будем полагать, что масса и радиус стержня пренебрежимо малы по сравнению соответственно с массой и радиусом диска. Допустим также, что в некоторой точке на расстоянии h от оси к диску приложена внешняя сила \mathbf{F} , модуль и направление которой могут, вообще говоря, меняться со временем. Масса диска равна M , радиус диска — R .

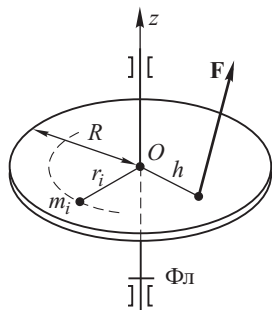


Рис. 7.8

Нашей целью является вывод уравнения движения диска, то есть уравнения, с помощью которого можно было бы находить функцию $\varphi(t)$, которая играет в данной задаче ту же роль, что и функция $\mathbf{r}(t)$ в задачах о движении материальной точки.

При выводе мы воспользуемся основным приемом, который применяется в механике для изучения движения тел конечных размеров. Все тело мысленно разбивается на совокупность малых частичек, которые можно рассматривать как материальные точки, взаимодействующие друг с другом и с другими телами. В результате задача сводится к задаче о движении системы материальных точек, для которых справедливы все законы движения, рассмотренные в предыдущих главах. Такой выделенный элемент макроскопического тела нередко называют физически бесконечно малым объемом, полагая, что все физические параметры внутри каждого из этих объемов совершенно однородны, а взаимодействие существенно лишь с внешним окружением. (Математически эта идея выражается удержанием лишь первого члена ряда Тэйлора во всех функциональных зависимостях.)

Итак, разобьем мысленно наш диск на совокупность материальных точек с массами m_i ($\sum m_i = M$). Мы уже отмечали, что при изучении вращательных движений целесообразно пользоваться не вторым законом Ньютона, но уравнением моментов. Для системы материальных точек уравнение для суммарного

момента импульса \mathbf{L} имеет вид

$$\frac{d\mathbf{L}}{dt} = \sum \mathbf{M}_{\text{вн}},$$

где $\mathbf{L} = \sum \mathbf{L}_i$, а \mathbf{L}_i — момент импульса i -й материальной точки системы относительно одного и того же центра O .

Выберем центр O в точке пересечения оси вращения с плоскостью диска. Все образующие диск материальные точки будут при его вращении двигаться по окружностям с радиусами r_i , где r_i — модуль радиуса-вектора, соединяющего соответствующую i -ю материальную точку с центром O . Момент импульса отдельной материальной точки определяется векторным произведением:

$$\mathbf{L}_i = [\mathbf{r}_i \mathbf{p}_i] = m_i [\mathbf{r}_i \mathbf{v}_i],$$

где \mathbf{p}_i — импульс, а \mathbf{v}_i — скорость i -й материальной точки. Отсюда следуют две важные особенности нашей задачи. Во-первых, поскольку векторное произведение направлено перпендикулярно плоскости, в которой лежат образующие его векторы, то все моменты импульса \mathbf{L}_i направлены вдоль оси Oz , так как плоскость векторов \mathbf{r}_i и \mathbf{v}_i перпендикулярна этой оси. Это означает, что у суммарного момента импульса диска \mathbf{L} отлична от нуля только одна компонента $L_z = \sum L_{iz}$. Во-вторых, при вращении диска все образующие его материальные точки движутся по окружностям, поэтому \mathbf{r}_i и \mathbf{v}_i всегда перпендикулярны друг другу, синус угла между ними равен единице, и для L_{iz} в соответствии с определением векторного произведения получаем

$$L_{iz} = m_i r_i v_i = m_i r_i^2 \omega. \quad (7.12)$$

Мы учли, что $v_i = r_i \omega$, где $\omega = d\varphi/dt$ — угловая скорость вращения, одинаковая для всех точек диска. Знак L_{iz} зависит от того, куда направлено векторное произведение $[\mathbf{r} \mathbf{v}]$ — вдоль выбранного положительного направления оси Oz или против него. Мы выбрали для определенности знак «плюс».

Вернемся теперь к уравнению моментов. Левая часть этого уравнения для отличной от нуля компоненты L_z с учетом (7.12) принимает вид

$$\frac{dL_z}{dt} = \frac{d \sum L_{iz}}{dt} = \left(\sum m_i r_i^2 \right) \frac{d\omega}{dt}. \quad (7.13)$$

Величина, равная сумме произведений элементарных масс на квадрат их расстояний от некоторой оси Oz , называется *моментом инерции тела относительно данной оси* и обозначается через I_{Oz} :

$$I_{Oz} = \sum m_i r_i^2, \quad (7.14)$$

суммирование производится по всем элементарным массам m_i , на которые мысленно разбито тело. Используя (7.12) и (7.14), вы-

ражение для z -компоненты суммарного момента импульса диска можно записать в виде

$$L_z = \sum_i L_{iz} = I_{Oz} \omega. \quad (7.15)$$

Таким образом, левая часть уравнения моментов может быть записана как

$$\frac{dL_z}{dt} = I_{Oz} \frac{d\omega}{dt},$$

откуда, подставляя $\omega = d\varphi/dt$, окончательно получаем

$$I_{Oz} \frac{d^2\varphi}{dt^2} = \sum_k M_{kz \text{ вн}}. \quad (7.16)$$

Сделаем одно замечание, касающееся принятой в механике терминологии. Проекция вектора момента силы \mathbf{M} на некоторую ось Oz , проходящую через центр O , относительно которого определен \mathbf{M} , называется *моментом силы относительно оси Oz* , который мы будем обозначать через $M_{\text{оси}}$. С учетом этого замечания уравнение (7.16) можно также представить в виде

$$I_{Oz} \frac{d\omega}{dt} = \sum M_{\text{оси}}. \quad (7.17)$$

Хотя мы вывели уравнение (7.17) для частного случая (тонкого диска), этот вывод посредством интегрирования по слоям обобщается на случай вращения вокруг неподвижной оси тел любой формы. Итак, основное уравнение для вращения абсолютно твердого тела вокруг неподвижной оси Oz можно записать в виде

$$I_{Oz} \frac{d^2\varphi}{dt^2} = \sum M_{\text{оси}}.$$

Как и второй закон Ньютона, оно является дифференциальным уравнением. Нахождение из этого уравнения функции $\varphi(t)$ — зависимости угла поворота тела от времени, решает одну из основных задач механики: предсказать положение движущегося тела в различные моменты времени при задании двух начальных условий: начального положения и скорости. В нашем случае начальные условия — это значение угла поворота в начальный момент времени, $\varphi(0) = \varphi_0$, и начальной угловой скорости $\omega(0) = \omega_0$. Значения φ_0 и ω_0 определяются условиями конкретной задачи.

Таким образом, задание начальных условий является одним из важнейших элементов строгой математической постановки конкретной задачи о вращении твердого тела.

Характерной отличительной особенностью задачи о вращении тела вокруг оси по сравнению с задачей о движении материальной точки является то, что теперь в основное уравнение входит

не масса m , а момент инерции I_{Oz} , и не сила \mathbf{F} , а момент силы относительно оси $M_{оси}$.

Подчеркнем, что момент инерции тела относительно некоторой оси существует безотносительно к тому, вращается тело вокруг этой оси или покоится. Если же тело вращается вокруг некоторой оси под действием внешних сил, то *момент инерции служит мерой инертности тела при вращении*. В самом деле, при одном и том же моменте внешних сил относительно оси вращения угловое ускорение будет меньше у того тела, у которого больше момент инерции. Таким образом, момент инерции тела при вращательном движении играет ту же роль меры инертности, что его масса при поступательном движении.

Как видно из определения (7.14), задание полной массы тела m еще ничего не говорит о величине его момента инерции I_{Oz} , который зависит от того, как расположены различные части тела (различные элементы его полной массы) относительно той или иной оси Oz , конкретное положение которой определяется условиями задачи. Поэтому при решении каждой конкретной задачи о вращении вопрос о величине I_{Oz} требует отдельного рассмотрения. Покажем на примере нашей исходной задачи о вращении тонкого диска, как определяется момент инерции. Здесь нам необходимо найти момент инерции однородного диска относительно оси, перпендикулярной к плоскости диска и проходящей через его центр (см. рис. 7.8). Каждый из элементарных объемов с массой m_i , на которые мы мысленно разбили весь диск, можно представить как произведение плотности тела ρ на соответствующий элементарный объем ΔV_i :

$$m_i = \rho \Delta V_i.$$

Следовательно, выражение (7.14) для момента инерции I_{Oz} можно представить в виде

$$I_{Oz} = \sum \rho \Delta V_i r_i^2 = \rho \sum r_i^2 \Delta V_i,$$

где мы учли, что для однородного диска плотность постоянна, и поэтому вынесли ρ за знак суммирования. При переходе к бесконечно малым значениям ΔV_i суммирование сводится к интегрированию:

$$I_{Oz} = \int \rho(\mathbf{r}) r^2 dV. \quad (7.18)$$

Ради большей общности результата мы вернули $\rho(\mathbf{r})$ под знак интеграла.

Симметрия задачи подсказывает наиболее простой путь вычисления I_{Oz} в случае однородного тела вращения. Разобьем наш диск на кольцевые слои толщиной dr . Объем такого слоя равен

$$dV = b \cdot 2\pi r dr,$$

где b — толщина диска. Подставляя это выражение в (7.18), получаем

$$I_{Oz} = 2\pi b\rho \int_0^R r^3 dr = 2\pi b\rho \frac{R^4}{4}.$$

Последний результат можно выразить через полную массу диска m , равную произведению плотности ρ на объем диска $b \cdot \pi R^2$. Окончательно получаем

$$I_{Oz} = \frac{mR^2}{2}. \quad (7.19)$$

Нахождение момента инерции в рассмотренном примере сильно упрощалось тем, что тело было однородным и симметричным, а момент инерции мы вычисляли относительно оси симметрии. Если бы мы захотели найти момент инерции диска относительно, например, оси, перпендикулярной к диску и проходящей через его край, то вычисления оказались бы более сложными. В подобных случаях нахождение момента инерции значительно облегчается, если воспользоваться следующей теоремой: «момент инерции I относительно произвольной оси равен сумме момента инерции I_C относительно оси, параллельной данной и проходящей через центр масс тела, и произведения массы тела m на квадрат расстояния a между осями», т. е.

$$I = I_C + ma^2. \quad (7.20)$$

Что касается центра масс твердого тела, то он определяется по уже известному правилу

$$\mathbf{R}_C = \frac{\int_V \mathbf{r}\rho(\mathbf{r}) dV}{\int_V \rho(\mathbf{r}) dV}.$$

Для доказательства соотношения (7.20) предположим, что «старая» и «новая» оси связаны вектором параллельного переноса \mathbf{R} (рис. 7.9):

$$\mathbf{r}_i = \mathbf{r}_{i0} + \mathbf{R},$$

где \mathbf{r}_i и \mathbf{r}_{i0} — старый и новый радиусы-векторы i -й точки в плоскости ее вращения. Тогда моменты инерции связаны следующим образом:

$$\begin{aligned} I &= \sum_i m_i r_i^2 = \sum_i m_i (\mathbf{r}_{i0} + \mathbf{R})^2 = R^2 \sum_i m_i + \\ &+ 2\mathbf{R} \sum_i m_i \mathbf{r}_{i0} + \sum_i m_i r_{i0}^2 = mR^2 + 2\mathbf{R} \cdot m\mathbf{R}_C + I_0. \end{aligned}$$

Если «старая» ось проходила через центр инерции, то $\mathbf{R}_C = 0$, а $I_0 = I_C$. Теорема доказана. Ее обычно называют *теоремой Штейнера* (по имени швейцарского геометра Я. Штейнера

(1796–1863)) или Гюйгенса–Штейнера. Отметим, что в ней отнюдь не предполагается какая-либо особая симметрия тела.

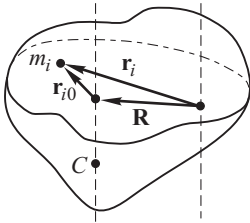


Рис. 7.9

В соответствии с названной теоремой, момент инерции диска относительно оси, перпендикулярной поверхности диска и проходящей через его край, равен найденному нами моменту инерции (7.19) относительно оси, проходящей через центр диска, плюс mR^2 (расстояние между осями равно радиусу диска R):

$$I_{O'z'} = \frac{mR^2}{2} + mR^2 = \frac{3}{2} mR^2.$$

В отличие от второго закона Ньютона, в правую часть уравнения вращательного движения входит не сила, а момент силы относительно оси. При этом он играет роль, аналогичную той, которую играет сила для поступательного движения.

В реальных задачах непосредственно сам момент силы обычно не задан, а заданы модули и направления внешних сил и точки приложения этих сил к телу, как, например, было в нашей задаче о вращении тонкого диска (см. рис. 7.8). Поэтому для определения $M_{\text{оси}}$ в конкретных задачах необходимо воспользоваться тем, что $M_{\text{оси}}$ является, по определению, проекцией момента силы \mathbf{M} относительно некоторого центра на ось вращения. Если ось вращения направлена вдоль координатной оси Oz , то

$$M_{\text{оси}} = M_z = [\mathbf{r}\mathbf{F}]_z.$$

Покажем, как выражается $M_{\text{оси}}$ через условия задачи о вращении тонкого диска (рис. 7.10). Будем полагать, что центр O , относительно которого определен момент силы \mathbf{F} , расположен в произвольной точке на оси вращения. Разложим вектор силы \mathbf{F} на три взаимно перпендикулярные составляющие, две из которых, \mathbf{F}_{\parallel} и \mathbf{F}_h , параллельная и перпендикулярная оси вращения, лежат в плоскости, проходящей через ось вращения и точку приложения внешней силы, а третья \mathbf{F}_{\perp} — перпендикулярна к этой плоскости (эта составляющая обозначена на рисунке кружком с крестиком). Если в плоскости диска представить себе окружность радиуса h с центром на оси Oz , то составляющая \mathbf{F}_{\perp} будет направлена по касательной к этой окружности. Момент силы \mathbf{F} относительно центра O равен сумме моментов составляющих: $\mathbf{M} = \mathbf{M}_{\parallel} + \mathbf{M}_h + \mathbf{M}_{\perp}$, где

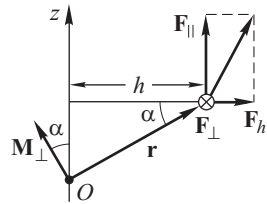


Рис. 7.10

$$\mathbf{M}_{\parallel} = [\mathbf{r}\mathbf{F}_{\parallel}], \quad \mathbf{M}_h = [\mathbf{r}\mathbf{F}_h], \quad \mathbf{M}_{\perp} = [\mathbf{r}\mathbf{F}_{\perp}].$$

Векторное произведение направлено перпендикулярно плоскости, в которой лежат образующие его векторы. Поэтому векторы \mathbf{M}_{\parallel} и \mathbf{M}_h перпендикулярны оси Oz , и следовательно, их проекции на эту ось равны нулю. Момент \mathbf{M}_{\perp} имеет модуль, равный rF_{\perp} . Если составляющая \mathbf{F}_{\perp} на рис. 7.10 направлена «от нас», то ее момент \mathbf{M}_{\perp} по правилу винта направлен так, как изображено на рисунке, т.е. образует с осью Oz угол α , косинус которого равен h/r . Следовательно, в этом случае проекция на ось Oz вектора \mathbf{M}_{\perp} положительна и имеет величину $M_{\perp} \cos \alpha = hF_{\perp}$. Если вектор \mathbf{F}_{\perp} направлен на рисунке «на нас», то \mathbf{M}_{\perp} следует направить на рисунке в противоположную сторону, и тогда проекция этого вектора на ось будет иметь величину $M_{\perp} \cos(\pi + \alpha) = -hF_{\perp}$. Другие компоненты \mathbf{M} (\mathbf{M}_{\parallel} и \mathbf{M}_h) проекций на ось Oz , как уже отмечалось, не имеют. Следовательно, $M_z = M_{\perp,z} = \pm hF_{\perp}$. Таким образом, получаем для момента силы относительно оси вращения:

$$M_{\text{оси}} = M_z = \pm hF_{\perp}. \quad (7.21)$$

Подчеркнем еще раз, что по правилу винта (см. (7.2)) знак «плюс» соответствует случаю, когда внешняя сила приводит к вращению вокруг направления оси Oz против часовой стрелки, а знак «минус» соответствует вращению тела в противоположном направлении.

7.4. Следствия уравнения моментов

Рассмотрим несколько характерных задач, представляющих движение либо равновесие тела с закрепленной осью.

Первый пример — вращение однородного цилиндра вокруг неподвижной оси, совпадающей с его геометрической осью симметрии Oz , в случае, когда по касательной к его поверхности и перпендикулярно к оси вращения на цилиндр действует постоянная сила \mathbf{F}_0 (рис. 7.11 а). Пусть масса цилиндра m , радиус — R , длина — l . Такая задача является примером вращения вала какого-либо механизма. В реальной ситуации между осью и валом всегда еще действуют силы трения. Мы будем полагать, что они достаточно малы, чтобы ими можно было пренебречь. Тогда единственный момент силы, не равный нулю, есть

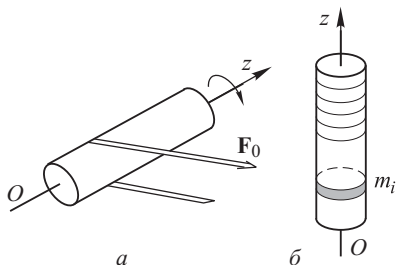


Рис. 7.11

$$M_{\text{оси}} = F_0 R.$$

Момент инерции цилиндра относительно оси симметрии легко найти, если мысленно представить цилиндр как совокупность тонких дисков с массами m_i , «нанизанных» на ось Oz (рис. 7.11 б), откуда сразу следует, что искомая величина дается все той же формулой (7.19):

$$I_{Oz} = \frac{1}{2} m R^2.$$

Разумеется, теперь в ответ входит полная масса цилиндра m . Таким образом, уравнение моментов (7.16) приобретает вид

$$\frac{d^2\varphi}{dt^2} = \frac{2F_0}{mR}. \quad (7.22)$$

Чтобы решение было однозначно определено, необходимо, как мы уже знаем, задать начальные условия: начальные значения угла поворота φ и угловой скорости $\omega = d\varphi/dt$. Выберем, например, начальные условия в виде: $\varphi(0) = 0$, $\omega(0) = 0$.

Уравнение (7.22) с математической точки зрения полностью аналогично уравнению, определяющему движение тела, падающего под действием силы тяжести, и его решение также находится без труда на основе элементарных правил вычисления производных от простейших функций. Удовлетворяющее начальным условиям решение уравнения (7.22) имеет вид

$$\varphi(t) = \frac{F_0}{mR} t^2.$$

Мы уже отмечали, что одной из особенностей вращательного движения тела вокруг неподвижной оси, отличающей его от поступательного движения, является зависимость вращения не непосредственно от величины приложенной к телу силы, а от момента этой силы относительно оси вращения. Так, в рассмотренном выше примере движение диска определяется не величиной силы F_0 , а ее моментом относительно оси вращения, то есть произведением силы на *плечо* (в нашем случае — $F_0 R$). И даже при очень малом значении модуля силы это произведение может иметь большую величину, если плечо силы достаточно велико.

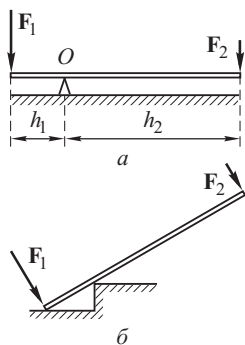


Рис. 7.12

Эта особенность вращательного движения вокруг неподвижной оси лежит в основе работы *рычага* — одного из древнейших орудий труда человека. Схематически любой рычаг можно представить, как тонкий стержень, который может вращаться вокруг неподвижной оси, перпендикулярной стержню и проходящей не через его центр. На рис. 7.12 а изображен такой стержень, который может вращаться вокруг горизонтальной оси, перпендикулярной стержню

(в нашем случае — перпендикулярно плоскости рисунка). Точка O стержня, через которую проходит ось вращения, делит стержень на неравные отрезки длиной h_1 и h_2 . Допустим стержень расположен горизонтально, и к концам стержня, перпендикулярно ему и оси вращения, приложены две силы — \mathbf{F}_1 и \mathbf{F}_2 . Будем считать стержень достаточно легким, так что действием силы тяжести по сравнению с силами \mathbf{F}_1 и \mathbf{F}_2 можно пренебречь.

Из уравнения моментов (7.5) можно найти условие равновесия стержня, то есть условие, при котором стержень будет оставаться неподвижным или вращаться с постоянной угловой скоростью. В нашем случае уравнение моментов имеет вид

$$I_0 \frac{d\omega}{dt} = F_2 h_2 - F_1 h_1, \quad (7.23)$$

где I_0 — момент инерции стержня относительно оси вращения. Моменты сил \mathbf{F}_1 и \mathbf{F}_2 относительно оси вращения входят в уравнение с разными знаками в соответствии с «правилом винта»: в одном случае сила стремится повернуть стержень по часовой стрелке, а в другом — против. Приравнявая левую часть уравнения (7.23) нулю, получаем условие равновесия в виде

$$F_1 h_1 = F_2 h_2.$$

Из него следует, что как бы ни была велика, например, сила \mathbf{F}_1 , ее всегда можно уравновесить малой силой \mathbf{F}_2 , если выбрать плечо h_2 достаточно большим по сравнению с h_1 .

Приспособления, использующие рассмотренное выше свойство вращательного движения вокруг неподвижной оси, получили название рычагов. В качестве примера рис. 7.12 б иллюстрирует, как свойство рычага используется при работе лопаты. «Правило рычага» было известно уже в глубокой древности, и впервые оно было, по-видимому, четко сформулировано в третьем веке до нашей эры знаменитым древнегреческим ученым и изобретателем Архимедом.

Выведем теперь выражение для кинетической энергии тела, вращающегося вокруг неподвижной оси, которое может оказаться полезным при изучении некоторых особенностей вращательного движения на основе законов сохранения. Пусть тело вращается вокруг оси Oz (рис. 7.13). Разобьем мысленно тело на элементарные массы m_i . Линейная скорость элементарной массы m_i равна $v_i = \omega r_i$, где r_i — расстояние массы m_i от оси Oz . Следовательно, для кинетической энергии i -й элементарной массы получается выражение

$$T_i = \frac{m_i v_i^2}{2} = \frac{1}{2} m_i \omega^2 r_i^2.$$

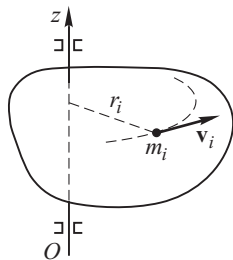


Рис. 7.13

Кинетическая энергия тела складывается из кинетических энергий его частей:

$$T = \sum T_i = \frac{1}{2} \omega^2 \sum m_i r_i^2.$$

Сумма в правой части этого соотношения представляет собой момент инерции тела I_{O_z} относительно оси вращения. Таким образом, кинетическая энергия тела, вращающегося вокруг неподвижной оси, равна

$$T = \frac{1}{2} I_{O_z} \omega^2 = \frac{L^2}{2I_{O_z}}. \quad (7.24)$$

Сравнивая соотношения (7.17) и (7.24), соответственно, со вторым законом Ньютона и выражением для кинетической энергии материальной точки, можно убедиться в существовании между ними глубокой аналогии. Поэтому в случае, когда движение тела есть изменение во времени не линейных, но угловых (ориентационных) координат, говорят, что \mathbf{L} есть *обобщенный импульс*, \mathbf{M} — *обобщенная сила*, а I_{O_z} — *обобщенная масса*, соответствующие *обобщенной координате* φ . (Эта терминология построена отнюдь не ради данного частного случая — на таком языке формулируется механика в самом общем виде, который называется аналитической механикой.)

7.5. Трехмерное движение твердого тела. Гироскопы

Замечание, сделанное в конце предыдущего параграфа, побуждает нас представить уравнения (7.17) и (7.24) в векторной форме, для чего целесообразно включить в рассмотрение понятие *вектора угловой скорости* и следующие из этого понятия формулы (2.13):

$$\frac{d\mathbf{r}}{dt} = [\boldsymbol{\omega}\mathbf{r}]; \quad \frac{d\mathbf{v}}{dt} = [\boldsymbol{\omega}\mathbf{v}]. \quad (7.25)$$

С учетом (7.25) формула (7.12) должна быть переписана в виде

$$\mathbf{L}_i = m_i r_i^2 \boldsymbol{\omega},$$

а (7.15) — соответственно,

$$\mathbf{L} = I_{O_z} \boldsymbol{\omega}, \quad (7.26)$$

наконец, уравнение моментов (7.17) принимает следующую форму:

$$I_{O_z} \frac{d\boldsymbol{\omega}}{dt} = \sum_i \mathbf{M}_i. \quad (7.27)$$

Сколь ни велик соблазн отнести уравнения (7.26), (7.27) к произвольному неоднородному движению, делать это надлежит с большой осторожностью, поскольку выводились они для частного случая тела с закрепленной осью, когда вектор $\boldsymbol{\omega}$ удерживается неизменным по направлению, а достигается это в общем случае несимметричного вращающегося тела посредством

момента реакции \mathbf{M}' в опорах оси, дополнительного к правой части (7.17), учитывающей только вращающие моменты.

В общем случае трехмерного движения твердого тела произвольной формы (точнее — произвольной геометрии массы) реви-зии подвергается прежде всего формула (7.26). Векторы $\boldsymbol{\omega}$ и \mathbf{L} оказываются, вообще говоря, не параллельны. Соответственно, уравнение моментов (7.15) лишь в случае осевой симметрии сводится к (7.27).

Оказывается, однако, что в каждом твердом теле имеются три взаимно перпендикулярные, проходящие через центр масс, оси, такие, что при вращении твердого тела вокруг них зависимость возникающего момента импульса от угловой скорости вращения имеет простой вид: $\mathbf{L} = I\boldsymbol{\omega}$, где I — момент инерции тела относительно оси вращения. Эти оси называются *главными осями инерции твердого тела*. Мы не будем останавливаться на доказательстве этого утверждения, а рассмотрим основные следствия из него.

Представим себе, что мы сообщили телу вращение с угловой скоростью $\boldsymbol{\omega}$ вокруг одной из его главных осей. Возникающий у тела момент импульса \mathbf{L} равен, согласно вышесказанному, $\mathbf{L} = I\boldsymbol{\omega}$. Если при этом на тело не действуют внешние силы или суммарный момент внешних сил равен нулю, то, согласно (7.27), $\mathbf{L} = I\boldsymbol{\omega} = \text{const}$, и, следовательно, $\boldsymbol{\omega} = \text{const}$. Другими словами, *в случае, когда момент действующих на тело внешних сил равен нулю, угловая скорость вращения тела вокруг любой из его главных осей инерции сохраняется по величине и направлению*.

Определение фактического положения главных осей в твердом теле облегчается правилом, которое гласит: все оси симметрии твердого тела являются его главными осями инерции. (В частном случае тела вращения происходит, как принято говорить, вырождение: любая ось, перпендикулярная оси симметрии, является главной осью, так что в этом случае их будет не три, а бесконечно много.) Поэтому поведение твердых тел, вращающихся вокруг своих осей симметрии, обладает рядом специфических особенностей, широко используемых в технике. Массивные симметричные тела, быстро вращающиеся вокруг оси симметрии, которая, в свою очередь, свободно ориентируется в пространстве, получили название *гироскопов*. Способность оси гироскопа сохранять при определенных условиях свое направление в пространстве лежит в основе их использования в качестве навигационных приборов. Само же это свойство обусловлено сохранением момента импульса вкпе с тем обстоятельством, что угловая координата, характеризующая вращение, меняется быстрее, нежели любая другая:

$$\omega \gg \Omega, \quad (7.28)$$

где Ω — угловая скорость поворота оси волчка в пространстве. Именно в этом смысле вращение гироскопа должно быть быстрым.

Рассмотрим гироскоп, имеющий вид массивного колеса, подвешенного несимметричным образом, как показано на рис. 7.14.

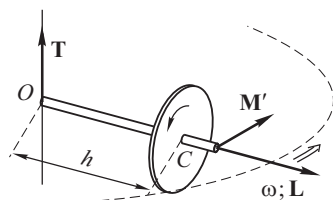


Рис. 7.14

(Для демонстрации удобно использовать велосипедное колесо, раскрученное электромоторчиком).

Угловая скорость ω и момент импульса \mathbf{L} направлены по оси. На гироскоп действует сила тяжести и натяжение нити \mathbf{T} . Вычисляя момент сил относительно центра инерции гироскопа, мы должны учесть вклад в него лишь от силы натяжения, так

как момент силы тяжести относительно точки C равен нулю. Момент силы натяжения равен $M' = Th$ и направлен так, как показано на рис. 7.14. Из уравнения моментов (7.5) следует, что приращение вектора \mathbf{L} будет происходить в направлении момента \mathbf{M}' , то есть ось вращения должна поворачиваться в горизонтальной плоскости — на рис. 7.14 это показано двойной стрелкой.

Ответ лишь на первый взгляд кажется парадоксальным. В наших рассуждениях мы молчаливо считали выполненным условие (7.28), в котором под Ω следует понимать именно скорость поворота вектора \mathbf{L} . Таким образом, для демонстрации нетривиального поведения гироскопа он должен быть как следует раскручен. В этом случае он не может опрокинуться, хотя и подвешен несимметричным образом. В самом деле, при опрокидывании вектор \mathbf{L} должен был бы повернуться в вертикальной плоскости. Но момент внешней силы для такого поворота должен быть направлен по вертикали, а значит, сама сила должна быть горизонтальна и направлена по нормали к оси. Но при подвешивании гироскопа на нити такая сила отсутствует. Вот если бы посредством жесткой муфты гироскоп был лишен степени свободы, позволяющей его оси поворачиваться в горизонтальной плоскости, то реакция опоры была бы ориентирована именно так, и колесо бы опрокинулось.

Мы не случайно упомянули именно велосипедное колесо. Большая устойчивость быстро движущегося велосипеда в сравнении с неподвижно стоящим, не в последнюю очередь обусловлена именно гироскопическим эффектом вращающихся колес.

Рассмотрим задачу о так называемой регулярной прецессии волчка (рис. 7.15). Как известно, быстро вращающийся волчок, будучи слегка выведен из вертикального положения, не падает, но приобретает дополнительное вращение в наклонном положении, угловая скорость которого Ω удовлетворяет неравенству (7.28). Когда же указанное неравенство, по мере замедления

вращения волчка, еще даже не нарушается, но перестает быть сильным ($>$ вместо \gg), на вращение накладываются сложные нерегулярные движения (их называют нутациями). Эту стадию динамики волчка мы рассматривать не будем ввиду математической сложности, однако объяснить регулярную прецессию для нас труда не составит.

Обратимся к рис. 7.15. Пусть C — центр инерции волчка, O — точка опоры, причем $|CO| = a$. Поскольку в точке O , помимо реакции опоры \mathbf{N} , приложена еще и сила трения $\mathbf{F}_{\text{тр}}$, нам, вообще говоря, не известная, момент сил проще вычислять относительно точки O . Он, очевидно, равен $mga \sin \theta$ и направлен от нас перпендикулярно плоскости рисунка. При условии (7.28) мы вправе связать момент импульса гироскопа \mathbf{L} исключительно с угловой скоростью $\boldsymbol{\omega}$ и считать его направленным строго по оси.

Из уравнения моментов (7.15) следует, что за малое время dt момент импульса \mathbf{L} получает приращение

$$d\mathbf{L} = \mathbf{M} dt,$$

причем, как и сам вектор \mathbf{M} , это малое приращение перпендикулярно \mathbf{L} .

Таким образом, эволюция вектора \mathbf{L} свелась к повороту, который в горизонтальной плоскости можно характеризовать углом

$$d\varphi = \frac{dL}{L \sin \theta} = \frac{mga \sin \theta}{L \sin \theta} dt = \frac{mga}{L} dt.$$

Получается, что ось волчка и в самом деле должна прецессировать, причем угловая скорость прецессии не зависит от угла θ :

$$\Omega \equiv \frac{d\varphi}{dt} = \frac{mga}{I_0 \omega}, \quad (7.29)$$

где I_0 — момент инерции волчка относительно оси симметрии. Условие (7.28) можно в данном случае представить в эквивалентном виде $mga \ll I_0 \omega^2$. Уравнение моментов можно переписать в векторной форме

$$\mathbf{M} = \frac{d\mathbf{L}}{dt} = [\boldsymbol{\Omega} \mathbf{L}],$$

а результат (7.29), соответственно, примет вид

$$\boldsymbol{\Omega} = -\frac{mga}{I_0 \omega}. \quad (7.30)$$

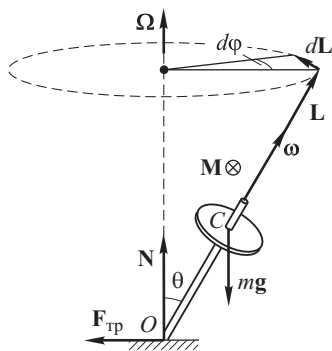


Рис. 7.15

7.6. Плоское движение твердого тела

По самому определению, плоское движение отвечает изменению во времени лишь трех координат твердого тела, например, двух декартовых x , y и одной ориентационной φ . Последняя может и выпасть, если движение — чисто поступательное. К плоскому сведется и существенно трехмерное движение, если оно является поступательным по одной из декартовых координат.

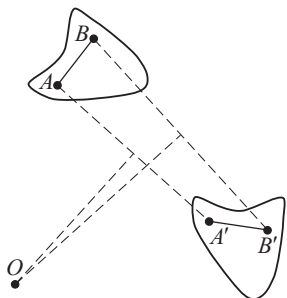


Рис. 7.16

При плоском движении, как легко усмотреть из рис. 7.16, любое перемещение можно представить как поворот вокруг некоторой точки O (точнее — оси, проходящей через O и ортогональной плоскости перемещения). Особый случай — поступательное движение, при котором точка O как бы перемещается в бесконечность.

Как следствие, малое перемещение представляется как поворот вокруг некоторой оси O на малый угол $d\varphi$.

Если этот поворот происходит за время dt , то величина $\omega(t) = d\varphi/dt$ называется *мгновенным значением угловой скорости*, а сама ось, проходящая через точку O — *мгновенной осью вращения*. Если, например, мы можем установить, что в данный момент времени какая-то точка тела имеет нулевую скорость, то именно через нее проходит мгновенная ось вращения; если же мы наблюдаем две таких точки, то $\mathbf{v} \equiv 0$ для всего нашего тела.

Из рис. 7.16 можно понять также и то, что плоское перемещение можно представить не только в виде поворота, но и в виде комбинации любого поступательного перемещения и поворота. Просто, перенеся тело, которому принадлежит отрезок AB , в любую заданную точку, мы проведем затем построение, эквивалентное тому, которое показано на рис. 7.16. Если мы вновь обратимся к малым перемещениям, то сдвинув любую точку тела на некоторый вектор $d\mathbf{r} = \{dx, dy\}$, мы затем осуществим поворот на угол $d\varphi$. Соответственно, деля на dt , получаем представление движения скоростью $\mathbf{v}(t)$ и угловой скоростью $\omega(t)$. Физический смысл $\mathbf{v}(t)$ — скорость системы отсчета, в которой рассматривается вращение.

Обратимся еще раз к рис. 7.16. Какова бы ни была упомянутая комбинация поступательного и вращательного перемещений, угол между отрезками AB и $A'B'$ остается инвариантом. Иными словами, как бы мы ни сместили мгновенную ось, тело придется повернуть на один и тот же угол. Еще раз возвращаясь к малым

перемещениям, приходим к важному выводу: *угловая скорость $\omega(t)$ при плоском движении не зависит от системы отсчета.*

Уравнения движения, в зависимости от постановки задачи, обычно пишутся либо для центра масс тела —

$$m \frac{d\mathbf{v}_C}{dt} = \sum \mathbf{F}_{\text{вн}}, \quad I_C \frac{d\omega}{dt} = \sum M_{\text{вн}}, \quad (7.31)$$

либо для мгновенной оси вращения —

$$I_O \frac{d\omega}{dt} = \sum M_{O \text{ вн}}, \quad (7.32)$$

с соответствующим пересчетом момента инерции и момента сил относительно этой оси.

Рассмотрим в качестве иллюстрации задачу о скатывании однородного сплошного цилиндра массы m и радиуса R с наклонной плоскости, образующей с горизонтом угол γ (рис. 7.17). Предполагаем, что цилиндр скатывается без проскальзывания, тогда скорость в точке O равна нулю. Мгновенная ось вращения совпадает с линией зацепления, проходящей через точку O , в которой к телу приложена сила трения. Второй закон Ньютона, с учетом направления движения, целесообразно записать в проекции на наклонную плоскость:

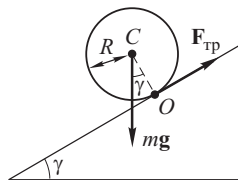


Рис. 7.17

$$m \frac{dv_C}{dt} = mr \frac{d\omega}{dt} = mg \sin \gamma - F_{\text{тр}}, \quad (7.33)$$

а уравнение моментов запишем относительно мгновенной оси:

$$I_O \frac{d\omega}{dt} = \left(\frac{1}{2} mr^2 + mr^2 \right) \frac{d\omega}{dt} = mgr \sin \gamma, \quad (7.34)$$

(мы использовали теорему Штейнера для вычисления момента инерции относительно мгновенной оси вращения). Ни сила трения, ни реакция опоры не дают вклада в уравнение (7.34), поскольку обе они проходят через точку O , а значит, их моменты относительно мгновенной оси равны нулю. Из (7.34) находим угловое ускорение

$$\frac{d\omega}{dt} = \frac{2}{3} \frac{g}{r} \sin \gamma,$$

следовательно, линейное ускорение равно

$$a_{\parallel} = r \frac{d\omega}{dt} = \frac{2}{3} g \sin \gamma.$$

Далее, подставляя это выражение в (7.33), получаем

$$F_{\text{тр}} = \frac{1}{3} mg \sin \gamma. \quad (7.35)$$

Таким образом, сила трения в данной задаче принимает вполне определенное значение. Но как мы знаем (см. гл. 5), оно не

может превосходить предела αN_{\perp} , где α — коэффициент трения, а N_{\perp} — сила реакции опоры, в нашем случае:

$$N_{\perp} = mg \cos \gamma.$$

Таким образом, (7.35) можно выполнить лишь при условии

$$\operatorname{tg} \gamma < 3\alpha. \quad (7.36)$$

По физическому смыслу (7.36) представляет собой *условие непроскальзывания* скатывающегося цилиндра.

Если цилиндр скатывается с высоты h , то его конечная скорость равна

$$v_{\text{кон}} = \sqrt{2a_{\parallel}l} = \left(2 \cdot \frac{2}{3} g \sin \gamma \cdot \frac{h}{\sin \gamma} \right)^{1/2} = \sqrt{\frac{4}{3} gh}.$$

Убедимся, что такое значение скорости удовлетворяет закону сохранения механической энергии. Согласно теореме Кёнига,

$$T_f = \frac{1}{2} mv_c^2 + \frac{1}{2} I_c \omega^2 = \frac{1}{2} mv^2 + \frac{1}{2} \frac{mr^2}{2} \left(\frac{v}{r} \right)^2 = mgh.$$

Элементарные вычисления дают то же самое значение $v_{\text{кон}}$.

Возникает вопрос: как может сохраняться чисто механическая энергия при наличии заведомо ненулевой силы трения? Но дело в том, что сила трения приложена в точке O , где скорость равна нулю, а потому и работы не совершает.

Таким образом, плоский характер движения столь существенно упрощает задачу о динамике тела конечных размеров, что она в принципе оказывается не намного сложнее задачи о динамике материальной точки.

Вопросы и задачи

1. Что такое момент импульса и момент силы?
2. Сохраняется ли при движении спутника по орбите его энергия, импульс, момент импульса?
3. Сохраняется ли суммарная энергия, суммарные импульс и момент импульса всех тел Солнечной системы?
4. Сколько степеней свободы у абсолютно твердого тела?
5. Что такое момент инерции твердого тела?
6. Что такое момент силы относительно оси?
7. Что такое главные оси инерции твердого тела?
8. Чему равен момент инерции тонкого однородного кольца массы m и радиуса R относительно оси, касательной к нему и перпендикулярной его плоскости?

Ответ: $I = 2mR^2$.

9. Однородный цилиндр массы m_1 и радиуса R вращается без трения вокруг горизонтальной оси под действием груза массы m_2 , прикрепленного к легкой нерастяжимой нити, намотанной на цилиндр (рис. 7.18). Найти угол поворота цилиндра φ в зависимости от времени, если при $t = 0$ $\varphi = 0$, $d\varphi/dt = 0$.

Решение. Уравнения движения для цилиндра и для груза имеют вид

$$I \frac{d^2 \varphi}{dt^2} = RT, \quad m_2 \frac{d^2 z}{dt^2} = m_2 g - T,$$

где T — натяжение нити, $I = m_1 R^2/2$ — момент инерции цилиндра. Перемещение груза и угол поворота цилиндра связаны соотношением $z = R\varphi$. Учитывая это соотношение и исключая из уравнений натяжение нити T , получаем:

$$(I + m_2 R^2) \frac{d^2 \varphi}{dt^2} = m_2 g R.$$

Решение этого уравнения, удовлетворяющее начальным условиям, имеет вид

$$\varphi(t) = \frac{gt^2}{2R(1 + m_1/(2m_2))}.$$

Проверьте решение по размерности и по очевидным результатам в предельных случаях ($g = 0$, $m_2 = 0$ и т. п.).

10. Вертикальный столб высотой l подпиливается у основания и падает на землю, поворачиваясь вокруг нижнего основания. Определить линейную скорость его верхнего конца в момент удара о землю.

Ответ: $v = \sqrt{3gl}$.

Указание: Воспользоваться законом сохранения энергии.

11. Однородный тонкий тяжелый стержень длины l висит на горизонтальной оси, проходящей через один из его концов. Какую начальную угловую скорость ω надо сообщить стержню, чтобы он повернулся на 90° ?

Ответ: $\omega = \sqrt{3g/l}$.

12. Небесное тело 1991ДА движется по вытянутой орбите, так что минимальное расстояние от него до Солнца равно радиусу орбиты Марса, а максимальное — радиусу орбиты Урана. Определите период обращения 1991ДА вокруг Солнца T , если известны периоды обращения Марса $T_1 = 1,88$ года и Урана $T_2 = 84$ года.

Ответ: $T = 2^{-3/2} (T_1^{2/3} + T_2^{2/3})^{3/2}$.

13. В 1978 г. у планеты Плутон обнаружен спутник — Харон. Плутон и Харон обращаются вокруг общего центра масс по круговым орбитам, причем расстояние между их центрами $R = 19640$ км, а период обращения $T = 6,4$ суток. Определите, какую часть массы Земли составляет суммарная масса системы Плутон–Харон. Считать известными радиус Земли $R = 6400$ км и ускорение свободного падения на поверхности Земли.

Ответ: $(M_{\text{Пл}} + M_{\text{Х}})/M_3 = 4\pi^2 R^3/T^2 g_{R3} = 0,024$.

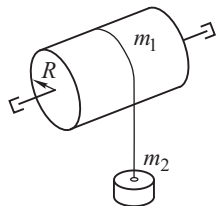


Рис. 7.18

Глава 8

ЭЛЕМЕНТЫ МЕХАНИКИ СПЛОШНЫХ СРЕД

8.1. Упругие деформации. Закон Гука

До сих пор мы изучали условия равновесия или характер динамики либо отдельной материальной точки, либо их совокупности. Абсолютно твердое тело рассматривалось как важный частный случай такой совокупности. Теперь мы обращаемся к сплошной среде. Такое понятие используется при решении задач, которые затруднительно свести к механике материальной точки. Различие относится, конечно же, не к предмету, но к методу рассмотрения.

Когда мы говорим об абсолютно твердом теле, мы просто пренебрегаем его деформациями, полагая, что их учет приведет лишь к некоторым малым поправкам в ответе динамической задачи. То же можно сказать и о любой материальной точке, хотя в этом приближении влияние деформаций и не столь очевидно. Но избежать деформаций невозможно в принципе, и это оказывается следствием самых фундаментальных физических законов. В самом деле, представим себе абсолютно твердый стержень. Приложим к одному его концу некоторую силу, которая приведет его в движение в продольном направлении. Поскольку стержень не может деформироваться, оба его конца должны начать двигаться одновременно, но это значит, что мы транслировали воздействие (сигнал) от одного конца к другому с бесконечной скоростью. Это решительно противоречит теории относительности, поскольку по современным понятиям эта скорость должна быть конечной и не может превышать скорости света (см. гл. 10).

Описанный мысленный эксперимент лишь показывает, что никакое тело нельзя считать абсолютно твердым в абсолютном смысле, т. е. безотносительно к постановке задачи. Корректный учет деформаций и механических напряжений в целом составляет предмет обширной и довольно сложной области механики, во многих отношениях смыкающейся с физикой твердого тела. Механика деформируемых тел имеет массу технических приложений, весьма важных в практическом отношении. В нашем же курсе мы вынуждены ограничиться наиболее простыми и фундаментальными положениями этой науки.

Во-первых, мы ограничимся рассмотрением *упругих*, или, что то же, *обратимых*, деформаций. Тем самым мы оставляем за

рамками нашего курса эффекты пластичности и механического разрушения. Упругие деформации, по определению, исчезают после снятия вызывающих их напряжений, то есть не сопровождаются ни диссипацией, ни перестройкой структуры вещества. Строго говоря, абсолютно упругих деформаций не бывает, но в массе практически интересных случаев мы с очень хорошей точностью можем использовать такое приближение.

Во-вторых, мы определим *твердое тело* так, как принято определять его в физике, а именно — *как кристаллическое состояние*. В бытовом понимании (и даже в механике, пока и поскольку мы не касаемся проблемы деформаций) твердыми телами можно считать также переохлажденные жидкости (стекла). Но их свойства с точки зрения материаловедения весьма нетривиальны, например, под нагрузкой они могут течь, только очень медленно. Еще большее своеобразие обнаруживает высокополимерная фаза. Но мы ограничимся рассмотрением упругих свойств кристаллов.

В-третьих, мы не будем касаться монокристаллов, а займемся исключительно *поликристаллическими* телами — к таковым относятся, например, все обычные изделия из различных металлов. Они не демонстрируют упорядоченности структуры, характерной для монокристаллов, но в кажущемся хаосе расположения элементарных монокристалликов сохраняются некоторые присущие им свойства.

По отношению к деформациям монокристалл и поликристалл могут быть весьма различны. Например, предел упругости монокристалла Al составляет примерно 40 Н/см^2 , а у технического алюминия тот же параметр равен 10^4 Н/см^2 . Если же говорить об упругих свойствах, то для монокристалла их можно характеризовать как минимум тремя константами — в случае кубической решетки, — а для более сложных кристаллов это число достигает 21. Упругие свойства поликристалла описываются всего двумя независимыми константами.

Простейшим и важнейшим законом, которому подчиняются упругие деформации, является закон Гука. В изначальном виде он формулировался предельно просто: при растяжении или сжатии твердого стержня малое изменение его длины оказывается пропорционально приложенной силе:

$$\Delta l = \frac{F}{k}. \quad (8.1)$$

При этом величина силы F должна быть много меньше некоторого *предела упругости* F_0 , за которым начинаются необратимые деформации. Существенно, что коэффициент k — один и тот же при растяжении и сжатии стержня. Однако закон Гука в виде (8.1) выглядит слишком неуниверсальным; получается,

что коэффициент k должен определяться особо для каждого отдельного тела, и даже для воздействия на одно и то же тело в разных направлениях. Привести его к более конструктивной форме нам поможет рис. 8.1. Возьмем два одинаковых стержня длины l с поперечным сечением S и растянем их силой F , сцепив последовательно, как показано на рис. 8.1 а. Каждый стержень растягивается силой F , каждый удлиняется на Δl , а вместе они удлиняются на $2\Delta l$. Отсюда ясно, что $\Delta l \propto l$, и (8.1) можно переписать следующим образом:

$$\frac{\Delta l}{l} = \frac{F}{k_1}.$$

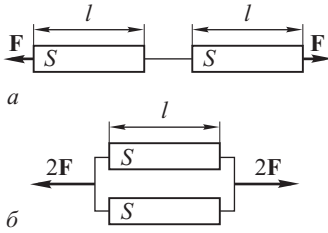


Рис. 8.1

Если теперь сцепить стержни параллельно (рис. 8.1 б) и приложить силу $2F$, каждый стержень снова растягивается силой F и удлиняется на Δl . Представим систему, изображенную на рис. 8.1 б, как стержень длины l с поперечным сечением $2S$. Отсюда следует, что растяжение задается не силой как таковой, но отношением силы к площади сечения:

$$\sigma = \frac{F}{S}. \quad (8.2)$$

Величина σ в механике деформируемых тел называется *напряжением* и измеряется в Н/м² (в технике часто пользуются внесистемной единицей кгс/см²). В телах сложной формы и при сложных нагрузках естественно использовать определение $\sigma = dF/dS$. При этом величина σ в общем случае характеризуется двумя векторами: ориентацией площадки dS и вектором силы $d\mathbf{F}$.

С учетом изложенных выше соображений закон Гука приобретает окончательный вид

$$\varepsilon \equiv \frac{\Delta l}{l} = \frac{\sigma}{E}, \quad (8.3)$$

где ε называется *деформацией*, а величина E теперь уже просто константа вещества, которую называют *модулем Юнга* (по имени английского ученого Т. Юнга (1773–1829)).

Линейность закона (8.3) отражает всего лишь тот факт, что в разложении функции общего вида

$$\sigma = E\varepsilon + E'\varepsilon^2 + E''\varepsilon^3 + \dots$$

первый член при упругих деформациях доминирует, и таким образом, существенной оказывается только одна константа вещества. Но как следствие продольной деформации, может из-

мениться и поперечный размер тела. Опыт показывает, что и этот закон для поликристаллических веществ в пределах упругих деформаций оказывается линейным, причем знак поперечной деформации противоположен знаку продольной:

$$\varepsilon_{\perp} \equiv \frac{\Delta l_{\perp}}{l_{\perp}} = -\mu \frac{\Delta l}{l} + \dots \approx -\mu \varepsilon. \quad (8.4)$$

Константа вещества μ называется *коэффициентом Пуассона* (по имени французского ученого С. Пуассона (1781–1840)). Важно отметить, что поперечная деформация вызвана продольным напряжением и сама дополнительных напряжений не вызывает.

Работа, совершенная при растяжении или сжатии образца, зависит исключительно от продольной деформации, в силу соотношения

$$\delta A = \mathbf{F} d\mathbf{l} = F dl_{\parallel}.$$

Совершенная работа запасается в виде энергии упругой деформации, которая освобождается при снятии нагрузки,

$$U = \Delta A = \int_0^{\Delta l} \sigma S_{\perp} dl = \frac{S_{\perp} l}{E} \int_0^{\sigma} \sigma d\sigma = \frac{1}{2} V \frac{\sigma^2}{E}.$$

Эта энергия распределена по всему объему деформированного тела, что дает основание ввести *плотность энергии упругой деформации*

$$w = \frac{1}{2} \frac{\sigma^2}{E} = \frac{1}{2} \sigma \varepsilon = \frac{1}{2} E \varepsilon^2. \quad (8.5)$$

Оказывается, две константы — E и μ — в линейном случае, характерном для упругих деформаций поликристаллов, исчерпывающим образом описывают упругие свойства вещества, то есть линейный отклик на любое воздействие, может быть выражен через эти две постоянные. Нам предстоит убедиться в этом на ряде хрестоматийных примеров. Первый пример — всестороннее сжатие твердого тела. Введем понятие *давления* — нагрузки, действующей по нормали к поверхности тела —

$$\mathcal{P} = \frac{dF_{\perp}}{dS}.$$

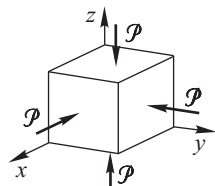


Рис. 8.2

Как известно, эта величина измеряется в паскалях: $1 \text{ Па} = 1 \text{ Н/м}^2$. Для облегчения последующих вычислений рассмотрим тело, имеющее форму параллелепипеда (рис. 8.2). Для одной продольной и двух поперечных деформаций, обусловленных

продольным напряжением вдоль каждой из его осей, получаем

$$\begin{aligned} \text{по оси } x \quad \varepsilon_x &= -\frac{\mathcal{P}}{E}; & \varepsilon_{y(x)}, \varepsilon_{z(x)} &= \mu \frac{\mathcal{P}}{E}, \\ \text{по оси } y \quad \varepsilon_y &= -\frac{\mathcal{P}}{E}; & \varepsilon_{z(y)}, \varepsilon_{x(y)} &= \mu \frac{\mathcal{P}}{E}, \\ \text{по оси } z \quad \varepsilon_z &= -\frac{\mathcal{P}}{E}; & \varepsilon_{x(z)}, \varepsilon_{y(z)} &= \mu \frac{\mathcal{P}}{E}. \end{aligned}$$

Таким образом,

$$\varepsilon_x = \varepsilon_y = \varepsilon_z = -\frac{\mathcal{P}}{E} (1 - 2\mu). \quad (8.6)$$

Соответственно,

$$\begin{aligned} \frac{\delta V}{V} &= \delta \ln V = \delta(\ln l_x + \ln l_y + \ln l_z) = \\ &= \varepsilon_x + \varepsilon_y + \varepsilon_z = -\frac{3\mathcal{P}}{E} (1 - 2\mu). \end{aligned} \quad (8.7)$$

При всестороннем сжатии обязательно должно быть $\delta V < 0$, в противном случае тело было бы неустойчиво по отношению к неограниченной самодеформации. Действительно, пусть $\alpha = dV/dp > 0$ и в результате случайной температурной флуктуации объем тела увеличился, это эквивалентно тому, что мы подвергли тело сжатию. Но увеличение объема приводит к еще большему увеличению давления, т. к. $\Delta p = (1/2)\Delta V > 0$ и т. д. Отсюда следует универсальное ограничение

$$0 < \mu < \frac{1}{2}. \quad (8.8)$$

Опыт показывает, например, что у поликристаллов $1/4 < \mu < 1/2$, а для пористых тел $\mu \rightarrow 0$.

8.2. Сдвиг и кручение

В отличие от растяжения и сжатия, деформация сдвига вызывается касательными напряжениями (рис. 8.3а). Пусть $\tau = dF_{\parallel}/dS$ — касательное напряжение (F_{\parallel} — сила, параллельная поверхности пробного кубика). Следствием деформации будет угол перекоса γ . При условии его малости связь между воздействием и откликом должна быть линейной —

$$\gamma = G\tau, \quad (8.9)$$

где коэффициент G именуется *модулем сдвига*, а все соотношение (8.9), как и (8.3), — *законом Гука*.

Для равновесия деформированного кубика к нему одновременно с касательным напряжением τ должно быть приложено еще одно, равное и противоположно направленное (показано

на рис. 8.3а штрихованной стрелкой), а также дополнительная пара сил, уравнивающая момент этих сил, показанная на рис. 8.3а точечными стрелками. Два верхних напряжения

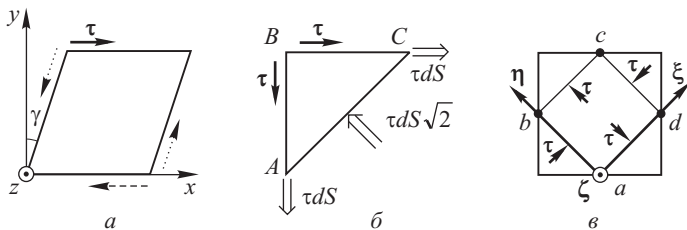


Рис. 8.3

обеспечиваются непосредственно воздействием, нижние — реакцией опоры. Рассмотрим равновесие уголка ABC (рис. 8.3б). Из построения равновесия сил, действующих на уголок (τdS вдоль граней BA и BC , $\tau dS\sqrt{2}$ — по биссектрисе угла ABC), следует, что на поверхности BA возникает такое же напряжение τ , но уже не касательное, а нормальное. Как следствие, внутри деформируемого кубика можно выделить кубик $abcd$ (рис. 8.3в), который подвергается растяжению по оси η и сжатию по оси ξ . Далее каждому способу рассмотрения упругой деформации будет удобно поставить в соответствие свою плотность энергии. Во-первых, в соответствии с рис. 8.3а,

$$dU = \frac{1}{2} \mathbf{F} \cdot d\mathbf{l} = \frac{1}{2} \tau dx dy \cdot dz \gamma \Rightarrow w = \frac{dU}{dV} = \frac{1}{2} \tau \cdot \gamma = \frac{\tau^2}{2G}.$$

С другой стороны (см. рис. 8.3в),

$$-\varepsilon_\xi = \varepsilon_\eta = \frac{\tau}{E} + \mu \frac{\tau}{E} \Rightarrow dU = 2 \cdot \tau d\eta d\xi \frac{|\varepsilon_\xi| d\xi}{2} \Rightarrow w = \tau |\varepsilon_\xi|.$$

Таким образом,

$$\frac{\tau^2}{2G} = \tau^2 \frac{1+\mu}{E} \Rightarrow G = \frac{E}{1+\mu}. \quad (8.10)$$

Итак, модуль сдвига удастся выразить через две ранее введенные упругие константы.

Рассмотрим деформацию кручения (рис. 8.4). Пусть к цилиндру радиуса R с закрепленным основанием приложен крутящий момент M . Если при этом свободный конец прокручивается на угол φ , то линейный закон, аналогичный (8.3), имеет вид

$$M = f \varphi, \quad (8.11)$$

где f именуется *модулем кручения*. Рассмотрев деформацию трубочки радиуса r и толщины $\delta r \ll r$, нетрудно убедиться, что для нее кручение эквивалентно деформации сдвига, причем угол

сдвига тривиальным образом связан с φ ,

$$\gamma h = \varphi r,$$

где h — длина образца. Вклад в полный момент от нашей трубочки равен

$$\delta M = \tau \cdot 2\pi r \delta r \cdot r, \quad \text{где} \quad \tau = G\gamma = G\varphi \frac{r}{h}.$$

Соответственно, для сплошного цилиндра

$$M = \int \delta M = \int_0^R \frac{2\pi G}{h} \varphi r^3 dr = \varphi \frac{\pi G R^4}{2h} \equiv f \cdot \varphi. \quad (8.12)$$

Таким образом, модуль кручения однородного стержня равен $\pi G R^4/(2h)$, а для тонкой трубочки — $2\pi r^3 \delta r \cdot G/h$. Мы вновь убедились в том, что линейный отклик на внешнее воздействие может быть выражен исключительно через модуль Юнга и коэффициент Пуассона. Интересным следствием закона (8.11) оказывается соотношение для так называемых крутильных колебаний. Пусть массивное симметричное тело радиуса R с моментом инерции I подвешено на проволоке длины l и радиуса r с модулем кручения f (рис. 8.5). Уравнение моментов (7.13) для этого случая приводится к виду

$$I\ddot{\varphi} = -M(\varphi) = -f\varphi,$$

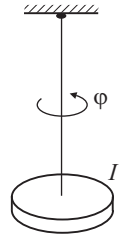


Рис. 8.5

а это уравнение типа (4.18), решение которого хорошо известно:

$$\varphi = \varphi_0 \cos(\omega t + \psi), \quad \text{где} \quad \omega^2 = \frac{f}{I} = \frac{\pi G r^4}{m l R^2}.$$

8.3. Течение идеальной жидкости. Уравнение непрерывности

В отличие от твердых тел жидкости и газы с точки зрения механики являются веществами текучими. Но если в твердых телах пластичность неотделима от диссипации, то в случае текучей среды диссипацией при рассмотрении многих важных процессов и явлений можно пренебречь. Такой подход называется приближением *идеальной жидкости* (заметим, что во многих случаях приближение идеальной жидкости описывает и динамику течения газа). Далее, очень многие практически важные задачи допускают представление о *несжимаемой жидкости*. Это понятие вошло в обиход и нередко используется в технических приложениях, но надо отдавать себе отчет в том, что сжимаемость жидкости, если понимать ее в контексте гл. 8.1,

заметно превышает таковую у твердого тела. Но как твердое тело — вполне реальное и подверженное деформациям — можно во многих динамических задачах с хорошей точностью считать абсолютно твердым, так и жидкость в задаче о ее течении нередко с достаточной точностью представима как несжимаемая. Обычно такое представление оправдано, если перепады скоростей в жидкости гораздо меньше скорости звука, и в этом смысле даже воздух при атмосферном давлении может рассматриваться как несжимаемая жидкость, если скорость его течения меньше 300 м/с.

Хотя, казалось бы, течение жидкости и движение тела конечных размеров весьма различны, общий «камертон» в системе основных понятий все же существует. Если мы выделим объем жидкости столь малый, что можно пренебречь его размерами и формой, и этот объем не перемещивается на характерном масштабе задачи (или за характерное время задачи) с другим веществом, то такой *жидкий элемент* можно рассматривать как материальную точку. Это позволяет определить скорость и ускорение элемента и вывести на этой основе уравнения течения (или, в частности, равновесия) жидкости. Траектория такого жидкого элемента называется *линией тока* (рис. 8.6*а*). Совокупность линий тока, близких в пределах, заданных характерным временем и масштабом задачи, называется *трубкой тока*. Предполагается, что трубка тока охватывается некоторым односвязным контуром ($C \rightarrow C'$ на рис. 8.6*б*). Корректность понятий жидкого элемента и линии тока обусловлена лишь малостью элементарного объема. Понятие трубки тока подразумевает, что на интересующем нас пространственном масштабе линии тока достаточно мало расходятся и, в частности, не образуют вихрей.

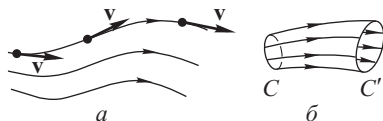


Рис. 8.6

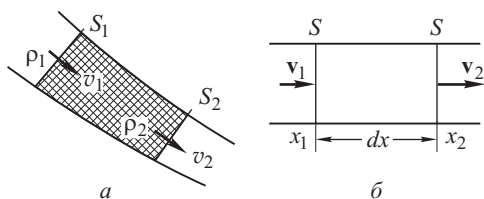


Рис. 8.7

При таком рассмотрении локальные характеристики — плотность вещества ρ и скорость \mathbf{v} — принято относить не к движущемуся жидкому элементу, а к потоку в целом и рассматривать как функции точки $\rho(\mathbf{r})$, $\mathbf{v}(\mathbf{r})$.

Если протекание вещества не сопровождается никакими реакциями, то масса вещества в потоке сохраняется. Рассмотрим некоторую трубку тока (рис. 8.7*а*). Пусть течение стационарно.

Выделим объем, ограниченный стенками трубки тока и двумя ее сечениями, нормальными к скорости жидкости. Сохранение массы в выделенном объеме выражается равенством массы жидкости, втекающей в сечение 1 и вытекающей из сечения 2 за время dt : $\rho_1 S_1 \mathbf{v}_1 dt = \rho_2 S_2 \mathbf{v}_2 dt$. Здесь $\rho_{1,2}$, $S_{1,2}$ и $\mathbf{v}_{1,2}$ — плотность жидкости, площадь нормального сечения трубки тока и скорость жидкости в точках 1 и 2. Тем самым закон сохранения массы принимает вид *уравнения неразрывности струи*

$$\rho S_{\perp} v = \text{const} \quad (8.13)$$

или для несжимаемой жидкости ($\rho = \text{const}$)

$$S_{\perp} v = \text{const}. \quad (8.14)$$

Если поток нестационарен, уравнения (8.13, 8.14) должны быть модифицированы. Представим себе одномерную трубку тока (рис. 8.7 б). Закон сохранения массы вещества означает, что масса жидкости, втекающая в некоторый объем, $\rho v(x) dt S$, равна массе жидкости, накапливающейся в этом объеме, $\dot{\rho} dt S dx$, плюс массе вытекающей жидкости $\rho v(x + dx) dt S$. Имея в виду, что все характерные параметры зависят, вообще говоря, от двух переменных x и t , введем понятие *частной производной*

$$\frac{\partial f(x, t)}{\partial t} \equiv \left. \frac{df}{dt} \right|_{x=\text{const}}, \quad \frac{\partial f(x, t)}{\partial x} \equiv \left. \frac{df}{dx} \right|_{t=\text{const}}.$$

Закон сохранения массы принимает вид

$$\rho v(x) S = \dot{\rho} S dx + \rho v(x + dx) dt S \implies \frac{\partial \rho}{\partial t} = - \frac{\partial(\rho v(x))}{\partial x}.$$

Традиционная форма записи этого закона называется *уравнением непрерывности*:

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho v)}{\partial x} = 0. \quad (8.15)$$

В случае неоднородного течения обобщение уравнения (8.15) достигается посредством введения оператора *дивергенции*. По определению, для любой векторной функции $\mathbf{j}(\mathbf{r}) = \mathbf{j}(x, y, z)$ дивергенция \mathbf{j} есть

$$\text{div } \mathbf{j} \equiv \frac{\partial j_x}{\partial x} + \frac{\partial j_y}{\partial y} + \frac{\partial j_z}{\partial z}.$$

В общем случае трехмерного течения уравнение (8.15) принимает вид

$$\frac{\partial \rho}{\partial t} + \text{div}(\rho \mathbf{v}) = 0. \quad (8.16)$$

Если же справедливо приближение несжимаемой жидкости, то изменением плотности как в пространстве, так и во времени можно пренебречь:

$$\text{div } \mathbf{v} = 0. \quad (8.17)$$

Уравнения (8.16), (8.17) выглядят, конечно, сложнее, чем, соответственно, (8.13), (8.14), но обладают и заметным преимуще-

ством — они локальны, т. е. не привязаны ни к какой трубке тока, и их решение, в принципе, просто некоторая функция точки в пространстве \mathbf{r} и времени t .

8.4. Архимедова сила. Уравнение Бернулли

В отличие от твердого тела, жидкость — среда бесструктурная, ее микроскопическое устройство не содержит такой «несущей конструкции», как кристаллическая решетка. Соответственно вопрос об ориентации поверхности, к которой приложено напряжение, здесь не возникает — все ориентации равноправны. Кроме того, в жидкости несравненно меньшую роль играют касательные напряжения. Если в твердом теле они ответственны за сдвиговые деформации и вполне проявляют себя в статике, то в жидкости касательные напряжения возникают исключительно в динамической задаче и лишь при учете диссипативного эффекта — вязкости.

Опыт показывает, что в гидростатике — науке о равновесии жидких тел — а равно и в консервативной (бездиссипативной) гидродинамике существует всего один вид напряжения — давление, обладающее свойством изотропности. Это и составляет содержание *закона Паскаля* (по имени французского ученого Блеза Паскаля (1623–1662)): давление жидкости и газа передается одинаково во всех направлениях. Иными словами, давление — скалярная функция:

$$\mathcal{P} = \mathcal{P}(\mathbf{r}). \quad (8.18)$$

Одно из важнейших следствий изотропии давления было известно задолго до Паскаля — это закон Архимеда (по имени древнегреческого ученого Архимеда (ок. 287–212 до н. э.)): тело, погруженное в жидкость, выталкивается с силой, равной весу вытесненной жидкости. Рассмотрим тело A , плавающее в сосуде с жидкостью (рис. 8.8а), и отдельно — сосуд, заполненный той же жидкостью до того же уровня, но без погруженного в него тела (рис. 8.8б). Если жидкость, заполняющая полость A' на рис. 8.8б, пребывает в равновесии, то значит силы, действующие на полость A' , как раз и удерживают ее собственный вес. Ввиду скалярности давления (8.18) и отсутствия касательных напряжений, неважно, чем заполнена полость A' , так что и на тело A будут действовать так же распределенные силы, имеющие ту же равнодействующую.

Хотя приведенные выше рассуждения и дают нам аргументы в пользу закона Архимеда, они не могут считаться доказа-

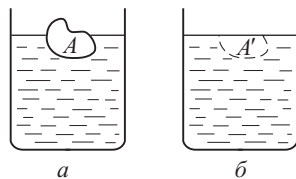


Рис. 8.8

тельством. В дополнение к ним хотелось бы дать современную трактовку архимедовой силы, пригодную для решения динамических задач. Мы сделаем это, опираясь на закон Паскаля. Пусть давление в жидкости или газе зависит от координаты:

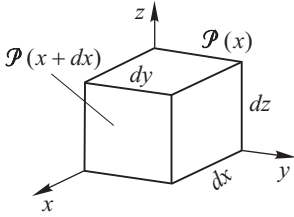


Рис. 8.9

$\mathcal{P} = \mathcal{P}(x)$ (рис. 8.9). Выделим мысленно в жидкости некоторый параллелепипед со сторонами dx , dy , dz , ориентированными по координатным осям. Тогда в направлении оси x на него будет действовать сила $\mathcal{P}(x) \cdot dy \, dz$, а навстречу ей — $\mathcal{P}(x + dx) \cdot dy \, dz$, суммарная же сила будет равна их разности, которая, в силу малости dx , может быть линеаризована:

$$F_x = \mathcal{P}(x) \cdot dy \, dz - \mathcal{P}(x + dx) \cdot dy \, dz \approx -\frac{d\mathcal{P}}{dx} \cdot dx \, dy \, dz.$$

Удобно ввести понятие *объемной плотности силы*:

$$f_x \equiv \frac{dF_x}{dV} = -\frac{d\mathcal{P}}{dx}. \quad (8.19)$$

Это и есть сила Архимеда. В случае зависимости давления от трех координат, вместо обычной производной в (8.19), следует использовать операцию градиента — см. (5.25).

Представим себе, например, несжимаемую жидкость в поле тяжести. При $\rho(x) = \text{const}$ объемная плотность силы тяжести есть просто ρg . Пусть ось x направлена вглубь жидкости. Запишем условие равновесия любого малого элемента объема dV :

$$\rho g \cdot dV - \frac{d\mathcal{P}}{dx} \cdot dV = 0.$$

Сокращая dV и интегрируя, получаем хрестоматийный результат:

$$\mathcal{P} = \mathcal{P}_0 + \rho g x, \quad (8.20)$$

где \mathcal{P}_0 — давление на поверхности жидкости. Из (8.20) можно получить решение упомянутых выше статических задач. Рассмотрим в качестве примера два открытых сосуда, соединенных посредством сифона (рис. 8.10). Давление жидкости в точке A , с точностью до отношения ширины трубки к высоте ее над уровнем жидкости в сосудах h , равно

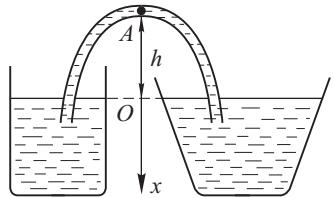


Рис. 8.10

$$\mathcal{P}_0 + \rho g x_A = \mathcal{P}_0 - \rho g h,$$

безотносительно к точке свободной поверхности, от которой мы отсчитываем x_A . Тем самым подтверждается, что в сосудах, соединенных так, как показано на рис. 8.10, жидкость в равновесии оказывается на одном уровне, а при нарушении этого условия должно возникнуть течение, перекачивающее жидкость в сосуд с более низким уровнем через трубку. На этом и основан принцип сифона.

Рассмотрим эффекты, обусловленные работой давления. Прежде всего представим себе жидкость или газ, занимающие в состоянии равновесия некий цилиндр с площадью основания S (рис. 8.11 а). Давление в пределах данного цилиндрического объема будем считать постоянным и равным \mathcal{P} . Пусть одно из оснований смещается на малое расстояние dx , однако при этом количество текучего вещества внутри цилиндра сохраняется, т.е. жидкость (газ) не течет через движущуюся границу. Сила давления вещества, заполняющего цилиндр, на границу равна $\mathcal{P} \cdot S$, соответственно, работа, совершаемая веществом при расширении, есть

$$\delta A = \mathcal{P} \cdot S \cdot dx = \mathcal{P} dV, \quad (8.21) \quad \text{Рис. 8.11}$$

где dV — изменение объема жидкости или газа. Если в процессе расширения как-то изменится и давление, $\mathcal{P} \rightarrow \mathcal{P} + d\mathcal{P}$, то это даст поправку в (8.21) всего лишь второго порядка. Пусть теперь слегка меняет объем жидкое или газообразное тело произвольной формы — на рис. 8.11 б начальное и конечное состояние показаны соответственно сплошной и штриховой линиями. В силу закона Паскаля, давление на любой участок границы действует в направлении нормали. Поэтому на любой малой площадочке, принадлежащей границе, мы можем построить цилиндр с образующей, параллельной направлению нормали ξ (рис. 8.11 б). Дальнейшие рассуждения аналогичны выводу (8.21); вклад всех цилиндров можно просуммировать (с учетом знака!) и получить тот же ответ:

$$\delta A = \mathcal{P} dV,$$

где под dV подразумевается полное изменение объема жидкого элемента.

Рассмотрим *стационарное* течение жидкости или газа *без диссипации*. Выделим некоторую трубку тока (рис. 8.12). Пусть

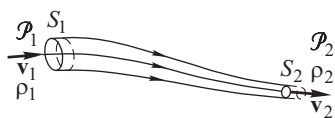


Рис. 8.12

S_1 и S_2 — два произвольных сечения, нормальных к потоку, ρ_1, v_1 и ρ_2, v_2 — соответственно, плотность и скорость в сечениях S_1 и S_2 . За время dt через сечение S_1 протекает объем $S_1 v_1 dt$, при этом втекающий

газ совершает работу $\mathcal{P}_1 S_1 v_1 dt$. За то же время из сечения S_2 вытекает объем $S_2 v_2 dt$, совершая при этом работу $\mathcal{P}_2 S_2 v_2 dt$ над всем вытекающим газом. При стационарном течении массы $\rho_2 S_2 v_2 dt$ и $\rho_1 S_1 v_1 dt$, очевидно, равны. Введем *плотность энергии* текущей среды:

$$\frac{dE}{dV} = \rho \varepsilon + \rho \frac{v^2}{2},$$

где ε включает отнесенную к единице массы внутреннюю энергию плюс энергию во внешних полях, т. е. всю энергию, кроме кинетической, плотность которой равна $\rho v^2/2$. Поскольку мы положили диссипативные эффекты несущественными, используем закон сохранения энергии:

$$\begin{aligned} S_2 v_2 dt \left(\rho_2 \varepsilon_2 + \rho_2 \frac{v_2^2}{2} \right) - S_1 v_1 dt \left(\rho_1 \varepsilon_1 + \rho_1 \frac{v_1^2}{2} \right) = \\ = \mathcal{P}_1 S_1 v_1 dt - \mathcal{P}_2 S_2 v_2 dt. \end{aligned} \quad (8.22)$$

Разделим это уравнение почленно на величину dm протекающей через сечение за время dt массы. Удобно первый член левой части и второй член правой части (8.22) разделить на $dm = \rho_2 S_2 v_2 dt$, а второй член левой части и первый член правой части (8.22) разделить на $dm = \rho_1 S_1 v_1 dt$. В результате получаем *уравнение Бернулли* (по имени математика и физика Д. Бернулли (1700–1782))

$$\frac{\mathcal{P}_1}{\rho_1} + \varepsilon_1 + \frac{v_1^2}{2} = \frac{\mathcal{P}_2}{\rho_2} + \varepsilon_2 + \frac{v_2^2}{2},$$

или, что то же,

$$\frac{\mathcal{P}}{\rho} + \varepsilon + \frac{v^2}{2} = \text{const.} \quad (8.23)$$

Важный частный случай уравнения Бернулли — течение в поле силы тяжести при неизменной массовой плотности внутренней энергии (для идеального газа это обусловлено постоянством температуры). Тогда $dE/dV = \rho gh + \rho v^2/2$, где h — высота по отношению к некоторому заранее определенному нулевому уровню. Как следствие, получаем

$$\frac{\mathcal{P}}{\rho} + gh + \frac{v^2}{2} = \text{const.} \quad (8.24)$$

Если к тому же, жидкость несжимаема, то (8.24) можно использовать в виде

$$\mathcal{P} + \rho gh + \rho \frac{v^2}{2} = \text{const.} \quad (8.25)$$

Именно эту форму записи чаще всего связывают с именем Бернулли; подчеркнем, однако, что уравнение (8.23) гораздо более универсально.

Уравнение (8.25) позволяет легко получить *формулу Торричелли* для скорости вытекания несжимаемой жидкости из сосуда

через малое отверстие (она, правда, была выведена за сто лет до уравнения Бернулли). Постановка задачи ясна из рис. 8.13. Пусть высота уровня воды в сосуде равна h_1 , а высота, на которой расположено отверстие, — h_2 , так что $h_1 - h_2 = h$. Отверстие должно быть достаточно малым, чтобы выполнялось условие $\dot{h} \ll v$. Из уравнения непрерывности (8.14) легко усмотреть, что для этого необходимо, чтобы сечение отверстия было много меньше сечения сосуда. Таким образом обеспечивается с необходимой точностью, во-первых, стационарность течения и тем самым применимость уравнения Бернулли. Во-вторых, при медленном вытекании мы можем избежать вихревых течений, что позволяет рассматривать наше течение как одну-единственную трубку тока (обозначена пунктиром на рис. 8.13). Это очень важный аспект данной задачи; напомним, что уравнения (8.23)–(8.25) задают инвариант только в пределах трубки тока, который не всегда можно распространить на поток как целое.

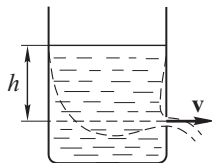


Рис. 8.13

Итак, $\rho(h_1) = \rho(h_2)$; на уровне h_1 $v \approx 0$, а на уровне h_2 , как следует из (8.24),

$$v = \sqrt{2gh}, \quad (8.26)$$

т.е. скорость вытекания несжимаемой жидкости из малого отверстия совпадает со скоростью тела, свободно падающего с высоты h . Это и есть формула Торричелли (по имени итальянского физика Э. Торричелли (1608–1647)).

Отметим в заключение, что величина $\rho v^2/2$ иногда называется динамическим давлением, а формула (8.25) трактуется как инвариантность суммы статического и динамического давлений. В стационарном потоке несжимаемой жидкости давление \mathcal{P} должно быть меньше там, где больше скорость. Так объясняют принцип работы пульверизатора, эффект «присасывания» кораблей при близком прохождении параллельными курсами и т. д.

8.5. Вязкость. Течение Пуазейля

До сих пор мы ничего не говорили о касательных напряжениях в жидкости или газе, ограничиваясь только изотропным давлением в рамках закона Паскаля. Оказывается, однако, что закон Паскаля является исчерпывающим лишь в гидростатике, а в случае неоднородных в пространстве течений вступает в игру диссипативный эффект — *вязкость*, вследствие которого как раз и возникают касательные напряжения.

Пусть в некоторой области потока жидкости два бесконечно близких ее слоя, движущихся в направлении оси x , соприка-

саются друг с другом на горизонтальной поверхности с площадью S (рис. 8.14). Опыт показывает, что возникающая на этой площадке сила трения F между слоями тем больше, чем больше площадь S и чем быстрее изменяется в этом месте скорость потока v в направлении, перпендикулярном к площадке S , то есть, в направлении оси y . Быстрота изменения скорости v как функции y характеризуется производной dv/dy . Окончательно, полученный из опыта результат можно записать в виде

$$F = \eta S \frac{dv}{dy}. \quad (8.27)$$

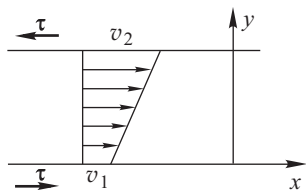


Рис. 8.14

Здесь F — сила, действующая со стороны вышележащего слоя на нижележащий, η — коэффициент пропорциональности, получивший название *коэффициента вязкости* жидкости (сокращенно его называют просто *вязкостью* жидкости). Размерность его вытекает из формулы (8.27): $[\eta] = [m]/[l][t]$; единицу измерения принято выражать как $1 \text{ Па} \cdot \text{с}$. Направление силы F (вправо или влево на рис. 8.14) зависит от того, быстрее или медленнее движется вышележащий слой относительно нижележащего. Из (8.27) следует выражение для касательных напряжений:

$$\tau = \eta \frac{dv}{dy}. \quad (8.28)$$

Коэффициент вязкости η имеет разные значения для различных жидкостей, и для определенной жидкости зависит от внешних условий, в первую очередь, от температуры. По своей природе силы трения в жидкости являются силами межмолекулярного взаимодействия, то есть электромагнитными силами, как и силы трения между твердыми телами.

Перейдем к рассмотрению задачи о вычислении расхода несжимаемой жидкости, текущей в горизонтальной круглой прямолинейной трубе с постоянной площадью поперечного сечения при заданном перепаде давлений. Расходом называется масса жидкости, протекающая в единицу времени через сечение трубы. Эта задача имеет чрезвычайно большое практическое значение: инженерный расчет нефтепроводов и даже обычного водопровода безусловно требует ее решения.

Будем полагать, что нам заданы длина трубы l , ее радиус R ,

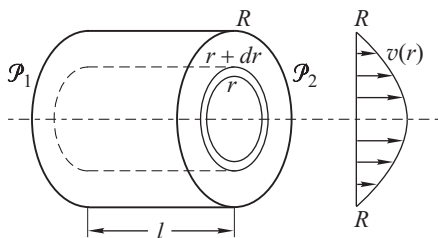


Рис. 8.15

давления на концах трубы \mathcal{P}_1 и \mathcal{P}_2 ($\mathcal{P}_1 > \mathcal{P}_2$), а также плотность жидкости ρ и ее вязкость η (рис. 8.15).

Наличие сил трения приводит к тому, что на разных расстояниях от центра трубы жидкость течет с разной скоростью. В частности, непосредственно у стенки жидкость должна быть неподвижна, иначе из (8.28) следовали бы бесконечные касательные напряжения.

Для вычисления массы жидкости, протекающей каждую секунду через все поперечное сечение трубы, мы разобьем это поперечное сечение на бесконечно малые кольцевые площадки с внутренним радиусом r и внешним $r + dr$ и вычислим сначала расход жидкости через каждое из этих бесконечно малых сечений, в которых скорость жидкости можно считать одинаковой. Просуммировав потом по всем бесконечно малым сечениям, мы определим полный расход жидкости.

Масса жидкости dm , протекающая каждую секунду через бесконечно малое поперечное сечение $2\pi r dr$ со скоростью $v(r)$, равна

$$\frac{dm}{dt} = 2\pi r dr \rho v(r). \quad (8.29)$$

Полный расход жидкости Q мы получим, проинтегрировав выражение (8.29) по r от 0 до R :

$$Q \equiv \frac{dm}{dt} = 2\pi \rho \int_0^R r v(r) dr, \quad (8.30)$$

где мы вынесли за знак интегрирования постоянную величину $2\pi\rho$. Чтобы вычислить интеграл в (8.30), необходимо знать зависимость скорости жидкости от радиуса, то есть конкретный вид функции $v(r)$. Для определения $v(r)$ мы воспользуемся уже известными нам законами механики. Рассмотрим в некоторый момент времени цилиндрический объем жидкости некоторого произвольного радиуса r и длины l (рис. 8.15). Заполняющую этот объем жидкость можно рассматривать как совокупность бесконечно малых жидких частиц, образующих систему взаимодействующих материальных точек. При стационарном течении жидкости в трубе все эти материальные точки движутся с независимыми от времени скоростями. Следовательно, центр масс всей этой системы также движется с постоянной скоростью. Уравнение для движения центра масс системы материальных точек имеет вид (см. гл. 6)

$$M \frac{d\mathbf{V}_{цм}}{dt} = \sum \mathbf{F}_{вн}, \quad (8.31)$$

где M — полная масса системы, $\mathbf{V}_{цм}$ — скорость центра масс, $\sum \mathbf{F}_{вн}$ — сумма внешних сил, приложенных в выбранный момент времени к рассматриваемой системе.

Так как в нашем случае $\mathbf{V}_{цм} = \text{const}$, то из (8.31) получаем

$$\sum \mathbf{F}_{вн} = 0.$$

Внешние силы — это силы давления $F_{\text{давл}}$, действующие на основания выбранного цилиндрического объема, и силы трения $F_{\text{тр}}$, действующие на боковую поверхность цилиндра со стороны окружающей жидкости — см. (8.27):

$$F_{\text{давл}} = \pi r^2 (\mathcal{P}_1 - \mathcal{P}_2), \quad F_{\text{тр}} = 2\pi r l \eta \frac{dv}{dr}.$$

Как мы показали, сумма этих сил равна нулю, то есть

$$2\pi r l \eta \frac{dv}{dr} + \pi r^2 (\mathcal{P}_1 - \mathcal{P}_2) = 0.$$

Это соотношение после простых преобразований можно записать в виде

$$dv = -\frac{\mathcal{P}_1 - \mathcal{P}_2}{2\eta l} r dr.$$

Интегрируя обе части написанного выше равенства, получим

$$v = -\frac{\mathcal{P}_1 - \mathcal{P}_2}{4\eta l} r^2 + \text{const}.$$

Постоянная интегрирования определится из условия, что при $r = R$ скорость v должна обращаться в нуль. Это дает

$$v = \frac{\mathcal{P}_1 - \mathcal{P}_2}{4\eta l} (R^2 - r^2). \quad (8.32)$$

Как мы видим, скорость жидкости максимальна на оси трубы и при удалении от оси меняется по параболическому закону (см. рис. 8.15).

Подставив (8.32) в (8.30), находим искомый расход жидкости

$$Q = \pi \rho \frac{\mathcal{P}_1 - \mathcal{P}_2}{2\eta l} \int_0^R (R^2 - r^2) r dr,$$

или

$$Q = \pi \rho \frac{\mathcal{P}_1 - \mathcal{P}_2}{8\eta l} R^4. \quad (8.33)$$

Это выражение для расхода жидкости называется *формулой Пуазейля* (по имени французского физика Ж. Пуазейля (1799–1869)). Отличительной чертой соотношения (8.33) является сильная зависимость расхода жидкости от радиуса трубы: расход пропорционален четвертой степени радиуса. (Сам Пуазейль эту формулу не выводил, а исследовал проблему только экспериментально, изучая движение жидкости в капиллярах). На формуле Пуазейля основан один из экспериментальных методов определения коэффициентов вязкости жидкостей.

8.6. Турбулентность

Формула Пуазейля применима только для ламинарных течений жидкости. *Ламинарным* называется такое течение, когда жидкие частицы движутся вдоль устойчивых траекторий. При достаточно больших скоростях ламинарное течение становится неустойчивым, хаотичным и переходит в так называемое *турбулентное* течение. При этом основные уравнения гидродинамики остаются в силе, но большинство результатов настоящей главы должно быть подвергнуто ревизии.

Характер турбулентных движений может быть весьма различным в зависимости от внешних условий. Из повседневного опыта нам более всего знакомо явление *гидродинамической турбулентности несжимаемой жидкости* — представление о таком режиме дает турбулентная струя, вытекающая из крана. Реализуется данное состояние жидкости в случае перепада скоростей течения, много меньшего скорости звука. Для него характерно образование множественных вихрей (М. В. Ломоносов подразделял течения жидкости на «покойные» и «коловратные»). На рис. 8.16 *а* схематически изображено образование вихрей при обтекании несжимаемой жидкостью шара, а на рис. 8.16 *б* — при течении в прямой трубе.

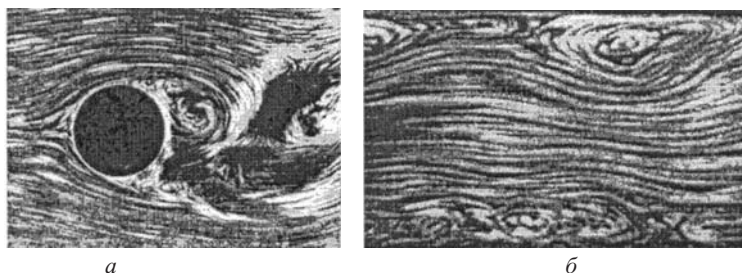


Рис. 8.16

Следует особо обратить внимание на то, что во всех оценках, относящихся к турбулентному состоянию, фигурирует не скорость течения как таковая, а именно перепад скоростей $|\Delta \mathbf{v}|$. Действительно, если некоторое течение происходит с постоянной всюду скоростью \mathbf{v} , мы можем, переходя в движущуюся систему отсчета, сделать эту скорость как угодно малой или большой, так что подлинная «индивидуальность» течения может быть отражена лишь величиной $|\Delta \mathbf{v}|$.

Хотя детальное рассмотрение физики турбулентных состояний далеко выходит за рамки нашего курса, мы все же можем кое-что сказать о переходе от ламинарного («покойного») к турбулентному («коловратному») течению. Принципиальным моментом здесь будет использование *методов подобия и размерно-*

сти. Дело в том, что размерность физических величин — не только один из важных методов контроля правильности ответов, но и метод получения новых результатов в физике вообще и в механике, в частности, хотя возможности его ограничены и предсказательная сила не абсолютна (чему последующее рассмотрение будет хорошим примером).

Попробуем проанализировать физические предпосылки перехода к турбулентному течению, представленному на рис. 8.16. Пока и поскольку течение остается ламинарным, у него сохраняется характерный пространственный масштаб a (радиус шара на рис. 8.16 а либо диаметр трубы на рис. 8.16 б). Турбулизация течения приводит к измельчению масштаба, а значит, и к росту пространственных производных от скорости, чему, как можно усмотреть из формул (8.27), (8.28), противодействуют эффекты, обусловленные вязкостью. Таким образом, *чем больше вязкость жидкости, тем более затруднен переход в турбулентное состояние*. Еще более очевидным представляется следующее утверждение: *чем больше перепад скоростей, тем легче переход из ламинарного режима в турбулентный* — скажем, в предельном случае пространственно-однородного течения турбулентности просто не может быть, поскольку оно эквивалентно состоянию покоя.

Как уже указывалось выше, размерность вязкости составляет

$$[\eta] = \frac{[m]}{[l][t]}.$$

Составим безразмерную комбинацию

$$\text{Re} = \frac{m}{lt\eta}.$$

Подставляя в это выражение значения характерных параметров:

$$m = \rho a^3, \quad l = a, \quad t = \frac{a}{v},$$

получим величину

$$\text{Re} = \frac{\rho v a}{\eta}, \quad (8.34)$$

которую принято называть *числом Рейнольдса* (по имени английского физика О. Рейнольдса (1842–1912)). Число Рейнольдса можно переписать в виде

$$\text{Re} = \frac{\rho v^2 l^3}{\eta v l^2},$$

откуда следует, что оно характеризует отношение кинетической энергии движущейся жидкости $T \sim \rho v^2 l^3 / 2$ к потере этой энергии, равной работе сил вязкости на характерной длине l , $A \approx \eta v l^2$. Помимо вязкости η , жидкость характеризуется еще и плотностью ρ , а течение как таковое — пространственным

масштабом l и характерным перепадом скоростей v . (Применительно к ситуации рис. 8.16 это просто скорость по порядку величины, так как вблизи неподвижных стенок скорость должна обращаться в нуль). Изложенные выше рассуждения позволяют заключить, что чем больше величина Re , тем благоприятнее ситуация для турбулентного течения, и напротив, для течений с малыми числами Рейнольдса должен, по-видимому, реализоваться ламинарный режим.

Данные эксперимента находятся с этими соображениями в полном согласии. Оказывается, действительно существует *критическое число Рейнольдса*, при превышении которого происходит переход из ламинарного режима в турбулентный. Но сама величина $Re_{кр}$, как показывает опыт, не универсальна — она зависит от геометрии системы. Например, в случае течения по трубе (рис. 8.16б) $Re_{кр} \sim 2 \cdot 10^3$, тогда как при вращении цилиндра в газе переход в турбулентный режим происходит при $Re_{кр} \sim 50$. В этом и состоит некоторая слабость чисто размерностной оценки.

Но у нее есть и очень сильная сторона. Сами по себе величины ρ , v , a , η могут меняться в очень широком диапазоне; скажем труба может быть капилляром, а может быть аэродинамической трубой диаметром в десятки метров — ответ, тем не менее, будет универсальным и опираться он будет всего лишь на одну безразмерную комбинацию — число Рейнольдса. Такие зависимости называются в физике *законами подобия*, а переход на их основании от одних экспериментальных ситуаций к другим принято именовать *скэйлингом*. Таким образом, если для некоторого физического явления известен закон подобия, мы можем ставить эксперименты в малом масштабе, так чтобы они были более дешевыми или, например, более безопасными, а затем проводить скэйлинг для получения ответа в интересующем нас масштабе. Поэтому методы подобия и размерности занимают достаточно почетное место в арсенале современной физики.

Вопросы и задачи

1. Каким числом констант описываются упругие свойства изотропного вещества?
2. Можно ли ввести модуль сдвига для стекла?
3. Из каких свойств жидкостей и газов следует закон Паскаля?
4. Что такое вязкость?
5. Допустимо ли использовать уравнение Бернулли для вязкой жидкости?
6. Оценить давление в центре Земли, полагая плотность вещества постоянной и равной $\rho \approx 5,5 \cdot 10^3$ г/см³. Радиус Земли $R \approx 6400$ км.

Решение. Поскольку объемная плотность силы тяжести, в отличие от простых «настошных» задач, зависит от расстояния до центра Земли, уравнение

равновесия, с учетом (8.19), следует в данном случае использовать в виде:

$$-\frac{d\mathcal{P}}{dr} - \rho g(r) = 0, \quad \text{где} \quad g(r) = \frac{\gamma M(r)}{r^2}, \quad M(r) = \frac{4}{3} \pi r^3 \rho.$$

Мы учли то обстоятельство, что в случае сферической симметрии все величины зависят только от r ; что сила тяжести направлена к центру Земли, т.е. против \mathbf{r} — отсюда и знак «—», и, наконец, что внешние слои Земли с радиусами, большими r , вкладывают в силу притяжения не дают (см. гл. 6). Окончательно,

$$\frac{d\mathcal{P}}{dr} = -\frac{4\pi}{3} \gamma \rho^2 r \Rightarrow \mathcal{P} = \text{const} - \frac{2\pi}{3} \gamma \rho^2 r^2 \approx \frac{2\pi}{3} \gamma \rho^2 (R^2 - r^2),$$

если учесть, что на поверхности Земли, т.е. при $r = R$, $\mathcal{P} \approx 0$. Подставляя $\rho = 5,5 \cdot 10^3 \text{ кг/м}^3$, $R = 6,4 \cdot 10^6 \text{ м}$, получаем $\mathcal{P}_{\max} = \mathcal{P}|_{r=0} \approx 2,6 \times 10^{11} \text{ Па} \approx 2,6 \cdot 10^6 \text{ атм}$. Большая точность в оценках бессмысленна, хотя бы по той причине, что плотность вещества не может быть инвариантной при таком перепаде давлений.

7. Определите отношение энергий деформации стального и пластмассового цилиндров одинаковых размеров для двух случаев: 1) цилиндры поставлены рядом друг с другом и сжаты параллельными плоскостями; 2) цилиндры поставлены друг на друга и также сжаты плоскостями. Модуль Юнга для стали равен $E_1 = 2 \cdot 10^5 \text{ Н/мм}^2$, для пластмассы — $E_2 = 10^2 \text{ Н/мм}^2$.

Ответ: 1) $U_1/U_2 = E_1/E_2 = 2 \cdot 10^3$; 2) $U_1/U_2 = E_2/E_1 = 5 \cdot 10^{-4}$.

8. В вертикально стоящий цилиндрический сосуд налита идеальная жидкость до уровня H (относительно дна сосуда). Площадь дна сосуда равна S . Определите время T , за которое вся жидкость выльется из сосуда, если в дне сосуда сделано малое отверстие площади σ . Считать режим течения несжимаемым.

Решение. Движение не стационарно — скорость истечения меняется, и верхняя граница жидкости движется. Уравнением Бернулли, полученным для стационарного течения, можно пользоваться, если эти изменения происходят медленно, как в случае $\sigma \ll S$. Обозначим высоту поверхности жидкости относительно дна сосуда в момент t через h , скорость жидкости на этой поверхности (скорость движения частиц, находящихся на границе) — через $v_1 = -dh/dt$, скорость жидкости в отверстии — через v_2 . Учитывая, что давление на верхней поверхности жидкости и на поверхности вытекающей струи примерно одинаково (равно атмосферному) и что в силу несжимаемости жидкости $v_1 S = v_2 \sigma$, получаем из уравнения Бернулли, рассматривая всю жидкость как единую трубку тока, $v_2^2 - v_1^2 = 2gh$. Учитывая, что $\sigma \ll S$, находим

$$-\frac{dh}{\sqrt{h}} = \frac{\sigma \sqrt{2g}}{S} dt.$$

После интегрирования отсюда получаем

$$t = \frac{S}{\sigma} \sqrt{\frac{2}{g}} (\sqrt{H} - \sqrt{h}).$$

Полагая здесь $h = 0$, определяем

$$T = \frac{S}{\sigma} \sqrt{\frac{2H}{g}}.$$

9. Из неплотно закрытого крана вытекает в единицу времени количество воды $q = 1 \text{ см}^3/\text{с}$. Вода попадает в сосуд и затем вытекает из него по горизонтальной трубке длины $l = 20 \text{ см}$. Со временем в сосуде устанавливается уровень воды на $h = 5 \text{ см}$ выше уровня трубки. Определите диаметр трубки. Вязкость воды $\eta = 10^{-3} \text{ Па} \cdot \text{с}$.

Ответ: $d = 2 \cdot [(8/\pi)(\eta l / \rho g h) q]^1/4 \approx 2 \text{ мм}$.

Глава 9

ЗАКОНЫ МЕХАНИКИ В НЕИНЕРЦИАЛЬНЫХ СИСТЕМАХ ОТСЧЕТА

9.1. Принцип относительности Галилея

Как меняются законы движения при переходе из одной системы отсчета в другую? Другими словами, меняется ли при этом (и как) основной закон механики — второй закон Ньютона —

$$m\mathbf{a} = \mathbf{F}.$$

Этот вопрос имеет очень важное значение, так как наблюдать за движением тел и использовать законы механики на практике приходится не в одной какой-то, раз и навсегда выбранной системе отсчета, а в различных системах, по-разному движущихся друг относительно друга. Особое значение придает данной проблеме то обстоятельство, что инерциальная система отсчета, в которой мы до настоящего времени формулировали законы механики, есть физическая идеализация, тогда как в природе мы всегда имеем дело с неинерциальными системами.

Рассмотрим случай, когда обе системы отсчета — исходная и движущаяся относительно нее — являются инерциальными системами. Допустим, что система отсчета K инерциальна, а система K' движется относительно первой поступательно с постоянной скоростью \mathbf{V} (рис. 9.1). Для простоты можно принять, что координатные оси x' , y' , z' соответственно параллельны осям x , y , z и что в начальный момент времени $t = 0$ начало O' совмещается с началом O . Будем также считать, что скорость \mathbf{V} параллельна оси x . При этих условиях ось x' все время будет совпадать с осью x . Такие упрощения в постановке задачи не лишают ее общности, так как переход к общим формулам может быть совершен дополнительным поворотом координатных осей и переносом начала координат. Радиус-вектор некоторой материальной точки m в исходной системе отсчета в момент времени t обозначим через \mathbf{r} , а радиус-вектор той же материальной точки в тот же момент времени

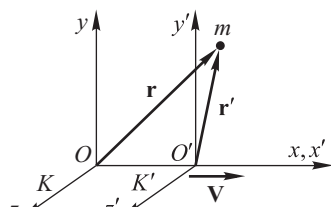


Рис. 9.1

в движущейся системе обозначим через \mathbf{r}' . Тогда координаты и время в системах K и K' будут связаны друг с другом соотношениями

$$\mathbf{r} = \mathbf{r}' + \mathbf{V}t', \quad t = t', \quad (9.1)$$

или в проекциях на оси

$$x = x' + Vt', \quad y = y', \quad z = z', \quad t = t'.$$

Мы уже обсуждали эти соотношения в гл. 2. Напомним, что они называются преобразованиями Галилея. Мы добавили к формулам преобразования координат дополнительную формулу $t = t'$, выражающую предположение Ньютона о том, что время является абсолютным, то есть текущим одинаково в любых системах отсчета.

С точки зрения нашего повседневного житейского опыта, преобразования Галилея кажутся очевидными. В самом деле, они фактически основаны на двух предположениях. Во-первых, предполагается, что в разных системах отсчета остаются неизменными длины одних и тех же твердых стержней, которые используются для измерения пространственных размеров и координат различных тел. Кроме того, преобразования Галилея предполагают также, что, например, показания часов у двух человек не станут различаться только из-за того, что один из них начнет идти быстрее другого, и это тоже, казалось бы, не вызывает сомнения. Но всегда ли здравый смысл достаточен для доказательства истины? Об этом пойдет речь в следующей главе, а сейчас поговорим о том, что означают преобразования Галилея с точки зрения формулировки законов механики в разных инерциальных системах. Различаются ли законы движения для наблюдателей в разных системах?

Дифференцируя соотношение (9.1) по времени t , получим

$$\frac{d\mathbf{r}}{dt} = \frac{d\mathbf{r}'}{dt'} + \mathbf{V},$$

или

$$\mathbf{v} = \mathbf{v}' + \mathbf{V}, \quad (9.2)$$

где \mathbf{v} — скорость материальной точки в системе K , а \mathbf{v}' — в системе K' . Эта формула выражает известное уже нам правило сложения скоростей в механике Ньютона.

Дифференцируя второй раз, получим (с учетом постоянства \mathbf{V})

$$\frac{d\mathbf{v}}{dt} = \frac{d\mathbf{v}'}{dt} = \frac{d\mathbf{v}'}{dt'},$$

или

$$\mathbf{a} = \mathbf{a}'. \quad (9.3)$$

Таким образом, ускорение *инвариантно* относительно преобразований Галилея.

С правилом сложения скоростей и с равенством (9.3) мы познакомились впервые в гл. 2. Поставим теперь вопрос: а как меняется сила при переходе из одной инерциальной системы в другую? Сила зависит от разности координат взаимодействующих материальных точек (для электромагнитных сил — еще и от разности их скоростей). Поэтому, в соответствии с (9.1) и (9.2), сила не меняется при переходе от одной системы отсчета к другой: $\mathbf{F} = \mathbf{F}'$. Такие соображения, сколь бы они не казались естественными, ни в коей мере не являются доказательством — они, например, должны быть пересмотрены в рамках релятивистской механики. Иначе говоря, сила инвариантна лишь относительно преобразований Галилея. Это утверждение должно рассматриваться как опытный факт. Так как и ускорение инвариантно, а масса материальной точки предполагается величиной постоянной, не зависящей от ее положения и скорости, то второй закон Ньютона в «штрихованной» системе принимает вид

$$m\mathbf{a}' = \mathbf{F}'.$$

Это уравнение в «штрихованной» системе отсчета K' имеет точно такой же вид, что и в «нештрихованной» системе K . Таким образом, *уравнения механики Ньютона инвариантны относительно преобразований Галилея*. Это утверждение составляет содержание *принципа относительности Галилея*.

Принцип относительности Галилея провозглашает полное равноправие всех инерциальных систем отсчета и его можно сформулировать также в виде следующего утверждения: *никакими механическими опытами, проведенными в пределах только данной системы отсчета, нельзя установить, находится ли она в состоянии покоя или в состоянии равномерного прямолинейного движения*. Находясь, например, в вагоне поезда, движущегося без толчков прямолинейно и равномерно, мы, не выглянув в окно, не сможем определить, движется вагон или покоится. Свободное падение тел, движение брошенных нами предметов и все другие механические процессы будут в этом случае происходить так же, как и в случае, если бы вагон был неподвижен.

9.2. Законы механики в неинерциальных системах отсчета. Силы инерции

Рассмотрим теперь вопрос о том, как формулируются законы механики в неинерциальных системах отсчета. Вопрос этот имеет важное практическое значение: система отсчета, связанная с поверхностью Земли, не является инерциальной — например, в системе отсчета, связанной с Солнцем, точки на поверхно-

сти Земли испытывают центростремительное ускорение, обусловленное вращением Земли вокруг собственной оси и вокруг Солнца. При определенных условиях этой неинерциальностью можно пренебречь, но при решении многих практических задач (например, запуск космического корабля) без учета этой неинерциальности не обойтись. К тому же, многие машины и механизмы реально работают в неинерциальных системах отсчета — в движущихся с ускорением вагонах, самолетах, космических кораблях и так далее.

Итак, посмотрим, как преобразуется основной закон механики — второй закон Ньютона — при переходе от инерциальной системы отсчета к неинерциальной. Остановимся отдельно на двух случаях: рассмотрим сначала простейший случай поступательного движения неинерциальной системы, а затем обсудим, как видоизменяются законы механики во вращающейся системе отсчета.

Поступательное движение неинерциальной системы отсчета. В повседневной жизни каждому из нас чуть ли не ежедневно приходится испытывать неприятные ощущения, когда при резком торможении автобуса или вагона метро какая-то сила бросает нас вперед. Чтобы понять происхождение этой силы, рассмотрим две системы отсчета: инерциальную систему K и систему K' , которая совершает поступательное движение относительно системы K с ускорением \mathbf{A} , зависящем в общем случае от времени (рис. 9.2). Пусть \mathbf{r} и \mathbf{r}' — радиусы-векторы

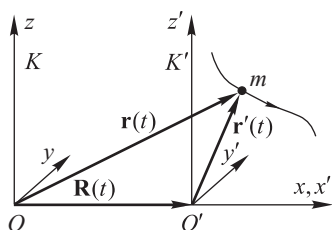


Рис. 9.2

материальной точки с массой m в инерциальной и неинерциальной системах соответственно, а $\mathbf{R}(t)$ — радиус-вектор начала отсчета системы K' относительно системы K .

Уравнение движения материальной точки в инерциальной системе отсчета K — второй закон Ньютона:

$$m\mathbf{a} = \mathbf{F},$$

где \mathbf{F} — результирующая сила, действующая на рассматриваемую материальную точку со стороны других тел, $\mathbf{a} = d^2\mathbf{r}/dt^2$ — ускорение материальной точки в инерциальной системе. Координаты и скорости материальной точки в системах K и K' связаны друг с другом соотношениями (9.1), (9.2). Если тело неподвижно относительно K' , то движение этой системы приводит к переносу тела относительно K . Скорость \mathbf{V} называется *переносной скоростью*. Скорость относительно системы K' , т.е. \mathbf{v}' так и называется — *относительная скорость*. Наконец, скорость \mathbf{v} в инерциальной системе K носит название *абсолютной ско-*

рости. Надо понимать условность последнего названия: это, конечно, просто скорость относительно системы K . Но K — система инерциальная, в некоторой степени «привилегированная», и только в этом условном смысле скорость может быть названа абсолютной.

Продифференцировав еще раз по времени обе части равенства (9.2), получаем соотношение между ускорениями

$$\mathbf{a} = \mathbf{A} + \mathbf{a}'. \quad (9.4)$$

Здесь, соответственно, \mathbf{a} — абсолютное ускорение, \mathbf{A} — переносное и \mathbf{a}' — относительное. Подставив полученное выражение для \mathbf{a} во второй закон Ньютона, перепишем его в виде

$$m\mathbf{a}' = \mathbf{F} - m\mathbf{A}. \quad (9.5)$$

Как мы видим, уравнение движения в неинерциальной системе отсчета (9.5) отличается от второго закона Ньютона в инерциальной системе (3.2) тем, что в правой части уравнения наряду с силой \mathbf{F} появляется еще добавочное слагаемое $-m\mathbf{A}$. Если ввести обозначение

$$-m\mathbf{A} = \mathbf{F}_{\text{ин}}, \quad (9.6)$$

то уравнение движения в неинерциальной системе примет такой же привычный вид, как и второй закон Ньютона,

$$m\mathbf{a}' = \mathbf{F} + \mathbf{F}_{\text{ин}}, \quad (9.7)$$

где определенное равенством (9.6) добавочное слагаемое $\mathbf{F}_{\text{ин}}$ называют *силой инерции*, конкретно — *поступательной силой инерции*. Далее мы познакомимся и с другими силами инерции.

В данном случае сила инерции является *пространственно-однородной*, т.е. сила инерции при поступательном движении неинерциальной системы отсчета имеет одно и то же значение для всех точек этой системы. Это следует из (9.6): сила инерции зависит только от ускорения, с которым начало неинерциальной системы K' движется относительно инерциальной системы K .

Итак, движение относительно рассматриваемой неинерциальной системы можно исследовать двумя способами. Можно определить закон движения материальной точки $\mathbf{r}(t)$ в некоторой инерциальной системе, используя второй закон Ньютона в его стандартном виде, а затем пересчитать его относительно неинерциальной системы, т.е. получить $\mathbf{r}'(t)$ из закона преобразования координат (9.1) (при условии, конечно, что закон движения неинерциальной системы $\mathbf{R}(t)$ известен). Но можно сразу решать задачу в неинерциальной системе отсчета с помощью видоизмененного второго закона Ньютона (9.7), в котором в правой части к реальной силе \mathbf{F} , определяемой взаимодействием рассматриваемого тела с другими телами, добавлена сила инерции $\mathbf{F}_{\text{ин}}$, опре-

деляемая соотношением (9.6). Появление этой добавочной силы при рассмотрении движения относительно неинерциальной системы отсчета — формальное следствие преобразования координат (8.7) и не отражает появления какого-либо нового воздействия на материальную точку со стороны других тел. В этом смысле силу инерции можно назвать фиктивной силой, хотя для наблюдателя в неинерциальной системе она будет приводить к таким же последствиям, как и реальная сила той же величины.

Мы уже упоминали об ощущениях наблюдателя в неинерциальной системе (в автобусе, например), связанных с ее торможением или ускорением. Поясним еще действие рассмотренной силы инерции на следующем примере. Рассмотрим тележку с укрепленным на ней кронштейном, к которому подвешен на нити шарик (рис. 9.3). Пока тележка покоится или движется без ускорения, нить расположена вертикально и сила тяжести \mathbf{P} уравнивается натяжением нити \mathbf{F}_n . Приведем теперь тележку в поступательное движение с ускорением \mathbf{A} . Нить отклонится от вертикали на такой угол, чтобы результирующая сил \mathbf{P} и \mathbf{F}_n сообщала шарiku ускорение, равное \mathbf{A} , т.е. в инерциальной системе отсчета угол отклонения нити определяется условием, которое, как и должно быть, является следствием второго закона Ньютона. В неинерциальной системе отсчета, связанной с тележкой, шарик покоится, несмотря на то, что результирующая сил \mathbf{P} и \mathbf{F}_n отлична от нуля. Отсутствие ускорения шарика по отношению к этой системе отсчета можно формально объяснить

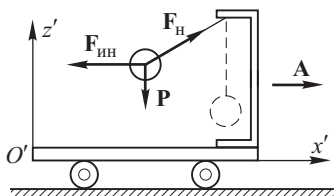


Рис. 9.3

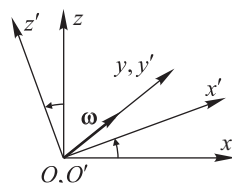


Рис. 9.4

тем, что кроме сил \mathbf{P} и \mathbf{F}_n , равных в сумме $m\mathbf{A}$, на шарик действует еще и сила инерции $\mathbf{F}_{ин} = -m\mathbf{A}$.

Силы инерции во вращающейся системе отсчета. Рассмотрим теперь случай вращающейся неинерциальной системы и для определенности рассмотрим ситуацию, когда неинерциальная система K' вращается с угловой скоростью ω вокруг оси y' , совпадающей с осью y инерциальной системы K . Будем полагать также, что начала отсчета этих систем совпадают (рис. 9.4). Наша цель — записать уравнение движения материальной точки в неинерциальной системе K' в виде второго закона Ньютона (9.7).

В инерциальной системе K уравнение движения имеет обычный вид

$$m\mathbf{a} = \mathbf{F},$$

где \mathbf{a} — ускорение рассматриваемой материальной точки в системе K . Вычитая его из уравнения (9.7) (из левой части одного — левую часть другого, из правой части — правую), получим

$$\mathbf{F}_{\text{ин}} = -m(\mathbf{a} - \mathbf{a}'). \quad (9.8)$$

Это выражение для силы инерции является обобщением выражения (9.6) для $\mathbf{F}_{\text{ин}}$ в случае прямолинейного движения неинерциальной системы отсчета. Для поступательного движения неинерциальной системы $\mathbf{a} - \mathbf{a}' = \mathbf{A}$. Итак, в общем случае сила инерции равна произведению массы тела на взятую с обратным знаком разность его ускорений по отношению к инерциальной и неинерциальной системам отсчета. При поступательном движении неинерциальной системы K' относительно инерциальной системы K эта разность равна переносному ускорению \mathbf{A} , т. е. ускорению начала отсчета системы K' . Рассмотрим в качестве примера относительно простой случай.

Пусть материальная точка во вращающейся системе K' движется в плоскости $x'z'$ по окружности радиуса R с постоянным модулем скорости v' (рис. 9.5). В неинерциальной системе K' ускорение материальной точки \mathbf{a}' — это центростремительное ускорение, направленное к центру вращения и равное по модулю $a' = v'^2/R$.

В инерциальной системе K материальная точка движется по той же самой окружности, но модуль ее скорости в этой системе определяется суммой $v' + v$, где $v = \omega R$ — скорость движения по окружности в инерциальной системе той точки вращающейся системы K' , в которой находится рассматриваемая материальная точка (мы рассматриваем случай, когда \mathbf{v} и \mathbf{v}' направлены в одну сторону). Следовательно, в инерциальной системе K ускорение материальной точки \mathbf{a} — это также центростремительное ускорение, направленное к центру вращения и по модулю равное

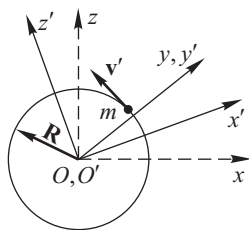


Рис. 9.5

$$a = \frac{(v' + \omega R)^2}{R} = \frac{v'^2}{R} + 2v'\omega + \omega^2 R. \quad (9.9)$$

Воспользуемся теперь определением (9.8) и найдем следующее выражение для модуля силы инерции в рассматриваемом случае:

$$F_{\text{ин}} = m(a - a') = m\omega^2 R + 2mv'\omega. \quad (9.10)$$

В соответствии с (9.8) направление найденной силы инерции противоположно направлению разности векторов $\mathbf{a} - \mathbf{a}'$, направленных в сторону оси вращения. Поскольку, согласно (9.9), модуль \mathbf{a} больше модуля \mathbf{a}' , то разность центростремительных ускорений $\mathbf{a} - \mathbf{a}'$ направлена к оси вращения, а следовательно, $\mathbf{F}_{\text{ин}}$ направлена от оси вращения. Окончательно, силу инерции (9.10) можно представить как сумму двух сил, имеющих специальные названия:

$$\mathbf{F}_{\text{ин}} = \mathbf{F}_{\text{цб}} + \mathbf{F}_{\text{к}}. \quad (9.11)$$

Здесь $\mathbf{F}_{\text{цб}}$ — *центробежная сила инерции*

$$\mathbf{F}_{\text{цб}} = m \omega^2 \mathbf{R}. \quad (9.12)$$

В данном случае \mathbf{R} — радиус-вектор, проведенный к материальной точке из начала координат, потому что точка движется в плоскости $x'O'z'$. В общем случае это вектор, проведенный от оси вращения к точке в плоскости траектории и его модуль R дает расстояние материальной точки от оси вращения.

Центробежная сила инерции (9.12) действует на тело во вращающейся системе отсчета независимо от того, покоится тело в этой системе или движется относительно нее. Величина и направление этой силы определяются движением системы отсчета, т. е. *переносным ускорением*. В этом отношении она подобна силе инерции при поступательном движении системы. Однако центробежная сила зависит еще и от положения тела во вращающейся системе; она не обладает пространственной однородностью. Центробежную силу инерции необходимо учитывать при точном решении задач о движении тел относительно земной поверхности — ее учет приводит, в частности, к небольшим поправкам к силе тяжести (эти поправки составляют доли процента).

Сила $\mathbf{F}_{\text{к}}$ в формуле (9.11) называется *силой Кориолиса* (по имени французского механика Г.Г. Кориолиса (1792–1843)), или *кориолисовой силой инерции*. Для рассмотренного нами частного случая, когда направление скорости движения материальной точки перпендикулярно оси вращения, ее модуль равен $2mv'\omega$ и она направлена вдоль \mathbf{R} . В общем случае выражение для силы Кориолиса имеет вид

$$\mathbf{F}_{\text{к}} = 2m[\mathbf{v}'\boldsymbol{\omega}], \quad (9.13)$$

где $\boldsymbol{\omega}$ — вектор угловой скорости. Отличительной особенностью силы Кориолиса (9.13) является то, что она действует только на движущиеся относительно вращающейся системы отсчета тела. Она пропорциональна векторному произведению \mathbf{v}' и $\boldsymbol{\omega}$, то есть перпендикулярна обоим этим векторам, и ее направление определяется по правилу винта.

Влияние силы Кориолиса следует учитывать, в частности, при истолковании некоторых явлений, связанных с движени-

ем тел относительно земной поверхности. Например, летящий снаряд испытывает отклонения, обусловленные кориолисовыми силами инерции. В северном полушарии при горизонтальном полете, независимо от направления, снаряд будет отклоняться вправо, а в южном полушарии — влево. По этой же причине в нашем полушарии у рек подмывается преимущественно правый берег, в южном полушарии — левый.

Итак, в инерциальной системе отсчета ускорение тела определяется «реальными» силами, действующими со стороны других тел (и полей). В неинерциальной системе для определения относительного ускорения, т.е. относительно этой системы, надо добавить силы инерции.

В общем случае система отсчета может двигаться относительно инерциальной (которую мы условно можем считать неподвижной, «абсолютной») поступательно и одновременно вращаться. Связь между радиусами-векторами частицы в двух системах по-прежнему определяется соотношением

$$\mathbf{r} = \mathbf{R} + \mathbf{r}'.$$

Скорость в абсолютной системе теперь складывается из трех составляющих. Изменение вектора \mathbf{r}' определяет движение относительно системы K' — это относительная скорость. Свой вклад в абсолютную скорость дает движение системы K' относительно системы K — это скорость начала координат системы K' . Но если даже тело неподвижно в системе K' и начало координат этой системы неподвижно в системе K , тело все же будет двигаться относительно K по окружности радиуса \mathbf{r}' . Как мы знаем, скорость в этом случае равна величине $[\boldsymbol{\omega} \mathbf{r}']$. Таким образом, переносная скорость складывается из двух составляющих, связанных соответственно с поступательным движением системы K' и с ее вращением. В результате получаем

$$\mathbf{v} \equiv \dot{\mathbf{r}} = \dot{\mathbf{R}} + \dot{\mathbf{r}}' = \mathbf{v}_{\text{пер}} + \mathbf{v}_{\text{отн}} = \mathbf{V} + \mathbf{v}_{\text{отн}} + [\boldsymbol{\omega} \mathbf{r}']. \quad (9.14)$$

Для нахождения связи ускорений нам надо продифференцировать по времени выражение (9.14). Получаем

$$\mathbf{a} \equiv \dot{\mathbf{v}} = \dot{\mathbf{V}} + \dot{\mathbf{v}}_{\text{отн}} + [\boldsymbol{\omega} \dot{\mathbf{r}}'] + [\dot{\boldsymbol{\omega}} \mathbf{r}']. \quad (9.15)$$

В этом случае также надо учесть наличие двух причин изменения вектора в системе K . Это надо иметь в виду при вычислении второго и третьего членов в формуле (9.14). Не забудем, что

$$\dot{\mathbf{r}}' = \mathbf{v}_{\text{отн}} + [\boldsymbol{\omega} \mathbf{r}'].$$

Нетрудно видеть, что мы должны получить такое выражение:

$$\mathbf{a} = \mathbf{A} + \mathbf{a}_{\text{отн}} + [\boldsymbol{\omega} \mathbf{v}_{\text{отн}}] + [\boldsymbol{\omega} \mathbf{v}_{\text{отн}}] + [\boldsymbol{\omega} [\boldsymbol{\omega} \mathbf{r}']] + [\dot{\boldsymbol{\omega}} \mathbf{r}'].$$

Учитывая, что вклад в векторное произведение дает только перпендикулярная одному из векторов составляющая второго вектора, из правил преобразования двойного векторного произведения

$$[\mathbf{a} [\mathbf{b} \mathbf{c}]] = \mathbf{b}(\mathbf{a} \mathbf{c}) - \mathbf{c}(\mathbf{a} \mathbf{b})$$

получаем

$$[\boldsymbol{\omega} [\boldsymbol{\omega} \mathbf{r}']] \equiv [\boldsymbol{\omega} [\boldsymbol{\omega} \mathbf{r}'_{\perp}]] = \boldsymbol{\omega}(\boldsymbol{\omega} \mathbf{r}'_{\perp}) - \mathbf{r}'_{\perp}(\boldsymbol{\omega} \boldsymbol{\omega}).$$

Скалярное произведение двух взаимно перпендикулярных векторов равно нулю, и от последнего выражения остается одно слагаемое

$$-\mathbf{r}'_{\perp}(\boldsymbol{\omega} \boldsymbol{\omega}) = -\mathbf{r}' \omega^2.$$

Окончательно имеем

$$\mathbf{a} = \mathbf{a}_{\text{отн}} + \mathbf{A} - \mathbf{r}' \omega^2 + 2[\boldsymbol{\omega} \mathbf{v}_{\text{отн}}] + [\dot{\boldsymbol{\omega}} \mathbf{r}'].$$

В соответствии с определением (9.8) получаем выражение для суммарной силы инерции в общем случае

$$\mathbf{F}_{\text{ин}} = -m(\mathbf{A} - \mathbf{r}' \omega^2 + 2[\boldsymbol{\omega} \mathbf{v}_{\text{отн}}] + [\dot{\boldsymbol{\omega}} \mathbf{r}']). \quad (9.16)$$

Напомним еще раз, что первый член в этой формуле определяет поступательную силу инерции, второй — центробежную, четвертый — силу инерции, возникающую при неравномерном вращении неинерциальной системы (этот член редко представляет интерес). Все эти силы инерции связаны с переносным ускорением. Особняком стоит третий член — кориолисова сила инерции, связанная и с движением (вращением) неинерциальной системы, и с движением тела в этой системе.

Одним из любопытных проявлений сил инерции является *состояние невесомости*. Представим себе, что кто-то стоит на весах в кабине покоящегося лифта (рис. 9.6 а). Вес — это сила, с которой тело действует на опору, по модулю равная (в соответствии с третьим законом Ньютона) силе, с которой опора действует на тело. Обозначим последнюю через \mathbf{N} . Когда лифт покоится, сила \mathbf{N} уравнивает силу тяжести, то есть $\mathbf{N} = m\mathbf{g}$. Теперь представим, что лифт свободно падает, то есть движется вниз с ускорением \mathbf{g} (рис. 9.6 б). Для наблюдателя в лифте (в неинерциальной системе отсчета) это означает появление «дополнительной» силы инерции $\mathbf{F}_{\text{ин}} = -m\mathbf{g}$. Эта сила направлена вверх и в точности уравнивает силу тяжести, так что ни-

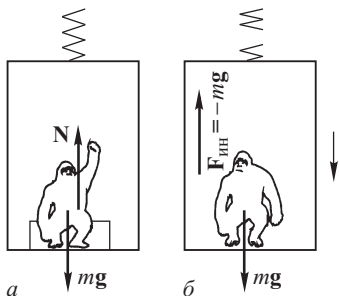


Рис. 9.6

полнительной» силы инерции $\mathbf{F}_{\text{ин}} = -m\mathbf{g}$. Эта сила направлена вверх и в точности уравнивает силу тяжести, так что ни-

какого взаимодействия человека с опорой весов не будет, что и означает обращение его веса в нуль. Таким образом, возникающее в рассмотренных условиях состояние невесомости связано с тем, что силы инерции компенсируют силу тяжести.

Аналогичная компенсация гравитационного притяжения к Земле силой инерции происходит и при рассмотрении явлений относительно космического корабля, вращающегося вокруг Земли. В этом случае притяжение к Земле компенсируется центробежной силой инерции (рис. 9.7). В самом деле, центробежная сила инерции, действующая на обитателя космического корабля, равна (см. (9.12))

$$\mathbf{F}_{\text{ин}} = m\omega^2\mathbf{R}, \quad (9.17)$$

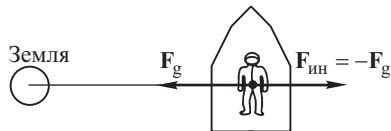


Рис. 9.7

где m — масса космонавта, ω — угловая скорость его вращения вместе с кораблем вокруг Земли и \mathbf{R} — его радиус-вектор относительно центра Земли. Предположим, что корабль летит по круговой орбите. Это означает, что летящий вместе с кораблем космонавт испытывает в системе, связанной с Землей, центростремительное ускорение $\omega^2 R$, направленное к центру Земли и связанное с силой притяжения к Земле \mathbf{F}_g вторым законом Ньютона (если считать Землю инерциальной системой)

$$m\omega^2 R = F_g.$$

Сравнивая это равенство с (9.17), мы видим, что сила инерции равна по величине и противоположна по направлению силе притяжения космонавта к Земле. То есть в этом случае, как и в примере с лифтом, возникает состояние невесомости. Спутник свободно падает на Землю, но благодаря большой начальной скорости «никак не может на нее упасть» и в результате движется по окружности.

Вопросы и задачи

1. Из каких предположений о свойствах пространства и времени следует принцип относительности Галилея?

2. Что такое силы инерции и какие типы сил инерции вы знаете?

3. Почему в южном полушарии левые берега рек подмыты сильнее правых?

4. Что такое состояние невесомости?

5. При спуске на 1-й этаж лифт тормозит, причем модуль его ускорения $a = 3 \text{ м/с}^2$. Определите вес находящегося в лифте человека, если его масса равна $m = 70 \text{ кг}$.

Ответ: $P = m(g + a) \approx 900 \text{ Н}$.

6. Кабины на колесе обозрения движутся в вертикальной плоскости по кругу радиуса $R = 20 \text{ м}$, делая один оборот за время $T = 5 \text{ мин}$ с постоянной

угловой скоростью. Определите величину и направление силы инерции, действующей на человека массы $m = 70$ кг, находящегося в кабине.

Ответ: Сила инерции $F_{\text{ин}} = m(2\pi/T)^2 R \approx 0,6$ Н направлена от центра окружности, по которой движется кабина. Говоря строже, в каждой точке кабины сила направлена к центру именно той окружности, по которой движется данная точка (рис. 9.8). Отметим, что поле сил в пределах каждой кабины в каждый данный момент строго однородно: кабина движется поступательно.

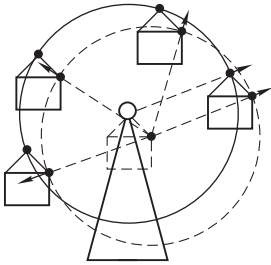


Рис. 9.8

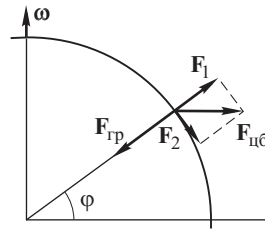


Рис. 9.9

7. Оцените, насколько вес тела в Москве (на широте $\varphi = 56^\circ$) меньше силы гравитационного притяжения к Земле.

Решение. На неподвижное относительно Земли тело действуют две силы — сила гравитации $F_{\text{гр}}$, в основном определяющая вес тела, и относительно малая центробежная сила

$$F_{\text{цб}} = m\omega^2 R \cos \varphi,$$

где ω — угловая скорость вращения Земли, а $R \approx 6400$ км — ее радиус). Основной вклад в изменение веса тела вносит (рис. 9.9) составляющая

$$F_1 = F_{\text{цб}} \cos \varphi = m\omega^2 R \cos^2 \varphi.$$

Относительное изменение веса, связанное с вращением Земли, равно

$$\frac{F_1}{F_{\text{гр}}} = \frac{m\omega^2 R \cos^2 \varphi}{mg} \sim 10^{-3}.$$

Сила F_2 приводит к отклонению отвеса от направления к центру Земли, почти не меняя веса тела.

8. Человек намеревается пройти от центра карусели до ее края по радиусу с постоянной скоростью $v = 3$ м/с. Карусель радиуса $R = 10$ м делает $n = 3$ оборота в минуту. При каком значении коэффициента трения α человек сможет осуществить свое намерение?

$$\text{Ответ: } \alpha \geq \sqrt{(\omega^2 R)^2 + (2v\omega)^2} / g \approx 0,22 \quad (\omega = 2\pi n/60 \text{ с}^{-1}).$$

9. Тело падает с высоты $h = 100$ м без начальной скорости. В какую сторону и на какое расстояние оно сместится под действием силы Кориолиса на широте Москвы?

Решение. Скорость падения по вертикали растет по закону $v = gt$. Соответственно кориолисово ускорение равно $a = 2v\omega \cos \varphi = 2gt\omega \cos \varphi$. Дважды интегрируя по времени кориолисово ускорение, получаем величину сноса $s = g\omega \cos \varphi (T^3/3)$. Полное время падения $T = (2h/g)^{1/2}$. В результате получаем снос к востоку:

$$s = \frac{(2h/g)^{3/2} g \omega \cos \varphi}{3} \approx 0,012 \text{ м}.$$

Глава 10

ВВЕДЕНИЕ В РЕЛЯТИВИСТСКУЮ МЕХАНИКУ

10.1. Постоянство скорости света для всех систем отсчета. Принцип относительности Эйнштейна и преобразования Лоренца

Одной из важнейших физических постоянных является скорость света в вакууме c , то есть скорость распространения электромагнитных волн в свободном от вещества пространстве. Эта скорость не зависит от частоты электромагнитных волн, и принятое сейчас ее значение равно $c = 299\,792\,458$ м/с. В громадном большинстве случаев эту величину c достаточной точностью можно принять равной $c = 3 \cdot 10^8$ м/с — погрешность при этом менее 0,001. И именно «триста тысяч километров в секунду» для скорости света запоминается большинством из нас на всю жизнь. Напомним, что 300 000 км — это, по порядку величины, расстояние от Земли до Луны (точнее, 380 000 км). Таким образом, радиосигнал с Земли достигает Луны через время немного большее, чем одна секунда.

Предположение о том, что свет распространяется не с бесконечной, а с конечной скоростью, высказывались за много столетий до того, как люди смогли доказать это экспериментально. Впервые это было сделано в XVII веке, когда астрономические наблюдения странных «нерегулярностей» в движении спутника Юпитера Ио удалось объяснить только на основе предположения о конечной скорости распространения света (кстати, эта первая попытка определить скорость света дала заниженный результат, $c \approx 214\,300$ км/с).

Вплоть до конца XIX столетия скорость света интересовала исследователей, главным образом, с точки зрения понимания природы электромагнитного излучения — физикам тогда было не ясно, могут ли электромагнитные волны распространяться в вакууме, или они распространяются в особой заполняющей пространство субстанции — эфире. Однако итогом исследования этой проблемы явилось открытие, перевернувшее все существовавшие до тех пор представления о пространстве и времени. В 1881 г. в результате знаменитых опытов американского ученого А. Майкельсона (1852–1931) был установлен удивительный факт — величина скорости света не зависит от того, относительно какой системы отсчета она определяется!

Этот опытный факт противоречит закону сложения скоростей Галилея, который мы рассматривали в предыдущей главе и который кажется очевидным и подтверждается нашими повседневными наблюдениями. Но свет не подчиняется этому естественному, казалось бы, правилу сложения скоростей — относительно всех наблюдателей, как бы они ни двигались, свет распространяется с одной и той же скоростью $c = 299\,793\text{ км/с}$. И то, что распространение света — это движение электромагнитного поля, а не частиц, состоящих из атомов, не играет здесь роли. При выводе закона сложения скоростей (9.2) не имела значения природа движущегося объекта.

И хотя невозможно отыскать что-либо подобное в накопленных нами ранее опыте и знаниях, тем не менее, мы должны признать этот опытный факт, помня, что именно опыт является решающим критерием истины. Вспомним, что мы сталкивались с подобной ситуацией в самом начале курса, когда обсуждали свойства пространства. Тогда мы отмечали, что представить себе кривизну трехмерного пространства нам — трехмерным существам — невозможно. Но мы поняли, что факт «наличия или отсутствия» кривизны можно установить опытным путем: измерять, например, сумму углов треугольника.

Какие же изменения необходимо внести в наше понимание свойств пространства и времени? И как в свете этих фактов относиться к преобразованиям Галилея? Можно ли их изменить так, чтобы они по-прежнему не противоречили здравому смыслу при их применении к привычным движениям окружающих нас тел и в то же время не противоречили факту постоянства скорости света во всех системах отсчета?

Принципиальное решение этих вопросов принадлежит Альберту Эйнштейну, создавшему в начале XX века *специальную теорию относительности* (СТО), связавшую необычный характер распространения света с фундаментальными свойствами пространства и времени, проявляющимися при движениях со скоростями, сравнимыми со скоростью света. В современной физической литературе ее чаще называют просто релятивистской механикой. Впоследствии Эйнштейн построил общую теорию относительности (ОТО), где исследуется связь свойств пространства и времени с гравитационными взаимодействиями.

Основу СТО составляют два постулата, которые носят название *принципа относительности Эйнштейна* и *принципа постоянства скорости света*.

Принцип относительности Эйнштейна является обобщением принципа относительности Галилея, рассмотренного в предыдущей главе, на все без исключения (а не только механические) явления природы. Согласно этому принципу все законы природы одинаковы во всех инерциальных системах отсчета. Принцип

относительности Эйнштейна можно сформулировать следующим образом: *все уравнения, выражающие законы природы, инвариантны по отношению к преобразованиям координат и времени от одной инерциальной системы отсчета к другой.* (Напомним, что инвариантностью уравнений называется неизменность их вида при замене в них координат и времени одной системы отсчета координатами и временем другой.) Понятно, что в соответствии с эйнштейновым принципом относительности никакими вообще опытами нельзя установить, движется «наша» система отсчета с постоянной скоростью или она неподвижна, точнее говоря, между этими состояниями нет никакого различия. Галилей эту невозможность постулировал в принципе только для механических опытов.

Принцип постоянства (точнее, инвариантности) скорости света утверждает, что *скорость света в пустоте одинакова для всех инерциальных систем отсчета.* Как мы вскоре убедимся, из этого следует, что c — максимальная из всех возможных физических скоростей.

Оба постулата являются отражением *опытных фактов*: скорость света не зависит от движения источника или приемника; она не зависит также от движения системы отсчета, в которой производятся эксперименты по ее измерению. В принципе относительности это отражено в признании того факта, что не только механические, но и электромагнитные (распространение света) явления, подчиняются во всех инерциальных системах отсчета одним и тем же законам.

Из сформулированных выше положений вытекает ряд важных выводов, касающихся свойств пространства и времени. Прежде всего, из них следуют новые правила перехода от одной инерциальной системы отсчета к другой, в рамках которых «очевидные» преобразования Галилея являются лишь некоторым частным случаем, реализуемым только при движениях со скоростями, много меньшими c . Для определения этих новых правил рассмотрим свет, распространяющийся от точечного источника, расположенного в начале неподвижной системы отсчета K (рис. 10.1 *а*).

Распространение света можно представить как распространение светового фронта, имеющего форму сферической поверхности в системе отсчета, относительно которой источник света непо-

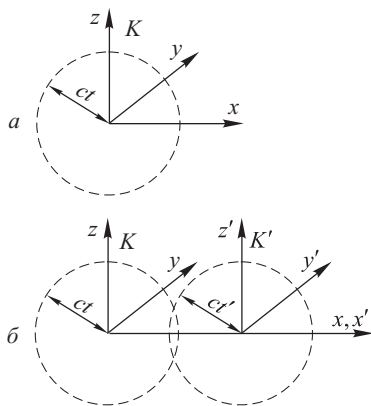


Рис. 10.1

движен. Но согласно принципу относительности Эйнштейна световой фронт должен быть сферическим также и тогда, когда он наблюдается в системе отсчета, находящейся в равномерном и прямолинейном движении относительно источника. Из этого условия мы и определим сейчас, каковы должны быть правила преобразования координат и времени при переходе от одной инерциальной системы к другой.

Если источник света находится в начале координат системы отсчета K , то для света, испускаемого в момент $t = 0$, уравнение сферического светового фронта имеет вид

$$x^2 + y^2 + z^2 = c^2 t^2. \quad (10.1)$$

Это уравнение описывает сферическую поверхность, радиус которой $R = ct$ увеличивается во времени со скоростью c .

Координаты и время, измеряемые наблюдателем в движущейся системе отсчета K' , обозначим буквами со штрихами: x' , y' , z' , t' . Положим, что начало отсчета времени t' совпадает с началом отсчета t и что в этот совпадающий нулевой момент времени начало координат системы K' совпадает с положением источника света в системе K . Пусть, для определенности, система K' движется в направлении $+x$ с постоянной скоростью V относительно системы K (рис. 10.1 б).

Как мы уже говорили, согласно второму постулату Эйнштейна, для наблюдателя в «штрихованной» системе световой фронт должен быть также сферическим, то есть уравнение светового фронта в движущейся системе должно иметь вид

$$x'^2 + y'^2 + z'^2 = c^2 t'^2, \quad (10.2)$$

причем величина скорости света c здесь та же, что и в системе отсчета K . Таким образом, преобразования координат и времени от одной нашей системы отсчета к другой обязаны обладать таким свойством, что, например, после замены с помощью этих преобразований в (10.2) «штрихованных» величин на «не штрихованные» мы должны вновь получить уравнение сферического фронта (10.1).

Легко убедиться, что преобразования Галилея (9.3) не удовлетворяют этому требованию. Напомним, что эти преобразования связывают координаты и время в двух разных системах отсчета следующими соотношениями:

$$x' = x - Vt, \quad y' = y, \quad z' = z, \quad t' = t. \quad (10.3)$$

Если мы подставим (10.3) в (10.2), то получим

$$x^2 - 2xVt + V^2t^2 + y^2 + z^2 = c^2t^2, \quad (10.4)$$

что, конечно, не согласуется с уравнением (10.1). Какими же должны быть новые преобразования? Во-первых, так как все си-

системы равноправны, переход из некоторой системы в любую другую должен описываться одними и теми же формулами (со своим значением V), а двукратное применение преобразований с заменой на втором шаге $+V$ на $-V$ должно возвращать нас в исходную систему. Таким свойством могут обладать только линейные по x и t преобразования. Бесполезно испытывать для этого соотношения типа $x' = x^{1/2}t^{1/2}$, $x' = \sin x$ или им подобные.

Во-вторых, при $V/c \rightarrow 0$ эти преобразования должны переходить в преобразования Галилея, справедливость которых для малых скоростей не может быть подвергнута сомнению.

Из уравнения (10.4) ясно видно, что мы не можем оставить без изменения преобразование $t' = t$, если хотим уничтожить в этом уравнении нежелательные слагаемые $-2xVt + V^2t^2$, потому что для их уничтожения необходимо обязательно что-то прибавить к t .

Попробуем сначала преобразование вида

$$x' = x - Vt, \quad y' = y, \quad z' = z, \quad t' = t + bx, \quad (10.5)$$

где b — постоянная, значение которой надо определить. Тогда уравнение (10.2) принимает вид

$$x^2 - 2Vxt + V^2t^2 + y^2 + z^2 = c^2t^2 + 2c^2bxt + c^2b^2x^2. \quad (10.6)$$

Заметим, что члены в левой и правой частях равенства, содержащие произведение xt , взаимно уничтожаются, если принять

$$b = -\frac{V}{c^2}, \quad \text{или} \quad t' = t - \frac{Vx}{c^2}. \quad (10.7)$$

При этом значении b уравнение (10.6) можно переписать следующим образом:

$$x^2 \left(1 - \frac{V^2}{c^2}\right) + y^2 + z^2 = c^2t^2 \left(1 - \frac{V^2}{c^2}\right). \quad (10.8)$$

Это уже ближе к уравнению (10.1), но еще остается нежелательный множитель $1 - V^2/c^2$, на который умножаются x^2 и t^2 .

Мы можем исключить и этот множитель, если окончательно запишем преобразование координат и времени в следующем виде:

$$x' = \frac{x - Vt}{(1 - V^2/c^2)^{1/2}}, \quad y' = y, \quad z' = z, \quad t' = \frac{t - (V/c^2)x}{(1 - V^2/c^2)^{1/2}}. \quad (10.9)$$

Это и есть знаменитые *преобразования Лоренца*, названные по имени голландского физика-теоретика Х.А. Лоренца (1853–1928), который в 1904 г. вывел формулы (10.9) и тем самым подготовил переход к теории относительности.

Нетрудно проверить, что при подстановке (10.9) в уравнение (10.2) преобразования Лоренца, как и должно быть, преобра-

зуют это уравнение в уравнение сферической поверхности (10.1) в неподвижной системе координат. Также легко убедиться, что при $V/c \rightarrow 0$ преобразования Лоренца переходят в преобразования Галилея (9.2).

10.2. Следствия из преобразований Лоренца. Сокращение длины и замедление времени

Из преобразований Лоренца вытекает ряд необычных с точки зрения ньютоновой механики следствий.

Длина тел в разных системах отсчета. Рассмотрим стержень, расположенный вдоль оси x и покоящийся относительно системы отсчета K' (рис. 10.2). Длина его в этой системе равна $l_0 = x'_2 - x'_1$, где x'_1 и x'_2 — не изменяющиеся со временем t'

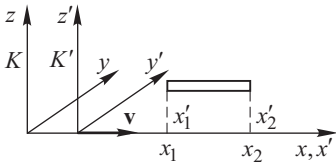


Рис. 10.2

координаты концов стержня. Относительно системы K стержень движется вместе со штрихованной системой со скоростью v . Для определения его длины в этой системе нужно отметить координаты концов стержня x_1 и x_2 в один и тот же момент времени $t_1 = t_2 = t$. Разность этих координат $l = x_2 - x_1$

даст длину стержня, измеренную в системе K . Чтобы найти соотношение между l_0 и l , следует взять ту из формул преобразований Лоренца, которая содержит x', x и t , то есть первую из формул (10.9). Согласно этой формуле,

$$x'_1 = \frac{x_1 - vt}{\sqrt{1 - v^2/c^2}}, \quad x'_2 = \frac{x_2 - vt}{\sqrt{1 - v^2/c^2}},$$

откуда получаем

$$x'_2 - x'_1 = \frac{x_2 - x_1}{\sqrt{1 - v^2/c^2}}$$

или окончательно

$$l = l_0 \sqrt{1 - \frac{v^2}{c^2}}. \quad (10.10)$$

Таким образом, длина стержня l , измеренная в системе, относительно которой он движется, оказывается меньше «собственной» длины l_0 , измеренной в системе, относительно которой стержень покоится. Поперечные размеры стержня в обеих системах одинаковы. Итак, для неподвижного наблюдателя размеры движущихся тел в направлении их движения сокращаются, и тем больше, чем больше скорость движения.

Длительность процессов в разных системах отсчета. Пусть в некоторой точке, неподвижной относительно движущей-

ся системы K' , происходит какой-то процесс, длящийся время $\Delta t_0 = t'_2 - t'_1$. Это может быть работа какого-либо прибора или механизма, колебание маятника часов, какое-нибудь изменение в свойствах тела и так далее. Началу процесса соответствует в этой системе координата $x'_1 = a$ и момент времени t'_1 , концу — та же самая координата $x'_2 = x'_1 = a$ и момент времени t'_2 . Относительно системы K точка, в которой происходит процесс, перемещается. Согласно формулам (10.9) началу и концу процесса в системе K соответствуют моменты времени

$$t_1 = \frac{t'_1 + (v/c^2)a}{\sqrt{1 - v^2/c^2}}, \quad t_2 = \frac{t'_2 + (v/c^2)a}{\sqrt{1 - v^2/c^2}},$$

откуда получаем

$$t_2 - t_1 = \frac{t'_2 - t'_1}{\sqrt{1 - v^2/c^2}}.$$

Введя обозначения $t_2 - t_1 = \Delta t$, получим окончательно:

$$\Delta t = \frac{\Delta t_0}{\sqrt{1 - v^2/c^2}}. \quad (10.11)$$

В этой формуле Δt_0 — длительность процесса, измеренная по часам в движущейся системе отсчета, где тело, с которым происходит процесс, покоится. Промежуток Δt измерен по часам системы, относительно которой тело движется со скоростью v . Иначе можно сказать, что Δt определено по часам, которые движутся относительно тела со скоростью v . Как следует из (10.11), промежуток времени Δt_0 , измеренный по часам, неподвижным относительно тела, оказывается меньше, чем промежуток времени Δt , измеренный по часам, движущимся относительно тела.

Заметим, что для релятивистских множителей (лоренц-факторов) движущейся со скоростью V системы отсчета и/или движущейся со скоростью v частицы приняты обозначения $\Gamma = 1/\sqrt{1 - V^2/c^2}$ и соответственно $\gamma = 1/\sqrt{1 - v^2/c^2}$. Если это не приводит к путанице, для обеих величин употребляется обозначение γ .

Рассматривая протекание процесса из системы K , можно определить Δt как его длительность, измеренную по неподвижным часам, а Δt_0 — как длительность, измеренную по часам, движущимся со скоростью v . Согласно (10.11) $\Delta t_0 < \Delta t$, поэтому можно сказать, что *движущиеся часы идут медленнее, чем покоящиеся часы* (имеется, конечно, в виду, что во всем, кроме скорости движения, часы совершенно идентичны).

Время Δt_0 , отсчитанное по часам, движущимся вместе с телом, называется «собственным временем» этого тела. Как видно

из (10.11), собственное время всегда меньше, чем время, отсчитанное по часам, движущимся относительно тела.

Эффект замедления времени симметричен по отношению к обоим рассматриваемым часам: для обоих наблюдателей из разных систем отсчета часы движущегося относительно него наблюдателя будут идти медленнее. Замедление времени является объективным следствием преобразований Лоренца, которые, в свою очередь, являются следствием постоянства скорости света во всех системах отсчета. Необходимо подчеркнуть то обстоятельство, что релятивистские эффекты отнюдь не умозрительны. На сегодняшний день СТО с очень хорошей точностью подтверждена экспериментально. Разумеется, при $V/c \rightarrow 0$ формулы (10.10), (10.11) преобразуются к тривиальному нерелятивистскому пределу. Для наблюдения нетривиальных эффектов необходимо исследовать объекты с $V \sim c$.

Примерами могут служить явления, наблюдаемые при изучении элементарных частиц. Одним из наиболее наглядных опытов, подтверждающих соотношение (10.11), является наблюдение в составе космических лучей одного из видов элементарных частиц, именуемых мюонами. Эти частицы нестабильны — они самопроизвольно распадаются на другие элементарные частицы. Время жизни мюонов, измеренное в условиях, когда они неподвижны (или движутся с малой скоростью), равно примерно $2 \cdot 10^{-6}$ с. Казалось бы, даже двигаясь почти со скоростью света, мюоны могут пройти от момента своего рождения до момента распада лишь путь, равный примерно $(3 \cdot 10^8 \text{ м/с})(2 \cdot 10^{-6} \text{ с}) = 600 \text{ м}$. Однако наблюдения показывают, что мюоны, образуясь в космических лучах в верхних слоях атмосферы на высоте 20–30 км, успевают, тем не менее, в большом количестве достигнуть земной поверхности. Это объясняется тем, что $2 \cdot 10^{-6}$ с — собственное время жизни мюона, то есть время, измеренное по часам, которые бы «двигались вместе с ним». Время, отсчитанное по часам экспериментатора, связанного с поверхностью Земли, оказывается гораздо большим из-за того, что скорость мюонов близка к скорости света. Поэтому не удивительно, что экспериментатор наблюдает пробег мюона, значительно превышающий 600 м. Интересно рассмотреть этот эффект с точки зрения наблюдателя, «движущегося вместе с мюоном». Для него расстояние, пролетаемое до поверхности Земли, сокращается до 600 м в соответствии с формулой (10.10), так что мюон успевает пролететь его за $2 \cdot 10^{-6}$ с, т. е. за «собственное время жизни».

Наиболее впечатляющее следствие преобразований Лоренца — *относительность одновременности* разнесенных в пространстве событий. Если два события A и B произошли одновременно в одной точке пространства, то в любой системе

координат $t_A = t_B$. Конкретные значения, например, t_A и t'_A могут быть различными, но в каждой системе останется справедливым равенство $t'_A = t'_B$. Если же при $t_A = t_B$ окажется, что $x_A \neq x_B$, то в любой другой системе, как это с очевидностью следует из преобразований Лоренца, $t_A \neq t_B$.

Почему это обстоятельство до Эйнштейна оставалось незамеченным? До Эйнштейна явно или неявно сохранялось представление о существовании абсолютного пространства и абсолютного времени. Но если нет абсолютной системы отсчета, нет и абсолютной одновременности. Исчезает не только абсолютное пространство, исчезает и абсолютное время, которое, по Ньютону, течет «всегда одинаково, безотносительно к чему-либо внешнему». Время СТО зависит от системы отсчета. Зависит от системы отсчета и промежуток времени между двумя событиями, и расстояние между двумя точками. В механике Галилея–Ньютона координаты точек зависят от системы отсчета, но расстояние между точками A и B :

$$(x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2 = l^2$$

от системы не зависит. В механике СТО эта величина перестает быть инвариантом. Независимым от системы отсчета становится *интервал* между событиями, определяемый соотношением

$$s_{AB}^2 = c^2(t_A - t_B)^2 - (x_A - x_B)^2 - (y_A - y_B)^2 - (z_A - z_B)^2.$$

Время становится в один ряд с пространственными координатами или, как сказал немецкий математик и физик Г. Минковский (1864–1909): «пространство само по себе и время само по себе погружаются в реку забвения, а остается жить лишь своеобразный их союз». Это проявляется особенно наглядно, если, следуя Минковскому, в качестве четвертой координаты выбрать не t , как таковое, а ict . Тогда интервал запишется в симметричной форме:

$$s^2 = - \sum_{i=1}^4 \Delta x_i^2,$$

где $x_1 = x$, $x_2 = y$, $x_3 = z$, $x_4 = ict$. Не следует, однако, воспринимать четырехмерное пространство Минковского как простой аналог нашего трехмерного мира. Все же четвертая координата сохраняет важнейшее отличие от трех остальных — *однонаправленность*, которой, в частности, обусловлены причинно-следственные связи. Путешествие вспять во времени как было, так и остается невозможным.

Ввиду того, что по Лоренцу, в отличие от Галилея, преобразуется, кроме координат, и время, заметно меняется закон сложения скоростей. Если в системе K тело движется со скоростью v ,

имеющей составляющие по осям координат v_x , v_y , v_z , а система K' движется со скоростью V вдоль оси x , для составляющих скорости тела в системе K' получаем

$$v'_x = \frac{dx'}{dt'} = \frac{\Gamma(dx - Vdt)}{\Gamma[dt - (V/c)^2 dx]} = \frac{v_x - V}{1 - Vv_x/c^2}; \quad (10.12)$$

$$v'_y = \frac{dy'}{dt'} = \frac{dy}{\Gamma[dt - (V/c)^2 dx]} = \frac{v_y \sqrt{1 - V^2/c^2}}{1 - Vv_x/c^2}; \quad (10.13)$$

$$v'_z = \frac{dz'}{dt'} = \frac{dz}{\Gamma[dt - (V/c)^2 dx]} = \frac{v_z \sqrt{1 - V^2/c^2}}{1 - Vv_x/c^2}. \quad (10.14)$$

Хотя координаты y' и z' равны соответственно y и z , составляющие скорости по этим осям в разных системах различны, так как различаются темпы течения времени.

Не представляется неожиданным факт, что если v_x по модулю равна скорости света $-c$, то эта величина не изменится при переходе в любую другую систему отсчета. Ведь именно инвариантность скорости света является критерием справедливости преобразований Лоренца.

10.3. Импульс и энергия в релятивистской механике

Закон сохранения импульса — один из основных законов природы; он отражает однородность пространства. В мире Минковского однородность пространства сохраняется. Видимо, должен быть по-прежнему справедлив закон сохранения импульса в замкнутой системе. Однако это требует определенной корректировки выражения для импульса тела.

Рассмотрим частный случай. Два одинаковых тела (масса каждого равна m) движутся навстречу друг другу, имея, с точки зрения некоторой системы отсчета K , равные скорости. Следовательно, $v_{xA} = -v_{xB}$, и $v_{yA} = -v_{yB}$. Пусть при столкновении тела объединяются в единое целое («слипаются»). В соответствии с законом сохранения импульса образовавшееся тело должно покоиться. В частности, его координата y теперь должна быть неизменной. В любой другой системе K' , события в которой связаны с событиями в системе K преобразованиями (10.9), координата y' тоже должна стать постоянной.

Рассмотрим ситуацию до столкновения в системе K' . Согласно формулам преобразования скоростей (10.13), получаем

$$v'_{yB} = \frac{v_{yB} \sqrt{1 - V^2/c^2}}{1 - Vv_{xB}/c^2}; \quad v'_{yA} = \frac{v_{yA} \sqrt{1 - V^2/c^2}}{1 - Vv_{xA}/c^2} = -\frac{v_{yB} \sqrt{1 - V^2/c^2}}{1 + Vv_{xB}/c^2}.$$

Модули y -составляющих скоростей теперь оказываются различными, и если сохранить классическое выражение для импульса, в частности, $p_y = mv_y$, суммарная y -составляющая импульса двух частиц не равна нулю. Тогда после «слипания» образовавшееся тело должно иметь отличную от нуля составляющую скорости по оси y . Видимо, надо изменять выражение для импульса тела.

Для того чтобы догадаться, как именно надо менять это выражение, рассмотрим следующую процедуру. Установим на каждом из тел часы. Они будут показывать собственное время данного тела. Теперь будем фиксировать координату тела y в различных системах и отмечать собственное время t_0 , которое соответствует этому значению y . Запишем «гибридную» скорость $v_y^* = dy/dt_0$. Так как координата y не преобразуется при переходе от системы к системе, а время мы фиксируем по собственным часам тела, во всех системах мы должны получить одно и то же значение v_y^* . Мы знаем, что промежутки времени по часам системы, в которой тело движется (dt), и промежутки, отсчитанные по собственным часам (dt_0), связаны соотношением (10.11). Применительно к нашему случаю — $dt_0 = dt/\gamma$.

Следовательно, во всех системах величина $\gamma dy/dt = dy/dt_0$ будет иметь одно и то же значение. Ясно, что равенство не нарушится, если обе его части умножить на одну и ту же величину m .

Применительно к нашей задаче (о «слипании» двух тел) это означает, что комбинации $\gamma m \mathbf{v}$ у двух частиц будут равными по модулю в любой системе отсчета.

Естественно именно эту величину принять за y -составляющую импульса. Обобщая, определим релятивистский импульс выражением

$$\mathbf{p} = \gamma m \mathbf{v} = \frac{m \mathbf{v}}{\sqrt{1 - v^2/c^2}}, \quad (10.15)$$

где \mathbf{v} — скорость тела в системе, в которой мы вычисляем импульс.

Опыт показывает, что определенный таким образом импульс обладает основными свойствами, присущими величине $m \mathbf{v}$ в механике Ньютона: он сохраняется в замкнутых системах тел, а скорость его изменения равна силе, действующей на тело. Силу при этом следует определять в соответствии с (5.25): $\mathbf{F} = -\Delta U$. При $v/c \ll 1$ выражение (10.15) переходит в обычное нерелятивистское определение импульса. Тем самым соблюдается *принцип соответствия*, за выполнением которого следует внимательно следить при любом расширении физических понятий. (Мы уже касались этого вопроса в связи с законом сохранения энергии.)

Для работы силы, то есть для изменения кинетической энергии тела, сохраняется выражение

$$\delta A = dT = \mathbf{F} \cdot \delta \mathbf{r} = \mathbf{F} \mathbf{v} dt, \quad \text{где} \quad \mathbf{F} = \dot{\mathbf{p}}.$$

Ограничиваясь случаем действия силы вдоль направления скорости, с учетом (10.15) получаем

$$dT = m\gamma^3 v a \cdot dt = m\gamma^3 v dv.$$

Тогда для кинетической энергии имеем

$$T = \int_0^v \frac{m d(v)^2}{2(1 - v^2/c^2)^{3/2}} = \frac{mc^2}{\sqrt{1 - v^2/c^2}} - mc^2 = (\gamma - 1)mc^2. \quad (10.16)$$

Выражение (10.16) трактуется так: энергия тела $E = \gamma mc^2$ складывается из кинетической энергии и *энергии покоя* $E_0 = mc^2$.

Закон сохранения энергии справедлив именно для величины E . Это верно как в случае упругих, так и неупругих взаимодействий, когда потери кинетической энергии компенсируются ростом энергии покоя взаимодействующих частиц или энергией частиц, излучаемых при взаимодействии, например, фотонов. В более общем случае, при рождении в процессе взаимодействия массивных частиц, необходимо, конечно, учитывать и кинетическую энергию и энергию покоя всех частиц как до взаимодействия, так и после него. Такие понятия, как потенциальная энергия или энергия связи, в релятивистском представлении включаются в энергию покоя.

Для энергии с помощью формулы (10.15) можно получить соотношение

$$E^2 = (pc)^2 + (mc^2)^2. \quad (10.17)$$

Связь импульса с энергией принимает вид

$$\mathbf{p} = \frac{E\mathbf{v}}{c^2}. \quad (10.18)$$

Введя обозначения $M = mc^2$, $\mathbf{P} = \mathbf{p}c$, $\beta = \mathbf{v}/c$, мы можем записать соотношения между массой, импульсом и энергией в несколько более компактной форме:

$$M^2 + P^2 = E^2; \quad \mathbf{P} = E\beta; \quad \mathbf{P} = m\gamma c^2\beta = M\gamma\beta.$$

Фактически эта форма соответствует неявному введению системы единиц, в которой $c = 1$, а масса, импульс и энергия измеряются в одинаковых единицах; например, в ядерной физике в качестве такой универсальной единицы измерения массы, импульса и энергии обычно используется электронвольт: $1 \text{ эВ} = 1,6 \cdot 10^{-19} \text{ Дж} = 1,6 \cdot 10^{-12} \text{ эрг}$.

Комбинация $E^2 - P^2 = M^2$ очевидным образом является инвариантом, т.е. не зависит от системы отсчета для отдель-

но взятой частицы. Менее очевидно, что такая комбинация — инвариант и для совокупности частиц. Доказательство этого утверждения не представляет особых трудностей, но достаточно громоздко. Между тем, обратим внимание на тот факт, что «тела» состоят из движущихся друг относительно друга (хотя бы в процессе колебаний) атомов, атомы — из ядра и электронов, и даже протоны или пионы — из кварков и антикварков. А в отношении этих, строго говоря, систем или совокупностей частиц мы без малейшего сомнения уверены в инвариантности величины $E^2 - P^2$. Использование этого обстоятельства помогает упростить решение многих задач.

Поскольку определение (10.15) и соотношения (10.16)–(10.18) оказываются в хорошем согласии с экспериментом, мы можем сделать еще один важный вывод: из вещественности физических величин \mathbf{p} и E следует, что *скорость любого физического тела не может превышать скорости света*. Если же частица (фотон, нейтрино) движется именно со скоростью c , из (10.15), (10.16) вытекает, что масса такой частицы (более строго, энергия покоя) должна быть равна нулю, а двигаться она может исключительно со скоростью света. Релятивистская динамика по мере развития дает дополнительные аргументы в пользу таких заключений, но уже сейчас можно сказать, что все они подтверждены прямыми экспериментами.

Вопросы и задачи

1. Что такое принцип относительности Эйнштейна, и какие опытные факты лежат в его основе?

2. Что такое «замедление времени», и из каких законов природы вытекает этот эффект?

3. Чему равна масса фотона?

4. Два космических корабля движутся по одной прямой со скоростью $v = 0,6 c$ друг относительно друга. На каждом корабле наблюдают прямое телевизионное изображение часов другого корабля. Как связаны показания собственных часов и показания изображения «чужих» часов?

Решение. Допустим, наблюдатель A в некоторый момент времени поставил на своих часах то же показание, которое он видел на телеэкране, например, t_0 . Отметим, что передатчик B послал изображение своих часов, на которых было это показание — t_0 , с некоторого расстояния L_0 , и мы на часах A установили время t_0 через промежуток времени L_0/c после того, как это время показали часы B .

Пока по часам B пройдет некоторое время τ_0 , по «неподвижным» часам системы A пройдет уже время $\tau = \gamma \tau_0$. Однако надо учесть, что корабль B движется, и его сигналы приходят из разных точек системы A . Если корабли сближаются, путь сигнала во втором случае меньше на величину $v\tau$, т.е. $L = L_0 - v\tau$. Через время $L/c = L_0/c - v\tau/c$ после того, как часы B показали $t_0 + \tau_0$, сигнал об этом дойдет до наблюдателя A .

Итак, наблюдатель A увидит на телевизионном экране показание $t_0 + \tau_0$ тогда, когда его часы покажут время

$$t_0 + \tau - \frac{v\tau}{c} = t_0 + \gamma \tau_0 \left(1 - \frac{v}{c}\right) = t_0 + \tau_0 \sqrt{\left(1 - \frac{v}{c}\right) \left(1 + \frac{v}{c}\right)} = t_0 + 0,5 \tau_0.$$

На телеэкране прошла минута, а по часам A — только полминуты. Часы на телевизоре идут *быстрее*, чем часы в кабине корабля (!). Точно такие же рассуждения применимы, конечно, и к космонавту B : у него тоже часы на экране, изображающем корабль A , будут идти вдвое быстрее собственных. После того, как космонавты разминутись, часы на экране пойдут медленнее часов в кабине. Нетрудно понять, что теперь промежутки времени будут связаны соотношением

$$\tau = \tau_0 \sqrt{\left(1 - \frac{v}{c}\right) \left(1 + \frac{v}{c}\right)}.$$

В нашем случае часы на экране пойдут вдвое медленнее собственных. Отметим, что это отношение не равно γ и в случае удаления кораблей друг от друга.

5. Стержень пролетает с большой скоростью мимо метки, расположенной в лабораторной системе отсчета K . Известно, что время прохождения стержня мимо метки равно $t = 3$ нс по часам системы K и $t = 5$ нс по часам системы K' , связанной со стержнем. Определить собственную длину стержня (т.е. его длину в системе K').

Решение. Рассмотрим два события — совмещение с меткой переднего и заднего концов стержня — в двух системах координат. В системе K эти события происходят в одной точке, и интервал между ними определяется равенством $s^2 = (ct)^2$. В системе K' расстояние между точками, в которых происходят эти события, как раз и есть собственная длина стержня, т.е. $(s')^2 = (ct')^2 - L_0^2$. В силу инвариантности интервала $s^2 = (s')^2$. И окончательно получаем $L_0 \sqrt{(ct')^2 - (ct)^2} = 1,2$ м.

6. Стержень, собственная длина которого равна $L_0 = 240$ см, пролетает с большой скоростью мимо метки, расположенной в лабораторной системе отсчета K . Время прохождения по лабораторным часам равно $t = 6$ нс. Какое время прохождения метки зафиксировано по часам, расположенным на стержне?

Ответ: $t_0 = \sqrt{t^2 + (L_0/c)^2} = 10$ нс.

7. При соударении достаточно энергичных протонов может идти реакция рождения протон-антипротонной пары $p + p \rightarrow p + p + p + \bar{p}$. На этом примере определите выигрыш от использования реакции на встречных пучках.

Решение. Если осуществляется реакция на встречных пучках, когда лабораторная система является системой центра инерции, т.е. системой, в которой суммарный импульс частиц равен нулю, вся кинетическая энергия протонов может перейти в энергию покоя пары, т.е. протоны должны быть разогнаны до энергии $T_0 = M$.

Если же разогнанный на ускорителе протон сталкивается с неподвижным, требуется значительно большая кинетическая энергия. Наиболее выгоден случай, когда продукты реакции движутся вместе, что очевидно, если мысленно пересечь в ц-систему (т.е. вернуться к варианту на встречных пучках). В ц-системе инвариант $E^2 - P^2$ равен $16M^2$ ($M_\Sigma = 4M$, $P_\Sigma = 0$).

Значит, он должен иметь это же значение и в лабораторной системе. Но таким же он должен быть и до соударения, когда один из протонов покоился, а второй имел энергию T и импульс

$$P = (E^2 - M^2)^{1/2} = [(T + M)^2 - M^2]^{1/2} = [T(2M + T)]^{1/2}.$$

Получаем $16M^2 = (2M + T)^2 - P^2$, откуда $T = 6M = 6T_0$.

Раздел второй

ЭЛЕКТРИЧЕСТВО И МАГНЕТИЗМ

ВВЕДЕНИЕ

По традиции раздел общей физики, в котором рассматриваются электромагнитные явления, называется «Электричество и магнетизм»; профессионалы предпочитают именовать данную науку электродинамикой. В нашем рассмотрении мы будем исходить из того, что школьный курс физики читателю уже известен. Это не должно отражаться на порядке изложения, поскольку наша цель — представить единый курс, а не избранные главы. Но мы достаточно широко будем пользоваться примерами и аналогиями, выходящими за пределы собственно электродинамики. Тем самым мы в какой-то степени погрешим против исторической последовательности в развитии электродинамики — впрочем, это неизбежное свойство любого учебного курса. Даже законы Ньютона отнюдь не первичны, но представляют собой обобщение большого числа известных ранее экспериментальных фактов.

В этой связи подчеркнем, что электродинамика, как и вся физика, — наука естественная, а следовательно — экспериментальная. Будучи самой фундаментальной из естественных наук, физика, как следствие, и более других формализована, но надо понимать, что аксиом и теорем в строгом смысле в ней нет. Все основные факты берутся из эксперимента, все предположения и выводы экспериментом в конечном счете и проверяются. Это утверждение, конечно, не следует трактовать слишком прямолинейно. Теория и опыт в физике не на каждом шагу контролируют друг друга, но лишь в некоторых принципиальных «узловых» точках, оставляя достаточно места и формальной логике, и математическому моделированию физической реальности. Не следует только путать модель и самую реальность.

В отличие от механики, которая (включая релятивистскую и квантовую механику) представляет движение или равновесие как результат заданного взаимодействия, в электродинамике мы впервые изучаем взаимодействие как таковое. Фундаментальных взаимодействий в физике всего четыре: сильное, слабое, гравитационное и электромагнитное. Первые два — предмет ядерной физики, и, по крайней мере, явно не проявляются в том круге явлений, который мы будем изучать. Гравитационное взаимодействие обычно широко используется для иллюстрации

законов ньютоновской механики. Из всех перечисленных оно оказывается самым слабым, но для макроскопических (особенно космических) масштабов остается единственно эффективным. Однако в тех случаях, когда вступает в игру электромагнитное взаимодействие, гравитацию, как правило, можно исключить из рассмотрения.

Мы постоянно сталкиваемся с электромагнитным взаимодействием, не всегда отдавая себе в этом отчет. Вся информация о внешнем мире: как пассивная (зрение, слух, обоняние, осязание), так и полученная через посредство любых приборов — обеспечена исключительно электромагнитным взаимодействием. Само наше существование, включая процесс мышления, базируется на химии, а последняя, в свою очередь, на электромагнитных явлениях. Было бы, конечно, недопустимым упрощением пытаться «свести» весь окружающий нас мир к элементарным явлениям электродинамической природы, но они, во всяком случае, играют роль тех кирпичиков, из которых этот мир построен.

Каждой естественной науке, в частности каждой области физики, присущ свой собственный формальный язык. Можно представить себе курс механики, в котором нет ни производных, ни интегралов. Читатель не затруднится с примером — это, скажем, школьный курс. Но имея опыт с разделом «Механика» нашего курса, читатель согласится, по-видимому, и с тем, что из-за такой формальной уступки уровню аудитории школьный курс механики много теряет по существу. Равным образом, например, естественный язык термодинамики — язык частных производных и криволинейных интегралов. Тут уже школьный курс оказывается в совсем тяжелом положении, и студенты ощущают в этой связи особую трудность термодинамики.

У электродинамики есть свой «родной язык» — это язык векторного анализа. Некоторые учебники без него обходятся, а затем он неожиданно возникает в самый последний момент (при формулировке локальной формы уравнений Максвелла) либо отодвигается до лучших времен — куда-нибудь в курс теоретической физики.

По самому принципу построения нашей книги мы не должны отсылать читателя к каким-либо последующим курсам, поскольку у студента технического вуза этот учебник физики может оказаться и последним в жизни. В то же время мы не предполагаем у читателя предварительного знакомства с операциями векторного дифференцирования и будем вводить таковые по ходу изложения. Разумеется, ни объем, ни задача настоящей книги не позволяют сопроводить этот математический материал какими-либо доказательствами; заинтересованному читателю мы порекомендуем обратиться к учебникам математического анализа.

Принципиально важным для любого курса электродинамики оказывается вопрос о системе единиц, которая используется при его построении. Следуя опять программе обучения инженерных вузов, мы представляем настоящий курс в системе СИ, хотя физики ею не всегда пользуются. В механике системы СИ и СГС равноправны, и соответствующие единицы отличаются только масштабом, а все формулы выглядят в обеих системах единиц одинаково. В электродинамике дело обстоит иначе, поскольку в системе СИ, кроме метра, килограмма и секунды, эталонируется еще одна физическая единица — ампер, единица силы тока. Принято это делать исключительно из соображений удобного масштаба практически важных физических величин, но в результате одни и те же законы выглядят в разных системах единиц немного по-разному, отличаясь размерными коэффициентами. И если в системе СГС «лишних» коэффициентов нет, то в СИ возникают понятия с точки зрения физики парадоксальные, как, например, диэлектрическая и магнитная проницаемости вакуума.

Полезно, однако, убедиться в том, что и любимая физиками система СГС оказывается «не без греха». Для начала отметим, что корректным образом любая новая размерность в физике вводится исходя из соотношения, связывающего новую физическую величину с уже известными, скажем, работу с силой и расстоянием. Но в традиционном курсе физики особняком стоит второй закон Ньютона, поскольку в него входят сразу две новые величины — сила и масса. Мало того, основное утверждение — пропорциональность ускорения приложенной силе — по существу бессодержательно, поскольку сила как физическая величина здесь же и определяется. Ничто, казалось бы, не мешает определить ее иначе, например, считать пропорциональной корню квадратному из ускорения.

В действительности, однако, все парадоксы без труда снимаются, если одновременно со вторым законом Ньютона рассматривать ему же принадлежащий закон всемирного тяготения. Тогда вдруг окажется, что три фундаментальных физических единицы нам не нужны — достаточно двух. Из двух зависимостей: $F \propto ma$ и $F_{12} \propto m_1 m_2 / r^2$ следует, что при гравитационном взаимодействии двух одинаковых тел $a \propto m / r^2$. Естественным оказывается определение единицы массы из величины ускорения, скажем, 1 см/с^2 при расстоянии 1 см между взаимодействующими телами. Далее столь же естественным образом из закона тяготения выделяется понятие силы, и второй закон Ньютона будет содержать, как и следует, лишь одну новую физическую величину.

Это, быть может, было бы несколько сложно для школьного курса, но и в учебниках для студентов такой схемой не пользу-

ются. Дело в том, что в этом случае закон всемирного тяготения следовало бы вводить без коэффициента: $F_{12} = m_1 m_2 / r^2$. Определенная отсюда единица массы оказалась бы равной примерно 15 000 кг, а единица силы, даже для базовых единиц сантиметр и секунда, была бы равна примерно 150 Н. Совершенно очевидно, что и далее определенные физические величины оказались бы по масштабу совершенно неприемлемы. Поэтому единица массы все-таки используется в качестве базовой, а платить за это приходится некоторой путаницей в основных законах механики и размерным коэффициентом в законе всемирного тяготения — гравитационной постоянной. Совершенно аналогичная ситуация возникает при использовании системы СИ в электродинамике.

Если бы мы пользовались системой СГС, то единица силы тока оказалась бы равной примерно $3,3 \cdot 10^{-10}$ А, а, к примеру, единица сопротивления — $9 \cdot 10^{11}$ Ом. Поэтому для практических надобностей система СГС была бы совершенно непригодна. Правда, основные физические законы в ней выглядят более простыми и естественными, но уравнения колебательного контура, электрических цепей, многие формулы электроники оказываются проще и удобнее в системе СИ. Не следует, однако, забывать, что ее породило исключительно удобство единиц измерения и практически важных формул, поэтому встречающиеся иногда рассуждения о физическом смысле величин ε_0 и μ_0 заведомо абсурдны.

Любой учебный курс общей физики должен включать, помимо основного текста, семинарские занятия и лабораторные работы, а также лекционные демонстрации. Всего перечисленного наш учебник никоим образом заменить не может, но некоторое количество задач, как и в предыдущем разделе, будет представлено и в курсе электродинамики.

Глава 1

ЭЛЕКТРИЧЕСКОЕ ПОЛЕ В ВАКУУМЕ

1.1. Электрический заряд

Как уже было замечено во введении, новая физическая величина естественно возникает из нового закона, количественно представляющего некоторое физическое явление. Понятие о заряде возникает на основании известного из опыта явления электризации. Исследуя его (поначалу на достаточно примитивном уровне), можно установить, что идентичным образом полученные электризованные тела отталкиваются друг от друга, и далее, что все так или иначе электризованные тела делятся на две группы — и отсюда происходит понятие знака заряда. За формулировкой «одноименные заряды отталкиваются, а разноименные притягиваются» стоит огромный массив информации, накопленный опытным путем и порождающий в общем очень сильные утверждения: что заряды бывают всего двух знаков и что мир наш в целом электрически нейтрален. Уже в конце XIX столетия было ясно, что нейтрален каждый атом в отдельности и что электрический заряд порождается всегда с «нулевой суммой» и притом дискретным образом: любой нескомпенсированный заряд кратен некоторому элементарному — заряду электрона

$$e = 1,60217733 \cdot 10^{-19} \text{ Кл},$$

где $1 \text{ Кл} = 1 \text{ А} \cdot \text{с}$ — 1 кулон — единица заряда (названная по имени французского физика Ш.О. Кулона (1736–1806)). В определение и, соответственно, в размерность кулона входит 1 А — единица тока, которая в системе СИ входит в число базовых и должна эталонироваться. (Вообще, как будет ясно из дальнейшего, электродинамические единицы в системе СИ определяются нередко достаточно причудливым образом, что обусловлено историческими причинами.) То обстоятельство, что именно электрон оказался в конечном счете отрицательно заряженной частицей, а протон — положительно заряженной, является абсолютно случайным, точнее, оно следует из произвольной первоначальной идентификации двух упомянутых выше групп заряженных тел. Но, один раз договорившись, мы должны, во избежание путаницы, безоговорочно соблюдать эту договоренность.

В данной и последующей главах мы будем заниматься исключительно статическими задачами; соответствующий раздел

электродинамики принято называть электростатикой (во избежание терминологической путаницы заметим, что так называемая магнитостатика — не совсем статика, ибо магнитное поле порождается движением зарядов).

Важнейший закон электростатики — закон Кулона — зачастую трактуется в узком смысле, только как закон взаимодействия заряженных тел. На самом деле он гораздо более фундаментален, поскольку вводит само количественное понятие электрического заряда.

Первоначальный термин «количество электричества» правильно отражал историю вопроса. Было установлено, что сила взаимодействия между двумя заряженными телами (неважно, были они одно- или разноименно заряжены) пропорциональна заряду каждого из них, определенному в произвольных единицах просто как сумма одинаковых заряженных тел, сложенных вместе.

Далее Кулон, и независимо — Г. Кавендиш (1731–1810), установили, что сила взаимодействия, безотносительно к знаку заряда, обратно пропорциональна квадрату расстояния между взаимодействующими телами:

$$F_{12} \propto r_{12}^{-2}. \quad (1.1)$$

Исходя из этих экспериментальных фактов и вводится количественная мера электрического заряда. Для этого (1.1) объединяется с эмпирическим понятием количества электричества:

$$F_{12} = \frac{q_1 q_2}{r_{12}^2}. \quad (1.2)$$

Именно так выглядит закон Кулона в системе СГС. Приравняв силу отталкивания двух одинаковых зарядов на расстоянии 1 см одной дине, мы получили бы единицу заряда в системе СГС. Но тут и оказывается, что для практических задач эта величина неудобна; например, таким будет заряд на обкладках конденсатора емкости 1 мкФ при напряжении $\sim 3 \cdot 10^{-4}$ В. В системе СИ, которой мы в дальнейшем будем пользоваться, уравнение (1.2) приобретает числовой коэффициент; его принято записывать в виде

$$F_{12} = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r_{12}^2}, \quad (1.3)$$

где сила F измерена в ньютонах, q — в кулонах, r — в метрах, а электрическая постоянная $\epsilon_0 = 8,85 \cdot 10^{-12}$ Ф/м. С очень хорошей точностью эту константу можно принять равной $10^{-9}/(36\pi)$ Ф/м. Если мы хотим, как обычно принято в статике, придать закону взаимодействия векторную форму, то нам нужно доопределить направление силы. Пусть вектор \mathbf{r}_{12} и, соответственно, орт \mathbf{r}_{12}/r_{12} направлены от заряда 1 к заряду 2.

Тогда в векторной форме закон Кулона приобретает окончательный вид

$$\mathbf{F}_{12} = -\frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r_{12}^3} \mathbf{r}_{12}. \quad (1.4)$$

Формула (1.4) правильно учитывает знак заряда — из нее как раз следует притяжение разноименных и отталкивание одноименных зарядов. Во всех формах закона Кулона мы молчаливо подразумевали, что размеры заряженных тел много меньше расстояния между ними, а распределение зарядов по объему или по поверхности тел достаточно однородно, чтобы не придавать взаимодействию каких-либо анизотропных свойств. Такое приближение называется *приближением точечных зарядов*; во многом оно близко приближению точечной массы в механике.

1.2. Электрическое поле

Выражения (1.3), (1.4) можно переписать в эквивалентных формах:

$$\begin{aligned} F_{12} &= q_1 E(q_2) = q_2 E(q_1); \\ \mathbf{F}_{12} &= q_1 \mathbf{E}(q_2); \quad \mathbf{F}_{21} = q_2 \mathbf{E}(q_1); \quad \mathbf{F}_{12} = -\mathbf{F}_{21}, \end{aligned} \quad (1.5)$$

где векторная величина \mathbf{E} называется напряженностью электрического поля или — в особенности, в современной литературе — просто *электрическим полем*. Возьмем точечный электрический заряд q и поместим начало координат в центре этого заряда. Сравнивая формулы (1.3), (1.4) с (1.5), для поля этого заряда получаем

$$E = \frac{q}{4\pi\epsilon_0 r^2}; \quad \mathbf{E} = \mathbf{r} \frac{q}{4\pi\epsilon_0 r^3}. \quad (1.6)$$

Пока это всего лишь тождественная запись, преобразование, которое в принципе может когда-либо оказаться удобным. Но, как будет следовать из всего нашего последующего курса, именно здесь мы вышли на одно из самых фундаментальных понятий. Поле, подобно веществу, оказывается видом материи. Пространство между электрическими зарядами отнюдь не пусто — оказывается, там, где \mathbf{E} отлично от нуля, в пространстве как-то распределена энергия, а следовательно, и масса. Уравнениями (1.3), (1.4) представлено как бы мгновенное дальнее действие, а на самом деле, именно поле оказывается промежуточным агентом, и когда мы от статической перейдем к временной задаче, предельная скорость взаимодействия окажется равной скорости света.

Единица напряженности поля — 1 вольт на метр (В/м); из (1.5) следует $1 \text{ В/м} = (1 \text{ Н})/(1 \text{ Кл})$. Полезно в этой связи отметить, что логика наименования и связей между единицами в системе СИ не адекватна логике общего курса физики, поэтому нам придется не раз употреблять обозначения, выраженные че-

рез единицы, которых мы еще не вводили. Правда, для вычисления полей в повседневной инженерной практике именно вольты на метры и приходится чаще всего делить.

В качестве принципиально важного обобщения опытных фактов выступает *принцип суперпозиции полей*, в наиболее явной форме выражающий векторную природу и векторные свойства электрического поля. Пусть в пространстве размещены любым способом, дискретным или непрерывным, заряды $q_1, q_2, q_3, \dots, q_n$. Тогда их действие на любой (т. н. пробный) заряд выражается формулой типа (1.5), где электрическое поле равно просто векторной сумме полей всех зарядов:

$$\mathbf{E}_\Sigma = \mathbf{E}_1 + \mathbf{E}_2 + \mathbf{E}_3 + \dots + \mathbf{E}_n.$$

В отличие от механики, где имеет место принцип суперпозиции сил, здесь ситуация даже проще, поскольку силу в механике характеризует, помимо величины и направления, линия приложения, тогда как электрическое поле является просто функцией точки пространства. Это позволяет, в частности, «оторвать» поля от порождающих их зарядов и не только складывать их в любой точке, но и представлять в виде любой векторной суммы, удобной для решения той или иной задачи. В качестве пробного заряда может, в частности, выступать любой из зарядов системы $q_1 \dots q_n$, скажем, q_m . Но в этом случае из векторной суммы должно быть исключено его собственное поле, поскольку самодействия по принципу Мюнхгаузена физика не допускает:

$$\mathbf{E}_\Sigma = \sum_{i \neq m} \mathbf{E}_i.$$

Еще одно важное замечание касается среды, в которой мы можем пользоваться принципом суперпозиции. Он абсолютно точен в вакууме, но лишь в качестве некоторого приближения — в сплошных средах, которые, в частности, будут предметом рассмотрения в следующей главе. Если принцип суперпозиции работает с хорошей точностью, среда называется линейной, если же реакция среды на внешнее поле такова, что линейность отклика нарушается, то не имеет смысла говорить и о принципе суперпозиции полей. Таковы, например, сегнетоэлектрики, а в случае магнитных полей — ферромагнетики.

Из школьного курса физики известно, что графически поле удобно представлять силовыми линиями — кривыми, к которым вектор электрического поля в каждой точке представляет направление касательной. Густота силовых линий (ее можно определить как число линий, пересекающих площадку единичной площади, нормальную к линиям), по определению, считается пропорциональной напряженности поля в данной точке. Силовые линии начинаются на положительных зарядах и кончатся

ся на отрицательных, либо могут уходить на бесконечность. При этом на положительно заряженном теле начинается, а на отрицательно заряженном заканчивается число силовых линий, пропорциональное заряду. Определяют его, например, так: на точечном заряде в 1 Кл начинается (или на нем заканчивается) $1/\varepsilon_0$ силовых линий.

Это, на первый взгляд, искусственное определение, как и все перечисленные свойства, неявным образом отражает чрезвычайно важный эквивалент закона Кулона — теорему Гаусса. Теоремой она является постольку, поскольку математически строго следует из закона Кулона, по смыслу же обладает большей общностью и входит в число основных законов электродинамики.

1.3. Теорема Гаусса

Обратимся к рис. 1.1. Пусть некоторый заряд q окружен гладкой замкнутой поверхностью S . Выделим на ней малый элемент dS . Пусть \mathbf{n} — единичный вектор нормали к dS , а φ — угол между векторами \mathbf{E} и \mathbf{n} . При вычислении некоторых поверхностных интегралов оказывается удобным представить дифференциал поверхности в векторной форме. По определению, вектором элемента площади $d\mathbf{S}$ называется $d\mathbf{S} \cdot \mathbf{n}$, соответственно,

$$(\mathbf{E} d\mathbf{S}) \equiv (\mathbf{E} \mathbf{n}) dS \equiv E \cos \varphi dS.$$

Также по определению, потоком вектора \mathbf{E} через поверхность S называется поверхностный интеграл

$$\oint \mathbf{E} d\mathbf{S} \equiv \int_S (\mathbf{E} \mathbf{n}) dS = \int_S E \cos \varphi dS. \quad (1.7)$$

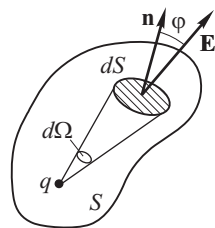


Рис. 1.1

При этом в случае замкнутой поверхности (рис. 1.1) поток в (1.7) будем определять по отношению к внешней нормали (в противном случае поток имел бы ту же величину, но противоположный знак). Для последующих выкладок нам полезно вспомнить еще определение телесного угла. Линейный угол, как известно, определяется отношением элемента дуги к радиусу. Телесный же угол Ω , по аналогии (см. рис. 1.1) определяется отношением площади, нормальной к радиусу-вектору, к квадрату этого радиуса. Удобно определить его через малый элемент:

$$d\Omega = \frac{dS_{\perp}}{r^2} = \frac{(d\mathbf{S} \mathbf{n})}{r^2} = \frac{dS \cos \varphi}{r^2}. \quad (1.8)$$

Полный линейный угол равен отношению длины окружности к радиусу, т. е. 2π , аналогично, полный телесный угол равен

отношению площади поверхности сферы к квадрату радиуса, т. е. 4π .

Теперь подставим первое — скалярное — равенство (1.6) в (1.7) и воспользуемся соотношением (1.8):

$$\oint \mathbf{E} d\mathbf{S} = \frac{q}{4\pi\epsilon_0} \int_S \frac{dS \cos \varphi}{r^2} = \frac{q}{4\pi\epsilon_0} \int d\Omega = \frac{q}{\epsilon_0}. \quad (1.9)$$

Таким образом, поток вектора \mathbf{E} через поверхность, окружающую заряд, равен, с точностью до коэффициента ϵ_0^{-1} , величине этого заряда. Два обстоятельства стоят того, чтобы их отметить. Во-первых, ни симметрия поверхности, ни точка расположения заряда абсолютно не существенны. Во-вторых, источником поля может быть и не один точечный заряд, но любая система зарядов. В этом случае следует руководствоваться принципом суперпозиции. Поток поля любого заряда, расположенного вне замкнутой поверхности, будет равен нулю (удобно иллюстрировать это на языке силовых линий, пронизывающих поверхность), а каждый из потоков от внутренних зарядов будет соответствовать формуле (1.9). Итак, поле может быть произвольной природы, но поток его через замкнутую поверхность задается лишь заключенным в ней суммарным зарядом (разумеется, с учетом знака):

$$\oint \mathbf{E} d\mathbf{S} = \sum_i \frac{q_i}{\epsilon_0}. \quad (1.10)$$

Это и есть *теорема Гаусса* (по имени немецкого математика, астронома и физика К.Ф. Гаусса (1777–1855)). Она, как уже отмечалось, принадлежит к числу основных законов электродинамики; вместе с тем, это хороший инструмент для решения задач электростатики, особенно в случае достаточно высокой симметрии зарядовой системы. Рассмотрим в качестве примера поле одномерного и двумерного «точечного заряда» — соответственно, равномерно заряженной плоскости и однородно заряженной нити (рис. 1.2а, б).

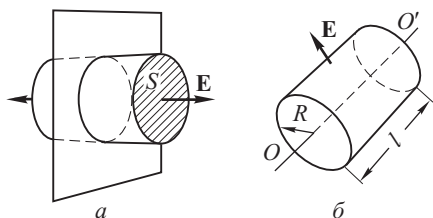


Рис. 1.2

Пусть заряд распределен по плоскости с плотностью $\sigma \equiv dQ/dS$ (рис. 1.2а). Возьмем цилиндрическую поверхность, образуемая которой совпадает с нормалью к плоскости,

Из симметрии задачи следует, что в первом случае вектор \mathbf{E} направлен по нормали к плоскости (причем его направление по отношению к ней одинаково в обоих полупространствах), а во втором — по радиусу (при том, что OO' — ось симметрии).

основание имеет произвольную форму площади S , причем заряженная плоскость пересекается с боковой поверхностью между двумя основаниями цилиндра. Заметим, что поскольку \mathbf{E} перпендикулярно плоскости, то и «густота» силовых линий, т. е. напряженность поля постоянна в каждом полупространстве, а в двух полупространствах, разграниченных заряженной плоскостью, отличается знаком. Поэтому интегральное соотношение (1.10) вырождается в алгебраическое:

$$(\mathbf{S} \mathbf{E}_1) - (\mathbf{S} \mathbf{E}_2) = E \cdot 2S = \varepsilon_0^{-1} \sigma \cdot S,$$

откуда следует:

$$E = \frac{\sigma}{2\varepsilon_0}. \quad (1.11)$$

Обратимся к геометрии рис. 1.26. Пусть $q^* \equiv dQ/dl$ — линейная плотность заряда нити. Окружим последнюю цилиндрической поверхностью длины l и радиуса R . Теперь (1.10) преобразуется к виду

$$E \cdot 2\pi Rl = \varepsilon_0^{-1} q^* l,$$

что дает:

$$E = \frac{q^*}{2\pi\varepsilon_0 R}. \quad (1.12)$$

Выражения (1.11), (1.12) могут быть получены и непосредственно из закона Кулона, но вычисления будут достаточно громоздкими. Для полноты описания представим теорему Гаусса еще и в локальной форме, опираясь не на интегральные соотношения, а на параметры поля в данной точке пространства. Для этого удобно использовать дифференциальный оператор — дивергенцию вектора, — который был введен в гл. 8 раздела «Механика»:

$$\operatorname{div} \mathbf{E} \equiv \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}.$$

Его часто записывают как скалярное произведение оператора векторного дифференцирования ∇ («набла») —

$$\nabla \equiv \mathbf{e}_x \frac{\partial}{\partial x} + \mathbf{e}_y \frac{\partial}{\partial y} + \mathbf{e}_z \frac{\partial}{\partial z}$$

— на векторную функцию:

$$\operatorname{div} \mathbf{E} = \nabla \cdot \mathbf{E}.$$

Здесь $\mathbf{e}_{x,y,z}$ — орты соответствующих осей.

В математическом анализе известна теорема Гаусса–Остроградского: поток вектора через замкнутую поверхность равен интегралу от его дивергенции по объему, ограниченному этой поверхностью, —

$$\oint_S \mathbf{E} \cdot d\mathbf{S} = \int_V \operatorname{div} \mathbf{E} dV. \quad (1.13)$$

Дополним выражение (1.10) понятием пространственной плотности заряда $\rho \equiv dQ/dV$. Тогда, очевидно,

$$\sum_i q_i = \int_V \rho \, dV.$$

Поскольку поверхность интегрирования (или заключенный в ней объем) выбраны произвольно, из (1.10), (1.13) следует равенство подынтегральных выражений:

$$\operatorname{div} \mathbf{E} = \frac{\rho}{\varepsilon_0}. \quad (1.14)$$

Математически уравнение (1.14) эквивалентно (1.10); использование теоремы Гаусса в локальной или интегральной форме зависит от характера задачи.

1.4. Потенциал. Понятие электрической емкости

Чтобы ввести корректно понятие электрического потенциала, мы должны прежде освоить логический прием, связанный с использованием пробного заряда. Предположим, что нам дана некоторая система точечных зарядов q_1, q_2, \dots, q_n , расположенных соответственно в точках $\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_n$. В качестве мысленного эксперимента будем перемещать в поле этих зарядов некий точечный заряд q , малый настолько, что с заранее заданной точностью можно будет пренебречь вызванным им механическим возмущением нашей зарядовой системы. (Иногда говорят об относительной малости его поля. Это некорректное условие, потому что на достаточно малом от заряда расстоянии, в силу закона Кулона, поле его будет как угодно велико, однако на его движении это никак не скажется). Вычислим работу по перемещению пробного заряда из некоторой точки 1 в некоторую точку 2 в предположении, что ни одна точка траектории не совпадает с точками \mathbf{R}_k .

$$A_{12} = - \int_1^2 \mathbf{F} \, d\mathbf{l} = -q \int_1^2 \mathbf{E} \, d\mathbf{l} = -q \sum_{i=1}^n \int_1^2 \mathbf{E}_i \, d\mathbf{l}.$$

Мы воспользовались принципом суперпозиции полей. А теперь рассмотрим один из членов этой суммы и учтем, что, согласно (1.4), $\mathbf{E}_i \parallel \mathbf{r}_i$, где \mathbf{r}_i — радиус-вектор пробного заряда в системе с началом координат в точке \mathbf{R}_i . Поэтому $\mathbf{E}_i \, d\mathbf{l} = E_i \, dl \cos \angle(\mathbf{E}_i, d\mathbf{l}) = E_i \, dr_i$. Теперь воспользуемся

законом Кулона в форме (1.3), откуда следует

$$\int_1^2 E_i dr_i = \frac{q_i}{4\pi\epsilon_0} \int_{r_1}^{r_2} \frac{dr}{r^2} \equiv \varphi_{i1} - \varphi_{i2}, \quad (1.15)$$

где мы ввели величину φ_i — потенциал i -го заряда:

$$\varphi_i = \frac{q_i}{4\pi\epsilon_0 r_i}. \quad (1.16)$$

Используя (1.15), (1.16), мы можем определить работу следующим образом:

$$A_{12} = q \cdot (\varphi_2 - \varphi_1), \quad \text{где} \quad \varphi(\mathbf{r}) = \sum_i \varphi_i(\mathbf{r}), \quad (1.17)$$

т. е. из принципа суперпозиции полей следует аддитивность потенциалов.

Величина $q\varphi$ имеет смысл потенциальной энергии. Как мы уже знаем из раздела «Механика», потенциальная энергия определена с точностью до произвольной аддитивной константы; точно так же определен и потенциал φ . Физический смысл имеет только разность потенциалов ($\varphi_2 - \varphi_1$), но в принципе, исключительно из соображений удобства, можно доопределить потенциал так, чтобы, скажем, он был равен нулю на бесконечности, как это фактически сделано в формуле (1.16). Тогда для данной точки он будет равен отношению работы по перемещению пробного заряда из этой точки на бесконечность к величине этого заряда.

Иногда, в модельных задачах, такое определение оказывается не вполне удобным. Рассмотрим еще раз примеры, представленные на рис. 1.2. Процедура интегрирования, аналогичная той, что привела нас к определению (1.16), дает:

— для одномерного случая (заряженная плоскость)

$$\varphi = \frac{\sigma x}{2\epsilon_0} \quad (x = 0 \text{ соответствует плоскости заряда});$$

— для двумерного случая

$$\varphi = q^* \ln \frac{r}{2\pi\epsilon_0}.$$

Появление в последнем случае размерной функции под знаком логарифма нас смущать не должно, так как разность потенциалов оказывается при этом правильной размерности: $(\varphi_2 - \varphi_1) \propto \ln(r_2/r_1)$. Можно учесть упомянутую выше произвольную константу в виде $\ln(r/R_0)$. Общее свойство обоих результатов состоит в том, что потенциал даже в принципе не может быть задан равным нулю на бесконечности. Но дело в том, что одно- и двумерные системы всегда представляют собой некоторую идеализацию, приближение. Удалившись от них на

достаточное расстояние, мы вернемся к существенно трехмерной картине и, соответственно, к общим формулам (1.16), (1.17).

Единица измерения потенциала или разности потенциалов, которая называется напряжением, — 1 вольт (по имени итальянского физика А. Вольты (1745–1827)): $1 \text{ В} = 1 \text{ Дж/1 Кл}$.

В задачах электростатики мы имеем дело исключительно с покоеющимися зарядами, находящимися в равновесии. Из курса механики нам известно, что такое состояние обусловлено либо минимумом потенциальной энергии, либо условием $U = \text{const}$. Представим себе проводник в электрическом поле. Поскольку заряды по нему могут течь свободно (это и есть главное свойство проводящего тела), а должны находиться в равновесии, статическая ситуация отвечает условию $\varphi = \text{const}$, т. е. *проводник должен быть эквипотенциален*. Это позволяет ввести специфичное для проводников понятие емкости.

Представим себе уединенный полый металлический шар, обстреливаемый пучком заряженных частиц. Если энергия частиц T , то они могут преодолеть разность потенциалов, не превышающую T/e , а значит, оседать на шаре частицы будут лишь до тех пор, пока на нем не соберется заряд $Q = 4\pi\epsilon_0 R T/e$, где R — радиус шара. Зависимость $Q \propto R \cdot \varphi_{\text{max}}$ дает основание говорить о радиусе как мере электрической емкости шара. То же примерно можно сказать и об уединенном проводящем теле произвольной геометрии: величина $C = Q/\varphi$ есть инвариант относительно величины заряда либо потенциала и характеризует тело как таковое. Ее называют *емкостью*. Проводником тело должно быть для того, чтобы мы вообще могли приписать ему определенный потенциал и пренебречь перепадом потенциала от точки к точке. Единица емкости называется фарадой; $1 \text{ Ф} = 1 \text{ Кл/1 В}$. Такой емкостью обладает, например, проводящий шар радиуса $9 \cdot 10^6 \text{ км}$. Для практических нужд используются микрофарады (10^{-6} Ф) и пикофарады (10^{-12} Ф).

Можно ввести в рассмотрение и взаимную емкость двух проводников (такую систему обычно называют конденсатором, а проводники — его обкладками). Она измеряется в тех же единицах, а определяется следующим образом: если система в целом нейтральна, а заряды обкладок равны $\pm Q$, и при этом потенциалы обкладок равны φ_1 и φ_2 , то емкость конденсатора

$$C_{12} = \frac{|Q|}{|\varphi_1 - \varphi_2|}.$$

В отличие от емкости уединенного проводника, это величина не универсальная, она, вообще говоря, зависит от формы, расположения и зарядов всех окружающих тел. Но в двух случаях эта характеристика однозначна: а) один проводник заключен внутри другого, и задача обладает достаточно высокой симметрией;

б) размеры обкладок велики в сравнении с расстоянием между ними и малы в сравнении с расстояниями до всех прочих тел.

Пример 1. Сферический конденсатор с радиусами обкладок R_1 и R_2 . Из теоремы Гаусса и симметрии задачи следует, что поле внутри меньшего шара и вне большего равно нулю, а в зазоре эквивалентно полю точечного заряда Q ; таким образом,

$$|\varphi_1 - \varphi_2| = \frac{1}{4\pi\epsilon_0} \left| \frac{1}{R_1} - \frac{1}{R_2} \right|,$$

откуда следует

$$C = \frac{4\pi\epsilon_0 R_1 R_2}{|R_1 - R_2|}.$$

Пример 2. Плоский конденсатор, площадь обкладок S , зазор d . Этот пример по существу уже рассмотрен (см. рис. 1.2 а):

$$\Delta\varphi = \frac{\sigma d}{\epsilon_0}; \quad Q = \sigma S,$$

откуда $C = \epsilon_0 S/d$. (Следует иметь в виду, что поля двух заряженных плоскостей дадут в сумме нуль вне конденсатора и удвоят перепад потенциала, в сравнении с одной заряженной плоскостью, внутри него.)

1.5. Уравнения Лапласа и Пуассона

Из определений (1.15)–(1.17) следует, что при малом перемещении точки наблюдения в пространстве $d\varphi = -\mathbf{E} \cdot d\mathbf{l}$, т. е. электрическое поле можно трактовать как вектор, ориентированный в направлении наиболее быстрого спада потенциала и равный $\partial\varphi/\partial l$ в этом направлении. Для такой связи между векторной и скалярной функцией существует специальная операция — градиент (мы уже ввели это понятие в гл. 5 раздела «Механика»). Градиент есть результат прямого применения уже известного нам векторного оператора «набла» к скалярной функции:

$$\nabla\varphi \equiv \mathbf{e}_x \frac{\partial\varphi}{\partial x} + \mathbf{e}_y \frac{\partial\varphi}{\partial y} + \mathbf{e}_z \frac{\partial\varphi}{\partial z}.$$

Вектор $\nabla\varphi$ показывает направление максимального роста скалярной функции. Электрическое поле направлено в обратную сторону:

$$\mathbf{E} = -\nabla\varphi. \quad (1.18)$$

Подставим (1.18) в дифференциальную форму теоремы Гаусса (1.14). Любопытно отметить, что с операторами векторного дифференцирования при этом можно обходиться, как с обычными векторами:

$$\nabla \cdot \nabla\varphi = \frac{\partial^2\varphi}{\partial x^2} + \frac{\partial^2\varphi}{\partial y^2} + \frac{\partial^2\varphi}{\partial z^2} \equiv \nabla^2\varphi.$$

Оператор ∇^2 называется *оператором Лапласа* (по имени французского астронома, физика и математика Пьера С. Лапласа (1749–1827)). Итак, мы видим, что уравнение (1.14) может быть представлено в скалярном виде, правда, порядок его при этом повышается до второго:

$$\nabla^2 \varphi = -\frac{\rho}{\varepsilon_0}. \quad (1.19)$$

Оно называется *уравнением Пуассона*. В принципе, его можно точно решить при любом заданном распределении $\rho(\mathbf{r})$. Если же мы ищем поле в свободном пространстве, то его следует описывать уравнением

$$\nabla^2 \varphi = 0. \quad (1.20)$$

Это *уравнение Лапласа*, относящееся к числу основных уравнений математической физики. Конечно, помимо самих уравнений, принципиально важны также и граничные условия — зачастую в них-то и заключена физическая постановка задачи.

Наиболее очевидным оказывается граничное условие на поверхности проводника. В статической задаче ток в проводнике течь не может, значит весь проводник должен находиться при одном и том же потенциале. То же можно сказать и о его поверхности — на ней $\varphi = \text{const}$. Такие поверхности в общей картине поля называются эквипотенциальными. Градиент потенциала, как направление самого быстрого изменения функции, должен быть к ним ортогонален. Следовательно, электрическое поле в статической задаче должно быть перпендикулярно к поверхности проводника, а внутри него — равно нулю. Действительно, $\nabla \text{const} \equiv 0$. Легко сообразить, что, нарушив любое из этих условий, мы тем самым спровоцировали бы ток в проводнике.

Выведем в рамках электростатики граничное условие на произвольной границе раздела (пока это будет относиться исключительно к полю в вакууме). Из механики нам известно, что в случае консервативных сил, когда можно ввести потенциальную энергию, работа по перемещению тела по любому замкнутому контуру равна нулю. Мы, по существу, доказали, что из закона Кулона и принципа суперпозиции полей следует потенциальность статического электрического поля, в частности,

$$\oint \mathbf{E} \cdot d\mathbf{l} = 0. \quad (1.21)$$

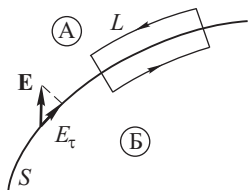


Рис. 1.3

Пусть дана некоторая поверхность S , разграничивающая области пространства А и Б (рис. 1.3). Построим контур L , как это показано на рисунке, а затем начнем стягивать с двух сторон, устремляя к нулю сторону, перпендикулярную поверхности S . Как следствие,

формула (1.21) перейдет в условие непрерывности на границе тангенциальной компоненты поля:

$$\Delta \mathbf{E}_\tau \equiv (\mathbf{E}_\tau)_B - (\mathbf{E}_\tau)_A = 0. \quad (1.22)$$

1.6. Электрический диполь

Самое простое определение электрического диполя — пара зарядов $+q$ и $-q$, на расстоянии l друг от друга (рис. 1.4 а); при этом величина $\mathbf{p} = q \cdot \mathbf{l}$ называется *вектором дипольного момента*. Мы не ограничимся таким определением — как будет ясно из дальнейшего, физический смысл электрического диполя и дипольного момента оказывается более понятен, если мы несколько обобщим наше определение.

Пусть дана некоторая система зарядов q_1, q_2, \dots, q_n , расположенных в точках $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n$. Достаточно очевидно, что на расстояниях, больших в сравнении с характерным расстоянием между зарядами, ее можно рассматривать как точечный заряд:

$$\mathbf{E} \approx \frac{Q}{r^3} \mathbf{r}; \quad \varphi \approx \frac{Q}{r}; \quad Q = \sum_i q_i.$$

Что же определяет поле системы зарядов, если $\sum_i q_i = 0$? Или, как ввести поправки, учитывающие конечный размер системы для ненулевого полного заряда? Следующим приближением по отношению характерной величины $|\mathbf{r}_i - \mathbf{r}_j|$ к расстоянию до точки наблюдения и будет так называемое *дипольное приближение*. Определим дипольный момент произвольной системы зарядов следующим образом:

$$\mathbf{p} = \sum_i q_i \mathbf{r}_i. \quad (1.23)$$

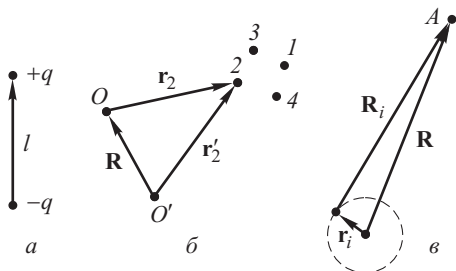


Рис. 1.4

Если зарядов всего два, и притом равных по величине и отличных по знаку, старое определение остается в силе — просто как следствие (1.23). Но определение (1.23) является существенно более общим, поскольку вводится для произвольной системы зарядов. Важно отметить, что формула (1.23) инвариантна относительно переноса начала координат, если полный заряд системы равен нулю. Это нетрудно показать с помощью рис. 1.4 б. Можно видеть, что при переносе начала отсчета из точки O в точку O'

имеем $\mathbf{r}'_i = \mathbf{r}_i + \mathbf{R}$, откуда следует

$$\mathbf{p}' = \mathbf{p} + \mathbf{R} \sum_i q_i.$$

При нулевой сумме зарядов дипольный момент и в самом деле оказывается инвариантом. И если мы хотим определить поле нейтральной в целом системы заряженных частиц (например, поле заряженного плоского конденсатора), то именно дипольный момент системы оказывается принципиально важным.

Вычислим поле системы зарядов в точке A , удаленной в сравнении с характерным пространственным масштабом системы (рис. 1.4 *в*). Начнем с потенциала, ограничившись первым членом разложения в ряд Тэйлора:

$$f(\mathbf{R} + \delta\mathbf{R}) \approx f(\mathbf{R}) + \delta x \cdot \frac{\partial f}{\partial x} + \delta y \cdot \frac{\partial f}{\partial y} + \dots = f(\mathbf{R}) + \delta\mathbf{R} \cdot \nabla f.$$

Воспользовавшись тем, что $r_i \ll R$ (см. рис. 1.4 *в*), получаем

$$\begin{aligned} 4\pi\epsilon_0\varphi(A) &= \sum_i \frac{q_i}{|\mathbf{R} - \mathbf{r}_i|} \approx \sum_i \frac{q_i}{R} - \sum_i q_i \mathbf{r}_i \cdot \nabla \frac{1}{R} = \\ &= \frac{Q}{R} - \mathbf{p} \cdot \nabla \frac{1}{R} \stackrel{Q=0}{=} \frac{(\mathbf{p} \cdot \mathbf{R})}{R^3}. \end{aligned} \quad (1.24)$$

Таким образом, у нейтральной системы с ненулевым дипольным моментом $\varphi \propto R^{-2}$. Теперь можно вычислить и поле:

$$\mathbf{E} = -\nabla\varphi = \frac{1}{4\pi\epsilon_0} \left(3 \frac{(\mathbf{p} \cdot \mathbf{R})}{R^4} \frac{\mathbf{R}}{R} - \frac{\mathbf{p}}{R^3} \right) = \frac{3(\mathbf{p} \cdot \mathbf{R})\mathbf{R} - R^2\mathbf{p}}{4\pi\epsilon_0 R^5}, \quad (1.25)$$

т. е. при нулевом полном заряде $E \propto R^{-3}$. При этом, в отличие от точечного заряда, поле зависит от угла. На примере (1.24), (1.25) можно убедиться, что зачастую полезно потратить время на подготовку корректных определений и затем использовать адекватный математический аппарат, если это позволяет избежать громоздких вычислений. Любая из этих формул, выведенная «в лоб», потребовала бы много больше работы.

Далее, если мы будем говорить о точечном диполе, понимать это надо как нейтральную зарядовую систему, поле которой рассматривается на столь больших расстояниях, что справедливы соотношения (1.24), (1.25). И если для представлений об элементарном магнитном диполе современная физика дает некоторые основания, то в данном случае таковые отсутствуют: на сегодняшний день у элементарных частиц электрические дипольные моменты не обнаружены.

Вычислим в дипольном приближении энергию системы зарядов во внешнем поле:

$$U = \sum_i q_i \varphi(\mathbf{R} + \mathbf{r}_i) = \varphi(\mathbf{R}) \sum_i q_i + \sum_i q_i \mathbf{r}_i \cdot \nabla \varphi = \\ = Q \varphi(\mathbf{R}) - (\mathbf{p} \mathbf{E}). \quad (1.26)$$

Из (1.26) следует, что сила, действующая на точечный диполь в электрическом поле, может быть представлена в виде

$$\mathbf{F} = \nabla(\mathbf{p} \mathbf{E}). \quad (1.27)$$

Можно также показать (предоставляем это читателю в качестве упражнения), что момент, действующий на дипольную систему во внешнем поле, равен

$$\mathbf{M} = [\mathbf{p} \mathbf{E}]. \quad (1.28)$$

Задачи

1. В пространстве между пластинами плоского конденсатора имеется свободный поток электронов, который создает равномерный объемный заряд. Расстояние между пластинами равно d , потенциал одной из пластин равен φ_0 . При каком значении объемной плотности заряда ρ потенциал и напряженность поля у другой пластины равны нулю?

Решение. Введем ось Ox , как показано на рис. 1.5, и воспользуемся теоремой Гаусса в форме (1.14). В нашей одномерной задаче $\mathbf{E} \parallel Ox$, следовательно, $\operatorname{div} \mathbf{E} = \partial E / \partial x$. Таким образом, $\partial E / \partial x = \rho / \varepsilon_0$, откуда $E = \text{const} + \rho x / \varepsilon_0$. Поскольку $E(d) = 0$, $\text{const} = -\rho d / \varepsilon_0$, т. е. $E = \rho(x - d) / \varepsilon_0$. Далее,

$$\varphi(x) = \varphi_0 - \int_0^x E dx; \quad \varphi(d) = \varphi_0 + \frac{\rho d^2}{2\varepsilon_0} = 0; \\ \rho = -\frac{2\varepsilon_0 \varphi_0}{d^2}.$$

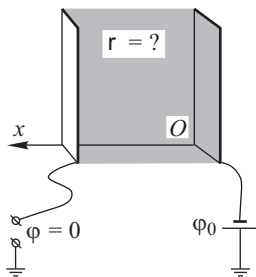


Рис. 1.5

2. Определить силу взаимодействия между точечным зарядом q и заземленным металлическим шаром. Радиус шара R , расстояние между зарядом q и центром шара равно d .

Эту задачу удобно решать *методом изображений*. Если в электрическом поле заменить любую эквипотенциальную поверхность проводником той же формы и создать на нем потенциал, равный потенциалу той же поверхности, поле не изменится. Можно действовать и наоборот. Оба приема базируются на теореме однозначности: поле системы проводников и/или распределение зарядов на их поверхности полностью определены, если: а) известен потенциал каждого проводника либо б) задан общий заряд каждого проводника.

Решение. Предположим, что мы можем реализовать нужную конфигурацию полей вне шара, заменив последний точечным зарядом Q , расположенным на линии «заряд q — центр шара» (учитывается осевая симметрия задачи) и сдвинутым на некоторое расстояние a от центра (рис. 1.6 а). Если такая

модель адекватна, то $\varphi = 0$ на поверхности сферы, в частности, в точках ее, ближайшей и удаленнейшей относительно заряда q : $\varphi(d - R) = \varphi(d + R) = 0$, или, подробнее,

$$\frac{q}{d - R} + \frac{Q}{R - a} = 0;$$

$$\frac{q}{d + R} + \frac{Q}{R + a} = 0,$$

откуда следует: $a = R^2/d$, $Q = -qR/d$.

Докажем, что при такой замене потенциал на всей сфере обратится в нуль. Воспользуемся осевой симметрией (рис. 1.6 б) и получим уравнение геометрического места точек в произвольном сечении, для которых $\varphi = 0$:

$$\frac{|Q|}{[(x - a)^2 + y^2]^{1/2}} = \frac{q}{[(d - x)^2 + y^2]^{1/2}}.$$

Подставим сюда уже вычисленные Q и a :

$$\frac{R^2/d^2}{(x - R^2/d)^2 + y^2} = \frac{1}{(d - x)^2 + y^2},$$

откуда в два действия следует $x^2 + y^2 = R^2$. Наше утверждение доказано, так что и окончательный ответ для силы получить нетрудно:

$$F = \frac{q|Q|}{4\pi\epsilon_0(d - a)^2} = \frac{q^2 R d}{4\pi\epsilon_0(d^2 - R^2)^2}.$$

3. Найти силу взаимодействия двух точечных диполей, если их дипольные моменты \mathbf{p}_1 и \mathbf{p}_2 направлены вдоль соединяющей их прямой, а расстояние между диполями равно d .

Ответ: $F_{12} = -F_{21} = 3p_1 \cdot \frac{p_2}{2\pi\epsilon_0 d^4}$. Сонаправленные

диполи притягиваются, антипараллельные — отталкиваются.

4. Найти электрическое поле \mathbf{E} в шаровой полости однородно заряженного шара (рис. 1.7). Объемная плотность заряда равна ρ . Расстояние между центром полости и центром шара равно l .

Ответ: поле в полости однородно; $\mathbf{E} = \frac{\rho \mathbf{l}}{3\epsilon_0}$.

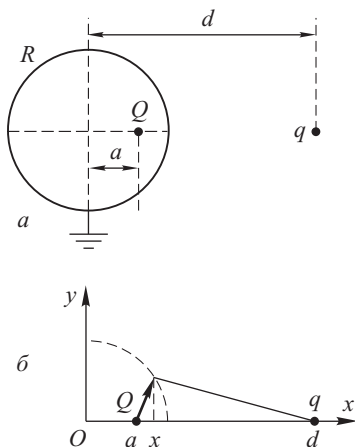


Рис. 1.6

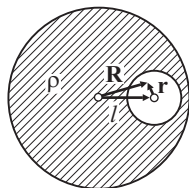


Рис. 1.7

Глава 2

ЭЛЕКТРИЧЕСКОЕ ПОЛЕ В ВЕЩЕСТВЕ

2.1. Проводники в электрическом поле

Мы уже касались вопроса о проводящем теле в постоянном электрическом поле в гл. 1; а именно, было установлено, что внутри проводника электрическое поле должно быть равно нулю, соответственно, на границе равна нулю его тангенциальная компонента — условие (1.22). Тогда, окружив некоторый участок границы цилиндрической поверхностью, подобно тому, как это сделано на рис. 1.2 *a*, и устремляя высоту цилиндра к нулю, мы приходим к выводу, что на поверхности должен быть сосредоточен некоторый электрический заряд. Теорема Гаусса (1.10) дает нам непосредственно поверхностную плотность этого заряда:

$$\sigma = \frac{dq}{dS} = E_n \cdot \varepsilon_0,$$

где E_n — нормальная к границе проводника компонента поля. Можно посмотреть на ту же ситуацию и по-другому. Пусть в некоторой области пространства внешними источниками задано ненулевое поле. Внесем туда электрически нейтральное проводящее тело. Теперь поле внутри него обращается в нуль. В силу принципа суперпозиции, это означает, что собственное поле, созданное зарядами на поверхности проводника, везде должно быть равно по величине и противоположно по направлению внешнему полю.

То, что заряды перераспределяются исключительно по поверхности, следует из уравнения (1.14). Действительно, если $\mathbf{E} \equiv 0$, то и $\rho \equiv 0$. Поскольку тело остается нейтральным как целое, поле поверхностных зарядов имеет характер дипольного (или даже включает следующие по порядку величины поправки, которых мы в нашем курсе не касались). Подчеркнем еще раз, что, вследствие равенства нулю поля внутри проводника, последний в рамках электростатики остается эквипотенциальным (что, в частности, используется в упомянутом выше методе изображений).

Поскольку читателю уже известно даже из школьного курса, что электрический заряд любого тела есть не что иное, как несбалансированность концентрации электронов и ионов в веществе, представляется естественным следующий вопрос. В реаль-

ном теле, в отличие от его математической идеализации, возмущение заряда не может представлять собой поверхность нулевой толщины; как же в таком случае обеспечить нулевое поле всюду внутри проводника? Дело в том, что в наших аргументах мы молчаливо апеллировали к закону Ома — в его качественной форме: ненулевое поле в проводящей среде порождает ток. Но такой простой подход годится лишь в случае нейтральной среды, а как раз в поверхностном слое, где нейтральность существенно нарушена, движение заряженных частиц подчиняется более сложным зависимостям, нежели закон Ома. И конечно, в действительности электрическое поле меняется в некотором тонком слое у поверхности от нуля до некоторого пограничного значения.

Детальное рассмотрение проблемы выходит за рамки настоящей книги, так что мы в этом случае лишь отметим нетривиальность данного вопроса.

Если проводник не нейтрален, т. е. несет на себе некоторый заряд, то, опираясь по-прежнему на принцип суперпозиции, можно представить себе, что в этом случае на поле внешних источников накладываются поле поляризации проводника и поле его собственного заряда. Поляризационный и собственный заряды распределены по поверхности таким образом, чтобы внутри проводника поле оставалось нулевым. Условие $E \equiv 0$ внутри проводника и соображения о его эквипотенциальности остаются в силе.

И последнее — по порядку, но не по значению. Во всех рассуждениях относительно поля в проводящей среде, включая уравнение (1.22) и условие нулевого поля и заряда внутри проводника, мы неявно имели в виду некоторое усреднение по пространству и/или по времени. В самом деле, внутри атома электрические поля составляющих его частиц при заряде порядка заряда электрона и пространственном масштабе порядка 10^{-10} м равны по порядку величины 10^{11} В/м. В межатомном пространстве поля слабее, ибо атомы в целом нейтральны, но, скажем, в твердом теле порядок величины сохраняется; в металлах или ионных кристаллах сделанная выше оценка вполне корректна.

Поэтому, говоря о среде как непрерывной субстанции или о поле в среде, которое должно фигурировать в уравнениях электростатики, мы тем самым подразумеваем усредненную картину, в которой структура вещества остается за рамками нашего рассмотрения. В отношении поля в диэлектриках это приближение будет использовано с большей определенностью.

2.2. Поляризация диэлектриков. Понятие электрической индукции

Из опыта хорошо известно, что любое вещество, если его внести в электрическое поле, проявляет ответную реакцию на такое воздействие. Даже в тех случаях, когда в стандартных задачах этим взаимодействием традиционно пренебрегают — например, когда в поле находится воздух в нормальных условиях, — речь идет лишь о чисто количественной малости взаимодействия, но не об его отсутствии. Равным образом можно утверждать, что не существует веществ, абсолютно не проводящих. Выделение диэлектриков в отдельный класс (а тем более — выделение в этом классе подкласса полупроводников) базируется на опыте и подкрепляется определенными количественными аргументами. Например, обыкновенная вода H_2O в зависимости от параметров задачи, отнюдь не меняя физико-химических свойств, может вести себя как проводящая среда или использоваться в качестве диэлектрика. Микроскопическое рассмотрение (см. раздел 5) внесет в эту классификацию большую определенность. Нам же сейчас, имея дело с диэлектриками в электростатике, важно будет лишь следующее условие: рассматривается простейший случай взаимодействия с полем веществ, проводимостью которых мы можем пренебречь и тем самым считать поле в среде ненулевым.

Элементарный школьный рецепт хорошо известен: при данном распределении свободных зарядов разделить силы, поля, потенциалы и т. д. на постоянную величину ϵ , характеризующую данное вещество. Как мы скоро убедимся, рецепт этот не случаен, более того, во многих случаях, важных с точки зрения технических приложений, например в конденсаторах и волноводах, он работает достаточно хорошо. Но с точки зрения физики, такой метод, как правило, некорректен. Попробуем разобратся в простейшей электродинамике диэлектриков, исходя из фундаментальных принципов.

Для последующих обобщений уравнений электрического поля в вакууме полезно будет ввести новую физическую величину. Определим вектор *электрической индукции в вакууме*, отличающийся от поля лишь коэффициентом:

$$\mathbf{D} = \epsilon_0 \mathbf{E}, \quad (2.1)$$

так что, например, теорема Гаусса, вместо выражений (1.10), (1.14), может быть представлена в виде:

$$\oint \mathbf{D} \, d\mathbf{S} = \sum_i q_i, \quad (2.2)$$

$$\operatorname{div} \mathbf{D} = \rho. \quad (2.3)$$

Пока и поскольку мы рассматриваем поле в вакууме, физический смысл величин \mathbf{E} и \mathbf{D} остается тождественным.

Пусть теперь предметом рассмотрения будет электрическое поле в диэлектрической среде. Представим себе реакцию среды. Если даже ее проводимость пренебрежимо мала, это не означает запрета на смещение во внешнем поле заряженных частиц относительно их равновесного положения в нейтральном атоме либо кристалле. Результатом смещения окажется некоторое ослабление внешнего поля, как говорят, экранирование. Сама же реакция среды — поляризация — может быть представлена распределенным в среде электрическим дипольным моментом. При не слишком больших напряженностях поля в веществе естественно ожидать, что дипольный момент и поле будут связаны линейной зависимостью (да и более сложную функциональную связь можно в слабых полях приближенно представить первым членом ряда Тэйлора). Поскольку обе величины являются векторными, естественно ожидать, что эти векторы будут параллельны. Итак, предлагается следующее соотношение между полем \mathbf{E} и плотностью дипольного момента \mathbf{P} :

$$\mathbf{P} \equiv \frac{d\mathbf{p}}{dV} = \beta \cdot \varepsilon_0 \mathbf{E}. \quad (2.4)$$

Величина ε_0 выделена в коэффициенте особо для удобства последующих вычислений; под \mathbf{E} понимается действующее поле в среде — только его и могут воспринимать заряженные частицы.

Сделаем два важных замечания. Во-первых, сколь ни естественны наши предположения при введении зависимости (2.4), нельзя исключить и обратных предположений — скажем, в сильных полях связь может оказаться существенно нелинейной. Поэтому физик следом за любыми умозрительными предположениями обязан обращаться к опыту, чтобы подтвердить их либо отвергнуть. В нашем случае совокупность экспериментальных данных говорит о том, что, хотя соотношение (2.4) и не универсально, оно очень хорошо работает в широком диапазоне параметров. Поэтому для того, чтобы разобраться в физике дела, мы вполне можем пользоваться такой моделью. Можно в принципе обобщить ее на случай анизотропного кристалла — тогда скалярную линейную связь придется заменить тензорной — но лучше все-таки для понимания принципиальных моментов в физике не перегружать модель необязательными обобщениями.

Во-вторых, оперируя понятием поля в веществе, мы абстрагируемся от очень сильных, но короткодействующих микроскопических полей в атоме или кристаллической решетке. Подразумевается усреднение в пространстве и/или во времени с таким пространственным и временным масштабом, что микроскопические эффекты из нашего поля зрения исчезают и остается лишь

умеренной напряженности электрическое поле, которое может быть измерено на макромасштабе и макроскопическим регистрирующим прибором. При этом диэлектрик рассматривается просто как сплошная среда.

Рассмотрим две модельных задачи, геометрия которых представлена на рис. 2.1 а, б. Прежде всего, пусть имеется малый параллелепипед, мысленно выделенный в объеме однородно поляризованного диэлектрика, так что вектор плотности дипольного момента \mathbf{P} (его называют также *вектором поляризации*) параллелен оси x . Пусть стороны параллелепипеда соответственно равны Δx , Δy , Δz . Тогда полный его дипольный момент равен $p_x = P_x \Delta x \Delta y \Delta z$. С другой стороны, нетрудно сообразить, что при однородной поляризации — смещении всех отрицательных зарядов относительно всех положительных — плотность заряда в объеме тела остается невозмущенной и лишь на торцевых гранях появляются поверхностные заряды $\pm \sigma_x$. Из определения дипольного момента (1.23) следует

$$p_x = \sum_i q_i x_i = (\sigma \Delta y \Delta z) \cdot \Delta x.$$

Сравнивая оба результата, получаем $P_x = \sigma_x$, или в общем случае границы поляризованного диэлектрика — скачок поляризации

$$\Delta(\mathbf{P})_n = \sigma, \quad (2.5)$$

где индекс n означает компоненту, нормальную к поверхности разрыва.

В случае неоднородной поляризации в объеме может возникнуть и ненулевое возмущение заряда. Такой заряд принято называть *связанным*, в отличие от свободных зарядов, с которыми мы до сих пор имели дело. Связанный заряд не может быть ни создан внешним источником, ни снят с тела, поскольку это лишь определенное состояние среды, обусловленное воздействием электрического поля и оставляющее тело электрически нейтральным. Смоделируем неоднородную поляризацию среды, соединив торцами два малых параллелепипеда, как показано на рис. 2.1 б. В соответствии с (2.5) на первом стыке возникает заряд

$$dq = dy dz \cdot \Delta \sigma_x \approx -dx dy dz \frac{\partial P_x}{\partial x}.$$

Относя эту величину к объему элементарного параллелепипеда, получаем ответ уже в достаточно универсальном виде плотности

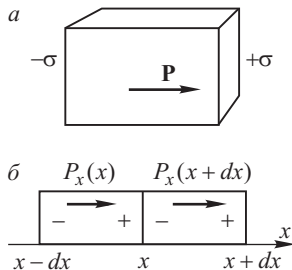


Рис. 2.1

заряда:

$$\rho_{\text{связ}} = -\frac{\partial P_x}{\partial x}.$$

Аналогичные ответы последуют при неоднородности поляризации в направлениях y и z . В общем же случае мы должны будем взять их сумму:

$$\rho_{\text{связ}} = -\left[\frac{\partial P_x}{\partial x} + \frac{\partial P_y}{\partial y} + \frac{\partial P_z}{\partial z}\right] = -\operatorname{div} \mathbf{P}. \quad (2.6)$$

Этот результат позволит нам получить достаточно общее уравнение для поля в диэлектрике. Обратимся к уравнению (1.14) и подставим в его правую часть плотность всех зарядов, как свободных, так и связанных:

$$\varepsilon_0 \operatorname{div} \mathbf{E} = \rho_{\text{своб}} + \rho_{\text{связ}} = \rho - \operatorname{div} \mathbf{P}. \quad (2.7)$$

Мы сохранили обозначение ρ за плотностью свободных зарядов. Введем понятие *индукции* электрического поля в случае диэлектрической среды, по отношению к которому данное выше определение (2.1) окажется естественным для вакуумного поля частным случаем:

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}. \quad (2.8)$$

Сравнивая (2.7) и (2.8), нетрудно убедиться, что электрическая индукция \mathbf{D} по-прежнему удовлетворяет уравнениям (2.2), (2.3). Зависимость, связывающая \mathbf{D} и \mathbf{E} , не столь универсальна; более того, она зависит еще и от того приближения, в котором мы описываем диэлектрическую среду. В случае простой модели (2.4) получаем

$$\mathbf{D} = \varepsilon_0(1 + \beta)\mathbf{E} \equiv \varepsilon_0 \varepsilon \mathbf{E}. \quad (2.9)$$

Величина β называется *поляризуемостью* среды, а $\varepsilon = 1 + \beta$ — *диэлектрической проницаемостью*. Таким образом, если среда однородна, то, действительно, поле в сравнении с вакуумной ситуацией просто ослабляется в ε раз. В частности, закон Кулона может быть переписан в виде

$$\mathbf{F}_{12} = \frac{q_1 q_2}{4\pi \varepsilon_0 r_{12}^2} \frac{\mathbf{r}_{21}}{r_{12}}.$$

Рассмотрим конденсатор, емкость которого в отсутствие диэлектрического заполнения — C_0 . Заполним пространство между обкладками диэлектриком с постоянной ε .

Если при этом конденсатор был отключен от источника, должны сохраниться заряды на обкладках $\mp Q$. Тем самым при заданном распределении свободных зарядов сохраняется величина индукции; поле же падает в ε раз. Исходя из соотношений

$$U = \frac{U_0}{\varepsilon} \equiv \frac{Q}{C},$$

получаем $C = \varepsilon C_0$.

Если конденсатор был подключен к источнику, то инвариантным остается напряжение на обкладках, а значит и электрическое поле в конденсаторе. Следовательно, индукция обязана возрасти в ε раз, а с ней, в согласии с (2.2), (2.3), и заряды на обкладках. В итоге снова приходим к тому же ответу — он вообще оказывается универсальным:

$$C = \varepsilon C_0. \quad (2.10)$$

Базируясь на подобных тривиальных примерах, авторы некоторых публикаций позволяют себе смелые заявления:

«Электрическая индукция — с точностью до коэффициента — есть поле свободных зарядов, без учета поляризации среды».

«Индукция и поле связаны между собой коэффициентом диэлектрической проницаемости вещества».

Оба эти утверждения в общем случае неверны. И дело не только в том, что связь между \mathbf{P} и \mathbf{E} может не быть линейной. Переход от поля свободных, т. е. заданных извне зарядов к истинному полю не есть результат простого суммирования внешнего воздействия и отклика, но еще и переход от точного к некоторому усредненному описанию. Как мы увидим далее на примере сегнетоэлектриков, пространственное распределение \mathbf{D} и \mathbf{E} может быть вообще существенно различным и притом в отсутствие свободных зарядов. Более того, эти два вектора могут оказаться противоположны по направлению. Формально это получается из того, что совпадение уравнений индукции и поля в пустоте еще не означает совпадения решений, ибо решать их следует при разных граничных условиях.

2.3. Граничные условия в электростатике

Располагая в качестве рабочего инструмента теоремой Гаусса в дифференциальной форме или уравнением Пуассона, мы, казалось бы, можем в приближении сплошной среды решить любую электростатическую задачу, т. е. задать заряды или потенциалы системы физических тел и найти поле во всем пространстве. Первое препятствие к тому — проблема состояния вещества в электрическом поле. Мы ее предельно упростили, полагая вещество либо идеально проводящим, либо идеальным диэлектриком, свойства которого сведены к одной-единственной константе. Но остается еще проблема границы вещества с вакуумом или двух веществ. Нам придется считать такие границы идеально гладкими и вдобавок предположить, что свойства среды, такие, как проводимость и диэлектрическая проницаемость, меняются на границе скачком. (Иногда адекватность таких приближений представляет собой проблему более сложную, нежели собственно решение задачи).

Для статического электрического поля всегда остается в силе условие потенциальности

$$\oint \mathbf{E} d\mathbf{l} = 0. \quad (2.11)$$

Именно это условие позволяет представить электрическое поле в виде градиента скалярной функции. Оно выполняется для поля любого неподвижного электрического заряда. Конструируя поле, взаимодействующее с веществом, т. е. учитывая поляризационные эффекты, мы ничего в этом смысле не меняем. Сколько и какие бы мы ни взяли заряды, суммарное поле обязано удовлетворять уравнению (2.11).

Обратимся к рис. 2.2 а. Пусть A — идеальная граница двух сред. Построим на двух касательных и двух нормалях к этой границе некоторый малый контур, как показано на рисунке, а затем перейдем для его высоты к пределу $dz \rightarrow 0$, оставляя при этом длину dl малой, но ненулевой, и следя за тем, чтобы верхняя и нижняя стороны контура оставались в разных средах. В результате (2.11) вырождается в соотношение

$$\mathbf{E}_{II} dl - \mathbf{E}_I dl = 0,$$

Рис. 2.2

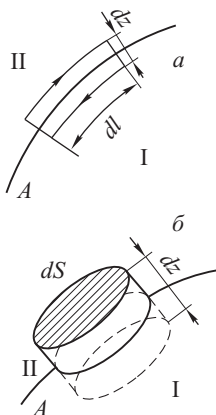
т. е. в условие сохранения проекции поля \mathbf{E} на направление касательной. Последнее выбиралось произвольно; повторив это дважды для двух ортогональных направлений в касательной плоскости, приходим к выводу о сохранении проекции поля \mathbf{E} на касательную плоскость:

$$\mathbf{E}_{I\tau} = \mathbf{E}_{II\tau}. \quad (2.12)$$

Еще одно граничное условие последует из уравнения (2.2). Построим на границе сред I и II малый цилиндр с образующей, параллельной нормали, и основаниями dS , параллельными касательной плоскости и расположенными в разных средах, как показано на рис. 2.2 б. Устремим высоту цилиндра dz к нулю, сохраняя площадь оснований малой, но ненулевой. Если на поверхности A нет поверхностного заряда, (2.2) вырождается в условие непрерывности на границе нормальной компоненты вектора индукции:

$$\mathbf{D}_{In} = \mathbf{D}_{II n}. \quad (2.13)$$

Вообще говоря, даже элементарных сведений о строении вещества и электронной природе электрического заряда достаточно, чтобы утверждать, что в строгом смысле поверхностного заряда не бывает. Невозможно конечный заряд поместить на



листок нулевой толщины. Но в строгом смысле не бывает и идеального скачка на границе, так что мы должны различать реальную физику и ее математическую модель. Например, корректно описать проводник в электрическом поле можно лишь привлекая понятие поверхностного заряда или входя в такие детали физики пограничного слоя, которые неоправданно усложняют задачу. Иногда, к тому же, необходимая нам точность решения допускает такое упрощение, как «стягивание» заряженного слоя конечной толщины в бесконечно тонкий листок. Итак, пусть на поверхности A располагается некоторый заряд с поверхностной плотностью σ . Тогда (2.2) можно переписать в виде

$$\mathbf{D}_{\parallel n} dS - \mathbf{D}_{\perp n} dS = \sigma dS,$$

и окончательно, вместо (2.13), получаем

$$\mathbf{D}_{\perp n} - \mathbf{D}_{\parallel n} = \sigma. \quad (2.14)$$

Таким образом, действительно, в отличие от тривиального вакуумного случая, граничные условия для поля и индукции не совпадают, так что утверждение, будто индукция — это просто вакуумное поле свободных зарядов, некорректно.

Условия (2.12), (2.13) подсказывают и способы измерения поля в веществе. Представим себе, что мы вырезали в образце вещества очень тонкую щель достаточно большого сечения, перпендикулярную силовым линиям, и вводим туда зонд, способный измерить в ней электрическое поле. Тогда, в силу (2.13), индукция в такой щели будет равна индукции в веществе, но внутри щели она с точностью до коэффициента ϵ_0 как раз и равна измеренному полю. Если бы мы вырезали щель не перпендикулярно, а параллельно силовым линиям, то, согласно (2.12), измеренное в ней поле было бы равно полю в веществе. Не столь нагляден ответ в случае, например, анизотропного кристалла, когда направления \mathbf{E} и \mathbf{D} , вообще говоря, не совпадают.

Забегая вперед, отметим, что область применимости полученных граничных условий не ограничивается электростатикой, более того, они оказываются совершенно универсальными.

2.4. Диэлектрики с квазиупругими и жесткими диполями

Говоря о поляризации среды, мы до настоящего времени ограничивались моделью самого общего вида, полагая наиболее естественной реакцией диэлектрика на внешнее электрическое поле переход в состояние с ненулевым дипольным моментом — см. формулы (2.4), (2.8). Детально разобраться в механизме явления мы могли бы лишь на микроскопическом уровне, но

некоторое разъяснение, по крайней мере для изотропных (жидких или газообразных) диэлектриков, можно дать и в рамках настоящего курса.

Поляризация изотропного диэлектрика может быть обусловлена одним из двух возможных механизмов (или двумя одновременно). Это, во-первых, приобретение дипольного момента нейтральной «квазиупругой» молекулой во внешнем поле и, во-вторых, ориентация во внешнем поле молекул с ненулевыми собственными дипольными моментами.

Нейтральная молекула называется квазиупругой, если в достаточно слабом поле ее дипольный момент $\mathbf{p}_{\text{мол}}$ пропорционален приложенному полю:

$$\mathbf{P} = n\mathbf{p}_{\text{мол}} = n\beta_{\text{мол}} \cdot \mathbf{E}. \quad (2.15)$$

Здесь n — концентрация молекул, $\beta_{\text{мол}}$ — коэффициент поляризуемости молекулы. Эта зависимость и в самом деле выглядит так, как если бы положительный и отрицательный заряды были связаны пружинкой, которая растягивалась бы пропорционально приложенной силе. Здесь полезно будет следующее замечание. Существует теорема (ее называют *теоремой Ирншоу*), которая утверждает: механическая система не может находиться в равновесии под действием одних только электростатических сил. Она исключает как возможность существования статического атома, так и подобную «пружинку» чисто электростатической природы. На самом деле атом — отнюдь не статическая, но динамическая система, поведение которой описывается законами квантовой механики, но для многих веществ с достаточно симметричными неполярными молекулами — N_2 , H_2 , CO_2 , CH_4 , CCl_4 — приближение (2.15) оказывается вполне адекватным. Пусть k — упругость молекулярной «пружинки»:

$$k \delta l = e\mathbf{E} \implies \mathbf{p}_{\text{мол}} = \frac{e^2}{k} \mathbf{E},$$

(мы предположили, что сила, растягивающая «пружинку», действует на элементарный электрический заряд). Окончательно,

$$\beta_{\text{мол}} = \frac{e^2}{k}; \quad \beta \equiv \frac{ne^2}{k\varepsilon_0}. \quad (2.16)$$

Отсюда сразу следует, в частности, известное соотношение для диэлектрической проницаемости воздуха в зависимости от давления:

$$\frac{d\varepsilon}{dP} = \frac{\beta_{\text{мол}}}{\varepsilon_0 k_B T},$$

где k_B — постоянная Больцмана, $P = nk_B T$.

Однако далеко не для всех диэлектриков применимо представление о квазиупругих диполях. Многие из них, например H_2O , NH_3 , H_2S , SO_2 , состоят из молекул, имеющих и в отсут-

ствие поля отличный от нуля дипольный момент, причем заметно превосходящий поправки, связанные с упругими свойствами молекул. Разумеется, в нулевом поле эти «жесткие» диполи ориентированы хаотически вследствие теплового движения молекул. При наложении внешнего поля диполи стремятся «выстроиться», а тепловое движение этому препятствует. Можно ожидать, что чем выше температура, тем ниже будет поляризуемость среды. И действительно, для таких полярных диэлектриков в умеренных полях поляризуемость с хорошей точностью удовлетворяет следующей зависимости от температуры:

$$\beta \propto \frac{p_{\text{мол}}^2}{T}, \quad (2.17)$$

где $p_{\text{мол}}$ — собственный дипольный момент молекулы. Эта зависимость может быть с достаточной строгостью получена теоретически для жидких и газообразных диэлектриков методами молекулярной физики. Особого внимания заслуживает сделанная выше оговорка относительно умеренных полей. Этот критерий не универсален; поле можно считать умеренным, если потенциальная энергия диполя, помещенного в такое поле, мала в сравнении с тепловой энергией молекулы. Привлекая для сравнения формулу (1.26), получаем условие «умеренности» электрического поля для данного полярного диэлектрика:

$$p_{\text{мол}} E \ll k_B T. \quad (2.18)$$

Если же $E \geq k_B T / p_{\text{мол}}$, результат окажется даже более очевиден, нежели формула (2.17). В этом предельном случае молекулы просто выстроятся по полю, и полный дипольный момент единицы объема можно будет оценить следующим образом:

$$\mathbf{P} \approx n \cdot \mathbf{p}_{\text{мол}}; \quad \beta \propto E^{-1}. \quad (2.19)$$

В подобных случаях принято говорить об эффекте насыщения. Конечно, и в сильных полях — пока и поскольку не вступает в игру диссоциация молекул — остается эффект квазиупругих диполей, но у полярных диэлектриков он относительно слаб. Для сравнения: ε у различных масел (неполярных диэлектриков) варьируется в пределах от 2 до 5, тогда как у воды — полярного диэлектрика — $\varepsilon = 81$.

2.5. Пьезоэлектрики и сегнетоэлектрики

Простые модели, описанные в предыдущем параграфе, по преимуществу применяются для описания диэлектрических свойств изотропных бесструктурных веществ, т. е. жидкостей, газов, аморфных твердых тел (стекол). Гораздо сложнее обстоит дело с монокристаллами, хотя идея квазиупругого отклика на

внешнее электрическое поле и здесь может быть работоспособна. Но, как и сам монокристалл, этот отклик является в принципе анизотропным, так что вместо константы вещества ε , приходится иметь дело с тензором диэлектрической проницаемости ε_{ik} :

$$D_i = \varepsilon_0 \sum_{k=1}^3 \varepsilon_{ik} E_k, \quad i = x, y, z \ (1, 2, 3). \quad (2.20)$$

В случае поликристалла анизотропия естественным образом усредняется, и свойства вещества можно снова представить единственной константой. Мы, однако, в текущем параграфе будем говорить о других свойствах кристаллов, связанных не с откликом на внешнее электрическое поле, но с собственным дипольным моментом.

Остановимся вначале на пьезоэлектрическом эффекте, открытом в 1880 г. французскими физиками Пьером (1859–1906) и Жаком Кюри. Классическим примером пьезоэлектрического кристалла является кварц SiO_2 в некоторых кристаллических модификациях.

Суть эффекта заключается в том, что при деформации кристалла последний приобретает электрический дипольный момент, который может быть зарегистрирован посредством электродов, прижатых к поверхности кристалла. Представляется достаточно очевидным, что пьезоэлектрический эффект должен быть свойством практически любого вещества с достаточно нетривиальной кристаллической структурой. Под действием механического напряжения происходит смещение ионов в кристалле и деформация электронных орбит, так что, вообще говоря, естественно ожидать ненулевого дипольного момента:

$$\mathbf{P} = \left\langle \sum_i e_i \mathbf{r}_i \right\rangle \neq 0.$$

Главным должен быть вопрос не о существовании эффекта как такового, но о его величине, о возможности его зарегистрировать при неразрушающих напряжениях. Именно по этому принципу и определяются вещества, называемые пьезоэлектриками.

Различают *продольный эффект* (самый простой), *поперечный* (вызванный пуассоновским поперечным откликом на продольную деформацию), а также *изгибный*. Возможен и *обратный* пьезоэлектрический эффект — изменение формы кристалла под действием внешнего электрического поля. Как прямой, так и обратный эффекты активно используются в технике, например, в звукозаписывающих и звуковоспроизводящих устройствах.

Сегнетоэлектрические вещества, в отличие от пьезоэлектриков, кардинально отличаются от простейших моделей — квазиупругой и ориентационной. Свое название они получили от

сегнетовой соли $\text{NaKCl}_4\text{H}_4\text{O}_6 \cdot 4\text{H}_2\text{O}$, которая в интервале температур от -15 до $+22,5^\circ\text{C}$ как раз и проявляет физические свойства, характерные для этого класса веществ. Эти основные свойства были с исчерпывающей полнотой исследованы советскими физиками И.В. Курчатовым (1903–1960) и П.П. Кобеко (1897–1954). Они в общем весьма напоминают постоянные магниты, известные каждому из нас с детства (поэтому другое их название — ферроэлектрики). Но поле, обусловленное сегнетоэлектриком, — не магнитное, а электрическое поле¹⁾. Диэлектрическая проницаемость сегнетоэлектриков ϵ очень велика, она достигает значений в несколько тысяч (сегнетова соль — до 10 000, метатитанат бария — до 7 000 и т. д.). Она не является константой вещества, но зависит от его предьстории, что выражается в так называемом *эффекте гистерезиса* (рис. 2.3).

При увеличении поля E от нуля индукция в первоначально неполяризованном диэлектрике возрастает от нуля до некоторого состояния насыщения, в котором, как обычно, $dD/dE \approx \epsilon_0\epsilon$ (на рисунке масштаб ради наглядности искажен). При уменьшении, а затем перемене знака электрического поля кривая $D(E)$ отнюдь не следует по пути 1, но, минуя нулевую точку, образует замкнутую петлю между двумя состояниями насыщения (путь 2–3). Это и есть *кривая гистерезиса*.

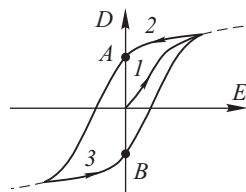


Рис. 2.3

При нулевом внешнем поле индукция, вообще говоря, ненулевая. Поэтому, подобно постоянному магниту, образец из сегнетоэлектрика может проявлять свойства диполя и в отсутствие внешнего поля. Строго говоря, это свойство присуще более широкому классу веществ, именуемых пьезоэлектриками; сегнетоэлектрики выделены среди них тем, что их дипольным моментом можно управлять посредством внешнего электрического поля.

Рассмотрим образец простейшей формы — палочку из сегнетоэлектрика. В принципе, даже лучше подходят на роль постоянных диполей электреты — некоторые органические смолы, которые, застывая в сильном электрическом поле, сохраняют «замороженной» поляризацию молекул. Такому состоянию соответствуют отличные от нуля как поле, так и индукция. Состояние вещества при этом не следует смешивать с откликом на внешнее поле; точки A и B на рис. 2.3 не отвечают случаю

¹⁾ Существенно также и то обстоятельство, что ферроэлектрические свойства веществ обусловлены нетривиальной кристаллической структурой, тогда как ферромагнетизм — сугубо квантовый эффект выстраивания электронов. Но это различие на феноменологическом уровне, т. е. на языке усредненных полей, никак не проявляется.

постоянного диполя. Обратимся к рис. 2.4. В пустоте силовые линии E и D идентичны, поскольку сами эти величины отличаются всего лишь коэффициентом ε_0 . Но внутри образца

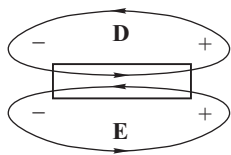


Рис. 2.4

поведение их различно. Нормальная составляющая вектора D на торцах непрерывна, так что образуются замкнутые петли. В то же время поле E должно удовлетворять условию $\oint E d\mathbf{l} = 0$ при интегрировании по любому замкнутому контуру; из этого соотношения и рис. 2.4 следует, что внутри образца векторы D и E направлены противоположным образом. Если мы имеем дело с сегнетоэлектриком, постоянный электродиполь должен попадать на рис. 2.3 в четные квадранты на кривой гистерезиса.

Из этого примера можно видеть, сколь поверхностным является популярное утверждение, что индукция — поле свободных зарядов. В нашем случае таковые вообще отсутствуют, но индукция нулю не равна. Связь ее с электрическим полем, как можно усмотреть из того же примера, вовсе не обязательно сводится к константе. Здесь, например, даже $\oint D d\mathbf{l} \neq 0$.

С реальными палочками из сегнетоэлектрика либо электрета возникает проблема экранировки. В воздухе или, скажем, в воде всегда есть свободные ионы, которые, оседая на полюсах образца, нейтрализуют торцевые заряды диполя, и перед каким-либо использованием последнего его приходится специальным образом очищать.

Мы лишены возможности в рамках курса электродинамики останавливаться более подробно на физике пьезо- либо сегнетоэлектриков, но как мы видим, они стоят упоминания даже в самом общем контексте, поскольку демонстрируют нетривиальность проблемы электрического поля в веществе.

2.6. Энергия электрического поля в вакууме и в веществе

Понятие энергии поля в начальных курсах физики по традиции дается на излишне примитивном уровне, и в результате закон сохранения энергии превращается в символ веры, что естественным наукам противопоказано. В действительности каждая новая область физики, каждый новый круг явлений в ходе исследования ставит этот вопрос заново. Мы ввели в курсе механики потенциальную и кинетическую энергию и установили, что в идеальной замкнутой механической системе сумма их есть инвариант или, как принято говорить, интеграл движения. Тогда же мы оговорили возможность перехода энергии в другие ви-

ды — это соответствует наиболее общему случаю диссипативных систем. И, строго говоря, пока мы не расширяем самого понятия энергии, у нас не может быть оснований постулировать какой-то универсальный закон сохранения. Но такое расширение возможно производить лишь шаг за шагом, по мере накопления новой экспериментальной информации. А значит, и закон сохранения энергии, если рассматривать его именно как физический закон, а не философский принцип, должен с развитием физики постоянно проверяться и дополняться. До настоящего времени этот процесс проходил вполне успешно.

Таким образом, корректная постановка вопроса об энергии электрического поля должна быть такова: можно ли ввести такую физическую величину, которая в сумме с механической энергией была бы инвариантна, а в отсутствие электромагнитных проявлений исчезала бы, оставляя справедливым известный нам закон сохранения механической энергии?

Реализацию этого принципа начнем с общеизвестного примера. Поставим мысленный эксперимент с плоским конденсатором, «раздвигая» его пластины от почти нулевого зазора до некоторой величины δ . Пусть заряд каждой пластины $\pm Q$, тогда поверхностная плотность заряда $\sigma = Q/S$, а поле каждой из пластин равно $E = \sigma/(2\varepsilon_0) = Q/(2S\varepsilon_0)$. Движение пластин будем считать адиабатически медленным, чтобы можно было пренебречь их кинетической энергией. Тогда работа, совершенная в ходе такого мысленного эксперимента, может трактоваться как энергия заряженного конденсатора:

$$W = F \delta = Q E \delta = \frac{Q^2 \delta}{2S\varepsilon_0} = \frac{Q^2}{2C}. \quad (2.21)$$

Ту же самую величину можно представить как $CU^2/2$ или $QU/2$. Все эти формы эквивалентны. Обратим внимание на то, что в наших вычислениях мы умножали заряд пластины не на полное поле в зазоре, но лишь на поле другой пластины, ибо собственное поле каждой из них действует на ее разрыв, но никак не на смещение. Результат (2.21) можно получить и по-другому. Представим себе конденсатор, изначально не заряженный, и начнем мысленно заряжать его, перенося заряд с одной пластины на другую малыми порциями dQ :

$$\delta A = dQ \cdot U = Q \frac{dQ}{C} \implies W \equiv A = \frac{Q^2}{2C}.$$

Следующий шаг будет связан с попыткой переосмысления формулы (2.21). Вводя объем конденсатора $V = S \delta$, ее можно переписать в виде

$$W = C \frac{(E\delta)^2}{2} = \frac{1}{2} \varepsilon_0 E^2 V. \quad (2.22)$$

В этой форме энергия заряженного конденсатора представляется как энергия электрического поля, заполняющего объем конденсатора. Можно попытаться ввести плотность энергии электрического поля

$$w = \frac{\varepsilon_0 E^2}{2}. \quad (2.23)$$

Как мы вскоре убедимся, именно эта форма представления электрической энергии оказывается наиболее глубокой и содержательной.

Вернемся к первичному представлению о точечных зарядах. Возьмем в качестве примера систему из двух точечных зарядов, пометив их индексами 1 и 2. По смыслу введенного нами понятия потенциала мы можем рассматривать энергию взаимодействия как энергию заряда 1 в поле заряда 2 или, напротив, энергию заряда 2 в поле заряда 1:

$$W = q_1 \varphi_{(21)} = q_2 \varphi_{(12)}.$$

Запишем энергию взаимодействия в симметричной форме:

$$W \equiv \frac{1}{2} (q_1 \varphi_{(21)} + q_2 \varphi_{(12)}). \quad (2.24)$$

Исходя из этого простого примера и принимая во внимание определение (1.16), мы можем обобщить (2.24) на произвольную систему точечных зарядов:

$$W = \frac{1}{2} \sum_{i \neq k} \frac{q_i q_k}{r_{ik}}. \quad (2.25)$$

Можно предложить и другое обобщение. Пусть электрический заряд распределен в пространстве с некоторой плотностью $\rho(\mathbf{r})$. Тогда, вместо (2.24) и (2.25), можем написать

$$W = \frac{1}{2} \int_V \rho(\mathbf{r}) \varphi(\mathbf{r}) dV. \quad (2.26)$$

Обратимся к уравнению Пуассона, записав его в виде

$$\varphi (\varphi''_{xx} + \varphi''_{yy} + \varphi''_{zz}) = -\frac{\varphi \rho}{\varepsilon_0}.$$

Проинтегрируем по всему пространству любое из слагаемых левой части методом интегрирования по частям — и с учетом того, что на бесконечности потенциал и поле любой системы зарядов обращаются в нуль

$$\begin{aligned} \int \varphi \varphi''_{xx} dx dy dz &= \int dy dz \varphi \frac{\partial \varphi'_x}{\partial x} dx = - \int (\varphi'_x)^2 dx dy dz = \\ &= - \int E_x^2 dV \text{ и т. д.} \end{aligned}$$

Отсюда следует:

$$W = \frac{1}{2} \int_V \rho(\mathbf{r}) \varphi(\mathbf{r}) dV = \frac{\varepsilon_0}{2} \int (E_x^2 + E_y^2 + E_z^2) dV, \quad (2.27)$$

а это совершенный эквивалент формулы (2.23), но получен он уже не в качестве догадки. Теперь мы можем считать задачу решенной. Мы пришли к выводу о том, что электрическая энергия есть энергия электрического поля, распределенная в пространстве с плотностью $\varepsilon_0 E^2/2$. Отсюда, в частности, следуют три важных качественных следствия.

1. В соответствии со знаменитой эйнштейновской формулой $W = mc^2$ электрическое поле имеет массу.

2. Энергия поля всегда положительна, безотносительно к знаку поля или создающих его зарядов.

Второе утверждение требует небольшого комментария. Стандартная школьная формула $W = q_1 q_2 / (4\pi \varepsilon_0 r_{12})$ учитывает лишь взаимодействие зарядов друг с другом, но не частей каждого из них между собой. Полное поле $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$, соответственно, $E^2 = E_1^2 + E_2^2 + 2(\mathbf{E}_1 \cdot \mathbf{E}_2)$. Таким образом полная энергия поля равна

$$W = W_1 + W_2 + \varepsilon_0 \int (\mathbf{E}_1 \cdot \mathbf{E}_2) dV.$$

Энергия взаимодействия представлена третьим членом этой суммы, знак которого и в самом деле не предопределен.

3. Энергия поля неаддитивна, иными словами, энергия суммы полей не равна сумме энергий. Таким образом, для электрической энергии не существует принципа суперпозиции.

Обращаясь к случаю поля в веществе, можно прежде всего сразу же написать ответ для простейшего случая линейной связи между полем и индукцией. Опираясь на приведенные выше примеры с полем в конденсаторе, нетрудно убедиться в том, что результат (2.21) остается в силе. Соответственно, формула (2.22) нуждается в минимальной модификации:

$$w = \frac{\varepsilon \varepsilon_0}{2} E^2. \quad (2.28)$$

В очень многих практически интересных случаях этой модификации, т. е. коэффициента ε , оказывается вполне достаточно. Соотношение (2.28) можно переписать также, используя в качестве параметра индукцию, которую в случае линейной связи как раз можно трактовать как поле свободных зарядов:

$$w = \frac{D^2}{2\varepsilon\varepsilon_0}.$$

Но, как мы видели хотя бы на примере сегнетоэлектриков, физика поля в веществе случаем линейной связи отнюдь не ис-

черпывается. Более того, в общем случае и расширение понятия энергии, о котором говорилось в начале текущего параграфа, — задача не простая. В лучшем случае энергию можно представить в следующей форме:

$$W = W_{(m)} + W_{(E)} + W_{(m-E)},$$

т. е. в виде суммы энергии вещества, энергии поля и энергии взаимодействия поля с веществом. Но даже и такое представление не всегда корректно, поэтому надо понимать, что наши простые и легко обозримые ответы все же представляют собой некоторые простые предельные случаи; в частности, подразумевается, что поле не слишком велико. Это вовсе не означает, что данный вопрос вообще не поддается исследованию. Просто в общем случае мы имели бы дело с серьезной физической задачей, включающей физику строения вещества на должном уровне.

Попытаемся дать хотя бы минимальное обобщение формулы (2.28). Умножим правую и левую части формулы (2.3), понимаемой как уравнение для поля в веществе, на φ :

$$\rho \varphi = \varphi \operatorname{div} \mathbf{D}.$$

Далее, воспользуемся следующей цепочкой тождественных преобразований:

$$\begin{aligned} \operatorname{div}(\varphi \mathbf{D}) &= \\ &= \left(\mathbf{e}_1 \frac{\partial}{\partial x} + \mathbf{e}_2 \frac{\partial}{\partial y} + \mathbf{e}_3 \frac{\partial}{\partial z} \right) (\varphi D_x \mathbf{e}_1 + \varphi D_y \mathbf{e}_2 + \varphi D_z \mathbf{e}_3) = \\ &= \left(D_x \frac{\partial \varphi}{\partial x} \right) + \left(\varphi \frac{\partial D_x}{\partial x} \right) + \dots = \varphi \operatorname{div} \mathbf{D} + \mathbf{D} \nabla \varphi, \end{aligned}$$

откуда следует (см. (2.3)):

$$\rho \varphi = \operatorname{div}(\varphi \mathbf{D}) - \mathbf{D} \nabla \varphi = \operatorname{div}(\varphi \mathbf{D}) + (\mathbf{D} \mathbf{E}).$$

Воспользуемся еще раз формулой (2.26):

$$W = \frac{1}{2} \int \rho \varphi dV = \frac{1}{2} \int \operatorname{div}(\varphi \mathbf{D}) dV + \frac{1}{2} \int (\mathbf{D} \mathbf{E}) dV.$$

Здесь первый интеграл можно преобразовать по теореме Гаусса в поток вектора $\varphi \mathbf{D}$ через некоторую удаленную поверхность. Площадь этой поверхности $\propto R^2$, тогда как подынтегральное выражение даже в случае нескомпенсированного заряда $\propto R^{-3}$, так что при интегрировании по всему пространству мы можем пренебречь этим интегралом, а из второго интеграла немедленно следует искомое обобщение выражения для плотности энергии электрического поля

$$w = \frac{1}{2} (\mathbf{D} \mathbf{E}). \quad (2.29)$$

Задачи

1. Как изменится энергия заряженного конденсатора с вакуумным зазором, если заполнить последний диэлектрической жидкостью с постоянной ϵ ?

Решение. Ответ зависит от режима процесса заполнения. Емкость конденсатора увеличится в ϵ раз. Если конденсатор был предварительно заряжен, а при заполнении от источника отключен, то мы должны считать инвариантом заряды на обкладках $\pm Q$. Тогда энергия $Q^2/2C$ уменьшится в ϵ раз. Напротив, если при заполнении конденсатор остается подключенным к источнику напряжения, то инвариантной останется разность потенциалов между обкладками, соответственно, энергия $CU^2/2$ увеличится в ϵ раз. Тот же результат можно получить, сравнивая плотность энергии в пустом конденсаторе и заполненном с помощью формулы (2.29), но в первом случае $D = \text{inv}$, $E \propto \epsilon^{-1}$, а во втором — $E = \text{inv}$, $D \propto \epsilon$.

Полный энергетический баланс подводится следующим образом. В первом случае поле конденсатора втягивает диэлектрик, совершая над ним работу; если потенциальная энергия диэлектрика в поле тяжести не изменяется, работа переходит в кинетическую энергию и в конечном счете в тепло. Во втором случае та же работа совершается источником напряжения (течет ток, изменяя заряды обкладок), одновременно источник увеличивает и энергию поля.

2. По сфере радиуса R распределен равномерно заряд Q . Определить давление на поверхность сферы, обусловленное взаимодействием зарядов.

Решение. Пожалуй, наиболее поучительно решение этой задачи, следующее из энергетических соображений. Энергия заряженной сферы $W = Q\varphi/2 = Q^2/(8\pi\epsilon_0 R)$. Мысленно дадим радиусу сферы малое приращение: $R \rightarrow R + \delta R$. Приращение энергии свяжем с работой искомого давления:

$$W + \delta A \approx W - \mathcal{P} \cdot 4\pi R^2 \delta R \equiv \frac{Q^2}{8\pi\epsilon_0(R + \delta R)} \approx W - \frac{Q^2 \delta R}{8\pi\epsilon_0 R^2}.$$

Отсюда получаем $\mathcal{P} = Q^2/(2(4\pi)^2\epsilon_0 R^4)$.

Интересно отметить, что, просто умножив поверхностную плотность заряда на сфере на величину поля вне сферы, мы бы получили ответ в два раза больший. Причина достаточно очевидна: внутри сферы поле равно нулю, а в тонком заряженном слое меняется от нуля до $Q^2/(4\pi\epsilon_0 R^2)$. Поэтому сила, действующая на элемент площади сферы, в действительности определяется некоторым интегралом от величины $E\rho dV$, взятым в радиальном направлении. Тем не менее, ответ для тонкого слоя оказывается универсальным.

3. Пластина пьезоэлектрика толщины $2d$ вследствие неоднородной деформации поляризована так, что поляризация ее в центре равна P_0 , направлена вдоль оси Ox и изменяется по закону $P = P_0(1 - x^2/d^2)$, где x отсчитывается от плоскости симметрии пластины. Определить напряженность электрического поля внутри и вне пластины, а также разность потенциалов U между ее поверхностями.

Ответ: $E(x) = -P_0\epsilon_0^{-1}(1 - x^2/d^2)$, $U = \int_{-d/2}^{d/2} E dx = 4P_0d/(3\epsilon_0)$; вне

пластинки поле равно нулю.

4. Определить электрическую энергию ядра урана ${}^{235}_{92}\text{U}$ при равномерном распределении заряда $Z = 92$ по объему сферы радиуса $R = 1,3 \cdot 10^{-15} \text{ A}^{1/3} \text{ м}$, где $A = 235$ — массовое число.

Ответ: $W = 3Z^2e^2/20\pi\epsilon_0 R \approx 1,5 \cdot 10^{-10} \text{ Дж}$.

Глава 3

ПОСТОЯННЫЙ ТОК

3.1. Ток как движение зарядов

Понятие постоянного тока столь же существенно в начальном курсе физики, как и понятие статического электричества. Исторически оба эти раздела электродинамики строились параллельно и в какой-то мере независимо; в единую науку их удалось объединить лишь в XIX веке.

По своим экспериментальным проявлениям электрический ток существенно отличается от статического электричества — как известно, для токонесущих систем (сред) характерны тепловой эффект, электрохимический эффект, магнитное поле тока. Как мы увидим далее, даже электрическая цепь с постоянным током не является полностью стационарной системой. Просто в системе, принципиально открытой, возможно не только статическое, но и так называемое *динамическое равновесие*, которое, грубо говоря, можно охарактеризовать формулой «приход равен расходу», а на языке математики это означает, что постоянство некоторой физической величины (например тока), или, что то же, ее нулевая временная производная, обеспечивается балансом положительной и отрицательной производных, обусловленных разными физическими процессами. Применительно к случаю постоянного тока можно привести такой пример: электроны в проводнике движутся с постоянной скоростью, если их ускорение в электрическом поле компенсируется трением (сопротивлением).

Вводя понятие электрического тока, мы будем исходить из известного факта, базирующегося на огромном массиве экспериментальной информации: *электрический ток есть направленное движение электрических зарядов*. (При хаотическом их движении, типа броуновского, ток в среднем может отсутствовать.)

В общем случае электромагнитные свойства токонесущей системы, конечно, включают в себя всю электростатику и полностью в нее переходят в пределе, когда заряды можно считать неподвижными. Упомянутое выше существенное различие эффектов обусловлено еще и тем, что типичные токовые системы, наиболее интересные в практическом отношении, обладают свойством квазинейтральности.

Начнем с примера, известного каждому школьнику. Линейные проводники, по которым текут параллельные токи, притягиваются. (Почему — об этом мы еще будем говорить в следующей главе.) А вот два параллельных электронных пучка всегда отталкиваются друг от друга. Но это как раз и не удивительно, поскольку прямо следует из законов электростатики. Отсюда можно заключить, что известное нам электростатическое (кулоновское) взаимодействие сильнее того взаимодействия между параллельными токами, которое еще предстоит изучить. Это верно и для металлического токонесущего проводника, но сам этот проводник, если представить себе его микроскопическое строение, является системой, как говорят, *квазинейтральной*. Это означает, что истинной нейтральности в веществе проводника нет (без движения заряженных частиц не могло бы быть и тока), но при усреднении даже по очень малому объему вещества эффективный заряд оказывается нулевым из-за очень высокой концентрации заряженных частиц — $\sim 10^{28} - 10^{29}$ част./м³. Таким образом, электростатическое взаимодействие между проводниками оказывается пренебрежимо малым (говорят об электрическом поле очень высокой мультипольности). Как следствие, эффекты, связанные с протеканием электрического тока, не маскируются электростатическими проявлениями.

Еще одно чрезвычайно важное упрощение при подходе к экспериментальному изучению и теоретическому описанию проводников с током следует из того, что для технических приложений особенно интересны *линейные*, или, как говорят, *квазиодномерные*, токонесущие системы — электрические цепи. (Еще один термин для обозначения этого класса объектов — токонесущие системы с сосредоточенными параметрами.) Для них типично рассмотрение эффектов или вообще интегральных, или зависящих лишь от одной координаты — вдоль провода, а распределение параметров по сечению проводника не имеет значения.

Определив таким образом характерные приближения в задачах о протекании электрического тока, мы все же постараемся ввести основные понятия и законы по возможности в общем виде, не связывая себя приближениями.

Пока и поскольку мы имели дело с током в линейном проводнике, можно было определить *силу тока* (или просто «ток») как заряд, протекающий в единицу времени через поперечное сечение проводника:

$$I = \frac{dq}{dt}. \quad (3.1)$$

Единица тока — ампер (1 А = 1 Кл/с) относится к числу основных единиц системы СИ. Чтобы избежать затруднений с определением тока (понятие переноса заряда в квазинейтральном проводнике требует некоторого напряжения фантазии), дадим

более конкретную иллюстрацию. Пусть, например, перемещаться в проводнике могут только электроны — так это и происходит в металлах — и пусть в единицу времени через поперечное сечение провода проходит N электронов. Тогда ток равен $I = Ne$. Направление тока по традиции определялось «от плюса к минусу», т. е. в нашем примере оно как раз противоположно движению электронов. Если же мы имеем дело сразу с несколькими сортами носителей заряда — как, например, в полупроводниках или электролитах, — то мы можем эту формулу обобщить:

$$I = \sum_{\alpha} N_{\alpha} Z_{\alpha} e, \quad (3.2)$$

где α — сорт носителя, Z — его зарядовое число, e — заряд электрона (заметим, что для самого электрона $Z = -1$). При этом за направление тока в соответствии с традицией принимается направление движения положительных носителей заряда.

Если нас интересует более детальная картина, мы вводим еще и *плотность тока* — величину, характеризующую распределение тока по сечению проводника:

$$j = \frac{d^2 q}{dS dt}; \quad \int j dS = I. \quad (3.3)$$

Очевидно, единица плотности тока — 1 А/м^2 . Попробуем и для этой величины получить аналог формулы (3.2). Если концентрация данного сорта носителей равна n_{α} , а скорость их в направлении, принятом за направление тока, — v_{α} , то в единицу времени единицу площади сечения проводника пересекут $n_{\alpha} v_{\alpha}$ частиц данного сорта (все частицы, заполняющие цилиндр с единичной площадью основания и высотой v_{α}).

Таким образом, сравнивая (3.2) и (3.3), приходим к следующей формуле:

$$j = \sum_{\alpha} n_{\alpha} Z_{\alpha} v_{\alpha} e. \quad (3.4)$$

Представляется естественным следующий шаг. Откажемся от условия квазиодномерности проводника. Пусть движение зарядов происходит в произвольной трехмерной проводящей среде. В этом случае выражение для плотности тока принимает вид

$$\mathbf{j} = \sum_{\alpha} n_{\alpha} Z_{\alpha} \mathbf{v}_{\alpha} e. \quad (3.5)$$

Теперь это — векторная величина. Если единственный род носителей — электроны, то $\mathbf{j} = -nev$, где все величины в правой части относятся к электронам проводимости.

3.2. Закон сохранения заряда. Постоянный ток

В первой главе мы уже говорили об одном из самых фундаментальных законов природы — законе дискретности заряда: любой заряд в любых проявлениях всегда кратен заряду электрона.

Столь же фундаментален и закон сохранения заряда. Он ниоткуда не может быть выведен и не может быть установлен ни в единичном эксперименте, ни даже в конечной серии каких-либо экспериментов. Он следует из всего накопленного на сегодняшний день опыта экспериментальной физики.

Электрический заряд есть строго сохраняющаяся величина. Любые изменения зарядов любых физических тел происходят как следствие разделения зарядов, но не их возникновения либо уничтожения.

В ядерных превращениях возможно и рождение заряженных частиц, но лишь — парами, при сохранении полного нулевого заряда. В атомной, молекулярной и макроскопической физике сохранение заряда сводится к сохранению числа частиц любого сорта, а значит — и их суммарного заряда.

К этому следует добавить еще один очень важный экспериментальный факт — нейтральность Вселенной, и более того, высокую степень нейтральности всех макроскопических тел. Заряды макротел, с которыми мы реально имеем дело в природе или в технике, всегда суть малые возмущения на фоне очень высокой степени нейтральности тела. Уединенный металлический шарик радиуса 1 см, потерявший сотую долю электронов, оказался бы заряжен до потенциала $\sim 10^{15}$ В, а его энергозапас представлял бы величину $\sim 10^{19}$ Дж!

Как известно, закон Кулона и закон всемирного тяготения функционально тождественны, но электростатическое взаимодействие гораздо сильнее, поэтому, к примеру, при взаимодействии двух электронов гравитационная составляющая силы ничтожно мала в сравнении с кулоновской. Но, вследствие нейтральности, макроскопическая динамика Вселенной определяется все-таки исключительно гравитационным взаимодействием.

Попробуем придать закону сохранения заряда динамическую форму. Представим себе некоторый элемент объема проводящей среды, ограниченный произвольной замкнутой поверхностью S (рис. 3.1). Изменение полного заряда, заключенного внутри поверхности, может быть следующим образом связано с зарядом, протекающим в единицу

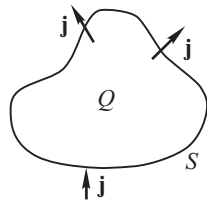


Рис. 3.1

дующим образом связано с зарядом, протекающим в единицу

времени через границу (током):

$$\frac{\partial Q}{\partial t} = - \int j_{\perp} dS = - \oint \mathbf{j} d\mathbf{S} = - \int \operatorname{div} \mathbf{j} dV \quad (3.6)$$

(мы преобразовали правую часть по теореме Гаусса). Но и левая часть может быть представлена в виде объемного интеграла, поскольку $Q \equiv \int \rho dV$. Элемент объема может быть выбран совершенно произвольно, поэтому равенство интегралов влечет за собой равенство подинтегральных выражений. Как следствие, получаем локальную (или дифференциальную) форму уравнения (3.6):

$$\frac{\partial \rho}{\partial t} + \operatorname{div} \mathbf{j} = 0. \quad (3.7)$$

Уравнения (3.6), (3.7) и выражают в электродинамике закон сохранения заряда. В математическом отношении они эквивалентны гидродинамическому уравнению непрерывности, что вполне естественно с точки зрения физики дела: в гидродинамике сохраняется масса вещества, а в электродинамике — заряд. Изменение любой из этих величин представимо в форме некоторого течения.

Теперь мы можем на новом уровне вернуться к приближению постоянного тока. В общем случае трехмерного течения зарядов мы должны потребовать выполнения условий

$$\mathbf{j}(t) = \text{const}, \quad \frac{\partial \rho}{\partial t} = 0.$$

Из последнего, в частности, следует

$$\operatorname{div} \mathbf{j} = 0. \quad (3.8)$$

Условие (3.8) часто отождествляют с условием постоянства тока. Но, как следует из нашего рассмотрения, оно является необходимым, но не достаточным. Например, оно выполняется с высокой точностью для переменных, но квазистационарных токов (см. главы 8, 10), если в цепи нет конденсатора.

Особо следует отметить, что условие постоянства тока подразумевает равенство нулю временной производной плотности заряда, но не самой плотности заряда. В неоднородных электрических цепях, а тем более при трехмерном протекании тока в среде, она вполне может быть отлична от нуля в отдельных точках или даже во всей проводящей среде. Устанавливается такое возмущение нейтральности на нестационарной стадии, следующей непосредственно за включением тока.

В случае системы с сосредоточенными параметрами — электрической цепи, составленной из проводов однородного сечения, — формулу (3.8) можно заметно упростить. Считая радиус кривизны провода много большим его поперечных размеров, бу-

дем рассматривать проводник как квазиодномерный. Тогда в (3.8) из всех производных остается только одна — по координате вдоль провода:

$$\frac{\partial j_{\parallel}}{\partial l} = 0.$$

То же самое условие можно получить напрямую из постоянства полного тока в сечении:

$$I = \text{const} = \int j_{\parallel} dS,$$

откуда $j_{\parallel} = \text{const}$. Если же проводник по длине неоднороден, из постоянства тока следует неинвариантность плотности тока — она должна меняться обратно пропорционально сечению провода.

3.3. Закон Ома. Закон Джоуля–Ленца

По-видимому, самый популярный закон электродинамики, известный даже людям, от физики и электротехники весьма далеким, — это закон линейной связи между током в одномерном проводнике и напряжением (разностью потенциалов) на его концах, установленный в 1826 г. немецким физиком Георгом Омом (1787–1854):

$$I = \frac{U}{R}. \quad (3.9)$$

Величина R в этом соотношении называется *сопротивлением* проводника и измеряется в системе СИ в омах: $1 \text{ Ом} = 1 \text{ В}/1 \text{ А}$. В отличие от электростатики, единицы СГС для данного круга задач чрезвычайно неудобны; скажем, $1 \text{ Ом} \approx 1,1 \cdot 10^{-12} \text{ СГС}$.

Для нашего курса формулировка (3.9) все же слишком примитивна. Мы заменим ее локальным эквивалентом, который в случае однородного распределения плотности тока по сечению линейного проводника тривиальным образом сводится к (3.9):

$$\mathbf{j} = \lambda \mathbf{E} \equiv \frac{\mathbf{E}}{\rho}. \quad (3.10)$$

Величина λ — иногда ее обозначают буквой σ — называется *проводимостью вещества*, а обратная ей величина ρ — *удельным сопротивлением*. Интегрируя выражение (3.10) по сечению проводника (тем самым мы от j переходим к I), а затем — по длине его (переходим от E к U), мы и в самом деле возвращаемся к (3.9); при этом, в соответствии с курсом физики средней школы,

$$R = \frac{\rho l}{S_{\perp}} = \left(\frac{\lambda S_{\perp}}{l} \right)^{-1}.$$

В отличие от (3.9), формула (3.10) устанавливает причинно-следственную связь между электрическим полем и током, ко-

торый оно порождает, в точке (точнее — в физически бесконечно малом объеме). В принципе, мы могли бы попытаться обосновать эту связь с помощью следующих рассуждений.

Какова бы ни была функциональная зависимость между \mathbf{j} и \mathbf{E} , в пределе достаточно слабых полей ее можно разложить в ряд Тэйлора, и тогда первый не исчезающий член такого разложения как раз и дает $\mathbf{j} \propto \mathbf{E}$. Разумеется, параллельность векторов \mathbf{j} и \mathbf{E} отсюда не следует; она следует из изотропности проводящей среды — просто, помимо направления поля, в задаче нет больше никакого другого выделенного направления. В более общем случае, например для анизотропного кристалла, связь должна быть тензорной, $j_i = \sum_k \lambda_{ik} E_k$, но по-прежнему линейной.

Однако доказательность этих рассуждений иллюзорна. Во-первых, не каждую функцию можно разложить в ряд Тэйлора, во-вторых, параметр разложения заранее не ясен, например, первый не исчезающий член при $E \rightarrow 0$ может оказаться пропорциональным $E^{1/2}$ или, скажем, все члены разложения обратятся в нуль вплоть до E^3 . В-третьих, в любом случае остается вопрос: сколь малыми должны быть поля и токи, чтобы зависимости (3.9), (3.10) оставались линейными? Представит ли этот предел хоть какой-то интерес?

Эксперимент отвечает на эти вопросы следующим образом: для твердых проводников формулы (3.9), (3.10) справедливы в достаточно широком диапазоне практически интересных напряжений и токов. Несколько уже область их применимости для электролитов, еще уже — для газов и плазмы, но и в этих средах закон Ома остается в числе самых фундаментальных. Как и все основные законы природы, закон Ома следует не из рассуждений, но из опыта; экспериментом же определяется и область его применения.

Формулу (3.10) можно попытаться вывести из законов механики для заряженной частицы в проводящей среде. Это достаточно просто сделать для жидкой или газообразной среды, моделируя трение электрона в среде посредством постоянной эффективной частоты столкновений $\nu_{\text{эф}}$. Столкновения могут происходить с ионами в электролите, с дефектами решетки в твердом теле и т. д. Для определенности отнормируем $\nu_{\text{эф}}$ так, как если бы электрон при каждом эффективном столкновении полностью терял свой импульс. Уравнение движения электрона

$$\frac{d}{dt} m\mathbf{v} = -e\mathbf{E} - \nu_{\text{эф}} m\mathbf{v}$$

в пределе стационарного течения электронов $d(m\mathbf{v})/dt \equiv 0$ имеет простое решение $\mathbf{v} = -e\mathbf{E}/\nu_{\text{эф}}m$. Подставляя его в

соотношение $\mathbf{j} = -nev$ и сравнивая с (3.10), получаем

$$\lambda = \frac{ne^2}{m\nu_{\text{эф}}}. \quad (3.11)$$

Аналогия между формулами (3.11) и (3.10) позволяет сделать ряд замечаний достаточно общего характера о применимости закона Ома.

1. В классическом виде закон Ома справедлив постольку, поскольку можно пренебречь магнитным полем в уравнении движения электрона (что мы и сделали). Если же магнитное поле тока существенно, то закон, как минимум, нуждается в модификации.

2. В общем случае произвольной природы сил трения величина $\nu_{\text{эф}}$ должна, в принципе, зависеть от скорости электрона в среде. Мы же посчитали ее постоянной, а это значит, что токовая скорость предполагается много меньше тепловой скорости электронов (при комнатной температуре — 60–70 км/с).

3. Формула (3.11) — и пока только она — подразумевает стационарность электронного течения (постоянный ток). Но ее, вместе с (3.9), (3.10), можно попытаться перенести и на более широкий класс явлений. Так, мы можем считать ток «безынерционным» и пользоваться законом Ома, если все величины медленно меняются во времени: $d(m\mathbf{v})/dt \ll \nu_{\text{эф}}m\mathbf{v}$. Это не единственное условие и не самое жесткое, но сама эта возможность очень важна.

4. Течение «газа» электронов проводимости возможно не только вследствие действия на них электрического поля, но и вследствие перепада давления — как в гидродинамике. И напротив, при одновременном действии поля и градиента температуры тока может и не быть. Закон Ома подразумевает, что все подобные (термоэлектрические) эффекты малы в сравнении с ускорением электронов в электрическом поле. Мы видим, что поле должно быть ограничено не только сверху, но и снизу. Более или менее корректное количественное сравнение нам придется отложить до раздела, посвященного строению вещества.

Опираясь на вывод (3.11) и на всю эту простую модель электрического сопротивления, мы можем рассмотреть также эффект джоулева тепловыделения, известный в школьной физике в форме закона Джоуля–Ленца (по имени английского физика Дж. Джоуля (1818–1889) и русского физика Э.Х. Ленца (1804–1865)). В соответствии с механикой, работу в единицу времени при перемещении одной частицы можно оценить следующим образом:

$$\frac{dA}{dt} = -(\mathbf{F}_{\text{тр}}\mathbf{v}) = m\nu_{\text{эф}}v^2.$$

Как правило, работа силы трения переходит в тепло (возможны и другие каналы потерь, но мы здесь не будем на этом

останавливаться). Для проводящей среды получаем следующее выражение для тепловыделения в единице объема:

$$\frac{dw}{dt} = nm\nu_{\text{эф}} v^2 = (nev)^2 \frac{m\nu_{\text{эф}}}{ne^2} = \frac{j^2}{\lambda} = \rho j^2. \quad (3.12)$$

Этот тепловой эффект принято называть *джоулевым теплом*. В однородном проводнике длины l с поперечным сечением S_{\perp} справедливо следующее выражение для тепловыделения:

$$\frac{dQ}{dt} = S_{\perp} l \cdot \rho j^2 = I^2 \frac{\rho l}{S_{\perp}} \equiv RI^2, \quad (3.13)$$

а это не что иное, как *закон Джоуля–Ленца*. Пока и поскольку речь идет о постоянном токе, тепловыделение с помощью закона Ома можно представить в различной форме:

$$\frac{dQ}{dt} = RI^2 = \frac{U^2}{R} = IU,$$

однако, если мы выходим за пределы этого приближения, остаются в силе только формулы (3.12), (3.13). Это дает основание для введения понятия активного сопротивления, очень важного для описания переменных токов и связанного именно с процессом диссипативной природы — джоулевым тепловыделением.

В заключение настоящего параграфа рассмотрим еще один важный вопрос, касающийся взаимосвязи электростатики и законов протекания тока. Мы будем базироваться на явлении, которое принято называть *максвелловской релаксацией* (по имени Джеймса К. Максвелла (1831–1879)).

Предположим, что в некоторой проводящей среде имеет место закон Ома (3.10); при этом мы не предполагаем стационарности протекания тока и толкуем данный закон расширенно — т. е. полагая все временные зависимости достаточно медленными. Пусть наша среда характеризуется постоянной проводимостью λ и диэлектрической проницаемостью $\varepsilon = 1$. Перепишем (3.7) в виде

$$\frac{\partial \rho}{\partial t} + \operatorname{div} \lambda \mathbf{E} = 0.$$

Вынося λ за знак всех производных и подставляя $\operatorname{div} \mathbf{E} = \rho/\varepsilon_0$, получаем

$$\frac{\partial \rho}{\partial t} + \frac{\lambda}{\varepsilon_0} \rho = 0. \quad (3.14)$$

Решение уравнения (3.14) хорошо известно:

$$\rho = \rho_0 \cdot e^{-\lambda t/\varepsilon_0}. \quad (3.15)$$

В простейших случаях соотношение (3.15) правильно описывает процесс релаксации заряда за счет протекания тока. Но из него следует, что в стационарном состоянии (включая слу-

чай постоянного тока) $\rho \equiv 0$. Можно рассуждать и по-другому: из $\partial \rho / \partial t = 0$ вытекает $\operatorname{div} \mathbf{j} = 0$. Поскольку $\lambda = \text{const}$, то и $\operatorname{div} \mathbf{E} = 0 \Rightarrow \rho = 0$. Между тем, это, строго говоря, не совсем так. Представим себе, в качестве простейшего примера, источник постоянного тока, нагруженный на резистор, с которым он соединен двумя проводниками, сопротивление которых пренебрежимо мало. Тогда между этими проводниками возникает разность потенциалов, равная падению напряжения в резисторе. И какова бы ни была их геометрия (параллельные провода, коаксиал и пр.), из теоремы Гаусса с неизбежностью следует, что проводники эти заряжены. Мы можем пользоваться законом Ома $\mathbf{j} = \lambda \mathbf{E}$ при условии, что заряд этот настолько мал (или проводимость так велика), что можно в первом приближении считать $\operatorname{div} \mathbf{E} = 0$ — и закон Ома будет справедлив с точностью, отвечающей такому приближению.

Таким образом, популярный в быту и инженерной практике термин «статическое электричество», оказывается, не лишенным смысла, если под электростатическими полями понимать поля, для которых $\operatorname{div} \mathbf{E} \neq 0$, а под омическими — обратный случай, который описывается уравнениями (3.8), (3.10).

3.4. Электродвижущая сила. Правила Кирхгофа

Обратимся к системам с сосредоточенными параметрами. Рассмотрим простейшую электрическую цепь, изображенную на рис. 3.2 а. Закон Ома в первоначальной формулировке (3.9) включает такое понятие, как напряжение или разность потенциалов на концах проводника. Если имеется в виду истинно постоянный ток, а значит и постоянное поле в проводнике, понятие потенциала представляется вполне корректным. И работа, совершаемая над зарядом на участке цепи, $\delta A = e \Delta \varphi$, тоже согласуется с таким пониманием. В состоянии динамического равновесия она должна быть равна диссипируемой энергии. Про-

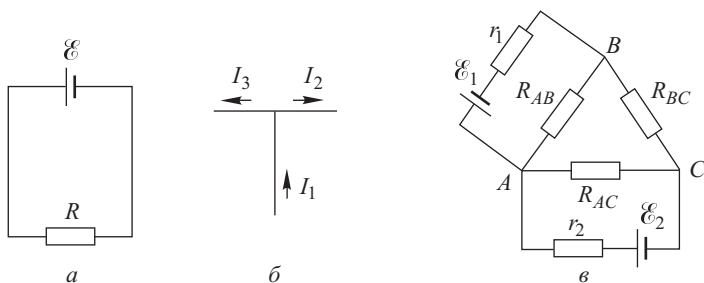


Рис. 3.2

суммируем обе величины по зарядам, прошедшим через данный участок цепи за время Δt :

$$\delta A = \delta q \cdot \Delta \varphi = I \Delta t \cdot U = R I^2 \Delta t,$$

что является очевидным тождеством. Но есть одно обстоятельство, которое не позволяет рассматривать поле в данной задаче как чисто потенциальное. Дело в том, что важнейшее свойство потенциальных полей,

$$\oint \mathbf{E} d\mathbf{l} = 0,$$

нарушается при интегрировании по контуру цепи:

$$\oint \mathbf{E} d\mathbf{l} = \lambda \cdot \oint \mathbf{j} d\mathbf{l} \neq 0,$$

потому что в отсутствие накопления заряда линии тока должны быть замкнуты (иначе как раз в точке разрыва получилось бы $dq/dt \neq 0$). В ситуации, изображенной на рис. 3.2 а, вся непотенциальность поля обусловлена источником тока (в данном случае — гальваническим элементом). Часто его невозможно столь же очевидным образом локализовать — например, в случае термоэлектричества. Но всегда источник тока характеризуется непотенциальностью, которую он вносит в электрическую цепь:

$$\mathcal{E} = \oint \mathbf{E} d\mathbf{l}. \quad (3.16)$$

Эта характеристика называется *электродвижущей силой* и обозначается как ЭДС. Оставаясь исключительно в рамках нашего курса, мы смогли бы вполне разобраться в работе лишь одного источника ЭДС — динамомашины, тогда как в примерах и задачах на постоянный ток более популярны как раз гальванические элементы. По необходимости оставляя их устройство и принцип работы вне рассмотрения, мы все же коснемся одного распространенного заблуждения. Имеет хождение следующая формулировка: «источник ЭДС обусловлен силами неэлектрической природы». Но так можно сказать только о гравитационных либо ядерных взаимодействиях. Химическое взаимодействие, фотоэффект, термо-ЭДС и тем более эффект динамо — все это явления вполне электрической природы. Общее их свойство, необходимое для любого источника тока, заключается в ином — нестационарности. Эта нестационарность может маскироваться длительным временем разрядки аккумулятора или динамическим равновесием в последнем звене энергетического производства — динамомашине, но она всегда имеет место. В противном случае, оставаясь в кругу чисто электромагнитных явлений, мы должны были бы стационарности электрического поля однозначно поставить в соответствие его потенциальность. К тому же выводу

приводит и другого рода соображение: ввиду неизбежной, хотя бы и сколь угодно слабой, диссипации, источник тока должен совершать работу.

Таким образом, постоянный ток не может рассматриваться иначе, как удобное приближение, адекватность которого обеспечивается сильным неравенством $\Delta t \ll \tau$, где Δt — характерный временной масштаб задачи, а τ — характерное время релаксации источника.

Для решения электротехнических задач, связанных с рассмотрением квазиодномерных цепей, подобных изображенной на рис. 3.2а, немецким физиком Густавом Р. Кирхгофом (1824–1887) были предложены специальные приемы, базирующиеся на фундаментальных физических законах. Собственно цепь на рис. 3.2а ни в каких иных рецептах, кроме закона Ома, не нуждается. Необходимость в дополнительных формальных правилах возникает при рассмотрении разветвленных цепей.

Первое правило Кирхгофа относится именно к точкам разветвления или узлам (рис. 3.2б). Из стационарности задачи следует, в частности, что в узловой точке не должен накапливаться заряд. Это значит, что заряд, поступающий в единицу времени в эту точку, должен быть равен заряду, уносимому из нее в единицу времени. Конкретно, для случая рис. 3.2б должно быть $I_1 = I_2 + I_3$. Общий рецепт строится по тому же принципу.

Будем считать, ради удобства, ток, втекающий в узел, положительным, а ток, из него вытекающий, — отрицательным. Тогда для любого узла сложной схемы должно иметь место соотношение

$$\sum_i I_i = 0. \quad (3.17)$$

Это и есть *первое правило Кирхгофа*.

Для иллюстрации второго правила Кирхгофа обратимся к рис. 3.2в. Любой замкнутый контур, который мы произвольным образом выделили бы из этой разветвленной схемы, должен был бы удовлетворить условию (3.16). Интеграл от потенциального поля по замкнутому контуру дал бы нуль, а фактически эта операция даст алгебраическую сумму всех ЭДС, включенных в контур. Но в случае системы с сосредоточенными параметрами (3.16) можно существенно упростить, заменив интеграл просто суммой по элементам контура:

$$\sum_k I_k R_k = \sum_k \mathcal{E}_k, \quad (3.18)$$

где как токи I_k , так и ЭДС \mathcal{E}_k выбираются с учетом знака — по соотношению направления тока и направления обхода контура. В этом состоит *второе правило Кирхгофа*. Направления (знаки) с самого начала можно выбрать произвольно, при последующих

вычислениях они будут исправлены автоматически. Например, для схемы рис. 3.2 в, полагая все токи и ЭДС направленными по часовой стрелке, можем написать

$$\mathcal{E}_1 - \mathcal{E}_2 = I_1 r_1 + I_2 r_2 + I_{BC} R_{BC}; \quad \mathcal{E}_1 = I_1 r_1 - I_{AB} R_{AB};$$

$$I_{AB} R_{AB} + I_{BC} R_{BC} + I_{CA} R_{CA} = 0;$$

$$-\mathcal{E}_2 = r_2 I_2 + I_{AB} R_{AB} + I_{BC} R_{BC} \dots$$

— всего шесть уравнений для всех возможных контуров и еще три уравнения для всех узлов схемы:

$$I_1 + I_{AB} - I_{BC} = 0 \text{ и т. д.}$$

Обычным свойством оказывается переопределенность системы уравнений, не ведущая, однако, ни к каким парадоксам. Так в нашем примере уравнений получается девять, а неизвестных токов, как следует из рис. 3.2 в, всего пять. При последующей работе с полученными уравнениями из них как раз и получится пять линейно независимых. При расчетах сложных схем правила Кирхгофа дают такие преимущества, что с этой переопределенностью приходится мириться.

Задачи

1. Пространство между пластинами плоского конденсатора заполнено многослойным диэлектриком (рис. 3.3), обладающим слабой электропроводностью. Диэлектрическая проницаемость и удельная проводимость изменяются от ε_1 , λ_1 у одной пластины до ε_2 , λ_2 у другой пластины. Конденсатор включен в цепь постоянной ЭДС. Определить величину и знак суммарного свободного заряда q , сосредоточенного в объеме диэлектрика, когда в цепи устанавливается постоянный электрический ток I , текущий через диэлектрик по направлению от пластины 1 к пластине 2.

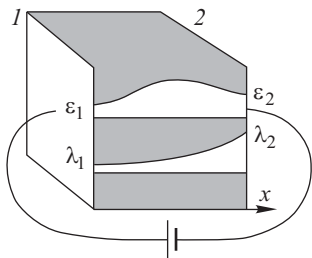


Рис. 3.3

Решение. Полезно обратить внимание на то, что никаких предположений касательно функций $\varepsilon(x)$, $\lambda(x)$ в задаче не сделано. Условие постоянства тока:

$$\frac{I}{S} = j = \lambda(x) E(x) = \text{const.}$$

Отсюда $E(x) = I/(S\lambda(x))$, тем самым $D(x) = \varepsilon_0(I/S) \varepsilon(x)/\lambda(x)$. Соответственно,

$$\rho = \text{div } \mathbf{D} = \frac{\partial}{\partial x} \varepsilon_0 \frac{I}{S} \frac{\varepsilon(x)}{\lambda(x)}.$$

Искомая величина получается интегрированием плотности заряда по объему:

$$q = S \int_1^2 \rho dx = \varepsilon_0 I \left(\frac{\varepsilon_2}{\lambda_2} - \frac{\varepsilon_1}{\lambda_1} \right).$$

2. Заземление концов телеграфной линии осуществлено посредством металлических шаров радиуса r_1 и r_2 соответственно (рис. 3.4). Удельная проводимость грунта вблизи них равна λ_1 и λ_2 . Найти сопротивление R земли между шарами. Считать почву в окрестности каждого из них однородной на расстояниях, больших по сравнению с радиусами шаров.

Решение. Для получения сопротивления заземления в компактной форме полезно пренебречь также асимметрией, связанной с подводящими проводами. Таким образом, для каждого (почти уединенного) шара задача о течении заряда оказывается с точностью до малых поправок сферически симметричной. Отсюда сразу следует:

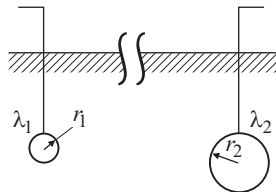


Рис. 3.4

$$I = \text{const} = 4\pi r^2 j(r) \Rightarrow E(r) = \frac{j}{\lambda} = \frac{I}{4\pi r^2 \lambda}.$$

При вычислении падения напряжения следует учесть, что интеграл $\int E dr$ быстро сходится, поэтому зависимостью $\lambda(r)$ можно пренебречь вдали от электродов, и, кроме того, падения напряжения на двух электродах войдут аддитивно:

$$U \equiv I R \approx \int_{r_1}^{\infty} \frac{I dr}{4\pi \lambda_1 r^2} + \int_{r_2}^{\infty} \frac{I dr}{4\pi \lambda_2 r^2},$$

откуда и следует

$$R = \frac{1}{4\pi} \left(\frac{1}{\lambda_1 r_1} + \frac{1}{\lambda_2 r_2} \right).$$

В силу сделанных выше предположений, результат представляет собой лишь оценку (впрочем, довольно точную). Но и на оценочном уровне универсальность сопротивления заземления — факт нетривиальный.

3. Определить сопротивление R_{AB} цепочки, изображенной на рис. 3.5 а.

Ответ: $R_{AB} = (7/5)r$.

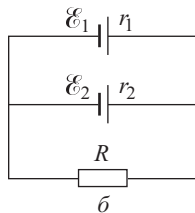
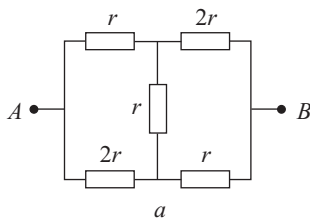


Рис. 3.5

4. Два гальванических элемента, ЭДС и внутренние сопротивления которых равны соответственно \mathcal{E}_1 , r_1 и \mathcal{E}_2 , r_2 , соединены параллельно и нагружены на сопротивление R (рис. 3.5 б). Определить падение напряжения на сопротивлении R .

Ответ: $U_R = \frac{\mathcal{E}_1 r_2 + \mathcal{E}_2 r_1}{(r_1 r_2 + r_1 R + r_2 R) R}.$

Глава 4

МАГНИТНОЕ ПОЛЕ ТОКА

4.1. Взаимодействие проводников с током. Понятие о магнитном поле

Говоря о магнитном поле тока, мы подразумеваем, если не оговорено противное, постоянный ток. В действительности, как будет ясно из последующих глав нашего курса, при не слишком жестких ограничениях все результаты настоящей главы могут быть перенесены и на случай переменного тока. Но и те ситуации, когда такой перенос затруднителен, достаточно типичны — это, например, длинные линии или электромагнитные волны в вакууме.

Мы уже упоминали в гл. 3 эффект взаимодействия квазинейтральных токонесущих проводников. В зависимости от величины тока, это взаимодействие может из чисто демонстрационного превратиться в очень сильное, например, результатом его может быть весьма эффектный и небезопасный взрыв соленоида. Цилиндрический проводник с продольным током — лайнер — может при уровне тока в несколько мегаампер создавать при схлопывании на ось давление порядка нескольких тысяч атмосфер.

Задолго до того как был построен первый гальванический элемент, и даже, пожалуй, раньше, чем были осознаны какие-либо электрические явления, человечество познакомилось с постоянными магнитами природного происхождения и, соответственно, с магнитными явлениями. Современная цивилизация сделала подобное знакомство неизбежным для любого ребенка.

Исследование свойств токонесущих проводников привело еще в начале XIX века к пониманию того, что они взаимодействуют не только между собой, но и с постоянными магнитами и взаимодействия эти — одной природы. Так складывалось представление о магнитном поле тока. Именно токовое магнитное поле, а не поле намагниченных образцов, оказалось в конечном счете тем первичным понятием, исходя из которого, мы представляем себе сегодня весь круг магнитных явлений, объединяющихся в современной электродинамике в единое целое с электрическими.

Реальная история физики развивалась таким образом, что первоначальное понимание сути магнетизма оказалось не вполне адекватным. Сначала, была построена, по аналогии с электростатикой, магнитостатика, в которой роль положительных и от-

рицательных зарядов исполняли северные и южные магнитные массы (названия были даны исходя из свойств стрелки компаса). Были введены понятия напряженности и индукции магнитного поля. Всю эту иерархию пришлось решительно пересмотреть на основании позднейших данных, и прежде всего — о магнитном поле тока.

Во-первых, оказалось, что магнитных зарядов в принципе нет. Вместо теоремы Гаусса (1.10) или (1.14), магнитное поле, точнее магнитная индукция \mathbf{B} , подчиняется более простому закону

$$\operatorname{div} \mathbf{B} = 0, \quad \text{или} \quad \oint_S \mathbf{B} d\mathbf{S} = 0. \quad (4.1)$$

Таким образом, поток магнитного поля любой природы через любую замкнутую поверхность равен нулю. Традиционная геометрическая интерпретация этого факта такова: силовые линии магнитного поля либо идут из бесконечности в бесконечность, либо замкнуты. (Обращаем особое внимание читателя на неполноту такой классификации — есть еще третья возможность, о которой большинство учебников просто забывает, — см. ниже.)

Во-вторых, как следует из возможности замкнутых силовых линий, магнитное поле и в стационарной задаче не является потенциальным. Можно посмотреть на это свойство и с другой стороны: поскольку теорема Гаусса (4.1) для магнитного поля вырождается, его источники должны фигурировать в каком-то другом уравнении — это будет так называемая теорема о циркуляции.

В § 1.6 мы рассмотрели дипольное приближение в электростатике. В частности, мы пользовались и понятием точечного диполя, но лишь как удобным приближением. Элементарным объектом при этом по умолчанию оставался электрический заряд. Именно заряд оказывается фундаментальным свойством элементарных частиц, тогда как элементарные «точечные» дипольные моменты у них не обнаружены. С точки зрения магнитных свойств элементарным объектом оказывается как раз диполь. У элементарных частиц нет магнитных зарядов, но многие из них обладают собственным магнитным дипольным моментом. (Так называемый магнитный монополю Дирака (по имени английского физика-теоретика П.А.М. Дирака (1902–1984)), входящий в некоторые теоретические построения как гипотетическая частица, в данном контексте рассматриваться не должен. Даже если он и существует реально, свойства его слишком далеко выходят за пределы классической, т. е. некантовой физики.)

Исторически понятие магнитного диполя, как и многие другие, было введено по полной аналогии с электростатикой. В дальнейшем, когда была признана неадекватность этой аналогии, пришлось изменить и саму методологию построения данного

раздела физики, опираясь на взаимодействие проводников с током. Но некоторые «следы» старинной магнитостатики сохранились и в современной электродинамике. Главный из них — так называемая перекрестная аналогия между характеристиками электрических и магнитных полей.

Суть ее в том, что *магнитной индукцией* \mathbf{B} называется векторная силовая характеристика магнитного поля, отражающая взаимодействие поля с током или магнитоактивным веществом (аналог \mathbf{E}).

Напряженностью магнитного поля \mathbf{H} называется величина, определяемая из конфигурации токов проводимости (или свободных токов) без учета отклика магнитоактивного вещества, или, что то же, микроскопических молекулярных токов (аналог \mathbf{D}).

Второе определение с неизбежностью несовершенно — такая же ситуация, как в сегнетоэлектриках, здесь еще более типична. По-настоящему корректное определение напряженности поля можно будет дать лишь на основании уравнений поля и граничных условий (как и для вектора \mathbf{D} в электростатике).

Очевидно, в вакууме величины \mathbf{B} и \mathbf{H} физически тождественны. Но в системе СИ их принято измерять в разных единицах. Коэффициент между единицей \mathbf{H} — 1 А/м и единицей \mathbf{B} — 1 Тл (Тесла; по имени сербского ученого Н. Тесла (1856–1943)) называется *магнитной постоянной* и обозначается символом $\mu_0 = 4\pi \cdot 10^{-7}$ СИ. Единица ее измерения по традиции именуется *генри на метр* (Г/м). Полезно заранее иметь в виду следующее равенство:

$$\varepsilon_0 \mu_0 = \frac{1}{c^2},$$

где $c = 3 \cdot 10^8$ м/с — скорость света. Строгий вывод этой формулы будет дан в главе «Электромагнитные волны».

Таким образом, в вакууме или в веществе, откликом которого на магнитное поле можно пренебречь, имеет место соотношение

$$\mathbf{B} = \mu_0 \mathbf{H}. \quad (4.2)$$

Соотношение (4.2) открывает цепочку самосогласованных формул, обобщающих опытные данные и корректно вводящих понятие магнитного поля. Для двух последующих законов нам еще понадобится определение элемента тока. Под таковым понимается малый участок квазиодномерного проводника, по которому течет ток I . В последующие формулы он войдет в виде $I d\mathbf{l}$, где $d\mathbf{l}$ — вектор элемента длины провода, поперечные масштабы которого пренебрежимо малы.

Итак, если электрическое поле действует на заряды как таковые и порождается зарядами, то магнитное поле действует на токи и токами порождается. Именно с вклада элемента тока в создание магнитного поля мы и начнем. Для этого рассмотрим

так называемый *закон Био и Савара* (по имени французских физиков Ж.Б. Био (1774–1862) и Ф. Савара (1791–1841)):

$$d\mathbf{H} = \frac{1}{4\pi r^3} [I d\mathbf{l}, \mathbf{r}]. \quad (4.3)$$

Для вычисления полного поля $\mathbf{H}(\mathbf{r})$ формулу (4.3) следует проинтегрировать по $d\mathbf{l}$. При этом векторное произведение автоматически согласует направление поля с «правилом буравчика».

Взаимодействие элемента тока с внешним магнитным полем описывается *законом Ампера* (по имени А.М. Ампера (1775–1836); иногда, быть может, более правильно его называют *законом Лапласа* по имени французского астронома, физика и математика П.С. Лапласа (1749–1827)):

$$d\mathbf{F} = [I d\mathbf{l}, \mathbf{B}]. \quad (4.4)$$

Полную силу и момент, действующие на проводник, и в этом случае можно получить, интегрируя по $d\mathbf{l}$. Формулы (4.3) и (4.4) вместе с соотношением (4.2) дают полное описание взаимодействия двух токонесущих проводников через посредство магнитного поля. Поучительна попытка Ампера объединить две основных формулы в одну:

$$d\mathbf{F}_{12} = I_1 [d\mathbf{l}_1, \mathbf{B}_2] = \frac{\mu_0}{4\pi} \frac{I_1 I_2}{r_{12}^3} [d\mathbf{l}_1 [d\mathbf{l}_2, \mathbf{r}_{12}]]. \quad (4.5)$$

Раскрывая двойное векторное произведение по известной формуле

$$[\mathbf{a}[\mathbf{b}\mathbf{c}]] = \mathbf{b}(\mathbf{a}\mathbf{c}) - \mathbf{c}(\mathbf{a}\mathbf{b}),$$

получаем

$$[d\mathbf{l}_1 [d\mathbf{l}_2, \mathbf{r}_{12}]] = d\mathbf{l}_2(d\mathbf{l}_1, \mathbf{r}_{12}) - \mathbf{r}_{12}(d\mathbf{l}_1, d\mathbf{l}_2).$$

Попробуем в этой формуле поменять местами индексы 1 и 2. Лишь второй член остается без изменений (с точностью, разумеется, до знака), но не первый. Таким образом,

$$d\mathbf{F}_{12} \neq -d\mathbf{F}_{21}.$$

Означает ли это, что или неверен третий закон Ньютона, или какая-нибудь из наших формул? Вовсе нет. Дело в том, что соотношения (4.3), (4.4) оперируют лишь с малыми элементами токонесущих проводов и дают в ответе лишь дифференциал вклада в силу. Говорить о третьем законе Ньютона можно только после интегрирования по $d\mathbf{l}_1$, $d\mathbf{l}_2$, т. е. по длине обоих контуров. При этом, как можно показать прямым вычислением,

$$\mathbf{F}_{12} = -\mathbf{F}_{21}.$$

Сама же формула (4.5) оказывается достаточно бессмысленной. Это не мешает ей, однако, также носить имя Ампера и притом именоваться законом взаимодействия токов.

Закон Ампера во многих случаях удобно переписать в локальной форме, к которой выражение (4.4) без труда можно свести. Введем элемент поперечного сечения проводника dS и представим элемент тока в виде $I d\mathbf{l} \equiv d\mathbf{l} \int \mathbf{j} dS$. Далее, принимая во внимание, что направления \mathbf{j} и $d\mathbf{l}$ для тонкого проводника совпадают, получаем элемент тока в следующей форме: $I d\mathbf{l} \equiv \mathbf{j} dV$. Поделив (4.4) на dV , получим плотность силы Ампера в проводящей среде:

$$\mathbf{f} \equiv \frac{d\mathbf{F}}{dV} = [\mathbf{j} \mathbf{B}]. \quad (4.6)$$

И, наконец, отнесем силу Ампера к одной заряженной частице. Для этого формулу (4.6) нужно просто разделить на концентрацию носителей тока n . Вспоминая соотношение (3.5) и ограничиваясь одним сортом носителей, получаем

$$\mathbf{F} = e[\mathbf{v} \mathbf{B}], \quad (4.7)$$

где e — заряд частицы с учетом знака. Эта зависимость называется *силой Лоренца* и представляет собой самостоятельную ценность. Можно было бы начать прямо с нее и отсюда выводить закон Ампера. Зависимость эта играет принципиально важную роль в механике заряженных частиц — как при движении в заданном внешнем поле, так и при изучении самосогласованной динамики.

Как можно видеть из (4.7), сила Лоренца относится к т. н. *гироскопическим силам* — при движении заряженной частицы в заданном однородном поле работа не совершается. Из нее, в частности, следует эффект циклотронного вращения. Пусть частица с зарядом e и массой m совершает нерелятивистское движение в однородном внешнем поле \mathbf{B} . Уравнение движения

$$m \frac{d\mathbf{v}}{dt} = e[\mathbf{v} \mathbf{B}]$$

расщепляется на два. Проекция уравнения на направление \mathbf{B} описывает равномерное движение, тогда как в плоскости, перпендикулярной \mathbf{B} , траекторией будет окружность, по которой частица будет двигаться равномерно с угловой скоростью

$$\omega_B = \frac{eB}{m}. \quad (4.8)$$

Ее принято называть *циклотронной частотой* или гирочастотой.

4.2. Магнитные поля простейших токовых конфигураций. Векторные свойства магнитного поля

Самая простая задача на вычисление магнитного поля — поле в центре тонкого кольца радиуса R с током I . Формула (4.3) при этом существенно упрощается; последующее интегрирование по

$d\mathbf{l}$ сводится к умножению на длину окружности, в то же время любой элемент $d\mathbf{l}$ дает вклад в вектор \mathbf{B} , направленный по оси. Как следствие, полное поле также оказывается параллельно оси, связано с направлением тока в кольце правилом буравчика и равно

$$H = \frac{I}{4\pi R^3} \cdot 2\pi R \cdot R = \frac{I}{2R}.$$

Соответственно, $B = \mu_0 I / (2R)$. Несколько сложнее будет вычислить поле прямого провода, но именно эта задача оказывается очень важной для последующих построений.

Обратимся к рис. 4.1 а. Пусть по бесконечному прямому проводу течет ток I . Из осевой симметрии задачи следует, что силовые линии магнитного поля — замкнутые концентрические окружности. (Другой вариант — радиальные лучи — не годится, потому что условие (4.1) оказалось бы нарушено.) Тот же самый ответ следует и из других соображений, а именно из геометрических свойств формулы (4.3) и из равнозначности двух полубесконечных лучей, образующих ось системы. Таким образом, $H = H(R)$, где R — расстояние до оси. Легко видеть, что $[d\mathbf{l}, \mathbf{r}] = dx \cdot R$ — см. рис. 4.1 б. Далее,

$$r = \frac{R}{\cos \alpha}; \quad dx = R d(\operatorname{tg} \alpha) = \frac{R d\alpha}{\cos^2 \alpha},$$

и окончательно —

$$H = \int_{-\infty}^{\infty} \frac{I}{4\pi} \frac{R dx}{r^3} = \int_{-\pi/2}^{\pi/2} \frac{I \cos \alpha d\alpha}{4\pi R} = \frac{I}{2\pi R}. \quad (4.9)$$

Результат (4.9) можно чисто формально переписать и таким образом:

$$\oint \mathbf{H} d\mathbf{l} = I, \quad (4.10)$$

поскольку интегрирование по контуру в данном случае сведется к тривиальному умножению на $2\pi R$. На этом простом примере мы познакомились с той формой закона Био и Савара, которая чаще всего употребляется в современной физике и называется *теоремой о циркуляции*. Название это достаточно условно; для вакуумного поля эквивалентность формул (4.10) и (4.3) может быть строго доказана методами векторного анализа, но в магнитоактивной среде обе эти формулы по сути постулируются в качестве определения напряженности магнитного поля. Не следует также забывать, что обе они подразумевают стационарную постановку задачи — в случае же переменных во времени полей их либо придется модифицировать, либо обеспечить их применимость какими-то сильными неравенствами.

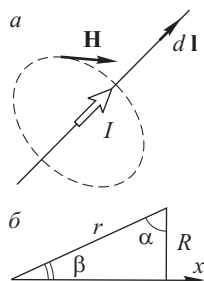


Рис. 4.1

При всех этих оговорках мы, тем не менее, будем придерживаться для (4.10) общепринятого названия: теорема о циркуляции. Но, имея в виду последующие ее применения, мы перепишем (4.10) в локальной (дифференциальной) форме, в которой она и фигурирует в числе основных уравнений теоретической электродинамики — уравнений Максвелла. Для этого нам придется, помимо операций дивергенции и градиента, познакомиться с новой операцией векторного дифференцирования, которая называется «ротор»:

$$\begin{aligned} \operatorname{rot} \mathbf{H} &\equiv [\nabla, \mathbf{H}] = \\ &= \mathbf{e}_1 \left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} \right) + \mathbf{e}_2 \left(\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} \right) + \mathbf{e}_3 \left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} \right). \end{aligned}$$

Операция может применяться к любому вектору $\mathbf{A}(\mathbf{r})$; мы используем вектор \mathbf{H} просто для удобства. Символическое представление операции rot в виде векторного произведения очень удобно для вывода некоторых важных соотношений. В частности, используя известные правила векторной алгебры, нетрудно получить

$$\operatorname{rot} (\nabla \psi) = [\nabla, \nabla \psi] \equiv 0,$$

$$\operatorname{rot} \operatorname{rot} \mathbf{A} = [\nabla[\nabla, \mathbf{A}]] = \nabla(\nabla, \mathbf{A}) - (\nabla, \nabla) \mathbf{A} = \nabla \operatorname{div} \mathbf{A} - \nabla^2 \mathbf{A}$$

— для любых функций $\psi(\mathbf{r})$ и $\mathbf{A}(\mathbf{r})$. Эта же операция позволяет придать весьма изящную форму одной из главных теорем векторного анализа — *теореме Стокса*.

Пусть дан некоторый произвольный контур L и векторная функция $\mathbf{H}(\mathbf{r})$. Пусть, далее, S — гладкая поверхность, натянутая на наш контур, т. е. построенная так, что он является для нее линией границы. Тогда при произвольной ее форме имеет место соотношение

$$\oint_L \mathbf{H} d\mathbf{l} = \int_S \operatorname{rot} \mathbf{H} d\mathbf{S}, \quad (4.11)$$

— циркуляция вектора \mathbf{H} по замкнутой границе поверхности равна потоку через эту поверхность его ротора. В этом и состоит содержание теоремы Стокса. Сравним формулы (4.11) и (4.10), представляя в последней $I \equiv \int \mathbf{j} d\mathbf{S}$. Правые части обеих формул должны быть равны для произвольных L и S . Это значит, что равны подынтегральные выражения:

$$\operatorname{rot} \mathbf{H} = \mathbf{j}. \quad (4.12)$$

Это и есть *теорема о циркуляции* в локальной форме, входящая наряду с (2.3) и (4.1) в систему уравнений Максвелла. Она еще будет в последующих главах дополнена током смещения; в настоящем виде этот закон является точным, по крайней мере, в случае токов и полей, не зависящих от времени.

Используя теорему Стокса для вывода теоремы о циркуляции в форме (4.12), мы, между прочим, доказали еще одно важное ее свойство: несущественно, каким именно образом линии тока пересекают контур — они могут пересекать его под любым углом, быть может, «вперед–назад», образуя токовые петли и т. д. Все это учитывается интегралом $\int \mathbf{j} d\mathbf{S}$ и никак не сказывается на формуле (4.10). В нее входит только заряд, в единицу времени пересекающий любую поверхность, натянутую на контур, а как это происходит — не имеет значения.

Поскольку уравнения (4.3), (4.12), (4.12) линейны, для вектора напряженности магнитного поля справедлив *принцип суперпозиции* — поле суммы токов равно сумме полей каждого из них. Это справедливо и для действующего поля \mathbf{B} , но лишь в вакууме или в линейной среде. Например, в ферромагнетиках принцип суперпозиции не работает — и не только для \mathbf{B} , но и для \mathbf{H} , ввиду нетривиальности граничных условий.

В качестве примера вычисления магнитного поля с помощью теоремы о циркуляции рассмотрим длинный цилиндрический соленоид, по обмотке которого протекает ток I (рис. 4.2). В отличие от тонкого кольца с током, поле внутри соленоида оказывается практически одномерным и параллельным его оси. Это следует из симметрии задачи, но может быть прямо выведено из закона Био и Савара при следующих условиях:

а) диаметр катушки должен быть много меньше ее длины;

б) шаг намотки должен быть много меньше диаметра соленоида;

в) намотка должна быть плотной, т. е. диаметр провода должен быть равен шагу намотки, в противном случае мы правильно определяем поле только для расстояний от стенок соленоида, много больших зазора между витками.

Иногда об этих условиях забывают. Другим примером «забывчивости» является популярное, но неправильное утверждение о том, что поле вне такого соленоида равно нулю. На самом деле оно такое же, как и у прямого провода:

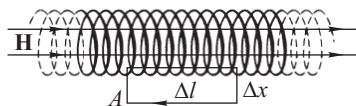


Рис. 4.2

$$H = \frac{I}{2\pi R},$$

где R — расстояние от оси (превышающее радиус катушки). Это следует из аксиальной симметрии задачи и теоремы о циркуляции. В самом деле, как уже говорилось выше, току безразлично, каким образом пересекать натянутую на контур интегрирования поверхность, важно лишь, что он пересекает ее один раз. Для

дальнейшего существенно, что силовые линии лежат в плоскости, перпендикулярной оси системы.

Определим поле внутри соленоида. Проведем любую плоскость через его ось и построим в этой плоскости прямоугольный контур A , как показано на рисунке. Одна из длинных сторон прямоугольника Δl расположена вне катушки; поле к ней перпендикулярно, и вклад ее в циркуляцию $\oint \mathbf{H} d\mathbf{l}$ равен нулю. Две боковых стороны Δx мы с самого начала выбираем малыми и в дальнейшем устремляем к нулю, поэтому их вкладом также можно пренебречь. Остается длинная сторона внутри катушки. Пусть плотность намотки (число витков на единицу длины соленоида) равна n . Если N — полное число витков, а l — длина соленоида, то $n = N/l$. Ток I пересекает контур A столько раз, сколько витков он захватывает, т. е. $n\Delta l$. Теорема о циркуляции (4.10) принимает вид простого соотношения

$$\mathbf{H} \cdot \Delta \mathbf{l} = n\Delta l I,$$

откуда окончательно следует

$$\mathbf{H} = nI = \frac{N}{l} I; \quad \mathbf{B} = \mu_0 nI. \quad (4.13)$$

Сколь бы ни были подробны наши рассуждения (которые при известном опыте попросту опускаются), довольно очевидно, что вывод этой хрестоматийной формулы оказался намного проще, чем если бы мы пользовались напрямую законом Био и Савара.

Рассмотрим еще один пример, ответ в котором будет чисто качественным, но принципиально важным для всей физики магнитных полей. Рассмотрим стационарную токовую конфигурацию, изображенную на рис. 4.3, — кольцевой ток, охватывающий прямой ток. (Она не является чисто умозрительной — используя сверхпроводящую технику, ее, в принципе, можно реально осуществить.) Как будут выглядеть магнитные силовые линии?

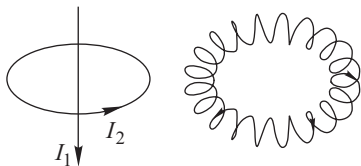


Рис. 4.3

Для прямого тока ответ известен — это будут concentрические окружности, центры которых лежат на оси. Силовые линии витка с током лежат, как известно, в плоскости, проходящей через ось. Это замкнутые линии неправильной формы, охватывающие виток, — картина, хорошо известная из школьного курса физики. (Лишь одна из них незамкнута и совпадает с осью; для нас она в данном случае неинтересна.) При попытке сложить эти два поля, следуя принципу суперпозиции, окажется, что результирующая силовая линия навивается на некоторую тороидальную поверхность и целиком на

ней расположена. Первое впечатление — это замкнутая на себя спираль.

Заметим, однако, что для замкнутости спирали необходима соизмеримость длины обхода тора и ее шага, т. е. сделав некоторое целое число обходов, силовая линия должна замкнуться сама на себя. Меняя радиус тора, мы получим семейство вложенных тороидальных поверхностей, причем замкнутые силовые линии будут соответствовать счетному подмножеству, а незамкнутые — континууму.

Это и есть «третий вариант» поведения магнитной силовой линии — *магнитная поверхность*. То есть силовая линия может не только идти из бесконечности в бесконечность или быть замкнутой, но и занимать некоторую поверхность, плотно ее заштриховывая, и не иметь при этом ни начала, ни конца. Как показано в нашем примере, поверхность эта может быть вложена в конечный объем. Такое поведение силовых линий — вовсе не экзотика; оно, например, используется в прикладных задачах для удержания заряженных частиц.

Ничего подобного нет в электростатике. Это лишнее свидетельство наивности первоначальных попыток построить магнито-статику по образу электростатики. На самом деле электрические и магнитные поля обладают разными геометрическими свойствами, что формально отражено в различных базовых уравнениях:

$$\operatorname{div} \mathbf{D} = \rho \quad \text{и} \quad \operatorname{rot} \mathbf{H} = \mathbf{j}.$$

В заключение данного параграфа дадим определение магнитного потока. Для любой гладкой поверхности S *магнитный поток* определяется следующим образом:

$$\Phi = \int \mathbf{B} d\mathbf{S}.$$

Это понятие оказывается очень удобным и широко используется в электродинамике. Если поверхность замкнута, то в силу (4.1) $\Phi \equiv 0$ — вне зависимости от того, расположены источники магнитного поля снаружи или внутри нее. Единица магнитного потока имеет специальное название — *вебер*:

$$1 \text{ Вб} = 1 \text{ А} \cdot \text{с} \cdot \text{Ом}.$$

Такая размерность следует из закона электромагнитной индукции, с которым нам еще предстоит познакомиться.

4.3. Вектор-потенциал

Материал, изложенный в данном параграфе, обычно не принято помещать в курс общей физики. Нам, однако, представляется, что экономия в объеме и сложности ряда последующих

вычислительных процедур оказывается столь значительной, что оправдывает такое начинание. (То же самое можно сказать и об операциях векторного дифференцирования. Они позволяют представить уравнения поля в чрезвычайно компактной и наглядной форме, существенно упрощают многие доказательства и выводы, и потому, на наш взгляд, стоит как можно раньше преодолеть некоторые трудности, связанные с освоением этого аппарата.)

Мы уже могли оценить, сколь конструктивным оказалось в электростатике понятие электрического потенциала. Определение его было дано в формулах (1.16), (1.17), а основное уравнение, которому потенциал подчиняется как функция точки, — (1.19). Правда, определение давалось для системы дискретных точечных зарядов, а уравнение соответствует описанию зарядового распределения непрерывной функцией $\rho(\mathbf{r})$. Однако все проблемы снимаются простым обобщением определения (1.16), (1.17):

$$\varphi = \frac{1}{4\pi\epsilon_0} \int \frac{\rho(\mathbf{r})}{r} dV. \quad (4.14)$$

Теперь попробуем подобным же образом ввести функцию, порождающую магнитное поле. Только, в отличие от электростатики, источником будет не скаляр ρ , а вектор \mathbf{j} . Рассмотрим некоторую проводящую среду с распределением плотности тока $\mathbf{j}(\mathbf{r})$. Введем вектор \mathbf{A} с компонентами

$$A_{x,y,z} = \frac{1}{4\pi} \int \frac{j_{x,y,z}(\mathbf{r})}{r} dV. \quad (4.15)$$

Сравнивая (4.14), (4.15) и (1.19), нетрудно убедиться, что компоненты вектора \mathbf{A} должны удовлетворять уравнению Пуассона:

$$\nabla^2 A_k \equiv \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) A_k = -j_k$$

(где $k = x, y, z$), что обычно принято записывать в виде

$$\nabla^2 \mathbf{A} = -\mathbf{j}. \quad (4.16)$$

Векторная величина \mathbf{A} и называется *вектор-потенциалом*. Но это название имеет смысл лишь в том случае, если эта функция и в самом деле порождает вектор магнитного поля. Чтобы в этом убедиться, вычислим прежде всего дивергенцию вектор-потенциала:

$$\operatorname{div} \mathbf{A} = \frac{1}{4\pi} \int \operatorname{div} \left(\frac{\mathbf{j}}{r} \right) dV = \frac{1}{4\pi} \oint_S \left(\frac{\mathbf{j}}{r} \right) d\mathbf{S}.$$

Мы преобразовали последний интеграл по теореме Гаусса. Поскольку объемное интегрирование предполагается по всей проводящей среде, где $\mathbf{j} \neq 0$, постольку поверхностный интеграл

должен браться по поверхности всех проводников, где, однако, $\hat{j}_\perp = 0$, а потому равен нулю и весь интеграл. Таким образом,

$$\operatorname{div} \mathbf{A} = 0.$$

Но тогда имеет место следующее соотношение (см. предыдущий параграф):

$$\operatorname{rot} \operatorname{rot} \mathbf{A} = \nabla \operatorname{div} \mathbf{A} - \nabla^2 \mathbf{A} \equiv -\nabla^2 \mathbf{A},$$

так что мы можем заменить левую часть в уравнении Пуассона (4.16) для вектор-потенциала:

$$\operatorname{rot}(\operatorname{rot} \mathbf{A}) = \mathbf{j}.$$

Сравнивая этот результат с дифференциальной формой теоремы о циркуляции (4.12), получаем окончательный результат:

$$\mathbf{H} = \operatorname{rot} \mathbf{A}. \quad (4.17)$$

Мы имеем возможность еще раз убедиться в различной векторной природе электрического и магнитного полей. Если статическое электрическое поле представимо как градиент некоторой скалярной функции (такие поля называют *потенциальными*), то магнитное поле в вакууме есть ротор некоторой векторной функции. Поля такого рода принято называть *соленоидальными*.

Скалярный электрический потенциал определен, как и потенциальная энергия, с точностью до константы. Произвол в определении вектор-потенциала больше: поскольку ротор градиента любой скалярной функции тождественно равен нулю, то и величина \mathbf{A} определена с точностью до $\nabla\psi$, где $\psi(\mathbf{r})$ — произвольная скалярная функция. Наблюдаемой же (измеряемой) величиной остается поле \mathbf{H} , т. е. ротор вектор-потенциала. Мы не будем, однако, углубляться в этот вопрос, потому что данное выше определение (4.15)–(4.17) вполне корректно и для целей нашего курса достаточно.

4.4. Магнитный диполь. Понятие о магнитном моменте

На первый взгляд, из закона Био и Савара (4.3) следует спадание магнитного поля на бесконечности, аналогичное закону Кулона ($\propto r^{-2}$). Для некоторых вырожденных ситуаций мы и в самом деле можем в случае электро- и магнитостатики получить ответы, функционально тождественные. Так, электрическое поле однородно заряженного длинного провода и магнитное поле длинного провода с током, хотя и различаются геометрией, зависят от радиуса по одному закону ($\propto 1/r$). Но в общем случае между ними существует принципиальное различие. Если уеди-

ненный заряд любого знака имеет право на существование, то элемент тока не может существовать сам по себе, ибо любой стационарный ток должен быть замкнут. Потому и мультипольность магнитного поля не может быть ниже дипольного приближения:

$$H \leq \text{const} \cdot r^{-3} \text{ при } r \rightarrow \infty.$$

(Конечно, на достаточно больших, например галактических, масштабах этим качеством обладает и электрическое поле, но это уже следствие нейтральности Вселенной, а не собственно электродинамики.)

Таким образом, элементарным объектом в физике магнитных полей оказывается диполь, и именно дипольным по геометрии и функциональной зависимости будет поле произвольной системы токов на больших расстояниях.

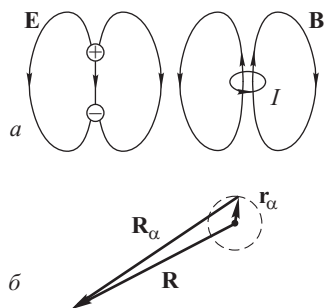


Рис. 4.4

По природе и по геометрии силовых линий электрический и магнитный диполи различаются (рис. 4.4 а). Поэтому перенос результатов гл. 1 на этот случай невозможен и вывод основных соотношений должен быть проведен заново. Здесь весьма полезным для нас окажется материал предыдущего параграфа, т. е. описание магнитного поля посредством вектор-потенциала. Рассмотрим квазистационарную токовую систему.

Можно в принципе смотреть на нее как на квазинейтральную систему зарядов, совершающих замкнутое финитное движение (поскольку квазистационарные токи должны быть замкнуты). Обратимся к формуле (4.15) и проведем следующую цепочку замен в подынтегральном выражении:

$$\mathbf{j} dV = nev dV = dq \langle \mathbf{v} \rangle_t.$$

Мы выбрали элемент dV столь малым, чтобы в него попали заряды только одного сорта и чтобы усреднения по объему не требовалось. В то же время усреднение по времени необходимо ввиду дискретности носителей заряда (в сущности, само понятие постоянного либо квазистационарного тока неявно подразумевает такое усреднение). Следуя логике этой замены, мы можем интегрирование в (4.15) заменить суммированием по дискретным зарядам, что в данном случае будет удобнее,

$$\mathbf{A} = \frac{1}{4\pi} \sum_{\alpha} \frac{q_{\alpha} \langle \mathbf{v}_{\alpha} \rangle}{R_{\alpha}}, \quad (4.18)$$

а затем, в соответствии с рис. 4.4 б, представим \mathbf{R}_{α} в виде

векторной суммы $\mathbf{R}_\alpha = \mathbf{R} - \mathbf{r}_\alpha$, где все $r_\alpha \ll R$. Тогда

$$\mathbf{A} = \frac{1}{4\pi} \sum_{\alpha} \frac{q_{\alpha} \langle \mathbf{v}_{\alpha} \rangle}{|\mathbf{R} - \mathbf{r}_{\alpha}|} \approx \frac{1}{4\pi R} \sum_{\alpha} \langle q_{\alpha} \mathbf{v}_{\alpha} \rangle - \frac{1}{4\pi} \sum_{\alpha} \left\langle q_{\alpha} \mathbf{v}_{\alpha} \left(\mathbf{r}_{\alpha} \nabla \frac{1}{R} \right) \right\rangle.$$

Первый член этой разности есть не что иное, как усредненная по времени временная производная полного дипольного момента системы,

$$\sum_{\alpha} \langle q_{\alpha} \mathbf{v}_{\alpha} \rangle = \left\langle \frac{d}{dt} \sum_{\alpha} q_{\alpha} \mathbf{r}_{\alpha} \right\rangle.$$

Неважно, чему он равен и по какому закону меняется. Важно, что система совершает финитное движение и любая величина, характеризующая ее состояние, изменяется в конечных пределах, а поэтому при усреднении по достаточно большому промежутку времени Δt ее производная стремится к нулю. Пусть F — такая величина, а τ — характерное время эволюции системы. Тогда

$$\left\langle \frac{dF}{dt} \right\rangle \leq \frac{\tau}{\Delta t} \left(\frac{dF}{dt} \right)_{\max} \rightarrow 0 \text{ при } \Delta t \gg \tau.$$

Нам еще не один раз придется воспользоваться этим результатом. Обратимся ко второму члену разности, принимая во внимание, что $\nabla 1/R = -\mathbf{R}/R^3$:

$$\mathbf{A} \approx \frac{1}{4\pi R^3} \sum_{\alpha} \langle q_{\alpha} \mathbf{v}_{\alpha} (\mathbf{r}_{\alpha} \mathbf{R}) \rangle, \quad (4.19)$$

в котором сделаем тождественное преобразование

$$\langle \mathbf{v}_{\alpha} (\mathbf{r}_{\alpha} \mathbf{R}) \rangle \equiv \left\langle \left(\frac{d}{dt} \{ \mathbf{r}_{\alpha} (\mathbf{r}_{\alpha} \mathbf{R}) \} - \mathbf{r}_{\alpha} (\mathbf{v}_{\alpha} \mathbf{R}) \right) \right\rangle.$$

Первый член в правой части — снова полная производная, дающая нуль при усреднении, поэтому, говоря о средних по времени значениях, можно приравнять левую часть второму члену в правой. Сделаем это в выражении (4.19), введя, соответственно, коэффициент $1/2$:

$$\mathbf{A} \approx \frac{1}{8\pi R^3} \sum_{\alpha} q_{\alpha} \langle \mathbf{v}_{\alpha} (\mathbf{r}_{\alpha} \mathbf{R}) - \mathbf{r}_{\alpha} (\mathbf{v}_{\alpha} \mathbf{R}) \rangle \equiv \frac{1}{8\pi R^3} \sum_{\alpha} q_{\alpha} \langle [\mathbf{R} [\mathbf{v}_{\alpha} \mathbf{r}_{\alpha}]] \rangle. \quad (4.20)$$

Если определить вектор магнитного момента системы как

$$\mathbf{m} = \frac{1}{2} \sum_{\alpha} \langle [\mathbf{r}_{\alpha}, q_{\alpha} \mathbf{v}_{\alpha}] \rangle, \quad (4.21)$$

то вектор-потенциал магнитного поля системы на больших, в сравнении с ее размерами, расстояниях, имеет вид

$$\mathbf{A} = \frac{[\mathbf{m} \mathbf{R}]}{4\pi R^3}. \quad (4.22)$$

Эквивалентная форма:

$$\mathbf{A} = -\frac{1}{4\pi} [\mathbf{m}, \nabla(1/R)]. \quad (4.23)$$

Итак, мы видим, что в теории магнитных полей дипольное приближение базируется на понятии вектор-потенциала — полезно сравнить выражения (1.24) и (4.22). Мы ввели понятие магнитного диполя и теперь можем для удобства считать его «точечным». В этом случае, т. е. при вычислении поля на расстояниях, существенно больших пространственного масштаба токовой конфигурации, последняя характеризуется единственным векторным параметром \mathbf{m} . Для вычисления поля воспользуемся формулой (4.23). Предварительно проведем операцию векторного дифференцирования в общем виде. Пусть \mathbf{a} — некоторый постоянный вектор, тогда

$$\text{rot} [\mathbf{a}, \mathbf{b}(\mathbf{r})] = [\nabla [\mathbf{a}, \mathbf{b}(\mathbf{r})]] = \mathbf{a}(\nabla \mathbf{b}) - (\mathbf{a} \nabla) \mathbf{b}.$$

В нашем случае $\mathbf{b} = \nabla(1/R)$, поэтому $(\nabla \mathbf{b}) \equiv \nabla^2(1/R) = 0$, потому что это уравнение Пуассона для потенциала точечного заряда при $R \neq 0$ (впрочем, это можно проверить и прямым вычислением). Таким образом,

$$\mathbf{H} = \text{rot} \mathbf{A} = -\left(\frac{\mathbf{m}}{4\pi}, \nabla\right) \frac{\mathbf{R}}{R^3} = -\mathbf{R} \left(\frac{\mathbf{m}}{4\pi}, \nabla R^{-3}\right) - R^{-3} \left(\frac{\mathbf{m}}{4\pi}, \nabla\right) \mathbf{R}.$$

Используя также цепочку преобразований

$$(\mathbf{m} \nabla) \mathbf{R} = \left(m_x \frac{\partial}{\partial x} + \dots\right) (\mathbf{e}_x x + \dots) = \mathbf{m},$$

получаем окончательный ответ:

$$\mathbf{H} = \text{rot} \mathbf{A} = \frac{3(\mathbf{m} \mathbf{R}) \mathbf{R} - R^2 \mathbf{m}}{4\pi R^5}. \quad (4.24)$$

Этот результат оказался идентичным формуле (1.25), в особенности если мы, имея в виду уже отмеченную перекрестную аналогию, перепишем последнюю для электрической индукции, т. е. просто умножим на ε_0 . Для справки отметим, что мы ввели т. н. амперовский магнитный момент; иногда вместо него используют кулоновский магнитный момент, отличающийся коэффициентом $\mu_0/4\pi$.

В отличие от электрического диполя, простейший магнитный диполь представляет собой тонкий замкнутый виток с током (рис. 4.4 а). Для вычисления его дипольного момента построим цепочку, обратную той, которую мы использовали при выводе формулы (4.18):

$$dqv = nev dV = \mathbf{j} dV = \mathbf{j} S_{\perp} dl = I dl,$$

где S_{\perp} — поперечное сечение провода, I — ток, протекающий через виток. Теперь сумму в (4.21) можно заменить интегралом по длине контура:

$$\mathbf{m} = \frac{1}{2} \oint [\mathbf{r}, I d\mathbf{l}].$$

Воспользуемся известным свойством векторного произведения: модуль $|\mathbf{a} \mathbf{b}|$ равен удвоенной площади треугольника, построенного на векторах \mathbf{a} и \mathbf{b} как на сторонах. Таким образом, $|\mathbf{r}, d\mathbf{l}| = 2 dS$, где S — площадь контура витка, и окончательно:

$$\mathbf{m} = I S \mathbf{n}. \quad (4.25)$$

Здесь \mathbf{n} — единичный вектор нормали к плоскости витка, ориентированный согласно правилу буравчика по отношению к направлению тока, или, что то же: с конца его ток в контуре должен выглядеть протекающим против часовой стрелки. Если мы имеем дело с катушкой из N витков, ее магнитный момент равен NIS ; подразумевается, что состояние, на котором определяется поле катушки, или масштаб изменения внешнего поля существенно превышают любой ее размер. Если же магнитный диполь — иной природы, например, постоянный магнит, то его очень часто удобно представить в виде витка или трубки с азимутальным током. Этот фиктивный ток называют *током намагничения*. В качестве источника поля он ведет себя так же, как и реальный ток, не следует только подставлять его в теорему о циркуляции, а тем более в закон Джоуля–Ленца.

Вернемся к дискретному представлению магнитного момента (4.21). Предположим, что диполь формируется циклическим движением одной частицы, либо ансамблем одинаковых частиц, либо, наконец, ансамблем различных частиц, но имеющих одинаковое отношение заряда к массе, которое мы обозначим как e/m . Опуская, ради простоты записи, индексы и операцию усреднения, имеем

$$\mathbf{m} = \frac{1}{2} \sum q [\mathbf{r} \mathbf{v}] \equiv \frac{e}{2m} \sum [\mathbf{r}, m \mathbf{v}] = \frac{e}{2m} \sum [\mathbf{r} \mathbf{p}].$$

Поскольку $[\mathbf{r} \mathbf{p}]$ есть момент импульса частицы, наша формула связывает магнитный момент и момент импульса системы:

$$\mathbf{m} = \frac{e}{2m} \mathbf{L}. \quad (4.26)$$

Величина $e/(2m)$ называется *гиромагнитным отношением*. Соотношение (4.26) оказывается настолько универсальным, что работает даже в квантовой механике — именно этим коэффициентом оказываются связаны механический и магнитный моменты электронных орбит в атоме. И только при переходе к элементарным частицам появляется необходимость в его модификации.

Задачи

1. По двум бесконечно длинным прямолинейным проводникам A и B , выполненным из немагнитного материала и ограниченными пересекающимися цилиндрическими поверхностями, текут в противоположных направлениях токи одинаковой плотности j (рис. 4.5а). Найти величину и направление магнитного поля в полости Γ .

Решение. Введем в сечении проводников вектор $\overrightarrow{AB} = \mathbf{d}$ (рис. 4.5б) и воспользуемся принципом суперпозиции полей. Поскольку плотности тока равны

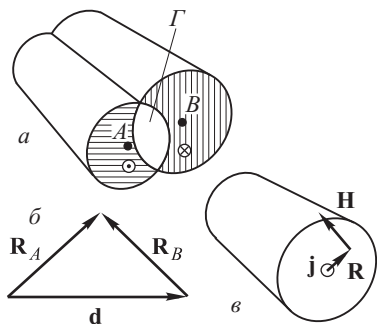


Рис. 4.5

в каждом из проводников и противоположно направлены, дополним мысленно каждый проводник до цилиндра, так что нулевой ток в полости получится как результат сложения встречных токов. Произвольная точка полости будет характеризоваться радиусами-векторами \mathbf{R}_A и \mathbf{R}_B ; $\mathbf{d} = \mathbf{R}_A - \mathbf{R}_B$. Внутри цилиндрического проводника, по которому течет однородный ток j (рис. 4.5в), поле в точке с радиусом R ищется из теоремы о циркуляции:

$$2\pi R H = \pi R^2 j \Rightarrow H = \frac{1}{2} R j.$$

Если же мы хотим определить вектор \mathbf{H} , то, как можно понять из рис. 4.5в, это соотношение следует переписать в виде векторного произведения: $\mathbf{H} = [\mathbf{j}, \mathbf{R}]/2$. Вычислим это поле в полости, имея в виду, что $\mathbf{j}_B = -\mathbf{j}_A$:

$$\mathbf{H}_\Pi = \frac{1}{2} ([\mathbf{j}_A \mathbf{R}_A] + [\mathbf{j}_B \mathbf{R}_B]) = \frac{1}{2} [\mathbf{j}_A, \mathbf{R}_A - \mathbf{R}_B] = \frac{1}{2} [\mathbf{j}_A \mathbf{d}].$$

Таким образом, поле в полости будет однородным и направлено в нашем случае снизу вверх.

2. По двум параллельным проводящим плоскостям текут антипараллельные токи с однородной линейной плотностью $dI/dl = i$ (рис. 4.6а). Определить величину и направление давления на каждую плоскость.

Решение. О направлении силы давления можно сказать сразу, что плоскости должны отталкиваться друг от друга. Вычислим поле, создаваемое токами каждой плоскости (рис. 4.6б).

Ввиду высокой симметрии задачи, силовые линии должны быть просто прямыми, параллельными плоскости. В каждом полупространстве поле однородно — иначе мы могли бы построить контур, который не пересекался бы током, но циркуляция поля по этому контуру была бы ненулевой. На рисунке контур $ABB'A'$ — вид на левую пластину сверху — пронизывается током $i \cdot AA'$, а циркуляция поля по нему равна $2 \cdot H \cdot AA'$, так что поле в каждом полупространстве равно $H_0 = i/2$. В системе параллельных плоскостей (рис. 4.6в) поле вне ее обратится в нуль, а между плоскостями будет равно $2H_0$, соответственно, индукция $B = \mu_0 i$. Но не все это поле надо подставлять в закон Ампера, а лишь половину его, потому что ток каждой из плоскостей выталкивается полем другой плоскости, но сам себя он никуда толкать не

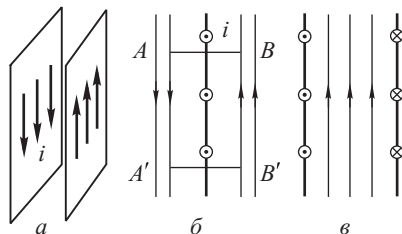


Рис. 4.6

может. Формулы (4.4) или (4.6) дают в случае нашей геометрии простой результат:

$$\mathcal{P} = i \cdot B_{\text{вн}} = \mu_0 \frac{i^2}{2}.$$

3. Определить траекторию движения произвольной заряженной частицы в скрещенных полях \mathbf{E}, \mathbf{H} ; напряженности поля E и H известны; $\mathbf{E} \perp \mathbf{H}$; $\varepsilon_0 E^2 \ll \mu_0 H^2$.

Решение. Движение заряженной частицы в магнитном поле определяется силой Лоренца (4.7). В направлении оси $Oz \parallel \mathbf{H}$ (рис. 4.7 а) на частицу не действует ни сила Лоренца, ни сила $e\mathbf{E}$, так что в этом направлении движение будет равномерным — $v_z = \text{const}$. Но оказывается, мы можем в данной задаче выделить не только это сносное движение, но и усредненное движение в плоскости xOy .

Формула (4.7) допускает интересную трактовку. Перейдем в систему отсчета, связанную с частицей. В этой системе частица имеет нулевую скорость, а значит, не работает и сила Лоренца. Но сама сила остается, и ее невозможно трактовать иначе, как перенормировку электрического поля в движущейся системе отсчета:

$$\mathbf{E}' = \mathbf{E} + [\mathbf{v} \mathbf{B}]. \quad (4.27)$$

Теперь попробуем найти такую систему отсчета, в которой $\mathbf{E}' = 0$. Умножаем (4.27) векторно на \mathbf{B} :

$$[\mathbf{E} \mathbf{B}] = -[[\mathbf{v} \mathbf{B}] \mathbf{B}] = \mathbf{v} B^2 - \mathbf{B}(\mathbf{v} \mathbf{B}).$$

Поскольку мы уже отделили движение параллельно магнитному полю, второй член справа можно без ограничения общности приравнять нулю. Таким образом, в системе отсчета, движущейся со скоростью

$$\langle \mathbf{v} \rangle = \frac{[\mathbf{E} \mathbf{B}]}{B^2}, \quad (4.28)$$

частица не будет вообще чувствовать электрическое поле, и ей останется лишь совершать циклотронное вращение согласно (4.8). Только это вращение и зависит от заряда и массы частицы, усредненное же движение описывается формулой (4.28) универсальным образом. Полная траектория есть результат сложения усредненного движения с $\mathbf{v} \parallel Ox \perp \mathbf{E}, \mathbf{H}$ и циклотронного вращения в плоскости xOy ; вид ее в зависимости от начальных условий представлен на рис. 4.7 б. В частности, она может быть и просто прямолинейной — это движение со скоростью (4.28) без вращения. Конечно, все это верно только при условии $\langle v \rangle = E/\mu_0 H \ll c = (\mu_0 \varepsilon_0)^{1/2}$, откуда и следует ограничение $\varepsilon_0 E^2 \ll \mu_0 H^2$. Описанный эффект называется *электрическим дрейфом*.

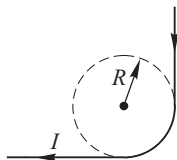


Рис. 4.8

4. Определить силу притяжения, приходящуюся на единицу длины каждого из двух тонких параллельных прямых проводов, если ток в них $I_1 = I_2 = 1$ А, а расстояние между проводами $x = 1$ м.

Ответ: $dF/dl = \mu_0 I_1 I_2 / (2\pi x) = 2 \cdot 10^{-7}$ Н/м.

5. Электрический ток I протекает по проводу, изогнутому так, как показано на рис. 4.8. Найти значение магнитной индукции в центре окружности (радиус окружности R).

Ответ: $B = \mu_0 I / (8\pi R)(\pi + 4)$.

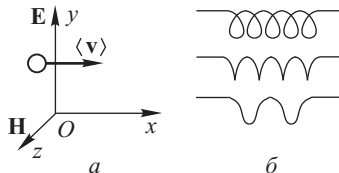


Рис. 4.7

Глава 5

МАГНИТНОЕ ПОЛЕ В ВЕЩЕСТВЕ

5.1. Микро- и макроскопическое описание поля в веществе

Уравнения магнитного поля, полученные в предыдущей главе, будут верны и в том случае, если пространство заполнено магнитоактивной средой, но при этом их следует считать микроскопическими уравнениями поля, порождаемого совокупно сторонними токами и откликом среды. Последний можно формализовать в виде некоторого наведенного распределения токов. Еще Ампер предполагал, что вещество магнетика представимо в виде совокупности магнитных диполей — элементарных витков с током, которые он называл молекулярными токами. Как показали последующие работы, эта гипотеза, безусловно, содержала элемент физической реальности, только отклик среды едва ли можно свести к молекулярному уровню. Лишь эффект парамагнетизма представим с некоторыми оговорками моделью Ампера. Диамагнетизм объясняется либо откликом на внешнее поле электронных оболочек атомов, либо откликом обобществленных электронов проводимости в кристалле. Наиболее сложна природа ферромагнетизма — этот круг явлений обусловлен квантовым эффектом упорядочения спиновых и орбитальных моментов электронов. Вообще адекватное представление магнетиков зачастую с необходимостью требует перехода на язык квантовой механики. Мы займемся этим в разделе, посвященном строению вещества, а здесь ограничимся феноменологическим описанием магнитных явлений, в минимальной степени адресуясь к их микроскопической природе.

Проблема состоит в том, что мы не можем реально использовать точные микроскопические уравнения, включающие «поштучно» все частицы магнетика, и должны перейти к какому-то усредненному описанию, в котором отклик частиц должен быть представлен как некоторая реакция сплошной среды. Ситуация подобна таковой в диэлектриках; здесь так же перестают быть тождественными понятия индукции и поля, только индуцируются уже не связанные заряды, а некоторые микроскопические токи. Микроскопичность в данном случае означает не только малый пространственный масштаб наведенных диполей. Токи эти — атомной природы. Они, например, не испытывают сопро-

тивления и не приводят к джоулеву тепловыделению. Мы их представим некоторым распределением $\mathbf{j}_{\text{микро}}(\mathbf{r})$, которое затем будем усреднять вместе с уравнениями.

Исходная система уравнений, включающая микроскопические токи и поля, имеет вид:

$$\operatorname{div} \mathbf{b} = 0; \quad \mathbf{b} = \mu_0 \mathbf{h}; \quad \operatorname{rot} \mathbf{h} = \mathbf{j} + \mathbf{j}_{\text{микро}}. \quad (5.1)$$

Здесь $\mathbf{j}(\mathbf{r})$ — распределение обычных макроскопических токов; их традиционно именуют токами проводимости.

От этого описания мы должны перейти к некоторому усреднению системы (5.1):

$$\mathbf{B} = \langle \mathbf{b} \rangle; \quad \operatorname{div} \mathbf{B} = 0; \quad \operatorname{rot} \mathbf{B} = \mu_0 (\mathbf{j} + \langle \mathbf{j}_{\text{микро}} \rangle). \quad (5.2)$$

Чтобы представить $\langle \mathbf{j}_{\text{микро}} \rangle$ какой-либо макроскопической принципиально измеряемой величиной, воспользуемся определением (4.15) для вектор-потенциала. После усреднения по физически бесконечно малому объему эта формула принимает вид:

$$\mathbf{A} = \frac{1}{4\pi} \int \frac{\mathbf{j} dV}{R} + \frac{1}{4\pi} \int \frac{\langle \mathbf{j}_{\text{микро}} \rangle dV}{R}.$$

С другой стороны, мы можем представить ту же самую величину \mathbf{A} через наведенный дипольный момент единицы объема \mathbf{J} — его называют также *намагничиванием*. Полагая $d\mathbf{m} = \mathbf{J} dV$, получаем из (4.22) альтернативное выражение для поправки к (4.15), учитывая отклик среды:

$$\delta \mathbf{A} = \int \frac{[d\mathbf{m}, \mathbf{R}]}{4\pi R^3} = \int \frac{[\mathbf{J} \mathbf{R}]}{4\pi R^3} dV.$$

Далее строим следующую цепочку формальных математических преобразований:

$$\operatorname{rot} \frac{\mathbf{J}}{R} \equiv \frac{[\nabla, \mathbf{J}]}{R} = \frac{\operatorname{rot} \mathbf{J}}{R} + \left[\mathbf{J}, \nabla \frac{1}{R} \right] = \frac{\operatorname{rot} \mathbf{J}}{R} - \frac{[\mathbf{J} \mathbf{R}]}{R^3},$$

что позволяет переписать нашу поправку в следующей форме:

$$\delta \mathbf{A} = \int \frac{\operatorname{rot} \mathbf{J}}{4\pi R} dV - \frac{1}{4\pi} \int \operatorname{rot} \left(\frac{\mathbf{J}}{R} \right) dV.$$

Интегрируя по всему пространству, последний интеграл можем положить равным нулю. Действительно, под знаком интеграла стоит операция rot , представляющая собой комбинацию производных, и если мы на достаточно больших расстояниях выходим за пределы магнетика, то в подстановках получаем нуль, потому что вне магнетика $\mathbf{J} \equiv 0$. А теперь приравняем два полученных выражения для $\delta \mathbf{A}$:

$$\int \frac{\operatorname{rot} \mathbf{J}}{4\pi R} dV = \int \frac{\langle \mathbf{j}_{\text{микро}} \rangle dV}{4\pi R}.$$

В отсутствие каких-либо специальных математических условий отсюда следует равенство подынтегральных выражений

$$\langle \mathbf{j}_{\text{микро}} \rangle = \text{rot } \mathbf{J},$$

которое мы и подставим в (5.2):

$$\text{rot } \mathbf{B} = \mu_0 (\mathbf{j} + \text{rot } \mathbf{J}). \quad (5.3)$$

Далее доопределим напряженность поля \mathbf{H} «вакуумным» уравнением, в которое включим только токи проводимости:

$$\text{rot } \mathbf{H} = \mathbf{j}, \quad (5.4)$$

как следствие получаем

$$\mathbf{B} = \mu_0 (\mathbf{H} + \mathbf{J}). \quad (5.5)$$

В действительности (5.3) и (5.4) связывают поля \mathbf{B} и \mathbf{H} с точностью до градиента произвольной скалярной функции, поскольку $\text{rot } (\nabla f) \equiv 0$. Мы знаем, однако, что магнитное поле не может быть потенциальным, а потому исключаем такую неопределенность. Во многих практически интересных случаях, в частности, во всех случаях пара- и диамагнетизма, можно с достаточной точностью аппроксимировать зависимость намагничения от поля токов простой линейной зависимостью:

$$\mathbf{J} = \chi \mathbf{H}.$$

Тогда и связь между полем и индукцией оказывается линейной:

$$\mathbf{B} = \mu \mu_0 \mathbf{H}; \quad \mu = 1 + \chi. \quad (5.6)$$

Величина χ называется *магнитной восприимчивостью*, а μ — *магнитной проницаемостью*. В случае линейной связи (5.6) это просто константа данного вещества. Не столь проста ситуация в случае ферромагнетиков. Во-первых, зависимость $B(H)$ становится нелинейной, поэтому в таких веществах не работает принцип суперпозиции полей. Во-вторых, она оказывается еще и неоднозначной: значение индукции зависит не только от величины поля, но и от истории магнетика. В-третьих, сама величина \mathbf{H} определяется не только внешними токами, но и геометрией образца. В гл. 2 при описании сегнетоэлектриков мы уже встречались с эффектами такого рода. Хотя уравнение (5.4) идентично уравнению поля в вакууме, его, как мы вскоре увидим, приходится решать при других граничных условиях, так что и свойства этой векторной величины будут другими, скажем, возможна ситуация, когда $\mathbf{j} \equiv 0$, но $\mathbf{H} \neq 0$. Соотношение (5.5) справедливо всегда при учете оговорок относительно природы вектора \mathbf{H} . Но формула (5.6) верна лишь для безграничной

сплошной среды. В одном и том же внешнем поле образцы из одного и того же материала с одной и той же константой μ , но разной формы, будут намагничены по-разному, т. е. разными будут величины $\mathbf{J}(\mathbf{r})$ и $\mathbf{B}(\mathbf{r})$.

Если, как это нередко бывает, образец магнетика представляет собой монокристалл, то, даже в случае линейной связи, одной константы для описания его свойств недостаточно. В общем случае магнитные восприимчивость и проницаемость оказываются величинами тензорной природы:

$$J_{\alpha} = \sum_{\beta} \chi_{\alpha\beta} H_{\beta}; \quad B_{\alpha} = \sum_{\beta} \mu_{\alpha\beta} H_{\beta}; \quad \alpha, \beta = x, y, z. \quad (5.7)$$

Геометрические особенности данного магнитоактивного образца, помимо анизотропных зависимостей $\chi_{\alpha\beta}$ или $\mu_{\alpha\beta}$, обусловленных строением вещества магнетика, представляются обычно так называемым *фактором размагничивания* $N_{\alpha\beta}$. Он как раз и учитывает форму образца и геометрию приложенного к нему внешнего поля. При использовании этого фактора можно уже без всяких оговорок рассматривать внешнее поле $\mathbf{H}_{(\text{вн})}$ как вакуумное поле токов проводимости, но действующее поле в веществе оказывается от него отличным:

$$H_{\alpha} = \mathbf{H}_{(\text{вн})\alpha} - \sum_{\beta} N_{\alpha\beta} J_{\beta}. \quad (5.8)$$

Из уравнения (5.2), а равно и всей процедуры с вектор-потенциалом следует, что для величины \mathbf{B} можно пользоваться теоремой о циркуляции или, что то же, законом Био и Савара, рассматривая $\langle \mathbf{j}_{\text{микро}} \rangle$ как некий ток, дополнительный к току проводимости. Это и есть ток намагничения. Необходимо только помнить, что сопротивление этому току равно нулю, поэтому он не фигурирует ни в законе Ома, ни в законе Джоуля–Ленца. Не следует также подставлять этот ток в теорему циркуляции для \mathbf{H} . Не всегда просто перевести на этот язык эффект размагничивающего фактора. Но иногда такая модель отклика магнетика на внешнее поле оказывается для решения задач удобнее, чем модель магнитной восприимчивости; очевидно, физически они совершенно эквивалентны.

5.2. Диа- и парамагнетизм

Если главное следствие поляризации диэлектрика — ослабление действующего поля \mathbf{E} в сравнении с полем свободных зарядов, то в парамагнетике, напротив, действующее поле \mathbf{B} больше приложенного внешнего поля. Правда, отличие это невелико: обычно у парамагнетиков величина χ не превышает 10^{-4}

для некоторых химических соединений, а у чистых металлов $\chi \sim 10^{-7}$.

Мы начали с парамагнетизма, поскольку на уровне модельных соображений, скажем, молекулярных токов Ампера, эффект этот может быть представлен проще, чем другие. Отличие знака эффекта от случая диэлектриков связано с принципиальным различием электрического и магнитного диполей (рис. 4.4 а). Как следствие, выстраивание диполей не ослабляет, а усиливает поле в веществе. В действительности простая модель выстроенных элементарных магнитиков более или менее адекватна парамагнитному веществу в газообразном состоянии, а также некоторым жидкостям. В металлах существенную роль играет спиновый парамагнетизм электронов проводимости, в полупроводниках — электронный или «дырочный» парамагнетизм, в сложных химических соединениях он может быть обусловлен ионами того или иного сорта. В парамагнитном состоянии оказываются и ферромагнетики, нагретые выше определенной температуры, т. н. температуры Кюри. Магнитная восприимчивость большинства парамагнетиков в зависимости от температуры подчиняется закону Кюри–Вейсса (по имени французского физика П.Э. Вейсса (1865–1940)):

$$\chi \propto \frac{1}{T - \Theta},$$

где Θ для ферромагнетика при высокой температуре как раз примерно и равна температуре Кюри (обычно порядка нескольких сотен градусов Кельвина); для истинных парамагнетиков эта величина заметно ниже и никакой реальной особенности магнитной восприимчивости не соответствует.

Казалось бы, парамагнетизмом и должна была бы ограничиться реакция вещества на магнитное поле, коль скоро он допускает настолько простое и естественное объяснение. Но известно, что существует никак не менее широкий класс веществ с диамагнитными свойствами. У них магнитная восприимчивость обычно того же порядка $\chi \sim 10^{-7}$ – 10^{-6} , но другого знака. Дело в том, что, как уже было замечено, магнитостатика — в принципе не совсем статика, ибо ток есть движение зарядов; соответственно, и реакция на внешнее магнитное поле носит не только статический, но и динамический характер. Забегая несколько вперед, можно сказать, что эффект диамагнетизма есть не что иное, как электромагнитная индукция на уровне молекулярных токов. Поэтому диамагнитные свойства присущи всем веществам без исключения, и если какое-то вещество все же оказывается парамагнетиком, то это значит, что собственный магнитный момент каждой частицы достаточно велик, чтобы обеспечить результирующий парамагнитный эффект.

Поведение электронных оболочек атомов во внешнем магнитном поле \mathbf{B} вполне может быть смоделировано в рамках классической физики, т. е. без привлечения квантовых эффектов. Пусть система заряженных частиц с одинаковым для всех частиц отношением заряда к массе e/m совершает в заданном поле финитное движение. Средняя по времени сила, действующая на систему, равна нулю,

$$\langle \mathbf{F} \rangle = \left\langle \sum_i e_i [\mathbf{v}_i \mathbf{B}] \right\rangle = \left\langle \frac{d}{dt} \sum_i e_i [\mathbf{r}_i \mathbf{B}] \right\rangle = 0,$$

в силу уже известного нам свойства равенства нулю средней производной величины, меняющейся в конечных пределах. Вычислим средний момент силы, действующий на систему:

$$\langle \mathbf{M} \rangle = \left\langle \frac{d}{dt} \mathbf{L} \right\rangle = \left\langle \sum_i e_i [\mathbf{r}_i [\mathbf{v}_i \mathbf{B}]] \right\rangle = \left\langle \sum_i e_i (\mathbf{v}_i (\mathbf{r}_i \mathbf{B}) - (\mathbf{r}_i \mathbf{v}_i) \mathbf{B}) \right\rangle. \quad (5.9)$$

Последний член в правой части при усреднении опять дает нуль, поскольку пропорционален $\langle d\mathbf{r}^2/dt \rangle$. Первый член в скобках может быть представлен в виде

$$\mathbf{v}_i (\mathbf{r}_i \mathbf{B}) = \frac{d}{dt} \{ \mathbf{r}_i (\mathbf{r}_i \mathbf{B}) \} - \mathbf{r}_i (\mathbf{v}_i \mathbf{B}).$$

Первый член в правой части опять дает нуль при усреднении, поэтому, пока и поскольку мы оперируем с усредненными уравнениями движения, правую часть (5.9) можно представить в виде полусуммы двух равных величин:

$$\left\langle \frac{d}{dt} \mathbf{L} \right\rangle = \sum_i \frac{e_i}{2} \{ \mathbf{v}_i (\mathbf{r}_i \mathbf{B}) - \mathbf{r}_i (\mathbf{v}_i \mathbf{B}) \} = \sum_i \frac{e_i}{2} [[\mathbf{r}_i \mathbf{v}_i] \mathbf{B}] \equiv [\mathbf{m} \mathbf{B}]. \quad (5.10)$$

Именно здесь мы воспользуемся нашим предположением о том, что все частицы имеют одинаковое отношение заряда к массе. Это значит, что мы пренебрегаем влиянием магнитного поля на движение ядер, что в нашем случае справедливо с точностью до отношения масс, которое даже у водорода порядка 2000. Обращаясь к выражению (4.26) для гиромагнитного отношения, перепишем (5.10) в виде

$$\left\langle \frac{d}{dt} \mathbf{L} \right\rangle = -\Omega \left[\frac{\mathbf{B}}{B}, \mathbf{L} \right]; \quad \Omega = \frac{eB}{2m}, \quad (5.11)$$

а это не что иное, как уравнение регулярной прецессии с угловой скоростью $-e\mathbf{B}/(2m)$. Сравнивая эту угловую скорость с угловой скоростью циклотронного вращения (4.8), можно указать на различие в знаке и на дополнительный коэффициент $1/2$. Угловая частота (5.11) называется *ларморовой частотой*, а факт прецессии системы заряженных частиц в магнитном поле со-

ставляет содержание *теоремы Лармора* (по имени английского физика Дж. Лармора (1857–1942)).

Именно ларморова прецессия оказывается ответственной за диамагнитный отклик вещества. Движение системы частиц — электронного облака атома — происходит таким образом, что в результате генерируется добавка к внешнему полю $\delta\mathbf{B}$, антипараллельная \mathbf{B} .

5.3. Ферромагнетизм

С феноменологической точки зрения ферромагнетики характеризуются следующими свойствами:

1. Если у пара- и диамагнетиков $|\mu - 1| \ll 1$, то у ферромагнетиков значение магнитной проницаемости достигает величин порядка 10^4 и более.

2. Если у диа- и парамагнетиков зависимость $\mathbf{B}(\mathbf{H})$ линейна и однозначна, то у ферромагнитных материалов она не только нелинейна, но и зависит от истории образца. В частности, они могут существовать в виде постоянных магнитов.

3. В отличие от диа- и парамагнетизма, явление ферромагнетизма не может быть смоделировано в рамках классической физики; природа его — исключительно квантовая. Таким образом, как раз для одного из самых популярных феноменов электродинамики — постоянного магнита — отсутствует возможность простого популярного объяснения.

По этой причине мы вынуждены будем вернуться к этому кругу явлений в заключительном разделе курса, здесь же будет дано лишь краткое феноменологическое описание. К ферромагнетикам относятся Fe, Co, Ni, редкоземельные металлы, а также многочисленные сплавы и интерметаллические соединения. Магнитные свойства их весьма разнообразны; мы, естественно, коснемся лишь некоторых, самых общих.

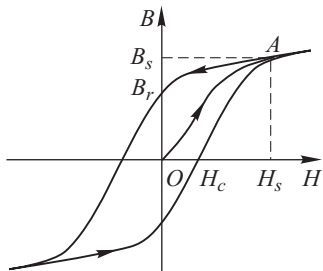


Рис. 5.1

Главная характеристика любого ферромагнитного материала — кривая гистерезиса, представляющая собой нелинейную зависимость намагничивания от внешнего поля. Она может быть представлена в форме $B(H)$ или $J(H)$. На рис. 5.1 показана кривая $B(H)$. Выходя из точки O , мы попадаем в конце концов в точку насыщения A , которой соответствует поле насыщения B_s . Если мы начнем

уменьшать поле H , не доходя до точки A , кривая $B(H)$ поведет себя сложным образом, формируя малые петли, которые, однако,

не будут замкнуты. Если увеличивать поле до значений, превышающих H_s , зависимость $B(H)$ становится линейной, а на кривой $J(H)$ наступает насыщение. Помимо поля и индукции насыщения, характерными величинами будут также остаточная намагниченность B_r и так называемая коэрцитивная сила H_c — см. рис. 5.1. Используется также понятие дифференциальной магнитной проницаемости $\mu' = \mu_0^{-1} dB/dH$. Оно удобно как характеристика начальной кривой $B(H)$ или ее прямолинейного участка. Нередко употребляется энергия потерь $\delta A = \oint H dB$ при прохождении одного цикла кривой гистерезиса.

Вещества называются магнитомягкими, если $H_s < 10^3$ А/м. Для них характерны высокие значения $\mu_{\max} \sim 10^4$ – 10^6 ; $\mu'_{\max} \sim 10^4$ – 10^5 . Коэрцитивная сила обычно невелика, $H_c \leq \leq 10$ – 10^3 А/м, невелики и потери на перемагничивание: $\delta A \leq \leq 1$ – 10 Дж/м³. У магнитотвердых материалов $H_s \sim 10^3$ – 10^5 А/м, для них характерны также высокие значения B_r , H_c , B_s , δA .

Таким образом, для магнитотвердых веществ характерна широкая кривая гистерезиса, тогда как поведение магнитомягких материалов зачастую можно аппроксимировать просто нелинейной зависимостью $B(H)$ — или даже линейной, но с насыщением намагничивания при $H \geq H_s$.

Возможна двоякая трактовка кривой гистерезиса, и это обстоятельство требует особого внимания. Мы в какой-то мере уже касались данного вопроса в гл. 2 в связи с проблемой сегнетоэлектриков. Как там мы рассматривали в качестве примера палочку из электрета, так здесь мы рассмотрим линейный постоянный магнит. Поскольку линии \mathbf{B} всегда замкнуты, они будут вести себя так же, как и у токового диполя — витка или соленоида (рис. 4.4а). В пустоте линии \mathbf{H} с ними совпадают, но внутри магнита они обязаны иметь противоположное направление. Действительно, в такой системе отсутствуют токи проводимости, а ток намагничивания не входит в теорему о циркуляции. Поэтому, взяв $\oint \mathbf{H} d\mathbf{l}$ по любой линии \mathbf{B} , мы должны получить ноль.

Таким образом, линии \mathbf{H} должны выглядеть так же, как и силовые линии электрического диполя на том же рис. 4.4а. В принципе это можно даже представить как свободные северный и южный магнитные заряды на торцах магнита. Последнему на кривой гистерезиса должна соответствовать точка в одном из четных квадрантов.

Мы пришли к такому выводу, имея дело с постоянным магнитом вполне определенной геометрии. Если бы, например, речь шла о намагниченном тороиде (\mathbf{B} направлено по обходу тора), то в этом случае из той же теоремы о циркуляции следовало бы $H \equiv 0$. Разница между этими двумя магнитами — в различных

размагничивающих факторах, которые прежде всего и зависят от геометрии тела. Обращаясь к кривой гистерезиса, необходимо отдавать себе отчет в том, что именно отложено по оси абсцисс: поле внешних токов или истинное поле в веществе. Вторая трактовка во многих отношениях более привлекательна, поскольку, как мы увидим в следующем параграфе, пара векторных величин \mathbf{B} , \mathbf{H} может непосредственно измеряться и представлять состояние вещества в магнитном поле. Но при этом мы уже не вправе считать \mathbf{H} просто полем токов проводимости, потому что уравнение (5.4) решается при граничных условиях, несколько отличных от таковых для вакуумного поля.

Природа ферромагнетизма обусловлена квантовым эффектом *обменного взаимодействия*. Этот термин при неосторожном употреблении может ввести читателя в заблуждение. В физике известно четыре типа взаимодействий: гравитационное, сильное, слабое и электромагнитное. Обменное взаимодействие этого списка отнюдь не расширяет. В данном конкретном случае речь идет об известном нам электромагнитном (кулоновском) взаимодействии между электронами, которому квантовые эффекты придают некоторые особенности. В частности, в веществах с определенной архитектурой электронных оболочек атомов данный эффект приводит к выстраиванию электронных магнитных моментов строго параллельным образом.

Сделаем простую оценку. Собственный магнитный момент электрона (т. е. *магнетон Бора*)

$$|\mathbf{m}_e| \equiv \mu_B = \frac{eh}{4\pi m} \approx 9,3 \cdot 10^{-24} \text{ Дж/Тл}, \quad (5.12)$$

где $e = 1,6 \cdot 10^{-19}$ Кл — заряд электрона, $m = 9,1 \cdot 10^{-31}$ кг — его масса, $h = 6,63 \cdot 10^{-34}$ Дж·с — постоянная Планка. (К единицам магнитного момента мы еще вернемся в следующей главе.) Далее, постоянная решетки a — характерное расстояние между атомами — большинства металлов есть величина порядка $2\text{--}3 \cdot 10^{-10}$ м. Предположим, что на каждый атом приходится один такой «выстроенный» электрон и оценим, сколь сильным окажется поле в веществе, в соответствии с формулой (5.5) при $H = 0$:

$$B = \mu_0 J \approx \mu_0 \frac{|\mathbf{m}_e|}{a^3} \sim 1 \text{ Тл}.$$

Это довольно сильное поле. Например, чтобы получить его на расстоянии 1 см от длинного тонкого прямолинейного проводника, по нему следует пропустить ток порядка 50 кА. В принципе такое спонтанное (самопроизвольное) намагничение вещества энергетически невыгодно, а наиболее выгодно как раз нулевое среднее поле, которое получается уже за счет макроскопического «беспорядка». Вещество разбивается на макроскопические (хотя

и очень мелкие) области — *домены*, суммарные магнитные моменты которых ориентированы хаотически, как изображено на рис. 5.2а. Намагничивание ферромагнетика означает постепенное, по мере роста H , выстраивание этих магнитных моментов, а насыщение — полное выстраивание, когда весь образец превращается в единственный домен (рис. 5.2б). Подчеркнем, что процесс намагничивания не связан с макроскопическим перемещением вещества — происходит лишь переориентация магнитных моментов электронов. Тем не менее, перемещение и последующая ликвидация доменных границ требует энергетических затрат. Отсюда и гистерезис, этим и обусловлена возможность существования постоянных магнитов (которые, как теперь легко сообразить, делаются из магнитотвердых материалов).

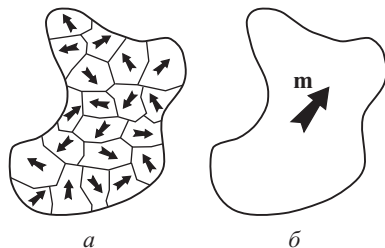


Рис. 5.2

У парамагнетиков нет такого эффекта выстраивания. Поэтому в них ориентация магнитных диполей происходит только за счет внешнего поля, на фоне теплового движения, и степень ее обычно очень невелика. Это, между прочим, следует из только что сделанной оценки. Дело в том, что магнитный момент атома, если это не нуль (у диамагнитного вещества) тоже равен магнетону Бора или кратен ему. Поэтому, если ориентация диполей в парамагнетике была бы столь же или почти столь же эффективной, то и намагниченность его была бы на уровне ферромагнетика. Заметим, что в принципе это возможно — и даже насыщение намагничивания можно наблюдать экспериментально, но лишь при условии

$$B > \frac{k_B T}{\mu_B},$$

где k_B — постоянная Больцмана, T — абсолютная температура. Оценив последнюю как 300 К, получим $B > 10^2$ Тл. Такие поля трудно получить даже в импульсном режиме, а тем более — в стационарном. Но при низких температурах эффект вполне наблюдаем.

Из рис. 5.1 можно понять, что, изменяя поле в пределах кривой гистерезиса, мы не сможем снять намагниченность. Это, однако, можно сделать, поместив образец в переменное поле H , амплитуда которого медленно уменьшается до нуля. При этом зависимость $\mathbf{B}(H)$ многократно обойдет нулевую точку, постепенно к ней приближаясь. Другой способ размагничивания связан с существованием у ферромагнетиков точки Кюри Θ ,

выше которой они переходят в парамагнитное состояние и теряют собственный магнитный момент. По порядку величины эта температура обычно не превышает сотен градусов Цельсия, что открывает возможность неразрушающего отжига.

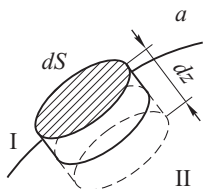
Ферромагнетики представляют собой лишь частный случай более широкого класса веществ, называемых *ферритами*, но это важнейший случай с точки зрения приложений. В рамках общего курса электродинамики мы должны будем им и ограничиться.

5.4. Граничные условия на поверхности раздела

Как и в электростатике, граничные условия принципиально важны как для формального описания поля, так и для понимания физики явления. В вакууме уравнения $\operatorname{div} \mathbf{B} = 0$ и $\operatorname{rot} \mathbf{H} = \mathbf{j}$ решались совместно; эти величины отличались только коэффициентом μ_0 . В общем случае они разорваны и их решения связаны соотношением (5.5), в котором величина \mathbf{J} не универсальна.

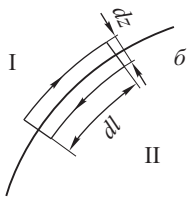
Необходимые граничные условия последуют из самих же основных уравнений. Рассмотрим границу раздела двух сред и построим малый цилиндр, основания которого площадью dS

расположены в средах I и II, а образующая перпендикулярна соприкасающейся плоскости (рис. 5.3 а). Поток магнитной индукции через замкнутую поверхность этого цилиндрика



$$\oint_S \mathbf{B} d\mathbf{S} = 0.$$

Затем, переходя для высоты к пределу $dz \rightarrow 0$ и считая основание dS достаточно малым, получаем после сокращения на dS



$$B_{In} = B_{II n}. \quad (5.13)$$

Это условие похоже на (2.13), но, в отличие от (2.14), здесь уже не может возникнуть никаких поверхностных зарядов — это принципиально невозможно.

К тому же, и смысл уравнения (5.13) другой: оно написано для действующего поля в произвольной среде.

Далее, согласно рис. 5.3 б, окружим поверхность раздела малым плоским контуром, две стороны которого имеют длину dl и расположены параллельно соприкасающейся плоскости, а две другие ей перпендикулярны и имеют длину dz . Согласно теореме о циркуляции имеем

$$\oint_L \mathbf{H} d\mathbf{l} = i \cdot dl,$$

где $i = dI/dl$ — линейная плотность поверхностного тока, перпендикулярного плоскости контура. Снова полагаем $dz \rightarrow 0$ и сокращаем на dl :

$$H_{I\parallel} - H_{II\parallel} = i_{\perp},$$

индексы « \parallel »; « \perp » определяют ориентацию относительно плоскости контура на рис. 5.3б. Повторив ту же операцию для компоненты поля, ориентированной перпендикулярно предыдущей в касательной плоскости, получим

$$H_{I\tau} - H_{II\tau} = i; \quad (5.14)$$

при желании это условие можно переписать и в векторной форме:

$$\mathbf{H}_{I\tau} - \mathbf{H}_{II\tau} = [i \mathbf{n}], \quad (5.15)$$

где \mathbf{n} — вектор нормали к поверхности раздела. Соотношения (5.14), (5.15) подразумевают конечный поверхностный ток в бесконечно тонком слое или, точнее, ненулевую толщину поверхностного слоя, на двух сторонах которого измеряются тангенциальные проекции вектора \mathbf{H} . Иногда, например, на поверхности сверхпроводника, только таким образом удастся свести задачу к чистой электродинамике; в остальных случаях (5.15) принимает более простую форму:

$$\mathbf{H}_{I\tau} = \mathbf{H}_{II\tau}. \quad (5.16)$$

Полезно еще раз обратить внимание на формальное сходство этой формулы с (2.12), но здесь речь идет не о действующем поле, но лишь о том, которое определяется токами проводимости; кроме того, какие-либо добавки в правой части (2.12) отсутствуют в принципе. Не следует сопоставлять (5.15) с (2.12), а (5.13) с (2.14), потому что электрическое и магнитное поля — объекты разной векторной природы.

Прежде всего рассмотрим граничные условия для магнитного поля на поверхности идеального проводника. Пусть поле внутри него равно нулю; в случае сверхпроводника это всегда верно (при переходе в сверхпроводящее состояние происходит выталкивание поля).

Если бы мы исходили из условий (5.13), (5.16), поле на границе проводника в вакууме должно было бы в точности обратиться в нуль. В действительности только условие (5.13) не допускает никакой модификации. Таким образом, нормальная компонента поля в вакууме у поверхности такого проводника равна нулю. К тангенциальной компоненте это не относится; просто из условия (5.15) следует некоторое распределение токов на поверхности проводника. В вакууме поле подходит к этой поверхности по касательной.

Отсюда, в частности, следует, что и к задачам такого рода применим метод изображений. Но поскольку здесь, в отличие от электростатики, в вакууме $\mathbf{B}_\tau \neq 0$, $\mathbf{B}_n = 0$, то в плоской поверхности проводника северный полюс магнита зеркально отразится в северный, а южный — в южный. Прямой провод с током отразится в прямой провод, по которому ток будет течь в противоположном направлении, и потому будет отталкиваться от проводящей плоскости.

Этот эффект используется для пространственной стабилизации импульсных сильноточных разрядов. Окружая область разряда коаксиальным хорошо проводящим (или даже сверхпроводящим) экраном, его как бы помещают в потенциальную яму и тем самым фиксируют, с точностью до малых отклонений, его положение в пространстве.

В случае непроводящего магнетика работают граничные условия (5.13), (5.16). Они дают нам, в частности, возможность определить корректным образом поле и индукцию в веществе магнетика посредством мысленного эксперимента или даже непосредственно их измерить, если образец допускает реальный эксперимент такого рода.

Представим себе, что мы вырезали в объеме магнетика малую щель, позволяющую ввести измерительный зонд. Щель должна быть столь мала, чтобы не исказить поле в магнетике, и иметь геометрию тонкого плоского листка, ориентированного перпендикулярно силовым линиям. Тогда внутри щели можно, в свою очередь, пренебречь влиянием краев этого листка, а на плоской поверхности справедливо граничное условие (5.13). Следовательно, индукция \mathbf{B} внутри щели должна быть равна таковой в объеме магнетика.

Повторим тот же мысленный эксперимент, но теперь вырежем щель так, чтобы силовые линии лежали в плоскости листка, или, точнее, чтобы плоскость листка совпадала с касательной плоскостью к силовым линиям в данной точке. Тогда, базирясь на граничном условии (5.16), нетрудно заключить: поле, измеренное в тонкой щели, ориентированной таким образом, совпадает с полем внутри магнетика.

Если эти измерения выполняются в реальности, щели и зонды в нужных местах готовятся заранее. Приходится, однако, признать, что так можно работать только с полями достаточно простой геометрии, когда линии \mathbf{B} и \mathbf{H} совпадают (либо когда мы умеем их различить — а это, строго говоря, можно сделать, лишь зная заранее \mathbf{B} и \mathbf{H}). Для нас более важно скорее другое. Именно в этом мысленном эксперименте мы окончательно уходим от представления о напряженности магнитного поля как поле свободных токов. Теперь два вектора \mathbf{B} и \mathbf{H} выступают как параметры состояния вещества в магнитном поле. На таком

понимании строится термодинамика магнетиков, нам же оно понадобится в следующей главе при обсуждении проблемы энергии поля.

5.5. Магнитные цепи

Данный параграф будет посвящен достаточно простому с точки зрения физики вопросу. Он, однако, важен для приложений, но обычно в курсах электротехники физические принципы метода магнитных цепей не особенно акцентируются.

Предположим, что нам дан замкнутый сердечник — рис. 5.4*а*, выполненный из магнитомягкого материала. Соответственно используем приближение $\mathbf{B} = \mu\mu_0\mathbf{H}$, причем $\mu \gg 1$. Длина по обходу сердечника l , сечение S , ширина зазора δ . Считается, что $l \gg S^{1/2} \gg \delta$. Число витков обмотки равно N , ток через обмотку I . Задача: найти поле в зазоре.

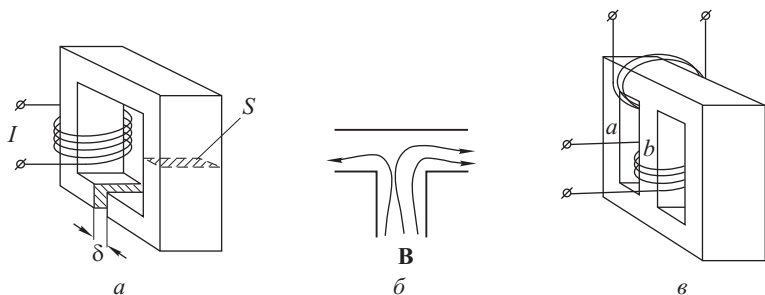


Рис. 5.4

Каждое из приведенных сильных неравенств принципиально важно для всей последующей вычислительной процедуры, об этом не следует забывать: $l \gg S^{1/2}$ позволяет пренебречь профилем поля по сечению, как следствие, сердечник уподобляется электрической цепи с сосредоточенными параметрами; $S^{1/2} \gg \delta$ дает нам возможность рассматривать поле в зазоре как одномерное, пренебрегая краевыми эффектами. Особую роль играет условие $\mu \gg 1$. Это значит, что в сердечнике $B \gg \mu_0 H$. При переходе через боковую границу сохраняется величина $\mathbf{H}_\tau \approx \mathbf{H}$, но индукция вне сердечника оказывается много меньше, чем внутри него: $B_{\text{вн}} = \mu_0 H \ll B$. И более того, предполагается, что магнитная проницаемость сердечника настолько больше единицы, что даже магнитный поток $\Phi = \int_S \mathbf{B} d\mathbf{S} \approx BS$, захваченный в сердечник, намного превосходит поток $\Phi_{\text{вн}} = \int B_{\text{вн}} d\mathbf{S}_{\text{вн}}$ вне его. Последнее предположение не будет слишком жестким в случае конкретной геометрии рис. 5.4*а*, но в общем случае оно

является наименее строгим, так как требует не просто сильных неравенств, но специального расчета для каждой частной конфигурации сердечника. В формально поставленных задачах это предположение принято заключать в формулировку «рассеяние поля отсутствует».

Полезно отметить, что именно условие $\Phi \gg \Phi_{\text{вн}}$ есть та цель, ради которой используются в катушках ферромагнитные сердечники. Это будет особенно важно в тех задачах, где существенны эффекты электромагнитной индукции (см. гл. 7) и в таких технических устройствах, как, например, трансформаторы.

Итак, если нет рассеяния, то Φ есть постоянная величина в любом сечении сердечника. Тогда $H = \Phi/(\mu\mu_0 S)$, а в зазоре $H^* = \Phi/(\mu_0 S)$ — связь между ними следует из сохранения нормальной компоненты индукции при переходе через границу зазора и малой ширины последнего. Используем теорему о циркуляции:

$$Hl + H^* \delta = NI.$$

Окончательный результат удобно записать в виде

$$\Phi \left(\frac{l}{\mu\mu_0 S} + \frac{\delta}{\mu_0 S} \right) = NI,$$

что абсолютно идентично закону Ома для некоторой замкнутой цепи при условии следующих переобозначений:

$$\Phi \Rightarrow I; \quad NI \Rightarrow \mathcal{E}; \quad \mu\mu_0 \Rightarrow \lambda; \quad \frac{l}{\mu\mu_0 S} \Rightarrow R. \quad (5.17)$$

Эта аналогия и составляет суть *метода магнитных цепей*. Строго говоря, все сильные неравенства, выполнения которых мы предварительно потребовали, не только легко переводятся на язык электрической цепи, но и подразумеваются там выполненными (обычно по умолчанию). Аналогию можно продолжить, поставив условие $\oint \mathbf{B} d\mathbf{S} = 0$ в соответствие закону сохранения заряда $\sum_i I_i = 0$ в разветвлении магнитной цепи (рис. 5.4б). Это позволяет перейти к рассмотрению более сложных сердечников.

Например, для случая, представленного на рис. 5.4в, вводим обозначения $\mu_0 H_a = \Phi_a/(\mu_a S_a)$; $\mu_0 H_b = \Phi_b/(\mu_b S_b)$, после чего можем записать

$$H_a l_a + H_b l_b = \sum_k N_k I_k \Rightarrow I_a R_a + I_b R_b = \sum_k \mathcal{E}_k. \quad (5.18)$$

Таким образом, в случае сердечников из магнитомягкого материала с большими μ , но при условии линейности отклика $B(H)$ мы можем использовать для расчетов законы Кирхгофа.

5.6. Эффект Холла

Весьма значительная группа явлений в твердом теле и плазме обязана своей природой эффекту, открытому американским физиком Э.Г. Холлом (1855–1938) в 1879 г. — когда еще неизвестны были ни электрон, ни сила Лоренца.

Идея опыта представлена на рис. 5.5. Электрический ток пропускается через образец, помещенный во внешнее магнитное поле \mathbf{B} таким образом, чтобы направление тока было перпендикулярно \mathbf{B} . Эффект состоял в том, что проводник поляризовался, т. е. между боковыми сторонами появлялась некоторая ЭДС, причем электрическое поле, ей соответствующее, оказывалось перпендикулярным как \mathbf{B} , так и \mathbf{j} . Варьируя материал образца, можно было убедиться, что холловская ЭДС может быть не только различной величины, но и разного знака.

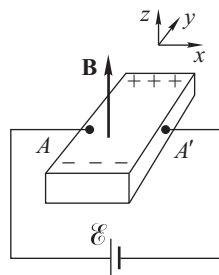


Рис. 5.5

Объяснить эффект удалось лишь на базе электронной теории. На любой носитель тока — электрон, ион в электролите или плазме, дырку в полупроводнике — действует при движении в магнитном поле сила Лоренца (4.7). Если бы речь шла об уединенной частице в магнитном поле, это привело бы лишь к циклотронному вращению (см. формулу (4.8)). Но в проводнике с высокой концентрацией носителей такое движение нарушило бы квазинейтральность. Поэтому в действительности немедленным откликом на боковое смещение, скажем, всех токовых электронов относительно всех ионов решетки будет поляризация образца. Поле в объеме установится как раз таким, чтобы скомпенсировать силу Лоренца:

$$e\mathbf{E} = e[\mathbf{v} \mathbf{B}].$$

Это удобно переписать в следующем виде:

$$\mathbf{E} = R_H [\mathbf{j} \mathbf{B}]; \quad R_H \equiv \frac{1}{ne}, \quad (5.19)$$

где n — концентрация носителей, e — их заряд с учетом и его знака. Величину R_H принято называть *постоянной Холла* — это константа вещества.

Тот же самый результат можно получить из других соображений. В магнитном поле заряженные частицы не смогли бы преодолеть промежуток AA' (рис. 5.5), если бы не сформировали электрическое поле, ортогональное приложенному магнитному. В такой конфигурации полей направленное движение уже возможно (см. задачу 3 к гл. 4 — формула (4.28)). Среднюю

скорость направленного движения

$$\langle \mathbf{v} \rangle = \frac{[\mathbf{E} \mathbf{B}]}{B^2}$$

следует приравнять скорости носителей $\mathbf{j}/(ne)$. Направление последней зависит от знака заряда носителей, тогда как дрейфовая скорость (4.28) от заряда не зависит. Поэтому направление холловского электрического поля определяется знаком заряда доминирующих носителей тока.

На первый взгляд, постоянная Холла как константа вещества имела смысл лишь до появления электронной теории электропроводности. В действительности наше объяснение оказывается во многих случаях слишком модельным. Понять это можно, представив себе в одном проводнике два вида носителей разного знака (пример — полупроводник) или одноименных, но с разным зарядовым числом Z (электролит). Формула (5.19) с очевидностью оказывается неадекватной, но сам эффект Холла остается в силе, а величина R_H может быть измерена экспериментально либо вычислена в рамках достаточно полной модели.

На сегодняшний день холловскими обычно называют любые явления, в которых существенны и нетривиальны эффекты вида $[\mathbf{j} \mathbf{B}]$. Из многочисленных приложений отметим, например, возможность прямого и очень точного измерения магнитного поля.

Задачи

1. На железный сердечник постоянного сечения длиной $l = 1$ м с зазором $\Delta = 1$ мм намотана катушка с числом витков $N = 1600$, по которой течет ток $I = 1$ А (рис. 5.6 а). Зависимость $B(H)$ материала сердечника представлена на рис. 5.6 б. Определить поле в зазоре.

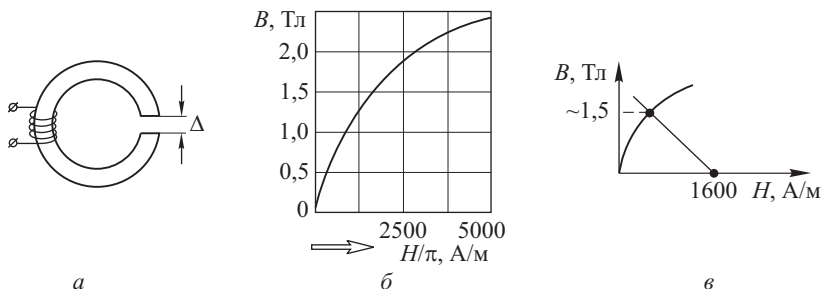


Рис. 5.6

Решение. В силу граничного условия для нормальной компоненты \mathbf{B} значение B в зазоре равно B в сердечнике. Используем теорему о циркуляции

$$\frac{B}{\mu_0} \Delta + H l = N I. \quad (5.20)$$

А теперь объединим график рисунка 5.6б и линейную зависимость $B(H)$, следующую из уравнения (5.20). Результат представлен на рис. 5.6в.

Ответ: $B \approx 1,5 \text{ Тл}$.

2. Безграничная плоская магнитная пленка толщины h включает одну доменную стенку G , разграничивающую две полуплоскости с противоположной намагниченностью $\mp \mathbf{J}_0$ (рис. 5.7). Вектор \mathbf{J}_0 ортогонален к пленке. Пленка помещена в однородное электрическое поле $\mathbf{E} \parallel \mathbf{J}_0$. Над границей раздела на расстоянии $L \gg h$ параллельно ей движется электрон с постоянной скоростью \mathbf{v} . При какой величине E такое движение возможно?

Решение. Скачок индукции \mathbf{B} на плоской доменной границе может быть связан с линейной плотностью тока намагничивания $i = dI/dl$ (см. задачу 2 гл. 4):

$$\Delta \mathbf{B} = 2\mathbf{J}_0; \quad \Delta B = i \Rightarrow i = 2J_0.$$

Поскольку электрон находится от пленки на расстоянии $L \gg h$, для него ток намагничивания будет выглядеть как линейный ток $I = ih$, создающий в точке, где находится электрон, поле

$$H = \frac{I}{2\pi L} = \frac{J_0 h}{\pi L}.$$

Согласно результату задачи 3 гл. 4 электрон может двигаться равномерно в скрещенных полях только при условии $\varepsilon_0 E^2 \ll \mu_0 H^2$. Таким образом, для нашего случая получаем ответ:

$$E \ll \left(\frac{\mu_0}{\varepsilon_0} \right)^{1/2} \frac{J_0 h}{\pi L}.$$

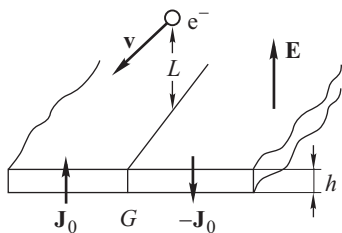


Рис. 5.7

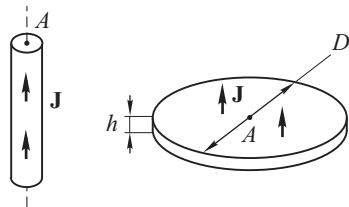


Рис. 5.8

3. Бесконечная плоская пластина изготовлена из однородно намагниченного ферромагнетика. Найти поля \mathbf{B} и \mathbf{H} внутри и вне пластины, если вектор \mathbf{J} а) параллелен и б) перпендикулярен плоскости пластины.

Ответ: а) внутри пластины $\mathbf{H} = 0$, $\mathbf{B} = \mu_0 \mathbf{J}$; вне пластины $\mathbf{H} = 0$, $\mathbf{B} = 0$, б) внутри пластины $\mathbf{B} = 0$, $\mathbf{H} = -\mathbf{J}$; вне пластины $\mathbf{B} = 0$, $\mathbf{H} = 0$.

4. Длинный однородный цилиндр изготовлен из материала с «замороженной» однородной намагниченностью, вектор которой параллелен его оси. Индукция в точке A оказалась равной $B_A = 0,1 \text{ Тл}$ (рис. 5.8). Найти индукцию B вблизи торца короткого цилиндра, изготовленного из того же материала, если $h = 5 \cdot 10^{-2} D$.

Ответ: $B = 2B_A (h/D) = 10^{-2} \text{ Тл}$.

Глава 6

ЭЛЕКТРОМАГНИТНАЯ ИНДУКЦИЯ

6.1. Индуктивность и взаимная индукция

В этом параграфе мы от рассмотрения сплошных сред в очередной раз возвращаемся к системам с сосредоточенными параметрами — электрическим цепям. В отличие от гл. 3, где в электрической цепи важна была только последовательность элементов и законы формирования контуров (в случае их неодносвязности), теперь для нас первостепенное значение приобретает геометрия контура, поскольку мы рассматриваем магнитное поле токов. И здесь весьма существенный выигрыш в объеме вычислений даст нам введенное в (4.3) понятие вектор-потенциала:

$$\mathbf{A} = \frac{1}{4\pi} \int \frac{\mathbf{j}(\mathbf{r})}{r} dV.$$

Рассмотрим два замкнутых контура 1 и 2. Пусть в каждом из них течет некоторый постоянный ток: I_1 и I_2 . Каждый контур создаст в пространстве магнитное поле, которое можно представить соответствующим вектор-потенциалом. Будем считать контуры линейными в том смысле, что поперечные размеры проводника и распределение в нем тока несущественны в сравнении с характерными масштабами задачи. Тогда $\mathbf{j} dV = j dS_{\perp} d\mathbf{l}$, и объемные интегралы сводятся к интегралам по контуру:

$$\mathbf{A}_1 = \frac{I_1}{4\pi} \oint_1 \frac{d\mathbf{l}_1}{r}; \quad \mathbf{A}_2 = \frac{I_2}{4\pi} \oint_2 \frac{d\mathbf{l}_2}{r}.$$

Вычислим магнитный поток, создаваемый током контура 1 в контуре 2, используя при этом теорему Стокса (по имени английского физика и математика Дж.Г. Стокса (1819–1903)):

$$\Phi_{12} = \int_2 \mathbf{B}_1 d\mathbf{S}_2 = \mu_0 \int_2 \text{rot} \mathbf{A}_1 d\mathbf{S}_2 = \mu_0 \oint_2 \mathbf{A}_1 d\mathbf{l}_2 = \frac{I_1 \mu_0}{4\pi} \oint_1 \oint_2 \frac{d\mathbf{l}_1 d\mathbf{l}_2}{R_{12}}. \quad (6.1)$$

Совершенно аналогично,

$$\Phi_{21} = \frac{I_2 \mu_0}{4\pi} \oint_2 \oint_1 \frac{d\mathbf{l}_2 d\mathbf{l}_1}{R_{21}}. \quad (6.2)$$

Коэффициенты при токе в формулах (6.1) и (6.2), как нетрудно видеть, идентичны:

$$\Phi_{12} = L_{12}I_1; \quad \Phi_{21} = L_{21}I_2; \quad L_{12} = L_{21} = \frac{\mu_0}{4\pi} \oint_1 \oint_2 \frac{d\mathbf{l}_1 d\mathbf{l}_2}{R_{12}}. \quad (6.3)$$

Величины L_{12} , L_{21} называются *взаимной индуктивностью* или *коэффициентами взаимной индукции*, а правило их равенства — *теоремой взаимности*, которая часто используется при решении задач или инженерных расчетах. Легко видеть, что эти коэффициенты зависят только от геометрии контуров, с точностью, впрочем, до знака, который задается направлением обхода в интегралах (6.1)–(6.3). Его, как правило, выбирают так, чтобы все величины L_{ik} были положительны, за исключением тех случаев, когда приходится суммировать магнитные потоки разного знака.

Теперь представим себе единственный замкнутый контур, по которому протекает ток I . Так же, как и в предыдущем примере, представим магнитный поток через контур, обязанный собственному току I в виде

$$\Phi = \frac{I\mu_0}{4\pi} \oint \oint \frac{d\mathbf{l} d\mathbf{l}'}{R} \equiv LI, \quad (6.4)$$

где $R = |\mathbf{R}_{d\mathbf{l}, d\mathbf{l}'}|$. Величина L называется *индуктивностью контура* или *коэффициентом самоиндукции*. Единицу измерения индуктивности или взаимной индукции принято называть генри; ее размерность: $1 \text{ Гн} = 1 \text{ В} \cdot \text{с} / \text{А}$. Причина именно такого выбора размерности проявится ниже (см. закон Фарадея). Другое определение той же единицы: $1 \text{ Гн} = 1 \text{ Вб} / \text{А}$. Как и фарада для измерения емкости, единица эта очень велика, поэтому фактически используются микрогенри или даже наногенри; последняя совпадает с очень хорошей точностью с единицей измерения индуктивности в системе СГС, а потому часто называется сантиметром.

Определение самоиндукции (6.4) удобно постольку, поскольку следует общему алгоритму (6.3), демонстрирует явно зависимость только от геометрии контура, но никогда не используется, в отличие от (6.3), для фактического расчета индуктивности. Дело в том, что именно при рассмотрении самоиндукции неадекватно приближение линейности контура. Как следствие, интеграл (6.4) оказывается несобственным.

Это впрочем, легко понять и на более элементарном уровне: вблизи идеально тонкого провода $B \propto H \propto 1/r$, а значит, поток через контур Φ расходится логарифмически. Но формула (6.4) удобна для доказательств общего вида, в чем мы еще убедимся в следующей главе; теперь же дадим два простых примера прямого вычисления индуктивности.

1. Индуктивность коаксиальной линии. На рис. 6.1 представлена система двух соосных проводников. Ток течет по внутреннему цилиндрическому проводнику радиуса r_1 и возвращается по внешнему цилиндрическому проводнику радиуса r_2 .

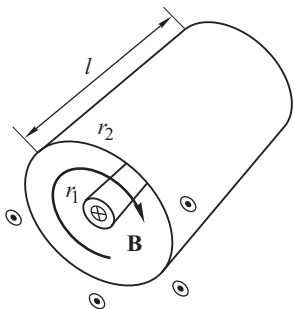


Рис. 6.1

Во избежание уже упомянутой проблемы расходимости полного потока, внутренний проводник, как и внешний, будем считать полым и тонким в сравнении с масштабами r_1 , r_2 . Тогда магнитное поле можно считать сосредоточенным в пространстве между цилиндрами. Внутри внутреннего цилиндра и вне внешнего оно обращается в нуль в силу теоремы о циркуляции и азимутальной симметрии системы.

В силу тех же аргументов, между цилиндрами поле идентично полю прямого провода. Отсюда нетрудно вычислить и поток, взяв элемент сечения контура $dS = l dr$:

$$\Phi = l \int_{r_1}^{r_2} \frac{\mu_0 I}{2\pi r} dr = \frac{\mu_0 I l}{2\pi} \ln \frac{r_2}{r_1} \Rightarrow \frac{dL}{dl} = \frac{\mu_0}{2\pi} \ln \frac{r_2}{r_1}. \quad (6.5)$$

Величина dL/dl есть по определению индуктивность на единицу длины коаксиальной линии.

2. Индуктивность соленоида. Пусть дан длинный соленоид, без сердечника или с магнитоягким сердечником, с плотной намоткой (см., например, рис. 4.2); радиус соленоида R , длина l , полное число витков N , число витков на единицу длины $n = N/l$, магнитная проницаемость сердечника μ . Согласно выражению (4.13), внутри соленоида $B = \mu \mu_0 n I$. Определяя поток через контур, следует принять во внимание, что каждая силовая линия охватывается обмоткой соленоида N раз: $\Phi = \pi R^2 N \mu \mu_0 n I$. Отсюда следует и результат, который может быть представлен в двух эквивалентных формах:

$$\frac{dL}{dl} = \pi \mu_0 \mu (nR)^2; \quad L = \pi \mu_0 \mu \frac{NR^2}{l}. \quad (6.6)$$

6.2. Принцип Ленца. Закон Фарадея

Из школьного курса химии известен принцип ле Шателье–Брауна, сформулированный французским ученым ле Шателье (1850–1936) и немецким физиком К.Ф. Брауном (1850–1918). Этот принцип в простейшей форме выглядит следующим образом: если некая смесь веществ пребывает в динамическом

равновесии за счет встречных экзо- и эндотермической реакций, то при нагревании быстрее начинает идти эндотермическая реакция, а при охлаждении — экзотермическая. В электродинамике близкий по идее принцип был сформулирован знаменитым российским ученым-естественником Э. Х. Ленцем.

Токи, возникающие в проводнике при движении его в магнитном поле, направлены так, что силы, действующие на эти токи, противодействуют движению проводника. С учетом иных возможностей изменения магнитного потока может быть предложена более общая формулировка.

Отклик динамической системы на перестройку магнитного поля всегда имеет тенденцию противодействовать этой перестройке.

Если уронить медную или серебряную монетку между полюсами мощного электромагнита, то невооруженным глазом можно наблюдать замедление — даже весьма значительное — скорости по сравнению со свободным падением. Это «работают» магнитное давление и натяжение силовых линий (см. гл. 7). Основная причина того, что эти механизмы вступают в игру, может быть связана с принципом Ленца. Равным образом объясняются этим принципом токи Фуко (по имени французского физика-экспериментатора Ж.Б.Л. Фуко (1819–1868)) в железном сердечнике трансформатора (из-за которых его обычно набирают из отдельных пластин) и работа динамомашины. Заметим что в первом из этих двух случаев нет ни движения проводника в поле, ни деформации контура, но лишь осцилляция магнитного поля в сердечнике как следствие непостоянства тока. Это и соответствует второй — более общей — формулировке принципа Ленца. Приходится, однако, встречаться с очень грубой его трактовкой, типа «реакция противоположна воздействию». Если бы это было правильно, то в природе, например, не существовало бы ни пара-, ни ферромагнетиков, но лишь диамагнетики. Потому и существенна оговорка относительно реакции динамической системы. Ларморова прецессия и вызванный ею диамагнетизм, и в самом деле, вполне согласуются с правилом Ленца, но, к примеру, ферромагнитные явления уже не могут быть сведены к такой схеме.

Круг явлений, подпадающий под принцип Ленца, принято называть *электромагнитной индукцией*. Основной количественный закон здесь был установлен английским физиком Майклом Фарадеем (1791–1867). Пусть имеется замкнутый контур любой природы (проводящий или непроводящий — даже мысленно очерченный), и пусть по любой причине — движение, деформация, затухание тока в контуре, изменение внешнего магнитного поля — изменяется во времени магнитный поток через означен-

ный контур. Тогда в контуре возникает ЭДС \mathcal{E} , которая может быть найдена из соотношения:

$$\mathcal{E} = \oint \mathbf{E} d\mathbf{l} = -\frac{d\Phi}{dt}. \quad (6.7)$$

Знак в (6.7) отражает то обстоятельство, что возбуждаемая в контуре ЭДС индукции по направлению такова, что противодействует изменению магнитного потока. В частности, при идеальной проводимости контура поток, который им охватывается, должен быть инвариантом. Это относится как к тому случаю, когда охватывающий контур является сверхпроводящим, так и к тому, когда пространственные и временные масштабы задачи позволяют считать его идеальным проводником. (Эти две ситуации не следует путать; мы еще вернемся к этому вопросу в гл. 9.)

Из (6.7) следует введенная в гл. 4 размерность магнитного потока $1 \text{ Вб} = 1 \text{ А} \cdot \text{с} \cdot \text{Ом}$. По этому поводу следует, впрочем, заметить, что расшифровка единиц в системе СИ в принципе может быть любой, которая не противоречила бы их определению, но обычно предпочитают пользоваться размерностью, следующей из наиболее употребительных законов и по возможности не опирающейся на размерность «вакуумных» констант μ_0 и ϵ_0 .

В некоторых простых случаях соотношение (6.7) может быть получено как следствие уже известных законов. Поставим следующий мысленный эксперимент (рис. 6.2): имеется проводящий контур, составленный из двух направляющих, замкнутых с одной стороны

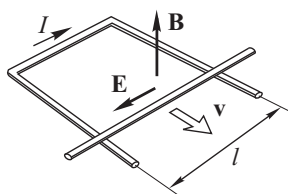


Рис. 6.2

неподвижной перемычкой, а с другой — подвижной, которая движется вдоль направляющих со скоростью \mathbf{v} . Расстояние между направляющими равно l , контур помещен в магнитное поле \mathbf{B} , ортогональное его плоскости. Рассмотрим электрон в подвижной перемычке, который движется вместе с ней, а значи-

т, на него действует сила Лоренца $\mathbf{F} = -e[\mathbf{v} \mathbf{B}]$. Как уже не раз отмечалось, это означает, что в системе отсчета, связанной с перемычкой, возникает электрическое поле $\mathbf{E} = [\mathbf{v} \mathbf{B}]$. Отсюда циркуляция $\oint \mathbf{E} d\mathbf{l} = vBl$. И поскольку $vl = dS/dt$, отсюда как раз следует формула (6.7). В общем случае, однако, ее не удастся вывести, так что этот закон должен рассматриваться как опытный факт.

Полезно лишний раз отметить (см. рис. 6.2), что ток, порождаемый индукционным электрическим полем, в свою очередь, порождает магнитное поле, по направлению противоположное внешнему полю \mathbf{B} , и тем самым как бы «стареется»

скомпенсировать увеличение магнитного потока через контур — в согласии с принципом Ленца. При обратном направлении движения рамки генерируется поле того же направления, что и приложенное извне.

6.3. Интегральная и локальная форма закона электромагнитной индукции

Рассмотренные до настоящего момента законы электродинамики явно распадаются на две группы. В одну из них входят так называемые материальные соотношения: закон Ома, зависимость $\mathbf{P}(\mathbf{E})$ или $\mathbf{J}(\mathbf{B})$ и т. д. Эти законы задаются физикой строения вещества; результаты, на них основанные, всегда не универсальны, причем это общее свойство их всех, а не только простых линейных моделей.

Вторая группа — чисто электродинамические законы, например теорема Гаусса или теорема о циркуляции (которую, правда, нам еще предстоит записать в окончательном виде). Эти законы, разумеется, также имеют свою область применимости, но она гораздо шире, чем у материальных соотношений, по сути, оставаясь в рамках классической физики, мы за нее заведомо не выходим. Более того, эти фундаментальные законы более совершенны, нежели закон Кулона или закон Био и Савара, их породившие, поскольку не связаны приближениями точечных зарядов или бесконечно тонких элементов тока. Справедливость их подтверждается как прямыми экспериментами, так и (даже в большей степени) согласованностью и непротиворечивостью всего массива информации, составляющего основу современной электродинамики.

Опыт показывает, что закон электромагнитной индукции (6.7) также относится к этой группе фундаментальных законов. Его, конечно, имеет смысл модифицировать применительно к системе с сосредоточенными параметрами — токовому контуру —

$$\mathcal{E} = \oint \mathbf{E} d\mathbf{l} = -L \frac{dI}{dt}, \quad (6.8)$$

если его параметры не меняются во времени, если же контур является перестраиваемым или, скажем, деформируемым, то соотношение (6.8) допускает естественное обобщение:

$$\mathcal{E} = \oint \mathbf{E} d\mathbf{l} = -\frac{d\Phi}{dt} = -\left(L \frac{dI}{dt} + I \frac{dL}{dt}\right). \quad (6.9)$$

В то же время, как и для всех законов такого рода, желательно получить локальную форму закона Фарадея, которая была бы свободна от понятия какого-либо материального контура.

Для этого воспользуемся уже упоминавшейся в гл. 4 теоремой Стокса:

$$\oint \mathbf{E} d\mathbf{l} = \int \text{rot} \mathbf{E} d\mathbf{S}$$

— циркуляция вектора по некоторому контуру равна потоку ротора этого вектора через любую гладкую поверхность, натянутую на контур. Объединяем теорему Стокса и закон Фарадея:

$$\int \text{rot} \mathbf{E} d\mathbf{S} = -\frac{d\Phi}{dt} = -\frac{d}{dt} \int \mathbf{B} d\mathbf{S}.$$

Ввиду произвольности контура и натянутой на него поверхности, можно просто приравнять подынтегральные выражения. Только временную производную надо будет представить в виде частной производной по времени, чтобы избежать в локальной форме уравнения эффектов, связанных, например, с движением вещества. Итак, вот *локальная форма закона Фарадея*:

$$\text{rot} \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}. \quad (6.10)$$

Знак « ∂ », который мы до сих пор старательно удерживали как символ направленности ЭДС индукции, приобрел в этой векторной формуле окончательный смысл. В зависимости от конкретной задачи, может быть использована любая форма закона электромагнитной индукции (6.7)–(6.10), но именно (6.10) представляет закон Фарадея в наиболее общем виде.

Задачи

1. На один сердечник намотаны две катушки (рис. 6.3). Индуктивности их равны соответственно $L_1 = 0,5$ Гн и $L_2 = 0,7$ Гн. Чему равна их взаимная индуктивность в отсутствие рассеяния магнитного потока?

Решение. Поскольку магнитный поток не рассеивается (это означает, помимо надлежащей геометрии сердечника, $\mu \gg 1$), то через каждый виток он один и тот же: $\Phi = \text{const}$. Пусть, для определенности, ток течет через первую обмотку:

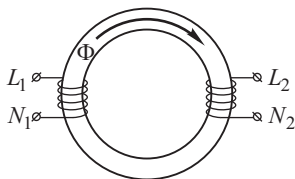


Рис. 6.3

$$\Phi = \frac{\Phi_1}{N_1} = \frac{\Phi_{12}}{N_2} \implies \frac{L_1 I}{N_1} = \frac{L_{12} I}{N_2}.$$

Аналогичным образом $L_{21}/N_1 = L_2/N_2$. По теореме взаимности $L_{21} = L_{12}$. Исключая число витков N_1, N_2 , получаем

$$L_{12} = (L_1 L_2)^{1/2} = 0,59 \text{ Гн}.$$

Простая формула $(L_1 L_2)^{1/2}$ представляет собой максимально возможный коэффициент взаимной индукции при заданной индуктивности двух катушек. Реально он всегда меньше, потому что невозможно полностью исключить рассеяние магнитного потока.

2. Внутри длинного соленоид с плотностью намотки n , вдали от его концов, расположен параллельно оси намагниченный стержень с магнитным моментом m_μ . Найти магнитный поток Φ , пронизывающий соленоид.

Решение. Прямое вычисление было бы громоздким, и в нем легко было бы ошибиться. Но мы можем сделать мысленно совершенно эквивалентную замену. Пусть вместо постоянного магнита в соленоид помещен виток с током, так что $IS = m_\mu$. Далее мы можем воспользоваться теоремой взаимности, пропустив ток I не через виток, а через обмотку соленоида. Тогда $B = \mu_0 n I = \mu_0 m_\mu n / S$, и магнитный поток через виток, как раз и равный искомому, составляет $\Phi = \mu_0 m_\mu n$. Представляется достаточно нетривиальным факт универсальности ответа; без теоремы взаимности установить его было бы затруднительно.

3. В ускорителе электронов — бетатроне — роль ускоряющего напряжения играет ЭДС индукции, возбуждаемая изменением магнитного потока, пронизывающего орбиту электрона. Электроны же при этом движутся по орбитам примерно постоянного радиуса. Определить необходимое для этого соотношение между средним по площади орбиты магнитным полем $\langle B \rangle(t)$, пронизывающим орбиту электрона, и магнитным полем $B_0(t)$ непосредственно на орбите. Поле считать нормальным к плоскости орбиты.

Решение. Если частица движется по круговой орбите, то, согласно (4.8), частота обращения есть $\omega_B = eB_0/m$. При импульсе частицы p радиус орбиты равен $R = v/\omega_B = p/eB_0$. Условие $R = \text{const}$ эквивалентно следующему: $p \propto B_0$ или, что то же, $\dot{p}/p = \dot{B}_0/B_0$. Ускорение электрона на орбите постоянного радиуса происходит в соответствии с законом $\dot{p} = eE$, где электрическое поле определяется из закона Фарадея (6.7):

$$\oint \mathbf{E} d\mathbf{l} = 2\pi R E = -\frac{d\Phi}{dt} = \pi R^2 \langle \dot{B} \rangle.$$

Таким образом, зависимость $E = R \langle \dot{B} \rangle / 2$ подставляется в соотношение

$$\frac{\dot{B}_0}{B_0} = \frac{eE}{eRB_0}.$$

Отсюда следует $\dot{B}_0(t) = (1/2) \langle \dot{B} \rangle(t)$ в любой момент времени, что эквивалентно $B_0(t) = \langle B \rangle(t)/2$.

4. Один и тот же ток идет по двум параллельным проводам в противоположные стороны. Провода имеют круглые сечения радиуса $r = 2$ мм, а расстояние между ними $d = 2$ см. Найти индуктивность единицы длины такой системы, учитывая поле только вне проводов.

Ответ: $L = \frac{\mu_0}{\pi} \ln \frac{d-r}{r} \approx 9 \cdot 10^{-7} \text{ Гн/м}.$

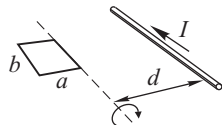


Рис. 6.4

5. Прямоугольная рамка $a \times b$ лежит в одной плоскости с прямым проводником, по которому течет ток I и который расположен параллельно стороне b , на расстоянии $d > a$ от ближайшей стороны (рис. 6.4). Какой заряд Q пройдет через любое сечение провода рамки, если она повернется вокруг ближайшей к проводу стороны b на угол π и останется в этом положении? Сечение провода рамки S , удельное сопротивление ρ . Самоиндукцией рамки пренебречь.

Ответ: $Q = \frac{b}{a+b} \frac{IS}{4\pi\rho} \ln \frac{d+a}{d-a}.$

Глава 7

ЭНЕРГИЯ И СИЛЫ В МАГНИТНОМ ПОЛЕ

7.1. Проводники в магнитном поле. Магнитное давление

С тем, что такое магнитное давление, мы уже познакомились в задаче 2 гл. 4. Напомним, что поле между проводящими плоскостями было равно $H = i$, где i — линейная плотность поверхностного тока, а давление на каждую плоскость составляло

$$\mathcal{P} = i B_{\text{вн}} = \frac{\mu_0 i^2}{2}.$$

Нетрудно сообразить, что такое же точно давление будет действовать на любую плоскую поверхность проводящего тела, если вне тела существует магнитное поле \mathbf{B} , параллельное поверхности, а в самом теле $\mathbf{B} = 0$. Действительно, величина i однозначно следует из теоремы о циркуляции, и все последующие соображения остаются в силе. Нужно только переписать формулу для давления на поверхность проводника, исходя из того, что задано все-таки именно поле в вакууме:

$$\mathcal{P} = \frac{\mu_0}{2} H^2 = \frac{B^2}{2\mu_0} = \frac{1}{2} (\mathbf{B}\mathbf{H}). \quad (7.1)$$

Существенно, во-первых, то обстоятельство, что магнитное давление обусловлено «непроникновением» магнитного поля в проводник. Если последний пребывает в сверхпроводящем состоянии, это всегда верно, но и в обратном, гораздо более обыденном случае, это совершенно реально. Как мы увидим далее (см. гл. 8), проникновение поля в проводящую среду — процесс не мгновенный, и чем выше проводимость, тем более он затянут во времени.

Во-вторых, при выводе формулы (7.1) мы использовали упрощенную геометрию: внешнее поле, параллельное плоской поверхности проводника. Это сужает круг явлений, которые мы могли бы обсуждать, либо создает опасность ошибок, проистекающих от расширительного толкования результата (7.1).

Прежде чем давать формальное описание сил, действующих в магнитном поле на проводящую среду, приведем чрезвычайно удобную форму для архимедовой силы, использующую операцию векторного дифференцирования. Рассмотрим некоторую текучую

среду, т. е. жидкость или газ, подчиняющуюся закону Паскаля, и выделим в ней малый объем — кубик со сторонами dx , dy , dz (рис. 7.1). Пусть мы можем пренебречь всеми объемными силами (силой тяжести и т. д.); принимаем во внимание только силу, связанную с неоднородностью давления, — а это и есть архимедова сила. Как было показано в гл. 9 курса механики, эффекты, обусловленные неоднородностью давления, удобно выводить из объемной плотности силы $\mathbf{f} = d\mathbf{F}/dV$:

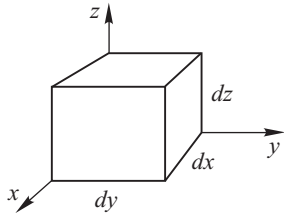


Рис. 7.1

$$f_x = -\frac{\partial \mathcal{P}}{\partial x} \text{ и т. д. } \implies \mathbf{f} = -\nabla \mathcal{P}. \quad (7.2)$$

С силой в форме (7.2) удобно работать; скажем, закон Архимеда можно мгновенно получить, уравновесив ею плотность силы тяжести $\rho \mathbf{g}$. Нам же она понадобится для сравнения именно с плотностью силы Ампера, которую мы предварительно перепишем следующим образом:

$$\mathbf{f} = [\mathbf{j} \mathbf{B}] = -[\mathbf{B} [\nabla \mathbf{H}]] = (\mathbf{B} \nabla) \mathbf{H} - \langle \nabla (\mathbf{B} \mathbf{H}) \rangle.$$

Проводя формальные векторные операции с оператором дифференцирования ∇ , следует заботиться о том, чтобы справа от него оставались лишь те величины, которые подлежат дифференцированию. Этому правилу не удовлетворяет последний член в правой части; чтобы подчеркнуть это, он заключен в кавычки. В стоящем справа от оператора скалярном произведении вектор \mathbf{B} дифференцироваться не должен. Если, однако, связь между \mathbf{B} и \mathbf{H} линейна — а это во всяком случае верно для всех диа- и парамагнитных проводников,, то мы можем представить формулу для плотности силы Ампера в виде

$$\mathbf{f} = -\nabla \frac{(\mathbf{B} \mathbf{H})}{2} + (\mathbf{B} \nabla) \mathbf{H}. \quad (7.3)$$

Второй член в правой части (7.3) называется «максвелловским натяжением силовых линий» (формально он и в самом деле имеет вид упругой силы), что же касается первого, то он оказался абсолютно идентичен архимедовой силе (7.2), если в нее подставить давление в виде (7.1). Таким образом, формула (7.1) оказалась точной, но теперь мы должны сделать два важных замечания относительно границ ее применимости.

1. «Натяжение силовых линий» обусловлено их кривизной. Действительно, если силовые линии — прямые, то поле может меняться лишь поперек силовых линий, но не вдоль них, поэтому $(\mathbf{B} \nabla) \mathbf{H} \equiv 0$. Следовательно, мы можем свести силу Ампера к магнитному давлению, только когда имеем право пренебречь

кривизной силовых линий. Это не значит, что они должны быть непременно прямыми — просто радиус их кривизны должен быть много больше толщины слоя, в котором магнитная индукция **B** меняется на величину порядка ее самой. Вот пример ситуации, когда натяжением пренебречь нельзя: магнитное поле в проводнике установилось таким образом, что токи в нем не текут. Допустим, вне проводника проходит прямой провод с током, и поле в проводящей среде соответствует формуле (4.9). Из (7.1), казалось бы, следует, что в проводнике возникает некоторый перепад давлений, а значит, и силы. Но из формулы (4.6) следует, что при $\mathbf{j} = 0$ должно быть и $\mathbf{f} = 0$. Действительно, можно показать, что в этом случае два члена в правой части (7.3) в точности компенсируют друг друга.

2. Возможная причина недоразумений — независимость формул (7.1) и (7.3) от проводимости среды. На первый взгляд может показаться, что и на диэлектрик должно действовать точно такое же давление. В этой связи полезно напомнить, что при выводе (7.3) мы существенным образом воспользовались двумя предположениями: теоремой о циркуляции: $\text{rot } \mathbf{H} = \mathbf{j}$ и линейной зависимостью **B(H)**. Как уже было замечено в гл. 4, теорема о циркуляции у нас пока не дописана; в нестационарном случае в правой части появится еще один член (и добавим, забегаая вперед, — роль его будет тем важнее, чем ниже проводимость вещества). Таким образом, для настоящего диэлектрика наш вывод справедлив только при условии абсолютной стационарности задачи, и, поскольку тока проводимости в непроводящем веществе просто не может быть, два члена в правой части (7.3) обязаны при этом точно друг друга компенсировать. Тем самым определяется класс допустимых магнитных конфигураций в линейной диэлектрике.

Положение дел меняется, если мы имеем дело с магнитоактивной средой (линейная зависимость **B(H)** представляет не более, чем частный случай). Тогда к силам Ампера, обязанным своей природой токам проводимости, добавятся силы, обусловленные намагниченностью образца. Это и будет темой последующего рассмотрения.

7.2. Диполь в магнитном поле

Хотя векторная природа электрического и магнитного полей различна и магнитный дипольный момент (4.21) определяется иначе, нежели электродипольный момент (1.23), обращает на себя внимание тождественность выражений для поля (4.24) и (1.25). Из рис. 4.4 можно усмотреть совершенную идентичность картины поля при переходе к приближению точечных диполей.

(Только условия «точечности» будут различны — в одном случае необходимо с надлежащей точностью обеспечить квазинейтральность $\operatorname{div} \mathbf{E} = 0$, а в другом — неразрывность токов $\operatorname{div} \mathbf{j} = 0$.)

Мы воспользуемся этим обстоятельством, чтобы избежать весьма трудоемкого вывода из определения (4.21) формул, аналогичных (1.27) и (1.28), а вместо этого сразу же воспроизведем их с соответствующими переобозначениями:

$$\mathbf{F} = \nabla(\mathbf{m} \mathbf{B}), \quad (7.4)$$

$$\mathbf{M} = [\mathbf{m} \mathbf{B}]. \quad (7.5)$$

Как известно из механики, консервативные силы связаны с потенциальной энергией соотношением $F_x = -dW/dx$ в одномерном случае, а в трехмерном мы должны повторить то же самое для F_y , F_z , так что $\mathbf{F} = -\nabla W$ — мы уже использовали это обстоятельство в электростатике. Таким образом, (7.4) может быть представлена в эквивалентном виде: энергия магнитного диполя во внешнем поле равна

$$W = -(\mathbf{m} \mathbf{B}). \quad (7.6)$$

Это соотношение является базовым для определения единицы дипольного момента — он измеряется в *джоулях на тесла* (Дж/Тл). Мы уже использовали эту единицу в гл. 5 для магнетона Бора.

Если бы речь шла только о заданном диполе — например постоянном магните малого размера, формулами (7.4) и (7.6) можно было бы и ограничиться, но в случае, когда мы имеем дело с наведенным дипольным моментом, в них неявно заключена опасность ошибки. Именно, если $\mathbf{m} \propto \mathbf{B}$ или вообще дипольный момент функционально как-то связан с полем, возникает вопрос, следует ли и его дифференцировать в (7.4). Правильно будет этого не делать, потому что взаимодействие диполя с полем — характеристика силовая, и она не может зависеть от природы диполя: на диполи с одинаковыми дипольными моментами должна действовать одинаковая сила безотносительно к их происхождению. В таких случаях принято указывать в формуле, на какой именно сомножитель действует оператор ∇ :

$$\mathbf{F} = \nabla(\mathbf{m} \check{\mathbf{B}}),$$

но лучше привести (7.4) к такому виду, в котором \mathbf{m} заведомо не дифференцируется. Воспользуемся векторным тождеством

$$[\mathbf{m}, \operatorname{rot} \mathbf{B}] = [\mathbf{m} [\nabla \mathbf{B}]] = \nabla(\mathbf{m} \check{\mathbf{B}}) - (\mathbf{m} \nabla) \mathbf{B},$$

которое позволяет записать силу (7.4) в следующем виде:

$$\mathbf{F} = \nabla(\mathbf{m} \check{\mathbf{B}}) = (\mathbf{m} \nabla) \mathbf{B} + [\mathbf{m}, \operatorname{rot} \mathbf{B}]. \quad (7.7)$$

В магнитоактивном веществе соотношение (7.7) естественно представить в форме объемной плотности силы

$$\mathbf{f} = (\mathbf{J} \nabla) \mathbf{B} + [\mathbf{J}, \text{rot} \mathbf{B}]. \quad (7.8)$$

В соответствии с вышесказанным, величина $\mathbf{J}(\mathbf{r})$ здесь не дифференцируется.

Дальнейшее преобразование формулы (7.8) требует явного учета зависимости $\mathbf{J}(\mathbf{B})$. Мы ограничимся простым случаем линейной связи, т. е. диа- или парамагнетиками:

$$\mathbf{J} = \mu_0^{-1} \mathbf{B} - \mathbf{H} = \frac{\mu - 1}{\mu \mu_0} \mathbf{B}.$$

Подставим это выражение в (7.8) и проведем векторное преобразование, аналогичное тому, что мы сделали при выводе (7.3):

$$(\mathbf{B} \nabla) \mathbf{B} + [\mathbf{B}, \text{rot} \mathbf{B}] = \frac{1}{2} \nabla B^2.$$

В итоге получаем

$$\mathbf{f} = \frac{\mu - 1}{2\mu\mu_0} \nabla B^2 = \frac{\mu - 1}{2} \nabla (\mathbf{B} \mathbf{H}). \quad (7.9)$$

Мы снова выразили силу, распределенную в среде, в форме градиента давления (7.2). Но этот результат еще может быть дополнен плотностью силы Ампера (7.3), если магнетик одновременно является проводником. Особенно интересно отметить, что эффективное «давление»

$$\langle \mathcal{P} \rangle = -\frac{\mu - 1}{2} (\mathbf{B} \mathbf{H}) \quad (7.10)$$

может быть как положительным при $\mu < 1$, так и отрицательным при $\mu > 1$. Как следствие, диамагнетик выталкивается из области более сильного магнитного поля, а парамагнетик, напротив, втягивается в нее. Этот эффект может быть с успехом продемонстрирован на примере растворов пара- или диамагнитных солей. По существу, мы получили объяснение и того общеизвестного факта, что предметы из магнитомягкого железа притягиваются постоянным магнитом. Конечно, конкретные формулы (7.9), (7.10) могут оказаться неадекватными из-за нелинейности зависимости $\mathbf{J}(\mathbf{B})$, но качественно это тот же самый эффект. (Что касается магнитотвердых материалов, то здесь может оказаться существенной их собственная намагниченность.)

7.3. Энергия магнитного поля

Понятие энергии магнитного поля можно ввести, следуя программе, изложенной в гл. 2: оперируя экспериментальной информацией о свойствах магнитного поля, предложить такой мыслен-

ный эксперимент, в котором появился бы интеграл движения, включающий аддитивно механическую энергию и в отсутствие магнитного поля к ней сводящийся. Тогда добавка, обусловленная магнитным полем, и должна трактоваться как магнитная энергия.

Невозможно, однако, предложить один мысленный эксперимент на все случаи жизни. Поэтому новое понятие неизбежно проверяется снова и снова по мере накопления экспериментального материала, иногда дополняется или модифицируется. В конечном счете все принципиальные моменты в физике оказываются — пусть даже косвенно — экспериментального происхождения.

В данном случае особенно деликатным оказывается вопрос о магнитном поле в веществе, поскольку не всегда просто разделить энергию поля, энергию вещества и энергию взаимодействия поля с веществом.

Хотя мы уже обсуждали подобного рода вопросы в гл. 2, это краткое введение представляется все же не лишним. А теперь поставим первый мысленный эксперимент. Обратимся к рис. 7.2 а.

Представим себе тонкий бесконечно длинный идеально проводящий цилиндр (его обычно называют лайнером), по которому течет однородный азимутальный ток с постоянной линейной плотностью. Внутри цилиндр пуст; никаких источников энергии либо каналов диссипации в задаче нет. Представим далее, что мы сжимаем лайнер, прикладывая извне давление \mathcal{P} . При сжатии от начального радиуса R на dR совершается следующая работа на единицу длины:

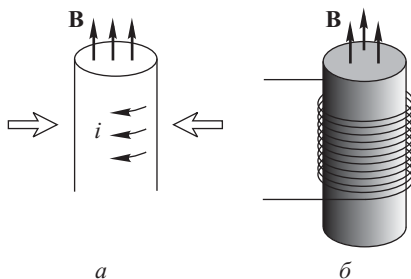


Рис. 7.2

совершается следующая работа на единицу длины:

$$\delta A = \mathcal{P} \cdot 2\pi R dR.$$

Пусть смещение происходит очень медленно, практически без ускорения, и пусть мы можем пренебречь всеми механическими напряжениями в веществе лайнера. Заметим, что ввиду высокой степени симметрии, магнитное поле вне лайнера — в точности равно нулю, а внутри него оно совершенно однородно и силовые линии параллельны оси. В этой ситуации совершенно неважно, как именно распределен ток по глубине оболочки — все равно имеет место механическое равновесие: $\mathcal{P} = (\mathbf{B} \mathbf{H})/2$.

Последующие вычисления будут базироваться на законе Фарадея в его самой простой формулировке: магнитный поток через любой идеально проводящий контур — инвариант. В нашей

постановке задачи это значит, что инвариантным является произведение поля на площадь поперечного сечения лайнера, т. е. $B, H \propto R^{-2}$. Тем самым $\mathcal{P} \propto R^{-4}$, и мы можем через механическую работу определить энергию поля внутри лайнера. Удобно сопоставить работу против сил магнитного давления при сжатии лайнера от бесконечного до текущего радиуса с энергией поля на единицу длины лайнера:

$$\begin{aligned} \frac{dW_M}{dl} &= \int_R^\infty \frac{1}{2} B(r) H(r) \cdot 2\pi r dr = \\ &= \pi B(R) H(R) \int_R^\infty \left(\frac{R}{r}\right)^4 dr^2 = \pi R^2 \cdot \frac{1}{2} B(R) H(R). \end{aligned}$$

Мы получили энергию в форме, уже привычной нам из электростатики. Исходя из этого модельного результата, можно предложить следующее выражение для *плотности магнитной энергии*:

$$w = \frac{(\mathbf{B} \mathbf{H})}{2}, \quad (7.11)$$

которое и подтверждается всем опытом современной электродинамики. Мы получили его сразу в наиболее универсальном виде, который можно будет переносить непосредственно на поле в веществе.

В самом деле, мысленный эксперимент станет еще более фантастичным, но не менее корректным, если мы заполним лайнер некоторой идеально сжимаемой магнитоактивной средой, а в результате получится то же выражение (7.11) для плотности энергии магнитного поля, но теперь уже $\mathbf{B} \neq \mu_0 \mathbf{H}$.

В качестве следующего объекта для рассуждений возьмем изображенный на рис. 7.2б бесконечно длинный идеально проводящий соленоид с плотной намоткой (число витков на единицу длины $n = N/l \gg R^{-1}$, где R — радиус соленоида, l — его длина). По обмотке соленоида течет ток I , при этом $H = nI$, внутри соленоид заполнен сердечником из однородного магнетика. Считается, что зависимость $B(H)$ известна; она, в частности, может быть и линейной $B = \mu_0 \mu H$, что несколько упрощает конечные формулы.

Энергия магнитного поля на единицу длины соленоида

$$\frac{dW_M}{dl} = \frac{BH}{2} \cdot \pi R^2 = \frac{1}{2} \Phi_0 nI,$$

где Φ_0 — поток магнитной индукции через отдельный виток; рассматривая соленоид как целое, учтем, что полный магнитный поток через его обмотку есть $N\Phi_0$, а полная энергия катушки

(контура) с током может быть представлена в виде

$$W = \frac{I\Phi}{2}. \quad (7.12)$$

Мы снова отдали предпочтение универсальной форме записи, наименее чувствительной к деталям постановки задачи. Но для очень широкого класса контуров без сердечника или с магнитомягким сердечником формулу (7.12) можно заменить более конкретной, предполагающей линейную зависимость $B(H)$:

$$\Phi = LI; \quad W = \frac{LI^2}{2}, \quad \text{где} \quad L \equiv \pi\mu_0\mu(nR)^2l = \pi\mu_0\mu \frac{(NR)^2}{l}. \quad (7.13)$$

Коэффициент L — не что иное, как индуктивность соленоида. Она зависит, как нам уже известно, только от геометрии контура и магнитной проницаемости вещества сердечника (если таковой имеется).

Проведенное выше вычисление энергии катушки с током и введенное понятие индуктивности требуют одной важной оговорки. Мы уже отметили в связи с выводом формулы (4.13), что поле вне бесконечного соленоида — отнюдь не нулевое, а с хорошей точностью — поле прямого тока $H_{\text{вн}} = I/(2\pi r)$. Сравним энергию поля в соленоиде, например, (7.13), с энергией внешнего поля. Мы немедленно сталкиваемся с парадоксом: хотя $H_{\text{вн}} \ll nI$, что обеспечено неравенством $n \gg R^{-1}$, энергия внешнего поля оказывается формально бесконечной (интеграл $\int H_{\text{вн}}^2 \cdot 2\pi r dr$ расходится логарифмически). В зависимости от геометрии системы, возможны два выхода из этой ситуации.

1. Пусть вся электрическая цепь достаточно симметрична. На рис. 7.3 представлен модельный пример полной аксиальной симметрии. Соленоид радиуса R включен последовательно с источником; цепь замкнута внешним цилиндрическим токопроводом радиуса $R_{\text{вн}}$, соосным с соленоидом. В этом случае справедливость результатов (7.12), (7.13) можно обеспечить следующим сильным неравенством:

$$BH \cdot \pi R^2 \gg \mu_0 \int_R^{R_{\text{вн}}} H_{\text{вн}}^2 \cdot 2\pi r dr.$$

Подставляя $H(I)$, $H_{\text{вн}}(I)$ и опуская, ввиду сильного неравенства, коэффициенты порядка единицы, получим

$$\mu(nR)^2 \gg \ln \frac{R_{\text{вн}}}{R}.$$

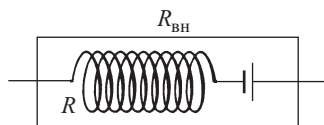


Рис. 7.3

(Читатель может проверить этот результат в качестве упражнения.)

2. Если, напротив, электрическая цепь совершенно не симметрична, то на больших расстояниях поле ее будет зависеть от координат примерно как поле диполя $H_{\text{вн}} \propto r^{-3}$, так что на самом деле интеграл энергии сходится при учете реальной трехмерности задачи. Но характерный пространственный масштаб спада поля в общем случае не определен, так что для уверенности в правильности формулы (7.13) его следует приравнять максимально возможному — длине соленоида l . Отсюда следует оценка

$$\mu(nR)^2 \gg \left(\frac{l}{R}\right)^2,$$

т. е. в обоих случаях условие на плотность намотки оказывается несколько более жестким, чем предполагалось вначале.

Такого рода соотношения приходится постоянно держать в поле зрения при работе с высокоточными системами. Мы хотим в данном случае обратить внимание на ограниченную применимость хрестоматийных формул (7.13), но это никоим образом не ставит под сомнение общий результат (7.11), да и формула (7.12) при правильном вычислении потока Φ через полный контур (а не только соленоид) остается в силе.

Дадим формальный вывод выражения для плотности энергии (7.11), базирующийся на аксиоматическом представлении о потенциальной энергии произвольной токовой конфигурации. Известному представлению энергии электрического поля

$$W_{\text{эл}} = \frac{1}{2} \int_V \rho(\mathbf{r}) \varphi(\mathbf{r}) dV$$

поставим в соответствие в качестве энергии магнитного поля следующее выражение:

$$W_{\text{м}} = \frac{1}{2} \int_V (\mathbf{j}, \mathbf{A}_*) dV = \frac{1}{2} \int_V (\mathbf{A}_*, \text{rot} \mathbf{H}) dV, \quad (7.14)$$

где интегралы берутся по объему всех входящих в систему проводников. Мы распространим интегрирование на все пространство; это не меняет результата, потому что вне проводников $\mathbf{j} \equiv 0$. При этом потенциальная функция \mathbf{A}_* должна порождать, в отличие от вектор-потенциала (4.15), действующее поле в веществе $\mathbf{B} = \langle \mathbf{b} \rangle$, так что, в соответствии с выводом макроскопических уравнений поля в 5.1,

$$\mathbf{A}_* = \frac{\mu_0}{4\pi} \left(\int \frac{\mathbf{j} dV}{R} + \int \frac{\langle \mathbf{j}_{\text{микро}} \rangle dV}{R} \right). \quad (7.15)$$

Для дальнейших выкладок нам понадобится следующая цепочка преобразований:

$$\operatorname{div}[\mathbf{H}, \mathbf{A}_*] = \nabla[\check{\mathbf{H}}, \mathbf{A}_*] + \nabla[\mathbf{H}, \check{\mathbf{A}}_*] = \mathbf{A}_* \cdot \operatorname{rot} \mathbf{H} - \mathbf{H} \cdot \operatorname{rot} \mathbf{A}_*,$$

с учетом которых (7.14) можно переписать в виде

$$W_{\text{м}} = \frac{1}{2} \int (\mathbf{A}_*, \operatorname{rot} \mathbf{H}) dV = \frac{1}{2} \int \operatorname{div}[\mathbf{H}, \mathbf{A}_*] dV + \frac{1}{2} \int \mathbf{H} \cdot \operatorname{rot} \mathbf{A}_* dV.$$

Но теперь мы можем первый из интегралов в правой части преобразовать с помощью теоремы Гаусса в поток вектора $[\mathbf{H}, \mathbf{A}_*]$ через бесконечно удаленную поверхность. Поскольку магнитное поле должно убывать как дипольное или быстрее, на бесконечности $H \propto r^{-3}$, $A_* \propto r^{-2}$, $S \propto r^2$, так что этот интеграл равен нулю. Таким образом,

$$W_{\text{м}} = \frac{1}{2} \int \mathbf{H} \cdot \operatorname{rot} \mathbf{A}_* dV = \int \frac{(\mathbf{H} \mathbf{B})}{2} dV, \quad (7.16)$$

что эквивалентно выражению (7.11). Тем самым мы фактически продемонстрировали, что энергия магнитного поля может быть, помимо прочего, представлена в виде (7.14), (7.15). Убедившись таким образом в правильности (7.14), попробуем дать вывод соотношений (7.12), (7.13), базирующийся уже не на отдельном частном примере, а на общем законе. Для простоты ограничимся случаем контура без сердечника; обобщения принципиальной трудности не составят, но потребуют более громоздких вычислений. В «вакуумном» случае $\mathbf{A}_* = \mu_0 \mathbf{A}$. Как и при выводе базовых выражений для индуктивных коэффициентов (6.3), (6.4), будем считать контур тонким. Тогда в интегралах (7.14), (7.15) $\mathbf{j} dV \Rightarrow I d\mathbf{l}$, что позволяет переписать их в виде

$$W = \frac{\mu_0}{2} \int (\mathbf{j} \mathbf{A}) dV = \frac{\mu_0}{8\pi} \int \int \frac{(\mathbf{j} \mathbf{j}')}{R} dV dV' = \frac{\mu_0 I I'}{8\pi} \oint \oint \frac{d\mathbf{l} d\mathbf{l}'}{R}. \quad (7.17)$$

Далее ответ зависит от того, что мы понимаем под штрихованными величинами в (7.17). Если имеется в виду всего один контур, то мы получаем просто формулу (7.13), где индуктивность задается выражением (6.4), а если контуров два, то нам придется составить все комбинации, проистекающие от суммирования во всем пространстве \mathbf{j} и \mathbf{j}' , \mathbf{A} и \mathbf{A}' :

$$W = \frac{1}{2} (L_1 I_1^2 + L_2 I_2^2 + 2L_{12} I_1 I_2) \quad (7.18)$$

(мы учли, что $L_{12} = L_{21}$). В случае многих контуров, вводя обозначение $L_{ii} \equiv L_i$, можем привести (7.18) к универсальному виду:

$$W = \frac{1}{2} \sum_{i,k} L_{ik} I_i I_k. \quad (7.19)$$

7.4. Подъемная сила электромагнита

Темой данного параграфа будет элементарный расчет подъемной силы электромагнита с сердечником из мягкого железа. Для определенности рассмотрим электромагнит типа простого ярма (рис. 7.4 а). Сечение ярма и якоря S будем считать везде

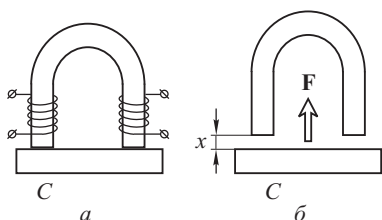


Рис. 7.4

одинаковым и настолько малым, чтобы имело смысл пользоваться приближением однородной магнитной цепи. Пусть магнитная проницаемость сердечника и якоря равна $\mu \gg 1$, длина по обходу (характерная длина силовой линии) l , число витков обмотки N , ток через обмотку I (детали геометрии сердечника и обмотки

при $\mu \gg 1$ и $l \gg S^{1/2}$ не имеют значения. Подчеркнем, что наша цель — не инженерный расчет, а установление на простейшем примере общих принципов рассмотрения таких систем.

В данном случае мы не можем пользоваться на только результатом (7.10), но и самой идеей его получения. Причина в том, что электромагнит — система динамическая, и корректное ее рассмотрение должно включать источник тока. Припишем последнему ЭДС \mathcal{E} , а обмоткам — сопротивление R .

Для рассмотрения таких систем весьма продуктивен принцип виртуальных перемещений, который мы уже использовали в решении задачи 2 гл. 2. Оторвем якорь от ярма, отодвинув его на расстояние dx (рис. 7.4 б) и подсчитаем, какая при этом должна быть совершена механическая работа. Полный энергетический баланс выглядит следующим образом:

$$-F dx = \delta A_{\mathcal{E}} - \delta A_I - dW_L,$$

где $\delta A_{\mathcal{E}}$ — работа источника ЭДС; δA_I — приращение джоулева тепла; dW_L — приращение магнитной энергии контура.

Величины $\delta A_{\mathcal{E}}$ и δA_I должны вычисляться за время перемещения на фоне постоянной мощности, потребляемой электромагнитом. Введем в рассмотрение также приращения тока в контуре dI и магнитного потока через сердечник $d\Phi$ в момент перемещения. Эти два приращения должны быть связаны законом Ома и законом Фарадея:

$$dI = -\frac{1}{R} \frac{d\Phi}{dt}.$$

Приращение работы источника равно

$$\delta A_{\mathcal{E}} = \mathcal{E}(I + dI) dt - \mathcal{E}I dt = \mathcal{E} dI dt = -\mathcal{E} \frac{dt}{R} \frac{d\Phi}{dt} = -I d\Phi.$$

Приращение джоулевых потерь относительно таковых в стационарном режиме

$$\delta A_I = R(I + dI)^2 dt - RI^2 dt \approx 2RI dI dt = -2I d\Phi.$$

И, наконец, приращение энергии магнитного поля, в отличие от диссипативных величин $\delta A_{\mathcal{E}}$ и δA_I , определяется только начальным и конечным состояниями:

$$dW_L = d\left(\frac{LI^2}{2}\right) = \frac{1}{2} I^2 dL = \frac{1}{2} I d\Phi.$$

Последняя цепочка равенств заслуживает особого внимания. Во-первых, при дифференцировании энергии поля мы положили $dI = 0$, поскольку она при изменении конфигурации зависит лишь от положения якоря при установившемся токе $I_0 = \mathcal{E}/R$. (Напротив, приращение тепловыделения и дополнительная работа источника обусловлены как раз приращением тока в контуре в процессе перемещения.) Во-вторых, величина dW_L формально совпадает с последующим ответом и потому иногда предлагается в качестве его обоснования. Как рецепт такой прием вполне допустим, но необходимо понимать, что за ним не стоит иного содержания, кроме проделанных выше выкладок. Итак,

$$-F dx = \frac{1}{2} I d\Phi \implies -F = \frac{I}{2} \frac{d\Phi}{dx} = \frac{dW_L}{dx}. \quad (7.20)$$

Далее для вычисления Φ заменим dx малой, но конечной величиной x и воспользуемся магнитной цепью с учетом того обстоятельства, что поток через контур в N раз больше потока через сердечник:

$$\Phi = \frac{NI \cdot N}{l/(\mu_0 \mu S) + 2x/(\mu_0 S)} = \mu_0 \mu S \frac{N^2 I}{l + 2\mu x}.$$

Максимальной величине F — подъемной силе — соответствует предел $x \rightarrow 0$:

$$F = -\frac{I}{2} \frac{d\Phi}{dx} \Big|_{x \rightarrow 0} = \mu_0 S \left(\frac{\mu N I}{l} \right)^2. \quad (7.21)$$

В заключение зададимся вопросом общего характера. Если сила Лоренца принципиально перпендикулярна скорости любой заряженной частицы, как вообще может совершаться работа при проникновении поля в вещество? Чему соответствует потенциальная энергия магнитоактивной среды?

Разрешение этого парадокса состоит в неперменном учете эффектов электромагнитной индукции. При любой перестройке поля изменение магнитной индукции $\partial \mathbf{B}/\partial t$ порождает электрическое поле \mathbf{E} , и работу совершает именно это поле. Но сценарий перестройки не входит в ответ, и конечные результаты оказываются универсальными, что и позволяет в очередной раз расширить формулировку закона сохранения энергии.

Задачи

1. Вдоль равновесного цилиндрического электронейтрального плазменного шнура (пинча) течет ток I (рис. 7.5). Определить температуру на оси пинча, считая температуру на его границе пренебрежимо малой. Плотность тока и концентрация частиц плазмы однородны по сечению, причем известно число частиц на единицу длины dN/dz .

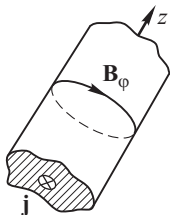


Рис. 7.5

Решение. Условие равновесия плазменного шнура — баланс силы Ампера и архимедовой силы. В плазме, как и в газе, давление можно выразить через концентрацию частиц и абсолютную температуру: $p = nk_B T$, где k_B — постоянная Больцмана. Таким образом, плотность архимедовой силы в нашей осесимметричной системе можно представить в виде

$$f_P = -\nabla nk_B T \Rightarrow (f_P)_r = -nk_B \frac{dT}{dr}.$$

Плотность силы Ампера также имеет только радиальную компоненту:

$$f_B = [jB] \Rightarrow (f_B)_r = -j_z B_\varphi = -jB.$$

Поле B при однородном распределении тока определяется с помощью теоремы о циркуляции (см. задачу 1 гл. 4): $B = (1/2)\mu_0 j \cdot r$. Из баланса сил следует

$$\frac{dT}{dr} = -\frac{\mu_0}{2} \frac{j^2}{nk_B} r.$$

Интегрируя это уравнение, получаем $T(r) \propto (\text{const} - r^2)$, причем из условия $T \approx 0$ на внешней границе пинча следует $\text{const} = R^2$, где R — радиус плазменного шнура. Вблизи его оси ($r = 0$) получаем ответ

$$T(0) = \frac{\mu_0}{4} \frac{j^2 R^2}{nk_B},$$

который приобретает окончательный вид после подстановки $j = I/\pi R^2$, $n = (dN/dz)/\pi R^2$:

$$T(0) = \frac{\mu_0}{4\pi} \frac{I^2}{k_B (dN/dz)}. \quad (7.22)$$

Формула (7.22) в магнитной гидродинамике называется по имени английского физика А. Беннета (1750–1799) *условием Беннета*. Интересно отметить, что при попытке решить задачу, приравнявая константе сумму газокINETического и магнитного давлений, мы получили бы результат, отличающийся от (7.22) в два раза — это как раз тот случай, когда в выражении для силы (7.3) нельзя пренебречь «натяжением силовых линий».

2. В результате некоторого космического события образовалась система, состоящая из звезды (масса M_0 , магнитный момент \mathbf{m}_0) и планеты (масса $M \ll M_0$, магнитный момент \mathbf{m}). Планета движется по круговой орбите радиуса R . Найти возможный разброс величины периода обращения в зависимости от ориентации магнитных моментов, считая плоскость орбиты перпендикулярной \mathbf{m}_0 .

Решение. Добавочная сила, действующая между диполями, может быть представлена в форме (7.4): $\mathbf{V} = \mu_0 \mathbf{H}$, где \mathbf{H} — поле диполя, которое дается формулой (4.24). В нашем случае вычисления существенно упрощаются, поскольку максимальные поправки к периоду обращения с очевидностью соот-

ветствуют параллельной и антипараллельной ориентациям диполей. При этом первый член в числителе (4.24) обращается в нуль, так что для энергии и силы взаимодействия получаются следующие выражения:

$$W = \mp \frac{\mu_0}{4\pi} \frac{m_0 m}{R^3} \Rightarrow \delta F_r = \mp \frac{3\mu_0}{4\pi} \frac{m_0 m}{R^4}.$$

Уравнение баланса сил во вращающейся системе отсчета:

$$M\omega^2 R = \gamma \frac{M_0 M}{R^2} \mp \frac{3\mu_0}{4\pi} \frac{m_0 m}{R^4},$$

где γ — гравитационная постоянная. Мы пренебрегли ради простоты ответа различием между массой планеты и приведенной массой, полагая, что наша поправка будет относительно больше, чем M/M_0 . Далее, представляя частоту обращения в виде $\omega^2 \approx \omega_0^2 \mp 2\omega_0 \delta\omega$, где $\omega_0^2 = \gamma M_0/R^3$, получаем

$$\frac{2\delta\omega}{\omega_0} = \frac{2\delta T}{T_0} = \frac{3\mu_0}{4\pi} \frac{m_0 m}{\gamma M_0 M R^2}.$$

Искомая величина разброса периода обращения $\Delta T = 2\delta T$, так что окончательный ответ выглядит следующим образом:

$$\Delta T = \frac{3\mu_0}{2} \frac{m_0 m}{M(\gamma M_0)^{5/2} R^{1/2}}.$$

3. Один из способов получения сверхсильных магнитных полей — взрывное сжатие металлического лайнера, подобного изображенному на рис. 7.2а, внутри которого предварительно создается магнитное поле с индукцией B_0 . Определить индукцию поля в цилиндре в момент максимального сжатия, если $B_0 = 5$ Тл, начальный внутренний радиус лайнера $R_0 = 5$ см, радиус в момент максимального сжатия $R_{\min} = 0,5$ см. Оболочку считать идеально проводящей. Оценить давление, необходимое для такого сжатия.

Ответ: $B_{\max} = 500$ Тл, $\mathcal{P} \sim 10^6$ атм.

4. Длинный цилиндрический стержень с магнитной проницаемостью μ и площадью сечения S расположен вдоль оси соленоида так, что один его конец находится внутри, а другой — вне соленоида. Найти силу F , с которой стержень втягивается в соленоид с собственным полем H .

Ответ: $F = \frac{1}{2} \mu_0 (\mu - 1) H^2 S$.

Глава 8

КВАЗИСТАЦИОНАРНОЕ ЭЛЕКТРОМАГНИТНОЕ ПОЛЕ

8.1. Условие квазистационарности электрической цепи. Релаксационные процессы

Основной предмет изучения в данной главе курса — электрические цепи переменного тока и, как частный случай — колебательный контур. В гл. 6 мы уже фактически рассматривали объекты такого рода в контексте закона электромагнитной индукции. При этом мы по умолчанию воспользовались некоторыми приближениями, которые обеспечили плавный переход от известных нам цепей постоянного тока.

Например, мы всюду вычисляли магнитное поле или вектор-потенциал таким образом, как если бы они порождались постоянным током в проводнике. В последующих главах мы сможем строго определить границы применимости такого рассмотрения, однако представление о них желательно иметь уже сейчас. Физика, как и любая естественная наука, постоянно оперирует с моделями реальности (и лишь на этом уровне может называться «точной»). И мы можем правильно оценивать наши результаты лишь постольку, поскольку нам известна область применимости модели.

Как уже было замечено, нам еще предстоит дополнить теорему о циркуляции с учетом как раз временной зависимости полей. Именно здесь возникает ограничение, которое принято называть условием квазистационарности, т. е. возможности распространять некоторые законы, установленные для стационарных цепей, на нестационарные цепи. («Квази-» означает «как бы...».)

Это условие можно получить, основываясь на том общеизвестном факте, что электромагнитный сигнал распространяется в вакууме со скоростью света $c = 3 \cdot 10^8$ м/с. В прозрачной среде (стекло) или специальной транспортирующей системе (волновод) эта скорость может быть меньше скорости света, но остается величиной того же порядка. Соответственно, если пространственный масштаб нашей системы (цепи) есть величина порядка l , то характерное время установления полей или токов в цепи (инерционное время) можно оценить как $\Delta t \sim l/c$.

Естественно предложить следующее неравенство в качестве условия квазистационарности:

$$\tau \gg \frac{l}{c}, \quad (8.1)$$

где τ — характерное время задачи. В частности, если речь идет о гармонических колебаниях либо переменном токе частоты ω , условие (8.1) эквивалентно следующему:

$$\omega \ll \frac{c}{l}. \quad (8.2)$$

При обратном соотношении мы заведомо должны рассматривать нашу цепь как систему с распределенными параметрами, т. е. просто решать общую задачу электродинамики. Вообще говоря, при рассмотрении электрических цепей обычно по умолчанию предполагается выполненным условие тонкого провода $r \ll l$, где r — характерный поперечный размер проводника. Тем самым обеспечивается независимость результата от распределения тока по сечению провода, а значит, и от самовоздействия через собственное магнитное поле. Но в одном из типичных элементов цепи — конденсаторе — это условие никогда не выполняется, поэтому приближение системы с сосредоточенными параметрами раньше всего отказывает именно в емкостном элементе.

Обсуждая условия квазистационарности, мы все время адресуемся к теореме о циркуляции. Полезно отметить также, какие законы, напротив, никак не меняются при переходе от стационарных полей к переменным. Во-первых, остается в силе теорема Гаусса как для электрического поля, так и для магнитного. (Это, правда, справедливо без оговорок лишь для точного — микроскопического — поля. Но мы уже говорили о неуниверсальности процедуры усреднения и ее результата.) Во-вторых, и это даже более интересно, остается в неприкосновенности закон электромагнитной индукции (6.10). Этот факт можно считать экспериментальным. Но тогда из теоремы Стокса следует, что сохраняет силу и интегральная форма закона Фарадея (6.7). Последнее обстоятельство выглядит отнюдь не тривиальным. Действительно, контур, быть может, не удовлетворяет условиям квазистационарности (8.1), (8.2), так что поле протекающих в нем токов нельзя рассматривать как безынерционное, тем не менее, интегральный закон (6.7) сохраняется в качестве одного из фундаментальных.

Пусть условие (8.1) выполнено. В качестве простейших примеров рассмотрим так называемые релаксационные или переходные процессы в цепях, представленных на рис. 8.1 а, б.

Пусть конденсатор C на рис. 8.1 а заряжен при разомкнутом ключе K и отсоединен от источника. В качестве начального условия можно взять, например, заряд на обкладках Q_0 . Замы-

каем ключ K и определяем момент замыкания как $t = 0$. Далее падение напряжения на сопротивлении подчиняется закону Ома, в котором напряжение есть просто разность потенциалов на обкладках:

$$U = \frac{Q}{C} = -RI.$$

Подставим $I = dQ/dt$ и в результате получаем простое дифференциальное уравнение, описывающее разрядку конденсатора через сопротивление R :

$$\frac{dQ}{dt} + \frac{Q}{\tau} = 0; \quad \tau \equiv RC.$$

Его решение при заданном начальном условии имеет вид

$$Q = Q_0 \exp\left(-\frac{t}{\tau}\right); \quad I = \frac{Q_0}{RC} \exp\left(-\frac{t}{\tau}\right). \quad (8.3)$$

Тот же результат можно получить и другим способом, скажем, составить закон Кирхгофа

$$\frac{1}{C} \int I dt + RI = 0,$$

а затем продифференцировать его по времени, и мы снова придем к тому же соотношению. Величина $\tau = RC$ называется *постоянной времени* или *временем релаксации* RC -цепочки. Именно эту величину следует подставлять в неравенство (8.1), оценивая пределы применимости модели.

Обратимся к RL -цепочке, представленной на рис. 8.1 б. Пусть вначале через катушку индуктивности L протекает некоторый ток I_0 , а затем в момент времени $t = 0$ одновременно размыкается ключ A и замыкается ключ B ; таким образом, в начальный момент задается величина тока I_0 в катушке. Уравнение для тока в цепи можно представить, например, в виде закона Ома, в котором роль напряжения выполняет ЭДС индукции в катушке:

$$\mathcal{E} = -\frac{d\Phi}{dt} = -L \frac{dI}{dt} = RI.$$

Решение похоже на (8.3):

$$I = I_0 \exp\left(-\frac{R}{L} t\right), \quad (8.4)$$

так что время релаксации $\tau_L = L/R$. Сколь бы ни было велико сопротивление, в течение этого времени ток остается по порядку величины близким к начальному. Значит, чем выше сопротивление, тем больше и пик напряжения, приходящийся на время релаксации. Этим и объясняется «искрение» при быст-

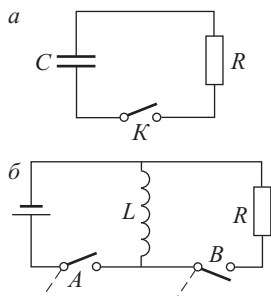


Рис. 8.1

ром обрыве тока в цепи конечной индуктивности: напряжение в максимуме оказывается выше пробойного для воздушного промежутка. Обратный эффект наблюдается при включении тока в цепи: индуктивность затягивает установление стационарного состояния. Пусть в момент $t = 0$ была замкнута цепь, состоящая из последовательно включенных источника постоянной ЭДС \mathcal{E}_0 , индуктивного элемента L и сопротивления R . Ток как функция времени подчиняется уравнению

$$L \frac{dI}{dt} + RI = \mathcal{E}_0.$$

Его общее решение состоит, как известно, из суммы любого частного решения и общего решения однородного уравнения

$$L \frac{dI}{dt} + RI = 0,$$

содержащего одну произвольную константу; последняя позволяет «привязать» решение к заданным начальным условиям. В нашем случае в качестве частного решения удобно выбрать просто константу $I_{\mathcal{E}} = \mathcal{E}_0/R$, а общее решение однородного уравнения имеет вид (8.4) с произвольной постоянной I_0 . Полагая ток в начальный момент времени равным нулю, получаем в итоге следующую зависимость $I(t)$:

$$I = \frac{\mathcal{E}_0}{R} \left[1 - \exp \left(-\frac{R}{L} t \right) \right], \quad (8.5)$$

демонстрирующую асимптотическое стремление к стационарному состоянию, тем более медленное, чем больше постоянная времени τ_L . Если бы мы заменили индуктивность конденсатором, то аналогичная зависимость получилась бы для напряжения на конденсаторе либо для заряда обкладок, а ток в цепи экспоненциально стремился бы к нулю. В этом случае постоянная времени в экспонентах равна RC . Предлагаем читателю самостоятельно в этом убедиться.

8.2. Колебательный контур

Подробное рассмотрение свойств колебательного контура еще предстоит в следующем разделе нашего курса, но и в курсе электродинамики невозможно оставить без внимания этот фундаментальный объект. Электрическая схема колебательного контура представлена на рис. 8.2. При этом подразумевается, что в сопротивление R включены не только резистивные элементы схемы (если они вообще существуют), но и «паразитное» сопротивление катушки индуктивности и соединительных проводов. При последовательном соединении элементов контура такое

объединение вполне корректно. Равным образом «паразитная» емкость катушки включена в C .

Вывод уравнения колебательного контура полезно предварить следующим замечанием. Если в схему включен источник ЭДС, то его можно включить в закон Ома или второй закон Кирхгофа как резистивный элемент, при том, что падение напряжения на нем должно быть равно $-\mathcal{E}$. Поэтому и уравнение контура, изображенного на рис. 8.2, можно представить в виде закона Кирхгофа с нулевой ЭДС на обходе:

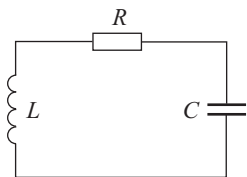


Рис. 8.2

$$L \frac{d^2 Q}{dt^2} + R \frac{dQ}{dt} + \frac{Q}{C} = 0. \quad (8.6)$$

(Мы учли, что заряд на обкладках и ток в контуре связаны соотношением $I = dQ/dt$.) Возможна эквивалентная форма:

$$L \frac{dI}{dt} + RI + \frac{1}{C} \int I dt = 0. \quad (8.7)$$

Дифференцируя (8.7) по времени, получаем снова уравнение (8.6), но уже для тока, $Q \Rightarrow I$. Таким образом, динамика контура в любом случае описывается дифференциальным уравнением второго порядка

$$\frac{d^2 x}{dt^2} + 2\delta \frac{dx}{dt} + \omega_0^2 x = 0, \quad (8.8)$$

где x — переменная величина (например, Q , I), $\delta = R/(2L)$, $\omega_0 = 1/\sqrt{LC}$.

Уравнение (8.8) представляет собой линейное дифференциальное уравнение второго порядка. Известны методы получения его решения; последнее в общем случае зависит от двух произвольных констант. В конкретной физической задаче эти константы доопределяются через начальные условия. Ввиду исключительной важности «колебательной» терминологии и понятийной системы для всего курса физики, мы уделим особое внимание свойствам решений уравнения (8.8).

Начнем с предельного случая нулевой диссипации $R \rightarrow 0 \Rightarrow \delta \rightarrow 0$.

Уравнение (8.8) в этом случае вырождается в простое уравнение гармонических колебаний, решение которого, включающее произвольные константы A , B или A , φ , может быть представлено одной из следующих эквивалентных форм:

$$\begin{aligned} x &= A \cos \omega_0 t + B \sin \omega_0 t; & x &= A \cos(\omega_0 t + \varphi); \\ x &= A \sin(\omega_0 t + \varphi); & x &= A e^{i\omega_0 t} + B e^{-i\omega_0 t}. \end{aligned}$$

Эквивалентность последнего выражения трем предыдущим обусловлена известными соотношениями — формулами Эйлера:

$$\cos \omega t = \frac{1}{2} (e^{i\omega t} + e^{-i\omega t}); \quad \sin \omega t = \frac{1}{2i} (e^{i\omega t} - e^{-i\omega t});$$

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

Такое представление осциллирующих функций экспонентами от мнимого аргумента, как мы вскоре убедимся, есть наиболее конструктивный метод описания волновых и колебательных процессов.

Коэффициенты A , B называются *амплитудными*, а величина φ , «сдвигающая» начало отсчета времени, именуется *фазой*; фазой называют также и полный аргумент осциллирующих функций $\omega_0 t + \varphi$.

Решение полного уравнения (8.8) отличается от этих чисто гармонических функций лишь учетом диссипации. В самом общем виде оно содержит две произвольные константы и может быть представлено, например, следующим образом:

$$x = A_1 e^{p_1 t} + A_2 e^{p_2 t}, \quad (8.9)$$

где $p_{1,2}$ суть корни алгебраического уравнения

$$p^2 + 2\delta \cdot p + \omega_0^2 = 0. \quad (8.10)$$

В случае достаточно слабой диссипации решения (8.9) не должны сильно отличаться от чисто колебательных. Тогда корни $p_{1,2}$ оказываются комплексными величинами, так что мы можем, используя формулы Эйлера, записать решение в эквивалентной форме:

$$x = A e^{-\delta t} \cos(\omega t + \varphi), \quad (8.11)$$

здесь A — амплитуда, φ — фаза колебательного решения, $\omega = (\omega_0^2 - \delta^2)^{1/2}$. Сразу видно, что при $\delta < \omega_0$ решение, действительно, оказывается осциллирующим во времени. В обратном случае оно представляет собой монотонно затухающую функцию, которую лучше всего представлять именно в виде (8.9). Величина $\delta = R/(2L)$ называется *декрементом затухания*; по физическому смыслу это обратное время затухания колебаний в $e=2,718\dots$ раз. Если $\delta \ll \omega_0$, колебания можно считать почти периодическими, подобно приведенным выше решениям для уравнения гармонических колебаний. Это неравенство удобно переписать как условие для характерных времен:

$$RC \ll \frac{L}{R} \quad \text{или просто} \quad R \ll \sqrt{\frac{L}{C}}. \quad (8.12)$$

При этом $\omega \approx \omega_0$. При условии (8.12) корни уравнения (8.10) оказываются непременно комплексными, а с ними и решения (8.9).

Надлежащим выбором констант интегрирования они «привязываются» к вещественным начальным условиям; соответственно, решение (8.11) представлено в формально вещественном виде. Но такое представление оказывается не слишком удобным. Применяя к той или иной осциллирующей функции линейные операции дифференцирования либо интегрирования, мы всегда вынуждены будем превращать \sin в \cos и наоборот, да еще и с учетом знака; дополнительно усложняет как процедуру, так и ответ экспонента, модулирующая в (8.11) тригонометрическую функцию. Намного проще та же операция с чисто экспоненциальными функциями типа (8.9). Экспонента при дифференцировании или интегрировании просто умножается на постоянный множитель, каждый раз один и тот же. С другой стороны, как мы уже видели, \cos представляет собой функционально вещественную часть экспоненты ($a \sin$ — мнимую), поэтому в линейной задаче (пока решения не приходится перемножать, возводить в степень и т. д.) их нетрудно поставить во взаимно однозначное соответствие. Таким образом, решение типового уравнения (8.8) ищется в виде

$$x = A \exp(i\omega t - \delta t + \varphi), \quad (8.13)$$

а затем физические величины (I , U) получают, взяв от него только вещественную или только мнимую часть. Чаще всего в этом даже нет необходимости, поскольку амплитуда и фаза, определенные из начальных условий, полностью описывают колебательный процесс.

При дифференцировании решения (8.13) последнее просто умножается на $(i\omega - \delta)$, при интегрировании — делится на ту же величину. Из соотношения

$$i \equiv e^{i\pi/2}$$

следует, что при умножении или делении на мнимую единицу полная фаза меняется на $\pi/2$ — это как раз и соответствует переходу синуса в косинус и наоборот.

При необходимости работать с квадратичными величинами (как-то энергия колебаний) представление (8.13) уже не годится, но экспоненциальную форму решения можно сохранить, заменив (8.13) следующим — уже вполне точным — решением:

$$X = \frac{1}{2} (x + x^*),$$

где звездочка — знак комплексного сопряжения. Подобно тому, как у механического маятника сумма потенциальной и кинетической энергий есть величина постоянная (в пренебрежении диссипацией!), а средние их значения по периоду равны, для колебательного контура при $\delta \ll \omega_0$ имеют место следующие

соотношения:

$$\frac{L}{2} \langle I^2 \rangle_T = \frac{C}{2} \langle U^2 \rangle_T; \quad \frac{LI^2}{2} + \frac{CU^2}{2} = W = \text{const.} \quad (8.14)$$

Читателю предлагается самому убедиться в справедливости (8.14); заметим, что, помимо данного выше рецепта, можно пользоваться и тригонометрическим представлением (8.11) вместе с известными свойствами средних:

$$\langle \cos^2 \omega t \rangle_T \equiv \frac{\omega}{2\pi} \int_{-\pi/\omega}^{\pi/\omega} \cos^2 \omega t \, dt = \langle \sin^2 \omega t \rangle_T = \frac{1}{2}.$$

При учете конечной величины затухания получаем $W \propto \exp(-2\delta t)$. Как характеристика малости затухания часто используется величина, именуемая в теории колебаний *добротностью* (обозначение — Q , от английского *quality*). По определению

$$Q = \frac{\omega_0}{2\delta} = \frac{\omega_0 L}{R} = \frac{1}{\sqrt{RC}} \sqrt{\frac{L}{R}}. \quad (8.15)$$

В любой из этих формул подразумевается, что добротность — характеристика колебательного контура. На самом деле это понятие гораздо шире, а потому мы приведем его к более универсальному виду. Пусть W — запас энергии в контуре. Оценим потери энергии за период, считая диссипацию слабой: $\delta \ll \omega$.

$$\Delta_T W = W(1 - e^{-2\delta \cdot T}) \approx W \cdot 2\delta \cdot T = \frac{2\pi W}{Q}.$$

Таким образом, добротность, с точностью до коэффициента — это обратное отношение энергии, запасенной в колебательной системе, к ее потерям за период:

$$Q = \frac{2\pi W}{\Delta_T W}. \quad (8.16)$$

8.3. Вынужденные колебания. Переменный ток

Из проведенного выше рассмотрения переходных процессов — формулы (8.3)–(8.5) — следует, что емкостные и индуктивные элементы можно включать в цепь так, как если бы это были сопротивления, но, в отличие от последних, падение напряжения на них в зависимости от тока задается не умножением, а другими линейными операциями:

$$U_C = \frac{1}{C} \int I(t) \, dt; \quad U_L = L \frac{dI}{dt}. \quad (8.17)$$

Поэтому, например, для случая гармонической функции $I(t)$ падение напряжения на конденсаторе или катушке индуктивности, помимо амплитудного коэффициента, характеризуется также фазовым сдвигом $\Delta\varphi = \pm\pi/2$. Такие элементы называются *реактивными*, в отличие от *активного* элемента — сопротивления. (Строго говоря, любой соленоид, если только он не сверхпроводящий, обладает активным сопротивлением, так что и сдвиг фаз оказывается меньше $\pi/2$.) Главное свойство чисто реактивного элемента состоит в том, что синусоидальный ток проходит его без энергетических потерь. В самом деле, подсчитаем выделение энергии в среднем за период:

$$\langle \dot{W} \rangle = \frac{\omega}{2\pi} \int_{-\pi/\omega}^{\pi/\omega} I(t)U(t) dt \propto \int_{-\pi/\omega}^{\pi/\omega} \cos \omega t \sin \omega t dt = 0.$$

В общем случае синусоидальных тока и напряжения, сдвинутых по фазе на φ :

$$I = I_0 \cos \omega t; \quad U = U_0 \cos(\omega t + \varphi),$$

среднее энергосодержание равно

$$\langle \dot{W} \rangle = \frac{\omega}{2\pi} \int_{-\pi/\omega}^{\pi/\omega} I(t)U(t) dt = \frac{1}{2} I_0 U_0 \cos \varphi. \quad (8.18)$$

Нетрудно заметить, что знак сдвига фазы не имеет значения. Это вполне физично, поскольку «опережение» или «отставание» по фазе суть понятия условные и зависят, например, от принятого в определении (8.13) знака частоты. Так же легко понять, что, пока и поскольку мы не включаем явно в нашу цепь динамо-машину, ограничиваясь пассивными элементами R , L , C , сдвиг фазы не может превышать $\pi/2$, точнее, при произвольном определении момента включения, он обязан принадлежать интервалу $[(2n - 1/2)\pi; (2n + 1/2)\pi]$, где n — любое натуральное число. В противном случае энергия в цепи не рассеивалась бы, а порождалась.

В частном случае активного сопротивления (8.18) дает

$$\langle \dot{W} \rangle = \frac{1}{2} I_0 U_0 \equiv R \langle I^2 \rangle_T = \frac{\langle U^2 \rangle_T}{R}.$$

Это соотношение принято привязывать к формуле (3.13) и таким образом определять эффективное напряжение сети; например, стандартному напряжению $U_{\text{эф}} = 220$ В соответствует амплитуда синусоидального напряжения $U_0 = \sqrt{2U_{\text{эф}}^2} \approx 311$ В. Так же опре-

деляется и эффективное значение тока — как постоянный ток, эквивалентный данному переменному по тепловому эффекту; соответственно, $I_{\text{эф}} = I_0/\sqrt{2}$.

Предположим, что мы имеем дело с цепочкой, изображенной на рис. 8.2, но теперь в контур последовательно включен источник синусоидальной ЭДС, так что, в отличие от (8.6), (8.7), динамика тока в цепи определяется уравнением:

$$L \frac{dI}{dt} + RI + \frac{1}{C} \int I dt = \mathcal{E}_0 e^{i\Omega t}. \quad (8.19)$$

Как уже говорилось, решение линейного уравнения с правой частью может быть представлено как сумма любого частного решения, удовлетворяющего этому уравнению, и общего решения однородного уравнения (т. е. без правой части); последнее в случае уравнений второго порядка зависит от двух произвольных констант, согласующих его с начальным моментом времени. Действительно, пусть $I_1(t)$ и $I_2(t)$ — два решения (8.19), тогда разность $I_1(t) - I_2(t)$ должна быть, в силу линейности, решением уравнения (8.7).

Общее решение однородного уравнения можно взять в форме (8.11); оно с течением времени будет затухать. Характерное время затухания $\tau \sim 1/\delta$. Частное решение будем искать в виде незатухающих колебаний с частотой приложенного напряжения —

$$I(t) = I \exp(i\Omega t); \quad I = I_0 e^{i\varphi}. \quad (8.20)$$

Аналогичная зависимость от времени должна получиться для падения напряжения на любом элементе, заряда конденсатора и прочих динамических величин. Ради удобства последующих операций мы полагаем величину I в (8.20) комплексной — помимо собственно амплитуды I_0 в нее входит и сдвиг фаз φ по отношению к внешней ЭДС. Подставляя (8.20) в (8.19), убеждаемся, что оно и в самом деле является решением. Комплексную амплитуду тока можно получить из уравнения

$$I \cdot \left(i\Omega L + R + \frac{1}{i\Omega C} \right) = \mathcal{E}_0.$$

Нетрудно усмотреть прямую связь с формулами (8.17) — получился просто закон Ома в частном случае синусоидального напряжения и тока. Но теперь мы можем грамотно ответить на вопрос о процессе установления тока при включении внешней ЭДС, тем более важный, что фаза $\mathcal{E}(t)$ в момент включения не определена. Итак, включение переменного тока сопровождается «звоном» собственных колебаний контура, затухающим с характерным временем $1/\delta$. Именно эти колебания обеспечивают согласование цепи с фазой источника. Если же мы, как обычно в электротехнике, интересуемся установившимися колебаниями

при $t \gg 1/\delta$, то можно использовать закон Ома, приписав пассивным элементам следующие комплексные сопротивления (такое «сопротивление» принято называть *импедансом*):

$$Z_L = i\Omega L; \quad Z_R = R; \quad Z_C = \frac{1}{i\Omega C}. \quad (8.21)$$

С помощью несложных преобразований из выражения для тока

$$I = \frac{\mathcal{E}_0}{R + i(\Omega L - 1/(\Omega C))}$$

следуют ответ для амплитуды

$$I_0 = (II^*)^{1/2} = \frac{\mathcal{E}_0}{\sqrt{R^2 + (\Omega L - 1/(\Omega C))^2}} \quad (8.22)$$

и фазы

$$\operatorname{tg} \varphi = \frac{\operatorname{Im} I}{\operatorname{Re} I} = \frac{\Omega L - 1/(\Omega C)}{R}. \quad (8.23)$$

Полученные нами ответы допускают удобную графическую интерпретацию — и, в принципе, их можно получать именно графически. Это так называемый *метод векторных диаграмм*. Интересующий нас случай демонстрируется на рис. 8.3. Представим каждый ток и каждое напряжение вектором, которому

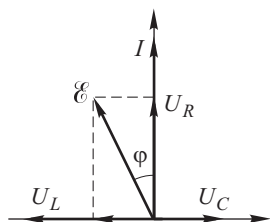


Рис. 8.3

в полярных координатах соответствуют модуль (амплитуда) и полярный угол (фаза). Зададим произвольно направление вектора I , а длина его задает масштаб. Напряжение U_R синфазно току, соответственно, откладываем $U_R = IR$ параллельно вектору I . Напряжение $U_L = \Omega LI$ откладывается перпендикулярно I с опережением по фазе (соответственно мнимой единице в коэффициенте), а $U_C = I/(\Omega C)$ — с отставанием. Сумма всех трех векторов напряжения даст вектор \mathcal{E} , так что из рисунка можно определить и фазу, и — по отношению масштабов длины — модуль импеданса $|Z| = \mathcal{E}_0/I_0$. Для простой схемы графический метод не дает существенных преимуществ в сравнении, например, с формулами (8.22), (8.23), но для сложных многосвязных схем такой анализ может быть заметно легче алгебраических вычислений.

Предположим, что мы при заданной амплитуде \mathcal{E}_0 варьируем частоту внешней ЭДС Ω . Из (8.22), а еще нагляднее — из рис. 8.3 — можно усмотреть, что при $U_L = U_C$ (по модулю) отклик — амплитуда тока — будет максимален, т. е. имеет место резонанс. При последовательном соединении элементов он называется *резонансом напряжений*. Резонансная частота

оказывается равной как раз собственной частоте контура без учета диссипации $\omega_0 = 1/\sqrt{LC}$. На этой частоте сопротивление цепи — чисто активное, $Z_{\text{рез}} = R$. Кривая отклика $I(\Omega)$ показана на рис. 8.4. (Вблизи резонансной частоты она обычно так и называется — *резонансной кривой*.) Чтобы сравнивать резонансные кривые различных контуров, принято сравнивать их ширину на уровне отклика составляющего $1/\sqrt{2}$ от максимума. Определим, какому сдвигу частоты этот уровень соответствует, полагая $\delta\Omega \ll \omega_0$:

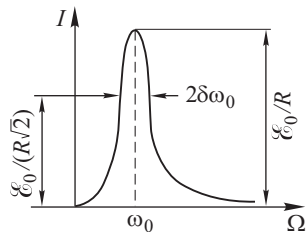


Рис. 8.4

$$\frac{I_{\max}}{I(\omega_0 \pm \delta\Omega)} = \sqrt{2} = \frac{1}{R} \sqrt{R^2 + (L(\omega_0 + \delta\Omega) - (\omega_0 + \delta\Omega)^{-1}C^{-1})^2}.$$

Вспоминая определение добротности (8.15), эту формулу можно привести к следующему виду:

$$\sqrt{2} \approx \sqrt{1 + (\delta\Omega/\omega_0)^2 \left(\frac{1}{\omega_0 RC} + \frac{\omega_0 L}{R} \right)^2} = \sqrt{1 + Q^2 \left(\frac{2\delta\Omega}{\omega_0} \right)^2}.$$

Окончательный ответ:

$$\Delta\Omega = 2\delta\Omega = \frac{\omega_0}{Q}. \quad (8.24)$$

Отсюда можно понять, почему добротности присвоен столь «значущий» термин. Еще один вид резонанса — его называют резонансом токов — представлен на рис. 8.5 а, б. Для простоты

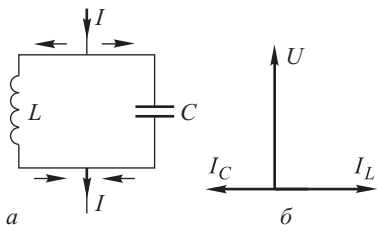


Рис. 8.5

будем считать сопротивление катушки индуктивности пренебрежимо малым. В этом предельном случае результат можно получить без вычислений, на основе лишь векторной диаграммы 8.5 б. На параллельно соединенных элементах общим должно быть напряжение, а токи I_C и I_L , соответственно, опережая и

отставая по фазе на $\pi/2$, вычисляются по уже известному нам рецепту. И если $I_L = I_C$ — это получается при $\Omega = 1/\sqrt{LC}$, — то при отличном от нуля напряжении ток в неразветвленном участке цепи равен нулю. В действительности при учете активного сопротивления получается просто более или менее острый минимум тока, при этом импеданс цепи опять соответствует чисто активному сопротивлению.

Как пример отклика электрической цепи на несинусоидальную переменную ЭДС рассмотрим цепочки, изображенные на рис. 8.6. На входные клеммы подается некоторый сигнал $\mathcal{E}(t)$ с характерным временем эволюции τ . Можно для оценок пользоваться также характерной частотой в спектре сигнала $\omega \sim \tau^{-1}$.

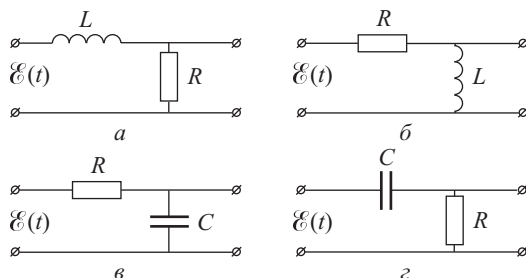


Рис. 8.6

Отклик — напряжение — снимается с выходных клемм, как показано на рисунках; предполагается, что сопротивление измерительного прибора столь велико, что он не вносит заметного вклада в уравнение цепи. Тогда динамика цепочек рис. 8.6 а, б описывается уравнением

$$L\dot{I} + RI = \mathcal{E}(t).$$

а) Если $RI \ll L\dot{I}$, т. е. $L/R \gg \tau$, то $\dot{I} \approx \mathcal{E}(t)/L \Rightarrow I \approx \approx L^{-1} \int \mathcal{E}(t) dt$, таким образом, с сопротивления будет снят сигнал

$$U_R \propto I \propto \int \mathcal{E}(t) dt.$$

б) В обратном случае $L/R \ll \tau$ получаем $I \approx \mathcal{E}(t)/R$. Снимая сигнал с индуктивности, имеем

$$U_L \propto \dot{I} \propto \frac{d\mathcal{E}(t)}{dt}.$$

В свою очередь, цепочкам рис. 8.6 в, г отвечает уравнение:

$$R\dot{Q} + \frac{Q}{C} = \mathcal{E}(t).$$

в) Если $RI \gg Q/C$, т. е. $RC \gg \tau$, то $I \approx \mathcal{E}(t)/R$, и с конденсатора можно снять сигнал

$$U_C \propto Q \propto \int \mathcal{E}(t) dt.$$

г) При обратном соотношении $RC \ll \tau$ получаем $Q \approx C\mathcal{E}(t)$; соответственно сигнал с сопротивления —

$$U_R \propto \dot{Q} \propto \frac{d\mathcal{E}(t)}{dt}.$$

Мы рассмотрели все четыре случая, базируясь на оценках, но те же результаты можно получить и более строго на основе фурье-разложения сигнала. Цепочки a , b на рис. 8.6 называют *интегрирующими*, а b , c — *дифференцирующими*. (Разумеется, при этом важна не только схема включения, но и соотношения характерных времен, заданные выше.) Их часто используют в качестве элементов различных измерительных схем. Конечно, точность выполняемых ими операций не слишком высока — как и у всех вообще аналоговых систем — зато они очень просты.

8.4. Скин-эффект

В настоящем параграфе мы рассмотрим особый случай квазистационарных процессов — токи высокой частоты, для которых остаются справедливыми условия (8.1), (8.2), но теряется понятие электрической цепи как совокупности элементов с сосредоточенными параметрами. Тем самым будут получены дополнительные к (8.1), (8.2) условия того, что систему проводников можно будет рассматривать как электрическую цепь.

Skin — по-английски «кожа», скин-эффект — протекание токов достаточно высокой частоты в тонком приповерхностном слое хорошего проводника. При этом в глубине проводящей среды ток может быть сколь угодно малым. Как мы вскоре убедимся, благоприятным условием для проявления скин-эффекта оказывается как можно более высокая проводимость вещества. Это может быть металл, а может и плазма, которая при термоядерных температурах $\sim 10^7$ – 10^8 К по проводимости превосходит металлы примерно на порядок. Однако при достаточно высоких частотах скин-эффект характерен даже для человеческого тела — при контакте с источником происходит прогрев тонкого слоя кожи, но исключается поражение электрическим током. Сколь высока должна быть частота и сколь хороша проводимость, станет ясно из оценок, которые вскоре будут сделаны, а пока мы примем этот факт как экспериментальный.

Трактовать это явление можно по-разному, но в общем, классический скин-эффект вполне объясняется классической электродинамикой, и мы этим ограничимся, но не пойдем традиционным формальным путем, а построим наше изложение на фундаментальном принципе Ленца и его количественном выражении — законе Фарадея (6.7):

$$\mathcal{E} = \oint \mathbf{E} d\mathbf{l} = -\frac{d\Phi}{dt}.$$

Ради упрощения последующего рассмотрения предположим, что у проводящей среды $\mu \approx 1$, что всегда верно для плазмы и с хорошей точностью — для большинства хорошо проводящих металлов.

Напомним экспериментальный факт, уже упомянутый в гл. 7, — замедление падения монеты (строго говоря, любого проводящего тела), если в процессе падения она пересекает силовые линии магнитного поля, например, если мы ее роняем между полюсами мощного электромагнита. Это происходит, как мы уже выяснили, вследствие возбуждения ЭДС индукции и, как следствие, генерации индукционных токов.

Если бы внешнее поле совсем не проникало в проводящую среду, монета вообще бы повисла в воздухе — это случай так называемого магнитного удержания или вмороженности поля. Но она падает, хотя и медленно, значит, поле как-то все же в проводник просачивается. Чтобы понять, как это происходит, рассмотрим модельную плоскую задачу. Геометрия ее представлена на рис. 8.7. Пусть в начальный момент времени вне проводника существует однородное магнитное поле \mathbf{B} , а в толще проводника поле равно нулю; скачок поля на границе поддерживается поверхностным током. Именно так должно случиться, согласно

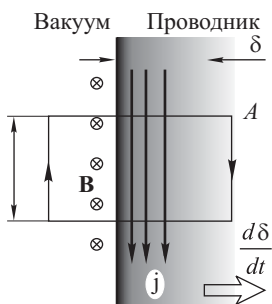


Рис. 8.7

принципу Ленца, при быстром включении внешнего магнитного поля. Связь между поверхностной (линейной) плотностью тока и полем вне проводника была нами установлена на основании теоремы о циркуляции в задаче 2 гл. 4: $H_{\text{вн}} = i/2$. Пусть этот ток распределен в слое толщины δ вблизи поверхности проводника, т. е. $j \sim i/\delta$. Таким образом, индукция внешнего поля и плотность тока связаны соотношением

$$B = \mu_0 H \approx \mu_0 \frac{dI}{dt} \approx \mu_0 j \delta.$$

Чтобы рассмотреть эффект проникновения магнитного поля в среду (или, что то же, расширения токового слоя вглубь), дополним наши оценки законом Ома в локальной форме: $\mathbf{j} = \lambda \mathbf{E}$. Заодно перейдем от общего закона (6.7) к конкретной оценке для контура A на рис. 8.7:

$$\frac{d\Phi}{dt} \approx B l \frac{d\delta}{dt} = E l, \quad (8.25)$$

что дает $B d\delta/dt \approx j/\lambda \approx B/(\lambda\mu_0\delta)$. Учитывая, что $\delta d\delta/dt = (1/2)(d\delta^2/dt)$, получаем окончательный и притом очень простой ответ

$$\delta^2 \approx (\lambda\mu_0)^{-1} t. \quad (8.26)$$

Этот результат абсолютно тождествен закону диффузии:

$$\langle x^2 \rangle \propto Dt, \quad (8.27)$$

где D — коэффициент диффузии. (В нашем курсе процессы переноса, включая диффузию, будут подробно рассмотрены в заключительном — пятом разделе.)

Таким образом, поле (и ток) просачивается в проводящую среду по диффузионному закону. Так и говорят: диффузия магнитного поля. Этот результат получен нами в модельной задаче как следствие простых оценок, но он является совершенно общим, и выражение для коэффициента диффузии поля оказывается верным вплоть до коэффициента $D_B = 1/(\lambda\mu_0)$.

И вот теперь на базе этого результата мы немедленно и практически без вычислений сможем понять природу скин-эффекта. Пусть на границе проводника создано быстропеременное поле (как следует из теоремы о циркуляции, по ней при этом протекает быстропеременный ток). Оценим глубину, на которую продиффундирует поле за период,

$$\delta_{\text{эф}} \approx (D_B T)^{1/2} = \left(\frac{2\pi}{\omega \lambda \mu_0} \right)^{1/2}. \quad (8.28)$$

Если этот масштаб много меньше l , где l — характерный размер проводника, то поле и ток окажутся локализованы в более или менее тонком слое у поверхности. Отсюда можно получить и обещанные условия на частоту и проводимость. Существенным оказывается произведение этих двух величин. Если

$$\omega \lambda \gg \frac{1}{\mu_0 l^2}, \quad (8.29)$$

то мы будем иметь дело не с однородным распределением тока по сечению проводника, которое до сих пор всегда молчаливо предполагали, но со скин-эффектом.

Точные решения системы уравнений электродинамики для плоского проводника, а равно и цилиндрического, если толщина скин-слоя много меньше его радиуса, дают

$$B \propto E \propto j \propto \exp\left(-\frac{x}{\delta_0}\right), \quad \text{где } \delta_0 = \frac{\delta_{\text{эф}}}{(\pi)^{1/2}},$$

что, безусловно, демонстрирует очень хорошее соответствие с нашей простой оценкой.

В случае ярко выраженного скин-эффекта существенным образом меняются как сопротивление, так и самоиндукция проводников. Как следствие, цепь стремится превратиться в систему с распределенными параметрами. Если в формуле (8.29) в качестве l фигурирует поперечный размер проводников, то это условие, не посягая на квазистационарность задачи — неравенства (8.1), (8.2) остаются в силе, в принципе ставит под вопрос само приближение электрической цепи. Таким образом,

при условии (8.29), возможно, придется перейти от сравнительно простой электротехники к полноценной электродинамической задаче.

Задачи

1. Последовательно соединенные дроссель и омическое сопротивление R присоединены к источнику постоянного напряжения с ЭДС \mathcal{E} (рис. 8.8а). Индуктивность дросселя, когда в него вставлен железный сердечник G , равна L_1 , а без сердечника — L_2 . Вначале сердечник был вставлен, и ток в цепи установился. В момент времени $t = 0$ сердечник очень быстро вынимают (за время, много меньшее времени релаксации). Определить ток в цепи $I(t)$ при $t > 0$.

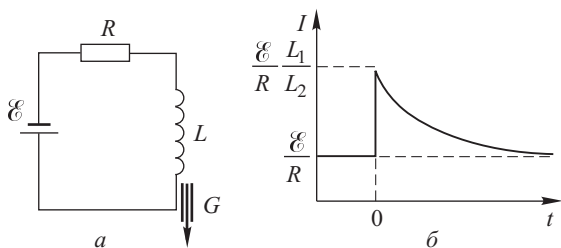


Рис. 8.8

Решение. При быстром удалении сердечника изменяется индуктивность дросселя и ток через него, но магнитный поток измениться мгновенно не может:

$$\Phi = L_1 \frac{\mathcal{E}}{R} = L_2 I_0,$$

где I_0 — ток через дроссель в момент $t = +0$. В последующие моменты времени цепь подчиняется уравнению:

$$\mathcal{E} = IR + L_2 \frac{dI}{dt}.$$

Решение этого линейного уравнения с правой частью имеет вид

$$I(t) = I \exp\left(-\frac{R}{L_2}t\right) + \frac{\mathcal{E}}{R}.$$

Это уравнение надо решить с начальным условием $I(t = 0) = I_0$, откуда следует ответ (рис. 8.8б)

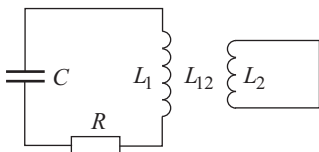


Рис. 8.9

$$I(t) = \frac{\mathcal{E}}{R} \left[1 + \frac{L_1 - L_2}{L_2} \exp\left(-\frac{R}{L_2}t\right) \right].$$

2. Вблизи катушки колебательного контура с параметрами L_1 , C , R расположена вторая катушка с индуктивностью L_2 (рис. 8.9). Взаимная индуктивность катушек равна L_{12} . Какой будет частота собственных колебаний контура, если выводы второй катушки замкнуты накоротко? Считать, что активное сопротивление второй катушки пренебрежимо мало. При каком условии резонанс невозможен?

Решение. Собственные колебания контура описываются системой уравнений:

$$L_{12} \frac{dI_1}{dt} + L_2 \frac{dI_2}{dt} = 0; \quad L_1 \frac{dI_1}{dt} + L_{12} \frac{dI_2}{dt} + RI_1 + C^{-1} \int I_1 dt = 0.$$

Исключая dI_2/dt , получаем

$$\left(L_1 - \frac{L_{12}^2}{L_2} \right) \frac{d^2 I_1}{dt^2} + R \frac{dI_1}{dt} + \frac{I_1}{C} = 0,$$

откуда и можно определить все интересующие нас величины:

$$\omega = \sqrt{\frac{1}{LC} - \left(\frac{R}{2L} \right)^2}, \quad \delta = \frac{R}{2L}, \quad \text{где } L = L_1 - \frac{L_{12}^2}{L_2}.$$

Заметим, что $L \geq 0$, потому что $L_{12} \leq L_1 L_2$. Резонанс недостижим, если $(R/2L)^2 > 1/(LC)$, т. е. если

$$\frac{R^2 C}{4} > L_1 - \frac{L_{12}^2}{L_2}.$$

3. Длинный соленоид с плотной намоткой размещен на цилиндрическом железном сердечнике с магнитной проницаемостью μ и проводимостью λ . Соленоид замкнут на конденсатор, в результате чего образован контур с резонансной частотой ω (рис. 8.10). Радиус сердечника r_0 , утечки в конденсаторе несущественны, обмотку и соединительные провода можно считать идеально проводящими. Определить добротность контура.

Решение 1. Будем считать, что $\omega \lambda \ll 1/(\mu_0 \mu r_0^2)$. При этом условии можно пренебречь скинированием токов Фуко в сердечнике. Тогда магнитное поле внутри сердечника можно считать однородным и записать закон Фарадея в виде

$$E \cdot 2\pi r = \omega \pi r^2 B_0 \sin \omega t \Rightarrow E = \frac{B_0}{2} \omega r \sin \omega t,$$

где $E(r)$ — вихревое электрическое поле. Подсчитаем мощность энерговыделения на единицу длины сердечника:

$$\frac{d}{dt}(\dot{W}) = \int E \cdot \lambda E dS = \frac{1}{4} \lambda \omega^2 B_0^2 \sin^2 \omega t \int_0^{r_0} 2\pi r^3 dr.$$

Таким образом, потери за период на единицу длины будут равны

$$\frac{d}{dt} \Delta W = \frac{\pi^2}{8} B_0^2 r_0^4 \omega \lambda = \frac{\pi^2}{2} \frac{\langle BH \rangle_T}{2} \mu \mu_0 r_0^4 \omega \lambda.$$

А с другой стороны, полная энергия, запасенная в контуре, равна

$$W = W_L + W_C = 2 \langle W_L \rangle_T = \langle BH \rangle_T \cdot \pi r_0^2 l.$$

Вспоминая определение добротности (8.16), получаем ответ

$$Q = \frac{8}{\mu \mu_0 \omega \lambda r_0^2}.$$

Решение 2. Рассмотрим ту же задачу в пределе высоких частот, когда токи Фуко заведомо скинированы, т. е. $\omega \lambda \gg 1/(\mu_0 \mu r_0^2)$. Теперь ток и поле

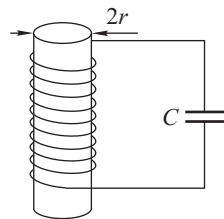


Рис. 8.10

в сердечнике будут сосредоточены в тонком слое $\delta_{\text{эф}} \ll r_0$, прилегающем к поверхности, где $\delta_{\text{эф}}$ задается формулой (8.28). Чтобы не смешивать эффекты разной природы, предположим, что магнитная проницаемость от частоты не зависит.

Полагая неравенства достаточно сильными, мы не будем решать точной задачи о распределении полей и энерговыведения, а ограничимся оценочными формулами, а именно:

$$H \cdot 2\pi r = \frac{B}{\mu\mu_0} \cdot 2\pi r = I \approx j \cdot 2\pi r \delta_{\text{эф}}.$$

и

$$\frac{d}{dt}(\dot{W}) \approx \frac{\langle j^2 \rangle_T}{\lambda} S_{\text{эф}} \approx \left\langle \frac{H}{\delta_{\text{эф}}} \frac{B}{\mu\mu_0 \delta_{\text{эф}}} \right\rangle_T S_{\text{эф}} \approx \frac{\omega}{2\pi} \langle HB \rangle_T S_{\text{эф}}.$$

Отсюда

$$Q \approx \frac{1}{\omega T} \approx 1.$$

4. Железный сердечник несет на себе две обмотки (рис. 8.11). Одна обмотка, из большого числа n витков, присоединена к источнику синусоидальной ЭДС \mathcal{E} . Другая обмотка состоит из одного однородного кольца, сопротивление которого — R . Точки A, B, C этого кольца отстоят друг от друга на равных расстояниях. Если к двум из этих точек присоединить амперметр переменного тока с сопротивлением r , то что он покажет? Рассмотреть два варианта подключения, обозначенных на рис. 8.11.

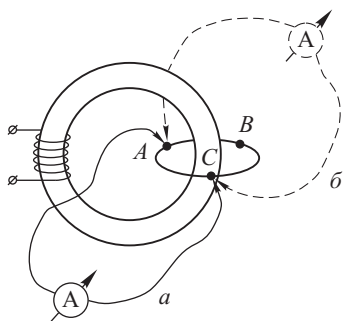


Рис. 8.11

Ответ: а) $I_A = 3\mathcal{E}/[n(9r + 2R)]$; б) $I_A = 6\mathcal{E}/[n(9r + 2R)]$.

5. Для схемы, изображенной на рис. 8.12, определить частоты источника ЭДС, соответствующие резонансам токов и напряжений. Построить график сдвига фазы тока I относительно ЭДС источника \mathcal{E} в зависимости от частоты источника, считая все активные сопротивления пренебрежимо малыми.

Ответ: а) резонанс токов $Z^{-1} = 0 \Rightarrow \omega_0 = 1/\sqrt{LC}$; б) резонансы напряжений $Z = 0 \Rightarrow \omega_1 = \sqrt{2 - \sqrt{3}}/\sqrt{LC} \approx 0,52/\sqrt{LC}$, $\omega_2 = \sqrt{2 + \sqrt{3}}/\sqrt{LC} \approx 1,93/\sqrt{LC}$. Зависимость сдвига фазы от частоты представлена на рис. 8.13.

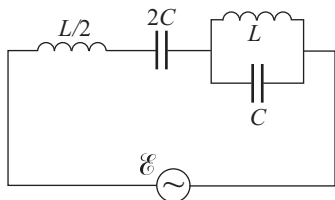


Рис. 8.12

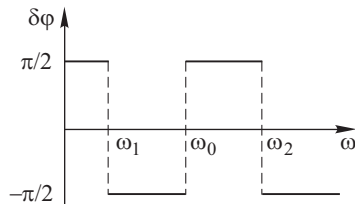


Рис. 8.13

Глава 9

УРАВНЕНИЯ МАКСВЕЛЛА

9.1. Ток смещения. Обобщение теоремы о циркуляции

Данную тему мы начнем с мысленного эксперимента. Представим себе заряженный сферический конденсатор (рис. 9.1), у которого пространство между обкладками заполнено однородным диэлектриком. Пусть заряды обкладок равны $\pm Q$. Предположим, что в некоторый момент времени диэлектрик практически мгновенно и всюду однородно переводится в проводящее состояние. В случае мысленного эксперимента неважно, как именно это сделано — скажем, импульсом ионизирующего излучения. Итак, в силу абсолютной симметрии постановки задачи, конденсатор начинает разряжаться также симметричным образом, т. е. вектор плотности тока направлен строго по радиусу:

$$\mathbf{j} = -\frac{\mathbf{r}}{r} \frac{\partial Q}{\partial t}.$$

Знак «—» учитывает, что ток разряжает конденсатор, понижая заряд, а зависимость $j \propto 1/r^2$ отвечает закону сохранения заряда, который должен выполняться, по крайней мере, в случае достаточно медленного процесса разрядки. Действительно, если при протекании тока нигде в диэлектрике не происходит накопления заряда, ток через любую сферическую поверхность, концентрическую с обкладками, должен быть одинаков, а значит, плотность тока обратно пропорциональна площади сферы.

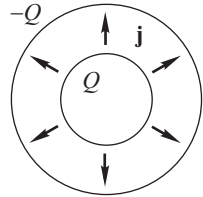


Рис. 9.1

При попытке применить теорему о циркуляции в форме (4.10) мы столкнулись бы с очевидным абсурдом. Единственным разрешенным направлением для \mathbf{H} в случае столь высокой симметрии было бы радиальное направление \mathbf{r}/r . Но эта возможность исключена теоремой Гаусса для магнитного поля (4.1), поскольку такая конфигурация отвечала бы магнитному заряду в центре симметрии. Таким образом, единственно разумный ответ в данной задаче — нулевое магнитное поле. А в то же время, приложив формулу (4.10) к любой токовой трубке, мы получили бы $\oint \mathbf{H} d\mathbf{l} \neq 0$. Значит, мы должны так модифицировать теорему (4.10), чтобы в данной задаче получить строго нуль в правой части.

Другую иллюстрацию той же проблемы представляет рис. 9.2. Квазистационарный ток $I(t)$ протекает по цепи,

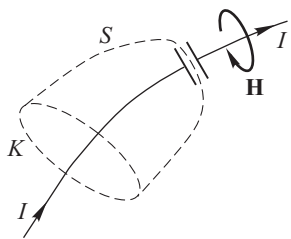


Рис. 9.2

содержащей конденсатор. Вдали от конденсатора магнитное поле у проводов должно определяться формулой (4.10). Натянем на контур K , охватывающий провод, такую поверхность S , которая проходила бы между пластинами конденсатора. Тогда поток заряда (ток) через нее будет равен нулю, но совершенно очевидно, что значение магнитного поля, а значит и циркуляция его по любому контуру от таких построений не

зависит. Таким образом, модификация теоремы (4.10) должна «замкнуть» ток через конденсатор.

Удобнее все же вносить поправки не в формулу (4.10) как таковую, а в соответствующее ей локальное уравнение (4.12):

$$\operatorname{rot} \mathbf{H} = \mathbf{j} + \dots ?$$

Вернемся к ситуации, изображенной на рис. 9.1. Поправка, обращающая для этого случая в нуль правую часть (4.12), должна иметь вид

$$-\mathbf{j} = \frac{1}{4\pi} \frac{\partial Q}{\partial t} \frac{\mathbf{r}}{r^2} \equiv \frac{\partial \mathbf{D}}{\partial t}. \quad (9.1)$$

Можно убедиться в том, что проблему, представленную на рис. 9.2, поправка (9.1) полностью разрешает. В самом деле,

$$\frac{\partial D}{\partial t} = \varepsilon_0 \frac{\partial U}{\partial t} \frac{d}{d} = \frac{\varepsilon_0}{d} \frac{\partial U}{\partial t} = \frac{C}{S} \frac{\partial U}{\partial t} = \frac{1}{S} \frac{\partial Q}{\partial t}.$$

Интеграл по поверхности S рис. 9.2 даст просто полный ток I . Величина, представленная формулой (9.1), называется *плотностью тока смещения*, а поток от нее — соответственно, *током смещения*. К тому же результату можно прийти и другим путем, исходя из закона сохранения заряда (3.7):

$$\frac{\partial \rho}{\partial t} + \operatorname{div} \mathbf{j} = 0.$$

Если мы возьмем дивергенцию от уравнения (4.12), то в левой части получим нуль: $\operatorname{div} (\operatorname{rot} \mathbf{H}) = \nabla [\nabla \mathbf{H}] \equiv 0$. Значит, нуль должен получиться и в правой части нового уравнения

$$\operatorname{div} [\mathbf{j} + \dots] = -\frac{\partial \rho}{\partial t} + \dots = \operatorname{div} \left[-\frac{\partial \mathbf{D}}{\partial t} + \dots \right],$$

что тождественно удовлетворяется как раз добавкой (9.1). Все приведенные рассуждения являются хорошими аргументами в пользу (9.1), но отнюдь не бесспорными доказательствами, и окончательное слово остается, как обычно в физике, за

экспериментом. Логика, которой мы следовали, для физики вообще характерна. Мы стартовали от простых эмпирических законов (Кулона, Био и Савара и т. п.), а пришли к соотношениям гораздо более общим и базирующимся на новых понятиях. После этого надлежит от аналитических методов вернуться к эксперименту и убедиться в правильности нашего понимания законов природы. Экспериментальное тестирование этих законов будет продолжаться и далее, но уже на предмет установления границ их применимости. Таким образом, основные уравнения электродинамики — уравнения Максвелла — должны рассматриваться, подобно законам Ньютона, именно как обобщение огромного массива экспериментальной информации.

9.2. Уравнения электромагнитного поля

Наиболее компактный вид *уравнения Максвелла* имеют в *локальной форме*:

$$\operatorname{rot} \mathbf{H} = \mathbf{j} + \frac{\partial \mathbf{D}}{\partial t}; \quad (9.2)$$

$$\operatorname{rot} \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}; \quad (9.3)$$

$$\operatorname{div} \mathbf{D} = \rho; \quad (9.4)$$

$$\operatorname{div} \mathbf{B} = 0. \quad (9.5)$$

Особенно простой и согласованной выглядит эта система уравнений для электромагнитного поля в вакууме, когда по физическому смыслу величины \mathbf{E} и \mathbf{D} , а также \mathbf{B} и \mathbf{H} тождественны (в системе СГС они просто равны друг другу). Но и в этом случае система уравнений еще не замкнута, поскольку должны быть еще определены величины $\mathbf{j}(\mathbf{E}, \mathbf{B})$ и $\rho(\mathbf{E}, \mathbf{B})$. Таким образом, уравнения поля, вообще говоря, должны решаться совместно с уравнениями, описывающими вещество. Последние, в частности для случая поля в среде, позволяют установить связь между полем и индукцией:

$$\mathbf{D} = \hat{\epsilon} \mathbf{E}; \quad (9.6)$$

$$\mathbf{B} = \hat{\mu} \mathbf{H}. \quad (9.7)$$

Здесь $\hat{\epsilon}(\mathbf{E}, \mathbf{B})$ и $\hat{\mu}(\mathbf{E}, \mathbf{B})$ — самый общий вид операторов, связывающих две величины. В простейшем случае, которым мы пока в основном ограничивались, это просто константы вещества. В анизотропных средах, скажем кристаллах, они могут превратиться в тензоры, могут вместо константы быть представлены интегральным оператором, могут быть комплексными и нелинейными по полям. Любое из этих представлений неуниверсально, границы применимости каждого из них существенно уже, чем собственно уравнений поля, и по традиции уравнения (9.6), (9.7)

вместе с законом Ома и т. п. выделяют в особую группу и называют *материальными уравнениями*, тогда как уравнениями Максвелла считаются именно (9.2)–(9.5).

Как мы уже могли убедиться, для решения многих, в особенности простых, задач может быть весьма полезна *интегральная форма уравнений Максвелла* — совершенный эквивалент формул (9.2)–(9.5):

$$\oint \mathbf{H} d\mathbf{l} = I_{\Sigma} + \frac{\partial}{\partial t} \int_S \mathbf{D} d\mathbf{S}; \quad (9.8)$$

$$\oint \mathbf{E} d\mathbf{l} = -\frac{\partial \Phi}{\partial t}; \quad (9.9)$$

$$\oint \mathbf{D} d\mathbf{S} = q_{\Sigma}; \quad (9.10)$$

$$\oint \mathbf{B} d\mathbf{S} = 0. \quad (9.11)$$

Любое уравнение можно брать в любой форме, локальной либо интегральной, руководствуясь исключительно соображениями удобства. Система уравнений Максвелла решается при следующих граничных условиях на поверхности:

$$[\mathbf{E}_{\tau}] = 0; \quad (9.12)$$

$$[\mathbf{B}_n] = 0; \quad (9.13)$$

$$[\mathbf{D}_n] = 0 \quad \text{или} \quad \sigma = \frac{dq}{dS}; \quad (9.14)$$

$$[\mathbf{H}_{\tau}] = 0 \quad \text{или} \quad i = \frac{dI}{dl}. \quad (9.15)$$

Напомним, что квадратные скобки здесь означают скачок соответствующей величины на поверхности. В правой части двух последних уравнений стоит нуль или, соответственно, поверхностная плотность заряда и тока в зависимости от того, рассматриваем ли мы истинную границу или представляем в качестве таковой слой с толщиной много меньшей характерного пространственного масштаба задачи.

Было бы заблуждением считать, что граничные условия (9.12)–(9.15) нами уже были выведены. Мы действительно получали каждое из них в главах 2 и 5, но при существенных упрощающих предположениях. Например, условие (9.12) получено для случая потенциального поля, которое тождественно обратило бы в нуль левую часть (9.3) или (9.9). В определенном смысле граничные условия следуют из самих уравнений (9.2)–(9.5) и точно так же должны рассматриваться как обобщение экспериментальных данных.

Электрическое поле при учете эффекта электромагнитной индукции теряет свойство потенциальности. Но оказывается,

добавление в систему уравнений введенного в гл. 4 вектор-потенциала позволяет в принципе представить два вектора \mathbf{E} и \mathbf{H} как функции двух потенциалов. Не останавливаясь на этом вопросе подробно, ограничимся случаем поля в вакууме:

$$\mathbf{H} = \text{rot } \mathbf{A}; \quad \mathbf{E} = -\nabla\varphi - \mu_0 \frac{\partial \mathbf{A}}{\partial t}. \quad (9.16)$$

Взяв ротор от выражения для \mathbf{E} , получим уравнение (9.3), а дивергенция \mathbf{H} оказывается нулевой, как и следует из (9.5) для поля в вакууме. Можно видеть, что, как мы уже заметили в гл. 6, ЭДС, т. е. непотенциальность электрического поля, обязательно связана с нестационарностью.

Из уравнений (9.16) вытекает весьма полезное свойство потенциалов, которое принято называть калибровочной инвариантностью. Физически наблюдаемыми (измеряемыми) величинами являются поля \mathbf{E} и \mathbf{H} , между тем, даже в электростатике потенциал φ определен лишь с точностью до аддитивной постоянной. Обратившись к первому уравнению (9.16), нетрудно заметить, что магнитное поле не изменится, если к величине \mathbf{A} мы добавим градиент любой скалярной функции ψ , поскольку ротор любого градиента равен нулю (отсюда ранее употреблявшийся термин — «градиентная инвариантность»). Но для того чтобы из второго уравнения (9.16) получилось то же самое поле \mathbf{E} , следует к скалярному потенциалу добавить величину $-\mu_0 \partial\psi / \partial t$. Таким образом, уравнения Максвелла в общем случае оказываются инвариантными относительно следующей замены:

$$\varphi' = \varphi - \mu_0 \frac{\partial\psi}{\partial t}, \quad (9.17)$$

$$\mathbf{A}' = \mathbf{A} + \nabla\psi.$$

Очевидно, добавление к φ произвольной константы, не зависящей от координат, есть частный случай преобразований (9.17).

9.3. Теорема Пойнтинга

В предшествующих главах мы расширили известное из механики понятие энергии, введя корректным образом энергию электрического, а затем и магнитного поля. Теперь, базируясь на уравнениях Максвелла, мы можем проследить, как эволюционирует энергия поля в динамической задаче. Это, в свою очередь, позволяет распространить закон сохранения энергии на такие среды, в которых существенно взаимодействие поля с веществом.

Итак, мы исходим из выражений (2.29) и (7.11). Таким образом, плотность энергии электромагнитного поля (если ее можно

выделить из общей энергии системы «поле-вещество») равна

$$w = \frac{(\mathbf{E}\mathbf{D})}{2} + \frac{(\mathbf{B}\mathbf{H})}{2}. \quad (9.18)$$

Предположим, что мы следим за эволюцией энергии поля в некотором объеме V , ограниченном поверхностью s :

$$\frac{\partial W}{\partial t} = \frac{\partial}{\partial t} \int_V w dV = \int_V \left[\frac{\partial}{\partial t} \frac{\mathbf{E}\mathbf{D}}{2} + \frac{\partial}{\partial t} \frac{\mathbf{B}\mathbf{H}}{2} \right] dV. \quad (9.19)$$

На данный момент ограничимся случаем линейной среды

$$\mathbf{D} \propto \mathbf{E}; \quad \mathbf{B} \propto \mathbf{H},$$

что позволит нам провести строгое доказательство последующих утверждений. Действительно, в этом случае производные квадратичных членов в (9.19) могут быть преобразованы следующим образом:

$$\frac{\partial W}{\partial t} = \int_V \left(\mathbf{E}, \frac{\partial \mathbf{D}}{\partial t} \right) + \left(\mathbf{H}, \frac{\partial \mathbf{B}}{\partial t} \right) dV. \quad (9.20)$$

Затем подставим соответствующие частные производные из уравнений Максвелла (9.2) и (9.3):

$$\frac{\partial W}{\partial t} = - \int_V (\mathbf{j}\mathbf{E}) dV + \int_V (\mathbf{E}, \text{rot } \mathbf{H}) - (\mathbf{H}, \text{rot } \mathbf{E}) dV. \quad (9.21)$$

Первый интеграл в (9.21) можно сопоставить с формулой (3.12), откуда сразу становится ясной его физическая природа — это джоулево тепловыделение в объеме V . Кстати, полезно заметить, что ток смещения напрямую к тепловому эффекту не ведет — это можно видеть из того же выражения (9.21). Второй интеграл — чисто «полевой» — может быть преобразован с помощью следующего векторного тождества:

$$\text{div} [\mathbf{E} \mathbf{H}] = \text{div} [\check{\mathbf{E}} \mathbf{H}] + \text{div} [\mathbf{E} \check{\mathbf{H}}] = (\mathbf{H}, \text{rot } \mathbf{E}) - (\mathbf{E}, \text{rot } \mathbf{H}).$$

Таким образом, объемный интеграл может быть сведен к поверхностному по теореме Гаусса:

$$\int_V (\mathbf{E}, \text{rot } \mathbf{H}) - (\mathbf{H}, \text{rot } \mathbf{E}) dV = - \int_V \text{div} [\mathbf{E} \mathbf{H}] dV = - \int_s [\mathbf{E} \mathbf{H}] ds,$$

а все выражение (9.21) может быть сведено к следующему:

$$\frac{\partial W}{\partial t} = - \int_V (\mathbf{j}\mathbf{E}) dV - \int_s \mathbf{S} ds. \quad (9.22)$$

Словами это можно выразить так: изменение со временем электромагнитной энергии в объеме обусловлено, во-первых,

джоулевой диссипацией и, во-вторых, «вытеканием» энергии через поверхность, ограничивающую данный объем. Вектор потока электромагнитной энергии

$$\mathbf{S} = [\mathbf{E} \mathbf{H}] \quad (9.23)$$

называется *вектором Пойнтинга* по имени английского физика Дж.Г. Пойнтинга (1852–1914), а закон эволюции энергии в замкнутом объеме (9.22) — *теоремой Пойнтинга*. Ее можно представить и в локальной форме. Для этого последний интеграл в правой части нужно оставить в виде объемного, а левую часть — также в виде объемного интеграла от плотности энергии. Поскольку объем интегрирования может быть совершенно произвольным, должны быть равны подынтегральные выражения:

$$\frac{\partial w}{\partial t} + \operatorname{div} \mathbf{S} = -\mathbf{j} \cdot \mathbf{E}. \quad (9.24)$$

В таком виде теорема Пойнтинга напоминает закон сохранения заряда (3.7), только в данном случае имеет место еще и сток энергии в тепло через джоулеву диссипацию. При отсутствии таковой мы получили бы точный закон сохранения электромагнитной энергии.

А теперь перейдем от линейной среды к общему случаю произвольной зависимости $\mathbf{D}(\mathbf{E})$, $\mathbf{B}(\mathbf{H})$. Пока и поскольку мы могли выделить энергию электромагнитного поля аддитивным образом, эта связь предполагалась линейной. Следовательно, в общем случае мы не сможем отделить друг от друга энергию поля, энергию вещества и энергию взаимодействия поля с веществом. Но поскольку расширение понятий и зависимостей представляется гораздо более естественным, нежели их полная ревизия, имеет смысл распространить и на этот случай действие теоремы Пойнтинга, доопределив все остальные добавки к энергии и потоку энергии как взаимодействие поля с веществом. При таком определении мы как бы узаконим проведенную выше операцию:

$$\frac{\partial}{\partial t} \left(\frac{(\mathbf{E} \mathbf{D})}{2} + \frac{(\mathbf{B} \mathbf{H})}{2} \right) = \left(\mathbf{E}, \frac{\partial \mathbf{D}}{\partial t} \right) + \left(\mathbf{H}, \frac{\partial \mathbf{B}}{\partial t} \right).$$

Таким образом, работа по изменению электрического и/или магнитного поля — каким бы образом она ни совершалась — может быть представлена в виде

$$\delta A = (\mathbf{E}, d\mathbf{D}) + (\mathbf{H}, d\mathbf{B}). \quad (9.25)$$

Рассмотрим пример. Предположим, что мы поместили ферромагнитный образец в периодическое внешнее магнитное поле. По прошествии каждого периода он неукоснительно возвращается в исходное состояние. Этот процесс требует энергетических

затрат, которые пропорциональны (с точностью до размагничивающего фактора) площади, охватываемой кривой гистерезиса:

$$\Delta Q_T \approx V \cdot \oint_{B(H)} H dB.$$

Далее рассмотрим некоторые простые примеры вычисления вектора Пойнтинга. Представим себе цилиндрический проводник, по которому течет ток с равномерно распределенной по сечению плотностью \mathbf{j} (рис. 9.3). Электрическое поле вне провода вблизи его поверхности должно быть, согласно условию (9.12), таким же, как и внутри, т. е. $\mathbf{E} = \mathbf{j}/\lambda$, магнитное поле $\mathbf{H} = I/(2\pi r) = \mathbf{j} \cdot r/2$, соответственно, вектор Пойнтинга нормален к поверхности проводника, направлен внутрь него и равен

$$|\mathbf{S}| = EH = \frac{j^2 r}{2\lambda}.$$

Энергия, «втекающая» в проводник на длине l

$$\dot{W}_{in} = |\mathbf{S}| \cdot 2\pi r l = \pi r^2 l \cdot \frac{j^2}{\lambda},$$

в точности соответствует темпу джоулевой диссипации, как и следует тому быть в стационарной задаче.

Другой пример представлен на рис. 9.4. Длинная двухпроводная линия соединяет источник постоянной ЭДС и нагрузку — сопротивление R . Полагая сопротивление подводящих проводов пренебрежимо малым, определим направление вектора Пойнтинга. Мы при этом ограничимся малой окрестностью проводов. Направление магнитного поля задается правилом буравчика (см. рис. 9.4). Электрическое поле есть не что иное, как поле двумерного диполя. Дело в том, что при идеальной проводимости соединительных элементов все падение напряжения приходится на сопротивление, так что каждый из проводов оказывается эквипотенциальным при разности потенциалов между ними IR . Отсюда и поле диполя. Определяя направление векторного произведения $[\mathbf{E} \mathbf{H}]$ в данной геометрии, нетрудно убедиться в том, что в любой точке вектор Пойнтинга будет «показывать» на нагрузку, что вполне естественно: в стационарной задаче поток энергии должен быть направлен туда, где происходит диссипация, поскольку $\dot{W} \equiv 0$.

Отметим важное обстоятельство. Из теоремы Гаусса следует, что оба провода в данной системе должны быть заряжены, что не

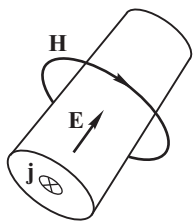


Рис. 9.3

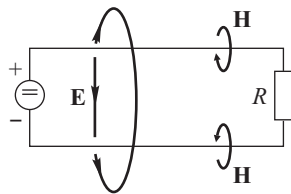


Рис. 9.4

всегда правильно понимается при решении задач на протекание тока в твердом проводнике. В отличие от электростатики, здесь не поле задается зарядом, а заряд — полем, которое следует из разности потенциалов IR . Но все закономерности, изученные нами ранее, продолжают работать, и забывать об этом не следует.

9.4. Импульс электромагнитного поля

Мы достаточно много внимания уделили проблеме энергии электромагнитного поля, но поле обладает, вообще говоря, также и импульсом, распределенным, подобно энергии, в пространстве с некоторой плотностью $d\mathbf{p}/dV$. Ввести его корректным образом было бы на уровне нашего курса непросто, но мы можем сделать это на основании достаточно очевидных аналогий.

Как мы уже выяснили, вектор Пойнтинга — это вектор потока энергии, т. е. модуль его есть энергия, переносимая электромагнитным полем в единицу времени через площадку единичной площади, ориентированную перпендикулярно вектору потока. (А если площадка ориентирована как-то иначе, для вычисления той же величины нужно взять проекцию вектора Пойнтинга на нормаль к площадке.)

Это не единственная потоковая величина, с которой мы имели дело. Например, вектор плотности тока \mathbf{j} можно с равным правом именовать вектором потока заряда, поскольку он представляет собой как раз заряд, протекающий в единицу времени через единицу площади поперечного сечения проводника. Как мы уже убедились (см. гл. 3), в случае единственного сорта носителей заряда его можно достаточно просто выразить через концентрацию заряженных частиц и их локальную скорость:

$$\mathbf{j} = nev.$$

Теперь представим себе, что вместо заряда, мы подставили в эту формулу массу частиц. Тогда, очевидно, мы получим выражение для потока массы

$$\mathbf{q}_m = nm\mathbf{v}.$$

В случае течения многокомпонентного газа (жидкости, плазмы) это выражение следовало бы просуммировать по сортам частиц, но для наших рассуждений в этом нет надобности. Нетрудно заметить, что поток массы представляет собой не что иное, как пространственную плотность импульса среды:

$$\mathbf{q}_m = nm\mathbf{v} \equiv \frac{d\mathbf{p}}{dV}. \quad (9.26)$$

Соотношение (9.26) остается в силе и в релятивистском случае, если под массой понимать не массу покоя, а так называемую ре-

лятивистскую массу m_r , которая как самостоятельная величина в настоящее время употребляется редко — предпочитают просто выражать ее через полную энергию частицы \mathcal{E} (к сожалению, стандартное обозначение энергии в релятивистской механике пересекается с не менее стандартным обозначением ЭДС):

$$m_r = \frac{m}{\sqrt{1 - v^2/c^2}} \equiv \frac{\mathcal{E}}{c^2}; \quad \mathbf{p} = \frac{m\mathbf{v}}{\sqrt{1 - v^2/c^2}} = \frac{\mathcal{E}}{c^2} \mathbf{v}.$$

Из этих общеизвестных формул и (9.26) следует:

$$\frac{d\mathbf{p}}{dV} = \frac{1}{c^2} \mathbf{q}_w, \quad \text{где} \quad \mathbf{q}_w \equiv n\mathcal{E}\mathbf{v}, \quad (9.27)$$

т. е. плотность импульса с точностью до коэффициента $1/c^2$ есть поток энергии частиц. И вот этот результат, абсолютно строгий в случае течения вещества, мы и перенесем на электромагнитное поле:

$$\frac{d\mathbf{p}_{\text{эм}}}{dV} = \frac{\mathbf{S}}{c^2} = \frac{[\mathbf{E} \mathbf{H}]}{c^2}. \quad (9.28)$$

Из формулы (9.28) можно извлечь несколько поучительных следствий. Во-первых, как нетрудно видеть, чисто электрическое или чисто магнитное поле импульсом не обладают. Во-вторых, если $\mathbf{E} \parallel \mathbf{H}$, плотность импульса также равна нулю. В-третьих, представим себе одну из названных ситуаций и предположим, что мы перешли в некоторую движущуюся систему отсчета. Поле, как мы теперь понимаем, имеет массу. Следовательно, в новой системе отсчета у поля окажется ненулевой плотность импульса, а это значит, в согласии с (9.28), в новой системе отсчета обязательно должны быть представлены оба поля \mathbf{E} , \mathbf{H} , и притом не параллельные друг другу. Справедливо и обратное: можно показать, что мы всегда можем перейти в такую систему отсчета, в которой импульс электромагнитного поля равен нулю,

т. е. существует или только \mathbf{E} , или только \mathbf{H} . Мы это уже проделали однажды — в задаче 3 гл. 4.

Рассмотрим электрическую цепь, представленную на рис. 9.5. Постоянный ток I протекает по длинному соленоиду радиуса r_2 с плотной намоткой n витков/м, далее —

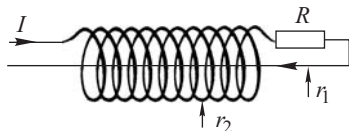


Рис. 9.5

через сопротивление R и возвращается по соосному с соленоидом прямому проводу радиуса r_1 . Найдём вектор Пойнтинга внутри катушки вдали от ее торцов, считая пренебрежимо малыми сопротивления как соленоида, так и соединительных проводов. При таком предположении разность потенциалов между обмоткой соленоида и центральным проводом равна $\Delta\varphi = IR$. Электрическое

поле внутри катушки направлено, в силу симметрии, по радиусу, а значит, $E(r) \propto 1/r$ (полезно еще раз обратить внимание на замечание в конце предыдущего параграфа). Зная разность потенциалов, его можно найти точно:

$$\mathbf{E}(\mathbf{r}) = -\frac{\mathbf{r}}{r^2} \frac{IR}{\ln(r_2/r_1)}.$$

Что касается магнитного поля, то оно имеет две компоненты: одна создается током, текущим по осевому проводу, а другая — током через обмотку соленоида:

$$H_\varphi = \frac{I}{2\pi r}; \quad H_z = nI.$$

Соответственно и вектор Пойнтинга имеет две компоненты. Одна из них соответствует направлению на нагрузку:

$$S_z = EH_\varphi = \frac{1}{2\pi} \frac{I^2 R}{r^2 \ln(r_2/r_1)}.$$

Ей отвечает полный поток энергии через сечение катушки

$$\int_{r_1}^{r_2} S_z \cdot 2\pi r \, dr = I^2 R.$$

Другая компонента вектора Пойнтинга направлена по азимуту:

$$S_\varphi = EH_z = \frac{I^2 R n}{r \ln(r_2/r_1)}.$$

Соответствующий поток энергии как будто бы «бесполезен» — он отнюдь не направлен на нагрузку, а просто циркулирует вокруг оси системы. Что это значит, можно понять, обратившись еще раз к соотношению (9.28). Оказывается, электромагнитное поле может иметь не только импульс, но и ненулевой момент импульса. Только что рассмотренная система дает тому самый простой пример.

Задачи

1. Цилиндрический нерелятивистский электронный пучок радиуса r_0 распространяется в свободном пространстве. Электроны пучка летят параллельно, энергия их w , а концентрация n . Найти величину и направление вектора Пойнтинга в любой точке пространства.

Решение. См. рис. 9.6. Вектор \mathbf{S} всюду будет параллелен оси пучка и направлен туда же, куда и скорость частиц. Электрическое поле ищется из теоремы Гаусса, магнитное — из теоремы о циркуляции. Результаты будут различными при $r < r_0$ (индекс *in*) и $r > r_0$ (индекс *ex*). Как всегда в симметричных конфигурациях, ток вне контура или заряд вне поверхности

интегрирования на величине поля не сказываются. Плотность тока в пучке $j = ne(2w/m)^{1/2}$. Далее,

$$\begin{aligned} E_{in} &= \frac{ner}{2\varepsilon_0}, & H_{in} &= \frac{rj}{2} = rne\sqrt{\frac{w}{2m}}, & S_{in} &= \frac{n^2 e^2 r^2}{4\varepsilon_0} \sqrt{\frac{2w}{m}}; \\ E_{ex} &= \frac{neR^2}{2\varepsilon_0 r}, & H_{ex} &= \frac{R^2 j}{2r} = \frac{R^2 ne}{r} \sqrt{\frac{w}{2m}}, & S_{ex} &= \frac{n^2 e^2 R^4}{4\varepsilon_0 r^2} \sqrt{\frac{2w}{m}}. \end{aligned}$$

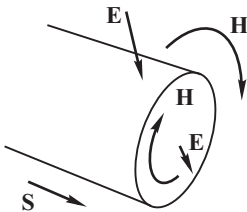


Рис. 9.6

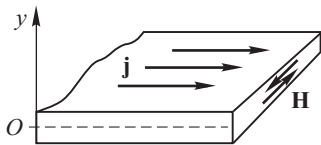


Рис. 9.7

2. По проводнику, имеющему форму плоской ленты, течет ток с плотностью j . Носителями тока являются электроны с концентрацией n . Найти величину и направление вектора Пойнтинга в произвольной точке внутри проводника вдали от края ленты. Считать толщину ленты много меньше ее ширины, сопротивление не учитывать; $\mu \approx 1$.

Решение. См. рис. 9.7. В плоскости симметрии ленты магнитное поле равно нулю, вне ее силовые линии лежат в плоскостях $y = \text{const}$, причем, как следует из теоремы о циркуляции, $B(y) = jy$. В отсутствие сопротивления единственный источник электрического поля — эффект Холла, но не во внешнем магнитном поле, которого здесь нет, а в собственном поле тока:

$$-ne\mathbf{E} + [\mathbf{j}\mathbf{B}] = 0; \quad \mathbf{j} \perp \mathbf{B} \Rightarrow E = \frac{j\mu_0}{ne} H = \frac{\mu_0 j^2 y}{ne}.$$

При этом вектор \mathbf{E} направлен извне по направлению к плоскости симметрии проводника. Таким образом, искомый вектор Пойнтинга оказывается антипараллелен вектору \mathbf{j} (т. е. сонаправлен со скоростью электронов — носителей тока) и равен $S = \mu_0 j^3 y^2 / (ne)$.

3. Плоский воздушный конденсатор, обкладками которого являются два одинаковых диска, заряжен до высокой разности потенциалов, а затем отключен от источника напряжения. В центре конденсатора происходит пробой — по оси проскакивает электрическая искра — как следствие, конденсатор разряжается. Считая разряд квазистационарным и пренебрегая краевыми эффектами, определить полное количество электромагнитной энергии, вытекающее за время разряда из пространства между обкладками.

Ответ: искомая величина равна нулю вследствие полной компенсации тока проводимости током смещения.

4. Обкладки плоского конденсатора имеют форму дисков радиуса R . Расстояние между дисками $d \ll R$. Пространство между ними заполнено однородным диэлектриком с диэлектрической и магнитной проницаемостями ε и μ . Конденсатор включен в цепь переменного тока $I = I_0 \cos \omega t$. Пренебрегая краевыми эффектами, определить отношение максимальной магнитной энергии в конденсаторе к максимальной электрической.

Ответ: $W_m/W_e = (1/8)\mu\mu_0\varepsilon\varepsilon_0(\omega R)^2$.

Глава 10

ЭЛЕКТРОМАГНИТНЫЕ ВОЛНЫ

10.1. Вопросы волновой динамики. Волновое уравнение

Мы начнем настоящую главу с достаточно краткой сводки результатов, относящихся к волновым движениям вообще. В какой-то части они будут напоминанием уже известного нам из курса механики или даже школьной физики. Придется, к сожалению, в некоторых вопросах забежать вперед, анонсируя результаты следующего раздела нашего курса. Подобные элементы неизбежно возникают в любом учебнике физики, ибо природа не знает, что мы разделили ее на главы и параграфы.

Одним из важнейших достижений физики XIX века явилось осознание общих свойств и, соответственно, возможности универсального подхода в изучении явлений самой различной природы. Это не относилось, конечно, к экспериментальным методам, но зато весьма способствовало адекватному пониманию и созданию общего для всей линейной физики математического формализма — уравнений математической физики.

В XX веке естественным продолжением этой линии стало взрывное развитие физики колебаний и волн. Пока и поскольку речь идет о линейных волнах, основные количественные закономерности — уравнения, соотношения характерных констант, законы подобия и т. п. — оказываются совершенно универсальными.

Итак, *волной* называется процесс, обладающий определенным свойством инвариантности, таким, что некоторая физическая величина представляется профилем, перемещающимся с постоянной скоростью:

$$f(x, t) \equiv f(x \pm ut), \quad (10.1)$$

где u — скорость волны. Если в данной конкретной среде соотношение (10.1) может быть выполнено без всяких оговорок, т. е. для произвольного профиля $f(x, t)$, среда называется бездисперсной. Но этот случай не является типичным. Куда более характерна ситуация, в которой (10.1) справедливо для некоторых частных функций $f_n(x, t)$ (их называют *собственными функциями* задачи), а эволюция произвольной функции лишь постольку определяется законом (10.1), поскольку эта функция

может быть представлена разложением по собственным функциям:

$$f(x, t) = \sum_n f_n(x - u_n t).$$

В качестве базиса собственных функций во многих физических задачах наиболее адекватным оказывается набор функций, представляющих гармонические колебания

$$f_n(x, t) = A_n \cos[k_n(x - u_n t)]. \quad (10.2)$$

Заметим, что мы по умолчанию определились со знаком в формуле (10.1). Знак «+» соответствовал бы волне, бегущей в отрицательном направлении оси Ox , мы же предпочли, чтобы волна распространялась вдоль оси Ox в положительном направлении. Противоположное направление распространения можно учесть выбором знака k или u_n .

Величину k принято называть *волновым числом*. Как можно видеть из формулы (10.2), в данный момент времени структура волны полностью воспроизводится при сдвиге на $\Delta x = 2\pi/k$; величину $\lambda = 2\pi/k$ называют *длиной волны*. (К сожалению, в очередной раз приходится отмечать терминологическую неряшливость физического сообщества: длину волны принято обозначать той же буквой, что и проводимость. В попытке избежать такого пересечения, там, где оно часто происходит, например в физике плазмы и физике твердого тела, проводимость обозначают буквой σ , что, правда, совпадает со стандартным обозначением плотности заряда.)

Выбор синусоидальных решений в качестве собственных функций не случаен. Как известно в математическом анализе, произвольная функция (при достаточно мягких ограничениях) действительно может быть по ним разложена — это называется разложением в ряд или интеграл Фурье. Мы рассмотрим этот вопрос детально в следующем разделе нашего курса, а в данной и последующей главе ограничимся рассмотрением именно синусоидальных решений. Каждое из них, как нетрудно усмотреть из формулы (10.2), характеризуется, помимо произвольной амплитуды и фазы, двумя параметрами: 1) волновым числом k или длиной волны $\lambda = 2\pi/k$ и 2) скоростью распространения волны u . Впрочем, гораздо более принято характеризовать каждое синусоидальное решение (как говорят, гармонику) не скоростью, а произведением ku , которое, как и в случае гармонических колебаний, называется частотой и обозначается буквой ω . Как можно видеть все из той же формулы (10.2), в данной точке пространства $x = \text{const}$ собственная функция в точности восстанавливается через время $\Delta t = 2\pi/\omega$ — отсюда и термин «частота». Таким образом, общий вид гармонической функции,

описывающей бегущую волну, может быть представлен следующим образом:

$$f(x, t) = A \cos(kx - \omega t + \varphi), \quad (10.3)$$

где A — амплитуда, φ — начальная фаза. Вместо косинуса, может фигурировать синус, а также, подобно тому, как описано в гл. 8 для гармонических колебаний, экспонента:

$$f(x, t) = A \exp[i(kx - \omega t + \varphi)]. \quad (10.4)$$

Аргументы для ее использования, а равно и алгоритм — те же, что и в гл. 8. Но, в отличие от (8.13), здесь частотное слагаемое в аргументе экспоненты принято брать со знаком «—». Причины этого расхождения, как можно видеть из всего изложенного, достаточно случайны. Иногда — в оптике или радиофизике — строят теорию на базе решений с противоположным знаком частоты, но в целом это нетипично, и мы будем пользоваться представлением (10.4), которое в целом в физике является общепринятым. При этом знак ω всегда считается положительным, если не оговорено противное, а знак k определяет направление распространения волны: по или против оси Ox . Скорость распространения ω/k обычно называется *фазовой скоростью*, поскольку именно с этой скоростью перемещается в пространстве точка постоянной фазы:

$$\varphi_0 = kx(t) - \omega t = \text{const.}$$

Бездисперсная среда определяется условием $u = \text{const}$ или $\omega \propto k$. Соответственно, зависимость $u(k)$ либо $\omega(k)$ называется *дисперсионным уравнением* или *законом дисперсии*. Мы в данном разделе ограничимся бездисперсными средами или вакуумом, поэтому для нас различие между (10.1) и (10.2) не будет иметь значения, но имея в виду оформление даже предварительных результатов на грамотном языке волновой физики, мы все же отдадим предпочтение решениям (10.4). Настало время определить, решения какого именно уравнения они представляют. Нетрудно сообразить, что это должно быть уравнение в частных производных, потому что все решения (10.1), (10.2), (10.4) суть функции двух переменных. Как мы уже не раз подчеркивали, уравнение будет одним и тем же для самых различных процессов: продольных упругих волн (звука), электромагнитных волн (света) — в чем мы скоро убедимся, поперечных колебаний струны или мембраны и т. п. Его так и принято называть — *волновое уравнение*:

$$\frac{\partial^2 f}{\partial t^2} = \frac{u^2}{\partial x^2} f. \quad (10.5)$$

Читатель может самостоятельно проверить прямой подстановкой, что все функции (10.1)–(10.4) и в самом деле являются его

решениями, при этом для (10.3), (10.4) имеет место тривиальный закон дисперсии:

$$\omega = k u. \quad (10.6)$$

10.2. Плоская монохроматическая волна

Уже к началу XX столетия было надежно установлено на основании многочисленных взаимодополняющих экспериментов, что радиоволны и свет суть явления одной природы — это электромагнитные волны, и различаются они только частотой — или длиной волны — см. (10.6). Физика XX века дополнила этот ряд рентгеновскими и γ -лучами, а также заполнила интервал длин волн 1 см–1 мкм, промежуточный между видимым светом и радиоволнами. Как следствие, радиотехническая и оптическая терминология в физике электромагнитных волн имеют равное употребление, дополняя друг друга.

Электромагнитные волны оптического диапазона формируют — в результате регистрации их сетчаткой и последующей обработки этого сигнала мозгом — наше восприятие цвета. Установлено, что каждому «чистому» цвету спектра соответствует волна определенной частоты. В повседневной практике цвет принято связывать с длиной волны (скажем, желтый — 550 нм, красный — 660 нм и т. д.), но в действительности это лишь привязка частоты к длине волны в вакууме в соответствии с (10.6). В иных прозрачных средах длина волны при данной частоте будет другой, как мы увидим в следующей главе, но восприятие цвета не меняется при наблюдении предмета через стекло или воду (вернее, меняется лишь постольку, поскольку среда оказывается не вполне прозрачной). Поэтому электромагнитную волну любого диапазона с заданной постоянной частотой принято называть монохроматической.

Это понятие естественным образом распространилось на волны любой природы. Если рассматривать самый простой пример волнового движения в реальном трехмерном пространстве, то это, несомненно, будет «одномерная» волна (10.4). При этом подразумевается, что ось Ox направлена как раз вдоль направления распространения волны. Если же мы имеем дело с произвольной системой координат либо с неким ансамблем таких волн, распространяющихся в разных направлениях, эту формулу придется несколько модифицировать. Введем вектор \mathbf{k} , модуль которого равен волновому числу, т. е. $2\pi/\lambda$, а направление совпадает с направлением распространения волны. Тогда слагаемое kx в фазе волны (10.4) можно представить как скалярное произведение $\mathbf{k}\mathbf{r}$, поскольку k_x есть единственная отличная от нуля компонента вектора \mathbf{k} . Если теперь представить себе волну, распространяющуюся в произвольном направлении, то зависимость ее фазы от

координат можно представить точно таким же образом. Действительно, координата вдоль вектора \mathbf{k} (назовем ее, например, ξ), в силу известных правил аналитической геометрии, может быть представлена в виде

$$\xi = x \cos(\angle \mathbf{k}, O x) + y \cos(\angle \mathbf{k}, O y) + z \cos(\angle \mathbf{k}, O z).$$

Соответственно получаем трехмерный аналог волны (10.4):

$$f(\mathbf{r}, t) = A_0 \exp[i(\mathbf{k}\mathbf{r} - \omega t + \varphi)]. \quad (10.7)$$

Мы, однако, не случайно поставили слово «одномерная» в кавычки. Дело в том, что зависимость (10.6) от координаты и в самом деле одномерная — распространение волны определяется единственным вектором \mathbf{k} . Но сама осциллирующая величина может быть не только скалярной, как в (10.6), но и векторной:

$$\mathbf{A}(\mathbf{r}, t) = \mathbf{A}_0 \exp[i(\mathbf{k}\mathbf{r} - \omega t + \varphi)]. \quad (10.8)$$

В любом случае — (10.7) или (10.8) — поверхность постоянной фазы есть перпендикулярная вектору \mathbf{k} плоскость, уравнение которой $\mathbf{k}\mathbf{r} = \text{const}$. Поэтому такая волна и называется *плоской*. Поверхность постоянной фазы перемещается в направлении \mathbf{k} со скоростью $\omega/k = \lambda\nu$, где $\nu = \omega/2\pi$ — линейная частота. В заданной плоскости, определяемой некоторым постоянным значением $\mathbf{k}\mathbf{r}$, физическая величина (\mathbf{A} или f) осциллирует с частотой ω . В принципе любое возмущение, эволюционирующее в рамках волнового уравнения, можно представить в виде ансамбля плоских волн; к этому вопросу мы вернемся в следующем разделе нашего курса.

Прежде, чем обобщать на трехмерную ситуацию само волновое уравнение, рассмотрим действие на функцию плоской волны операций векторного дифференцирования. Получим градиент функции (10.7)

$$\begin{aligned} \nabla f(\mathbf{r}, t) &= A_0 \left(\mathbf{e}_x \frac{\partial}{\partial x} + \mathbf{e}_y \frac{\partial}{\partial y} + \mathbf{e}_z \frac{\partial}{\partial z} \right) e^{i(\mathbf{k}\mathbf{r} - \omega t)} = \\ &= A_0 (i\mathbf{e}_x k_x + i\mathbf{e}_y k_y + i\mathbf{e}_z k_z) e^{i(\mathbf{k}\mathbf{r} - \omega t)} \equiv i\mathbf{k} f(\mathbf{r}, t). \end{aligned}$$

А теперь вычислим дивергенцию функции (10.8)

$$\begin{aligned} \text{div } \mathbf{A}(\mathbf{r}, t) &= \left(\mathbf{e}_x \frac{\partial}{\partial x} + \mathbf{e}_y \frac{\partial}{\partial y} + \mathbf{e}_z \frac{\partial}{\partial z} \right) \mathbf{A}_0 e^{i(\mathbf{k}\mathbf{r} - \omega t)} = \\ &= (k_x A_{0x} + k_y A_{0y} + k_z A_{0z}) e^{i(\mathbf{k}\mathbf{r} - \omega t)} \equiv i\mathbf{k} \cdot \mathbf{A}(\mathbf{r}, t). \end{aligned}$$

Предлагаем читателю убедиться самостоятельно в качестве упражнения в справедливости следующих равенств:

$$\begin{aligned} \text{rot } \mathbf{A}(\mathbf{r}, t) &= i[\mathbf{k}, \mathbf{A}(\mathbf{r}, t)]; \quad \nabla^2 \mathbf{A}(\mathbf{r}, t) = -k^2 \mathbf{A}(\mathbf{r}, t); \\ \nabla^2 f(\mathbf{r}, t) &= -k^2 f(\mathbf{r}, t). \end{aligned}$$

Таким образом, все операции дифференцирования для функций плоской волны могут быть сведены к следующим простым правилам:

$$\frac{\partial}{\partial t} \Rightarrow -i\omega; \quad \nabla \Rightarrow i\mathbf{k}. \quad (10.9)$$

Само же волновое уравнение в трехмерном пространстве должно выглядеть следующим образом:

$$\frac{\partial^2 f}{\partial t^2} = u^2 \nabla^2 f \quad \text{или} \quad \frac{\partial^2 \mathbf{A}}{\partial t^2} = u^2 \nabla^2 \mathbf{A}, \quad (10.10)$$

что и является естественным обобщением (10.5). Закон дисперсии (10.6) при этом остается в силе.

Далее следует важное замечание. Для некоторых волн какие-то физические величины и в самом деле можно описать уравнением (10.5) или первым из уравнений (10.10), например возмущение плотности в звуковой волне. Но никогда не удастся свести волновое движение к возмущению только скалярной величины. Поэтому именно векторное волновое уравнение и представляет собой истинное обобщение одномерной модели (10.5). Для него типичными оказываются два случая.

1. Возмущение векторной величины \mathbf{A} параллельно \mathbf{k} . Такие волны называются *продольными* (пример — звук в газе или жидкости). Из формул (10.9) можно видеть, что в этом случае $\text{rot } \mathbf{A} \equiv 0$.

2. $\mathbf{A} \perp \mathbf{k}$, что соответствует *поперечной* волне. Для нее $\text{div } \mathbf{A} \equiv 0$, что также следует из (10.9). Таков, например, поперечный звук в твердом теле; другой пример — электромагнитные волны, которые мы рассмотрим в следующем параграфе. Для поперечных волн существенна, помимо направления распространения, еще и поляризация, то есть направление вектора \mathbf{A} в плоскости, нормальной к вектору \mathbf{k} . Таким образом, поперечная плоская волна характеризуется, помимо амплитуды и частоты (длины волны), двумя ортами направления: \mathbf{k}/k и \mathbf{A}/A .

10.3. Электромагнитные волны в вакууме. Скорость света

Постановка задачи об электромагнитном поле в вакууме для нас отнюдь не нова. И закон Кулона (1.3), и магнитное поле прямого тока (4.9), и многие другие результаты, представленные в нашем курсе, соответствуют именно полю в вакууме и описываются решениями системы уравнений Максвелла (9.2)–(9.5), в которых, как и должно быть в пустоте, $\rho = 0$; $\mathbf{j} = 0$. Но, как мы уже неоднократно отмечали, важны не только уравнения, но и в не меньшей степени граничные условия: где-то вне области решения, но, безусловно, оказывая на него влияние, располагается

заряд или проводник с током, отсюда и получается — в наших примерах — $E \propto |\mathbf{r} - \mathbf{r}_q|^{-2}$ или $H \propto 1/r$.

А теперь мы рассмотрим пространство абсолютно пустое и совершенно однородное, а решение будем искать такое, в котором ни одна точка пространства не выделена. Конечно, в дальнейшем мы будем с приемлемой точностью переносить полученные результаты на реальные ситуации, а из базовых решений впоследствии (см. следующий раздел) мы сможем выстроить любое реальное волновое возмущение и проследить его эволюцию. Вопрос ставится достаточно принципиальный: мы ведь стартовали с законов Кулона и Био–Савара, связывающих поля с порождающими их источниками. Возможно ли существование электромагнитных полей в пространстве при совершенном отсутствии источников? Оказывается, возможно.

В пустоте систему уравнений (9.2)–(9.5) нетрудно преобразовать таким образом, чтобы в ней фигурировали только действующие поля \mathbf{E} и \mathbf{B} . В отсутствие зарядов теорема Гаусса для электрического и магнитного поля выглядит совершенно одинаково:

$$\operatorname{div} \mathbf{E} = 0; \quad \operatorname{div} \mathbf{B} = 0. \quad (10.11)$$

Связь между электрическим и магнитным полями задается парой уравнений (9.2), (9.3) при $\rho = 0$, $\mathbf{j} = 0$:

$$\operatorname{rot} \mathbf{B} = \varepsilon_0 \mu_0 \frac{\partial \mathbf{E}}{\partial t}; \quad \operatorname{rot} \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}. \quad (10.12)$$

Возьмем от первого из них ротор и воспользуемся векторным тождеством, справедливым для любой вектор-функции $\mathbf{A}(\mathbf{r})$

$$\operatorname{rot}(\operatorname{rot} \mathbf{A}) = [\nabla [\nabla \mathbf{A}]] = \nabla \operatorname{div} \mathbf{A} - \nabla^2 \mathbf{A},$$

аналогичным известной формуле векторной алгебры $\mathbf{b}(\mathbf{ac}) - \mathbf{c}(\mathbf{ab})$. Таким, образом, первое из уравнений (10.12) преобразуется к виду

$$-\nabla^2 \mathbf{B} = \varepsilon_0 \mu_0 \frac{\partial \operatorname{rot} \mathbf{E}}{\partial t},$$

(мы учли, что $\operatorname{div} \mathbf{B} = 0$). Подставляя второе из уравнений (10.12), получаем окончательно

$$\frac{\partial^2 \mathbf{B}}{\partial t^2} = c^2 \nabla^2 \mathbf{B}; \quad c^2 \equiv \frac{1}{\varepsilon_0 \mu_0} \approx 9 \cdot 10^{16} \text{ (м/с)}^2. \quad (10.13)$$

Получилось не что иное, как второе уравнение (10.10), причем величина c имеет смысл скорости распространения электромагнитных волн. Сделаем наоборот: возьмем ротор от второго из уравнений (10.12):

$$\operatorname{rot}(\operatorname{rot} \mathbf{E}) = \nabla \operatorname{div} \mathbf{E} - \nabla^2 \mathbf{E} = \frac{\partial}{\partial t} \operatorname{rot} \mathbf{B}.$$

Учтем, что $\operatorname{div} \mathbf{E} = 0$, и подставим первое уравнение. Получаем

совершенный аналог (10.13):

$$\frac{\partial^2 \mathbf{E}}{\partial t^2} = c^2 \nabla^2 \mathbf{E}. \quad (10.14)$$

Из уравнений (10.13), (10.14) можно сделать вывод, что в отсутствие источников электрическое и магнитное поля неразрывно связаны, подчиняются одним и тем же уравнениям и формируют движение типа бегущей волны (10.8). Мы не случайно обозначили скорость распространения через c . Поскольку электромагнитные волны и свет — одно и то же явление, то и скорость распространения электромагнитных волн в вакууме как раз равна максимально возможной в физике скорости распространения взаимодействия — скорости света

$$c = 2,9979 \dots \cdot 10^8 \text{ м/с}.$$

Эта фундаментальная мировая константа связана с базовыми константами системы СИ соотношением $c = 1/\sqrt{\varepsilon_0 \mu_0}$.

Решения уравнений (10.13), (10.14) уже готовы — и они действительно, как мы и потребовали выше, однородны во всем пространстве. Это решения типа (10.8):

$$\begin{aligned} \mathbf{E}(\mathbf{r}, t) &= \mathbf{E}_0 \exp[i(\mathbf{k}\mathbf{r} - \omega t + \varphi)]; \\ \mathbf{B}(\mathbf{r}, t) &= \mathbf{B}_0 \exp[i(\mathbf{k}\mathbf{r} - \omega t + \varphi)]. \end{aligned} \quad (10.15)$$

Потребовав для них выполнения условий (10.11) вкуче с правилом (10.9), приходим к выводу, что электромагнитные волны по своей геометрии являются поперечными:

$$\mathbf{E}_0 \perp \mathbf{k}, \quad \mathbf{B}_0 \perp \mathbf{k}.$$

Поэтому уравнения (10.13), (10.14) в принципе можно представить четырьмя уравнениями типа (10.5), в которых ось Ox параллельна вектору \mathbf{k} , а в качестве величины f фигурируют компоненты поля E_y, E_z, B_y, B_z в плоскости, ортогональной направлению распространения, т. е. \mathbf{k} .

Поля \mathbf{E} и \mathbf{B} в (10.15) не являются независимыми. Полезно, помимо уже известного нам линейного закона дисперсии (10.6), получить связь между \mathbf{E} и \mathbf{B} , для чего мы воспользуемся уравнениями (10.11), (10.12) и правилом (10.9). Тогда, пара уравнений (10.12) для решений (10.15) переходит в следующую систему алгебраических уравнений:

$$[\mathbf{k} \mathbf{E}] = \omega \mathbf{B}; \quad [\mathbf{k} \mathbf{B}] = -\mu_0 \varepsilon_0 \omega \mathbf{E}. \quad (10.16)$$

Из соотношений (10.16) видно прежде всего, что векторы $\mathbf{E}(\mathbf{r}, t)$ и $\mathbf{B}(\mathbf{r}, t)$ колеблются «в фазе», поскольку все мнимые единицы из этих формул выпали, и остались только алгебраические связи. Из них же можно усмотреть, что векторы $\mathbf{k}, \mathbf{E}, \mathbf{B}$ образуют

правую тройку (подобно координатной системе x, y, z), так что координатную зависимость вместе с ориентацией полей в данный момент времени можно передать рис. 10.1.

Сократим в (10.16) экспоненциальные осциллирующие множители, тогда это будут уравнения для амплитуд \mathbf{E}_0 и \mathbf{B}_0 . Подставим первое уравнение во второе. Результат имеет вид

$$[\mathbf{k} [\mathbf{k} \mathbf{E}_0]] = \mathbf{k} (\mathbf{k} \mathbf{E}_0) - k^2 \mathbf{E}_0 = \\ = -\mu_0 \varepsilon_0 \omega^2 \mathbf{E}_0. \quad (10.17)$$

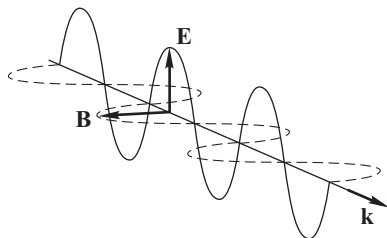


Рис. 10.1

Первый член разности в (10.17) равен нулю в силу поперечности электромагнитной волны. Оставшиеся члены после сокращения \mathbf{E}_0 дают в ответе уже известный нам закон дисперсии

$$\omega^2 = k^2 c^2.$$

Выпадение из ответа амплитуды — в данном случае \mathbf{E}_0 — вообще характерное свойство т. н. линейной физики. Оно происходит из самых основных законов, а именно, из того обстоятельства, что уравнения Максвелла (9.2)–(9.5) полностью сохраняют силу при умножении всех полей, токов и зарядов на один и тот же постоянный коэффициент. Поэтому, если токи и заряды связаны с полями линейной зависимостью или, как в данном случае, вовсе равны нулю, решение (10.15) можно умножать на любые коэффициенты, более того, любая линейная комбинация решений вида (10.15) с разными амплитудами и волновыми векторами (а тем самым, и частотами) также будет решением исходной системы уравнений. Это и есть «линейная физика». Только для этого предельного случая справедлив принцип суперпозиции полей, от которого приходится отказаться в случае нелинейности функций $\rho(\mathbf{E}, \mathbf{B}), \mathbf{j}(\mathbf{E}, \mathbf{B})$ либо операторов $\hat{\varepsilon}, \hat{\mu}$. Нам уже знакомы такие примеры нелинейных сред, как сегнетоэлектрики и ферромагнетики; в этой связи заметим, что нелинейные проблемы в современной физике представляются скорее правилом, нежели исключением. Даже вакуум в квантовой электродинамике оказывается нелинейной средой.

10.4. Энергия и импульс электромагнитной волны

Плотность энергии в электромагнитной волне задается стандартным выражением (9.18)

$$w = \frac{1}{2} \{(\mathbf{E} \mathbf{D}) + (\mathbf{B} \mathbf{H})\}.$$

Это выражение квадратично по полю, поэтому подставлять в него непосредственно формулы (10.15) не следует — нужно воспользоваться их вещественным эквивалентом

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 \cos(\mathbf{k}\mathbf{r} - \omega t + \varphi); \quad \mathbf{B}(\mathbf{r}, t) = \mathbf{B}_0 \cos(\mathbf{k}\mathbf{r} - \omega t + \varphi) \quad (10.18)$$

(обратим еще раз внимание на синфазность электрического и магнитного полей). Плотность энергии волны будет, следовательно, осциллировать в пространстве и во времени. Обычно в таких случаях оперируют средним ее значением по периоду либо по длине волны. Эти величины равны друг другу. Действительно,

$$\langle \cos^2(\mathbf{k}\mathbf{r} - \omega t + \varphi) \rangle_T = \langle \cos^2(\mathbf{k}\mathbf{r} - \omega t + \varphi) \rangle_\lambda = \frac{1}{2}.$$

Таким образом, из (10.18) можно получить следующее значение плотности энергии в электромагнитной волне:

$$\langle w \rangle = \frac{1}{4} \left(\varepsilon_0 E_0^2 + \frac{B_0^2}{\mu_0} \right). \quad (10.19)$$

Воспользуемся первым из уравнений (10.16). Помимо ориентации основной тройки векторов, оно дает, с учетом поперечности волны, соотношение между амплитудами полей

$$B_0 = \frac{k}{\omega} E_0.$$

Подставим его в (10.19). Получаем две эквивалентных формы ответа:

$$\langle w \rangle = \frac{\varepsilon_0}{4} E_0^2 \left(1 + \frac{k^2}{\varepsilon_0 \mu_0 \omega^2} \right) = \frac{B_0^2}{4\mu_0} \left(\frac{\varepsilon_0 \mu_0 \omega^2}{k^2} + 1 \right).$$

В силу закона дисперсии, дополнительное к единице слагаемое в каждой скобке также равно единице. Из этого следует, что средняя по периоду электрическая энергия равна средней магнитной, а поскольку поля синфазны, плотности электрической и магнитной энергии равны друг другу в любой момент времени в любой точке пространства. Окончательный ответ для усредненной плотности энергии мы по-прежнему приводим в двух формах:

$$\langle w \rangle = \frac{\varepsilon_0}{2} E_0^2 = \frac{1}{2\mu_0} B_0^2. \quad (10.20)$$

Первая из них более употребительна, по крайней мере в оптике. Дело в том, что при взаимодействии света с веществом, и в частности при его регистрации, большую роль играет именно электрический вектор. Взаимодействие происходит почти исключительно с электронами. Электрическое поле действует на элек-

трон с известной силой $e\mathbf{E}$, а магнитное — с силой Лоренца, которую можно оценить следующим образом:

$$|e[\mathbf{v} \mathbf{B}]| \sim \frac{kv}{\omega} eE = \frac{v}{c} eE,$$

так что для нерелятивистского электрона ($v \ll c$) она дает лишь малую поправку. Поэтому принято плотность энергии, поляризацию света, сдвиг фазы и т. д. привязывать именно к электрическому полю.

Продолжим исследование энергетических характеристик волны. Вычислим вектор Пойнтинга

$$\mathbf{S} = [\mathbf{E} \mathbf{H}] = \frac{1}{\mu_0 \omega} [\mathbf{E} [\mathbf{k} \mathbf{E}]] = \frac{1}{\mu_0 \omega} (\mathbf{k} E^2 - (\mathbf{k} \mathbf{E}) \mathbf{E}).$$

Для поперечной волны последний член в правой части равен нулю, что позволяет упростить выражение:

$$\mathbf{S} = \frac{\mathbf{k}}{k} \frac{k}{\mu_0 \varepsilon_0 \omega} \varepsilon_0 E^2 = \frac{\mathbf{k}}{k} w c.$$

Поток энергии осциллирует во времени и пространстве, но всегда направлен вдоль вектора \mathbf{k} . Средний по периоду или длине волны поток энергии в направлении \mathbf{k} равен

$$\langle \mathbf{S} \rangle = \frac{\mathbf{k}}{k} \langle w \rangle c. \quad (10.21)$$

Теперь не представит труда определить и среднюю плотность импульса поля в электромагнитной волне, исходя из соотношения (9.28):

$$\left\langle \frac{d\mathbf{p}}{dV} \right\rangle = \frac{\mathbf{S}}{c^2} = \frac{\mathbf{k}}{k} \frac{\langle w \rangle}{c}. \quad (10.22)$$

Его можно трактовать следующим образом: «плотность массы» поля $\langle w \rangle / c^2$ умножается на скорость ее «течения» c . Это соотношение оказалось одним из основополагающих в разработке идеи световых квантов.

Представим себе, что плоская монохроматическая электромагнитная волна падает нормально на абсолютно черную плоскую поверхность, где энергия поля полностью поглощается (рис. 10.2). За промежуток времени t поверхности достигнет фронт (поверхность постоянной фазы), первоначально отстоявший от плоскости на расстоянии ct . За это время поле, заключенное между указанным фронтом и плоскостью, полностью передаст последней свои импульс и энергию. Таким образом, участок черной плоскости единичной площади

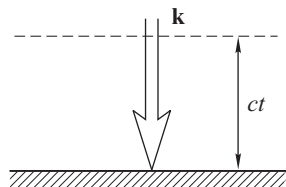


Рис. 10.2

воспримет за это время импульс

$$\frac{d\Delta \mathbf{p}}{dS} = \frac{\mathbf{k}}{k} \frac{\langle w \rangle}{c} \cdot ct.$$

Следовательно, волна, падающая нормально на черную поверхность, оказывает на нее давление

$$\mathcal{P} = \frac{d^2 p_{\perp}}{dS dt} = \langle w \rangle. \quad (10.23)$$

Эффект светового давления был подтвержден экспериментально в 1899 г. выдающимся русским физиком П. Н. Лебедевым (1866–1912). В случае черной облучаемой поверхности оно равно средней плотности энергии в падающей волне, в случае зеркально отражающей поверхности оно должно быть в два раза больше, поскольку отраженная волна уносит тот же по абсолютной величине импульс, но противоположного знака. Эту ситуацию мы рассмотрим подробнее, не ограничиваясь случаем нормального падения.

Идеальное зеркало — это идеально проводящая поверхность. Любые зеркала, используемые в обиходе и в технике, действуют именно по принципу отражения электромагнитной волны от проводящей поверхности. Мы не будем подробно разбирать всю относящуюся к делу электродинамику, ограничившись классической модельной задачей — отражением плоской монохроматической волны при косом падении на идеально проводящую плоскость в случае двух различных поляризаций (рис. 10.3 а, б). Угол φ между вектором \mathbf{k} и нормалью к плоскости принято называть *углом падения*.

Сам факт появления отраженной волны отнюдь не постулируется, но с необходимостью следует из граничных условий. Внутри идеального проводника должно быть $\mathbf{E} \equiv 0$. Значит, на границе, т. е. при $z = 0$, должно быть $\mathbf{E}_{\tau} \equiv 0$. Поскольку это заведомо не так в падающей волне (которой мы присвоим индекс i), приходится ввести еще одну волну — отраженную (индекс r):

$$\mathbf{E}_i = \mathbf{E}_{i0} \exp[i(\mathbf{k}\mathbf{r} - \omega t)]; \quad \mathbf{E}_r = \mathbf{E}_{r0} \exp[i(\mathbf{k}_1\mathbf{r} - \omega_1 t)]. \quad (10.24)$$

Если бы частота ω_1 была отлична от ω , мы не смогли бы выполнить граничное условие нулевого поля в любой момент времени. Поэтому частота отраженной волны, а значит, и волновое число $k = \omega/c$ должны совпадать с таковыми у падающей волны. Мы опустили в формулах (10.24) начальную фазу, поскольку у одной

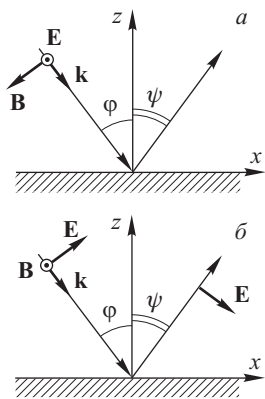


Рис. 10.3

волны она устраняется выбором начала отсчета времени, а у другой возможный фазовый сдвиг будет учтен знаком амплитуды.

Рассмотрим геометрию задачи, представленную на рис. 10.3а. При такой поляризации вектор электрического поля параллелен проводящей плоскости, вектор магнитного поля лежит в плоскости падения. Условие $\mathbf{E} \equiv 0$ в начале координат $\mathbf{r} = 0$ в терминах формул (10.24) означает просто

$$\mathbf{E}_{r0} = -\mathbf{E}_{i0},$$

или, что то же, фазовый сдвиг в π при равных амплитудах — обычно говорят: «потеря полуволны». Далее будем перемещаться вдоль границы, меняя координату x . Условие $\mathbf{E} \equiv 0$ должно быть выполнено в любой точке границы, но зависимость от x полей \mathbf{E}_i , \mathbf{E}_r при одинаковых волновых числах и $z = y = 0$ будет различной: $\mathbf{k}\mathbf{r} = kx \sin \varphi$; $\mathbf{k}_1\mathbf{r} = kx \sin \psi$, где ψ — угол отражения на рис 10.3а. Единственная возможность надлежащим образом выполнить граничное условие на проводящей поверхности обеспечивается известным школьным правилом: угол падения равен углу отражения, $\varphi = \psi$.

Рассмотрим случай другой поляризации, представленный на рис. 10.3б. Теперь тангенциальные компоненты электрического поля равны $E_{i\tau} = E_{i0} \cos \varphi$; $E_{r\tau} = E_{r0} \cos \psi$. Условие сохранения частоты при отражении универсально; обеспечив нулевое касательное поле в одной точке и требуя сохранения этого нуля в любой точке проводящей плоскости, получаем снова $\varphi = \psi$. Возвращаясь к выражениям для тангенциальных компонент поля, еще раз получаем $\mathbf{E}_{r0} = -\mathbf{E}_{i0}$, т. е. и в такой геометрии отражение происходит с потерей полуволны. Резюмируя, можно сказать, что мы построили количественную теорию идеального зеркала.

Важное замечание: граничное условие (9.12) — именно для тангенциальной компоненты \mathbf{E} — выбрано нами не случайно. Анализируя полученные нами условия, читатель может самостоятельно убедиться, что они не могут обеспечить ни непрерывности \mathbf{H}_τ , ни непрерывности \mathbf{D}_n . Но на поверхности проводника полем волны индуцируются поверхностные токи и заряды, которые и балансируют соответствующие скачки — см. условия (9.14), (9.15) в расширенной форме. Условие (9.12) является более жестким — никакая реакция среды не может его подправить.

10.5. Излучение электромагнитных волн

Самый популярный излучатель в школьной физике — точечный источник света. Попробуем понять, может ли действительно излучатель электромагнитных волн обладать такими свойствами. При этом необходимо особо подчеркнуть, что дело не только

в малости генератора излучения. Одно из важнейших свойств точечного источника — отсутствие какого-либо выделенного направления и, следовательно, абсолютная изотропия излучения.

Таким образом, подобный источник может порождать только сферическую волну. Но сферической электромагнитной волны просто не может быть. Это нетрудно понять, вернувшись к мысленному эксперименту, представленному на рис. 9.1. В сферически симметричной конфигурации электрическое поле \mathbf{E} может быть направлено только по радиусу, а магнитное поле должно быть тождественно равно нулю. Это и неудивительно, поскольку любые перемещения в сферически симметричной зарядовой системе, сохраняющие ее симметрию, оставляют поле \mathbf{E} вне ее неизменным в силу теоремы Гаусса. Следовательно, вне такой системы $\dot{\mathbf{E}} = 0$ и $\mathbf{j} = 0$, так что и $\mathbf{H} = 0$. Но тогда тождественно равен нулю и вектор Пойнтинга, так что об электромагнитной волне вообще говорить не приходится.

Это не означает, что модель точечного источника в оптике является ошибочной. Просто во многих задачах можно рассматривать поле малого излучателя усредненным во времени. Будучи в каждый момент асимметричным, оно за характерное время регистрации может в среднем оказаться симметричным в достаточной степени, чтобы мы считали излучение изотропным, а излучатель — истинно точечным. При этом, правда, постановка вопроса о направлении электрического и магнитного поля излучения теряет смысл.

Что же касается стандартного утверждения типа «линза преобразует плоскую волну в сферическую, сходящуюся в фокусе», то оно верно лишь как некоторое приближение, в котором малым параметром служит телесный угол Ω , охватывающий сходящийся пучок света. Истинная волна, сформированная линзой, мало отличается от сферической, пока и поскольку $\Omega \ll 1$. Во всяком случае, из уравнений Максвелла следует, что ни сферической, ни цилиндрической электромагнитной волны в точном смысле не существует, а значит, не может быть слишком симметричным излучатель.

Оказывается, самым простым и фундаментальным типом излучателя является электрический диполь с переменным во времени дипольным моментом. Формальный уровень нашего курса не дает возможности изложить этот вопрос с должной степенью строгости, но мы можем сопроводить это утверждение достаточно убедительными аргументами.

Обратимся к уравнениям (10.10). Заметим, что первое из них в принципе можно рассматривать как часть второго, если под f понимать любую компоненту вектора \mathbf{A} . Вернемся к сферически симметричной модели и поищем соответствующее решение урав-

ний (10.10). Оператор Лапласа применительно к скалярной величине f при условии, что последняя зависит только от радиуса r , но не от полярного или азимутального угла, может быть записан в простой форме, которую читатель может получить в виде упражнения, перейдя от декартовых координат к полярным:

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} = \frac{1}{r^2} \frac{\partial}{\partial r} r^2 \frac{\partial f}{\partial r} = \frac{2}{r} \frac{\partial f}{\partial r} + \frac{\partial^2 f}{\partial r^2}.$$

Подставим его в первое уравнение (10.10), после чего умножим обе части уравнения на r . Результат

$$r \frac{\partial^2 f}{\partial r^2} + 2 \frac{\partial f}{\partial r} = \frac{1}{u^2} \frac{\partial^2 (rf)}{\partial t^2}$$

можно преобразовать к виду

$$\frac{\partial^2 g}{\partial r^2} = \frac{1}{u^2} \frac{\partial^2 g}{\partial t^2}, \quad g \equiv rf.$$

Полученное уравнение формально одномерное, подобно (10.5), а следовательно, имеет решение

$$g = rf = g \left(t - \frac{r}{u} \right). \quad (10.25)$$

Решение (10.25) представляет, во-первых, пример истинно сферической волны (но она существует лишь для скалярной величины или потенциального вектора)

$$f(r, t) = \frac{f(t - r/u)}{r},$$

а во-вторых, если рассматривать f еще и как компоненту вектора, моделирует так называемые запаздывающие потенциалы. Если, например, речь идет об электромагнитном поле, то потенциалы \mathbf{A} и φ , созданные соответственно токами и зарядами некоторой ограниченной в пространстве системы, устанавливаются на далеких от нее расстояниях с задержкой во времени r/c .

Представим себе дипольную систему, которую мы на больших расстояниях можем описать просто дипольным моментом $\mathbf{p}(\mathbf{r}, t)$. Для удобства описания излучения волн немецкий физик Генрих Герц (1857–1894) ввел специальный вектор (его так и называют — *вектор Герца*)

$$\mathbf{P}(\mathbf{r}, t) = \frac{\mathbf{p}(t - r/c)}{r},$$

который, как следует из (10.25), является решением волнового уравнения при $u = c$, т. е. представляет динамику типа электромагнитной волны. Если мы заставим диполь осциллировать во времени, $\mathbf{p} = \mathbf{p}_0 \cos \omega t$, то в достаточно удаленной от диполя области пространства эти осцилляции будут регистрироваться именно как бегущая электромагнитная волна.

Далее проведем мысленный эксперимент. Пусть наш диполь движется с постоянной скоростью. Тогда он, очевидно, излучать не будет; поле его будет просто стационарным полем в движущейся системе отсчета. Значит, поле излучения определяет не скорость, а, по крайней мере, следующая временная производная, т. е. ускорение, пропорциональное $\omega^2 \mathbf{p}$. Точный расчет подтверждает такое заключение, а полная мощность излучения колеблющегося диполя дается выражением

$$\dot{W} = \frac{1}{12\pi\epsilon_0} \frac{p_0^2 \omega^4}{c^3}. \quad (10.26)$$

Это излучение обладает заметной анизотропией. Среднее за период значение вектора Пойнтинга оказывается пропорциональным $\sin^2 \theta$, где θ — полярный угол, который отсчитывается от направления дипольного момента (рис. 10.4).

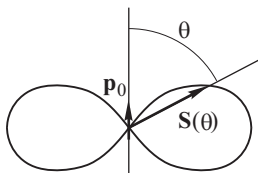


Рис. 10.4

Из всего вышеизложенного следует, в частности, принципиальная трудность резерфордской планетарной модели атома. Если бы электрон просто вращался по орбите вокруг ядра, атом в принципе существовал бы конечное время — пока вся начальная кинетическая энергия не ушла бы с излучением. Тогда произошло бы падение электрона на ядро. Существенно, что в этом случае свойства отдельных атомов не были бы идентичны, а зависели бы от возраста. Это соображение послужило главным толчком к созданию атомной модели Бора, а далее — и всей квантовой механики.

Формула (10.26), а равно и рис. 10.4, подразумевают, что размер диполя l меньше длины волны λ : $l \ll c/\omega$. Такой диполь называют *элементарным диполем* или *диполем Герца*. Излучение может быть надежно идентифицировано в так называемой волновой зоне — при $r \gg c/\omega$. В технических целях используются обычно другие излучатели — антенны, у которых характерный размер, с тем или иным коэффициентом порядка единицы, соответствует как раз длине волны: $l \sim 2\pi c/\omega$. Если излучение элементарного диполя может быть вызвано различными причинами, включая ускорение частиц, то в случае антенны это всегда результат подачи на диполь мощного внешнего сигнала.

Электродипольное излучение — наиболее типичный механизм генерации электромагнитных волн и в природе, и в технике, но в принципе существуют и другие, связанные с изменением во времени магнитодипольного момента, электрического и магнитного квадрупольных моментов и т. д.

Задачи

1. Плоская монохроматическая электромагнитная волна частоты ω падает нормально на плоскую гладкую поверхность проводника (см. рис. 10.2). Проводимость материала λ , магнитная проницаемость $\mu \approx 1$. Оценить коэффициент отражения по мощности и амплитуде.

Решение. Отличие коэффициента отражения от единицы обусловлено джоулевой диссипацией в скин-слое, т. е. на глубине вплоть до $\delta_{\text{эф}}$, которая дается формулой (8.28):

$$\delta_{\text{эф}} \approx \sqrt{\frac{2\pi}{\omega \lambda \mu_0}}.$$

Удобно вычислять связанные с этим потери, относя их к единице площади поверхности зеркала, а затем получить коэффициент отражения, нормируя эту диссипацию на поток энергии. Итак,

$$\frac{d^2 W}{dS dt} \approx \frac{j^2}{\lambda} \delta_{\text{эф}} = \lambda E^2 \sqrt{\frac{2\pi}{\omega \lambda \mu_0}}.$$

Модуль вектора Пойнтинга, равный потоку энергии на поверхность зеркала, можно представить в виде

$$|S| = \frac{k}{\mu_0 \omega} E^2 = \frac{E^2}{c \mu_0}.$$

Отнеся интенсивность потерь к потоку энергии, получим коэффициент поглощения по энергии; вычитая последний из единицы, получим коэффициент отражения по мощности R . (К сожалению, это совершенно стандартное обозначение пересекается с принятым в электротехнике символом омического сопротивления). Таким образом,

$$R \approx 1 - \sqrt{\frac{2\pi \lambda}{\mu_0 \omega}} \cdot c \mu_0 = 1 - c \sqrt{\frac{2\pi \lambda \mu_0}{\omega}}.$$

Поскольку отраженная волна представляет собой электромагнитную волну той же частоты, что и падающая, и с тем же соотношением между \mathbf{E} и \mathbf{H} , амплитудный коэффициент отражения равен просто квадратному корню из коэффициента отражения по мощности: $r = \sqrt{R}$. Надо заметить, что использованные нами соображения корректны лишь в случае радиоволн достаточно низкой частоты. В оптическом и даже инфракрасном диапазоне частот механизмы отражения и потерь будут несколько иными. Мы еще коснемся этого вопроса в заключительной главе.

Точность полученных нами результатов не слишком велика — именно потому, что не слишком точно оцененную величину приходится вычитать из единицы. Гораздо более надежна оценка коэффициента поглощения по мощности

$$1 - R = c \sqrt{\frac{2\pi \lambda \mu_0}{\omega}},$$

которая позволяет правильно оценить порядок величины потерь при отражении. Качественным ответом задачи будет либо малость этого коэффициента в сравнении с единицей, либо соизмеримость.

2. Электрон совершает циклотронное вращение в однородном магнитном поле \mathbf{B} . Получить зависимость его энергии от времени и оценить, сколько оборотов он сделает до остановки.

Решение. Мы неоднократно подчеркивали гироскопический характер силы Лоренца, в силу которого она не совершает работы над заряженной частицей. Но при движении частицы по круговой орбите (а это движение с ускорением) происходит, в соответствии с (10.26), излучение электромагнитных волн, оно-то и тормозит частицу. Мы сделаем лишь одно упрощающее предположение: пусть энергия, потерянная на излучение за один оборот, будет много меньше кинетической энергии электрона. Необходимое для этого условие последует в конце решения.

Дипольный момент единственной частицы равен (см. (1.23)): $\mathbf{p} = -e\mathbf{r}$. Удобно поместить начало координат в центре циклотронной орбиты, тогда $p_0 = ev/\omega_B$, где v — скорость электрона на орбите, а $\omega_B = eB/m$ — циклотронная частота. Поскольку диполь в данном случае вращается, а не осциллирует, интенсивность потерь (10.26) нужно удвоить (складывая $\omega^4 \langle p_x^2 \rangle$ и $\omega^4 \langle p_y^2 \rangle$):

$$\dot{W} = \frac{1}{6\pi\epsilon_0} \frac{p_0^2 \omega_B^4}{c^3} = \frac{e^2 v^2 \omega_B^2}{6\pi\epsilon_0 c^3} = \frac{W}{3\pi} \frac{e^2 \omega_B^2}{m\epsilon_0 c^3},$$

где W — кинетическая энергия. Составляя уравнение $dW/dt = -\dot{W}$, находим, что кинетическая энергия электрона должна убывать по экспоненциальному закону

$$W \propto \exp\left(-\frac{1}{3\pi} \frac{e^2 \omega_B^2}{m\epsilon_0 c^3} t\right).$$

Как обычно при экспоненциальном затухании, характерное время остановки не зависит от начальной энергии — по смыслу это масштаб времени, по истечении которого энергия становится существенно меньше начальной. Это время можно выразить через характерное число оборотов:

$$\tau = 3\pi \frac{m^2 \epsilon_0 c^3}{e^2 \omega_B^2} \equiv \frac{N}{\omega_B}.$$

Отсюда и получаем ответ

$$N = 3\pi \frac{m^2 \epsilon_0 c^3}{e^3 B}.$$

Наше предположение о малости потерь за период эквивалентно сильному неравенству $N \gg 1$, что приводит к следующему условию на поле:

$$B \ll 3\pi m^2 \epsilon_0 c^3 e^{-3} \sim 10^{12} \text{ Тл}$$

— оно с запасом выполнено при всех доступных значениях поля. Соответственно, с большим запасом выполнено условие $N \gg 1$, так что при рассмотрении не слишком большого числа оборотов частицы в магнитном поле мы вполне можем пользоваться приближением стационарной орбиты (как это до сих пор и делалось).

3. Плоская монохроматическая электромагнитная волна падает нормально на отражающую поверхность, частично поглощается, а частично отражается. В пространстве перед зеркалом образуется суперпозиция падающей и отраженной волн, причем т. н. коэффициент стоячей волны — отношение амплитуды в пучности к таковой в узле — равен $n = 10$. Определить коэффициент отражения по мощности.

Ответ: $R = \left(\frac{n-1}{n+1}\right)^2 \approx 0,67.$

Глава 11

ЭЛЕКТРОМАГНИТНЫЕ ВОЛНЫ В ПРОЗРАЧНЫХ СРЕДАХ

11.1. Распространение волн в сплошной среде. Показатель преломления

Предположим, что мы поставили задачу об электромагнитном поле в отсутствие источников, но не в вакууме, а в некоторой линейной среде, которую можно охарактеризовать константами ε , μ . Конечно, безграничность такой среды еще более проблематична, нежели безграничность пустого пространства в предыдущей главе. Но как всегда в физических задачах, такую идеализацию следует понимать в смысле некоторого сильного неравенства. В данном случае условие применимости следует из ответа: характерный размер образца диэлектрика или характерный масштаб его неоднородности должен многократно превосходить длину волны.

Вместо (10.12) получаем систему уравнений:

$$\operatorname{rot} \mathbf{B} = \varepsilon \varepsilon_0 \mu \mu_0 \frac{\partial \mathbf{E}}{\partial t}; \quad \operatorname{rot} \mathbf{E} = - \frac{\partial \mathbf{B}}{\partial t}. \quad (11.1)$$

Произошла, как мы видим, совершенно незначительная модификация — всего лишь в одном коэффициенте. Соответственно, и (10.13), (10.14) изменяются лишь в одном коэффициенте:

$$\frac{\partial^2 \mathbf{B}}{\partial t^2} = \frac{c^2}{\varepsilon \mu} \nabla^2 \mathbf{B}; \quad \frac{\partial^2 \mathbf{E}}{\partial t^2} = \frac{c^2}{\varepsilon \mu} \nabla^2 \mathbf{E}. \quad (11.2)$$

Итак, мы опять пришли к волновому уравнению, и его решением снова будет монохроматическая плоская волна (10.15) при том же законе дисперсии (10.6), но скорость распространения волн будет меньше, чем в вакууме:

$$\frac{\omega}{k} = \frac{c}{n}, \quad \text{где} \quad n = \sqrt{\varepsilon \mu}. \quad (11.3)$$

Величина n называется *коэффициентом преломления* — как мы вскоре убедимся, она как раз соответствует коэффициенту преломления в оптике. Природа устроила так, что во всех без исключения средах, прозрачных в оптическом диапазоне, величина μ с хорошей точностью равна единице, поэтому в оптике обычно без оговорок полагают $n = \sqrt{\varepsilon}$.

Не следует, однако, отождествлять диэлектрическую проницаемость, которую мы вводили в электростатике, с константой ϵ в уравнениях (11.1)–(11.3). Это одно и то же по физическому смыслу, но не по величине. Известно, что у чистой воды в статике $\epsilon = 81$. Казалось бы, отсюда следует $n = 9$, но в действительности показатель преломления воды равен примерно 1,33 в красной области спектра и 1,34 — в фиолетовой.

Мы с детства любимся радугой, существование которой обусловлено именно этой небольшой разницей. Такую же «радугу» можно наблюдать при преломлении света в стеклянных призмах — даже просто на скошенном краю толстого зеркала. Эти бытовые наблюдения свидетельствуют о фундаментальном факте зависимости показателя преломления от частоты. Данное свойство называется дисперсией; мы его уже касались в связи с соотношением (10.6). Не следует преувеличивать полноты информации, представленной нашим повседневным опытом. Порядок цветов в радуге любой природы таков, что из него следует рост n с частотой. Но во-первых, как раз при $\omega = 0$ величина $\sqrt{\epsilon}$ оказывается почти в 7 раз больше, чем в оптике, т. е. при $\omega \sim 10^{15} \text{ с}^{-1}$. А во-вторых, для электромагнитных волн существенно более высокой частоты — жестких рентгеновских лучей — коэффициент преломления любого вещества оказывается очень близок к единице. Дисперсия электромагнитных волн — явление сложное, функция $n(\omega)$ немонотонна и включает области непрозрачности.

Но и тот простой случай, который представляют уравнения (11.1)–(11.3), реализуется во многих ситуациях, представляющих практический интерес, а потому стоит изучения. Приведем некоторые соотношения, демонстрирующие отличие волн в прозрачной среде с постоянным коэффициентом преломления, от волн в вакууме.

Как уже было отмечено в гл. 10, восприятие цвета обусловлено частотой. Мы знаем, что в стационарной оптической системе чистый спектральный цвет не меняется при переходе из одной среды в другую. (Это не относится к нестационарным системам, включающим, например, движущиеся зеркала или среды с переменным во времени коэффициентом преломления, а также к нелинейным средам.) Факт сохранения частоты в линейной стационарной системе — по сути, закон сохранения энергии для световых квантов. Не касаясь этой стороны дела, мы его примем в качестве экспериментального. Как следствие, разная фазовая скорость волн в разных средах обусловлена зависимостью от n волнового числа или, что то же, длины волны:

$$k(n) = nk_0; \quad \lambda(n) = \frac{\lambda_0}{n}, \quad (11.4)$$

где $k_0 = 2\pi/\lambda_0 = \omega/c$ — характеристики волны данной частоты в вакууме. Из соотношений (11.4) естественно следует определение относительного коэффициента преломления двух сред:

$$n_{12} = \frac{k(n_1)}{k(n_2)} = \frac{n_1}{n_2}. \quad (11.5)$$

И это определение соответствует стандартному, принятому в оптике. Получим далее соотношения между величинами электрического и магнитного полей в волне, аналогичные (10.16). Подставим решения вида (10.15) в уравнения (11.1) и постараемся записать ответ в форме, как можно более симметричной:

$$B = \frac{n}{c} E = \sqrt{\varepsilon\varepsilon_0\mu\mu_0} E; \quad H = \sqrt{\frac{\varepsilon\varepsilon_0}{\mu\mu_0}} E. \quad (11.6)$$

Отсюда следует, что плотность энергии электромагнитной волны может быть представлена в следующем виде:

$$w = \frac{\varepsilon\varepsilon_0}{2} \langle E^2 \rangle + \frac{\mu\mu_0}{2} \langle H^2 \rangle = \varepsilon\varepsilon_0 \langle E^2 \rangle = \frac{1}{\mu\mu_0} \langle B^2 \rangle. \quad (11.7)$$

Представляется достаточно интересным тот факт, что электрическая и магнитная энергия поперечной волны остаются, как и в вакууме, равны друг другу. Исходя из (11.6), мы можем вычислить и вектор Пойнтинга

$$\langle \mathbf{S} \rangle = \frac{\mathbf{k}}{k} \sqrt{\frac{\varepsilon\varepsilon_0}{\mu\mu_0}} E^2. \quad (11.8)$$

И, наконец, плотность импульса электромагнитной волны в прозрачной среде оказывается равной

$$\frac{d\mathbf{p}}{dV} = \frac{\mathbf{S}}{c^2} = \frac{\mathbf{k}}{k} \sqrt{\frac{\varepsilon}{\mu}} \varepsilon_0^{3/2} \mu_0^{1/2} E^2. \quad (11.9)$$

Ради общности мы всюду сохранили величину $\mu \neq 1$, хотя, как уже говорилось, для прозрачных сред магнитная проницаемость, как правило, несущественна.

11.2. Отражение и преломление на плоской границе

Рассмотрим задачу о прохождении электромагнитной волны через плоскую границу двух диэлектрических сред I и II (одной из них может, в частности, быть и вакуум). Эта ситуация представлена на рис. 11.1 *a, б* для случая двух различных поляризаций. Направление электрического и магнитного векторов соответствует правилу, согласно которому \mathbf{k} , \mathbf{E} , \mathbf{B} образуют правую тройку. Мы использовали в качестве магнитного вектора \mathbf{H} , поскольку именно для него будем писать соответствующее граничное условие. Заметим, что магнитное поле ориентировано

на рис. 11.1а так же, как электрическое поле на рис. 11.1б (с точностью до знака), а на рис. 11.1б — ортогонально плоскости рисунка. Результаты, которые мы получим в данном параграфе, можно переносить и на случай искривленной поверхности раздела. Она лишь должна быть гладкой, а радиус кривизны ее

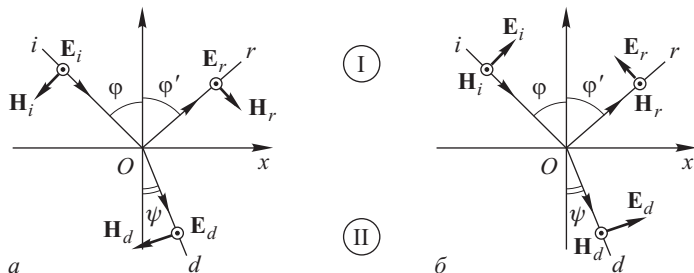


Рис. 11.1

должен многократно превосходить характерный пространственный масштаб электромагнитного поля — длину волны. Таким образом мы опишем действие на электромагнитную волну, в частности, поверхности линзы. Все расчеты линз, тонких и толстых, а также и сложных оптических систем базируются именно на законе преломления. Мы будем пользоваться не понятием луча, принятым в геометрической оптике, а более корректным с точки зрения электродинамики понятием волнового фронта; «лучи» — падающий, отраженный и преломленный — изображенные на рис. 11.1, представляют в наших терминах нормали к волновому фронту, направление которых задается вектором \mathbf{k} .

Пусть показатели преломления сред I и II равны, соответственно, n_1 , n_2 (в вакууме — просто единице). Мы показали на каждом из рис. 11.1 три волны: падающую (индекс i), отраженную (индекс r), и преломленную (индекс d). Это экспериментальный факт, известный каждому школьнику, но даже и не зная этого заранее, можно было его предсказать, исходя из уравнений Максвелла. В случае падения электромагнитной волны на проводящую поверхность (см. § 10.4) имеет место единственное нетривиальное граничное условие — закон сохранения тангенциальной компоненты электрического поля (9.12), а все остальные выполнялись должным образом за счет зарядов и токов, индуцированных на поверхности проводника. Теперь такой возможности уже нет, поскольку мы имеем дело с диэлектрическими средами, а поэтому всего лишь одной волны, помимо падающей, нам просто не хватит. В каждой из двух ситуаций на рис. 11.1 нам придется выполнить условия непрерывности \mathbf{E}_τ и \mathbf{H}_τ . Третье условие — сохранение \mathbf{B}_n в случае, изображенном

на рис. 11.1*а*, и \mathbf{D}_n в случае рис. 11.1*б* — будет выполнено автоматически как следствие закона преломления.

Еще один вопрос, на который целесообразно ответить заранее: правомерно ли разделение постановки задачи именно на те два случая, которые представлены на рис. 11.1*а*, *б*? Не могут ли возникнуть отраженные либо преломленные волны с поляризацией, ортогональной таковой в падающей волне? Ответ: не могут, и это прямое следствие уравнений Максвелла и граничных условий. В силу линейности задачи, мы можем расщепить решение уравнений Максвелла на два линейно независимых, соответствующих двум различным поляризациям. Выбирая решение с одной из поляризаций — той же, что и у падающей волны, мы оперируем с полями трех волн, что позволяет выполнить граничные условия (9.12) и (9.15). При попытке выполнить их для другой поляризации нам опять «не хватит» переменных, т. к. в нашем распоряжении будут только две волны, без падающей, так что единственным возможным решением с такой поляризацией окажется нулевое поле. Разумеется, эти рассуждения находятся в полном соответствии с экспериментальными данными.

Пусть все три волны записаны в виде (10.15). Очевидно, для линейной среды зависимость $\mathbf{H}(\mathbf{r}, t)$ или $\mathbf{V}(\mathbf{r}, t)$ будет иметь точно такой же вид. Воспользуемся для случая *а* граничным условием (9.12)

$$\mathbf{E}_I = \mathbf{E}_{II} \implies E_i + E_r = E_d,$$

а для случая *б* — (9.15)

$$\mathbf{H}_I = \mathbf{H}_{II} \implies H_i + H_r = H_d$$

(мы учли, что в обоих случаях в соотношение входят параллельные векторы). Поскольку дальнейшие действия для обеих поляризаций совершенно идентичны, мы ограничимся случаем рис. 11.1*а*. Пусть в какой-то момент времени граничное условие выполнено. Однако оно сразу же нарушится, если зависимость от времени не будет одинаковой для всех трех полей. Это означает, что частота всех трех волн должна быть одинаковой (и действительно, отражение от прозрачной среды и преломление в ней «сохраняют цвет»). Далее введем в плоскости падения вдоль границы сред координату x . Из поперечности волн и параллельности векторов \mathbf{E}_i , \mathbf{E}_r , \mathbf{E}_d следует, что все три волновых вектора \mathbf{k}_i , \mathbf{k}_r , \mathbf{k}_d лежат в одной плоскости — плоскости падения. Вдоль оси Ox произведения $\mathbf{k}\mathbf{r}$ в формулах типа (10.15) вырождаются в $k_x \cdot x$. Таким образом, граничное условие при равных частотах сводится к следующему:

$$E_{i0} \exp(ikx \sin \varphi) + E_{r0} \exp(ikx \sin \varphi') = E_{d0} \exp(ik'x \sin \psi).$$

Мы воспользовались обозначениями углов рис. 11.1 и учли, что из равенства частот для падающей и отраженной волн следует

равенство волновых чисел. Для преломленной волны волновое число k' определяется формулой (11.5). Теперь потребуем, чтобы наше граничное условие выполнялось в любой точке оси Ox . Для этого необходимо, чтобы экспоненциальные множители были тождественно равны друг другу, а значит, равны должны быть их аргументы:

$$k \sin \varphi = k \sin \varphi' = k(n_{II}/n_I) \sin \psi.$$

Мы получили аналитически из законов электродинамики хорошо известные правила вычисления углов отражения и преломления:

$$\varphi = \varphi'; \quad \frac{\sin \varphi}{\sin \psi} = n = \frac{n_{II}}{n_I}. \quad (11.10)$$

Соотношения (11.10), в частности, подтверждают, что введенный нами коэффициент преломления $n = \sqrt{\varepsilon\mu}$ и в самом деле имеет право так называться. Далее из (11.10) можно извлекать следствия, хорошо известные из школьного курса физики: эффект полного внутреннего отражения, отклонение световой волны призмой, фокусное расстояние тонкой линзы и т. д. Однако решение задачи о прохождении волной границы двух сред не исчерпывается этим результатом. Те же самые граничные условия (9.12)–(9.15) позволяют нам продвинуться существенно дальше — определить коэффициенты отражения и прозрачности на границе раздела двух диэлектриков. Впервые это было сделано О. Френелем (более чем за десять лет до рождения Максвелла).

11.3. Формулы Френеля

Предметом нашего рассмотрения по-прежнему остаются задачи, представленные на рис. 11.1а, б. Для упрощения последующих выкладок положим для обеих сред $\mu = 1$; по сути, все сколько-нибудь интересные ситуации укладываются в рамки такого ограничения. Далее, мы проведем вычисления для границы диэлектрика с вакуумом: $n_I = 1$, $n_{II} \equiv n$, а ответ представим в такой форме, что его можно будет отнести к границе сред с произвольными коэффициентами преломления.

В дальнейшем, выписывая граничные условия, мы сразу будем сокращать экспоненциальные осциллирующие множители, поскольку равенство частот и формулы (11.10) гарантируют нам такую возможность. Теперь выполнение граничных условий потребует лишь определенного соотношения между амплитудами волн. В этой связи заметим, что направление векторов **Е** на рис. 11.1а и **Н** на рис. 11.1б выбрано нами из соображений как можно большей симметрии получающихся уравнений —

в принципе для одной из вторичных волн — отраженной либо преломленной — ответ может включать знак «—», что, конечно, должно получиться совершенно автоматически.

Введем амплитудные коэффициенты отражения r и прозрачности t следующим образом:

$$E_{r0} = r E_{i0}; \quad E_{d0} = t E_{i0}. \quad (11.11)$$

Мы ввели их через отношение амплитуд электрического поля в духе сделанного выше замечания о том, что при регистрации электромагнитных волн происходит измерение именно электрического поля. Если вычислительная процедура приведет к тому, что один из них окажется отрицательным, это будет означать просто фазовый сдвиг на π . В наших предположениях $H = n(\varepsilon_0/\mu_0)^{1/2}E$, при этом имеет смысл удерживать только фактор n , все прочее можно будет в наших уравнениях сократить. Итак, для конкретной геометрии рис. 11.1 *a* (электрическое поле ортогонально плоскости падения) граничные условия (9.12), (9.15) преобразуются к виду

$$[\mathbf{E}_\tau] = 0 \implies E_i + E_r = E_d;$$

$$\begin{aligned} [\mathbf{H}_\tau] = 0 \implies -H_i \cos \varphi + H_r \cos \varphi &= -H_d \cos \psi \implies \\ \implies -E_i \cos \varphi + E_r \cos \varphi &= -n E_d \cos \psi. \end{aligned}$$

Эту систему уравнений с помощью введенных нами коэффициентов можно свести к следующей:

$$1 + r = t; \quad (1 - r) \cos \varphi = t \frac{\sin \varphi}{\sin \psi} \cos \psi. \quad (11.12)$$

Как и было обещано, система уравнений переписана в универсальном виде, в ней фигурируют только функции углов падения и преломления. Решение (11.12) не представляет труда. Мы приводим сразу же ответ:

$$r_\perp = -\frac{\sin(\varphi - \psi)}{\sin(\varphi + \psi)}; \quad t_\perp = \frac{2 \sin \psi \cos \varphi}{\sin(\varphi + \psi)}. \quad (11.13)$$

Здесь индекс \perp означает поляризацию электрического вектора нормально к плоскости падения. Как мы и предвидели, высокая симметрия рис. 11.1 *a* не соблюдена — отражение происходит с потерей популяций.

Надо заметить, что граничных условий в общей сложности четыре. Но условие (9.14) в данном случае вырождено, поскольку электрический вектор не имеет составляющей, нормальной к поверхности раздела. Обратимся к условию (9.13), которое для случая $\mu = 1$ может быть переписано для напряженности поля: $[\mathbf{H}_n] = 0$, или, что то же (см. рис. 11.1 *a*):

$$(H_i + H_r) \sin \varphi = H_d \sin \psi \implies (E_i + E_r) \sin \varphi = E_d \sin \psi.$$

Это условие тождественно удовлетворяется как следствие закона преломления и граничного условия на тангенциальные компоненты электрического поля. Тем самым мы убедились, что ответ (11.13) вполне корректен.

Рассмотрим случай поляризации электрического вектора в плоскости падения (рис 11.1б). Те же самые условия (9.12), (9.15) теперь запишутся в виде

$$\begin{aligned} [\mathbf{E}_\tau] &= 0 \implies E_i \cos \varphi - E_r \cos \varphi = E_d \cos \psi; \\ [\mathbf{H}_\tau] &= 0 \implies H_i + H_r = H_d \implies E_i + E_r = \frac{\sin \varphi}{\sin \psi} E_d. \end{aligned}$$

В терминах коэффициентов отражения и прозрачности эта система уравнений может быть представлена следующим образом:

$$t = \frac{\sin \varphi}{\sin \psi} (1 + r); \quad \frac{\cos \varphi}{\cos \psi} (1 - r) = \frac{\sin \varphi}{\sin \psi} (1 + r). \quad (11.14)$$

Решение (11.14), как и (11.12), сводится к стандартным тригонометрическим преобразованиям, хотя и несколько более громоздким. Ответ:

$$r_{||} = \frac{\operatorname{tg}(\varphi - \psi)}{\operatorname{tg}(\varphi + \psi)}; \quad t_{||} = \frac{2 \sin \psi \cos \varphi}{\sin(\varphi + \psi) \cos(\varphi - \psi)}. \quad (11.15)$$

Предоставляем читателю самостоятельно убедиться в том, что граничное условие (9.14) в данном случае выполняется автоматически, а условие (9.13) вырождено.

Таким образом, задача решена полностью. Хорошей проверкой наших ответов будет сведение их к общему предельному случаю нормального падения, которому мы припишем коэффициенты r_0 , t_0 . Из рис. 11.1 следует, что значения коэффициента отражения, полученные предельным переходом из (11.13), (11.15), должны отличаться знаком — на рис. 11.1а знак «—» учитывает потерю полуволны, которая на рис. 11.1б последует автоматически, потому что здесь сама амплитуда E_r — другого знака. С учетом этого обстоятельства искомые величины оказываются равны следующим:

$$r_0 = \frac{n - 1}{n + 1}; \quad t_0 = \frac{2}{n + 1}. \quad (11.16)$$

Формулы (11.13), (11.15), (11.16) — именно они и называются *формулами Френеля* (по имени французского физика О.Ж. Френеля (1788–1827)) — в достаточной мере универсальны с точки зрения соотношения коэффициентов преломления, т. е. они равным образом могут применяться при падении волны из вакуума на границу с диэлектриком, из диэлектрика в вакуум или на границу между двумя диэлектриками. При этом углы φ , ψ и коэффициенты преломления n_1 , n_{II} , n связаны законом (11.10).

11.4. Поляризационные эффекты. Поток энергии через границу

Сравнивая между собой коэффициенты отражения и прозрачности для различных поляризаций падающей волны — формулы (11.13) и (11.15), мы находим, что они отличаются друг от друга для всех углов падения, кроме нулевого — формула (11.16). Нам еще предстоит иметь дело с эффектом поляризации света в следующем разделе нашего курса, здесь же мы рассмотрим некоторые следствия такого рода из формул Френеля.

Дело в том, что естественный свет — и равным образом любые электромагнитные волны естественного происхождения — всегда неполяризованы. С точки зрения решений уравнений Максвелла это значит, что свет (излучение) представляет собой линейную смесь волн самых различных поляризаций. Мы уже обсуждали это свойство в связи с проблемой точечного источника. Разумеется, каждый отдельный квант, излучаемый возбужденным атомом, как-то поляризован, но если характерное время регистрации много больше времени замены одного такого кванта другим, с другой поляризацией, то мы этого не увидим. Это свойство присуще также и излучению искусственного происхождения, например, если его источник — тепловой природы: лампа накаливания, газоразрядная трубка (но не лазер!).

Теперь представим себе, что такое неполяризованное излучение падает, как это показано на рис. 11.1, на границу двух сред с разными n . Ввиду различия коэффициентов (11.13) и (11.15), как отраженный, так и преломленный свет уже приобретет некоторую степень поляризации. Последнюю можно характеризовать, например, отношением средних квадратов электрического поля, соответствующего различным поляризациям. Неполяризованный свет (или иное электромагнитное излучение) отвечает условию

$$\langle E_{\perp}^2 \rangle = \langle E_{\parallel}^2 \rangle.$$

Напомним, что в данном случае усреднение проводится по промежутку времени порядка или большему характерного времени регистрации. Для вторичных волн из (11.13) и (11.15) получаем

$$\frac{\langle E_{\perp}^2 \rangle_r}{\langle E_{\parallel}^2 \rangle_r} = \left(\frac{r_{\perp}}{r_{\parallel}} \right)^2 = \left[\frac{\cos(\varphi - \psi)}{\cos(\varphi + \psi)} \right]^2; \quad \frac{\langle E_{\perp}^2 \rangle_d}{\langle E_{\parallel}^2 \rangle_d} = \left(\frac{t_{\perp}}{t_{\parallel}} \right)^2 = \cos^2(\varphi - \psi). \quad (11.17)$$

Дадим для справки формальное определение степени поляризации электромагнитного излучения Δ :

$$\Delta = \frac{|\langle E_{\perp}^2 \rangle - \langle E_{\parallel}^2 \rangle|}{\langle E_{\perp}^2 \rangle + \langle E_{\parallel}^2 \rangle}.$$

Особый интерес представляет возможность полной поляризации отраженного света при $\varphi + \psi = \pi/2$ — уже из (11.15) видно, что при этом $r_{\parallel} = 0$. Это означает, что в отраженном свете вектор \mathbf{E} строго перпендикулярен плоскости падения. С учетом (11.10) находим соответствующий угол падения

$$\varphi_B = \arctg n. \quad (11.18)$$

Этот угол называется *углом Брюстера*, что и отражено индексом. Заметим, что максимальная поляризация отраженного света не соответствует максимальной поляризации преломленного — это можно увидеть из второй формулы (11.17). Еще отметим следующее обстоятельство. Если диэлектрическая пластинка освещается белым светом, то условие (11.18) должно включать зависимость $n(\omega)$ или, что то же, $n(\lambda)$, и угол Брюстера будет разным для волн разной частоты.

Таким образом, простейший поляроид, доступный каждому, — это просто стеклянная пластинка, которую следует располагать так, чтобы свет падал на нее по возможности близко к φ_B . С помощью такого поляроида нетрудно, например, заметить поляризацию света, отраженного от водной поверхности. Если же набрать пакет из таких пластинок и наблюдать источник света на просвет, то можно получить достаточно высокую степень поляризации, $\sim \cos^{4N}(\varphi - \psi)$, где N — число параллельных пластинок. (Фактор 4, вместо двойки, есть следствие того, что у каждой пластинки — две поверхности раздела сред.) Такой поляроид впервые был предложен русским физиком А. Г. Столетовым (1839–1896) и называется *стопой Столетова*.

Если r и t — амплитудные коэффициенты отражения и прозрачности, то r^2 и t^2 — то же для интенсивности. В этой связи обращает на себя внимание следующее обстоятельство. Вне зависимости от поляризации, из обеих формул (11.13) и (11.15) следует $r^2 + t^2 \neq 1$. Легче всего это увидеть из формул (11.16) для нормального падения, и в этом же случае легче всего прокомментировать. Никакого парадокса здесь нет, поскольку в задаче такого рода следует подводить баланс не энергии как таковой, а потока энергии.

Предположим, что электромагнитная волна падает из вакуума на границу диэлектрика. Падающий на границу поток энергии равен, согласно (11.8),

$$|\langle \mathbf{S} \rangle_i| = \sqrt{\frac{\varepsilon_0}{\mu_0}} \langle E_i^2 \rangle,$$

отраженный, соответственно,

$$|\langle \mathbf{S} \rangle_r| = r_0^2 \sqrt{\frac{\varepsilon_0}{\mu_0}} \langle E_i^2 \rangle,$$

а уходящий с преломленной волной,

$$|\langle \mathbf{S} \rangle_d| = t_0^2 \sqrt{\frac{\varepsilon \varepsilon_0}{\mu_0}} \langle E_i^2 \rangle = n t_0^2 |\langle \mathbf{S} \rangle_i|.$$

Таким образом, баланс потоков энергии в данной задаче требует следующего равенства:

$$r_0^2 + n t_0^2 = 1, \quad (11.19)$$

которое и в самом деле выполнено для коэффициентов отражения и прозрачности (11.16).

11.5. Электромагнитные волны и теория относительности

Исходя из самой процедуры вывода волнового уравнения (10.13) или (10.14), электромагнитные волны представляют собой поля \mathbf{E} и \mathbf{B} в вакууме и тем самым принципиально отличаются от других известных волновых движений, например, звуковых волн. Общность формального описания — волновые уравнения, их фундаментальные решения и пр. — в течение многих десятилетий побуждала физическое сообщество к поиску более глубокой физической аналогии. Поскольку звуковые волны представляют собой волновые движения в упругой твердой или жидкой среде, высказывались предположения о существовании некоторой гипотетической среды — эфира, колебания которой и описываются уравнениями Максвелла. Среда эта должна быть абсолютно упругой (прозрачной) и заполнять все физическое пространство, включая находящееся там вещество. Строились достаточно убедительные теоретические модели (автором одной из наиболее совершенных был Г. Герц), которые могли объяснить некоторую совокупность экспериментальных фактов, но рано или поздно приходили в противоречие с данными очередного эксперимента. Как особо важные следует отметить прежде всего опыты Майкельсона, экспериментально доказавшего, что скорость света c не зависит от того, распространяется сигнал по- или против скорости движения Земли по орбите (~ 30 км/с). Параллельно был поставлен опыт Физо, впоследствии повторенный в усовершенствованном варианте Майкельсоном, из которого следовало, что закон «сложения скоростей» эфира и световых волн должен принципиально отличаться от такового в механике Ньютона.

Вопрос о сложении скоростей вообще оказался совершенно принципиальным. Напомним, что в ньютоновой механике переход между двумя инерциальными системами отсчета задается

преобразованием Галилея

$$\mathbf{r}' = \mathbf{r} - \mathbf{V}t; \quad t' = t, \quad (11.20)$$

где \mathbf{V} — скорость новой системы отсчета относительно старой. Второе условие, в частности, декларирует, что время течет одинаково во всех инерциальных системах. В течение почти двух столетий физический эксперимент, выполненный с доступной по тому времени точностью, подтверждал адекватность преобразований (11.20).

Механика Ньютона инвариантна по отношению к преобразованию (11.20). Это и выражено по сути первым законом Ньютона, который можно прочесть следующим образом: покой есть эквивалент равномерного движения, и все законы природы одинаковы во всех инерциальных системах отсчета. Но уравнение (10.5), хотя бы оно и следовало из ньютоновой механики, отнюдь не инвариантно в смысле (11.20). В этом нет парадокса, потому что любые механические колебания привязаны к колебательной системе, а звук, соответственно, к твердой или текучей среде (под последней мы понимаем жидкость или газ). Разумеется, среди возможных систем отсчета, переходы между которыми описываются формулами (11.20), одна выделена, а именно, та, в которой среда покоится, и следовательно, для волновых движений не существует какого-либо преимущественного направления.

Предположим, эфира как такового не существует. Все известные модели его отвергнуты экспериментом. Электромагнитные волны суть поля в вакууме. Вакуум не имеет скорости, и понятие направления для него не имеет смысла. Значит, уравнения Максвелла (9.2)–(9.5) или (9.8)–(9.11) должны выглядеть одинаково во всех экспериментальных системах. Но уравнения Максвелла не инвариантны относительно преобразований (11.20)!

Описанные трудности и явились побудительным мотивом к созданию специальной теории относительности (СТО). В настоящее время ее более принято называть релятивистской механикой, и главный экспериментальный материал для ее подтверждения поставляют ядерная физика и физика элементарных частиц. На данный момент эти данные не менее убедительны, чем обоснования ньютоновой механики. Но до появления знаменитой работы Эйнштейна (1905 г.) их практически не было, зато в течение полувека информация к размышлению поставлялась оптическими измерениями. И сама работа Эйнштейна называлась «К электродинамике движущихся тел», и главным постулатом в ней — в духе сказанного выше — была независимость скорости света от (инерциальной) системы отсчета.

Как известно, последовательное развитие этой идеи вкупе с принципом эквивалентности инерциальных систем потребовало

замены преобразований Галилея более сложными преобразованиями Лоренца. Для удобства записи принято оси Ox , Ox' направлять вдоль относительной скорости \mathbf{V} . Тогда, вместо (11.20), получаются следующие формулы перехода:

$$x' = \gamma(x - Vt); \quad y' = y; \quad z' = z; \quad t' = \gamma\left(t - \frac{V}{c^2}x\right), \quad (11.21)$$

где γ — так называемый релятивистский фактор:

$$\gamma = \frac{1}{\sqrt{1 - V^2/c^2}}. \quad (11.22)$$

В отличие от преобразований Галилея, время уже не является общим для всех инерциальных систем и преобразуется вместе с координатами. При $V \ll c$ преобразования (11.21), (11.22) переходят в (11.20), и тем самым релятивистская механика возвращается к механике Ньютона. Аналогичный предельный переход совершают выражения для энергии и импульса свободной частицы

$$\mathcal{E} = \gamma mc^2; \quad \mathbf{p} = \gamma m \mathbf{v},$$

(m — масса покоя частицы), с тем, правда, отличием, что в выражении для энергии возникает аддитивная добавка mc^2 . Иногда пользуются понятием релятивистской массы, связанной с массой покоя коэффициентом γ , и тогда выражение $\mathcal{E} = mc^2$ становится универсальным; в последнее время, однако, такого раздвоения понятия массы физики предпочитают избегать.

Если бы мы просто применили (11.21), (11.22) к уравнениям Максвелла, никак не преобразуя полей, инвариантности все равно не получилось бы. Но еще в задаче 3 гл. 4 мы из физических соображений ввели преобразование электрического поля (4.27) при переходе к движущейся системе отсчета. Теперь же мы предложим (правда, без вывода) полную систему преобразования полей при произвольной относительной скорости. В тех же координатных осях

$$E'_{x'} = E_x; \quad E'_{y'} = \gamma(E_y - V B_z); \quad E'_{z'} = \gamma(E_z + V B_y); \quad (11.23)$$

$$H'_{x'} = H_x; \quad H'_{y'} = \gamma(H_y + V D_z); \quad H'_{z'} = \gamma(H_z - V D_y); \quad (11.24)$$

Существенно, что заряд оказывается при таких преобразованиях инвариантным, $q' = q$. Электромагнитное поле, как можно видеть из (11.23), (11.24), преобразуется по закону более сложному, чем простая вектор-функция, которая подчиняется преобразованию Лоренца. Электрическое и магнитное поля оказываются связанными в этих преобразованиях и в зависимости от системы координат переходят друг в друга. Тем не менее,

формулы (11.23), (11.24) имеют вполне определенный «геометрический» смысл: величины, которые мы считали векторными, оказываются компонентами некоторого тензора — четырехмерного, поскольку он определен в пространстве (x, y, z, t) . Мы лишь для того отмечаем это обстоятельство, чтобы предостеречь читателя: в рамках преобразований Галилея такого преобразования полей, которое обеспечивало бы инвариантность уравнений Максвелла, найти не удастся.

Таким образом, единственная среда с $n \equiv 1$ есть вакуум. (Разумеется, в конкретных физических задачах мы очень часто можем с приемлемой для нас точностью полагать $n = 1$ и в достаточно разреженном веществе, например в атмосферном воздухе.) Идея эфира оказалась несостоятельной, но и простые преобразования перехода между инерциальными системами для электромагнитных явлений неприменимы. Уравнения Максвелла (9.2)–(9.5) или (9.8)–(9.11) оказываются существенно релятивистскими, хотя получая их, мы этих вопросов вообще не касались. В электродинамике принципиально важную роль играют электромагнитные волны, а это объект в определенном смысле ультрарелятивистский, ибо скорость их в точности равна скорости света.

Физика XX в. накопила огромный экспериментальный материал, вполне подтверждающий СТО и, в частности, справедливость соотношений (11.23), (11.24).

Задачи

1. Определить степень поляризации преломленного света, если первичный поток неполяризованного света падает под углом Брюстера из пустоты на поверхность стекла с показателем преломления $n = 1,5$.

Решение. Из условия $\operatorname{tg} \varphi = n$ находим $\varphi = 56^\circ 19'$. Искомая величина Δ_t может быть выражена через коэффициенты прозрачности:

$$\Delta_t = \frac{|\langle t_{\perp}^2 \rangle - \langle t_{\parallel}^2 \rangle|}{\langle t_{\perp}^2 \rangle + \langle t_{\parallel}^2 \rangle}.$$

Далее, учитывая, что $\sin \varphi = n \sin \psi = n \cos \varphi$, подстановкой в (11.13), (11.15) получаем

$$\Delta_t = \frac{\cos^{-2}(\varphi - \psi) - 1}{\cos^{-2}(\varphi + \psi) - 1} = \frac{1 - 4n^2/(1 + n^2)}{1 + 4n^2/(1 + n^2)} \approx 0,08.$$

2. Переменное электромагнитное поле с характерной частотой ω создается на границе слабопроводящей диэлектрической среды с параметрами ϵ , μ , $\lambda \ll \ll \epsilon_0 \omega$. Определить характерную глубину проникновения.

Решение. Задача тем отличается от задачи о скин-эффекте, что в этой ситуации, согласно условию, ток смещения доминирует над током проводимости. Действительно, для полей вида (10.15) это сразу следует из уравнений (9.2), (9.3):

$$[\mathbf{k} \mathbf{E}] = \omega \mathbf{B}; \quad (\mu \mu_0 \omega)^{-1} [\mathbf{k} [\mathbf{k} \mathbf{E}]] = -(\mu \mu_0 \omega)^{-1} k^2 \mathbf{E} = -i \lambda \mathbf{E} - \omega \epsilon_0 \mathbf{E}. \quad (11.25)$$

Поэтому в данном случае мы будем иметь дело не с квазистационарным профилем тока, но с проникающей в среду электромагнитной волной. Но, в отличие от (10.15), для нее частота и волновое число не могут быть одновременно вещественными, как нетрудно усмотреть из второго уравнения (11.25). Постановка задачи соответствует пространственному затуханию волны, поэтому комплексным нужно выбрать k , причем, в силу малости проводимости, относительно малой будет и мнимая часть:

$$k \equiv k_0 + i\kappa \implies E, B \propto \exp(-\kappa x); \quad \kappa \ll k_0,$$

где x — координата по направлению в глубину среды от границы. Во втором уравнении (11.25) сокращаем амплитуду \mathbf{E} :

$$\varepsilon\mu\varepsilon_0\mu_0 + i\frac{\lambda\mu_0}{\omega} = \frac{k^2}{\omega^2} \approx \frac{k_0^2}{\omega^2} + \frac{2ik_0\kappa}{\omega^2}.$$

Из главных членов этого равенства следует, как обычно, $\omega/k = c/n$; приравняв члены первого порядка малости, получаем

$$\kappa = \frac{\lambda\mu_0 c}{2n}.$$

Таким образом, эффективная глубина проникновения, определенная через показатель экспоненты, оказывается равной

$$\delta_{\text{эф}} = \kappa^{-1} = \frac{2n}{\lambda\mu_0 c}.$$

Эта величина заметно отличается от квазистационарной глубины скин-слоя (8.28). Заметим, что из условия $\kappa \ll k$, которое, в свою очередь, есть следствие малой проводимости, вытекает, что глубина проникновения многократно превосходит длину волны.

3. Электромагнитное излучение заданной интенсивности $\langle E_i^2 \rangle$ падает по нормали из вакуума на поверхность диэлектрика с диэлектрической проницаемостью ε . Определить давление на поверхность раздела сред.

$$\text{Ответ: } \mathcal{P} = \frac{2r^2|S_i|}{c^2} = \frac{2}{c^2} \left(\frac{n-1}{n+1} \right)^2 \sqrt{\frac{\varepsilon_0}{\mu_0}} \langle E_i^2 \rangle.$$

4. Естественный свет падает под углом Брюстера на поверхность стекла с показателем преломления $n = 1,5$. Найти интенсивность $\langle E_r^2 \rangle$ отраженной волны, полагая величину $\langle E_i^2 \rangle$ известной.

$$\text{Ответ: } \langle E_r^2 \rangle = \frac{\langle E_i^2 \rangle}{2} \frac{(n^2 - 1)^2}{(n^2 + 1)^2} \approx 0,07 \langle E_i^2 \rangle.$$

Глава 12

ЭЛЕМЕНТЫ ФИЗИКИ ПЛАЗМЫ

12.1. Газовый разряд

В каждом разделе общей физики существуют объекты, наилучшим образом демонстрирующие физические законы, которые в данном разделе являются предметом рассмотрения. Так, в молекулярной физике традиционным объектом для мысленных экспериментов и частных примеров общих закономерностей является идеальный или слабонеидеальный газ, в квантовой физике — атом водорода, а затем кристалл и т. д. В электродинамике один из самых удобных объектов такого рода — состояние вещества, называемое плазмой, поскольку главные свойства вещества в состоянии плазмы задаются именно электродинамикой. При этом, в отличие, скажем, от ионных кристаллов или электролитов, прочие законы физики оказываются в какой-то мере подчиненными, вторичными, и многие совершенно принципиальные задачи физики плазмы могут рассматриваться в рамках одной только электродинамики.

Мы начнем, однако, не с физики плазмы как таковой, но с процесса протекания тока в газе — газового разряда. Физику газового разряда нередко смешивают с физикой плазмы. Как мы увидим далее, это не вполне корректно, но четкой границы между ними и в самом деле нет. Можно сказать, что в типичных условиях газовый разряд выделен исключительной неравновесностью: объект существует лишь при определенных условиях протекания через среду электрического тока (а иногда как необходимое условие добавляется внешний источник ионизирующего излучения). Все его свойства и закономерности чрезвычайно чувствительны к режиму, заданному извне. Но поскольку газовый разряд — главный метод получения плазмы в лаборатории и наиболее доступный объект плазменной природы (не считая Солнца и звезд), следует, хотя бы кратко, остановиться на физике этого явления.

Из чего бы газ ни состоял, в состоянии, близком к равновесному, он всегда — плохой проводник. Причина вполне очевидна: в газе молекулы почти никак друг с другом не связаны, каждая из них совершает хаотическое (броуновское) движение, и нет сколько-нибудь эффективных механизмов переноса заряда. При прохождении мощных импульсов тока через тонкие проводники,

когда вследствие мощного энерговыделения вещество проводника последовательно проходит все состояния от твердого до плазменного, паро-газовой фазе сопутствует так называемая «пауза проводимости».

Однако в любом газе, даже если представить его абсолютно изолированным от внешней среды, всегда есть хотя бы очень малая доля молекул, настолько энергичных, что при межмолекулярных столкновениях происходит отрыв электронов от нейтральных атомов и образование электрон-ионных пар. Таким образом в газе появляются возможные носители тока. Реальное же их число всегда значительно выше того, которое формально следует из термодинамического равновесия, поскольку во Вселенной существует определенный фон космического излучения. Оно и поддерживает в любом газе некоторую стационарную степень ионизации, а значит, и проводимости.

В гл. 3 мы моделировали механизм сопротивления, связав его с эффективной частотой столкновений электрона (см. формулу (3.11)):

$$\lambda = \frac{ne^2}{m\nu_{\text{эф}}}$$

Если для большинства проводников эта формула — всего лишь наглядная модель микроскопической природы закона Ома, то для слабоионизованного газа она работает как очень хорошее приближение к реальности. Не следует, однако, думать, что это выражение есть уже ответ на все вопросы. Физическая задача часто в том и состоит, чтобы правильно определить величину $\nu_{\text{эф}}$, включая зависимость от полей \mathbf{E} и \mathbf{B} , так как закон Ома в газах чаще всего оказывается нелинейным. Кроме электронов, ток могут переносить и ионы. Обычно этот канал менее эффективен, но в некоторых газах, например O_2 , электроны эффективно адсорбируются нейтральными атомами, и тогда ток переносится двумя сортами ионов. В случае носителей нескольких сортов скорость каждого из них (индекс α) связывается с электрическим полем коэффициентом, который принято называть *подвижностью* и обозначать через μ_α :

$$\mathbf{v}_\alpha = \mu_\alpha \mathbf{E}; \quad \mu_\alpha = \frac{Z_\alpha e}{m_\alpha \nu_\alpha}; \quad Z_e = -1; \quad (12.1)$$

— см. вывод формулы (3.11). Здесь $Z_\alpha e$ — заряд соответствующего иона; электроны для общности тоже включены в их число. Соотношение (12.1) позволяет обобщить формулу (3.11):

$$\lambda = \sum_{\alpha} \frac{(Ze)_\alpha^2 n_\alpha}{m_\alpha \nu_\alpha}. \quad (12.2)$$

Опыт показывает, однако, что простой закон Ома, хотя бы и нелинейный, описывает протекание тока в газе лишь при до-

статочных малых полях. Пример вольт-амперной характеристики газового разряда представлен на рис. 12.1. Для наших рассуждений ее удобнее рассматривать в локальной форме $j(E)$.

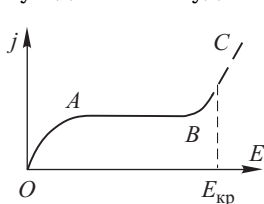


Рис. 12.1

Участок OA соответствует пределу слабого поля и описывается формулой (12.2). Далее в некотором интервале значений поля (участок AB) плотность тока — почти константа, затем при некотором критическом значении поля $E_{кр}$ происходит пробой, и ток многократно возрастает, чему отвечает участок кривой BC . Разряд в области OAB называется несамостоятельным, а при $E > E_{кр}$ — самостоятельным. Происхождение такой терминологии будет ясно из дальнейшего. Теория, описывающая столь нетривиальное поведение тока в зависимости от поля, была впервые построена английским физиком Дж. Таунсендом (1868–1957). Мы воспроизведем ее основные положения и результаты.

Главная причина неадекватности закона Ома в том, что не только усредненная величина $\nu_{эф}$, но и рождение носителей тока существенно для физики разряда. Рассмотрим для простоты чисто электронную динамику. Внешние источники (космические лучи) рожают в единицу времени определенное число электронов. Можно было бы ожидать, что когда плотность тока достигнет такой величины, то именно столько их и будет приходить на анод; рост тока, соответствующий (12.2), прекратится, и мы как раз выйдем на участок AB вольт-амперной характеристики. Но такое представление оказывается слишком примитивным, тем более, оно не может объяснить эффект пробоя. На самом деле уже в области AB в игру вступает новое явление — образование электронной лавины. Будем считать, что газ заполняет некоторый диод (на рис. 12.2 он показан плоско-параллельным, K — катод, A — анод). Будучи ускорены в электрическом поле E , электроны при соударениях с нейтральными атомами и молекулами производят вторичную ионизацию. Главный источник первичных электронов

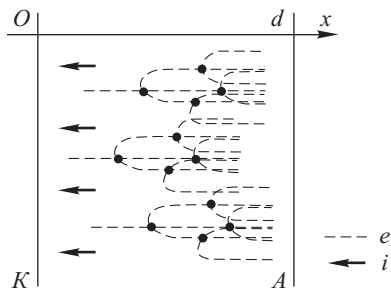


Рис. 12.2

в такой системе — катод, так как внешнее излучение эффективнее взаимодействует с конденсированным веществом, нежели с разреженным газовым наполнением. Но электроны, выбитые из анода, тут же втягиваются обратно электрическим полем, тогда

как с катода они стартуют в направлении анода. Пусть в единицу времени на катоде рождается число электронов, равное N_0 . Ради упрощения вычислений мы в дальнейшем пренебрегаем рождением электронов в газовом объеме. Поскольку длина пробега от нулевой скорости примерно одинакова в любой области, а практически всю энергию, полученную при ускорении в электрическом поле, электрон каждый раз отдает на вторичную ионизацию, число вторичных электрон-ионных пар на длине пробега, а значит и на единицу длины, примерно одинаково и может характеризоваться некоторым параметром β :

$$\beta \equiv \frac{dN_{\pm}}{dx},$$

где ось Ox направлена от катода к аноду (рис. 12.2), координата катода $x = 0$, анода — $x = d$. Уравнение, описывающее изменение числа электронов с координатой x , следует из условия

$$dN = N\beta dx.$$

Решение его, как хорошо известно, — экспонента:

$$N \propto \exp(\beta x); \implies N(d) = N(0) \exp(\beta d). \quad (12.3)$$

Этот процесс экспоненциального накопления электронов в катод-анодном промежутке схематически показан на рис. 12.2. Для определенности мы «привязали» вторичные электроны к самому концу свободного пробега. Напомним еще раз, что число электронов у нас нормировано на единицу времени. Чтобы получить число вторичных электрон-ионных пар, нужно вычесть из (12.3) число электронов, порождаемое в единицу времени внешним источником:

$$N_{\mp} = N(0)[\exp(\beta d) - 1].$$

В установившемся режиме разряда именно таким будет число ионов, приходящих в единицу времени на катод. Они не слишком эффективны в процессе объемной ионизации — из-за их малых скоростей, однако из массивного катода они могут выбивать вторичные электроны — для этого энергии у них достаточно. Пусть каждый из них выбивает в среднем γ вторичных электронов. (Заметим, что коэффициенты β и γ могут быть как больше, так и меньше единицы.) Тогда число электронов, выбитых ионами из катода в единицу времени, равно

$$\gamma N(0) [\exp(\beta d) - 1],$$

а полное их число, образованное ионами и внешним источником в единицу времени

$$N(0) = N_0 + \gamma N(0) [\exp(\beta d) - 1].$$

Отсюда и следует окончательно:

$$N(0) = \frac{N_0}{1 - \gamma[\exp(\beta d) - 1]}; \quad N(d) = \frac{N_0 \exp(\beta d)}{1 - \gamma[\exp(\beta d) - 1]}. \quad (12.4)$$

Если знаменатели в формулах (12.4) положительны, мы имеем дело с *несамостоятельным разрядом*, поскольку $I \propto N_0$, но если мы достигаем критического условия

$$\gamma[\exp(\beta d) - 1] = 1,$$

то величина N_0 уже не существенна, поскольку полный ток в (12.4) формально обращается в бесконечность. Это и есть пробой, иными словами — переход в режим *самостоятельного разряда*.

Почему это происходит при росте напряженности поля (рис. 12.1)? Экспериментально установлено, что «коэффициент размножения» β (на рис. 12.2 он равен трем) есть некоторая монотонная функция отношения поля к давлению газа:

$$\beta = F(E/P).$$

Таким образом, по мере роста поля или снижения давления мы и попадаем в режим самостоятельного разряда:

$$\gamma[\exp(\beta(E_{\text{кр}})d) - 1] = 1. \quad (12.5)$$

В отличие от несамостоятельного, самостоятельный разряд уже не нуждается во внешнем ионизаторе; последний, однако, играет важную роль в процессе зажигания. Как следствие, для вольт-амперных характеристик самостоятельных разрядов характерен гистерезис: напряжение гашения всегда ниже напряжения зажигания.

Теория Таунсенда дает правильное представление о *несамостоятельном разряде* (его часто и называют таунсендовским) и о переходе в режим *самостоятельного разряда*, но последующая динамика разряда уже никак не будет универсальной. Прежде всего, в различных режимах (которые зачастую зависят не только от величины поля E , но и от «сценария» зажигания разряда) весьма различными оказываются режимы поддержания самостоятельного разряда.

В *тлеющем разряде*, хорошо нам известном по газоразрядным трубкам, картина явления в наибольшей степени близка той, которую мы только что рассмотрели. Главную роль в поддержании разряда играют ударная ионизация в объеме газа и вторичная эмиссия на катоде; последняя обусловлена не только ионной бомбардировкой, но и фотоионизацией за счет электромагнитного излучения самого разряда. Пожалуй, главным отличием от таунсендовского режима оказывается очень серьезная зависимость от пространственных координат всех характеристик

разряда: поля, потенциала, концентрации частиц, всех основных процессов — да и чисто визуальная.

Коронный разряд («огни святого Эльма») образуется обычно на остриях и тонких проволочках. Его отличает еще бо́льшая пространственная неоднородность, так что лишь в некоторой области пространства вблизи острия идет самостоятельный разряд, а в остальном пространстве пробоя газа не происходит, и ток через газ просто «навязан» областью пробоя.

Для *искрового разряда* (в частности, и молнии) характерны очень сильные поля и короткие характерные времена. Разряд происходит по треку предварительно образовавшегося канала, т. н. *стримера*. За его образование отвечают множественные электронные лавины, развитие которых происходит не совсем так, как в тлеющем разряде. Начало лавине дает акт фотоионизации фотоном, испущенным какой-то предыдущей лавиной. Поскольку фотоны, в отличие от электронов, распространяются со скоростью света, развитие стримера идет намного быстрее, чем единичной электронной лавины.

Для *дугового разряда* характерны, напротив, большие токи и сравнительно небольшие значения падения напряжения. Электроны, поддерживающие ток, поставляются термоэлектронной эмиссией с катода (возможна, впрочем, и «холодная» дуга). Здесь важную роль играет материал электродов, испаряющийся и ионизирующийся в процессе разряда.

Существует немало и других режимов. Сколько-нибудь детальное их рассмотрение едва ли уместно в общем курсе физики. Отметим как общее свойство всех разрядов следующее обстоятельство. В процессе газового разряда вещество превращается в плазму — агрегатное состояние, располагающееся выше, чем газ, на температурной шкале. Для перехода в состояние плазмы по обычному каналу фазовых переходов необходимо нагреть вещество до температуры 5 000–10 000 К, что как раз соответствует характерной температуре поверхности Солнца и звезд. Газовый разряд, ввиду исключительной его неравновесности, позволяет осуществить этот переход более дешевой ценой при гораздо более умеренных температурах и поддерживать вещество в этом состоянии также относительно простым и доступным способом.

12.2. Понятие плазмы

Плазму нередко определяют как ионизованный газ. Это верно по сути, но с равным успехом и газ можно определить как испаренную жидкость, а последнюю — как расплавленное твердое тело. При достаточно высокой степени ионизации газ, как мы вскоре убедимся, приобретает существенно новые свойства,

позволяющие с полным основанием говорить о новом агрегатном состоянии. Правда, поскольку ионизация — процесс непрерывный и может быть проведен адиабатически медленно, здесь нет выраженного фазового перехода.

Газовый разряд — сугубо неравновесный канал превращения газа в плазму. В равновесном состоянии плазма — объект весьма высокотемпературный. Поэтому температуру ее обычно измеряют в «атомных» энергетических единицах — *электронвольтах*:

$$1 \text{ эВ} = e \cdot 1 \text{ В} = 1,6 \cdot 10^{-19} \text{ Дж} = 11\,600 \text{ К}.$$

Энергия ионизации атома водорода равна примерно 13,6 эВ; температура плазмы в современных лабораторных установках, предназначенных для исследований по управляемому термоядерному синтезу, достигает 10^3 – 10^4 эВ, тогда как температура поверхности Солнца — всего лишь около 0,5 эВ. Такую плазму называют холодной.

Может возникнуть вопрос: как может вещество в равновесном состоянии полной или почти полной ионизации иметь температуру, на порядок меньшую энергии ионизации? Дело в том, что ионизация поддерживается относительно малой долей высокоэнергичных частиц — «хвостами» функции распределения по скоростям.

Для сравнения укажем, что для излучения фотона в красной области спектра энергия возбуждения излучающего атома должна превышать 1,5 эВ, но температура красного каления обычно заметно ниже 0,1 эВ.

Ввиду высокой степени равновесной ионизации плазма, в отличие от газа при умеренной температуре, оказывается очень хорошим проводником. Даже холодная плазма поверхности Солнца по проводимости приближается к металлам, а высокотемпературная термоядерная плазма по этому параметру значительно их превосходит. Например, проводимость водородной плазмы при температуре 10 кэВ примерно в 20 раз выше, чем у меди при нормальных условиях. В то же время плазма — среда текучая. Отсюда следует первое отличие плазменной фазы от газовой: у нее невозможно разделить механические и электродинамические свойства. Электромагнитные поля вызывают течение плазмы и управляют им, и напротив, течение плазмы порождает электромагнитное поле.

Мы уже рассматривали в гл. 7 силы, действующие на проводник в магнитном поле, и, в частности, установили, что в случае прямых силовых линий и нулевого поля в проводящей среде на поверхность проводника действует давление (см. (7.1))

$$\mathcal{P}_B = \frac{B^2}{2\mu_0}. \quad (12.6)$$

Мы положили $\mu = 1$, что для плазмы, как и для газа, всегда верно с очень хорошей точностью. Именно для горячей плазмы эффект магнитного давления особенно выразителен. Во-первых, из-за высокой проводимости коэффициент диффузии поля $D_B = 1/(\lambda\mu_0)$ будет очень мал. А во-вторых, малы и характерные времена плазменного эксперимента: от 10^{-8} до 10^{-3} с. Как следствие, эффективная глубина проникновения поля в плазму (см. оценку (8.27)) может быть много меньше характерного размера плазменного облака.

Отсюда следует возможность магнитного удержания плазмы. Дело в том, что высокотемпературную плазму нельзя непосредственно заключить в твердый сосуд — ее погубит рекомбинация на стенках; скажем, в случае водородной плазмы $H^+ + e \Rightarrow H$. Поэтому плазму обычно удерживают бесконтактным методом — «магнитной стенкой» (соответственно, ее можно толкать «магнитным поршнем»). Создается конфигурация поля, в которой плазменное облако оказывается как бы в яме в некоторой окрестности минимума (B^2). Такие устройства называются *магнитными ловушками*. Разумеется, одно лишь магнитное давление (12.6) равновесия обеспечить не может. Необходимо учесть также газокINETическое давление плазмы $\mathcal{P} = nk_B T$, где k_B — постоянная Больцмана. Условие равновесия, включая и пограничный слой, где поле проникает в плазму, есть условие постоянного полного давления:

$$nk_B T(\mathbf{r}) + \frac{B^2(\mathbf{r})}{2\mu_0} = \text{const.} \quad (12.7)$$

Препятствуют равновесию, помимо диффузии поля, еще и многочисленные неустойчивости, исследование которых представляет собой довольно заметную часть вообще всей физики плазмы. В большинстве экспериментальных ситуаций макроскопическая динамика плазмы тождественна так называемой магнитной гидродинамике, которая, в частности, описывает равновесие и течение жидких металлов. Таким образом, хотя плазма и располагается по температурной шкале выше газов, по своим электромеханическим свойствам она оказывается близка конденсированному веществу и заметно отличается от газа.

Рассматривая электромагнитные поля в плазме, мы не можем считать ρ и \mathbf{j} заданными извне — они должны определяться самосогласованным образом как функции электрического и магнитного полей. Поскольку плазма, как и газ, — среда бесструктурная, заряды и токи, наведенные полями, представляют собой просто свободные заряды и токи проводимости. Поэтому их можно прямо включить в уравнения Максвелла, и не делая различия по физике между полем и индукцией, т. е. положив $\varepsilon = \mu = 1$, оставить в базовых уравнениях только действующие

поля \mathbf{E} и \mathbf{B} :

$$\operatorname{div} \mathbf{E} = \frac{1}{\varepsilon_0} \rho(\mathbf{E}, \mathbf{B}); \quad (12.8)$$

$$\operatorname{div} \mathbf{B} = 0; \quad (12.9)$$

$$\operatorname{rot} \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}; \quad (12.10)$$

$$\frac{1}{\mu_0} \operatorname{rot} \mathbf{B} = \mathbf{j}(\mathbf{E}, \mathbf{B}) + \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}. \quad (12.11)$$

К числу наиболее важных свойств плазмы относится квазинейтральность.

Представим себе, что некоторое количество газа переводится тем или иным способом в состояние, полностью или частично ионизованное. При этом полный заряд его остается равным нулю.

В принципе, возможно существование и заряженной плазмы (пример — облако пространственного заряда в электронной лампе). Но оказывается, что для систем с выраженными плазменными свойствами как в природе, так и в лабораторных условиях свойство нейтральности в целом гораздо более типично. Более того, плазма достаточно высокой концентрации оказывается нейтральной и в локальном смысле:

$$n_e = Z n_i, \quad (12.12)$$

где $n_{e,i}$ — концентрации компонентов, а Z — заряд иона (кратность ионизации), который в дальнейшем мы для простоты будем считать равным единице (для водородной или дейтериевой плазмы это вообще единственная возможность). Именно свойство (12.12) и называется *квазинейтральностью*. Оно определяет многие специфические для плазмы физические эффекты и свойства.

Покажем на типичном примере, почему квазинейтральность трудно нарушить в макроскопических масштабах.

Возьмем кубик объемом $V = 1 \text{ см}^3$ и предположим, что его заполняет плазма, достаточно типичная для лабораторных экспериментов:

$$n_i = n_e = 10^{20} \text{ м}^{-3}.$$

Пусть мы хотим нарушить равенство концентраций в данном объеме всего на один процент, т. е. создать избыток концентрации любого из компонентов

$$\delta n \sim 0,01 n = 10^{18} \text{ м}^{-3}.$$

Оценим возникающее при этом электрическое поле, используя теорему Гаусса:

$$\oint \mathbf{E} d\mathbf{S} \approx E \cdot S = \frac{q}{\varepsilon_0} = \frac{\delta n e V}{\varepsilon_0}.$$

Вычисления дают $E \approx 3 \cdot 10^7$ В/м. Можно той же оценке придать и другой смысл. Пусть кубик нейтрален в целом,

$$\int n_e dV = \int n_i dV,$$

но попытаемся нарушить локальное равенство (12.12), например, у одной грани кубика пусть будет $n_i - n_e = -10^{-2}n_{e,i}$, а у другой $+10^{-2}$. Получим ту же цифру. В первом случае электростатическое отталкивание эквивалентно огромному добавочному давлению, взрывающему объем, заполненный плазмой, а во втором притяжение разноименных зарядов эффективно выравнивает локальную плотность заряда. И в обоих случаях, чтобы поддержать нашу систему в неравновесном состоянии, необходимы внешние поля, не меньшие, чем 10^7 В/м. Если же внешнее поле такого порядка величины не достигает, условие (12.12) в нашем конкретном случае будет выполнено с точностью $\delta n \leq 0,01n_{i,e}$.

Конечно, само определение локальной концентрации подразумевает некоторое усреднение числа частиц по пространству и по времени. Поэтому и утверждение о квазинейтральности не может быть абсолютным, хотя бы из-за дискретности вещества. Однако, как мы увидим в следующем параграфе, главный фактор нарушения квазинейтральности, а тем самым и масштабы такого усреднения связаны с движением частиц плазмы.

12.3. Ленгмюровские колебания и дебаевское экранирование

Попытаемся оценить минимальный масштаб времени, ниже которого квазинейтральность не может быть обеспечена. Рассмотрим плоский слой плазмы и дадим в нем всем электронам смещение относительно ионов в перпендикулярном слою направлении (рис. 12.3). Если x — пространственное смещение ($x \ll l$), то заряд на единицу площади в узких приграничных слоях оказывается равным $\sigma = nex$. Поле каждой из двух параллельных заряженных плоскостей равно, согласно (1.11), $\sigma/2\varepsilon_0$; вне плазменного слоя эти поля компенсируют друг друга, а внутри слоя складываются (все это очень напоминает плоский конденсатор). Таким образом, электрическое поле в слое

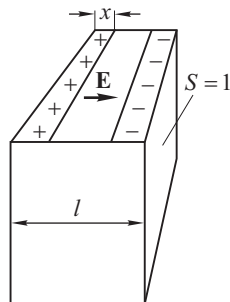


Рис. 12.3

$$E = \frac{nex}{\varepsilon_0}.$$

Принимая во внимание соотношение масс $M_i/m_e \gg 1$ (почти 2000 даже для водорода), пренебрегаем смещением ионов, и тогда электронное уравнение движения оказывается очень простым:

$$m \frac{d^2 x}{dt^2} = -eE = -\frac{n e^2 x}{\varepsilon_0}. \quad (12.13)$$

Это уравнение гармонических колебаний, частота которых равна

$$\omega \equiv \omega_{pe} = \left(\frac{n e^2}{\varepsilon_0 m} \right)^{1/2}. \quad (12.14)$$

Колебания электронов относительно ионов, которые мы таким образом получили, называются *ленгмюровскими* (по имени американского физика И. Ленгмюра (1881–1957)) или просто *плазменными колебаниями*. Последнее обстоятельство подчеркивает очень важную роль, которую играют движения такого рода в микроскопической динамике плазмы. Частота (12.14) называется соответственно *электронной плазменной* или *электронной ленгмюровской частотой*.

Поскольку электроны как целое могут перемещаться относительно ионов, постольку может нарушаться квазинейтральность, но лишь на очень короткое время, порядка периода плазменных колебаний. Следовательно, масштаб усреднения по времени, подразумеваемый в формуле квазинейтральности (12.12), может быть оценен как

$$\Delta t \gg \omega_{pe}^{-1}.$$

Определение пространственного масштаба усреднения потребовало бы привлечения определенных сведений из термодинамики (см. раздел 5), да и формальная сторона дела связана с достаточно сложной вычислительной процедурой. Но сам результат может быть продемонстрирован без особых затруднений на основании одних лишь соображений размерности. Наиболее характерная величина размерности скорости в микродинамике любой текучей среды (жидкость, газ, плазма) — тепловая скорость частиц, в данном случае — электронов: $v_{Te} = (k_B T/m)^{1/2}$. Отношение ее к плазменной частоте как раз и дает величину размерности длины, а в теории плазмы именно эта величина оказывается характерным масштабом разделения зарядов. Ее называют *дебаевской длиной* или *дебаевским радиусом*, по имени голландского физика П. Дебая (1884–1966):

$$r_D = \frac{v_{Te}}{\omega_{pe}} = \sqrt{\frac{\varepsilon_0 k_B T}{n e^2}}. \quad (12.15)$$

Масштаб пространственного усреднения в (12.12) должен быть существенно больше r_D . Но существует еще один характерный

масштаб нарушения квазинейтральности, упомянутый выше, — это просто характерное расстояние между частицами $n^{-1/3}$. Соотношение между этими двумя масштабами оказывается принципиально важным для микроскопических свойств плазмы. Дело в том, что если $r_D \gg n^{-1/3}$, то, в отличие от газа, где взаимодействие частиц сводится к достаточно редким столкновениям, частицы плазмы достаточно эффективно взаимодействуют в соответствии с законом Кулона.

В обратном случае кулоновское поле экранировано вследствие квазинейтральности, и тогда микродинамика плазмы лишь количественно отличается от газовой. Обычно из названных параметров образуют безразмерную комбинацию, которую называют *числом частиц в дебаевской сфере* или просто *дебаевским числом*:

$$N_D = \frac{4\pi}{3} n r_D^3 = \frac{4\pi}{3} \frac{(\varepsilon_0 k_B T)^{3/2}}{\sqrt{n} e^3}. \quad (12.16)$$

Если $N_D \gg 1$, то в сфере действия кулоновского потенциала каждой частицы оказывается большое число других частиц. В этом случае коллективные эффекты, т. е. согласованные движения ансамблей частиц, превалируют над столкновениями. Эта ситуация очень далека от газа, и даже от жидкой фазы, но напоминает скорее твердое тело — кристалл. Таким образом, и в макро-, и в микродинамике плазма, будучи более высокотемпературной фазой, нежели газ, обнаруживает свойства, присущие конденсированным средам, что и дает основания говорить о четвертом состоянии вещества. Из (12.16) можно видеть, что условие $N_D \gg 1$ может быть выполнено для плазмы либо достаточно горячей, либо достаточно разреженной. Поэтому удовлетворяют этому условию объекты столь различной природы, как «термоядерная» плазма в лабораторных условиях и холодная, но очень редкая плазма межзвездной среды.

12.4. Электромагнитные волны в плазме

В данном параграфе будет рассмотрена поперечная электромагнитная волна в плазме. Как мы вскоре убедимся, плазма для такой волны — диспергирующая среда, т. е. закон дисперсии $\omega(k)$ будет отличен от линейного. Поле волны представим в стандартной форме

$$\mathbf{E} = \mathbf{E}_0 \exp[i(\mathbf{k}\mathbf{r} - \omega t)].$$

В горячей плазме (1–10 кэВ или, что то же, $(10\text{--}100) \cdot 10^6$ К) столкновения частиц настолько редки, что безусловно доминируют коллективные эффекты. Проще говоря, мы можем рассматривать отклик ансамбля частиц на быстропеременное электромагнитное поле, совершенно не принимая во внимание какой-либо

диссипации. Относительно частоты осцилляций мы предположим, что за период ионы существенно сместиться не могут и успевают отреагировать на изменение поля лишь электроны. Напишем уравнение движения электронов в поле волны

$$m \frac{d^2 \mathbf{r}_E}{dt^2} = -e \mathbf{E}_0 \exp[i(\mathbf{k}\mathbf{r} - \omega t)]. \quad (12.17)$$

Здесь под $\mathbf{r}_E = (x_E, y_E, z_E)$ подразумевается лишь осцилляция электрона в волне. (Мы пренебрегли в (12.17) членом $[\mathbf{v} \mathbf{B}]$, поскольку в случае нерелятивистского движения частицы он дал бы малую поправку к правой части порядка v/c .) Уравнение (12.17) легко интегрируется:

$$\mathbf{r}_E = \frac{e}{m\omega^2} \mathbf{E}_0 \exp[i(\mathbf{k}\mathbf{r} - \omega t)]. \quad (12.18)$$

Далее в соответствии с соотношениями (2.8), (2.9) запишем общее выражение для диэлектрической проницаемости

$$\varepsilon_0 \varepsilon \mathbf{E} \equiv \mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}, \quad (12.19)$$

в котором поляризация среды представляется объемной плотностью дипольного момента

$$\mathbf{P} = \sum_{\alpha} q_{\alpha} \mathbf{r}_{\alpha}.$$

Для его вычисления воспользуемся решением (12.18). Пусть n — число электронов в единице объема, тогда

$$\mathbf{P} = -n e \mathbf{r}_E = -\frac{n e^2}{m \varepsilon_0 \omega^2} \mathbf{E},$$

откуда и следует

$$\varepsilon = 1 - \frac{\omega_{pe}^2}{\omega^2}, \quad (12.20)$$

где величина ω_{pe} — уже известная нам ленгмюровская или плазменная частота (12.14).

Теперь воспользуемся выражением (11.3) для показателя преломления электромагнитной волны в среде:

$$N^2 \equiv \frac{k^2 c^2}{\omega^2} = \varepsilon \mu. \quad (12.21)$$

(Поскольку n — стандартное обозначение концентрации, нам пришлось обозначить показатель преломления буквой N .) Положим опять $\mu = 1$. Получаем закон дисперсии

$$\frac{k^2 c^2}{\omega^2} = 1 - \frac{\omega_{pe}^2}{\omega^2}, \quad (12.22)$$

или, что то же,

$$\omega^2 = \omega_{pe}^2 + k^2 c^2. \quad (12.23)$$

Дисперсия (12.23) представлена на рис. 12.4а. Видно, что в пределе высоких частот мы приближаемся к «вакуумному» линейному закону $\omega = kc$, но при $k \rightarrow 0$ ситуация иная; в частности, распространение в плазме электромагнитных волн с $\omega < \omega_{pe}$ невозможно.

Закон дисперсии (12.23) имеет место не только в плазменном случае, но и, скажем, при распространении волн в волноводах и, что для нас более существенно, он справедлив для любого вещества при достаточно высоких частотах (при этом ω_{pe} определяется формально через концентрацию электронов в веществе). Это не так уж и удивительно, поскольку чисто электронный отклик на поле высокой частоты вполне естествен. Даже у водорода отношение масс $M_i/m_e \approx 1840$, а отклик на поле (см. (12.18)) обратно пропорционален массе. Но если для плазмы порядок минимальной частоты такого поля определяет именно ω_{pe} , то в общем случае ответ не универсален, поскольку не ясна заранее роль диссипативных эффектов, которыми мы пренебрегли.

Опираясь на соотношение (12.22), мы можем рассмотреть эффект отражения волны от плазмы критической плотности. Обратимся к рис. 12.4б. Пусть в плазму с монотонно нарастающей плотностью (или электронной концентрацией) $n(x)$ посылается электромагнитная волна. Следуя в направлении роста концентрации, волна сохраняет частоту ω_0 , но длина волны или волновое число при этом сохраняться не могут. При $n > n_{кр} = \varepsilon_0 \omega_0^2 m e^{-2}$ эта волна вообще существовать не может. Но она не может и исчезнуть, поскольку никаких механизмов диссипации мы в рассмотрение не вводили. Обратим внимание на то, что в области критической концентрации $n_{кр}$, как следует из (12.23), $k = 0$, а при меньших концентрациях, согласно той же формуле, k определено с точностью до знака. Отсюда и следует, что наша волна должна просто отразиться от области критической концентрации и выйти в обратном направлении.

Именно на этом эффекте основана дальняя радиосвязь на коротких волнах — точнее, на многократном отражении волны от поверхности Земли и ионосферного плазменного «зеркала» (рис. 12.5). А вот на УКВ, как известно, связь возможна только в пределах прямой видимости. Тем и отличаются ультракороткие волны от коротких, что у них частота выше максимальной плазменной частоты для ионосферы.

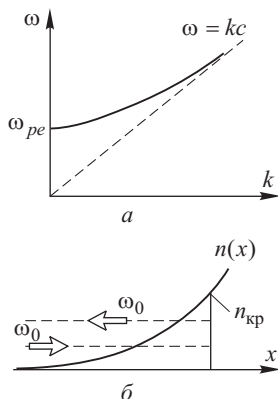


Рис. 12.4

Мы уже упоминали в гл. 11, что для достаточно жестких (коротковолновых) рентгеновских лучей коэффициент преломления

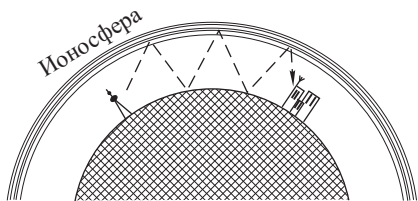


Рис. 12.5

любого вещества оказывается близок к единице. Имелся в виду закон дисперсии (12.22). Заметим еще раз, что соответствующий такой дисперсии коэффициент преломления меньше единицы. Поэтому и полное отражение оказывается не внутренним, как в оптике, а внешним. Это

используется для создания «рентгеноводов». В отличие от световода — стеклянной нити — рентгеновод представляет собой полую кварцевую трубочку. Принцип его работы тот же, что и у световода, — многократное полное отражение. И еще один эффект такого рода — металлический блеск. Здесь уже дело касается квантовой плазмы электронов проводимости. (Хотя, конечно, следует заметить, что чисто плазменным эффект отражения от металлов оказывается лишь в инфракрасной области — будь это так же и в оптическом диапазоне, все металлы были бы белыми, включая золото и медь).

Нарисовав картинку $n(x)$ и следя за распространением волны, мы молчаливо предполагаем, что длина волны $2\pi/k$ существенно меньше характерного масштаба неоднородности. Но это предположение нарушается, по крайней мере, вблизи $n_{кр}$, где $2\pi/k \rightarrow \infty$. В этой области осциллирующее поле сшивается с экспоненциально затухающим решением при одном и том же характерном масштабе c/ω_{pe} . Называют его *бесстолкновительной скин-длиной*, которая, как нетрудно видеть, зависит только от концентрации электронов, но не от частоты поля или проводимости среды.

В заключение заметим, что в плазме могут распространяться не только поперечные, но и продольные электромагнитные волны. Действительно, при отличной от единицы диэлектрической проницаемости уравнение (10.17) можно переписать в виде

$$[\mathbf{k} [\mathbf{k} \mathbf{E}_0]] = \mathbf{k}(\mathbf{k} \mathbf{E}_0) - k^2 \mathbf{E}_0 = -\mu_0 \varepsilon \varepsilon_0 \omega^2 \mathbf{E}_0.$$

Предположим, что $\mathbf{k} \parallel \mathbf{E}$. Тогда левая часть оказывается в точности равной нулю, и получившееся простое уравнение допускает два решения — тривиальное $\mathbf{E} \equiv 0$ и содержательное

$$\varepsilon(\omega, \mathbf{k}) = 0. \quad (12.24)$$

Уравнение (12.24) представляет общий вид дисперсионного уравнения для продольных волн. В частности, приравнявая нулю

диэлектрическую проницаемость (12.20), получаем $\omega = \omega_{pe}$ — это не что иное, как ленгмюровские колебания, рассмотренные в предыдущем параграфе на основании более простой модели.

Задачи

1. Плоский конденсатор заполнен плазмой со средней концентрацией электронов и ионов n_0 и температурой T . Расстояние между пластинами a , разность потенциалов U . Пренебрегая током в плазме и считая $eU \ll k_B T$, определить пространственную зависимость потенциала между обкладками.

Решение. Из механики известна барометрическая формула Больцмана

$$n = n_0 \exp\left(-\frac{mgx}{k_B T}\right),$$

где m — масса отдельной молекулы, g — ускорение силы тяжести, ось Ox направлена вертикально вверх. Эта формула — частный случай гораздо более общего закона, который называется статистическим распределением Больцмана. В случае произвольной потенциальной функции $U(\mathbf{r})$ он имеет вид

$$n = n_0 \exp\left(-\frac{U}{k_B T}\right).$$

Распределение Больцмана для электронов и ионов в нашем случае можно линеаризовать:

$$\begin{aligned} n_i &= n_0 \exp\left(-\frac{e\varphi}{k_B T}\right) \approx n_0 \left(1 - \frac{e\varphi}{k_B T}\right); \\ n_e &= n_0 \exp\left(\frac{e\varphi}{k_B T}\right) \approx n_0 \left(1 + \frac{e\varphi}{k_B T}\right), \end{aligned}$$

а затем подставить в уравнение Пуассона:

$$\frac{d^2 \varphi}{dx^2} = \frac{e}{\varepsilon_0} (n_e - n_i) \approx \frac{2ne^2}{\varepsilon_0 k_B T} \equiv \frac{2}{r_D^2} \varphi.$$

Уравнение получилось линейным и однородным, поэтому его решение в точности соответствует формуле (8.9), только в данном случае корни характеристического уравнения вещественны:

$$\varphi(x) = \varphi_1 \exp\left(\frac{\sqrt{2}}{r_D} x\right) + \varphi_2 \exp\left(-\frac{\sqrt{2}}{r_D} x\right).$$

Принимая во внимание граничные условия

$$x = 0, \quad \varphi = 0; \quad x = a, \quad \varphi = U,$$

приходим к следующему ответу:

$$\varphi(x) = U \frac{\operatorname{sh}(x\sqrt{2}/r_D)}{\operatorname{sh}(a\sqrt{2}/r_D)}.$$

Таким образом, на этой относительно простой одномерной задаче мы продемонстрировали, что из объединения электростатики с тепловым движением частиц в плазме действительно рождается пространственный масштаб r_D — дебаевская длина.

2. Площадь электродов плоского газонаполненного диода $S = 10 \text{ см}^2$, межэлектродный зазор $a = 10 \text{ см}$. В режиме несамостоятельного разряда ток насыщения $I_s = 10^{-6} \text{ А}$. Какое количество элементарных зарядов того и другого знака создается ежесекундно внешним ионизатором в 1 см^3 ?

Ответ: $N_{\mp} = \frac{I_s}{eSa} \approx 6 \cdot 10^9 \text{ см}^{-3} \cdot \text{с}^{-1}$.

3. Мощный источник тока создает в тонкой цилиндрической плазменной оболочке ток $I = 5 \cdot 10^6 \text{ А}$, параллельный оси и равномерно распределенный по азимуту. Внутри оболочки предварительно создано магнитное поле $B_0 = 0,1 \text{ Тл}$. Начальный радиус цилиндра $R_0 = 20 \text{ см}$. В дальнейшем под действием тока оболочка сжимается по радиусу. Считая ее идеально проводящей, оценить, при каком радиусе ускорение оболочки поменяет знак.

Ответ: $R = \frac{2\pi B_0 R_0^2}{\mu_0 I} \approx 4 \text{ мм}$.

4. В плазме находится дипольный излучатель, на который подается переменное напряжение с частотой ω . При какой концентрации электронов плазмы он перестает излучать электромагнитные волны?

Ответ: $n > \frac{\varepsilon_0 m \omega^2}{e^2}$.

5. Электромагнитная волна падает на поверхность плазмы, концентрация которой растет вглубь, а на поверхности много меньше критической. Угол падения равен θ . Какой концентрации плазмы n соответствует поверхность, от которой произойдет отражение? Будет ли угол отражения равен углу падения?

Ответ: угол отражения будет равен углу падения, $n = \frac{\varepsilon_0 m \omega^2}{e^2} \cos^2 \theta$.

Раздел третий

ФИЗИКА КОЛЕБАНИЙ И ВОЛН. ВОЛНОВАЯ ОПТИКА

ВВЕДЕНИЕ

Раздел посвящен изучению колебательно-волновых закономерностей, пронизывающих всю современную физику. В механике и акустике, в радиофизике и оптике, в квантовой физике и физике твердого тела — всюду мы сталкиваемся с колебаниями и волнами. Здесь особенно ярко находит отражение современная тенденция сближения и взаимопроникновения различных областей науки. Единый подход к изучению колебаний и волн различной физической природы, основанный на общности уравнений, описывающих колебательно-волновые закономерности, позволяет выявить глубокие связи между различными, на первый взгляд, явлениями. Этот подход оказался плодотворным по крайней мере в двух отношениях. Во-первых, он позволил обобщить и распространить на различные области физики идеи и методы, возникшие первоначально в какой-либо одной области (например, распространить идею Габора электронной голографии сначала на оптическую, а затем на радио- и акустическую голографию). Во-вторых, он позволил выявить плодотворные колебательно-волновые аналогии: идеи и методы решения задач физики колебаний перенести на задачи волновые и наоборот (например, фильтрация временных сигналов-колебаний и пространственная фильтрация волн). Знания и интуиция, которые мы приобретаем при изучении колебаний и волн одной физической природы естественным образом используются при изучении явлений казалось бы совершенно иной природы. Например, изучая поведение волн света, мы начинаем глубже понимать, как ведут себя «волны вероятности», которыми оперирует квантовая физика, или волны звука, которые изучаются в акустике. Таким образом, изучая колебания и волны, мы будем обращать внимание не только на то, что «волнуется» и что «колеблется», а главным образом на то, как и почему происходят колебания и возникают волны.

Глава 1

КИНЕМАТИКА КОЛЕБАНИЙ

1.1. Гармонические колебания

Колебание — это более или менее регулярно повторяющийся процесс. Таково очень нестрогое, «качественное» определение понятия «колебание». Можно привести множество примеров колебательных процессов, относящихся к различным областям физики (и не только физики). Колеблется маятник часов; колеблется груз, подвешенный на пружине; колеблется взволнованная поверхность воды и гитарная струна; колеблется заряд на пластинах конденсатора и магнитное поле в катушке индуктивности колебательного контура; более или менее периодически изменяется температура воздуха (зимой холоднее — летом теплее) и количество автомобилей на улицах города (больше в часы пик — меньше поздней ночью); периодически меняется экономическая ситуация в жизни общества: кризисные явления сменяются подъемом экономики. Колеблется давление (или плотность воздуха), вызывая колебания ушной мембраны — и мы слышим голос певца на оперной сцене. Таких примеров можно привести как угодно много. В настоящей главе мы не будем выяснять причины возникновения колебаний в той или иной физической системе. Здесь мы лишь познакомимся с наиболее часто встречающимися простейшими видами колебательных движений, основными характеристиками колебательных процессов, с математическим способом описания колебаний. Оставляя пока в стороне физическую природу колеблющейся величины, будем обозначать ее буквой S , подчеркивая при этом, что любой колебательный процесс развивается во времени: $S = S(t)$. (Величиной S может быть и угол отклонения маятника часов, и величина растяжения-сжатия пружины при колебаниях груза, и изменяющийся заряд на пластине конденсатора, и плотность воздуха при распространении звуковой волны. В общем, S — любая колеблющаяся величина.)

Наиболее простая функция, описывающая периодический процесс, имеет вид

$$S(t) = a \cos(\omega t + \varphi), \quad (1.1)$$

где a , ω и φ — константы. Функциональную зависимость (1.1) физики называют *уравнением гармонических колебаний*, ее гра-

фик изображен на рис. 1.1. Она хороша, разумеется, не только потому, что имеет довольно простой математический вид. Более существенно то обстоятельство, что реальные колебания во многих физических системах хорошо описываются этой функцией, т. е. близки к гармоническим колебаниям. Имеются и другие веские причины, по которым гармонические колебания играют исключительную роль в современной физике. Об этом мы еще будем говорить.

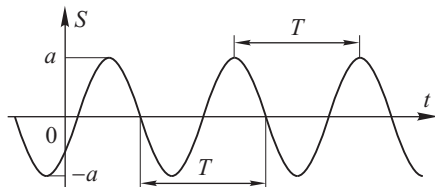


Рис. 1.1

Легко проверить, что функция (1.1) является периодической, т. е. для любого момента времени t имеет место равенство $S(t) = S(t + T)$, где T называется периодом колебаний: $T = (2\pi/\omega)n$.

Обычно *периодом колебаний* называют минимальное значение T ($n = 1$)

$$T = \frac{2\pi}{\omega}. \quad (1.2)$$

Величина ω называется *круговой частотой*. Положительная константа a — *амплитуда колебания* (это максимальное отклонение величины S от равновесного значения $S = 0$). Аргумент косинуса в (1.1) — угол Φ , выраженный в радианах, называется *фазой колебания*:

$$\Phi = \omega t + \varphi, \quad (1.3)$$

а значение Φ при $t = 0$, т. е. величину φ , называют *начальной фазой*. Соотношение (1.3) — это линейная связь между фазой колебания Φ и круговой частотой ω , из которой следует

$$\omega = \dot{\Phi},$$

т. е. круговая частота — это производная фазы Φ по времени.

Число колебаний в секунду

$$\nu = \frac{1}{T} \quad (1.4)$$

называют *циклической частотой* (иногда просто частотой). Единица циклической частоты (1/с) — *герц* (Гц). Например, $\nu = 50 \text{ с}^{-1}$ — колебание с частотой 50 Гц. Очевидно,

$$\omega = 2\pi\nu \quad (1.5)$$

(колебанию с частотой 1 Гц соответствует изменение фазы — угла поворота, — равное 2π в секунду).

Сделаем еще одно важное замечание, касающееся гармонических колебаний: очевидно, ни один реальный процесс не может,

строго говоря, описываться функцией вида (1.1). Ведь синусоида не имеет ни начала, ни конца — и хотя бы поэтому бесконечно длящийся колебательный процесс неосуществим.

1.2. Векторная интерпретация и комплексное представление гармонических колебаний.

Фазовая плоскость

Векторная интерпретация. При решении многих задач, связанных с изучением колебательных процессов, удобно использовать следующую геометрическую интерпретацию гармонических колебаний.

Гармоническое колебание (1.1) будем изображать вектором \mathbf{S} , длина которого равна амплитуде колебания a , а угол между вектором и горизонтальной осью x — начальной фазе колебания φ (рис. 1.2).

При этом частота ω гармонического колебания предполагается заданной. Смысл такого представления состоит в следующем. Вообразим, что вектор \mathbf{S} вращается вокруг точки O с угловой скоростью ω в положительном направлении, т.е. против часовой стрелки, а мы сделали мгновенную фотографию в момент времени $t = 0$, когда угол наклона вектора $\Phi(t) = \omega t + \varphi$ равен φ . Заметим, что проекция вектора \mathbf{S} на ось x при вращении изменяется по закону $x(t) = a \cos(\omega t + \varphi)$, т.е. совершает гармонические колебания.

Геометрическое изображение гармонического колебания $S(t)$ в виде вектора удобно использовать при решении задачи сложения колебаний. Пусть мы имеем две *скалярные величины* S_1 и S_2 , изменяющиеся по гармоническому закону с одинаковой частотой ω :

$$S_1(t) = a_1 \cos(\omega t + \varphi_1), \quad S_2(t) = a_2 \cos(\omega t + \varphi_2).$$

Необходимо найти колебание $S(t)$ (скалярную величину), являющуюся суммой колебаний $S_1(t)$ и $S_2(t)$:

$$S(t) = S_1(t) + S_2(t). \quad (1.6)$$

Изобразим колебания $S_1(t)$ и $S_2(t)$ в виде векторов \mathbf{S}_1 и \mathbf{S}_2 (рис. 1.3). Изобразим также суммарный вектор \mathbf{S} . Векторы \mathbf{S}_1 , \mathbf{S}_2 и \mathbf{S} образуют треугольник, причем внешний угол треугольника $\Delta\varphi$ равен разности фаз колебаний S_2 и S_1 . Представим себе, что векторы \mathbf{S}_1 и \mathbf{S}_2 вращаются с одной и той же угловой скоростью против часовой стрелки.

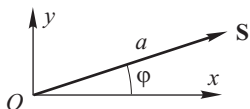


Рис. 1.2

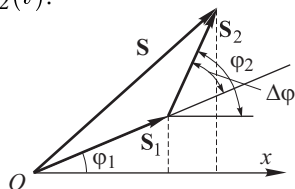


Рис. 1.3

Ясно, что угол $\Delta\varphi$ между векторами \mathbf{S}_1 и \mathbf{S}_2 остается при таком вращении неизменным, и суммарный вектор \mathbf{S} повернется через время t (как и \mathbf{S}_1 и \mathbf{S}_2) на угол ωt — т. е. весь треугольник векторов вращается как целое. Причем очевидно, что проекция суммарного вектора \mathbf{S} на ось x в произвольный момент времени t равна сумме проекций векторов \mathbf{S}_1 и \mathbf{S}_2

$$a \cos(\omega t + \varphi) = a_1 \cos(\omega t + \varphi_1) + a_2 \cos(\omega t + \varphi_2),$$

здесь a — длина вектора \mathbf{S} , а φ — его угол наклона при $t = 0$, т. е. начальная фаза суммарного колебания.

Итак, сумма гармонических колебаний одинаковой частоты является гармоническим колебанием той же частоты. Амплитуда a суммарного колебания может быть найдена из треугольника векторов (по теореме косинусов):

$$a^2 = a_1^2 + a_2^2 + 2a_1 a_2 \cos \Delta\varphi \quad (1.7)$$

где $\Delta\varphi = \varphi_2 - \varphi_1$ — разность фаз складываемых колебаний.

Заметим, что во многих случаях измеряемой величиной является величина, пропорциональная квадрату амплитуды колебания, которая называется *интенсивностью* $I = a^2$.

Из (1.7) следуют важные частные случаи:

1. Амплитуды слагаемых колебаний равны: $a_1 = a_2 = a_0$. Тогда:

$$a^2 = 2a_0^2(1 + \cos \Delta\varphi). \quad (1.8)$$

2. Разность фаз слагаемых колебаний равна $2\pi n$. Такие колебания называются *синфазными*. При $\Delta\varphi = 2\pi n$ находим из (1.7): $a = a_1 + a_2$.

Амплитуда суммарного колебания при этом максимальна и равна сумме амплитуд слагаемых колебаний. (При $\Delta\varphi = 0$ на векторной диаграмме мы получим два коллинеарных вектора \mathbf{S}_1 и \mathbf{S}_2 и длина суммарного вектора просто равна сумме длин слагаемых векторов.)

3. Разность фаз слагаемых колебаний равна $(2n + 1)\pi$. Такие колебания называют *противофазными*. При $\Delta\varphi = (2n + 1)\pi$ получаем из (1.7): $a^2 = (a_1 - a_2)^2$.

При равных амплитудах $a_1 = a_2$ имеем в этом случае $a = 0$, т. е. колебания «гасят» друг друга. На векторной диаграмме получаем два противоположно направленных вектора \mathbf{S}_1 и \mathbf{S}_2 .

На рис. 1.4 представлен график зависимости интенсивности I суммарного колебания от разности фаз $\Delta\varphi$ слагаемых колебаний (1.7).

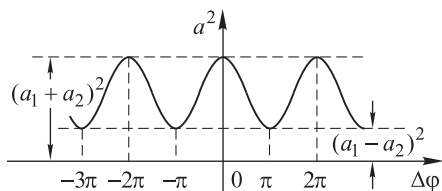


Рис. 1.4

Фазовая плоскость. При изучении колебаний полезным оказывается изображение колебательного процесса на *фазовой плоскости*. Координаты на фазовой плоскости — величина S (ось абсцисс) и производная по времени \dot{S} (ось ординат). В каждый момент времени t с помощью соотношений

$$\begin{aligned} S &= a \cos(\omega t + \varphi), \\ \dot{S} &= -a\omega \sin(\omega t + \varphi) \end{aligned} \quad (1.9)$$

можно найти значения S и \dot{S} и, следовательно, положение точки M на фазовой плоскости, изображающей состояние колебательной системы в данный момент времени. С течением времени величины S и \dot{S} изменяются (в соответствии с (1.9)), изменяется, следовательно, и положение изображающей точки, которая описывает так называемую *фазовую траекторию*. Уравнение фазовой траектории гармонического колебательного процесса легко найти, исключая из уравнения (1.9) время t :

$$\frac{S^2}{a^2} + \frac{\dot{S}^2}{(\omega a)^2} = 1. \quad (1.10)$$

Согласно (1.10), фазовая траектория гармонического колебания — эллипс с полуосями a и ωa . Из (1.9) ясно, что изображающая точка M движется по фазовой траектории по часовой стрелке.

На рис. 1.5 *a* изображено семейство фазовых траекторий, отличающихся амплитудой колебания.

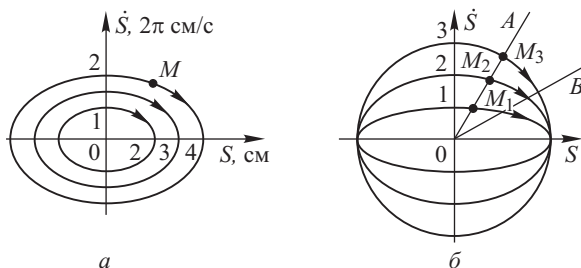


Рис. 1.5

Комплексная форма записи гармонических колебаний.

В дальнейшем при исследовании колебаний мы часто будем использовать комплексную форму записи гармонических колебаний. Напомним, что комплексное число z может быть записано в виде

$$z = x + iy, \quad (1.11)$$

$i = \sqrt{-1}$ — мнимая единица, x и y — собственно действительная

и мнимая части комплексного числа, $x = \operatorname{Re} z$, $y = \operatorname{Im} z$. Величина

$$|z| = \sqrt{x^2 + y^2}$$

называется модулем комплексного числа, а $\varphi = \operatorname{arctg}(y/x)$ — аргументом комплексного числа z . Комплексные числа принято изображать на так называемой *комплексной плоскости*: комплексному числу z поставим в соответствие точку M , абсцисса которой равна действительной части комплексного числа x , а ордината — мнимой части y . Вектор, проведенный из начала координат в точку M , является графическим изображением комплексного числа z . Очевидно, длина вектора равна модулю комплексного числа, а угол наклона вектора φ определяет аргумент комплексного числа (рис. 1.6).

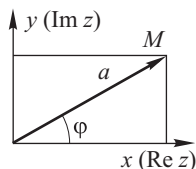


Рис. 1.6

Комплексные числа складываются по правилу сложения векторов: сумма двух комплексных чисел $z_1 = x_1 + iy_1$ и $z_2 = x_2 + iy_2$ есть комплексное число $z = x + iy$, где $x = x_1 + x_2$, $y = y_1 + y_2$ (именно такие проекции имеет вектор, равный сумме векторов $S_1(x_1, y_1)$ и $S_2(x_2, y_2)$).

Вспомним теперь о векторном изображении гармонического колебания. Мы отмечали ранее, что гармоническое колебание можно представить как проекцию вращающегося с угловой скоростью ω вектора S на ось абсцисс $S(t) = a \cos(\omega t + \varphi)$. При этом его проекция на ось ординат меняется по закону $S_1(t) = a \sin(\omega t + \varphi)$. Полагая, что $S(t)$ есть действительная часть комплексного числа z , а $S_1(t)$ — его мнимая часть, запишем

$$z = a \cos(\omega t + \varphi) + ia \sin(\omega t + \varphi). \quad (1.12)$$

Или, используя формулу Эйлера $e^{i\alpha} = \cos \alpha + i \sin \alpha$:

$$z = ae^{i(\omega t + \varphi)}. \quad (1.13)$$

Последнее выражение и представляет собой *комплексную форму записи гармонического колебания*, имеющего амплитуду a , частоту ω и начальную фазу φ . Реальный колебательный процесс $S(t)$ связан с комплексной формой записи $z(t)$ очевидным равенством

$$S(t) = \operatorname{Re} z(t). \quad (1.14)$$

Перепишем выражение (1.13), выделив множитель, зависящий от времени $e^{i\omega t}$:

$$z = ae^{i\varphi} e^{i\omega t} = ce^{i\omega t}.$$

Комплексное число $c = ae^{i\varphi}$ называется *комплексной амплитудой колебания*. Она содержит информацию как об амплитуде колебания, так и о начальной фазе.

Числа z_1 и z_2 называются *комплексно-сопряженными*, если их действительные части равны, а мнимые отличаются знаком: $z_1 = x + iy$, $z_2 = x - iy$. Комплексно-сопряженные числа имеют одинаковые модули и аргументы противоположного знака: $z_1 = ae^{i\varphi}$, $z_2 = ae^{-i\varphi}$. Тот факт, что числа z_1 и z_2 являются комплексно-сопряженными принято записывать в виде: $z_1 = z_2^*$.

1.3. Некоторые важные задачи сложения гармонических колебаний

Рассмотрим задачи сложения гармонических колебаний, необходимые как для дальнейшего углубленного изучения общих вопросов теории колебаний, так и для решения конкретных волновых задач.

1. Сумма N колебаний, фазы которых составляют арифметическую прогрессию

$$S = \sum_{n=1}^N a_0 e^{i(n-1)\alpha} = a_0 + a_0 e^{i\alpha} + a_0 e^{i2\alpha} + \dots \quad (1.15)$$

Построим векторную диаграмму (рис. 1.7). Первое слагаемое в (1.15) изображается вектором длины a_0 , направленным вдоль действительной оси; второе слагаемое $a_0 e^{i\alpha}$ — вектор той же длины, составляющий угол α с действительной осью; третий вектор составляет с действительной осью угол 2α (и угол α со вторым вектором) и т.д. Суммарное колебание — вектор S , проведенный из начала 1-го вектора в конец последнего N -го вектора. Складываемые векторы образуют часть равностороннего многоугольника, вписанного в окружность, радиус которой равен $R = a_0 / (2 \sin(\alpha/2))$.

Рис. 1.7

Из векторной диаграммы легко найти модуль вектора S :

$$|S| = 2R \left| \sin \frac{2\pi - N\alpha}{2} \right| = 2R \left| \sin \frac{N\alpha}{2} \right| = a_0 \left| \frac{\sin(N\alpha/2)}{\sin(\alpha/2)} \right|, \quad (1.16)$$

2. Спираль Френеля — векторная диаграмма для вычисления интеграла

$$S = \int_0^{x_0} e^{i\beta x} dx.$$

Разобьем весь отрезок интегрирования $[0, x_0]$ на N малых интервалов Δx ($x_0 = N\Delta x$). Начало n -го интервала находится в точке $x_n = n\Delta x$. Искомый интеграл приближенно равен сумме N векторов $\Delta x e^{i\beta n\Delta x}$, фазы которых образуют арифметическую прогрессию. Радиус окружности R , часть которой (при $\Delta x \rightarrow 0$) составляет сумма $\sum e^{i\beta n\Delta x} \cdot \Delta x$, находится так: при $\beta n\Delta x = \pi$ n -й вектор s_n составляет угол π с действительной осью (с вектором s_0 , направленным вдоль оси x) и следовательно, находится на верхнем конце вертикального диаметра (рис. 1.8); суммарная длина цепочки векторов $n\Delta x$ равна длине полуокружности πR : $n\Delta x = \pi R$, откуда имеем $R = 1/\beta$.

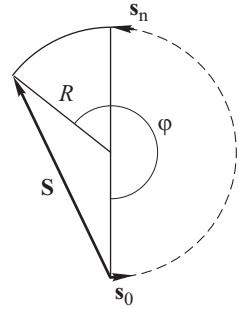


Рис. 1.8

Так как дуга окружности, стягиваемая суммарным вектором S , имеет длину x_0 , то угол φ равен: $\varphi = x_0/R$; окончательно получаем:

$$S = \frac{2}{\beta} \left| \sin \frac{x_0\beta}{2} \right| e^{ix_0\beta/2}. \quad (1.17)$$

Как видно из векторной диаграммы (и из формулы (1.17)), величина интеграла является периодической функцией верхнего предела, изменяясь от максимального значения (равного $2/\beta$, когда $x_0 = (\pi/2\beta)(2n \pm 1)$) до нуля при $x_0 = 2\pi n/\beta$. Ясно, что при стремлении верхнего предела к бесконечности, величина исследуемого интеграла не стремится к определенному пределу: конец суммарного вектора с ростом x_0 будет двигаться по окружности радиуса $1/\beta$. Представим себе теперь, что длина каждого последующего элементарного вектора в цепочке векторов $\Delta x e^{i\beta n\Delta x}$ меньше длины предыдущего вектора; пусть, например, β является комплексным числом $\beta = \beta' + i\beta''$, где $\beta'' > 0$. Тогда длина n -го вектора есть $\Delta x e^{-\beta'' n\Delta x}$, т.е. с ростом n длины элементарных векторов экспоненциально уменьшаются. Ясно, что если имеется сколь угодно малое затухание амплитуд складываемых колебаний (т.е. имеется сколь угодно малое $\beta'' > 0$), то цепочка векторов $\Delta x e^{i\beta n\Delta x}$ при $\Delta x \rightarrow 0$ образует уже не дугу окружности, а скручивающуюся спираль, фокус которой (точка, в которую эта спираль скручивается) находится в центре первоначальной окружности (рис. 1.9).

Значение интеграла $\int_0^\infty e^{i\beta x} dx$ определяется тогда вектором, проведенным из точки O в фокус спирали и следовательно,

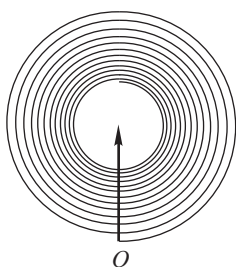


Рис. 1.9

$$\int_0^\infty e^{i\beta x} dx = \frac{i}{\beta} \quad (1.18)$$

(множитель $i = e^{i\pi/2}$ показывает, что суммарный вектор составляет угол $\pi/2$ с действительной осью). Каждый раз, когда какая-либо физическая задача приводит нас к интегралу вида (1.18), физически правильное его значение получается, если предположить наличие бесконечно малого затухания $\beta'' > 0$ и затем в полученном результате перейти к пределу при $\beta'' \rightarrow 0$, в результате мы получаем равенство (1.18). Изображенная на рис. 1.9 векторная диаграмма называется *спиралью Френеля*.

3. Спираль Корню — векторная диаграмма для вычисления интеграла

$$\int_0^{x_0} e^{i\beta x^2} dx. \quad (1.19)$$

Разбив отрезок интегрирования на малые интервалы Δx (так что $x_n = n\Delta x$) мы заменим интеграл (1.19) суммой

$$S \approx \sum_{n=0}^N e^{i\beta n^2 (\Delta x)^2} \Delta x \quad \left(N = \frac{x_0}{\Delta x} \right).$$

В отличие от предыдущего примера, здесь речь идет о сложении колебаний, фазы которых изменяются по квадратичному закону (пропорционально n^2). Если в предыдущем примере угол между двумя последовательными векторами в цепочке остается постоянной величиной ($\Delta\varphi = \beta(n+1)\Delta x - \beta n\Delta x = \beta\Delta x$), то в данном примере этот угол с ростом n растет линейно:

$$\delta\varphi_n = \beta[\delta x(n+1)]^2 - \beta[\delta x \cdot n]^2 = \beta(2n+1)(\delta x)^2.$$

Таким образом, закручивание цепочки векторов происходит здесь неравномерно: сначала (пока $\varphi_n = \beta(\Delta x)^2 n^2 \ll \pi$) кривая очень пологая и векторы $s_n = e^{i\beta(\Delta x)^2 n^2}$ выстраиваются (почти) в прямую линию — (элементарные колебания складываются почти синфазно (рис. 1.10). Затем, при увеличении n угол между двумя последовательными векторами становится весьма заметным — кривая начинает заметно скручиваться. Длина дуги первого витка x_1 , отмеченного на рис. 1.10 жирной линией, определяется

условием $\beta x^2 = 2\pi$ (n -й вектор \mathbf{s}_n коллинеарен с первым вектором \mathbf{s}_0). Откуда

$$x_1 = \sqrt{\frac{2\pi}{\beta}}. \quad (1.20)$$

Продолжая дальше прибавлять элементарные векторы, мы будем проходить все более мелкие витки скручивающейся спирали. Конечно, пока речь идет о дискретной сумме колебаний, получается не дуга, а скручивающаяся ломаная, которая переходит в плавную кривую при $\Delta x \rightarrow 0$, при этом сумма переходит в интеграл (1.19).

Очевидно, значение интеграла определяется вектором, проведенным из точки O в точку A , положение которой определяется условием: длина дуги, которую стягивает суммарный вектор, равна x_0 .

С ростом x_0 конец суммарного вектора перемещается по виткам скручивающейся спирали и при $x_0 \rightarrow \infty$ попадает в точку, называемую *фокусом спирали*.

Векторная диаграмма позволяет понять, как изменяется значение интеграла (1.19) при изменении верхнего предела. Пока конец суммарного вектора перемещается по первому витку, т.е. x_0 изменяется в пределах: $0 < x_0 < x_1 = \sqrt{2\pi/\beta}$, изменение x_0 приводит к заметному изменению интеграла (длина результирующего вектора на диаграмме).

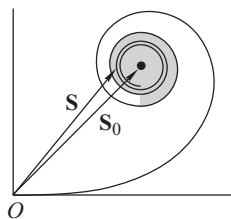


Рис. 1.11

Если же $x_0 > \sqrt{2\pi/\beta}$, т.е. конец суммарного вектора перешел на второй, третий и т.д. витки спирали (заштрихованная область на рис. 1.11), то дальнейшее увеличение верхнего предела уже не может привести к заметному изменению интеграла (1.19), так как конец результирующего вектора уже не выходит из небольшой заштрихованной области, окружающей фокус спирали. Результирующий вектор \mathbf{S} мало отличается при этом от вектора \mathbf{S}_0 , проведенного в фокус спирали и мы имеем оценку:

$$\int_0^{x_0} e^{i\beta x^2} dx \approx \int_0^{\infty} e^{i\beta x^2} dx \quad \text{при } x_0 > \sqrt{\frac{2\pi}{\beta}}.$$

Интеграл (1.19) легко вычисляется следующим образом. Вычисляем сначала

$$S_0^2 = \int_0^{\infty} \int_0^{\infty} e^{i\beta(x^2+y^2)} dx dy.$$

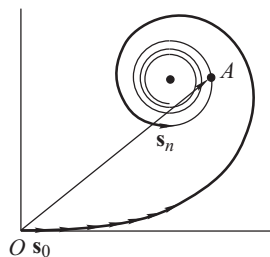


Рис. 1.10

Переходя к полярным координатам

$$\begin{aligned} x &= r \cos \theta, \\ y &= r \sin \theta, \quad dx dy = r dr d\theta, \end{aligned}$$

имеем

$$S_0^2 = \int_0^\infty \int_0^{\pi/2} e^{i\beta r^2} dr d\theta = \frac{\pi}{4} \int_0^\infty e^{i\beta \xi} d\xi \quad (\xi = r^2).$$

Используя (1.18), получаем $S_0 = (\pi i/4)/\beta$, и следовательно,

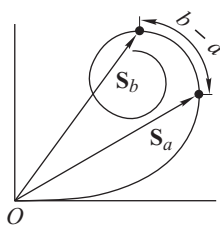


Рис. 1.12

$$S_0 = \int_0^\infty e^{i\beta x^2} dx = \frac{1}{2} \sqrt{\frac{\pi}{\beta}} e^{i\pi/4}. \quad (1.21)$$

Таким образом, вектор, проведенный из начала координат в фокус спирали, имеет длину $(1/2)\sqrt{\pi/\beta}$ и составляет с действительной осью угол $\pi/4$. Диаграмма, изображенная на рис. 1.12, позволяет графически определить как значение интеграла (1.19), так и величину

$$S_{ab} = \int_a^b e^{i\beta x^2} dx. \quad (1.22)$$

Действительно, вектор, определяющий S_{ab} , есть разность двух векторов: $S_b = \int_0^b e^{i\beta x^2} dx$ и $S_a = \int_0^a e^{i\beta x^2} dx$ (рис. 1.12). (Очевидно, вектор S_{ab} стягивает дугу, длина которой равна $|b - a|$.) Поскольку элементарный вектор $-\Delta x e^{i\beta x^2}$ составляет угол π с

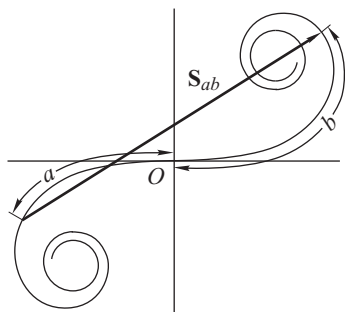


Рис. 1.13

вектором $\Delta x e^{i\beta x^2}$ ($-1 = e^{i\pi}$), то проведенные построения легко обобщаются на случай произвольных (в том числе отрицательных) пределов интегрирования. Полная векторная диаграмма (для произвольных a и b), изображенная на рис. 1.13, была впервые построена французским физиком М. Корню (1841–1902) и называется *спиралью Корню*. Вектор S_{ab} , изображенный на рис. 1.13, соответствует случаю $a < 0$, $b > 0$. Очевидно, величина интеграла $\int_{-\infty}^\infty e^{i\beta x^2} dx$ определяется

вектором, соединяющим фокусы спирали

$$\int_{-\infty}^{\infty} e^{i\beta x^2} dx = \sqrt{\frac{\pi}{\beta}} e^{i\pi/4}. \quad (1.23)$$

Если $a < -x_1$ и $b > x_1$, то концы вектора S_{ab} лежат в небольших окрестностях, окружающих фокусы, и значение интеграла мало меняется с изменением пределов интегрирования, поэтому имеет место следующая оценка:

$$\int_a^b e^{i\beta x^2} dx \approx \sqrt{\frac{\pi}{\beta}} e^{i\pi/4}$$

при $a < -\sqrt{2\pi/\beta}$, $b > \sqrt{2\pi/\beta}$.

Наконец, следует отметить еще одну характерную особенность спирали Корню. Если оба предела интегрирования a и b лежат в окрестности одного и того же фокуса $a > x_1$, $b > x_1$, то модуль интеграла (1.22) не превышает диаметра заштрихованного круга (т.е. мал по сравнению с $|S_0|$, формула (1.21)). Этот случай изображен на рис. 1.14.

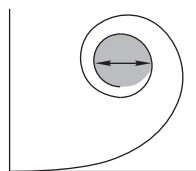


Рис. 1.14

1.4. Модулированные колебания

Амплитудная и фазовая модуляции. Любой колебательный процесс, отличный от гармонического, называется *модулированным колебанием*. Примеры таких процессов (их осциллограммы) приведены на рис. 1.15.

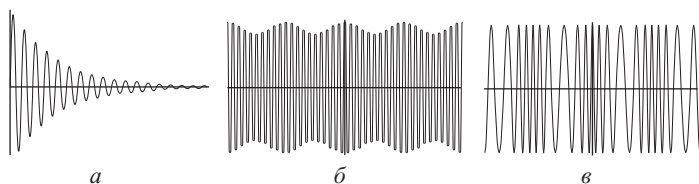


Рис. 1.15

Будем записывать модулированное колебание в виде

$$S(t) = a(t) \cos(\omega_0 t + \varphi(t)). \quad (1.24)$$

В отличие от гармонического колебания, здесь $a(t)$ и $\varphi(t)$ — меняющиеся во времени величины. Форма записи (1.24) особенно целесообразна в том случае, когда $a(t)$ и $\varphi(t)$ — медленно меняющиеся функции времени. Что означает медленность изме-

нения функций $a(t)$ и $\varphi(t)$? Рассмотрим наряду с модулированным колебанием (1.24) гармоническое колебание

$$S_0(t) = a_0 \cos(\omega_0 t + \varphi_0). \quad (1.25)$$

Здесь ω_0 — частота гармонического колебания, то есть та же самая константа, которая входит в выражение (1.24).

Рассмотрим интервал времени τ , существенно превышающий период гармонического колебания $S_0(t)$:

$$\tau \gg \frac{2\pi}{\omega_0} \quad (1.26)$$

и пусть на этом интервале функции $a(t)$ и $\varphi(t)$ остаются практически неизменными, причем

$$a(t) \approx a_0 \quad \text{и} \quad \varphi(t) \approx \varphi_0. \quad (1.27)$$

Итак, мы рассматриваем модулированное колебание, в котором на интервалах времени τ , больших по сравнению с периодом T_0 соответствующего гармонического колебания, величины $a(t)$ и $\varphi(t)$ остаются практически неизменными. В большинстве физических интересных ситуаций достаточно полагать, что сильное неравенство (1.26) выполнено, если интервал времени τ на один-два порядка превышает период колебаний $2\pi/\omega_0$. Такое модулированное колебание называется *квазигармоническим*. В этом случае медленно меняющиеся величины $a(t)$ и $\varphi(t)$ принято называть амплитудой и, соответственно, начальной фазой модулированного колебания. Итак, квазигармоническое колебание можно характеризовать двумя параметрами: периодом колебаний $T_0 = 2\pi/\omega_0$ и временем $\tau \gg T_0$, характеризующим быстроту изменения амплитуды $a(t)$ и (или) начальной фазы $\varphi(t)$. Для описания модулированных колебаний используется следующая терминология: говорят, что функция $a(t)$ описывает закон *амплитудной модуляции*, а функция $\varphi(t)$ — закон *фазовой модуляции*.

Если $\varphi(t) = \varphi_0 = \text{const}$, то

$$S(t) = a(t) \cos(\omega_0 t + \varphi_0), \quad (1.28)$$

где $a(t) \geq 0$. Такое колебание называют *модулированным по амплитуде*.

Если $a(t) = a_0 = \text{const}$, то

$$S(t) = a_0 \cos[\omega_0 t + \varphi(t)]. \quad (1.29)$$

Это — колебание, *модулированное по фазе*. В общем случае имеем как амплитудную, так и фазовую модуляцию, т. е. колебание вида (1.24).

Так же, как и гармонические колебания, квазигармонические процессы изображают графически в виде векторов. Если

гармоническое колебание $S_0(t)$, описываемое формулой (1.26), изображается вектором \mathbf{S}_0 , имеющим фиксированную длину a_0 и направление φ_0 (т. е. мы как бы смотрим на вращающийся с угловой скоростью ω_0 вектор \mathbf{S}_0 из системы координат, которая вращается с той же угловой скоростью — и поэтому видим неподвижный вектор), то модулированное колебание на той же векторной диаграмме естественно изобразить в виде вектора, длина которого $a(t)$ и (или) угол наклона $\varphi(t)$ медленно изменяются (медленно, если речь идет о квазигармоническом колебании). В частности, амплитудно-модулированное колебание (1.28) изображается вектором неизменного направления φ_0 , длина которого изменяется (рис. 1.16 а), а колебание (1.29), модулированное по фазе — вектором неизменной длины, угол наклона которого $\varphi(t)$ изменяется: качания вектора на рис. 1.16 б.

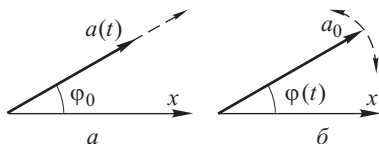


Рис. 1.16

Аргумент косинуса в (1.29) называют *фазой модулированного колебания*

$$\Phi(t) = \omega_0 t + \varphi(t),$$

причем в отличие от гармонического колебания, скорость изменения фазы $\dot{\Phi}$ (величина, которую можно назвать частотой ω), является функцией времени

$$\omega(t) = \dot{\Phi}(t) = \omega_0 + \dot{\varphi}(t). \quad (1.30)$$

Если выполнено условие квазигармоничности (1.26), то $\dot{\varphi}(t) \ll \omega_0$, т. е. меняющаяся во времени частота $\omega(t)$ мало отклоняется от частоты ω_0 гармонического колебания (1.25).

Отметим также, что если $\varphi(t)$ меняется по линейному закону $\varphi(t) = \Omega t$, то мы получаем из (1.30)

$$\omega = \omega_0 + \Omega, \quad (1.31)$$

т. е. гармоническое колебание со смещенной частотой, причем $\Omega \ll \omega_0$ при условии (1.26).

В системе координат, в которой гармоническое колебание частоты ω_0 изображается неподвижным вектором, колебание с частотой $\omega = \omega_0 + \Omega$ изображается вектором, медленно вращающимся против часовой стрелки с частотой Ω , если $\Omega > 0$ (и по часовой стрелке при $\Omega < 0$).

Осциллограммы процессов на рис. 1.15 а, б являются примерами амплитудно-модулированных колебаний, а на рис. 1.15 в — пример колебания, модулированного по фазе.

Биения. Рассмотрим модулированное колебание

$$S(t) = a(t) \cos \omega_0 t, \quad (1.32)$$

где закон модуляции $a(t)$ имеет вид

$$a(t) = a_0 \cos \Omega t. \quad (1.33)$$

Будем полагать изменения $a(t)$ медленными, т. е. множитель $\cos \omega_0 t$ множество раз изменяет знак (проходит много периодов несущего колебания) прежде, чем значение $a(t)$ заметно изменится. Другими словами, период изменения функции $a(t)$ (равный $T = 2\pi/\Omega$) много больше периода *несущего колебания* $2\pi/\omega_0 = T_0$:

$$\omega_0 \gg \Omega. \quad (1.34)$$

Колебание (1.32) с законом модуляции (1.33) изображено на рис. 1.17. Такое квазигармоническое колебание (конечно, не очень строго) называют гармоническим колебанием с медленно меняющейся амплитудой.

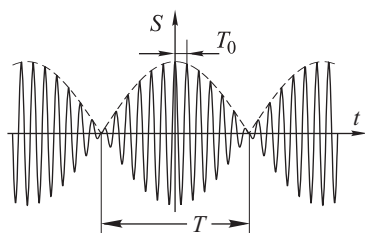


Рис. 1.17

Рассмотрим физический пример. Пусть рупор громкоговорителя является источником гармонического звукового колебания, например, он излучает звуковую волну с частотой $\nu_0 = 440$ Гц ($\omega_0 = 2\pi\nu_0$). Это — нота «ля» первой октавы. Мы слышим ровный

музыкальный тон — гармоническое колебание $S_0(t) = a_0 \cos \omega_0 t$. Как получить модулированное колебание (1.32)? Будем периодически то открывать, то закрывать перегородкой излучающий рупор, громкость звука будет периодически изменяться (например, с частотой 1 раз в секунду ($\Omega = 2\pi$ рад/с)). Звуковые колебания мы будем воспринимать как чистый звуковой тон с частотой ω_0 , громкость которого периодически изменяется.

А теперь обратим внимание на элементарное тригонометрическое тождество $\cos \omega_1 t + \cos \omega_2 t = 2 \cos \left(\frac{\omega_1 - \omega_2}{2} t \right) \cos \left(\frac{\omega_1 + \omega_2}{2} t \right)$, которое показывает, что то же модулированное колебание (1.32) с законом модуляции (1.33) можно получить, сложив два гармонических колебания, немного отличающиеся по частоте: $\omega_1 = \omega_0 + \Omega$ и $\omega_2 = \omega_0 - \Omega$; ($\omega_1 - \omega_2 = 2\Omega$). Мы имеем

$$a \cos((\omega_0 + \Omega)t) + a \cos((\omega_0 - \Omega)t) = 2a \cos \Omega t \cos \omega_0 t. \quad (1.35)$$

Согласно (1.35) мы получаем колебание с частотой ω_0 , амплитуда которого $a(t) = 2a \cos \Omega t$ медленно изменяется, т. е. тот же результат. Этот эффект — возникновение колебания с ме-

нящейся амплитудой при сложении гармонических колебаний с близкими частотами называется *биениями*.

Точнее, амплитудой колебания в этом случае следует назвать величину $2a|\cos\Omega t|$ — ведь амплитуда по смыслу существенно положительная величина, она определяет «размах» колебаний. Но период функции $2a|\cos\Omega t|$ (показанной на рис. 1.17 штриховой линией) вдвое меньше периода функции $2a\cos\Omega t$, поэтому *период биений* (интервал времени, через который повторяется максимальное значение амплитуды колебаний $|a(t)|$) равен $T = 2\pi/(\omega_1 - \omega_2)$. Например, при разности частот слагаемых колебаний $\Delta\nu = 1$ Гц период биений равен 1 секунде.

Эффект можно получить, следовательно, не только описанным выше способом (периодически перекрывая рупор громкоговорителя), но и расположив рядом два рупора, излучающих гармонические звуковые колебания с близкими частотами (например, с частотами 440 и 441 Гц). В результате суперпозиции звуковых колебаний возникают периодические (с частотой 1 Гц) изменения громкости звука.

1.5. Спектральное разложение

Изучая колебания и волны, мы множество раз будем приходить к необходимости представлять произвольный колебательный процесс в виде суммы гармонических колебаний различных частот

$$S(t) = \sum_n a_n \cos(\omega_n t + \varphi_n). \quad (1.36)$$

Равенство называется *спектральным разложением сигнала* $S(t)$. Оно утверждает, что можно, «правильно» подобрав частоты ω_n слагаемых гармонических колебаний, их амплитуды a_n и начальные фазы φ_n , образовать в сумме сигнал $S(t)$. Специальный раздел математики — *Фурье-анализ* — занимается математической стороной проблем, связанных с возможностью представления функций $S(t)$ в виде (1.36). Почему в качестве элементарных слагаемых во многих физических задачах оказывается целесообразно использовать гармонические колебания — об этом мы будем говорить в гл. 2. Здесь же отметим замечательные математические свойства гармонических функций.

Во-первых, гармоническое колебание частоты ω не может быть представлено суперпозицией гармонических колебаний $\sum a_n \cos(\omega_n t + \varphi_n)$ любых других частот ω_n , какие бы коэффициенты a_n и начальные фазы φ_n мы ни старались подобрать. Математически это свойство называют *ортogonalностью*: функция $\cos(\omega t + \varphi)$ не имеет «проекции» на любую другую функцию $\cos(\omega' t + \varphi')$ при $\omega \neq \omega'$, подобно тому, как вектор,

параллельный оси x , невозможно представить в виде суммы векторов, параллельных осям y и z .

Второе важнейшее математическое свойство — единственность представления (1.36): существует единственный набор необходимых частот ω_n и единственный набор отвечающих этим частотам амплитуд a_n и фаз φ_n , обеспечивающих представление функции $S(t)$ в виде суперпозиции гармонических функций.

Наконец, не вдаваясь в математические детали, отметим еще одно важное обстоятельство: любой физически реализуемый колебательный процесс может быть представлен в виде суммы (быть может в виде непрерывной суммы — интеграла) гармонических колебаний.

Комплексная форма спектральных разложений. Ранее мы говорили, что гармоническое колебание $a \cos(\omega t + \varphi)$ можно записать в комплексной форме

$$z(t) = a e^{i(\omega t + \varphi)} = c e^{i\omega t},$$

где $c = a e^{i\varphi}$.

Комплексное представление часто используют и в спектральных разложениях: в качестве элементов разложения (тех элементарных кирпичиков, из которых складывается суммарный колебательный процесс $S(t)$) используют комплексные функции $c_n e^{i\omega_n t}$.

$$S(t) = \sum_{n=-\infty}^{\infty} c_n e^{i\omega_n t}, \quad (1.37)$$

где c_n — в общем случае комплексные коэффициенты: $c_n = |c_n| e^{i\varphi_n}$.

Подчеркнем, что как в случае, когда используется разложение в ряд (1.36), так и в случае, когда в качестве элементов разложения используются функции $e^{i\omega t}$, мы будем говорить (если не делается специальных оговорок) о разложении в спектр реальных сигналов, описываемых действительными функциями $S(t)$.

Каково соотношение между спектральными разложениями (1.36) и (1.37)? Воспользуемся соотношением: $\cos \alpha = (e^{i\alpha} + e^{-i\alpha})/2$, представив каждое слагаемое в (1.36) в виде

$$a_n \cos(\omega_n t + \varphi_n) = \frac{a_n}{2} e^{i\varphi_n} e^{i\omega_n t} + \frac{a_n}{2} e^{-i\varphi_n} e^{-i\omega_n t}. \quad (1.38)$$

Мы видим, что разложения (1.36) и (1.37) будут эквивалентны, если каждому слагаемому в (1.36) поставить в соответствие два слагаемых в (1.37), т. е. каждой реальной гармонике (гармоническому колебанию с частотой ω_n) поставить в соответствие два комплексных слагаемых, введя в (1.37) суммирование по отрицательным значениям n , согласно правилу $\omega_{-n} = -\omega_n$

причем

$$c_0 = a_0, \quad c_n = \frac{a_n}{2} e^{i\varphi_n}, \quad c_{-n} = \frac{a_n}{2} e^{-i\varphi_n} \quad (n = 1, 2, \dots). \quad (1.39)$$

При комплексной форме спектрального разложения необходимо, таким образом, ввести отрицательные частоты $-\omega_n$, которые не имеют какого-либо самостоятельного физического смысла. Они возникают лишь «благодаря» соотношению (1.38).

При этом из (1.39) следует, что соответствующие отрицательным частотам $-\omega_n$ коэффициенты c_{-n} могут быть получены из соответствующих коэффициентов c_n для положительных частот:

$$c_{-n} = c_n^* \quad (1.40)$$

(т. е. c_n и c_{-n} — комплексно-сопряженные числа).

Обратите внимание, что задание комплексного спектра $\{c_n\}$ дает исчерпывающую информацию о сигнале $S(t)$, как об амплитудах слагаемых гармонических колебаний $a_n(t)$, так и об их начальных фазах φ_n .

Спектр функции $f(t)$ принято изображать в виде графика (рис. 1.18): длина стрелочки на каждой частоте ω_n определяется модулем коэффициента c_n (т. е. амплитудой соответствующего гармонического колебания). Следует указать также фазы φ_n спектральных компонент.

Соответствующее разложение (1.36) (в ряд косинусов) представлено на графике рис. 1.18 б: здесь нет отрицательных частот, а длины стрелочек на положительных частотах в соответствии с (1.39) удваиваются.

При этом постоянные составляющие (на частоте $\omega = 0$) в разложениях (1.37), (1.36) одинаковы: $a_0 = c_0$.

Подчеркнем еще раз, что мы говорим о разложении в ряд Фурье (либо ряд (1.37) либо ряд (1.36)) *действительных функций* $S(t)$.

Рассмотрим вначале лишь несколько простых физически интересных примеров.

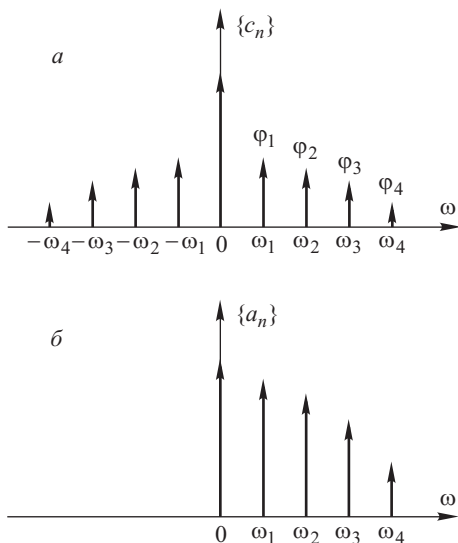


Рис. 1.18

1. $f(t) = a_0 \cos^2 \omega_0 t$. Используя известное тригонометрическое тождество, запишем

$$f(t) = \frac{a_0}{2} + \frac{a_0}{2} \cos 2\omega_0 t = \frac{a_0}{2} + \frac{a_0}{4} e^{i2\omega_0 t} + \frac{a_0}{4} e^{-i2\omega_0 t}. \quad (1.41)$$

Первое из равенств (1.41) — разложение в ряд (1.36), второе — в ряд (1.37). Константу $a_0/2$ можно рассматривать как гармоническое колебание нулевой частоты, поэтому равенство (1.41) — это есть представление колебания $f(t)$ в виде суммы двух гармонических колебаний одинаковой амплитуды $a_0/2$ с частотами $\omega_1 = 0$ и $\omega_2 = 2\omega_0$ (во втором случае с частотами $2\omega_0$ и $-2\omega_0$). Спектр процесса $f(t)$ представлен на рис. 1.19а (разложение в ряд (1.37) и на рис. 1.19б (разложение в ряд (1.36)).

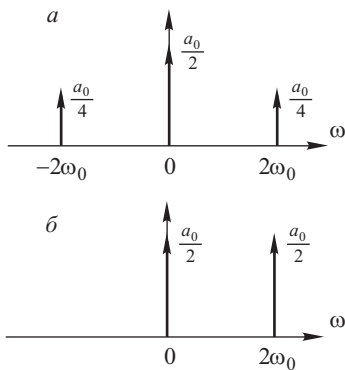


Рис. 1.19

2.

$$\begin{aligned} f(t) &= a(t) \cos \omega_0 t, \\ a(t) &= a_0(1 + m \cos \Omega t). \end{aligned} \quad (1.42)$$

Это — пример амплитудно-модулированного колебания. Константа $m < 1$ называется *глубиной модуляции*. Мы имеем

$$\begin{aligned} f(t) &= a_0(1 + m \cos \Omega t) \cos \omega_0 t = \\ &= a_0 \cos \omega_0 t + \frac{ma_0}{2} \cos(\omega_0 + \Omega)t + \frac{ma_0}{2} \cos(\omega_0 - \Omega)t. \end{aligned} \quad (1.43)$$

Итак, амплитудно-модулированное колебание с законом модуляции (1.42) представляется в виде суммы трех гармонических колебаний (трех гармоник)

$$\begin{aligned} f_1(t) &= a_0 \cos \omega_0 t, & f_2(t) &= \frac{ma_0}{2} \cos(\omega_0 + \Omega)t, \\ f_3(t) &= \frac{ma_0}{2} \cos(\omega_0 - \Omega)t \end{aligned}$$

с частотами соответственно ω_0 , $\omega_0 + \Omega$, $\omega_0 - \Omega$ и амплитудами a_0 , $ma_0/2$, $ma_0/2$. Колебание $f_1(t)$ называется *несущим колебанием*, а $f_2(t)$ и $f_3(t)$ — *боковыми гармониками*. Условие квазигармоничности колебания $f(t)$: $\Omega \ll \omega_0$. В этом случае целесообразно рассматривать колебания $f_2(t)$ и $f_3(t)$ — как колебания частоты ω_0 , начальная фаза которых меняется по закону $\varphi_2(t) = +\Omega t$ и $\varphi_3(t) = -\Omega t$. Другими словами на векторной диаграмме, где несущее колебание изображается неподвижным вектором \mathbf{S}_0 , колебания $f_1(t)$, $f_2(t)$ изображаются соответственно векторами \mathbf{S}_1 и \mathbf{S}_2 , которые вращаются (против и по часовой

стрелке) с угловой скоростью Ω (с периодом $T = 2\pi/\Omega$), длины этих векторов $ma_0/2$.

На рис. 1.20 показаны последовательные стадии векторного сложения несущего колебания с боковыми гармониками.

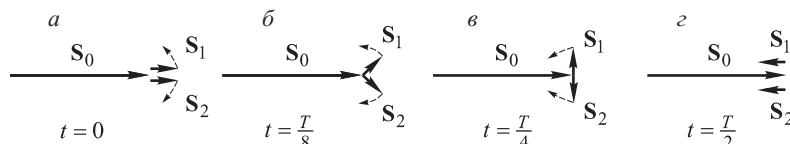


Рис. 1.20

В процессе колебаний направление суммарного вектора $\mathbf{S} = \mathbf{S}_0 + \mathbf{S}_1 + \mathbf{S}_2$ остается неизменным, изменяется лишь его длина (от максимального значения $a_0(1+m)$, до минимального значения $a_0(1-m)$), что соответствует амплитудной модуляции.

3. Рассмотрим пример сигнала, модулированного по фазе

$$f(t) = a_0 \cos(\omega_0 t + \varphi(t)), \quad \text{где} \quad \varphi(t) = m \cos \Omega t. \quad (1.44)$$

Константа m — *глубина модуляции фазы* — определяет диапазон изменения начальной фазы (от $-m$ до $+m$) или (если обратиться к векторной диаграмме) «амплитуду качания» вектора \mathbf{S} (рис. 1.16 б). Используя известное тригонометрическое тождество

$$\cos(\alpha + \beta) = \cos \alpha \cos \beta - \sin \alpha \sin \beta,$$

запишем $f(t)$ в виде

$$f(t) = a_0(\cos \omega_0 t \cos \varphi(t) - \sin \omega_0 t \sin \varphi(t)).$$

В общем случае закон модуляции (1.44) приводит к довольно сложному спектру (с большим числом слагаемых гармонических колебаний). Мы рассмотрим случай $m \ll 1$ (малая глубина модуляции фазы), когда можно использовать приближенные выражения: $\cos \varphi(t) \approx 1$, $\sin \varphi(t) \approx \varphi(t)$ (мы отбрасываем величины порядка m^2 и выше).

Тогда

$$f(t) = a_0 \cos \omega_0 t - a_0 m \sin \omega_0 t \cos \Omega t$$

или (так как $2 \sin \alpha \cos \beta = \sin(\alpha + \beta) + \sin(\alpha - \beta)$):

$$f(t) = a_0 \cos \omega_0 t + \frac{ma_0}{2} \left[\cos(\omega_0 + \Omega)t + \frac{\pi}{2} \right] + \frac{ma_0}{2} \left[\cos(\omega_0 - \Omega)t + \frac{\pi}{2} \right]. \quad (1.45)$$

Это и есть искомое представление колебания $f(t)$ в виде суммы гармонических колебаний.

Сравним формулы (1.43) и (1.45). Первая из них — разложение в спектр колебания, модулированного по амплитуде,

вторая — колебания, модулированного по фазе. Эти колебания сильно различаются по форме (сравните осциллограммы на рис. 1.15 б и в), однако их спектры весьма похожи (рис. 1.21).

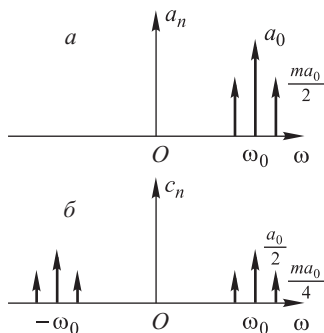


Рис. 1.21

В обоих случаях в правой части формул (1.43), (1.45) три слагаемых, три гармонических колебания, имеющих одинаковые частоты (ω_0 , $\omega_0 \pm \Omega$) и амплитуды (a_0 — несущие колебания, $ma_0/2$ — боковые гармоники). Различие выглядит небольшим: боковые гармоники отличаются фазовым сдвигом $\pi/2$, однако это различие приводит к кардинальному отличию в форме (в осциллограмме $f(t)$) результирующего сигнала. Векторные диаграммы на рис. 1.22 поясняют этот результат: поворот векторов \mathbf{S}_1 и \mathbf{S}_2 относительно вектора \mathbf{S}_0 на $\pi/2$, (рис. 1.22 а) приводит к тому, что с течением времени (рис. 1.22 б, в, г) суммарный вектор (изображен штриховой линией) изменяет угол наклона, не изменяя (с точностью до величины порядка m^2) своей длины, что и соответствует фазовой модуляции. Этот пример подчеркивает, какую огромную роль играют фазовые соотношения при сложении колебаний.

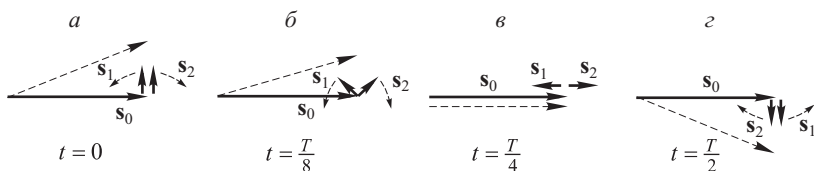


Рис. 1.22

Итак, изменив фазу несущего колебания (или боковых гармоник) на $\pi/2$, мы можем преобразовать колебание, модулированное по фазе, в амплитудно-модулированное колебание. Это известный в радиотехнике «прием с изменением фазы несущей».

Читатель может самостоятельно проанализировать «прием без несущей»: что представляет собой модулированное колебание $f(t)$, если «убрать» первое слагаемое — несущее колебание $a_0 \cos \omega_0 t$?

Спектр периодического процесса. Рассмотрим теперь периодический колебательный процесс общего вида $f(t) = f(t + T)$, где T — период процесса. В этом случае функция $f(t)$ может быть представлена суммой гармонических колебаний с кратными

частотами $\omega_n = n\omega_0$, где $T = 2\pi/\omega_0$.

$$f(t) = \sum_n c_n e^{in\omega_0 t}. \quad (1.46)$$

Действительно, для любого t функция (1.46) повторяет свое значение через время T , поскольку $e^{in\omega_0(t+T)} = e^{in\omega_0 T} \cdot e^{in\omega_0 t} = e^{i2\pi n} e^{in\omega_0 t} = e^{in\omega_0 t}$. Спектр $\{c_n\}$ можно найти следующим образом: домножим обе части равенства (1.46) на $e^{-im\omega_0 t}$ и проинтегрируем по t за время, равное периоду (от $-T/2$ до $+T/2$). Получим

$$\int_{-T/2}^{T/2} f(t) e^{-im\omega_0 t} dt = \sum_n c_n \cdot \int_{-T/2}^{T/2} e^{i(n-m)\omega_0 t} dt.$$

Легко проверить, что интеграл в правой части равенства есть

$$\int_{-T/2}^{T/2} e^{i(n-m)\omega_0 t} dt = \begin{cases} 0 & \text{при } n \neq m, \\ T & \text{при } n = m \end{cases}$$

(время интегрирования $(-T/2, T/2)$ равно целому числу периодов функций $\cos[(n-m)\omega_0 t]$ и $\sin[(n-m)\omega_0 t]$, поэтому $\int_{-T/2}^{T/2} \{\cos[(n-m)\omega_0 t] + i \sin[(n-m)\omega_0 t]\} dt = 0$ при $n \neq m$). Следовательно, получаем

$$c_m = \frac{1}{T} \int_{-T/2}^{T/2} f(t) e^{-im\omega_0 t} dt. \quad (1.47)$$

Формула (1.47) дает правило нахождения коэффициентов разложения периодической функции в ряд Фурье.

Задача 1. Найти спектр периодической последовательности прямоугольных импульсов длительности τ , с периодом следования импульсов $T > \tau$.

Используя (1.47), находим

$$c_n = \frac{1}{T} \int_{-\tau/2}^{\tau/2} 1 \cdot e^{-in\omega_0 t} dt$$

(на интервале интегрирования $T/2 < t < T/2$ функция $f(t)$ отлична от нуля и равна единице лишь в области $|t| < \tau/2$).

Далее

$$c_n = \frac{1}{T} \int_{-\tau/2}^{\tau/2} \frac{e^{-in\omega_0 t}}{-in\omega_0} d(-in\omega_0 t) = \frac{1}{T} \frac{1}{-in\omega_0} e^{-in\omega_0 t} \Big|_{-\tau/2}^{\tau/2} =$$

$$= 2 \frac{\tau}{2T} \left[\frac{e^{in\omega_0 \tau/2} - e^{-in\omega_0 \tau/2}}{2in\omega_0 \tau/2} \right].$$

Окончательно находим

$$c_n = \frac{\tau}{T} \frac{\sin(n\omega_0 \tau/2)}{n\omega_0 \tau/2}. \quad (1.48)$$

Спектр $\{c_n\}$ показан на рис. 1.23. Штриховой кривой изображена функция

$$C(\omega) = \frac{\tau}{T} \frac{\sin(\omega \tau/2)}{\omega \tau/2}.$$

Очевидно, при $\omega = n\omega_0$ эта функция принимает значение равное c_n : $c_n = C(n\omega_0)$. Полуширина $\Delta\omega$ главного максимума этой функции определяется условием $\sin(\omega \tau/2) = 0$:

$$\Delta\omega \cdot \frac{\tau}{2} = \pi \quad \text{или} \quad \Delta\omega \cdot \tau = 2\pi.$$

Рисунок соответствует ситуации, когда $3\omega_0 = \Delta\omega$, $\tau = 3T$. Как видно из рисунка, спектральные гармоники, имеющие заметную амплитуду, сосредоточены в интервале частот $|\omega| \lesssim \Delta\omega = 2\pi/\tau$.

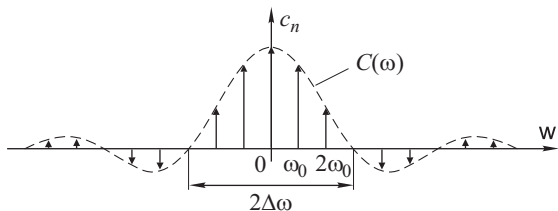


Рис. 1.23

Спектр непериодического сигнала. Соотношение неопределенностей. Рассмотрим задачу разложения в спектр произвольного сигнала $f(t)$.

Оказывается, что произвольный сигнал не может быть представлен в виде (1.36) либо (1.37), т.е. в виде суммы гармонических колебаний с дискретным набором частот ω_n . В общем случае необходим непрерывный набор гармоник, т.е. необходимо суммировать гармонические колебания, частоты которых непрерывно заполняют некоторый (быть может бесконечный) интервал частот. То есть необходимо иметь не только колебания с частота-

ми $\omega_1, \omega_2, \dots, \omega_n$, но также и все частоты в промежутке между ними. При этом ряд (1.37) заменится интегралом Фурье

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} C(\omega) e^{i\omega t} d\omega. \quad (1.49)$$

Множитель $C(\omega) = a(\omega) e^{i\varphi(\omega)}$ показывает с каким весом (т. е. с какой амплитудой $a(\omega)$ и с какой начальной фазой $\varphi(\omega)$) необходимо складывать гармонические колебания разных частот, чтобы при суммировании (интегрировании) образовать заданный сигнал $f(t)$. Функция $C(\omega)$ называется *спектром* (или *преобразованием Фурье*) сигнала $f(t)$.

Как найти спектр, если сигнал $f(t)$ известен? Приведем без доказательства (которое можно найти в учебниках по Фурье-анализу) формулу

$$C(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt. \quad (1.50)$$

Соотношение (1.50) математики называют *прямым преобразованием Фурье*, а формулу (1.49) — *обратным преобразованием Фурье*. Связь (1.49) между функциями $f(t)$ и $C(\omega)$ символически записывается в виде $f(t) \leftrightarrow C(\omega)$.

Как ясно из (1.50) спектр действительной функции $f(t) = f^*(t)$ обладает определенной симметрией: $C(\omega) = C^*(-\omega)$ и, следовательно, $|C(\omega)| = |C(-\omega)|$ (* — знак комплексного сопряжения). Математики называют это свойство *эрмитовостью*.

Задача 2. Разложение в спектр прямоугольного импульса длительности τ (рис. 1.24 а). Используя (1.50), получаем

$$C(\omega) = \int_{-\tau/2}^{\tau/2} e^{-i\omega t} dt = \frac{1}{-i\omega} \int_{-\tau/2}^{\tau/2} e^{-i\omega t} d(-i\omega t) = \tau \frac{\sin(\omega\tau/2)}{\omega\tau/2}. \quad (1.51)$$

Функция $C(\omega)$ показана на рис. 1.24 б.

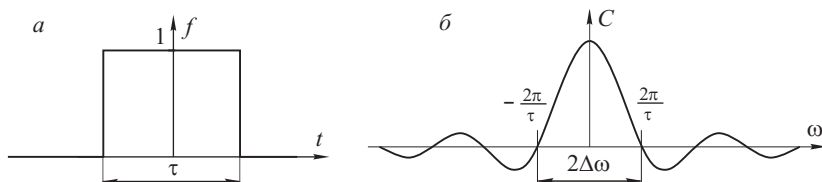


Рис. 1.24

Полезно сравнить спектр отдельного импульса со спектром периодической последовательности одинаковых импульсов

(см. рис. 1.23): спектр импульса $C(\omega)$ (с множителем $1/T$) представляет собой «огibaющую» частоту спектральных компонент c_n периодической последовательности импульсов. Вместо дискретного спектра $\{c_n\}$ получаем непрерывный спектр $C(\omega)$.

Модуль функции $C(\omega)$ определяет амплитуды гармонических колебаний разных частот, сумма которых образует импульс $f(t)$. Как видно из графика, основной вклад дают гармонические колебания, частоты которых заполняют интервал $|\Delta\omega| < 2\pi/\tau$. Это — полуширина главного максимума функции $\frac{\sin(\omega\tau/2)}{\omega\tau/2}$. Диапазон частот $\Delta\omega$ можно назвать *шириной спектра* $C(\omega)$.

Мы получили замечательное соотношение, связывающее между собой длительность сигнала с шириной его спектра:

$$\tau \cdot \Delta\omega \approx 2\pi. \quad (1.52)$$

Это соотношение имеет универсальный характер. Оно оказывается справедливым по порядку величины для произвольного сигнала $f(t)$. Чем больше длительность сигнала (либо больше интервал времени, в течение которого происходит его заметное изменение), тем уже спектр сигнала $\Delta\omega$, и, наоборот, чем короче сигнал (или быстрее происходит изменение сигнала), тем шире его спектр, т. е. требуется более широкий интервал частот гармонических колебаний, образующих в сумме данный сигнал. В этом состоит смысл замечательного соотношения (1.52), которое называется *соотношением неопределенностей*.

Задача 3. Спектр функции $f_0(t)$ есть $C_0(\omega)$: $f_0(t) \leftrightarrow C_0(\omega)$. Найти спектр $C(\omega)$ комплексной функции $f(t) = f_0(t)e^{i\omega_0 t}$.

Согласно (1.50) запишем

$$C(\omega) = \int_{-\infty}^{\infty} f_0(t)e^{i\omega_0 t} e^{-i\omega t} dt = \int_{-\infty}^{\infty} f_0(t)e^{-i(\omega - \omega_0)t} dt = C_0(\omega - \omega_0). \quad (1.53)$$

т. е. спектр переносится по оси частот вправо на величину ω_0 (рис. 1.25).

Из формулы (1.53) ясно, что если спектр функции $f_0(t)$ локализован в области частот $|\omega| \leq \Omega$ (рис. 1.25 а), то спектр

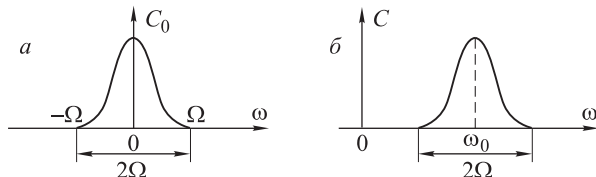


Рис. 1.25

комплексной функции $f(t)$ не содержит отрицательных частот (рис. 1.25 б): $C(\omega) \equiv 0$ при $\omega < 0$, если $\omega_0 > \Omega$.

Задача 4. Спектр функции $f_0(t)$ есть $C_0(\omega)$: $f_0(t) \leftrightarrow C_0(\omega)$. Найти спектр функции $f(t) = f_0(t - \tau)$.

Имеем

$$C(\omega) = \int_{-\infty}^{\infty} f_0(t - \tau) e^{-i\omega t} dt.$$

После замены переменных $t' = t - \tau$ ($dt = dt'$) получаем

$$C(\omega) = \int_{-\infty}^{\infty} f_0(t') e^{-i\omega(t'+\tau)} dt'.$$

Множитель $e^{-i\omega\tau}$ (не зависящий от переменной интегрирования t') выносится из под знака интеграла

$$C(\omega) = e^{-i\omega\tau} \int_{-\infty}^{\infty} f_0(t') e^{-i\omega t'} dt' = C_0(\omega) e^{-i\omega\tau} \quad (1.54)$$

или, символически $f_0(t - \tau) \leftrightarrow C_0(\omega) e^{-i\omega\tau}$.

Задача 5. Пусть $f_0(t) \leftrightarrow C_0(\omega)$. Найти спектр $C(\omega)$ функции $f(t) = f_0(t) \cos \omega_0 t$.

Используя формулу Эйлера, запишем

$$f(t) = \frac{1}{2} f_0(t) e^{i\omega_0 t} + \frac{1}{2} f_0(t) e^{-i\omega_0 t}.$$

Согласно (1.53), имеем

$$C(\omega) = \frac{1}{2} C_0(\omega - \omega_0) + \frac{1}{2} C_0(\omega + \omega_0), \quad (1.55)$$

т.е. спектр $C_0(\omega)$ (умноженный на 1/2) переносится по оси частот влево и вправо на несущую частоту ω_0 . В частности, пусть

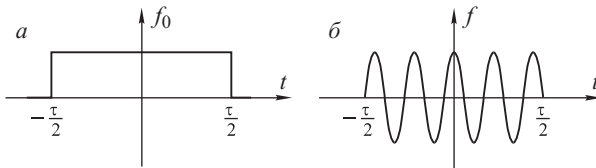


Рис. 1.26

$f_0(t)$ — прямоугольный импульс длительности τ (рис. 1.26 а)

$$f_0(t) = p_\tau(t) = \begin{cases} 1 & \text{при } |t| \leq \tau/2, \\ 0 & \text{при } |t| > \tau/2. \end{cases}$$

Тогда $f(t) = f_0(t) \cos \omega_0 t$ — обрывок косинусоиды (цуг) длительности τ (рис. 1.26 б). Согласно (1.55), получаем

$$C(\omega) = \frac{\tau}{2} \left\{ \frac{\sin [(\omega - \omega_0)\tau/2]}{(\omega - \omega_0)\tau/2} + \frac{\sin [(\omega + \omega_0)\tau/2]}{(\omega + \omega_0)\tau/2} \right\}, \quad (1.56)$$

где $C_0(\omega) = \tau \frac{\sin(\omega\tau/2)}{\omega\tau/2}$ — спектр импульса $f_0(t)$.

Спектры $C_0(\omega)$ и $C(\omega)$ представлены на рис. 1.27.

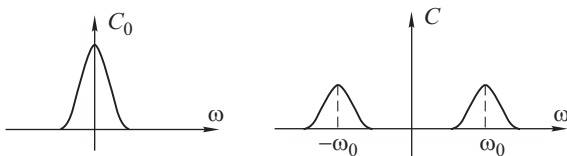


Рис. 1.27

Задача 6. Найти спектр $C(\omega)$ сигнала $f(t)$, являющегося периодическим повторением сигнала $f_0(t) \leftrightarrow C_0(\omega)$:

$$f(t) = \sum_{n=0}^{N-1} f_0(t - nT) \quad (\text{при конечном } N).$$

Используя (1.54), имеем

$$C(\omega) = C_0(\omega) \sum_{n=0}^{N-1} e^{-in\omega T}.$$

Выражение

$$S(\omega) = \sum_{n=0}^{N-1} e^{-in\omega T}$$

можно интерпретировать как сумму колебаний, фазы которых составляют арифметическую прогрессию

$$S = \sum_{n=0}^{N-1} e^{-in\theta} \quad (\theta = \omega T).$$

Используя (1.16), получаем

$$C(\omega) = C_0(\omega) \frac{\sin(N\omega T/2)}{\sin(\omega T/2)} e^{i[(N-1)/2]\omega T}. \quad (1.57)$$

Имеем набор узких (при $N \gg 1$) спектральных максимумов высотой равной N (так как $\frac{\sin(N\omega T/2)}{\sin(\omega T/2)} = N$ на частотах $\omega_n = n\omega_0$, кратных $\omega_0 = 2\pi/T$) с полушириной $\Delta\omega = 2\pi/(NT)$, которую можно найти из условия $\sin(N\omega T/2) = 0$ (рис. 1.28 а). Макси-

мумы промодулированы «огibaющей» — функцией $C_0(\omega)$, спектральная ширина которой определяется длительностью сигнала $f_0(t)$: $\Delta\omega_0 \approx 2\pi/\tau$. Функция $C(\omega)$ имеет вид, показанный на рис. 1.28 б (функция $C_0(\omega)$ изображена на рисунке штриховой линией).

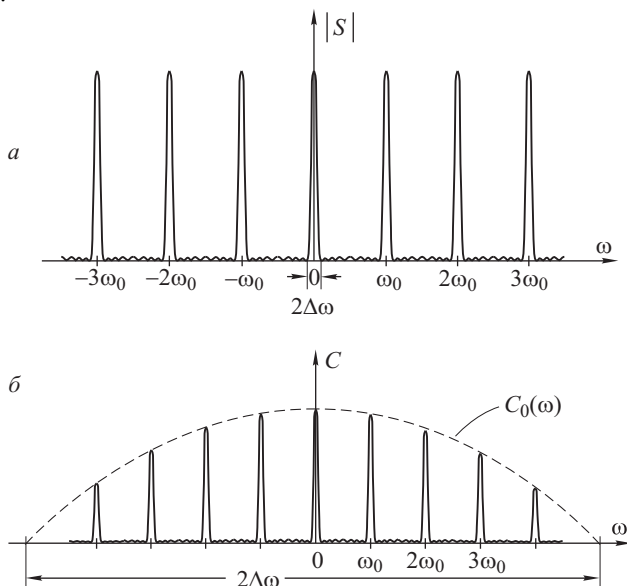


Рис. 1.28

При $N \rightarrow \infty$ высота спектральных максимумов стремится к бесконечности, а их ширина $\Delta\omega$ к нулю: мы получаем «частотол» δ -функций.

1.6. Векторные колебания. Фигуры Лиссажу

Обратим внимание на формулу (1.6) — правило, по которому мы складываем колебания $S_1(t)$ и $S_2(t)$. Это правило сложения скалярных величин, т. е. в каждый момент времени t мы просто алгебраически складываем две скалярные величины. Мы пользовались графическим изображением колебаний (в виде векторов) лишь для того, чтобы облегчить себе задачу сложения двух скалярных меняющихся во времени величин S_1 и S_2 .

Теперь мы будем говорить о другой проблеме. Представим себе, что колеблющейся величиной является вектор; например, колеблется вектор напряженности электрического поля в данной точке пространства, или вектор магнитного поля, или какая-либо другая векторная величина. При этом в процессе колебаний может изменяться не только модуль вектора, но и его направление

в пространстве. Начнем с простой задачи. Как сложить два колеблющихся вектора \mathbf{S}_1 и \mathbf{S}_2 :

$$\mathbf{S}_1 = \mathbf{A}_1 \cos(\omega t + \varphi_1) \text{ и } \mathbf{S}_2 = \mathbf{A}_2 \cos(\omega t + \varphi_2), \quad (1.58)$$

где \mathbf{A}_1 и \mathbf{A}_2 — два взаимно-перпендикулярных постоянных вектора, скажем вектор \mathbf{A}_1 направлен вдоль оси x , а вектор \mathbf{A}_2 — вдоль оси y . То есть в процессе колебаний вектор \mathbf{S}_1 остается параллельным оси x , а вектор \mathbf{S}_2 — параллельным оси y .

Мы убедимся, что суммарный вектор $\mathbf{S} = \mathbf{S}_1 + \mathbf{S}_2$ в общем случае изменяет с течением времени свое направление в пространстве. Очевидно, что в любой момент времени проекции суммарного вектора на оси x и y есть:

$$\begin{aligned} S_x &= A_1 \cos(\omega t + \varphi_1), \\ S_y &= A_2 \cos(\omega t + \varphi_2). \end{aligned} \quad (1.59)$$

Меняется, в общем случае, и ориентация суммарного вектора ($\operatorname{tg} \alpha = S_y/S_x$, α — угол между направлением вектора и осью x), и его длина. Ясно, что при этом \mathbf{S} остается в плоскости xy (его z -проекция равна нулю). Конец вектора \mathbf{S} описывает с течением времени некоторую плоскую траекторию, и формулы (1.59) можно рассматривать как уравнение этой траектории в параметрической форме.

Это — уравнение эллипса, вписанного в прямоугольник со сторонами $(2A_1, 2A_2)$ (рис. 1.29а). Ориентация главных полуосей эллипса зависит от сдвига фаз $\varphi_2 - \varphi_1$. Колебание суммарного вектора называют в этом случае эллиптически-поляризованным. На рис. 1.29 показаны частные случаи *эллиптически поляризованных колебаний*.

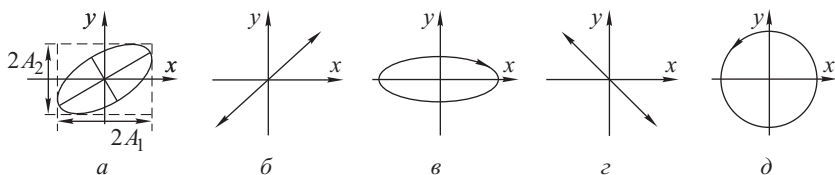


Рис. 1.29

При $\varphi_2 - \varphi_1 = 0, 2\pi, 4\pi, \dots$ эллипс поляризации вырождается в прямую: ориентация суммарного вектора не меняется (рис. 1.29б). Уравнение прямой в этом случае легко найти из (1.59):

$$y = \frac{A_2}{A_1} x.$$

Угол наклона вектора $\operatorname{tg} \alpha = y/x = A_2/A_1$ не меняется со временем.

При $\varphi_2 - \varphi_1 = \pi, 3\pi, 5\pi, \dots$ получаем

$$y = -\frac{A_2}{A_1} x.$$

Этот случай показан на рис. 1.29 г.

И в том, и в другом случае вектор \mathbf{S} в процессе колебаний не меняет своей ориентации в пространстве. Такие колебания называются *линейно-поляризованными*.

Главные оси эллипса совпадают с осями x, y , если

1) $\varphi_2 - \varphi_1 = \pi/2 + 2\pi n$ ($n = 1, 2, \dots$), либо

2) $\varphi_2 - \varphi_1 = 3\pi/2 + 2\pi n$ ($n = 1, 2, \dots$).

В обоих случаях уравнение эллипса имеет вид (рис. 1.29 в):

$$\frac{x^2}{A_1^2} + \frac{y^2}{A_2^2} = 1. \quad (1.60)$$

Однако движение конца суммарного вектора по эллипсу (1.60) в случаях 1 и 2 различны. (Вопрос для читателя: в чем состоит это различие?)

Более сложной оказывается траектория движения конца суммарного вектора \mathbf{S} , если частоты двух взаимно перпендикулярных колебаний не равны. В частности, траектория оказывается замкнутой, если частоты складываемых колебаний относятся как целые числа:

$$S_x(t) = A_1 \cos(n\omega t + \varphi_1), \quad S_y(t) = A_2 \cos(m\omega t + \varphi_2), \quad (1.61)$$

т. е. $\omega_1/\omega_2 = n/m$. Действительно, промежуток времени $T = 2\pi/\omega$ содержит как целое число периодов колебания $S_x(t)$, так и целое число периодов колебания $S_y(t)$, т. е. значения $S_x(t)$ и $S_y(t)$ повторяются через интервал времени T ($T = nT_1 = mT_2$) — конец вектора \mathbf{S} попадает в ту же точку на плоскости x, y через интервал времени T .

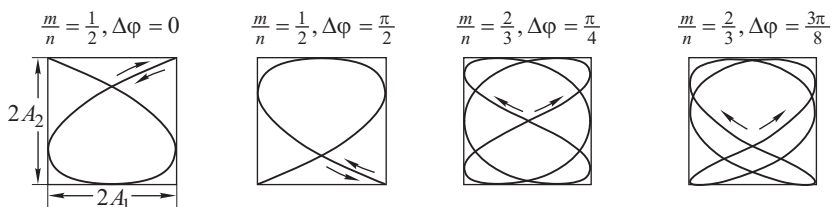


Рис. 1.30

Замкнутая траектория, которую описывает конец суммарного вектора \mathbf{S} , называется *фигурой Лиссажу*. На рис. 1.30 приведены примеры фигур Лиссажу для некоторых значений ω_1/ω_2 и $\Delta\varphi = \varphi_2 - \varphi_1$.

Фигуры Лиссажу можно наблюдать на экране осциллографа, подавая на горизонтальный вход осциллографа (т. е. на пластины конденсатора, создающего горизонтальное электрическое поле) напряжение, меняющееся по закону $x(t) = A_1 \cos(n\omega t + \varphi_1)$, а на пластины конденсатора, создающие вертикальное электрическое поле, напряжение $y(t) = A_2 \cos(m\omega t + \varphi_2)$.

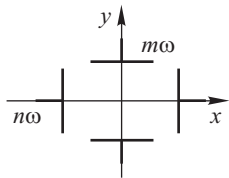


Рис. 1.31

Электроны, пролетающие между обкладками этих конденсаторов навстречу читателю (рис. 1.31), отклоняются суммарным электрическим полем $\mathbf{S}(t) = \mathbf{S}_1(t) + \mathbf{S}_2(t)$, оставляя на экране осциллографа след, представляющий собой фигуру Лиссажу (например, одну из фигур, изображенных на рис. 1.30).

Задачи

1. Нарисуйте графики $S(t)$ (осциллограммы) следующих гармонических колебаний: $S(t) = 2 \cos(100\pi t)$, $S(t) = 5 \cos(50\pi t + \pi/3)$, $S(t) = \cos(100\pi t + \pi/2)$.

2. Для осциллограмм, изображенных на рис. 1.32, найдите циклическую частоту и начальную фазу гармонических колебаний. Напишите уравнение $S(t)$ соответствующих гармонических колебаний.

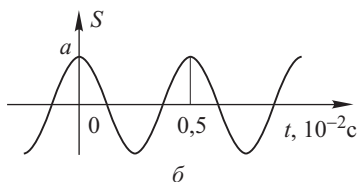
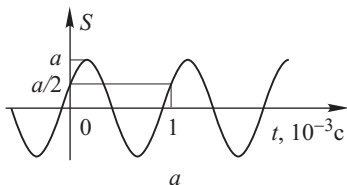


Рис. 1.32

3. Найти разность фаз двух колебаний одинаковой частоты и амплитуды, если амплитуда суммарного колебания равна: а) амплитуде слагаемых колебаний; б) в $\sqrt{2}$ раз больше амплитуды слагаемых колебаний.

4. При сложении трех гармонических колебаний одинаковой частоты и амплитуды оказалось, что суммарное колебание синфазно (т. е. совпадает по фазе) с одним из слагаемых, а его амплитуда вдвое больше амплитуд слагаемых колебаний. Найти разность фаз слагаемых колебаний.

5. Используя графическое изображение, найти сумму N гармонических колебаний одинаковой частоты и амплитуды, фазы которых составляют арифметическую прогрессию: $\Delta\varphi_n = n\delta$ ($n = 0, 1, \dots, N-1$).

6. Выразить начальную фазу суммарного колебания, изображенного на рис. 1.3, через амплитуды и фазы слагаемых колебаний.

7. Докажите, что точка M , изображающая состояние системы на фазовой плоскости, движется с течением времени по часовой стрелке (рис. 1.5а).

8. Докажите, что семейство фазовых траекторий на рис. 1.5а описывает гармонические колебания одной и той же частоты ω . Определите величину ω .

9. Чем отличаются гармонические колебания, отвечающие фазовым траекториям на рис. 1.5б и в. Найти отношение амплитуд и частот колебаний на этих траекториях.

10. Изображающие точки M_1 , M_2 и M_3 находятся в момент времени $t = 0$ на прямой OA (рис. 1.5б). Через время Δt точка M_2 находится на прямой OB . Найти положение точек M_1 и M_3 .

11. Нарисуйте графики (осциллограммы) следующих модулированных колебаний:

а) $S(t) = a(t) \cos(\omega_0 t + \varphi_0)$, где $a(t) = a_0 |\cos \Omega t|$ при $\Omega \ll \omega_0$;

б) $S(t) = a_0 (1 + m \cos \Omega t) \cos \omega_0 t$ при $m < 1$;

в) $S(t) = a_0 \cos(\omega_0 t + at^2)$.

12. Найдите результат сложения гармонических колебаний с близкими частотами ω и $(\omega + \Omega)$, амплитуды которых a_1 и a_2 отличаются. Каково максимальное и минимальное значение квадрата амплитуды (интенсивности) суммарного колебания?

13. При сложении гармонических колебаний с близкими частотами интенсивность суммарного колебания изменяется вдвое. Каково отношение интенсивностей слагаемых колебаний?

Глава 2

КОЛЕБАНИЯ В ЛИНЕЙНЫХ СИСТЕМАХ

2.1. Примеры простейших колебательных систем. Общность уравнений, описывающих колебания различной физической природы

1. Математический маятник — идеализированная колебательная система — материальная точка массы m , подвешенная на нерастяжимой невесомой нити длины l . Выведенный из положения равновесия маятник начинает качаться, совершать колебательное движение. На рис. 2.1 показаны силы, действующие на материальную точку в момент, когда угол отклонения маятника равен α . Уравнение движения (второй закон Ньютона) имеет вид

$$m\mathbf{a} = m\mathbf{g} + \mathbf{T},$$

где \mathbf{T} — сила натяжения нити, $m\mathbf{g}$ — сила тяжести.

Проецируя это уравнение на направление касательной к траектории движения (дуги окружности), запишем

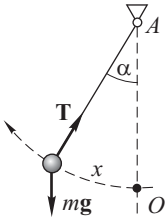


Рис. 2.1

$$ma_\tau = -mg \sin \alpha \quad (2.1)$$

(сила натяжения нити T нормальна к траектории, ее проекция на касательную равна нулю). Здесь a_τ — касательная к окружности составляющая ускорения: $a_\tau = \ddot{x}$ (x — длина дуги, отсчитываемая от положения равновесия — точки O , \ddot{x} — вторая производная функции $x(t)$).

Положительному отклонению $x > 0$ соответствует, согласно рис. 2.1, отклонение влево, при этом проекция силы тяжести на касательную направлена вправо — она отрицательна. Это очень существенно: куда бы ни отклонялся маятник от положения равновесия, возникает сила, возвращающая маятник в это положение — сила имеет знак, противоположный смещению. Точка O — *устойчивое положение равновесия*.

Поскольку $x = l\alpha$ и $a_\tau = \ddot{x} = l\ddot{\alpha}$, то из (2.1) получаем

$$\ddot{\alpha} + \omega_0^2 \sin \alpha = 0, \text{ где } \omega_0 = \sqrt{\frac{g}{l}}. \quad (2.2)$$

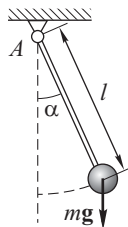
При малых отклонениях от положения равновесия $\sin \alpha \approx \alpha$, и мы приходим к уравнению малых колебаний

$$\ddot{\alpha} + \omega_0^2 \alpha = 0. \quad (2.3)$$

2. Физический маятник — это реальная колебательная система (рис. 2.2). Момент инерции маятника относительно оси A равен J . Второй закон Ньютона для вращательного движения тела относительно неподвижной оси имеет вид

$$J \ddot{\alpha} = M, \quad (2.4)$$

где M — суммарный момент внешних сил относительно оси A . В нашем случае (без учета сил трения в оси и сопротивления воздуха) $M = -mgl \sin \alpha$. Рассматривая малые колебания ($\sin \alpha \approx \alpha$), получим из (2.4)



$$\ddot{\alpha} + \omega_0^2 \alpha = 0, \quad \text{где} \quad \omega_0 = \sqrt{\frac{mgl}{J}} \quad (2.5) \quad \text{Рис. 2.2}$$

(при $J = ml^2$ — момент инерции материальной точки относительно оси A — имеем прежний результат (2.2)).

3. Пружинный маятник. Тело массы m прикреплено пружиной жесткости k к стенке. Начало координат — точка O — соответствует положению равновесия, в котором пружина не деформирована — нет сил, смещающих тело влево или вправо. При смещении тела из этого положения возникает упругая сила:

$$F = -kx, \quad (2.6)$$

где x — величина деформации (растяжения или сжатия) пружины, равная в нашем случае смещению от положения $x = 0$.

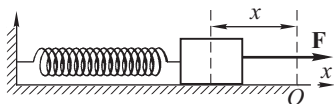


Рис. 2.3

Знак силы, согласно (2.6), противоположен знаку смещения. При $x < 0$ (как показано на рис. 2.3) $F > 0$: сила возвращает тело в положение равновесия. Второй закон Ньютона имеет вид: $m\ddot{x} = -kx$ или

$$\ddot{x} + \omega_0^2 x = 0, \quad \text{где} \quad \omega_0 = \sqrt{\frac{k}{m}}. \quad (2.7)$$

Обратите внимание на математическую тождественность уравнений (2.3), (2.5) и (2.7)!

4. Колебательный контур. Это — пример колебательной системы совершенно иной физической природы. На рис. 2.4 изображен контур, содержащий конденсатор емкости C и катушку индуктивности L . Полагаем, что сопротивление R отсутствует

(в предыдущих примерах мы полагали, что отсутствуют силы трения и сопротивления воздуха при движении).

Согласно закону Ома для замкнутой цепи сумма падений напряжения на всех участках цепи равна сумме ЭДС. Падение напряжения — разность потенциалов между обкладками конденсатора, равна q/C , а ЭДС индукции в катушке

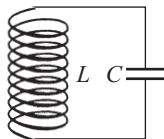


Рис. 2.4

$\mathcal{E}_{\text{инд}} = -L(dI/dt)$, где I — сила тока.

Итак: $q/C = -L(dI/dt)$. Но $I = dq/dt$, следовательно, $dI/dt = d^2q/dt^2 = \ddot{q}$; таким образом $L\ddot{q} + q/C = 0$, или

$$\ddot{q} + \omega_0^2 q = 0, \quad \text{где} \quad \omega_0 = \frac{1}{\sqrt{LC}}. \quad (2.8)$$

Снова получаем уравнение того же вида, что и (2.3), (2.5), (2.7).

5. Электрон в атоме Томсона. Рассмотрим модель атома Томсона, предложенную в 1903 г. Дж.Дж. Томсоном (1856–1940) — до открытия Э. Резерфорда (1871–1937) в 1911 г. — положительно заряженное шарообразное «облако» (его размер R , согласно Томсону, определяет размер атома), в котором под действием кулоновской силы движется отрицательно заряженный, «точечный» электрон. Атом электрически нейтрален, т. е. положительный заряд облака e равен по абсолютной величине отрицательному заряду электрона.

При смещении относительно центра облака на электрон действует сила F , направленная к центру и возвращающая электрон к положению равновесия (рис. 2.5).

Эта сила равна $\mathbf{F} = -e\mathbf{E}$, где \mathbf{E} — напряженность электрического поля, созданного облаком положительного заряда на расстоянии r от его центра. Применим теорему Гаусса к сфере радиуса r : поток вектора \mathbf{E} через поверхность этой сферы равен $(1/\epsilon_0)q$, где q — полный заряд, находящийся внутри сферы. Полагая, что заряд облака e распределен равномерно, получим $q = (r/R)^3 e$. Теорема Гаусса дает: $E \cdot 4\pi r^2 = (1/\epsilon_0)(r/R)^3 e$, или $E(r) = [e/(4\pi\epsilon_0 R^3)]r$. Если \mathbf{r} — радиус-вектор, проведенный из центра в точку, где находится электрон, то сила \mathbf{F} есть

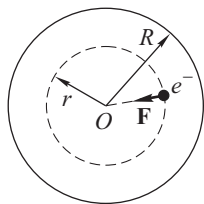


Рис. 2.5

$$\mathbf{F} = -\frac{e^2}{4\pi\epsilon_0 R^3} \mathbf{r}.$$

Второй закон Ньютона (уравнение движения электрона) имеет вид

$$m\ddot{\mathbf{r}} = -\frac{e^2}{4\pi\epsilon_0 R^3} \mathbf{r} \quad (2.9)$$

или

$$\ddot{r} + \omega_0^2 r = 0, \quad (2.10)$$

где $\omega_0 = e/\sqrt{4\pi\epsilon_0 m R^3}$. И снова получаем уравнение того же математического вида!

Гармонический осциллятор. А теперь пора сделать некоторые обобщения. Мы рассмотрели пять примеров из разных областей физики, хотя их можно привести десятки.

Физическое различие между рассмотренными системами в конечном счете проявляется в том, что параметр ω_0 , входящий в уравнения (2.3), (2.5), (2.7), (2.8) и (2.10), определяется различными физическими величинами: ускорением свободного падения g и длиной маятника l в первом примере; ускорением свободного падения, длиной стержня, массой и моментом инерции — во втором; жесткостью пружины k и массой тела m — в третьем; индуктивностью L катушки и емкостью C конденсатора — в четвертом; зарядом электрона, его массой и размером атома — в пятом. Однако одно важнейшее обстоятельство объединяет поведение этих физически разных систем. Каждый раз возникает ситуация, когда система, будучи выведенной из положения равновесия, отвечает на это возникновением силы, стремящейся вернуть систему к равновесию, причем во всех приведенных примерах сила оказывается пропорциональной смещению (т. е. имеет место линейная связь между силой и смещением). Эта линейность, как правило, обеспечивается дополнительным предположением, а именно: малостью отклонения от положения равновесия.

Конечно, понятия «смещение» и «сила» не в каждом случае следует понимать буквально: скажем, в примере с электрическим контуром состояние покоя — это отсутствие каких-либо токов и напряжений в контуре, т. е. положение равновесия это $q = 0$, $I = 0$. Здесь «смещение» из положения равновесия — это отличный от нуля заряд q на пластинах конденсатора. Возникающая при этом разность потенциалов между пластинами приводит к возникновению электрического поля в проводнике и, следовательно, к возникновению тока в цепи — т. е. к постепенному разряду конденсатора (постепенному, потому что быстрому (мгновенному) разряду препятствует возникающая в катушке ЭДС индукции). Так что «силой», возвращающей систему к положению равновесия, является по существу сила, действующая на заряд со стороны электрического поля в проводнике.

Любая физическая система, поведение которой подчиняется уравнению

$$\ddot{S} + \omega_0^2 S = 0, \quad (2.11)$$

называется *гармоническим осциллятором*, а уравнение (2.11) называется *уравнением гармонического осциллятора*.

Итак, согласно (2.11), вторая производная от смещения осциллятора из положения равновесия \ddot{S} пропорциональна (и противоположна по знаку) величине смещения S . Какова же зависимость $S(t)$, удовлетворяющая уравнению (2.11)?

2.2. Свободные колебания гармонического осциллятора

Легко убедиться, что решением уравнения (2.11) является гармоническое колебание (поэтому-то система, описываемая этим уравнением, и называется гармоническим осциллятором).

Действительно, дифференцируя функцию

$$S(t) = a \cos(\omega t + \varphi) \quad (2.12)$$

дважды, находим:

$$\dot{S} = -a\omega \sin(\omega t + \varphi), \quad \ddot{S}(t) = -a\omega^2 \cos(\omega t + \varphi). \quad (2.13)$$

Подставляя (2.12) и (2.13) в (2.11), убеждаемся, что уравнение обращается в тождество, причем значения a и φ могут быть любыми. Другими словами, (2.12) — это множество решений, или, как говорят математики, общее решение, содержащее две произвольные постоянные. Итак, любая из описанных нами физических систем, а также множество других систем, подчиняющихся уравнению (2.11), будучи выведенными из положения равновесия, совершают гармонические колебания. Эти колебания называются *свободными*, поскольку достаточно вывести систему из положения равновесия — и колебания далее происходят «сами по себе», без какого-либо внешнего воздействия.

Произвольные постоянные a и φ — амплитуда и начальная фаза — определяются начальными условиями, тем, как именно система выводится из положения равновесия. Например, маятник можно отклонить от положения равновесия на некоторый угол α_0 и затем отпустить. Этот момент (когда маятник отпускают) и есть начальный момент времени $t = 0$. Мы имеем в этом случае начальные условия: $\alpha(t) = \alpha_0$ при $t = 0$; это условие записывают в виде

$$\alpha(t)|_{t=0} = \alpha_0. \quad (2.14)$$

Одного условия (2.14) недостаточно для определения двух постоянных a и φ . Второе условие — задание начальной скорости. В нашем примере (маятник, отклонив, отпускают) мы имеем

$$\dot{\alpha}(t)|_{t=0} = 0. \quad (2.15)$$

Имея общее решение

$$\alpha(t) = a \cos(\omega t + \varphi) \quad (2.16)$$

и два начальных условия (2.14) и (2.15), можно выбрать из множества решений (2.16) единственное, удовлетворяющее этим условиям. Легко проверить, что оно имеет вид

$$\alpha(t) = \alpha_0 \cos \omega t.$$

Аналогичным образом можно определить колебательный процесс, если маятнику, не отклоняя его от положения равновесия, сообщить начальную скорость $\dot{\alpha}_0$. Тогда начальные условия имеют вид:

$$\alpha(t)|_{t=0} = 0, \quad \dot{\alpha}(t)|_{t=0} = \dot{\alpha}_0,$$

а соответствующее решение

$$\alpha(t) = \frac{\dot{\alpha}_0}{\omega} \sin \omega t.$$

В самом общем случае задача сводится к выбору из множества решений (2.12) единственного решения, удовлетворяющего начальным условиям

$$S(t)|_{t=0} = S_0, \quad \dot{S}(t)|_{t=0} = \dot{S}_0. \quad (2.17)$$

Два условия (2.17) позволяют найти две произвольные постоянные a и φ . Еще раз обратим внимание на замечательную особенность колебаний, совершаемых гармоническим осциллятором: частота колебаний определяется только параметрами самой колебательной системы (l и g — для математического маятника, m и k — для пружинного маятника и т. д.) и не зависит от амплитуды колебаний, т. е. не зависит от начальных условий (от того, каким образом осциллятор выведен из положения равновесия). Разумеется, все сказанное справедливо при условии, что амплитуда остается достаточно малой, что и обеспечивает линейную связь между возвращающей силой и величиной смещения из положения равновесия.

2.3. Превращения энергии при свободных колебаниях гармонического осциллятора

Почему все-таки возникают колебания? Одно обстоятельство мы уже отмечали: при смещении осциллятора из положения равновесия возникает сила, стремящаяся вернуть систему в это положение. Но ведь можно представить себе ситуацию, при которой система, вернувшись в положение равновесия, в этом положении и останется — никаких колебаний не будет. Значит, важно еще одно обстоятельство: инерционность системы, благодаря которой система «проскакивает» положение равновесия и отклоняется от него в противоположную сторону.

Обратимся к рассмотренным выше примерам. Восстанавливающей силой, возвращающей систему в положение равновесия, является: при колебаниях маятника — проекция силы тяжести; при упругих колебаниях груза на пружине — сила Гука; в колебательном контуре — электрическое поле в проводнике, возникающее при заряде конденсатора. Возвращение электрона к положению равновесия в атоме Томпсона обусловлено электрическим полем положительно заряженного «облака».

Инерционность же осциллятора, благодаря которой система, не «останавливаясь», проскакивает положение равновесия, обусловлена тем, что по мере приближения к равновесию осциллятор набирает «скорость»: если величина $S(t)$ при приближении к положению равновесия уменьшается, то «скорость» $\dot{S}(t)$ при этом увеличивается и система «по инерции» отклоняется в противоположную сторону (в колебательном контуре «скорость» — это скорость изменения заряда, т. е. сила тока в проводнике). Ток не исчезает при нулевом заряде — он существует благодаря «инерции» катушки индуктивности.

Если бы обкладки конденсатора были соединены через сопротивление R (без индуктивности: $L = 0$), то по закону Ома мы имели бы $q/C + IR = 0$, и при $q = 0$, $I = 0$, т. е. ток исчезает одновременно с зарядом — никаких колебаний! Инерционность в других примерах связана с наличием массы колеблющегося груза или электрона: тело массы M , набравшее скорость, не может остановиться мгновенно — вот почему положение равновесия и проскакивается!

Процесс колебаний осциллятора полезно рассмотреть с энергетической точки зрения. Вернемся к физическому маятнику, уравнение движения которого (для малых α) имеет вид

$$J\ddot{\alpha} = -mgl\alpha.$$

Умножив обе части равенства на $\dot{\alpha}$, находим

$$J\ddot{\alpha}\dot{\alpha} = -mgl\alpha\dot{\alpha} \quad \text{или} \quad \frac{d}{dt} \left(\frac{J\dot{\alpha}^2}{2} + \frac{mgl\alpha^2}{2} \right) = 0.$$

Мы получили закон сохранения энергии

$$E_0 = \frac{J\dot{\alpha}^2}{2} + \frac{mgl\alpha^2}{2} = \text{const.} \quad (2.18)$$

Первое слагаемое — кинетическая энергия маятника, второе — потенциальная энергия (при отклонении на угол α центр масс поднимается на высоту $h = l(1 - \cos \alpha) \approx l\alpha^2/2$). Согласно (2.18) в процессе колебаний суммарная энергия остается неизменной.

Подставив в (2.18) выражения $\alpha(t) = \alpha_0 \cos(\omega_0 t + \varphi)$ и $\dot{\alpha}(t) = -\alpha_0 \omega_0 \sin(\omega_0 t + \varphi)$, легко получить (упражнение для читателя)

$$E_0 = \frac{J\alpha_0^2\omega_0^2}{2} = \text{const.} \quad (2.19)$$

(Для математического маятника $J = ml^2$ имеем $E_0 = ml^2\alpha_0^2\omega_0^2/2 = \text{const.}$)

Составляющие полной энергии в процессе колебаний, очевидно, изменяются по закону

$$\begin{aligned} E_{\text{кин}} &= \frac{J\alpha_0^2\omega_0^2}{2} \sin^2(\omega_0 t + \varphi) = \frac{E_0}{2} - \frac{E_0}{2} \cos(2\omega_0 t + 2\varphi); \\ E_{\text{пот}} &= \frac{mgl\alpha_0^2}{2} \cos^2(\omega_0 t + \varphi) = \frac{E_0}{2} + \frac{E_0}{2} \cos(2\omega_0 t + 2\varphi). \end{aligned} \quad (2.20)$$

Как следует из (2.20), как кинетическая, так и потенциальная энергия колеблются с частотой $2\omega_0$, причем изменяются в противофазе: в моменты, когда максимальна кинетическая энергия, потенциальная энергия обращается в нуль — это момент прохождения маятником положения равновесия. В момент максимального отклонения от положения равновесия равна нулю кинетическая энергия — при этом максимальна его потенциальная энергия. Согласно определению среднее за время T значение меняющейся во времени величины $E(t)$ есть

$$\langle E(t) \rangle = \frac{1}{T} \int_0^T E(t) dt.$$

Из (2.20) ясно, что среднее за период колебаний значение кинетической энергии осциллятора равно среднему значению потенциальной энергии и равно половине полной энергии:

$$\langle E_{\text{кин}} \rangle = \langle E_{\text{пот}} \rangle = \frac{E_0}{2}.$$

Аналогичным образом можно рассмотреть энергетические изменения в идеальном ($R=0$) колебательном контуре. Умножив уравнение колебаний $q/C + L\ddot{q} = 0$ на \dot{q} , имеем

$$\frac{\dot{q}q}{C} + L\dot{q}\ddot{q} = 0 \quad \text{или} \quad \frac{d}{dt} \left(\frac{q^2}{2C} + \frac{L\dot{q}^2}{2} \right) = 0.$$

Мы получаем закон сохранения энергии

$$E_0 = \frac{q^2}{2C} + \frac{L\dot{q}^2}{2} = \text{const.} \quad (2.21)$$

Первое слагаемое $q^2/(2C)$ — энергия конденсатора (т. е. энергия его электрического поля), второе $L\dot{q}^2/2$ — энергия

катушки индуктивности (т. е. энергия магнитного поля). Изменения заряда $q(t)$ и тока $\dot{q}(t)$ происходят по закону $q = q_0 \cos(\omega_0 t + \varphi)$, $\dot{q}(t) = -q_0 \omega_0 \sin(\omega_0 t + \varphi) = -I_0 \sin(\omega_0 t + \varphi)$, где $I_0 = q_0 \omega_0$ — амплитуда колебаний тока.

Используя последние выражения, а также $\omega_0^2 = 1/(LC)$, можно получить из (2.21)

$$E_0 = \frac{Lq_0^2\omega_0^2}{2} = \frac{LI_0^2}{2} = \frac{q_0^2}{2C}.$$

Первое выражение $LI_0^2/2$ — это максимальное значение магнитной энергии, второе $q_0^2/(2C)$ — максимальное значение электрической энергии. Законы изменения электрической и магнитной энергии имеют вид

$$E_{\text{э}} = \frac{q_0^2}{2C} \cos^2(\omega_0 t + \varphi) = \frac{E_0}{2} + \frac{E_0}{2} \cos(2\omega_0 t + 2\varphi);$$

$$E_{\text{м}} = \frac{Lq_0\omega_0^2}{2} \sin^2(\omega_0 t + \varphi) = \frac{E_0}{2} - \frac{E_0}{2} \cos(2\omega_0 t + 2\varphi).$$

Таким образом, электрическая и магнитная энергии изменяются с частотой $2\omega_0$, причем изменяются в противофазе: в те моменты времени, когда максимальна электрическая энергия, магнитная энергия обращается в нуль, и наоборот, причем $\langle E_{\text{э}} \rangle = \langle E_{\text{м}} \rangle = E_0/2$.

Проследим за энергетическими превращениями при колебаниях электрона в атоме Томсона. При равномерном распределении положительного заряда внутри облака легко найти распределение потенциала электрического поля как функции расстояния от центра O . Принимая потенциал центра облака за нуль, найдем

$$\varphi(r) = \frac{-e}{8\pi\epsilon_0 R^3} r^2.$$

Потенциальная энергия электрона в этом поле (учитывая, что заряд электрона равен $-e$) равна

$$E_{\text{п}} = \frac{e^2}{8\pi\epsilon_0 R^3} r^2.$$

Кинетическая энергия есть

$$E_{\text{к}} = \frac{mr^2}{2}.$$

Закон сохранения энергии

$$E_0 = \frac{e^2 r^2}{8\pi\epsilon_0 R^3} + \frac{mr^2}{2} = \text{const} \quad (2.22)$$

можно получить (как и в предыдущих примерах) из уравнения колебаний (2.9): после умножения (2.9) на \dot{r} , находим

$$m\dot{r}\ddot{r} + \frac{e^2}{4\pi\epsilon_0 R^3} r\dot{r} = 0 \quad \text{или} \quad \frac{d}{dt} \left(\frac{m\dot{r}^2}{2} + \frac{e^2 r^2}{8\pi\epsilon_0 R^3} \right) = 0.$$

Интегрируя последнее равенство, приходим к закону сохранения энергии (2.22). В процессе колебаний изменяются как кинетическая энергия электрона, так и его потенциальная энергия. При максимальном смещении электрона от центра $r = r_0$ его потенциальная энергия максимальна, при этом кинетическая энергия обращается в нуль, при прохождении положения равновесия $r = 0$ максимальна кинетическая энергия, при этом его потенциальная энергия равна нулю.

2.4. Затухающие колебания гармонического осциллятора

Осциллятор, совершающий гармонические колебания — это идеализация. Мы уже отмечали, что бесконечно длящийся процесс вида $S(t) = a \cos(\omega t + \varphi)$ невозможен. Свободные колебания любого реального осциллятора рано или поздно затухают и прекращаются.

Примем теперь во внимание то, что мы сознательно упускали до сих пор, а именно наличие «сопротивления» движению осциллятора, которого не избежать в любой реальной системе. В качающемся маятнике — это трение в оси и сопротивление среды, в которой колеблется маятник. То и другое создает тормозящий момент сил. В пружинном маятнике к сопротивлению среды добавляется сила трения, возникающая при скольжении груза. В колебательном контуре имеется омическое сопротивление, обусловленное, по сути, столкновениями электронов с атомами кристаллической решетки в проводнике. Колеблющийся электрон в атоме Томсона излучает электромагнитную энергию, уносящую энергию колебаний — это также приводит к затуханию (т. н. радиационное затухание).

Рассмотрим влияние «тормозящей силы», зависящей от скорости. Во многих случаях (при малых скоростях) эта зависимость оказывается линейной.

Пример 1. Пружинный маятник (см. рис. 2.3). Уравнение движения (второй закон Ньютона) при наличии вязкого трения, пропорционального скорости ($F_{\text{тр}} = -\beta\dot{x}$) имеет вид

$$m\ddot{x} = -kx - \beta\dot{x} \quad \text{или} \quad \ddot{x} + 2\delta\dot{x} + \omega_0^2 x = 0, \quad (2.23)$$

где $\delta = \beta/(2m)$, $\omega_0 = \sqrt{k/m}$.

Пример 2. Колебательный контур при наличии омического сопротивления. Закон Ома для замкнутой цепи:

$$IR + \frac{q}{C} = -L \frac{dI}{dt} \quad (2.24)$$

(сумма падений напряжения на сопротивлении IR и на конденсаторе q/C равна ЭДС индукции $-L dI/dt$ в катушке индуктивности).

Так как $I = \dot{q}$ и $dI/dt = \ddot{q}$, то последнее уравнение преобразуется к виду

$$\ddot{q} + 2\delta\dot{q} + \omega_0^2 q = 0, \quad (2.25)$$

где $\delta = R/(2L)$ и $\omega_0 = 1/\sqrt{LC}$.

Уравнения (2.23) и (2.25) оказались математически тождественными, хотя описывают поведение совершенно различных физических систем.

Можно убедиться (упражнение для читателя), что при наличии силы сопротивления, пропорциональной скорости, $F_{\text{сопр}} = -\beta l \dot{\alpha}$, уравнение колебаний физического маятника (см. рис. 2.2) имеет вид

$$\ddot{\alpha} + 2\delta\dot{\alpha} + \omega_0^2 \alpha = 0, \quad (2.26)$$

где $\delta = \beta l^2/(2J)$, $\omega_0 = \sqrt{mgl/J}$.

Угол отклонения маятника от положения равновесия описывается уравнением того же вида, что и изменение заряда на конденсаторе в колебательном контуре с сопротивлением. Следует обратить внимание на общность уравнений (2.23)–(2.26), описывающих колебания различной физической природы.

К аналогичным уравнениям приводит множество различных физических задач (в частности, уравнение колебаний классического электрона в атоме с учетом радиационного затухания имеет тот же вид, что и уравнение, описывающее колебания маятника (2.26), учитывающее вязкое трение).

Итак, во всех рассмотренных случаях колебания описываются одним и тем же уравнением

$$\ddot{S} + 2\delta\dot{S} + \omega_0^2 S = 0, \quad (2.27)$$

в котором коэффициент δ называют *коэффициентом затухания* (или просто затуханием), а величину ω_0 — *собственной частотой осциллятора*. Это та самая частота, с которой колебался бы осциллятор, если бы вязкое трение отсутствовало. Важно обратить внимание, что параметр δ в уравнении (2.27) имеет размерность с^{-1} , т. е. совпадает с размерностью параметра ω_0 .

Общность уравнений, описывающих поведение различных физических систем, подсказывает путь изучения этих систем, основанный на аналогиях.

Параметры, входящие в уравнения (2.24)–(2.26) сведены в табл. 2.1.

Таблица 2.1

Параметр	Физический маятник	Пружинный маятник	Колебательный контур
Колеблющаяся величина $S(t)$	Угол отклонения $\alpha(t)$	Деформация пружины $x(t)$	Заряд конденсатора $q(t)$
Затухание δ	$\beta l^2/(2J)$	$\beta/(2m)$	$R/(2L)$
Собственная частота ω_0	$\sqrt{mgl/J}$	$\sqrt{k/m}$	$1/\sqrt{LC}$

Метод подобий позволяет заключить, что если параметры δ и ω_0 в физически различных системах, описываемых уравнением (2.27), оказываются одинаковыми, то одинаковы и закономерности поведения этих систем — одинаков характер изменения величины $S(t)$ (будь то заряд конденсатора, деформация пружины либо угол отклонения маятника). Разумеется, одинаковы должны быть при этом и начальные условия.

Каков вид функции $S(t)$, удовлетворяющей уравнению (2.27)? Мы убедимся далее, что затухающий процесс $S(t)$ имеет качественно различный характер в двух случаях: 1) $\delta < \omega_0$ (это случай малого затухания) и 2) $\delta > \omega_0$ — большое затухание. В случае, если выполняется равенство $\delta = \omega_0$, режим поведения осциллятора называется *критическим*. Итак, рассмотрим подробнее различные ситуации.

Осциллятор с малым затуханием ($\delta < \omega_0$). Качественно ясно, что процесс $S(t)$ должен иметь вид, изображенный на рис. 2.6: колебания, размах которых («амплитуда» $a(t)$) постепенно уменьшается.

Убедимся, что функция

$$S(t) = Ae^{-\delta t} \cos(\omega t + \varphi), \quad (2.28)$$

где $\omega = \sqrt{\omega_0^2 - \delta^2}$, является решением уравнения (2.26) при произвольных значениях A и φ . Действительно, получаем из (2.28):

$$\dot{S} = -A\delta e^{-\delta t} \cos(\omega t + \varphi) - A\omega e^{-\delta t} \sin(\omega t + \varphi), \quad (2.29)$$

$$\ddot{S} = A\delta^2 e^{-\delta t} \cos(\omega t + \varphi) + 2A\omega\delta e^{-\delta t} \sin(\omega t + \varphi) - A\omega^2 e^{-\delta t} \cos(\omega t + \varphi).$$

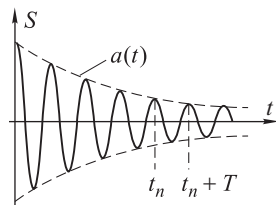


Рис. 2.6

Подставляя функцию (2.28), а также ее производные (2.29) в уравнение (2.27), легко убедиться (упражнение для читателя), что уравнение обращается в тождество, т. е. функция (2.28) действительно является решением уравнения (2.27). Это решение, зависящее от двух произвольных постоянных, является (как доказывается в теории дифференциальных уравнений) общим решением, т. е. любое решение (при $\delta < \omega_0$) может быть получено из (2.28) при соответствующем выборе констант A и φ .

Исследуем подробнее структуру решения (2.28). Функция $S(t)$ представляет собой произведение двух функций: гармонического колебания, частота которого

$$\omega = \sqrt{\omega_0^2 - \delta^2} \quad (2.30)$$

меньше частоты ω_0 того гармонического колебания, которое имел бы осциллятор без затухания ($\delta = 0$), и функции

$$a(t) = Ae^{-\delta t}, \quad (2.31)$$

которая описывает постепенное экспоненциальное уменьшение амплитуды колебания. Функция $a(t)$ представляет собой закон амплитудной модуляции колебания (2.28) (штриховая кривая на рис. 2.6).

Затухающий колебательный процесс не является в строгом смысле периодическим. Действительно, рассмотрим момент времени t_n , когда $\cos(\omega t_n + \varphi) = 1$ и следовательно $S(t_n) = a(t_n) = Ae^{-\delta t_n}$. Ясно, что при любом $t > t_n$ $S(t) < S(t_n)$, т. е. значение функции $S(t_n)$ не повторяется ни при каких $t > t_n$. Ясно в то же время, что благодаря гармоническому множителю $\cos(\omega t + \varphi)$ функция $S(t)$ обладает определенной повторяемостью: в частности, повторяются через равные промежутки времени как нулевые значения функции $S(t)$, так и ее максимумы и минимумы (докажите это!).

Значения амплитуды колебания в два момента времени t_n и $t_n + \tau$ связаны соотношением

$$\frac{a(t_n)}{a(t_n + \tau)} = \frac{Ae^{-\delta t_n}}{Ae^{-\delta(t_n + \tau)}} = e^{\delta \tau}.$$

При $\tau = 1/\delta$ имеем

$$\frac{a(t_n)}{a(t_n + \tau)} = e,$$

т. е. амплитуда колебаний уменьшается в e раз через промежуток времени $\tau = 1/\delta$.

Интервал времени $\tau = 1/\delta$ называют *постоянной времени осциллятора*. Это оценка времени, в течение которого продолжается процесс свободных колебаний осциллятора, выведенного

из положения равновесия. Разумеется, по истечении времени τ колебания продолжают, но амплитуда, спадая по экспоненциальному закону (2.31), становится столь малой, что практически можно полагать, что колебания прекратились (скажем, через время $\tau = 5/\delta$ амплитуда падает более чем в 100 раз).

Рассмотрим теперь два момента времени t_n и $t_n + T$, разделенных промежутком времени $T = 2\pi/\omega$, равным периоду колебания (см. рис. 2.6). Тогда получаем

$$\frac{a_n}{a_{n+1}} = e^{\delta T}$$

или

$$d = \ln \frac{a_n}{a_{n+1}} = \delta T. \quad (2.32)$$

Параметр d (логарифм отношения амплитуд при двух последовательных отклонениях осциллятора от положения равновесия) называется *логарифмическим декрементом затухания*. Он показывает, на сколько изменяется амплитуда колебаний за 1 период. Например, при $d = 0,01$ амплитуда колебаний изменяется за 1 период приблизительно на 1% (убедитесь в этом, используя определение (2.32)). Пусть N --- число колебаний функции (2.28), после которых амплитуда уменьшается в e раз. Тогда $\tau = NT$ или, используя (2.32):

$$d = \delta T = \frac{T}{\tau} = \frac{1}{N}, \quad (2.33)$$

т. е. логарифмический декремент есть величина, обратная числу колебаний, в течение которых амплитуда уменьшается в e раз. В приведенном выше примере (при $d = 0,01$) проходит $N = 100$ колебаний прежде, чем амплитуда уменьшится в e раз.

При исследовании осцилляторов интерес представляет как абсолютная скорость затухания свободных затухающих колебаний, характеризующая постоянной времени $\tau = 1/\delta$ и измеряемая в секундах, так и относительная скорость затухания, характеризующая параметром $N = 1/d$ (и измеряемая числом колебаний, совершаемых осциллятором за время затухания).

Введем еще один параметр, которым принято характеризовать затухающий гармонический осциллятор. Величина Q , определяемая равенством

$$Q = \frac{\pi}{d} \quad (2.34)$$

называется *добротностью осциллятора*. Используя (2.33), получаем:

$$Q = \frac{\pi}{\delta T} = \pi N = \frac{\omega_0}{2\delta}.$$

Последнее равенство справедливо для осциллятора с малым затуханием: $\omega_0 \gg \delta$, поэтому

$$\omega = \frac{2\pi}{T} \approx \omega_0.$$

Величина Q , как и d , является характеристикой относительной скорости затухания колебательного процесса: чем больше добротность, тем большее число колебаний совершает свободный осциллятор, прежде чем его колебания заметно затухают (амплитуда уменьшается в « e » раз). Ниже мы выясним энергетический смысл понятия добротности.

Рассмотрим энергетические превращения при затухающих колебаниях груза, колеблющегося на пружине. Вновь вернемся к уравнению колебаний (второй закон Ньютона):

$$m\ddot{x} + kx = -\beta\dot{x}, \quad (2.35)$$

где слагаемое в правой части $F_{\text{тр}} = -\beta\dot{x}$ — сила вязкого трения, пропорциональная скорости. Домножив обе части уравнения (2.35) на \dot{x} :

$$m\dot{x}\ddot{x} + kx\dot{x} = -\beta\dot{x}^2,$$

преобразуем последнее равенство к виду $\frac{d}{dt} \left(\frac{m\dot{x}^2}{2} + \frac{kx^2}{2} \right) = -\beta\dot{x}^2$ или

$$d \left(\frac{m\dot{x}^2}{2} + \frac{kx^2}{2} \right) = -\beta\dot{x}^2 dt. \quad (2.36)$$

Левая часть равенства — изменение полной энергии колебаний, суммы кинетической энергии груза $m\dot{x}^2/2$ и потенциальной энергии деформированной пружины $kx^2/2$.

Правая часть равенства — произведение силы трения $F_{\text{тр}} = -\beta\dot{x}$ на перемещение $dx = \dot{x} dt$ — представляет собой работу силы трения $dA_{\text{тр}} = F_{\text{тр}} dx$. Эта работа отрицательна, поскольку направления силы и перемещения dx имеют разные знаки в любой момент времени. Таким образом, равенство (2.36) описывает энергетические превращения при колебаниях пружинного маятника. Полная энергия колебаний E уменьшается за счет работы силы трения:

$$dE = -\beta\dot{x}^2 dt. \quad (2.37)$$

Проинтегрируем обе части равенства (2.37) за период колебания. В левой части в результате интегрирования получаем полное изменение энергии колебаний ΔE (убыль энергии) за период:

$$\int_t^{t+T} dE = E(t+T) - E(t) = -\Delta E. \quad (2.38)$$

Интеграл в правой части (работа силы трения за период колебания)

$$A_{\text{тр}} = - \int_t^{t+T} \beta\dot{x}^2 dt$$

преобразуем следующим образом:

$$A_{\text{тр}} = - \int_t^{t+T} \beta \dot{x}^2 dt = \frac{2\beta T}{m} \left(\frac{1}{T} \int_t^{t+T} \frac{m \dot{x}^2}{2} dt \right). \quad (2.39)$$

В скобках в правой части — средняя за период колебаний кинетическая энергия $\langle E_{\text{кин}} \rangle$. При малом затухании $\delta \ll \omega_0$ можно считать, что средняя за период кинетическая энергия равна средней потенциальной энергии $\langle E_{\text{пот}} \rangle$, а полная энергия E_0 есть их сумма, т. е. вдвое больше средней кинетической энергии (как и для осциллятора без затухания).

Тогда получаем из (2.38) и (2.39) $\Delta E = (\beta T/m) E_0$, или (так как $\beta/m = 2\delta$, где δ — коэффициент затухания)

$$\frac{E_0}{\Delta E} = \frac{m}{\beta T} = \frac{1}{2\delta T} = \frac{Q}{2\pi}. \quad (2.40)$$

Итак, отношение средней энергии колебаний к потерям энергии за период равно добротности осциллятора (с множителем $1/2\pi$).

Исследуем энергетические превращения в колебательном контуре. Перепишем уравнение, определяющее процесс колебания в контуре (закон Ома для замкнутой цепи, содержащей индуктивность, емкость и омическое сопротивление) в виде $L\ddot{q} + q/C = -R\dot{q}$.

Домножив последнее равенство на \dot{q} , получим

$$L\dot{q}\ddot{q} + \frac{q\dot{q}}{C} = -R\dot{q}^2$$

или

$$\frac{d}{dt} \left(\frac{L\dot{q}^2}{2} + \frac{q^2}{2C} \right) = -R\dot{q}^2. \quad (2.41)$$

В левой части (2.41) — производная по времени от полной энергии контура (суммы магнитной энергии катушки индуктивности и электрической энергии конденсатора).

В правой части — мощность джоулевых потерь. Итак: убыль полной энергии контура в единицу времени равна мощности джоулевых потерь — количеству тепла, выделяющемуся на омическом сопротивлении.

Интегрирование (2.41) за время одного периода (от t до $t + T$) дает в левой части равенства изменение полной энергии за период ΔE .

Интеграл в правой части (джоулевы потери за период) запишем в виде

$$\int_t^{t+T} R\dot{q} dt = 2\frac{RT}{L} \left(\frac{1}{T} \int_t^{t+T} \frac{L\dot{q}^2}{2} dt \right).$$

Выражение в скобках — среднее за период значение магнитной энергии $\langle E_m \rangle$, равной (при малом затухании) половине полной энергии E_0 . Поэтому

$$\Delta E = \frac{RT}{L} E_0 = 2\delta T E_0.$$

Вновь получаем (как и для пружинного маятника)

$$\frac{E_0}{\Delta E} = \frac{Q}{2\pi}, \quad (2.42)$$

где $Q = \pi/\delta T$ — добротность колебательного контура.

Мы рассмотрели лишь два примера, но вывод, заключенный в равенстве (2.42), является общим для любого осциллятора с затуханием. *Добротность осциллятора* есть (с точностью до множителя 2π) *отношение полной энергии колебаний к потерям энергии за период*. Соотношение (2.42) можно считать определением понятия *добротности осциллятора*.

Осциллятор с большим затуханием ($\delta > \omega_0$). Обратите внимание, что функция $S(t)$, описываемая формулой (2.28), не имеет в этом случае ясного физического смысла: ведь частота $\omega = \sqrt{\omega_0^2 - \delta^2}$ становится при $\delta > \omega_0$ мнимым числом. Итак, прежнее решение, имеющее вид затухающего колебательного процесса, теперь не годится.

Напомним, что в теории дифференциальных уравнений предлагается следующий способ решения. Поставим в соответствие уравнению (2.27) алгебраическое уравнение: $\lambda^2 + 2\delta\lambda + \omega_0^2 = 0$ (это уравнение называется *характеристическим*). Его корни

$$\lambda_{1,2} = -\delta \pm \sqrt{\delta^2 - \omega_0^2} \quad (2.43)$$

действительные (при $\delta > \omega_0$), причем оба корня отрицательны. Общее решение уравнения (2.27) имеет вид

$$S(t) = Ae^{\lambda_1 t} + Be^{\lambda_2 t}. \quad (2.44)$$

Непосредственной подстановкой легко убедиться, что функция (2.44) действительно является решением уравнения (2.27) при произвольных значениях констант A и B .

Эти две константы могут быть найдены, если заданы два начальных условия (например, функция $S(t)$ и ее производная $\dot{S}(t)$ при $t = 0$).

Итак, решение представляется в виде суммы двух монотонно убывающих экспоненциальных функций (поскольку оба показателя экспонент λ_1 и λ_2 отрицательны). Ясно, что их сумма не может привести к затухающему колебательному решению (т. е. функция $S(t)$ не может, бесконечное число раз изме-

няя знак, проходить через нуль). Действительно, пусть, например, $|\lambda_1| > |\lambda_2|$. Очевидно, что каковы бы ни были начальные условия, при достаточно большом t будет выполнено условие $|Be^{\lambda_2 t}| \gg \gg |Ae^{\lambda_1 t}|$ и решение приближенно (при больших t) будет иметь вид $S(t) \approx Be^{\lambda_2 t}$, т. е. монотонно убывающее решение. Можно показать, что система стремится к положению равновесия, совершив не более одного-двух (в зависимости от начальных условий) колебательных движений. Такой процесс называется аperiodическим. На рис. 2.7 приведен пример процесса $S(t)$, представляющего собой сумму двух затухающих экспонент $Ae^{\lambda_1 t}$ и $Be^{\lambda_2 t}$ (штриховые кривые).

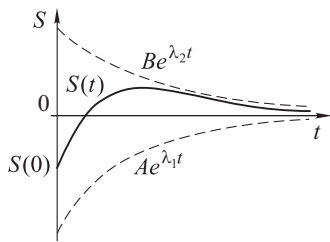


Рис. 2.7

В критическом случае $\delta = \omega_0$ получаем два равных корня характеристического уравнения

$$\lambda_{1,2} = -\delta.$$

В этом случае общее решение уравнения (2.27) имеет вид

$$S(t) = (A + Bt)e^{-\delta t} \quad (2.45)$$

(проверьте непосредственной подстановкой в (2.27), что функция (2.45) является решением при произвольных A и B).

Фазовая плоскость затухающего осциллятора. Напомним, что гармонический колебательный процесс на фазовой плоскости (S — ось абсцисс, \dot{S} — ось ординат) изображается эллипсом, причем полуоси эллипса определяются амплитудой колебания (см. формулу (1.10) и рис. 1.5). Качественно ясно, что траектория, изображающая затухающий колебательный процесс (2.28), должна иметь вид скручивающейся спирали (рис. 2.8). Действительно, поскольку амплитуда колебания (2.28) $a(t) = Ae^{-\delta t}$ постепенно уменьшается, изображающая точка переходит со временем с одного эллипса (отвечающего гармоническому колебанию с определенной амплитудой) на другой эллипс (отвечающий гармоническому колебанию с меньшей амплитудой).

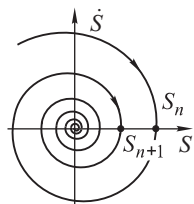


Рис. 2.8

Шаг спирали

$$\Delta a = a_n - a_{n+1} = Ae^{-\delta t} - Ae^{-\delta(t+T)} = Ae^{-\delta t} (1 - e^{-\delta T}).$$

Для осциллятора с малым затуханием $\delta T \ll 1$ получим, разлагая экспоненту $e^{-\delta T}$ в ряд $e^{-\delta T} \approx 1 - \delta T$: $\Delta a = Ae^{-\delta t} d = a(t)d$. Следует обратить внимание, что скручивание спирали (шаг Δa) определяется логарифмическим декрементом затухания, причем с ростом времени t изображающая точка перемещается на все более мелкие витки спирали, расстояние между которыми Δa экспоненциально уменьшается.

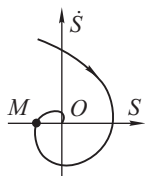


Рис. 2.9

Для осциллятора с большим затуханием $\delta > \omega_0$ характер фазовых траекторий оказывается иным.

Как видно из рис. 2.9, начиная с некоторого момента времени (когда изображающая точка находится в положении M) приближение осциллятора к положению равновесия становится монотонным.

2.5. Связанные осцилляторы

До сих пор мы рассматривали колебательные системы, состояние которых описывается одной переменной (например, угол отклонения маятника от положения равновесия $\alpha(t)$, заряд конденсатора в LC контуре $q(t)$ и т. д.). Сейчас мы рассмотрим системы с двумя степенями свободы: состояние таких систем описывается двумя переменными.

На рис. 2.10а изображены два осциллятора — два маятника, связанные пружиной жесткости k . На рис. 2.10б — два пружинных маятника: тело массы m_1 на пружине жесткости k_1 и тело массы m_2 на пружине жесткости k_2 , эти два осциллятора

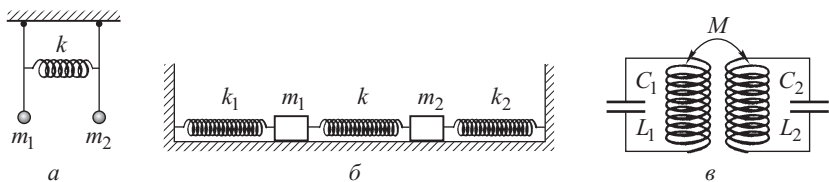


Рис. 2.10

связаны пружиной жесткости k . В этих двух примерах мы имеем в каждом случае два связанных осциллятора. Связь в обоих случаях осуществляется с помощью пружины k . Ее деформация в процессе колебаний приводит к взаимному влиянию друг на друга колеблющихся осцилляторов. На рис. 2.10в — два связанных осциллятора совершенно иной физической природы: два LC контура, причем взаимное влияние осцилляторов друг на друга осуществляется за счет взаимной индуктивности M катушек L_1 и L_2 .

Во всех приведенных примерах мы имеем системы с двумя степенями свободы: для описания состояния каждой из них необходимо иметь две переменные, например, углы $\alpha_1(t)$ и $\alpha_2(t)$ отклонения маятников на рис. 2.10а, координаты тел $x_1(t)$ и $x_2(t)$ на рис. 2.10б, заряды конденсаторов $q_1(t)$ и $q_2(t)$ на рис. 2.10в.

Мы постараемся подметить общие закономерности поведения колебательных систем с двумя степенями свободы, рассмотрев конкретный пример.

Пусть два маятника связаны пружиной (рис. 2.11). Будем рассматривать малые колебания, так что $\sin \alpha_1 \approx \alpha_1$ и $\sin \alpha_2 \approx \alpha_2$.

Пружина жесткости k закреплена на расстоянии h от точек подвеса O маятников, причем при $\alpha_1 = \alpha_2 = 0$ пружина не деформирована.

Уравнение движения ($J\ddot{\alpha} = M$) для первого маятника имеет вид

$$m_1 l_1^2 \ddot{\alpha}_1 = -m_1 g l_1 \alpha_1 + k(h\alpha_2 - h\alpha_1)h, \quad (2.46)$$

где $m_1 l_1^2$ — момент инерции относительно оси O_1 , $-m_1 g l_1 \sin \alpha_1 \approx -m_1 g l_1 \alpha_1$ — момент силы тяжести, $h(\alpha_2 - \alpha_1)$ — деформация (удлинение) пружины, $kh^2(\alpha_2 - \alpha_1)$ — момент упругой силы относительно той же оси.

Аналогично, для второго маятника

$$m_2 l_2^2 \ddot{\alpha}_2 = -m_2 g l_2 \alpha_2 + kh^2(\alpha_1 - \alpha_2). \quad (2.47)$$

Уравнения (2.46) и (2.47) преобразуем к виду

$$\begin{aligned} \ddot{\alpha}_1 + \omega_{01}^2 \alpha_1 &= -\sigma_1^2 (\alpha_1 - \alpha_2), \\ \ddot{\alpha}_2 + \omega_{02}^2 \alpha_2 &= -\sigma_2^2 (\alpha_2 - \alpha_1), \end{aligned} \quad (2.48)$$

где $\omega_{01} = \sqrt{g/l_1}$ и $\omega_{02} = \sqrt{g/l_2}$ — собственные частоты каждого маятника (именно с такой частотой совершал бы гармонические колебания каждый из маятников, если бы не было связи между ними), $\sigma_1^2 = kh^2/m_1 l_1^2$, $\sigma_2^2 = kh^2/m_2 l_2^2$ — коэффициенты, описывающие взаимодействие маятников, обусловленное пружиной.

В общем случае, как мы увидим позднее, колебания маятников $\alpha_1(t)$ и $\alpha_2(t)$ не являются гармоническими. Однако попробуем подобрать такие начальные условия, при которых оба маятника колеблются с одной и той же частотой, совершая гармонические колебания. Рассмотрим для простоты случай одинаковых осцилляторов $\omega_{01} = \omega_{02} = \omega_0$ и $\sigma_1 = \sigma_2 = \sigma$. Тогда мы

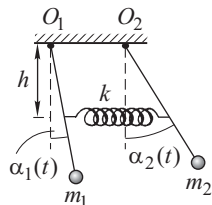


Рис. 2.11

имеем из (2.48):

$$\begin{aligned}\ddot{\alpha}_1 + \omega_0^2 \alpha_1 &= -\sigma^2(\alpha_1 - \alpha_2), \\ \ddot{\alpha}_2 + \omega_0^2 \alpha_2 &= -\sigma^2(\alpha_2 - \alpha_1).\end{aligned}\quad (2.49)$$

Легко сообразить, что если, отклонив одинаковые маятники на один и тот же угол в одну сторону, отпустить их (рис. 2.12), то пружина в процессе колебаний остается недеформированной (как и в положении равновесия), т. е. не оказывает никакого влияния на колебания маятников. Каждый из них будет колебаться с одной и той же частотой ω_0 и $\alpha(t) = \alpha_1(t) = \alpha_2(t) = \alpha_0 \cos \omega_0 t$ (где α_0 — начальный угол отклонения).

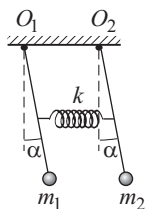


Рис. 2.12

То же самое ясно и из системы уравнений (2.49): при $\alpha_1(t) = \alpha_2(t) = \alpha(t)$ правые части уравнений (2.49) обращаются в нуль и для α_1 и α_2 имеем тождественные уравнения

$$\ddot{\alpha}_1 + \omega_0^2 \alpha_1 = 0, \quad \ddot{\alpha}_2 + \omega_0^2 \alpha_2 = 0.$$

Решение, как мы знаем, — гармоническое колебание частоты ω_0 :

$$\alpha_1(t) = \alpha_2(t) = A \cos(\omega_0 t + \varphi).$$

Можно угадать и еще один возможный тип колебаний в системе, приводящий к простому гармоническому колебанию каждого из маятников с одной и той же частотой. Отклонив маятники на один и тот же угол в противоположные стороны, отпустим их (рис. 2.13).

Ясно, что одинаковые маятники будут качаться таким образом, что в любой момент времени $\alpha_1(t) = -\alpha_2(t)$. Проверим наше предположение, обратившись вновь к системе (2.49).

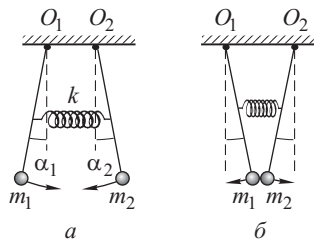


Рис. 2.13

Подставляя в правую часть первого из уравнений $\alpha_2(t) = -\alpha_1(t)$, получим

$$\ddot{\alpha}_1 + \omega_0^2 \alpha_1 = -\sigma^2 \cdot 2\alpha_1.$$

Аналогично, в правую часть второго из уравнений (2.49) подставим

$$\alpha_1(t) = -\alpha_2(t).$$

Мы получаем систему двух независимых уравнений:

$$\begin{aligned}\ddot{\alpha}_1 + (\omega_0^2 + \sigma^2) \alpha_1 &= 0, \\ \ddot{\alpha}_2 + (\omega_0^2 + \sigma^2) \alpha_2 &= 0.\end{aligned}\quad (2.50)$$

Из системы (2.50) следует, что в рассматриваемом случае каждый из двух связанных осцилляторов совершает гармонические колебания с частотой

$$\omega = \sqrt{\omega_0^2 + \sigma^2}$$

(которая оказалась отличной от собственной частоты колебаний ω_0 каждого из маятников).

Найденные нами типы колебаний связанных осцилляторов называются *нормальными колебаниями* или *нормальными модами*, или просто *модами*.

Итак, первый нормальный тип колебаний (или первая мода): оба маятника совершают гармонические колебания с частотой $\omega_1 = \omega_0$, имеют одну и ту же амплитуду и начальную фазу, т. е.

$$\begin{aligned}\alpha_{11}(t) &= A_1 \cos(\omega_1 t + \varphi_1), \\ \alpha_{12}(t) &= B_1 \cos(\omega_1 t + \varphi_1),\end{aligned}\tag{2.51}$$

причем $B_1 = A_1$.

Второй тип колебаний (вторая мода): оба маятника совершают гармонические колебания с частотой $\omega_2 = \sqrt{\omega_0^2 + \sigma^2}$, причем отклонения от положения равновесия в любой момент времени противоположны:

$$\begin{aligned}\alpha_{21}(t) &= A_2 \cos(\omega_2 t + \varphi_2), \\ \alpha_{22}(t) &= B_2 \cos(\omega_2 t + \varphi_2),\end{aligned}\tag{2.52}$$

где $B_2 = -A_2$.

Итак, каждая мода имеет свою собственную характерную частоту ($\omega_1 = \omega_0$ для первой моды и $\omega_2 = \sqrt{\omega_0^2 + \tau^2}$ для второй моды) и свою характерную форму, т. е. взаимное расположение колеблющихся осцилляторов в процессе колебаний.

А каков характер колебаний связанных осцилляторов в самом общем случае?

Можно показать, что в общем случае колебание системы представляет собой суперпозицию нормальных типов (мод), а именно

$$\begin{aligned}\alpha_1(t) &= A_1 \cos(\omega_1 t + \varphi_1) + A_2 \cos(\omega_2 t + \varphi_2), \\ \alpha_2(t) &= B_1 \cos(\omega_1 t + \varphi_1) + B_2 \cos(\omega_2 t + \varphi_2).\end{aligned}\tag{2.53}$$

Из шести констант (A_1 , A_2 , B_1 , B_2 , φ_1 , φ_2) две, как мы знаем, не являются независимыми:

$$B_1 = A_1 \quad \text{и} \quad B_2 = -A_2.$$

Прочие же константы — A_1 , A_2 , φ_1 и φ_2 , определяются начальными условиями: начальными углами отклонений маятников и их начальными скоростями.

Напомним, что ω_1 и ω_2 — это найденные нами частоты собственных мод, т. е. частоты гармонических колебаний, которые можно возбудить в системе с помощью определенных начальных условий, о которых говорилось ранее.

При произвольных же начальных условиях, как следует из (2.53), колебания осцилляторов не являются гармоническими (они представляют собой суперпозицию двух гармонических колебаний разных частот ω_1 и ω_2).

Для проверки справедливости нашего утверждения достаточно подставить функции $\alpha_1(t)$ и $\alpha_2(t)$ в уравнения (2.49) и убедиться в том, что уравнения обращаются в тождество при любых $A_1, A_2, \varphi_1, \varphi_2$.

2.6. Вынужденные колебания гармонического осциллятора (гармоническая внешняя сила)

Ранее мы изучали свободные колебания системы, выведенной из положения равновесия и предоставленной после этого самой себе. Рассмотрим теперь колебательную систему, на которую действует внешняя сила, меняющаяся во времени по гармоническому закону. Колебания, происходящие под действием внешней силы, называют *вынужденными*.

Рассмотрим пружинный маятник (см. рис. 2.3): на тело массы m , кроме упругой силы $f_{\text{упр}} = -kx$ (это — внутренняя сила в колебательной системе), действует внешняя сила $f(t) = f_0 \cos \omega t$. Второй закон Ньютона имеет вид $m\ddot{x} = -kx + f_0 \cos \omega t$, и мы приходим к уравнению

$$\ddot{x} + \omega_0^2 x = a_0 \cos \omega t, \quad (2.54)$$

где $a_0 = f_0/m$, $\omega_0 = \sqrt{k/m}$.

При наличии вязкого трения, пропорционального скорости, вместо (2.54) получаем

$$\ddot{x} + 2\delta\dot{x} + \omega_0^2 x = a_0 \cos \omega t. \quad (2.55)$$

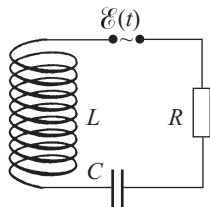


Рис. 2.14

На рис. 2.14 колебательная система другой физической природы: электрический контур, содержащий катушку индуктивности L , конденсатор емкости C и омическое сопротивление R . На этот раз контур подключен к внешней ЭДС, меняющейся по закону $\mathcal{E}(t) = \mathcal{E}_0 \cos \omega t$. Закон Ома для замкнутой цепи, изображенной на рисунке, имеет вид

$$\frac{q}{C} + RI = -L \frac{dI}{dt} + \mathcal{E}(t).$$

По сравнению с (2.24) в последнем уравнении к ЭДС индукции $-L(dI/dt)$, возникающей в катушке при изменении в ней тока, добавляется внешняя ЭДС $\mathcal{E}(t)$. Вновь учитывая, что $I = \dot{q}$, приходим к уравнению

$$\ddot{q} + 2\delta\dot{q} + \omega_0^2 q = a_0 \cos \omega t, \quad (2.56)$$

где $a_0 = \mathcal{E}_0/L$ и по-прежнему $\delta = R/(2L)$ и $\omega_0 = 1/\sqrt{LC}$.

Получено уравнение, математически тождественное уравнению (2.55). Для идеального контура (пренебрегая омическим сопротивлением) вместо (2.56) имеем уравнение, аналогичное (2.54):

$$\ddot{q} + \omega_0^2 q = a_0 \cos \omega t. \quad (2.57)$$

Как и ранее, независимо от физической природы осциллятора, будем обозначать искомую переменную через $S(t)$.

Уравнение, описывающее процесс вынужденных колебаний осциллятора, находящегося под действием гармонической внешней силы, имеет вид

$$\ddot{S} + 2\delta\dot{S} + \omega_0^2 S = a_0 \cos \omega t. \quad (2.58)$$

В дальнейшем удобно, введя вместо константы a_0 константу $a = a_0/\omega_0^2$, записать уравнение вынужденных колебаний (2.60) в виде

$$\ddot{S} + 2\delta\dot{S} + \omega_0^2 S = \omega_0^2 a \cos \omega t. \quad (2.59)$$

Фактически такая замена означает, что мы переходим к изучению колебаний величины S , размерность которой совпадает с размерностью «внешней силы» $a \cos \omega t$.

В частности, если речь идет о колебательном контуре, подключенном к внешней ЭДС $\mathcal{E}(t) = \mathcal{E}_0 \cos \omega t$, то в уравнении (2.59) $S(t)$ есть напряжение на конденсаторе. Если исследуются вынужденные колебания пружинного маятника, на который действует внешняя гармоническая сила $f(t) = f_0 \cos \omega t$, то $S(t)$ в (2.59) — это величина $kx(t)$, т. е. изменение упругой силы при колебаниях тела на пружине. (Предоставляем убедиться в этом читателю.)

В теории дифференциальных уравнений доказывается, что общее решение уравнения (2.59), которое математики называют неоднородным, представляет собой сумму общего решения однородного уравнения (2.27) и некоторого «частного» решения неоднородного уравнения (2.58):

$$S(t) = S_0(t) + S_1(t).$$

Общее решение *однородного уравнения* (2.27) (т. е. *уравнения с нулевой правой частью*) уже известно. Это решение есть не что иное, как процесс свободных колебаний осциллятора.

В зависимости от соотношения параметров δ и ω_0 , оно имеет вид либо (2.28) при $\delta < \omega_0$, либо (2.44) при $\delta > \omega_0$. Характер колебательного процесса $S_0(t)$ зависит от двух произвольных постоянных. Эти постоянные определяются начальными условиями, т. е. способом, которым возбуждаются свободные колебания (начальным отклонением осциллятора от положения равновесия и его начальной скоростью).

Но в любом случае, при наличии даже как угодно малого затухания (а в реальном осцилляторе затухание всегда имеет место) свободные колебания постепенно затухают и соответствующее решение $S_0(t)$ стремится к нулю. Рано или поздно практически остается лишь второе слагаемое $S(t) \approx S_1(t)$. Именно это частное решение неоднородного уравнения и называют «вынужденными колебаниями», которые уже не зависят от начальных условий и определяются лишь внешней силой (ее амплитудой и частотой).

Перейдем к комплексной записи: наряду с уравнением

$$\ddot{S}_1 + 2\delta\dot{S}_1 + \omega_0^2 S_1 = \omega_0^2 a \cos \omega t \quad (2.60)$$

рассмотрим уравнение

$$\ddot{S}_2 + 2\delta\dot{S}_2 + \omega_0^2 S_2 = \omega_0^2 a \sin \omega t, \quad (2.61)$$

а также уравнение

$$\ddot{g} + 2\delta\dot{g} + \omega_0^2 g = \omega_0^2 a e^{i\omega t}. \quad (2.62)$$

Заметим, что правая часть уравнения (2.62) является линейной комбинацией правых частей (2.60) и (2.61): $e^{i\omega t} = \cos \omega t + i \sin \omega t$.

Очевидно поэтому, что и решение (2.62) — функция $g(t)$ — является аналогичной линейной комбинацией решений $S_1(t)$ и $S_2(t)$:

$$g(t) = S_1(t) + iS_2(t).$$

Это легко проверить, домножив левую и правую части (2.61) на мнимую единицу i и сложив после этого левые и правые части (2.60) и (2.61), учтя при этом, что $\dot{S}_1(t) + i\dot{S}_2(t) = \dot{g}(t)$ и $\ddot{S}_1 + i\ddot{S}_2 = \ddot{g}(t)$.

Из сказанного ясно, что если будет найдено комплексное решение $g(t)$ уравнения (2.62), то искомого действительного решения — реальный процесс вынужденных колебаний $S_1(t)$ — есть просто реальная часть комплексной функции $g(t)$

$$S_1(t) = \operatorname{Re} g(t). \quad (2.63)$$

Решение же $g(t)$ уравнения (2.62) легко угадать. Будем искать это решение в виде

$$g(t) = H a e^{i\omega t}, \quad (2.64)$$

т.е. мы полагаем, что гармоническое внешнее воздействие возбуждает в осцилляторе гармоническое колебание, частота которого совпадает с частотой ω внешнего воздействия, а комплексный множитель $H(\omega) = A(\omega)e^{i\psi(\omega)}$ определяет амплитуду $A(\omega)$ и начальную фазу $\psi(\omega)$ вынужденных колебаний. Это предположение необходимо проверить. Подставив предполагаемое решение (2.64) в уравнение (2.62), убедимся, что можно подобрать комплексный множитель $H(\omega)$ так, чтобы уравнение (2.62) обратилось в тождество.

Действительно, уравнение (2.62) является линейным, т.е. содержит неизвестную функцию и ее производные только в первой степени. В частности, над функцией $g(t)$ в левой части (2.62) выполняется лишь линейная операция дифференцирования, которая оставляет неизменным вид экспоненциального множителя $e^{i\omega t}$

$$\frac{d}{dt} e^{i\omega t} = i\omega e^{i\omega t}, \quad \frac{d^2}{dt^2} e^{i\omega t} = -\omega^2 e^{i\omega t}. \quad (2.65)$$

Подставляя (2.64) в (2.62) и учитывая тождества (2.65), находим

$$-\omega^2 a H e^{i\omega t} + 2\delta(i\omega a H) e^{i\omega t} + \omega_0^2 a H e^{i\omega t} = \omega_0^2 a e^{i\omega t}.$$

Множитель $a e^{i\omega t}$ содержится как в правой части последнего уравнения, так и в его левой части. После сокращения на этот множитель мы приходим к алгебраическому уравнению для определения H , откуда получаем

$$H(\omega) = \frac{\omega_0^2}{(\omega_0^2 - \omega^2) + i2\delta\omega}. \quad (2.66)$$

Запись $H(\omega)$ подчеркивает, что найденное нами комплексное число H зависит от частоты ω вынуждающей силы. Таким образом, мы показали, что функция $g(t)$, определяемая формулой (2.64), действительно является решением уравнения (2.62), если $H(\omega)$ определяется соотношением (2.66).

Предоставляем читателю самостоятельно показать, что модуль комплексного числа $H(\omega)$ есть

$$A(\omega) = |H(\omega)| = \frac{\omega_0^2}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4\delta^2\omega^2}}, \quad (2.67)$$

а его аргумент $\psi(\omega)$ определяется формулой

$$\psi(\omega) = -\operatorname{arctg} \frac{2\delta\omega}{\omega_0^2 - \omega^2}. \quad (2.68)$$

Итак, мы нашли комплексное решение уравнения (2.62), имеющее вид

$$g(t) = aA(\omega)e^{i(\omega t + \psi(\omega))}, \quad (2.69)$$

где $A(\omega)$ и $\psi(\omega)$ определяются выражениями (2.67) и (2.68).

Соответствующий реальный колебательный процесс $S_1(t)$ имеет вид

$$S_1(t) = aA(\omega) \cos(\omega t + \psi(\omega)). \quad (2.70)$$

Подчеркнем еще раз, что речь идет о гармоническом осцилляторе, т. е. о линейной колебательной системе (в которой сила, возвращающая осциллятор в положение равновесия, прямо пропорциональна смещению). Мы получили исключительной важности результат.

Гармоническая внешняя сила частоты ω возбуждает в осцилляторе вынужденные колебания, которые происходят также по гармоническому закону и имеют ту же частоту ω .

Функция $A(\omega)$ определяет амплитуду вынужденных колебаний — ее зависимость от частоты внешней силы; ее называют *амплитудной характеристикой осциллятора* (точнее $A(\omega)$ есть отношение амплитуды вынужденных колебаний к амплитуде a внешнего воздействия, что ясно из (2.70)), а функция $\psi(\omega)$ (2.68) дает сдвиг по фазе между вынужденными колебаниями и колебаниями внешней силы и называется *фазовой характеристикой*.

Комплексную функцию $H(\omega) = A(\omega)e^{i\psi(\omega)}$ называют *частотной (или амплитудно-фазовой) характеристикой* (иногда используют термин *передаточная функция*).

Исследуем подробнее полученные нами зависимости амплитуды вынужденных колебаний и сдвига фазы от частоты внешней силы.

Зависимость интенсивности колебаний (квадрата амплитуды) от частоты ω внешней силы имеет вид

$$A^2(\omega) = \frac{\omega_0^4}{(\omega_0^2 - \omega^2)^2 + 4\delta^2\omega^2}. \quad (2.71)$$

Разделив числитель и знаменатель на $4\delta^2\omega_0^2$ и используя соотношение $Q = \omega_0/2\delta$ (см. ниже, задача 4), получаем

$$A^2(\omega) = \frac{Q^2}{Q^2 \left(1 - \omega^2/\omega_0^2\right)^2 + \omega^2/\omega_0^2}. \quad (2.72)$$

Функцию $A(\omega)$ часто называют *резонансной характеристикой*. Семейство резонансных характеристик для различных значений Q показано на рис. 2.15.

Обратим внимание на важную особенность резонансных характеристик. При $\omega = 0$ имеем $A = 1$. Это означает, что при постоянной внешней силе упругая сила просто равна ей и система остается в равновесии, а если речь идет о колебательном контуре, то при постоянной внешней ЭДС напряжение на конденсаторе равно ЭДС — никаких колебаний нет.

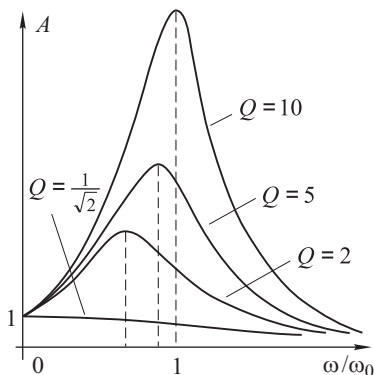


Рис. 2.15

Напомним, что мы рассматриваем колебательную систему в то время, когда переходный процесс, возникающий при включении внешней силы, уже закончился.

При $\omega = \omega_0$ (частота внешней силы совпадает с «собственной частотой» ω_0 колебательной системы) амплитуда вынужденных колебаний в Q раз больше амплитуды «внешней силы». Совпадение частоты внешней силы с собственной частотой ω_0 колебательной системы называют *резонансом*.

Как видно из приведенных графиков, амплитуда вынужденных колебаний максимальна, если речь идет о колебательной системе с большой добротностью $Q \gg 1$. Если Q невелико, то максимум амплитуды вынужденных колебаний наступает при частоте $\omega < \omega_0$. В общем случае, как видно из (2.72), максимум $A(\omega)$ наступает, когда минимален знаменатель дроби.

Найдем минимум знаменателя, приравняв нулю его производную по ω^2 :

$$-2Q^2 \left(1 - \frac{\omega^2}{\omega_0^2}\right) \frac{1}{\omega_0^2} + \frac{1}{\omega_0^2} = 0,$$

откуда

$$\omega^2 = \omega_0^2 \left(1 - \frac{1}{2Q^2}\right). \quad (2.73)$$

Последнее выражение и определяет частоту внешней силы, при которой амплитуда вынужденных колебаний максимальна. Величина максимальной амплитуды равна

$$A_{\max} = a \frac{Q}{\sqrt{1 - 1/(4Q^2)}} \quad (2.74)$$

(эта величина больше, чем на частоте $\omega = \omega_0$, когда амплитуда вынужденных колебаний равна aQ).

Таким образом, максимум амплитуды приходится на частоту ω , несколько меньшую ω_0 , однако при $Q \gg 1$ мы имеем из (2.73) и (2.74) уже отмеченный результат: максимальная амплитуда, равная a_0Q , наступает при $\omega = \omega_0$.

Наконец, рассмотрим вопрос об остроте резонансной кривой колебательной системы с большой добротностью. При каком отклонении частоты от значения $\omega = \omega_0$ амплитуда вынужденных колебаний заметно уменьшается (скажем в $\sqrt{2}$ раз)?

Как видно из (2.72), это происходит при

$$Q^2 \left[1 - \left(\frac{\omega}{\omega_0} \right)^2 \right]^2 + \left(\frac{\omega}{\omega_0} \right)^2 = 2,$$

откуда находим

$$\left(\frac{\omega}{\omega_0} \right)^2 = \left(1 - \frac{1}{2Q^2} \right) \pm \frac{1}{Q} \sqrt{1 + \frac{1}{4Q^2}}.$$

С точностью до членов порядка $1/Q^2$ имеем

$$\frac{\omega}{\omega_0} = 1 \pm \frac{1}{2Q} \quad \text{или} \quad \omega = \omega_0 \left(1 \pm \frac{1}{2Q} \right),$$

и мы получаем

$$\Delta\omega = \omega - \omega_0 = \pm \frac{\omega_0}{2Q} = \pm \delta. \quad (2.75)$$

Итак мы нашли *полуширину резонансной кривой* на уровне $1/\sqrt{2}$, которая равна величине затухания δ .

Заметим наконец, что при $\omega \rightarrow \infty$ амплитуда вынужденных колебаний стремится к нулю: знаменатель в (2.72) при $(\omega/\omega_0)^2 \gg 1$ имеет вид

$$Q^2 \left[1 - \left(\frac{\omega}{\omega_0} \right)^2 \right]^2 + \left(\frac{\omega}{\omega_0} \right)^2 \approx Q^2 \left(\frac{\omega}{\omega_0} \right)^4,$$

если $Q \gg 1$. Поэтому получаем

$$A(\omega) \approx \left(\frac{\omega_0}{\omega} \right)^2 \quad (2.76)$$

при $\omega \gg \omega_0$

Выражение для фазовой характеристики (2.68) можно переписать в виде

$$\psi = -\arctg \frac{\omega/\omega_0}{Q[1 - (\omega/\omega_0)^2]}. \quad (2.77)$$

Эта зависимость для разных значений Q показана на рис. 2.16.

Следует обратить внимание, что с ростом добротности фаза ψ меняется все круче вблизи резонансной частоты, изменяясь от значения $\psi \approx 0$ при $\omega < \omega_0$ до значения $\psi \approx -\pi$ при $\omega > \omega_0$, т. е. если колебательная система имеет высокую добротность, то при частотах $\omega < \omega_0$ вынужденные колебания происходят практически синфазно с колебаниями внешней силы, а при $\omega > \omega_0$ — в противофазе с колебаниями внешней силы.

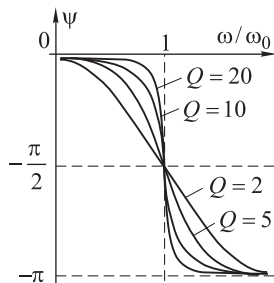


Рис. 2.16

2.7. Гармонические колебания в линейных стационарных системах

При изучении общих свойств линейных систем (*фильтров*) обычно не интересуются их конкретным устройством. Систему изображают в виде блок-схемы (рис. 2.17). Внешнее воздействие $f(t)$ называют «входным сигналом» фильтра, а искомую зависимость $S(t)$ — выходным сигналом (или откликом) фильтра.



Рис. 2.17

Квадратик L представляет собой некоторое устройство, преобразующее входной сигнал $f(t)$ в выходной сигнал $S(t)$. Тот факт, что $S(t)$ является откликом на внешнее воздействие $f(t)$, будем записывать в виде равенства $S(t) = L[f(t)]$, которое означает, что в результате действия входного сигнала $f(t)$ на линейную систему L получаем сигнал на выходе $S(t)$.

Гармонический осциллятор является одним из примеров *линейной колебательной системы*, т. е. системы, колебания в которой подчиняются линейному уравнению (напомним, уравнение линейно, если искомая функция $S(t)$ и ее производные входят в уравнение только в первой степени). Другие примеры линейных систем показаны на рис. 2.18.

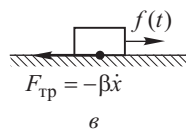
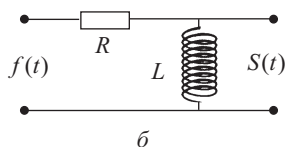
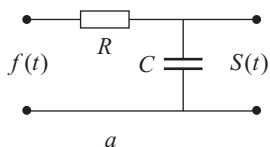


Рис. 2.18

На рис. 2.18а изображена RC -цепочка, колебания в которой возбуждаются внешней ЭДС $f(t)$. Искомой функцией является изменение напряжения на конденсаторе $S(t)$.

На рис. 2.18б показана RL -цепочка. Ищется закон изменения напряжения на катушке индуктивности.

На рис. 2.18в — простой механический пример: изучается движение тела массы m под действием внешней силы, меняющейся во времени по закону $f(t)$, тело движется в среде с вязким трением (сила трения пропорциональна скорости $F_{\text{тр}} = -\beta\dot{x}$). Ищется закон изменения скорости $V(t)$.

Покажем, что во всех этих случаях поведение систем также описывается линейным уравнением с постоянными коэффициентами, и гармоническое внешнее воздействие приводит к возникновению в этих системах вынужденных гармонических колебаний $g(t)$, происходящих с той же частотой, т. е. $g(t) = H(\omega)e^{i\omega t}$. Найдем также частотные характеристики $H(\omega)$ этих систем.

1. Рассмотрим RC -фильтр (см. рис. 2.18а). Закон Кирхгофа имеет вид

$$\dot{q}R + \frac{q}{C} = f(t)$$

(сумма падений напряжения на сопротивлении $\dot{q}R$ и на конденсаторе q/C равна внешней ЭДС $f(t)$). Искомой величиной (выходным сигналом фильтра) является напряжение на конденсаторе $S(t) = q/C$. Тогда напряжение на сопротивлении $\dot{q}R = RC \cdot \dot{S}$ и мы получаем

$$\dot{S} + \frac{1}{RC} S = \frac{1}{RC} f(t).$$

Вновь приходим к линейному уравнению с постоянными коэффициентами.

Каков процесс вынужденных колебаний, если внешним воздействием является гармоническое колебание единичной амплитуды? Уравнение, описывающее процесс колебаний имеет вид

$$\dot{S} + \frac{1}{RC} S = \frac{1}{RC} \cos \omega t.$$

Так же, как и в случае гармонического осциллятора, перейдем к комплексной форме записи:

$$\dot{g} + \frac{1}{RC} g = \frac{1}{RC} e^{i\omega t}. \quad (2.78)$$

Предположим, что и в этом случае, так же как и в гармоническом осцилляторе, вынужденные колебания $g(t)$ происходят по гармоническому закону

$$g(t) = H(\omega)e^{i\omega t}. \quad (2.79)$$

Проверим наше предположение, подставив функцию (2.79) в уравнение (2.78). Мы получаем

$$H(\omega)(i\omega)e^{i\omega t} + \frac{1}{RC} H(\omega)e^{i\omega t} = \frac{1}{RC} e^{i\omega t}.$$

После сокращения на множитель $e^{i\omega t}$ находим частотную характеристику $H(\omega)$

$$H(\omega) = \frac{1}{1 + i\omega RC}. \quad (2.80)$$

Итак, мы показали, что выходной сигнал фильтра, описываемый формулой (2.79), действительно представляет собой гармоническое колебание, если множитель $H(\omega)$ определяется формулой (2.80). Соответственно амплитудная $A(\omega) = |H(\omega)|$ и фазовая $\varphi(\omega) = \arg H(\omega)$ характеристики имеют вид

$$|H(\omega)| = \frac{1}{\sqrt{1 + (\omega RC)^2}}, \quad (2.81)$$

$$\varphi(\omega) = \arctg(\omega RC).$$

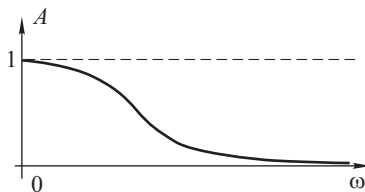


Рис. 2.19

Амплитудная характеристика RC -фильтра показана на рис. 2.19.

2. Аналогичным образом, используя закон Кирхгофа для RL -фильтра (см. рис. 2.18б) и имея ввиду, что выходной сигнал (напряжение на катушке индуктивности) $S(t) = L(dI/dt)$, находим

$$RI + L\dot{I} = \text{const.}$$

Переходя к комплексной форме записи, получаем

$$\dot{g}_1 + \frac{R}{L} g_1 = \frac{1}{L} e^{i\omega t} \quad (I(t) = \text{Re} g_1(t)). \quad (2.82)$$

Ищем решение последнего уравнения в виде $g_1(t) = H_1(\omega)e^{i\omega t}$. После подстановки в уравнение (2.82) получаем

$$H_1(\omega) = \frac{1}{R + i\omega L} \quad \text{и} \quad g_1(t) = \frac{1}{R + i\omega L} e^{i\omega t}.$$

Соответственно

$$g(t) = L\dot{g}_1(t) = \frac{i\omega L}{R + i\omega L} e^{i\omega t}. \quad (2.83)$$

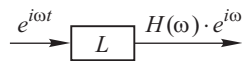
Итак, выходной сигнал $g(t)$ представляет собой гармоническое колебание, причем частотная характеристика RL -фильтра имеет вид

$$H(\omega) = \frac{i\omega L}{R + i\omega L}. \quad (2.84)$$

Далее находим:

$$|H(\omega)| = \frac{\omega L}{\sqrt{R^2 + (\omega L)^2}}; \quad \varphi(\omega) = \operatorname{arctg} \frac{R}{\omega L}.$$

Мы можем сделать важный вывод, общий для всех *линейных стационарных систем* (систем, описываемых линейными уравнениями с постоянными коэффициентами): *гармоническое внешнее воздействие возбуждает в линейной стационарной системе колебания, которые происходят по гармоническому закону с той же частотой*. Этот общий вывод можно обобщить с помощью символического равенства



$$L[e^{i\omega t}] = H(\omega) e^{i\omega t} \quad (2.85)$$

Рис. 2.20 и изобразить с помощью блок-схемы, представленной на рис. 2.20.

В рассмотренных выше примерах (гармонический осциллятор, RC - и RL -фильтры, механический пример) мы показали, как находится частотная характеристика $H(\omega)$, определяющая зависимость амплитуды и фазы вынужденных колебаний от частоты внешнего гармонического воздействия.

2.8. Спектральный анализ линейных колебательных систем

Рассмотрим более общую задачу: колебания гармонического осциллятора возбуждаются внешней силой, меняющейся во времени по произвольному закону $f(t)$.

Вместо уравнения (2.59) имеем в этом случае

$$\ddot{S} + 2\delta\dot{S} + \omega_0^2 S = \omega_0^2 f(t). \quad (2.86)$$

Отличие (2.86) от (2.59) состоит в том, что в правой части последнего уравнения вместо гармонической внешней силы $a_0 \cos \omega t$ — произвольное воздействие $f(t)$.

Фундаментальное свойство всех линейных систем (независимо от их конкретного устройства) состоит в следующем: результат нескольких одновременных воздействий можно найти, суммируя результаты, к которым приводит каждое отдельное воздействие.

Убедимся в этом на примере гармонического осциллятора. Пусть внешняя сила описывается функцией $f_1(t)$. Возникающий при этом процесс колебаний — функцией $S_1(t)$. Мы имеем уравнение

$$\ddot{S}_1 + 2\delta\dot{S}_1 + \omega_0^2 S_1 = \omega_0^2 f_1(t), \quad (2.87)$$

что символически запишем в виде: $L[f_1(t)] = S_1(t)$.

Аналогично, при внешнем воздействии $f_2(t)$ колебания осциллятора описываются функцией $S_2(t)$, подчиняющейся уравнению

$$\ddot{S}_2 + 2\delta\dot{S}_2 + \omega_0^2 S_2 = \omega_0^2 f_2(t) \quad (2.88)$$

или

$$L[f_2(t)] = S_2(t).$$

Как колеблется осциллятор, если внешняя сила изменяется по закону

$$f(t) = c_1 f_1(t) + c_2 f_2(t),$$

где c_1 и c_2 — произвольные постоянные?

Искомый процесс вынужденных колебаний $S(t)$ описывается уравнением

$$\ddot{S} + 2\delta\dot{S} + \omega_0^2 S = \omega_0^2 (c_1 f_1(t) + c_2 f_2(t)). \quad (2.89)$$

Домножив левые и правые части (2.87) и (2.88) соответственно на c_1 и c_2 и сложив их, убеждаемся, что искомая функция $S(t)$ может быть представлена в виде

$$S(t) = c_1 S_1(t) + c_2 S_2(t). \quad (2.90)$$

Обобщая сказанное выше на произвольную линейную систему, приходим к равенству

$$L[c_1 S_1(t) + c_2 S_2(t)] = c_1 L[f_1(t)] + c_2 L[f_2(t)]. \quad (2.91)$$

Итак, *отклик линейной системы на линейную комбинацию внешних воздействий равен линейной комбинации откликов на каждое отдельное воздействие.*

Можно обобщить (2.91), рассматривая произвольную линейную суперпозицию вида:

$$f(t) = \sum_{n=1}^{\infty} c_n f_n(t). \quad (2.92)$$

Тогда вместо (2.91) имеем

$$S(t) = L[f(t)] = \sum c_n L[f_n(t)]. \quad (2.93)$$

Сказанное выше подсказывает путь решения общей задачи изучения колебаний линейной системы под действием произвольной внешней силы $f(t)$.

Целесообразно представить эту силу в виде суммы (2.92) некоторых элементарных слагаемых — элементарных «кирпичиков» (в общем случае сумма может быть бесконечной).

Затем необходимо найти отклик $S_n(t)$ системы на каждое отдельное слагаемое $f_n(t)$

$$S_n(t) = L[f_n(t)],$$

и, наконец, записать искомый процесс вынужденных колебаний в виде линейной суперпозиции откликов

$$S(t) = \sum c_n S_n(t).$$

Мы описали суть спектрального подхода к исследованию линейных систем.

Разумеется, выбор «базиса» — $f_n(t)$ — тех элементарных «кирпичиков», из которых мы складываем «здание» заданной внешней силы $f(t)$, неоднозначен. Это — вопрос физической целесообразности, он зависит от того, насколько просто можно найти реакцию системы на каждое отдельное слагаемое. Мы исследуем сейчас поведение линейных стационарных систем (т. е. линейных систем, свойства которых не меняются со временем: например, колебательный контур с постоянными L , C , R или маятник неизменной длины и массы). Мы выяснили, что по отношению к таким системам особую роль играют *гармонические функции* (записанные в комплексной форме) $e^{i\omega t}$.

Рассматривая гармонический осциллятор, а также примеры других линейных систем, мы убедились в том, что гармоническая внешняя сила $f(t) = e^{i\omega t}$ возбуждает в линейной системе колебания, происходящие также по гармоническому закону с той же частотой ω .

Символически этот результат мы записали в виде равенства (2.85) и изобразили в виде блок-схемы (рис. 2.20).

Напомним: запись (2.85) означает, что отклик линейной системы L на входной сигнал $e^{i\omega t}$ есть $H(\omega)e^{i\omega t}$.

Множитель $H(\omega)$ зависит от свойств линейной системы и называется *частотной характеристикой системы* (мы нашли ее ранее для гармонического осциллятора (см. формулу (2.66), RC - и RL -фильтров и механической системы).

Используя математическую терминологию, можно сказать, что гармонические функции $e^{i\omega t}$ являются собственными функциями линейной системы L . Этот факт и отражает равенство (2.85): гармоническая функция $e^{i\omega t}$ преобразуется системой в гармоническую же функцию $H(\omega)e^{i\omega t}$ (множитель $H(\omega)$ математики называют собственным значением, отвечающим собственной функции $e^{i\omega t}$).

Еще раз подчеркнем, что частота ω в равенстве (2.85) может быть любой, при любых значениях ω равенство (2.85) справедливо, — и каждому значению ω отвечает число $H(\omega)$ — соответствующее собственное значение.

Оказывается, что гармонические функции $e^{i\omega t}$ (при любом ω) обладают двумя важнейшими математическими свойствами, аналогичными свойствам действительных функций $\cos \omega t$. Во-первых, эти функции ортогональны. Смысл этого понятия состоит

в том, что функция $e^{i\omega t}$ при данном, фиксированном значении ω не может быть получена сложением гармонических функций $c_n e^{i\omega_n t}$ любых других частот $\omega_n \neq \omega$, какие бы коэффициенты $c_n(\omega)$ мы не пытались подобрать. Можно сказать, что функция $e^{i\omega t}$ не имеет «проекции» на какую-либо другую функцию $e^{i\omega_n t}$ с $\omega_n \neq \omega$. Второе важнейшее свойство — полнота: любую функцию $f(t)$ можно представить единственным образом в виде суммы гармонических функций. Об этом мы уже говорили ранее.

Все сказанное позволяет утверждать, что в качестве базиса — элементарных слагаемых, из которых целесообразно складывать произвольное внешнее воздействие (если речь идет о стационарных линейных системах) — удобно использовать гармонические функции.

Ясен теперь путь решения общей задачи — возбуждение вынужденных колебаний в линейной системе произвольной внешней силой.

Первый шаг — представление внешней силы в виде суммы гармонических колебаний (либо в виде дискретной суммы — ряда Фурье, либо в виде интеграла Фурье):

$$f(t) = \sum c_n e^{i\omega_n t} \quad \text{или} \quad f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} C(\omega) e^{i\omega t} d\omega.$$

Второй шаг — нахождение частотной характеристики $H(\omega)$ колебательной системы.

Далее, поскольку каждая гармоника $c_n e^{i\omega_n t}$ возбуждают в системе колебание $H(\omega_n) c_n e^{i\omega_n t}$, то суммарная внешняя сила $f(t)$ возбуждает колебание вида

$$S(t) = \sum c_n H(\omega_n) e^{i\omega_n t}. \quad (2.94)$$

Или, в случае непрерывного спектра $C(\omega)$,

$$S(t) = \int C(\omega) H(\omega) e^{i\omega t} d\omega. \quad (2.95)$$

Процесс вынужденных колебаний $S(t)$ (как и любой процесс) можно представить либо в виде суммы гармонических колебаний

$$S(t) = \sum b_n e^{i\omega_n t}, \quad (2.96)$$

либо в виде непрерывной суммы (интеграла Фурье)

$$S(t) = \frac{1}{2\pi} \int B(\omega) e^{i\omega t} d\omega. \quad (2.97)$$

Из сравнения (2.94) и (2.96) следует, что

$$b_n = c_n H(\omega_n), \quad (2.98)$$

а из сравнения (2.95) и (2.97):

$$B(\omega) = C(\omega)H(\omega). \quad (2.99)$$

Равенства (2.98) и (2.99) лежат в основе спектрального анализа линейных колебательных систем. Они позволяют найти спектр вынужденных колебаний, т. е. либо набор коэффициентов b_n (в дискретном случае), либо непрерывный спектр $B(\omega)$, если известен спектр внешней силы и частотная характеристика колебательной системы.

Для иллюстрации изложенного метода рассмотрим одну из важных проблем, которая решается с помощью линейных фильтров — проблему *селекции*, которая возникает, в частности, при приеме радиосигналов.

Пусть на вход колебательного контура приемника поступают сигналы двух радиостанций, ведущих передачи на несущих частотах ω_0 и ω_1 . Это модулированные колебания

$$f_s(t) = a_1(t) \cos(\omega_0 t + \varphi_1(t))$$

и

$$f_n(t) = a_2(t) \cos(\omega_1 t + \varphi_2(t))$$

Их спектры — $C_s(\omega)$ и $C_n(\omega)$. Требуется выделить полезный сигнал $f_s(t)$ и отсеять помехи (сигнал $f_n(t)$).

Требования, предъявляемые к частотной характеристике контура вытекают из соотношений (2.98) и (2.99). Во-первых контур необходимо настроить на несущую частоту сигнала $f_s(t)$, т. е. резонансная частота контура ω_p должна совпадать с ω_0 : $\omega_p = \omega_0$. При этом добротность контура Q должна быть достаточно большой, чтобы в пределы полосы пропускания контура $\Delta\omega_k = \omega_0/Q$ не попали спектральные компоненты помех ω_1 , $|\omega_1 - \omega_0| > \Delta\omega_k$. С другой стороны, чтобы полезный сигнал был принят без искажений, необходимо, чтобы полоса частот полезного сигнала $\Delta\Omega$, определяемая характерным временем τ изменения функций $a(t)$ и $\varphi(t)$, описывающих закон модуляции ($\Delta\Omega \cdot \tau \approx 2\pi$ — соотношение неопределенностей), была меньше полосы пропускания контура $\Delta\Omega < \Delta\omega_k$.

Всем этим условиям удовлетворяет частотная характеристика, показанная на рис. 2.21 (показаны дискретные спектры $C_s(\omega)$ и $C_n(\omega)$ сигналов $f_s(t)$ и $f_n(\omega)$).

Спектр выходного сигнала находим с помощью (2.99)

$$B(\omega) = [C_s(\omega) + C_n(\omega)] \cdot H(\omega) \approx C_s(\omega) \cdot H(\omega_p) \approx Q \cdot C_s(\omega)$$

(спектральные компоненты на частотах $\omega \approx \omega_1$ оказываются подавленными, поскольку $C_n(\omega) \cdot H(\omega) \approx 0$).

Действительно, из выражения для частотной характеристики резонансного контура следует (при $\omega/\omega_0 \gg 1$): $H(\omega) \approx 1$ при

$\omega \ll \omega_p$, $H(\omega) \approx Q$ при $\omega \approx \omega_p$ и $H(\omega) \approx (\omega_0/\omega)^2$ при $\omega \gg \omega_p$. Поэтому первое слагаемое $C_s(\omega)H(\omega) \approx C_s(\omega)Q$, а второе $C_n(\omega)H(\omega) \approx C_n(\omega)(\omega_0/\omega)^2$ при $\omega \gg \omega_0$ пренебрежимо мало.

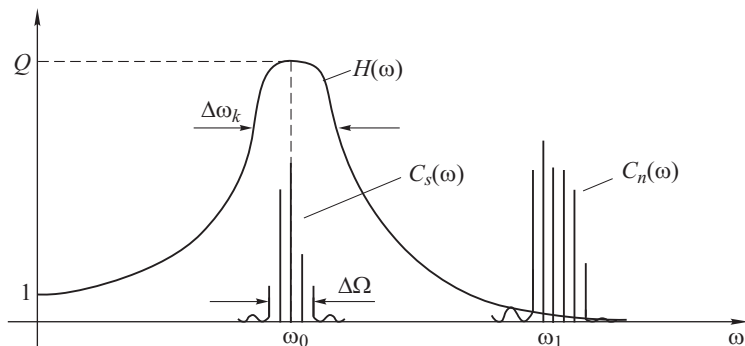


Рис. 2.21

Таким образом, выходной сигнал оказывается отфильтрованным: он содержит лишь спектральные компоненты полезного сигнала $f_s(t)$.

Демодуляция и квадратичное детектирование. После того как с помощью линейного фильтра осуществлена задача селекции — выделение из множества сигналов интересующего нас полезного сигнала (модулированного колебания $f(t) = a(t) \cos(\omega_0 t + \varphi(t))$) необходимо решить проблему демодуляции — преобразования высокочастотного колебательного процесса $f(t)$ (гармонические составляющие которого заполняют интервал частот $\omega_0 - \Delta\Omega < \omega < \omega_0 + \Delta\Omega$ в окрестности несущей частоты ω_0) в низкочастотные звуковые колебания $a(t)$ и $\varphi(t)$, содержащие непосредственно передаваемую информацию (речь, музыка, телевизионное изображение и т.п.). Такую трансформацию частот невозможно осуществить с помощью линейного фильтра. Мы рассмотрим метод *квадратичного детектирования*.

Идеальный квадратичный детектор имеет вольт-амперную характеристику $i = \text{const} \cdot V^2$. Демодулятор должен включать в себя также инерционное устройство, реагирующее на низкочастотные колебания, из которых состоят законы модуляции $a(t)$ и $\varphi(t)$, но не успевающее следить за колебаниями, происходящими с частотами порядка ω_0 .

Пусть функция $a(t)$, описывающая закон амплитудной модуляции, содержит низкочастотные колебания

$$a(t) = \sum c_n e^{in\Omega t}.$$

Тогда модулированное колебание

$$f(t) = a(t) \cos \omega_0 t$$

содержит частоты $\omega_0 \pm n\Omega$:

$$f(t) = \sum c_n e^{i(\omega_0 + n\Omega)t},$$

а ток детектора (пропорциональный $f^2(t)$) есть

$$i(t) \sim \sum c_n e^{i(\omega_0 + n\Omega)t} \cdot \sum c_m e^{i(\omega_0 + m\Omega)t}.$$

Он содержит комбинационные частоты вида $(\omega_0 + n\Omega) \pm (\omega_0 + m\Omega)$, которые распадаются на высокие частоты $2\omega_0 + (n + m)\Omega$ и низкие частоты $(n - m)\Omega$, за которыми только и может следить регистрирующий прибор.

Мы не будем далее интересоваться конкретным устройством демодулятора — квадратичного детектора и будем полагать, что сигнал на выходе детектора представляет собой квадрат входного сигнала, усредненный за некоторое время Δt , определяемое инерционностью регистрирующего устройства:

$$f(t) \rightarrow \boxed{D} \rightarrow g(t) = \overline{f^2(t)} = \frac{1}{\Delta t} \int_{t-\Delta t/2}^{t+\Delta t/2} f^2(t) dt. \quad (2.100)$$

Будем полагать, что время усреднения Δt удовлетворяет условию

$$\frac{2\pi}{\omega_0} \ll \Delta t \ll \tau, \quad (2.101)$$

где τ — характерное время изменения функций $a(t)$ и $f(t)$, описывающих закон модуляции сигнала $f(t)$.

Рассмотрим примеры.

1. На вход квадратичного детектора подано амплитудно-модулированное колебание $f(t) = a(t) \cos \omega_0 t$. Найдём сигнал на выходе детектора

$$g(t) = \frac{1}{\Delta t} \int_{t-\Delta t/2}^{t+\Delta t/2} a^2(t') \cos^2 \omega_0 t' dt'.$$

Поскольку $\Delta t \ll \tau$ (τ — время, в течение которого функция $a(t)$ может заметно измениться), то на интервале усреднения Δt функцию $a^2(t')$ можно считать константой, равной значению $a^2(t)$ в середине интервала усреднения: $a^2(t') \approx a^2(t)$

при $t - \Delta t/2 < t' < t + \Delta t/2$. Тогда получаем

$$\begin{aligned}
 g(t) &\approx a^2(t) \frac{1}{\Delta t} \int_{t-\Delta t/2}^{t+\Delta t/2} \cos^2 \omega_0 t' dt' = \\
 &= a^2(t) \frac{1}{\Delta t} \int_{t-\Delta t/2}^{t+\Delta t/2} \frac{1}{2} (1 + \cos 2\omega_0 t') dt' = \\
 &= \frac{1}{2} a^2(t) + \frac{1}{2} a^2(t) \frac{1}{\Delta t} \int_{t-\Delta t/2}^{t+\Delta t/2} \cos 2\omega_0 t' dt'.
 \end{aligned}$$

Последнее слагаемое пренебрежимо мало по сравнению с первым, если интервал усреднения содержит большое число периодов знакопеременной функции $\cos 2\omega_0 t'$ (т.е. при $\Delta t \gg \pi/\omega_0$). Окончательно получаем

$$g(t) \approx \frac{1}{2} a^2(t), \quad (2.102)$$

т.е. сигнал на выходе детектора содержит информацию о законе амплитудной модуляции входного сигнала.

На рис. 2.22 изображены колебания $f(t)$ и $g(t)$ на входе и выходе квадратичного детектора при $a(t) = a_0(1 + m \cos \Omega t)$.

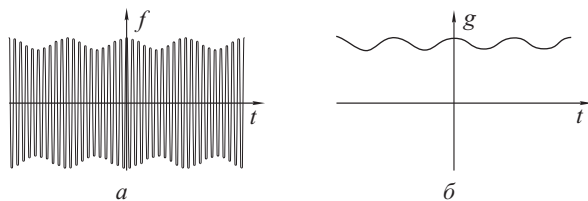


Рис. 2.22

2. Пусть на вход квадратичного детектора поступает фазово-модулированное колебание

$$f(t) = a_0 \cos(\omega_0 t + \varphi(t)).$$

Сигнал на выходе детектора

$$g(t) = \frac{1}{\Delta t} \int_{t-\Delta t/2}^{t+\Delta t/2} a_0^2 \cos^2(\omega_0 t' + \varphi(t')) dt'.$$

Поскольку на интервале усреднения функция $\varphi(t')$, описывающая закон фазовой модуляции, остается константой (равной

значению $\varphi(t)$ в середине интервала усреднения $\varphi(t') \approx \varphi(t) \approx \varphi_0$, то получаем

$$g(t) \approx \frac{1}{2} a_0^2 = \text{const},$$

т. е. закон фазовой модуляции оказывается утерянным при квадратичном детектировании.

Решить проблему можно, исказив определенным образом амплитудно-фазовые соотношения в спектре фазово-модулированного колебания (превратив таким образом колебание, модулированное по фазе в амплитудно-модулированное колебание). После чего искаженный таким образом сигнал можно подать на вход квадратичного детектора.

Вернемся в качестве примера к модулированному колебанию с законом модуляции $\varphi(t) = m \cos \Omega t$ (при $m \ll 1$). Разложение в спектр этого колебания имеет вид (см. (1.45)):

$$f(t) = a_0 \cos \omega_0 t + \frac{ma_0}{2} \cos \left[(\omega_0 + \omega)t + \frac{\pi}{2} \right] + \\ + \frac{ma_0}{2} \cos \left[(\omega_0 - \Omega)t + \frac{\pi}{2} \right].$$

Прием без несущей. Рассмотрим *прием без несущей*. После устранения несущего колебания $a_0 \cos \omega_0 t$ получаем сигнал

$$\tilde{f}(t) = \frac{ma_0}{2} \cos \left[(\omega_0 + \omega)t + \frac{\pi}{2} \right] + \frac{ma_0}{2} \cos \left[(\omega_0 - \Omega)t + \frac{\pi}{2} \right] = \\ = ma_0 \cos \Omega t \cdot \sin \omega_0 t.$$

На выходе квадратичного детектора имеем

$$g(t) = \frac{1}{\Delta t} \int_{t-\Delta t/2}^{t+\Delta t/2} \tilde{f}^2(t') dt' \approx \frac{m^2 a_0^2}{2} \cos^2 \Omega t,$$

закон фазовой модуляции (правда в искаженном виде) выявлен: изменение выходного сигнала детектора происходит с удвоенной частотой:

$$g(t) = \frac{m^2 a_0^2}{4} (1 + \cos 2\Omega t). \quad (2.103)$$

Прием с изменением фазы несущей. После изменения фазы несущего колебания на $\pi/2$:

$$a_0 \cos \omega_0 t \rightarrow a_0 \cos \left(\omega_0 t + \frac{\pi}{2} \right)$$

имеем

$$\begin{aligned}\tilde{f}(t) = a_0 \cos \left(\omega_0 t + \frac{\pi}{2} \right) + \\ + \frac{ma_0}{2} \cos \left[(\omega_0 + \Omega)t + \frac{\pi}{2} \right] + \frac{ma_0}{2} \cos \left[(\omega_0 - \Omega)t + \frac{\pi}{2} \right].\end{aligned}$$

Теперь фазовые соотношения между всеми гармониками оказались такими же, как и в амплитудно-модулированном колебании (1.43). Действительно, после простых преобразований получаем

$$\tilde{f}(t) = -a_0(1 + m \cos \Omega t) \sin \omega_0 t.$$

Мы преобразовали колебание, модулированное по фазе (с законом фазовой модуляции $\varphi(t) = m \cos \Omega t$), в амплитудно-модулированное колебание $\tilde{f}(t)$ с законом модуляции

$$a(t) = a_0(1 + m \cos \Omega t).$$

На выходе квадратичного детектора, согласно (2.102), получаем

$$g(t) \approx \frac{1}{2} a^2(t) = \frac{1}{2} a_0^2 (1 + m \cos \Omega t)^2 \approx \frac{1}{2} a_0^2 + a_0^2 m \cos \Omega t,$$

т. е. переменная составляющая сигнала на выходе детектора воспроизводит без искажения закон фазовой модуляции исходного сигнала $f(t)$.

Задачи

1. Вывести закон сохранения энергии для упругого маятника. Найти зависимость от времени кинетической и потенциальной энергии.
2. Выразить добротность колебательного контура с малым затуханием через его параметры L , C и R .
3. Решить аналогичную задачу для пружинного маятника с параметрами m , k и α .
4. Покажите, что добротность осциллятора выражается через параметры ω_0 и δ с помощью равенства $Q = \omega_0/2\delta$.
5. Определите добротность колебательной системы, осциллограмма колебания которой показана на рис. 1.15 а.
6. Найти частотные характеристики фильтров, изображенных на рис. 2.18.

Глава 3

ПАРАМЕТРИЧЕСКИЕ КОЛЕБАНИЯ. АНГАРМОНИЧЕСКИЙ ОСЦИЛЛЯТОР. АВТОКОЛЕБАНИЯ

3.1. Параметрические колебания

Выше мы рассмотрели вынужденные колебания осциллятора, возбуждаемые меняющейся во времени внешней силой (внешней ЭДС, включенной в контур L , C , R , либо силой, приложенной к маятнику). Внешнее воздействие на осциллятор может иметь, однако, совершенно иной характер и сводиться к периодическому изменению со временем параметров самой колебательной системы (например, к изменению индуктивности колебательного контура либо длины маятника). Возбужденные таким образом колебания называются *параметрическими*, а сам механизм возбуждения — *параметрическим резонансом*.

Прежде всего, попытаемся ответить на вопрос: можно ли раскачать уже имеющиеся в системе малые колебания, периодически изменяя определенным образом какой-либо ее параметр. В качестве примера рассмотрим колебательный контур, в котором происходят малые свободные колебания. Изменения тока в его цепи происходят по гармоническому закону и показаны на рис. 3.1. Пусть внешнее воздействие сводится к изменению в определенные моменты времени индуктивности контура.

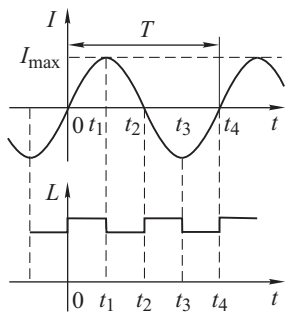


Рис. 3.1

Закон, по которому происходит изменение индуктивности, показан на рис. 3.1: в моменты, когда ток в катушке индуктивности принимает максимальное значение I_{\max} (моменты t_1 и t_3), индуктивность скачком уменьшается на малую величину ΔL , а в моменты, когда ток равен нулю, индуктивность скачком восстанавливается до прежнего значения L (моменты t_2 и t_4). Изменения ΔL будем полагать малыми ($\Delta L \ll L$).

Заметим, что при скачкообразном изменении индуктивности (время τ , в течение которого происходит изменение L , мало

в сравнении с периодом свободных колебаний) магнитный поток в катушке сохраняется. В противном случае, если бы поток Φ изменялся на конечную величину $\Delta\Phi$ за очень малое время τ , в катушке возникала бы ЭДС индукции $\mathcal{E}_{\text{инд}} = -\Delta\Phi/\tau \rightarrow \infty$, что приводило бы к бесконечно большим токам в контуре. Учитывая это обстоятельство, легко подсчитать изменение энергии контура ΔW (энергии катушки индуктивности) при скачкообразных изменениях индуктивности:

$$\Delta W = \Delta\left(\frac{LI^2}{2}\right) = \Delta\left(\frac{\Phi^2}{2L}\right) = -\left(\frac{\Phi^2}{2L^2}\right)\Delta L,$$

где $\Phi = LI \approx \text{const}$ за время τ .

Окончательное выражение для ΔW запишем в виде

$$\Delta W = -\frac{I^2}{2} \Delta L. \quad (3.1)$$

Скачкообразное уменьшение индуктивности ($\Delta L < 0$) происходит в моменты времени, когда $I = I_{\text{max}}$. При этом в контур вносится энергия ($\Delta W > 0$), равная

$$\Delta W = \frac{I_{\text{max}}^2}{2} |\Delta L|. \quad (3.2)$$

Эта энергия черпается из внешнего источника: быстро уменьшить индуктивность можно, например, быстро увеличив длину катушки — растягивая ее, либо быстро извлекая из катушки железный сердечник. В любом случае для этого следует совершить работу: витки катушки притягиваются между собой амперовой силой, железный сердечник втягивается внутрь катушки. Итак, уменьшая индуктивность, мы совершаем работу, которая и идет на увеличение энергии контура, т. е. на раскачку колебаний. Скачкообразное увеличение индуктивности мы производим в моменты времени, когда ток в катушке равен нулю (моменты t_2 и t_4 на графике). При этом, согласно (3.1), $\Delta W = 0$; энергия не меняется, поскольку изменение индуктивности в этом случае происходит без совершения работы.

Поскольку подкачку энергии мы производим дважды за период, то полное увеличение энергии за период колебания равно

$$\Delta W = I_{\text{max}}^2 \Delta L. \quad (3.3)$$

С другой стороны, в процессе колебаний энергия расходуется на джоулево тепло. Джоулевы потери за период

$$W_{\text{т}} = \overline{I^2} RT,$$

где $\overline{I^2}$ — среднее значение квадрата тока за период, которое связано с максимальным значением I_{max} известным соотношением $\overline{I^2} = I_{\text{max}}^2/2$ (последнее справедливо, если контур обладает

малым затуханием). Итак,

$$W_T = \frac{I_{\max}^2 R}{2} T. \quad (3.4)$$

Условие *параметрической раскачки колебаний* — превышение подводимой энергии над потерями $\Delta W \geq W_T$. С учетом (3.3) и (3.4) получаем $\Delta L > RT/2$.

Последнее неравенство можно записать в виде

$$\frac{\Delta L}{L} > \frac{R}{2L} T = \delta T = \frac{\pi}{Q}. \quad (3.5)$$

В последнем равенстве Q — добротность колебательной системы. Если подкачку энергии осуществлять один раз за период, то для раскачки колебаний нужны вдвое большие относительные изменения индуктивности $\Delta L/L$.

Аналогичным образом можно рассмотреть параметрическую раскачку колебаний маятника. Раскачиваясь на качелях (сгибая и выпрямляя ноги), мы по существу изменяем длину маятника.

Подумайте, каковы оптимальные моменты времени, в которые следует изменять длину маятника. Считая добротность маятника заданной, выведите условие параметрической раскачки его колебаний, аналогичное условию (3.5).

При выполнении неравенства (3.5) подводимая к колебательному контуру энергия превышает потери — амплитуда колебаний постепенно нарастает, т. е. имеет место параметрическое усиление колебаний. Естественно возникает вопрос: почему амплитуда не нарастает до бесконечности? Дело в том, что рано или поздно нарушается одно из важнейших условий, которые мы приняли при выводе уравнения, описывающего процесс колебаний, а именно, независимость параметров системы от амплитуды колебаний. В частности, омическое сопротивление контура R мы полагали неизменным, не зависящим от величины протекающего тока. При достаточно больших токах это условие нарушается: сопротивление R нарастает (резистор нагревается, растут джоулевы потери), поскольку растет температура.

Итак, если R становится зависящим от величины протекающего тока $I = I(R)$, уравнение, описывающее процесс колебаний, становится нелинейным.

Рост сопротивления при больших токах приводит к нарушению неравенства (3.5). Рано или поздно наступает стационарный режим колебаний, при котором подводимая энергия в точности компенсируется джоулевыми потерями. Фазовая траектория при установившемся, стационарном режиме колебаний показана на рис. 3.2: ток в катушке индуктивности \dot{q} дважды за период скачкообразно нарастает

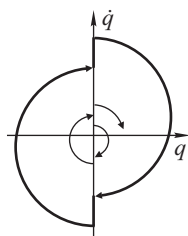


Рис. 3.2

на стационарном режиме колебаний показана на рис. 3.2: ток в катушке индуктивности \dot{q} дважды за период скачкообразно нарастает

в моменты скачкообразного уменьшения индуктивности (поток $\Phi = LI$ сохраняется!). Сделав полный оборот по фазовой траектории, изображающая точка попадает в ту же начальную точку на плоскости $q \dot{q}$ (т. е. значения q и \dot{q} через период колебаний восстанавливаются).

Мы рассматривали вопрос параметрической раскачки уже имеющихся малых колебаний. Но почему вообще возникают колебания? Ведь вначале маятник или колебательный контур находятся в покое. Откуда берется эта «начальная затравка»? Здесь принципиальную роль играют флуктуации. В частности, тепловое хаотическое движение зарядов в колебательном контуре приводит к хаотическим малым колебаниям, которые всегда существуют в контуре — это и есть та «затравка», которая далее параметрически раскачивается до наступления стационарного режима.

Отметим в заключение, что параметрические методы используются не только в механических или электрических системах. В настоящее время подобные методы используются даже для генерации световых колебаний (параметрические лазеры).

3.2. Анггармонический осциллятор

До сих пор мы изучали поведение колебательных систем, совершающих малые колебания; мы рассматривали малые отклонения системы от положения равновесия, когда «силу», возвращающую систему к положению равновесия, можно приближенно считать пропорциональной величине смещения. При этом уравнение, описывающее колебательную систему, является линейным, колебание — гармоническим, а период колебаний не зависит от амплитуды (это свойство гармонического осциллятора называют *изохронностью*).

В случае не малого отклонения от положения равновесия уравнение, описывающее процесс колебаний, оказывается в общем случае нелинейным, колебательный процесс уже не является гармоническим, а система теряет свойство изохронности: период колебаний оказывается зависящим от амплитуды (т. е. от величины начального отклонения и (или) начальной скорости).

Рассмотрим в качестве примера математический маятник. Если угол отклонения маятника от положения равновесия не полагать малым, то уравнение, описывающее процесс колебания, нелинейно (см. § 2.1):

$$\ddot{\alpha} + \omega_0^2 \sin \alpha = 0. \quad (3.6)$$

Исследование нелинейных систем (и, в частности, системы, описываемой уравнением (3.6)) представляет собой чрезвычайно

сложную математическую проблему, которая далеко выходит за рамки этой книги. Мы рассмотрим, не прибегая к сложным математическим методам, лишь одну проблему: зависимость периода колебаний от амплитуды. При этом мы ограничимся приближением

$$\sin \alpha \approx \alpha - \frac{\alpha^3}{6}, \quad (3.7)$$

учтя лишь первую поправку к выражению $\sin \alpha \approx \alpha$ (которое приводит к уравнению гармонического осциллятора и независимости периода колебаний от амплитуды).

Понятно, что учет слагаемого $-\alpha^3/6$ приводит к уменьшению величины возвращающей силы: $f(\alpha) = mg(\alpha - \alpha^3/6)$ по сравнению с линейной зависимостью $f(\alpha) = mg\alpha$; следовательно к уменьшению ускорения и скорости, а значит — к росту периода колебаний при увеличении амплитуды.

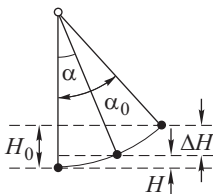


Рис. 3.3

Пусть максимальное отклонение маятника от положения равновесия (амплитуда колебаний) равна α_0 . Из закона сохранения энергии

$$\frac{mv^2}{2} = mg\Delta H$$

(см. рис. 3.3) находим скорость маятника в момент, когда его отклонение равно α :

$$v = \sqrt{2g\Delta H}, \text{ где } \Delta H = H_0 - H = 2l \left(\sin^2 \frac{\alpha_0}{2} - \sin^2 \frac{\alpha}{2} \right). \quad (3.8)$$

Если за время dt угол поворота маятника равен $d\alpha$, а длина пройденного за это время пути $dx = l d\alpha$, то $dt = dx/v = l d\alpha/v$ или, используя (3.8),

$$dt = \frac{1}{2} \sqrt{\frac{l}{g}} \frac{d\alpha}{\sqrt{\sin^2(\alpha_0/2) - \sin^2(\alpha/2)}}. \quad (3.9)$$

Интегрирование этого выражения в пределах от 0 до α_0 дает четверть периода колебаний $T/4$, поэтому

$$T = 2 \sqrt{\frac{l}{g}} \int_0^{\alpha_0} \frac{d\alpha}{\sqrt{\sin^2(\alpha_0/2) - \sin^2(\alpha/2)}}.$$

При использовании линейного приближения $\sin \alpha \approx \alpha$ мы получили бы

$$T_0 = 4 \sqrt{\frac{l}{g}} \int_0^1 \frac{dy}{\sqrt{1-y^2}},$$

где $y = \alpha/\alpha_0$. Нет надобности вычислять этот интеграл: результат известен — ведь линейное приближение приводит к уравнению гармонического осциллятора, период колебаний которого $T_0 = 2\pi\sqrt{l/g}$, поэтому

$$\int_0^1 \frac{dy}{\sqrt{1-y^2}} = \frac{\pi}{2}. \quad (3.10)$$

Не приводя подробных (хотя и очевидных) преобразований, запишем выражение для периода T , которое получается при использовании приближения (3.7) (т. е. первой нелинейной поправки)

$$T = 4\sqrt{\frac{l}{g}} \int_0^1 \frac{dy}{\sqrt{1-y^2}} + \frac{1}{6}\sqrt{\frac{l}{g}} \alpha_0^2 \int_0^1 \frac{1+y^2}{\sqrt{1-y^2}} dy.$$

Первое слагаемое — это период колебаний гармонического осциллятора (т. е. период малых колебаний маятника). Второе слагаемое — поправка, приводящая к увеличению периода, если амплитуда не слишком мала.

Относительная величина поправки $\Delta T/T_0$ равна

$$\frac{\Delta T}{T_0} \approx \frac{\alpha_0^2}{24} \int_0^1 \frac{1+y^2}{\sqrt{1-y^2}} dy. \quad (3.11)$$

Важно подчеркнуть, что поправка пропорциональна квадрату амплитуды колебаний.

Мы не будем вычислять интеграл, входящий в оценку (3.11). Очевидно, что подынтегральная функция в (3.11) ни при каком y не превышает подынтегральную функцию в (3.10) более чем вдвое (на интервале от нуля до единицы), поэтому и величина интеграла менее чем вдвое отличается от $\pi/2$, т. е. $\Delta T/T_0 < \alpha_0^2(\pi/24)$. Более точные вычисления приводят к оценке $\Delta T/T_0 \approx (1/16)\alpha_0^2$.

Подобная ситуация возникает и в пружинном маятнике, если жесткость пружины уменьшается при увеличении деформации. График на рис. 3.4 соответствует именно этому случаю: жесткость пружины $k = |df/dx|$ уменьшается с ростом x (такую характеристику принято называть *мягкой*). В слу-

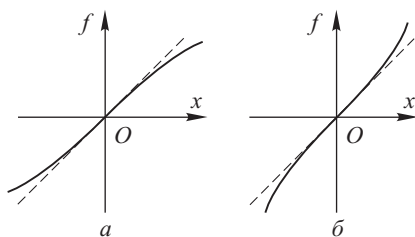


Рис. 3.4

чае жесткой характеристики (рис. 3.4 б) $k = df/dx$ растет при увеличении деформации — в этом случае рост амплитуды приводит к уменьшению периода колебаний.

Колебательный контур, в катушку индуктивности которого вставлен железный сердечник, также представляет собой ангармонический осциллятор: при больших амплитудах (большом токе через катушку) индуктивность L начинает зависеть от величины тока, при этом магнитный поток через катушку $\Phi = L(I)I$ уже не является линейной функцией величины тока. Отклонение от линейности обычно описывается функцией

$$I = L_0\Phi - L_1\Phi^3, \quad (3.12)$$

где константы L_0 и L_1 определяются параметрами контура. Зависимость тока в контуре от величины потока совершенно подобна в этом случае зависимости возвращающей силы от угла отклонения математического маятника $f = mg\alpha - (mg/6)\alpha^3$, поэтому период колебаний увеличивается с ростом амплитуды.

3.3. Автоколебания

Существуют системы, в которых стационарный колебательный процесс возникает несмотря на то, что нет периодической внешней силы, заставляющей систему колебаться (с частотой, навязанной внешней силой — вынужденные колебания) и нет периодического изменения параметров системы (параметрические колебания). В такой системе имеется постоянный, не меняющийся во времени источник энергии (например, постоянная ЭДС в электрической системе, либо постоянное поле тяжести, в котором находится маятник), однако в самой системе имеется механизм, обеспечивающий периодический характер поступления энергии от источника к элементу, совершающему колебания. Колебания в таких системах называются *автоколебаниями*. Важнейшим признаком автоколебаний является независимость установившегося стационарного режима колебаний (амплитуды) от начальных условий. Это отличает автоколебания от свободных незатухающих колебаний осциллятора.

Типичным примером автоколебательной системы являются маятниковые часы с гирей — так называемые «ходики». В этом случае убыль энергии (затухание колебаний маятника из-за трения) компенсируется за счет потенциальной энергии гири в поле тяжести, причем поступление энергии происходит толчками — отдельными порциями ΔE — один раз за каждый период колебаний. Порция энергии ΔE определяется опусканием гири на одно звено цепи, на которой эта гиря подвешена. Механизм, позволяющий гире опускаться в определенные моменты времени,

согласован с колебаниями маятника системой шестеренок и так называемым анкерным механизмом. Оказывается, что именно величина ΔE определяет амплитуду установившихся колебаний вне зависимости от начального отклонения. Пусть, например, передача энергии маятнику (толчок, возникающий при опускании гири на одно звено) происходит в момент наибольших отклонений от положения равновесия. Если непосредственно перед очередным, n -м толчком, энергия маятника (в данном случае потенциальная энергия отклоненного от положения равновесия маятника) равна $E_n^{(-)}$, то после толчка его энергия $E_n^{(+)}$ есть

$$E_n^{(+)} = E_n^{(-)} + \Delta E.$$

Совершив одно полное колебание, маятник потеряет часть энергии из-за трения. Полагая, что сила трения пропорциональна скорости (вязкое трение), получим, что перед очередным, $(n + 1)$ -м толчком, энергия маятника станет равной

$$E_{n+1}^{(-)} = E_n^{(+)} e^{-2\delta T}, \quad (3.13)$$

а после $(n + 1)$ -го толчка получим $E_{n+1}^{(+)} = E_{n+1}^{(-)} + \Delta E$, или с учетом (3.13):

$$E_{n+1}^{(+)} = E_n^{(+)} e^{-2\delta T} + \Delta E. \quad (3.14)$$

Стационарный режим соответствует условию $E_{n+1}^{(+)} = E_n^{(+)} \equiv E$, откуда находим

$$E = \frac{\Delta E}{1 - e^{-2\delta T}}. \quad (3.15)$$

При малом затухании δ , таком, что $2\delta T \ll 1$, получим $e^{-2\delta T} \approx 1 - 2\delta T$ и тогда

$$E = \frac{\Delta E}{2\delta T} = \frac{1}{2\pi} Q \Delta E \quad \text{или} \quad Q = \frac{2\pi E}{\Delta E} \quad (3.16)$$

(Q — добротность).

Отметим, что полученное значение энергии установившихся колебаний E не зависит от начальной энергии и начального отклонения от положения равновесия. Чтобы наглядно показать, как происходит «забывание» начальных условий, сделаем следующее построение. Выберем прямоугольную систему координат, откладывая по оси абсцисс (ось Ox) величину $x = E_n$, а по оси ординат (ось Oy) величину $y = E_{n+1}$ (рис. 3.5).

В этой системе координат уравнение (3.14) есть уравнение прямой

$$y = e^{-2\delta T} x + \Delta E. \quad (3.17)$$

Кроме этой прямой, проведем на графике биссектрису, уравнение которой

$$y = x. \quad (3.18)$$

Очевидно, точка пересечения этих прямых определяет стационарный режим автоколебаний $x = y = \Delta E / (1 - e^{-2\delta T})$, что ясно из сравнения последнего равенства с (3.15).

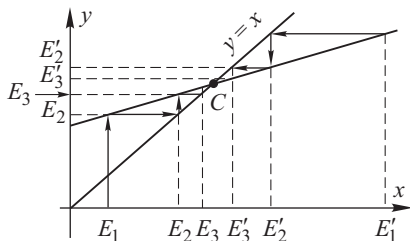


Рис. 3.5

Пусть начальная энергия маятника, соответствующая начальному отклонению, есть E_1 (отмечено на оси Ox , рис. 3.5).

Маятник совершает полное колебание и после первого толчка его энергия E_2 найдется с помощью (3.14)

$$E_2 = E_1 e^{-2\delta T} + \Delta E.$$

Значение E_2 найдем с помощью прямой (3.17) на графике. Для нахождения графически следующей точки последовательности E_3 ,

$$E_3 = E_2 e^{-2\delta T} + \Delta E,$$

превратим «ординату» E_2 в абсциссу E_2 с помощью биссектрисы $y = x$. Этой новой абсциссе $x = E_2$ соответствует ордината $y = E_3$, которая находится с помощью прямой (3.17).

Повторяя далее многократно всю процедуру, получаем ломаную кривую, заключенную между прямыми (3.17) и (3.18), которая неограниченно приближается к точке C их пересечения — стационарному режиму.

Очевидно, какую бы начальную точку мы ни взяли (например, E_1' на графике) мы неизменно получаем ломаную, приближающуюся к одной и той же точке C — стационарному режиму, что и означает забывание начальных условий в процессе установления автоколебаний.

Построенная нами в соответствии с описанной процедурой ломаная кривая называется *лестницей Ламерея*.

3.4. Автогенератор Ван-дер-Поля

Рассмотрим теперь возникновение автоколебаний в генераторе, построенном на основе электронной лампы — триода. Такой генератор называется *автогенератором Ван-дер-Поля*. Его схема показана на рис. 3.6.

Собственно колебательным элементом генератора является колебательный контур L, C, R , находящийся в сеточной цепи

триода. Напряжение $U_C = q/C$ на конденсаторе контура, подается на сетку лампы. Изменение этого напряжения управляет изменениями анодного тока I_a (в анодной цепи лампы) и следовательно, изменением магнитного потока в катушке L_a , находящейся в анодной цепи. Катушки L_a и L (катушка колебательного контура) индуктивно связаны между собой (т. е. магнитный поток катушки L_a пронизывает катушку L и его изменения создают в последней ЭДС взаимной индукции $\mathcal{E} = -L_{12}(dI_a/dt)$, где L_{12} — взаимная индуктивность катушек L и L_a .

С помощью этого осуществляется механизм *обратной связи*, т. е. обратное влияние анодного тока лампы на процесс колебаний контура.

Рассмотрим качественно процесс раскочки малых колебаний, которые всегда имеют место в контуре из-за теплового хаотического движения зарядов. Пусть в данный момент времени ток I в контуре направлен так, как показано на рис. 3.6 (положительное направление обхода по контуру) — при этом растет заряд q на пластинах конденсатора, $I = dq/dt > 0$; растет, следовательно, напряжение на конденсаторе U_C .

Подробнее характеристики лампы мы обсудим позднее, а пока отметим лишь, что при малых напряжениях на сетке U_C рост напряжения U_C приводит к росту анодного тока I_a . Последнее приводит к росту магнитного потока Φ_a , созданного катушкой L_a : $\Phi_a = L_a I_a$. На рисунке показано направление анодного тока, а также направление, в котором этот ток обтекает витки катушки L_a . Соответствующий магнитный поток Φ_a направлен вверх (по правилу буравчика). Собственный же магнитный поток Φ катушки L (по тому же правилу) направлен вниз. Нарастающий поток Φ_a , пронизывая витки катушки L контура, создает в ней ЭДС.

Каково ее направление, т. е. в каком направлении возникшая ЭДС индукции вызывает движение зарядов в контуре?

Согласно правилу Ленца возникающий индукционный ток должен создать собственный магнитный поток, направление которого должно быть таким, чтобы скомпенсировать изменение потока Φ_a (который направлен вверх и нарастает). Ясно поэтому, что собственный магнитный поток тока I в контуре (направленный вниз!) также должен нарастать, а это значит, что ток I в контуре должен увеличиться.

Заметим попутно, что ЭДС взаимной индукции $-L_{12}(dI_a/dt)$ оказывается при этом положительной величиной: увеличивая ток I , она гонит заряды в положительном направлении обхода контура, откуда следует (поскольку $dI_a/dt > 0$), что $L_{12} < 0$.

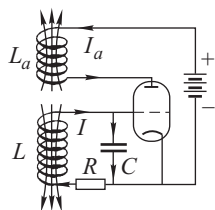


Рис. 3.6

Итак, обратная связь (влияние анодного тока на колебательный контур) приводит к росту тока в контуре, к дополнительному росту заряда конденсатора и напряжению U_c , на сетке лампы. Это в свою очередь приведет к росту анодного тока и т. д.

Таким образом, имеющиеся в контуре малые колебания начинают раскачиваться, их амплитуда постепенно растет. Если

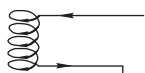


Рис. 3.7

изменить намотку катушки L_a на противоположную (как показано на катушке, изображенной на рис. 3.7), то рост Φ_a приводит к появлению ЭДС в катушке L , которая приводит не к росту, а к уменьшению тока в контуре и, следовательно, к быстрому подавлению колебаний. Обратная связь, осуществляемая катушкой L_a , в первом случае называется положительной (при этом $L_{12} < 0$), а во втором — отрицательной ($L_{12} > 0$). Итак, при положительной обратной связи возникшие в контуре малые колебания начинают нарастать, т. е. состояние покоя, при котором ток в контуре и заряд конденсатора равны нулю, неустойчиво. В случае отрицательной обратной связи, наоборот, малые колебания подавляются, т. е. состояние покоя устойчиво.

Рассмотрим теперь более детально процесс раскачки автоколебаний. Уравнение, описывающее процесс колебаний в контуре, имеет в нашем случае вид

$$RI + U_c = -L \frac{dI}{dt} - L_{12} \frac{dI_a}{dt}, \quad (3.19)$$

по сравнению с уравнением, описывающим процесс свободных колебаний, в правой части к ЭДС индукции $-L(dI/dt)$ добавляется ЭДС $-L_{12}(dI_a/dt)$. Подчеркнем, что взаимная индуктивность L_{12} может быть как положительной, так и отрицательной. При положительной обратной связи $L_{12} < 0$.

Перепишем уравнение (3.19) в виде (имея в виду, что $dI/dt = C d^2 U_c / dt^2$):

$$\ddot{U}_c + 2\delta \dot{U}_c + \omega_0^2 U_c = -\frac{L_{12}}{LC} \frac{dI_a}{dt}, \quad (3.20)$$

где $\delta = R/(2L)$, а $\omega_0 = 1/\sqrt{LC}$.

По внешней форме последнее уравнение похоже на уравнение вынужденных колебаний (2.78), в котором правая часть играет роль внешней, вынуждающей силы. Однако принципиальное отличие нашей задачи состоит в том, что эта правая часть зависит в свою очередь от изменений искомой функции U_c , (тогда как в уравнении вынужденных колебаний правая часть — это заданная функция времени, никак не связанная (и не зависящая) от колебаний напряжения на конденсаторе контура).

Рассмотрим идеализированную зависимость анодного тока лампы I_a от напряжения на сетке U_c , изображенную на рис. 3.8.

При отрицательном напряжении на сетке лампы, таком что $U_c < U_1$, анодный ток равен нулю — лампа заперта. При достаточно малых изменениях U_c , таких, что $U_1 \leq U_c \leq U_2$ анодный ток растет пропорционально сеточному напряжению. На этом участке, как видно из графика, зависимость $I_a(U_c)$ имеет вид

$$I_a = I_0 + S_0 U_c, \quad (3.21)$$

где $S_0 = dI_a/dU_c$ — тангенс угла наклона прямой на участке $U_1 U_2$ — константа, называемая *крутизной характеристики*. Мы полагаем, что на начальной стадии, в процессе раскачки, амплитуда колебаний настолько мала, что напряжение U_c не выходит за границы, где справедлива линейная зависимость (3.21). Наконец, при $U_c > U_2$ имеет место насыщение: анодный ток I_a достигает максимального значения I_s и при дальнейшем росте U_c , уже не меняется.

Итак, на начальной стадии, пока амплитуда колебаний достаточно мала, мы имеем для правой части уравнения (3.20)

$$-\frac{L_{12}}{LC} \frac{dI_a}{dt} = -\frac{L_{12}}{LC} \frac{dI_a}{dU_c} \frac{dU_c}{dt} = -\frac{L_{12}S_0}{LC} \frac{dU_c}{dt}.$$

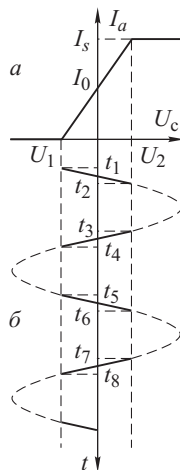


Рис. 3.8

Переносим последнее выражение в левую часть уравнения (где уже есть слагаемое, содержащее производную dU_c/dt), получаем

$$\frac{d^2 U_c}{dt^2} + 2 \left(\delta + \frac{L_{12}S_0}{2LC} \right) \frac{dU_c}{dt} + \omega_0^2 U_c = 0. \quad (3.22)$$

Введем обозначение

$$\delta^* = \frac{R}{2L} + \frac{L_{12}S_0}{2LC}, \quad (3.23)$$

после чего уравнение (3.20) приобретает вид

$$\ddot{U}_c + 2\delta^* \dot{U}_c + \omega_0^2 U_c = 0, \quad (3.24)$$

который совершенно подобен уравнению, описывающему свободные затухающие колебания гармонического осциллятора, причем роль затухания δ играет величина δ^* , определяемая формулой (3.23). Решение (3.24) хорошо известно:

$$U_c = A e^{-\delta^* t} \cos(\omega t + \varphi), \quad (3.25)$$

где $\omega = \sqrt{\omega_0^2 - \delta^{*2}}$, A и φ — произвольные постоянные, зависящие от начальных условий. Еще раз напомним, что коэффициент $\delta = R/(2L)$, определяющий изменение амплитуды колебаний

при свободных колебаниях осциллятора, существенно положительная величина, поэтому показатель экспоненты $e^{-\delta t}$ всегда отрицателен, он-то и определяет экспоненциальное затухание колебаний.

Совершенно иная ситуация в нашем случае. Как следует из (3.23), величина δ^* может быть как положительной, так и отрицательной. При $\delta^* < 0$ показатель экспоненты в (3.25) оказывается положительным и мы получаем экспоненциальный рост амплитуды колебаний $a(t) = Ae^{|\delta^*|t}$.

Необходимым условием выполнения неравенства $\delta^* < 0$ является условие

$$L_{12} < 0, \quad (3.26)$$

т. е. наличие положительной обратной связи (это мы уже выяснили при качественном обсуждении роли обратной связи). Условия (3.26), однако, недостаточно. Как видно из (3.23) для достижения отрицательности δ^* необходимо также, чтобы выполнялось неравенство

$$\frac{R}{L} < \frac{|L_{12}|S_0}{LC} \quad \text{или} \quad R < \frac{|L_{12}|S_0}{C}. \quad (3.27)$$

Смысл последнего условия ясен: величина сопротивления контура R определяет джоулевы потери энергии, а величина обратной связи $|L_{12}|$ — ту подкачку энергии в контур, которую обеспечивает ЭДС взаимной индукции. Для раскачки колебаний необходимо превышение подводимой энергии над потерями.

Наконец, при $L_{12} > 0$, как видно из (3.23), величина δ^* всегда положительна, поэтому амплитуда $a(t) = Ae^{-\delta^*t}$ всегда затухает.

Надо, однако, помнить, что все сказанное справедливо лишь до тех пор, пока напряжение на конденсаторе не выходит за границы малой области. Формула (3.25) неприменима на более поздних стадиях процесса, когда амплитуда становится достаточно большой.

Действительно, если выполнены условия (3.26) и (3.27) и величина δ^* оказывается отрицательной, то, согласно (3.25), амплитуда колебаний должна неограниченно экспоненциально нарастать; при этом невозможно ответить на вопрос: какова установившаяся амплитуда автоколебаний? Кроме того, из формулы (3.25) следует, что при $\delta^* = 0$ имеют место незатухающие колебания $A \cos(\omega t + \varphi)$, амплитуда A и начальная фаза φ которых зависят от начальных условий. Ни то, ни другое в действительности не имеет места. Причина несоответствия состоит в том, что даже при использовании идеализированной характеристики лампы $I_a(U_c)$, изображенной на рис. 3.8, крутизна S не является в действительности константой (равной тангенсу угла

наклона прямой на линейном участке). Правильная зависимость $I_a(U_c)$ должна описывать, что происходит, когда U_c становится достаточно большой отрицательной величиной ($U_c < U_1$) (и при этом $I_a = 0$) и что происходит при достаточно больших положительных U_c ($U_c > U_2$) и при этом $I_a = I_s$ (насыщение).

Итак, с учетом нелинейности характеристики лампы при больших амплитудах колебаний, правая часть уравнения (3.20) — величина $-(L_{12}/LC)(dI_a/dt)$ должна быть записана в виде

$$-\frac{L_{12}}{LC} \frac{dI_a}{dt} = -\frac{L_{12}S(U_c)}{LC} \dot{U}_c,$$

где $S = S(U_c)$ — крутизна, зависящая от напряжения на конденсаторе — функция показанная на рис. 3.9.

Соответственно, в уравнении колебаний (3.24) величина δ^* становится функцией напряжения на конденсаторе

$$\delta^* = \frac{R}{2L} + \frac{L_{12}S(U_c)}{2LC}, \quad (3.28)$$

т.е. уравнение (3.24) становится нелинейным.

Только учитывая нелинейность, можно исследовать установившийся процесс автоколебаний.

Обратите внимание на формулу (3.28): условие $\delta^* < 0$ нарушается, как только величина U_c выходит из области $U_1 < U_c < U_2$; при этом $S(U_c) = 0$ и $\delta^* = R/2L$ становится существенно положительной величиной.

Итак, рассмотрим ситуацию, когда амплитуда колебаний стала достаточно большой, так что значение U_c при колебаниях выходят за рамки рассмотренной области (рис. 3.8). Начнем с момента времени t_1 , когда $U_c = 0$ и начинает расти. Пропорционально, в соответствии с законом (3.21), изображенным на рис. 3.8 растет анодный ток вплоть до момента времени t_2 , когда наступает насыщение $I_a = I_s$. Далее (см. рис. 3.8б) значение U_c продолжает при колебаниях нарастать, а затем уменьшается, оставаясь, вплоть до момента времени t_3 в области насыщения.

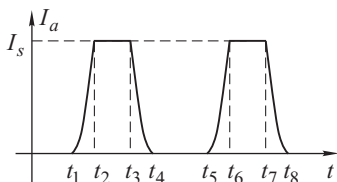


Рис. 3.10

значение U_c уменьшается, оставаясь на этом интервале внутри линейного коридора, поэтому и анодный ток I_a пропорционально

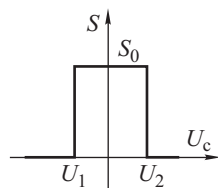


Рис. 3.9

При таких изменениях U_c в интервале времени от t_2 до t_3 анодный ток остается неизменным и равным значению I_s . Это горизонтальный участок зависимости $I_a(t)$ на рис. 3.10. Далее, от момента времени t_3 и до момента времени t_4 значение

падает (падающий участок $I_a(t)$ на рисунке, и к моменту времени t_4 , когда напряжение на конденсаторе U_C становится равным запирающему напряжению U_1 , анодный ток становится равным нулю.

Аналогичным образом с помощью рисунков 3.8 и 3.10 можно проследить процесс и далее.

Итак, мы видим, к чему приводит нелинейность характеристики лампы: при большом размахе колебаний напряжения на конденсаторе U_C (которые происходят, как показано на рисунке, по гармоническому закону) изменение анодного тока со временем имеет существенно другую форму: анодный ток представляет собой последовательность почти прямоугольных импульсов с крутыми участками роста и спада (на отрезках времени $t_1 t_2$ и $t_3 t_4$) и горизонтальным участком постоянных значений I_a (отрезок времени $t_2 t_3$, $I_a = I_s$ — ток насыщения и отрезок $t_4 t_5$ — нулевое значение I_a).

График зависимости $-(L_{12}/LC)\dot{I}_a$ от t (где L_{12} , L и C — константы) приведен на рис. 3.11: при $L_{12} < 0$ (положительная обратная связь, рис. 3.11 а) и при $L_{12} > 0$ (отрицательная обратная связь, рис. 3.11 б).

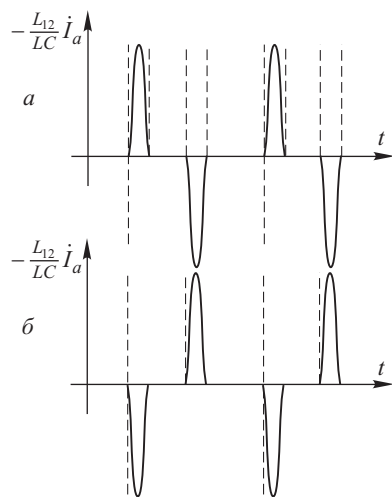


Рис. 3.11

Мы видим, что «внешняя ЭДС», раскачивающая в контуре колебания, имеет вид коротких толчков разного знака.

При положительной обратной связи ($L_{12} < 0$, рис. 3.11 а) толчки, как мы выяснили ранее, действуют в «правильном направлении» и раскачивают колебания. При отрицательной обратной связи ($L_{12} > 0$, рис. 3.11 б), наоборот, колебания затухают еще быстрее, чем без обратной связи.

Найдем установившуюся амплитуду колебаний тока. Толчки происходят, как видно из графика $U_C(t)$, в моменты, когда напряжение на конденсаторе равно нулю и, следовательно, ток $I = C\dot{U}_C$ максимален, т. е. равен амплитудному значению I . После толчка, создающего скачок тока ΔI , получим $I_1 = I + \Delta I$. За короткое время толчка $\tau = t_2 - t_1 = t_4 - t_3 = \dots$ и т. д. суммарный поток, пронизывающий катушку L , сохраняется

$$\Delta\Phi = L\Delta I + L_{12}\Delta I_a = 0,$$

откуда $|\Delta I| = \Delta I_a L_{12}/L$. За время действия толчка анодный ток, как видно из графика $I_a(t)$ на рис. 3.10 изменяется от 0 до максимальной величины I_s , т. е. $\Delta I_a = I_s$ и мы получаем

$$|\Delta I| = L_{12} \frac{I_s}{L}, \quad I_1 = I + \frac{|L_{12}| I_s}{L}.$$

Далее, до очередного толчка, контур свободен от внешнего воздействия и, в силу экспоненциального затухания, амплитуда уменьшается, ее значение перед очередным толчком равно

$$I_2 = I_1 e^{-\delta T/2}$$

($T/2$ — время между толчками).

Условие стационарности: значение амплитуды перед каждым очередным толчком (а значит и после каждого очередного толчка) не меняется, т. е. $I_2 = I_1 = I$

$$I = \left(I + \frac{L_{12} I_s}{L} \right) e^{-\delta T/2},$$

откуда находим стационарное, установившееся значение амплитуды автоколебаний

$$I = \frac{L_{12} I_s / L}{1 - e^{-\delta T/2}} e^{-\delta T/2}.$$

Если контур обладает малым затуханием (добротность велика), то скачки тока ΔI оказываются малыми по сравнению с установившимся значением I . Используя $e^{-\delta T/2} \approx 1 - \delta T/2$, находим $I = 2L_{12} I_s / (L \delta T)$.

Заметим в заключение, что все основные свойства автоколебательной системы, рассмотренные нами на примере лампового генератора, сохраняются, если использовать вместо лампы полупроводниковый триод.

Глава 4

КИНЕМАТИКА ВОЛН

Понятие волны — одно из важнейших в физике. Современная физика изучает и волны на море, и сейсмические волны в Земле, и волны звука в твердых телах, жидкостях и газах, и электромагнитные волны (в том числе свет) и, наконец, волны вероятности, которые используются в квантовой физике для описания поведения электронов, атомов и других микрообъектов. Важно подчеркнуть, что наука о волнах использует единый математический аппарат для описания волн любой физической природы — в этом и проявляется общность волновых закономерностей.

4.1. Основные понятия и определения. Простейшие типы волн. Волновое уравнение

Что такое волна? Очень трудно дать исчерпывающее определение этого понятия, не обращаясь к математическим формулам, ибо любое определение не может охватить всего многообразия процессов, которые принято называть *волнами*. Для первоначального знакомства с волнами достаточно понимать их как процесс распространения в пространстве колебаний, возникающих в результате некоторого возмущения. Таким возмущением может быть и изменение давления в каком-либо месте упругой среды, и изменение электрического или магнитного поля в некоторой точке пространства, и изменение уровня поверхности жидкости в каком-либо месте (например, от брошенного на гладкую поверхность воды камня). Во всех этих случаях возмущение начинает с конечной скоростью распространяться в пространстве.

Волновой процесс развивается и во времени, и в пространстве, поэтому в отличие от колебательного процесса, который описывается функцией времени $S(t)$, волновой процесс должен описываться функцией координат и времени $S(x, y, z, t)$ (или, коротко, $S(\mathbf{r}, t)$). На первом этапе мы не будем интересоваться физической природой волны, т. е. неважно, понимаем ли мы под S изменение давления или плотности среды при распространении в ней звуковой волны, либо S — это напряженность электрического или магнитного поля при распространении электромагнитной волны, либо это волна какой-либо иной природы.

Начнем математическое описание с наиболее простой ситуации — плоских волновых возмущений: будем считать, что волновое возмущение S зависит только от одной координаты (скажем от координаты z). Пусть в какой-то момент времени t мы определили распределение возмущения в пространстве, т. е. определили S как функцию координаты z (рис. 4.1 а). (Мы договорились, что S зависит только от z , т. е. если рассмотреть любую плоскость, перпендикулярную оси z (плоскость $z = \text{const}$), то в любой точке этой плоскости величина S одна и та же, независимо от координат x, y этой точки). Можно сказать, что картинка на рис. 4.1 а — это мгновенная фотография волнового возмущения, сделанная в момент времени t .

Будем следить за развитием ситуации в области $z > 0$. Сделаем мгновенную фотографию в другой момент времени $t + \Delta t$. Что мы должны на ней увидеть? Интуиция подсказывает нам, что мы должны увидеть картину, изображенную на рис. 4.1 б: возмущение переместилось в пространстве за время Δt на расстояние $\Delta z = v\Delta t$, где v — скорость распространения возмущения. Предположим, что пространственная форма возмущения не изменилась, т. е. вторая фотография изображает тот же «волновой пакет», только сдвинутый в пространстве на расстояние Δz :

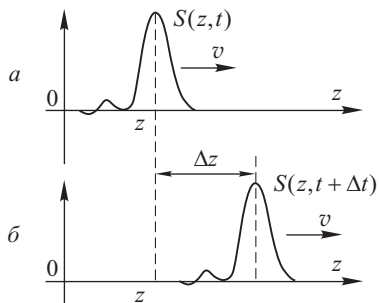


Рис. 4.1

$$S(z, t) = S(z + \Delta z, t + \Delta t). \quad (4.1)$$

Равенство (4.1) означает, что изменения во времени возмущения S (колебание величины S) в плоскости $z + \Delta z$ происходят точно так же, как и в плоскости z , лишь запаздывая по времени на Δt . Какова должна быть зависимость S от координаты z и времени t , чтобы равенство (4.1) было справедливо? Легко догадаться, что она должна иметь вид

$$S(z, t) = S(z - vt), \quad (4.2)$$

т. е. функция S должна зависеть от комбинации аргументов z и t вида: $z - vt$. Действительно в этом случае равенство (4.1) выполняется, поскольку если заменить z на $z + \Delta z$, а t на $t + \Delta t$, то величина $z - vt$ не изменится:

$$z - vt = (z + \Delta z) - v(t + \Delta t)$$

в силу того, что $\Delta z = v\Delta t$.

Итак, зависимость вида (4.2) соответствует нашему интуитивному представлению о волне, как о возмущении, которое распространяется в пространстве с конечной скоростью. В нашем случае $\Delta z = v \Delta t > 0$, т. е. волна бежит вправо, в положительном направлении оси z . Аналогичным образом можно показать, что функция $S(z + vt)$ описывает волну, бегущую справа налево.

Обозначая аргумент функции S через ξ ($\xi = z \pm vt$), найдем связь между производными S по координате и времени:

$$\begin{aligned} \frac{\partial S}{\partial t} &= \frac{dS}{d\xi} \frac{\partial \xi}{\partial t} = \pm v \frac{dS}{d\xi}; & \frac{\partial S}{\partial z} &= \frac{dS}{d\xi} \frac{\partial \xi}{\partial z} = \frac{dS}{d\xi}, \\ \frac{\partial^2 S}{\partial t^2} &= \pm \frac{\partial}{\partial t} \left(v \frac{dS}{d\xi} \right) = v^2 \frac{d^2 S}{d\xi^2}; & \frac{\partial^2 S}{\partial z^2} &= \frac{\partial}{\partial z} \left(\frac{dS}{d\xi} \right) = \frac{d^2 S}{d\xi^2}. \end{aligned} \quad (4.3)$$

Сравнивая вторые производные функции S по t и по z (4.3), замечаем, что они отличаются множителем v^2 , и мы получаем

$$\frac{\partial^2 S}{\partial z^2} - \frac{1}{v^2} \frac{\partial^2 S}{\partial t^2} = 0. \quad (4.4)$$

Соотношение (4.4) называется *волновым уравнением*. В более общем случае, когда возмущение S зависит не только от z -координаты, но также от x - и y -координат, волновое уравнение выглядит следующим образом:

$$\frac{\partial^2 S}{\partial x^2} + \frac{\partial^2 S}{\partial y^2} + \frac{\partial^2 S}{\partial z^2} = \frac{1}{v^2} \frac{\partial^2 S}{\partial t^2}.$$

Левая часть последнего уравнения коротко записывается с помощью оператора Лапласа $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$, волновое уравнение приобретает компактный вид:

$$\nabla^2 S - \frac{1}{v^2} \frac{\partial^2 S}{\partial t^2} = 0. \quad (4.5)$$

Вообще говоря, уравнение (4.5) описывает волновые процессы более общего характера, чем только волны вида $S(z \pm vt)$. Мы можем дать теперь несколько формальное, но более общее определение понятия волны. Любая физическая величина $S(\mathbf{r}, t)$, подчиняющаяся уравнению (4.5), может быть названа *волной*.

Рассмотрим несколько важных примеров простейших типов волн.

Плоская монохроматическая волна. Пусть возмущение описывается функцией координат и времени вида

$$S(z, t) = a \cos(\omega t - k z - \varphi), \quad (4.6)$$

где ω , k и φ — константы. Смысл константы ω ясен: при фиксированном z (т. е. в любой плоскости, перпендикулярной оси z)

мы имеем гармонический колебательный процесс частоты ω и амплитуды a .

Можно убедиться непосредственной подстановкой, что функция (4.6) удовлетворяет волновому уравнению (4.4), т. е. возмущение (4.6) представляет собой волну.

Действительно, дифференцируя (4.6) дважды по t и по z , получаем

$$\frac{\partial^2 S}{\partial z^2} = \frac{1}{(\omega/k)^2} \frac{\partial^2 S}{\partial t^2}. \quad (4.7)$$

Из сравнения (4.7) и (4.4) ясно, что возмущение (4.6) действительно удовлетворяет волновому уравнению, если

$$k = \frac{\omega}{v}. \quad (4.8)$$

Константа k , определяемая последним равенством, называется *волновым числом*. Константа a в (4.6) — *амплитуда волны*, а аргумент косинуса $\omega t - kz - \varphi = \Phi$ — *фаза волны*, φ — начальная фаза (значение Φ в момент $t = 0$ в плоскости $z = 0$). Очевидно, в любой фиксированной плоскости $z = \text{const}$ (в любой точке $P(x, y)$ этой плоскости) фаза Φ в любой момент времени одна и та же, т. е. колебания величины S происходят синхронно (обычно говорят *синфазно*). Скажем, в точках P_1 и P_2 плоскости $z = \text{const}$ (рис. 4.2) колебания происходят по одному и тому же закону с одной и той же частотой, амплитудой и одной и той же начальной фазой φ . Величина S во всех точках данной плоскости $z = \text{const}$ одна и та же в любой момент времени. Поверхности, на которых колебания возмущения S происходят синфазно, называются *волновыми поверхностями*. В данном случае волновые поверхности — плоскости (перпендикулярные оси z). Поэтому волна (4.6) называется *плоской*. Поскольку колебания происходят по гармоническому закону, волна (4.6) также называется *гармонической* (в оптике принят термин «монохроматическая», т. е. одноцветная).

Пусть некоторый наблюдатель перемещается в направлении оси z со скоростью V_Φ так, что величина фазы волны $\Phi = \omega t - kz - \varphi$ в процессе перемещения не меняется

$$\Phi = \omega t - kz - \varphi = \text{const}. \quad (4.9)$$

Дифференцируя (4.9) по t , находим

$$\frac{dz}{dt} = V_\Phi = \frac{\omega}{k}.$$

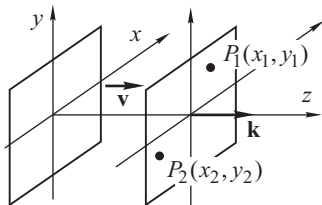


Рис. 4.2

Полученная нами величина V_ϕ называется *фазовой скоростью* (скорость, с которой перемещается поверхность постоянной фазы колебаний — волновая поверхность). Мы видим, что она совпадает с константой v , входящей в волновое уравнение (4.4).

Введем еще одно важное понятие — *волновой вектор* \mathbf{k} . Это вектор, совпадающий по направлению с нормалью к волновой поверхности (и вектором скорости, с которой перемещается волновая поверхность, рис. 4.2), причем модуль вектора \mathbf{k} равен волновому числу $|\mathbf{k}| = k = \omega/v$.

Разность фаз $\Delta\varphi$ колебаний на двух волновых поверхностях, разделенных расстоянием Δz , равна $\Delta\varphi = k\Delta z$. Найдем расстояние между волновыми поверхностями, колебания на которых происходят с разностью фаз $\Delta\varphi = 2\pi$ (т. е. колебания на этих поверхностях происходят синхронно, в такт). Мы получаем $\Delta z = 2\pi/k$. Это расстояние Δz называется *длиной волны* и обозначается буквой λ :

$$\lambda = \frac{2\pi}{k}. \quad (4.10)$$

Используя определение (4.8) волнового числа k , находим

$$\lambda = \frac{2\pi}{\omega} v = vT, \quad (4.11)$$

где $T = 2\pi/\omega$ — период колебания. Из (4.11) ясно, что длина волны есть расстояние, на которое перемещается волновая поверхность за время, равное одному периоду колебания.

Рассмотрим теперь плоскую монохроматическую волну, распространяющуюся в произвольном направлении (например вдоль оси z' , рис. 4.3). Волновой вектор \mathbf{k} этой волны совпадает по направлению с осью z' , а волновые поверхности —

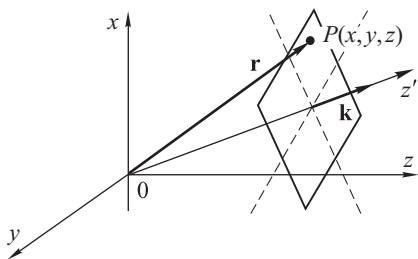


Рис. 4.3

плоскости, нормаль к которым совпадает по направлению с вектором \mathbf{k} . Напомним, что уравнение таких плоскостей имеет вид

$$\mathbf{k} \cdot \mathbf{r} = k_x x + k_y y + k_z z = \text{const},$$

где $\mathbf{k} \cdot \mathbf{r}$ — скалярное произведение вектора $\mathbf{k} (k_x, k_y, k_z)$ и радиуса-вектора \mathbf{r} точки наблюдения $P: \mathbf{r}(x, y, z)$. Итак, мы получили уравнение плоской волны, распространяющейся в произвольном направлении

$$S(\mathbf{r}, t) = a \cos(\omega t - \mathbf{k} \cdot \mathbf{r} - \varphi). \quad (4.12)$$

Как следует из (4.12), во всех точках (x, y, z) , принадлежащих плоскости $\mathbf{k} \cdot \mathbf{r} = \text{const}$, колебания происходят синфазно. Предоставляем читателю убедиться непосредственной постановкой, что функция (4.12) удовлетворяет волновому уравнению (4.5), если

$$k_x^2 + k_y^2 + k_z^2 = \left(\frac{\omega}{v}\right)^2 = k^2, \quad (4.13)$$

причем частота волны может быть любой. Направление распространения волны также произвольно: проекции волнового вектора \mathbf{k} можно изменять как угодно, лишь бы сумма их квадратов удовлетворяла равенству (4.13).

Таким образом, уравнение (4.12) описывает целую совокупность, множество решений волнового уравнения — множество плоских волн, отличающихся частотой колебаний и направлением распространения.

Плоские монохроматические волны (4.12) (или, в частном случае, волны (4.6)) называют также *бегущими плоскими волнами*, подчеркивая таким образом, что волновые поверхности перемещаются в пространстве, и скорость их перемещения — это и есть *фазовая* скорость.

Стоячая волна. Рассмотрим волновой процесс, представляющий собой суперпозицию, сумму волн, одна из которых бежит в направлении оси z :

$$S_1(z, t) = a \cos(\omega t - kz),$$

а вторая (с той же амплитудой и частотой) — во встречном направлении

$$S_2(z, t) = a \cos(\omega t + kz).$$

Легко убедиться (упражнение для читателя), что сумма $S = S_1 + S_2$ есть решение волнового уравнения (4.4). Складывая S_1 и S_2 , находим

$$S(z, t) = A(z) \cos \omega t, \quad (4.14)$$

где

$$A(z) = 2a \cos kz.$$

Согласно (4.14) колебания во всех точках происходят с одной и той же частотой ω , однако амплитуда колебаний $A(z)$ в разных точках (в разных плоскостях z) различна (рис. 4.4).

В частности, в плоскостях, для которых

$$kz = \pi n + \frac{\pi}{2} \quad (n = 0, \pm 1, \pm 2, \dots), \quad (4.15)$$

амплитуда колебаний равна нулю. Точки нулевой амплитуды называются *узлами* стоячей волны. Как видно из (4.15), расстояние между соседними узлами равно $\lambda/2$. В плоскостях, для которых

$kz = n\pi$ ($n = 0, \pm 1, \pm 2, \dots$), амплитуда колебаний максимальна и равна $2a$ (напомним, что a — амплитуды встречных бегущих волн).

Точки максимальной амплитуды называются *пучностями*, которые располагаются точно посередине между узлами, т. е. на расстоянии $\lambda/4$ от ближайших узлов (и $\lambda/2$ друг от друга).

Заметим, что при тех значениях z , где величина $A(z)$ отрицательна, функцию $S(z, t)$ можно записать в виде

$$S(z, t) = |A(z)| \cos(\omega t + \pi),$$

откуда ясно, что отрицательность амплитуды $A(z)$ означает лишь, что колебания происходят в противофазе (со сдвигом фазы на π) по сравнению с колебаниями при тех z , где $A(z) > 0$

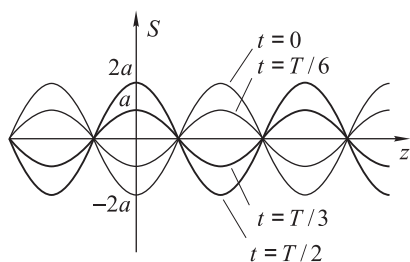


Рис. 4.4

(рис. 4.4). Вообще, как видно из (4.14), в точках нулевой амплитуды колебаний происходит скачкообразное изменение фазы на величину π . Например, если при $|z| < \lambda/4$ колебания происходят по закону $S = A(z) \cos \omega t$, то при $\lambda/4 < z < 3\lambda/4$ имеем: $S(z) = |A(z)| \cos(\omega t + \pi)$ и т. д.

В суммарном волновом процессе (4.14), образованном встречными волнами одинаковой частоты и амплитуды, нет перемещающихся волновых поверхностей (так как нет аргумента вида $\omega t \pm kz$), поэтому волна (4.14) называется *стоячей волной*.

Сферическая волна. Волна, описываемая уравнением

$$S(r, t) = \frac{a}{r} \cos(\omega t - kr - \varphi_0), \quad (4.16)$$

называется *сферической*. В уравнении (4.16) r — расстояние от начала координат (где расположен источник волны) до точки наблюдения (x, y, z)

$$r = \sqrt{x^2 + y^2 + z^2},$$

$k = \omega/v$ — волновое число. Обратная пропорциональность амплитуды радиусу — свойство сферически симметричного решения уравнения (4.5). Понятен смысл названия «сферическая волна»: согласно (4.16) колебания S происходят синфазно во всех точках пространства, отстоящих от источника на заданном расстоянии $r = \text{const}$, т. е. волновые поверхности — это сферы, центр которых (источник волны) находится в начале координат. Обратим внимание на различие уравнений (4.12) и (4.16). В уравнении плоской волны аргумент косинуса содержит ска-

лярное произведение векторов \mathbf{k} и \mathbf{r} . В уравнении сферической волны — произведение волнового числа k на расстояние r точки наблюдения от источника волны. Дифференцируя равенство $\omega t - kr = \text{const}$, получаем $dr/dt = \omega/k > 0$, т. е. волновые поверхности бегут от источника ($dr > 0$). Легко видеть, что уравнение

$$S(r, t) = \frac{a_0}{r} \cos(\omega t + kr - \varphi_0)$$

описывает сферическую волну, в которой волновые поверхности бегут к началу координат ($dr/dt = -\omega/k < 0$). Это — *сходящаяся сферическая волна*.

Исследуем колебания, созданные сферической волной в плоскости, отстоящей от источника на большом расстоянии z . При этом часто используют приближенное выражение, существенно упрощающее многие расчеты. Пусть размер области в плоскости наблюдения, которой мы интересуемся, мал по сравнению с z (эта область выделена на рис. 4.5). Тогда амплитуду колебаний a_0/r в этой области можно считать константой: $a_0/r \approx a_0/z = a = \text{const}$ во всех точках, для которых $x^2 + y^2 \ll z^2$.

Сложнее обстоит дело с оценкой фазы колебаний kr . Пользуясь приближенным выражением для r , мы должны быть уверены, что ошибка в вычислении r мала по сравнению с длиной волны λ . Действительно, ошибка при вычислении расстояния r , равная $\lambda/2$, приводит к ошибке в вычислении фазы колебаний, равной π . Это недопустимо, когда речь идет о сложении колебаний: например, вместо синфазного сложения волн мы получим противофазное — амплитуда суммарного колебания будет найдена неверно. Воспользуемся следующим приближением:

$$r = \sqrt{z^2 + x^2 + y^2} = z \left(1 + \frac{x^2 + y^2}{z^2} \right)^{1/2} \approx z + \frac{x^2 + y^2}{2z}$$

(приближенное выражение $\sqrt{1+\varepsilon} \approx 1 + \varepsilon/2$ справедливо при $\varepsilon \ll 1$). Мы полагаем при этом, что отброшенные члены разложения радикала в степенной ряд дают поправку, малую в сравнении с λ . Например, достаточно потребовать выполнения неравенства $(x^2 + y^2)^2/(6z^3) \ll \lambda$. В этом случае, учитывая все вышесказанное, получаем

$$S = a \cos[\omega t - \varphi(x, y)], \quad \text{где} \quad \varphi(x, y) = kz - \frac{k}{2z}(x^2 + y^2) - \varphi_0 \quad (4.17)$$

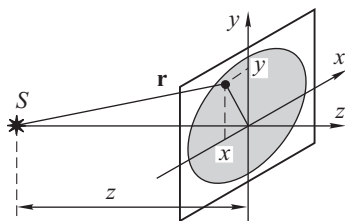


Рис. 4.5

(плоскость наблюдения $z = \text{const}$ предполагается фиксированной). Поскольку роль играет лишь относительная разность фаз, то можно, приняв фазу колебаний в начале координат плоскости наблюдения нулевой, записать

$$\varphi(x, y) \approx \frac{k}{2z} (x^2 + y^2).$$

Используя сделанные приближения, мы, по существу, полагаем, что сферическую форму волнового фронта можно на малом участке заменить параболической, поэтому приближение (4.17) называется *параболическим*. В дальнейшем всюду, где это необходимо, мы будем пользоваться параболическим приближением без каких-либо оговорок.

4.2. Монохроматические волны. Комплексная амплитуда. Уравнение Гельмгольца

Выше мы рассмотрели простейшие примеры монохроматических волн. В самом общем случае уравнение монохроматической волны (не обязательно плоской, сферической или цилиндрической) имеет вид

$$S(\mathbf{r}, t) = a(\mathbf{r}) \cos(\omega t - \varphi(\mathbf{r})), \quad (4.18)$$

где \mathbf{r} — радиус-вектор точки наблюдения.

В любой фиксированной точке наблюдения волна (4.18) создает гармонические колебания частоты ω с амплитудой $a(\mathbf{r})$ и начальной фазой $\varphi(\mathbf{r})$. Функции $a(\mathbf{r})$ и $\varphi(\mathbf{r})$ определяют пространственную структуру волны, в том числе и форму волновых поверхностей.

При изучении монохроматических волн, так же, как и при изучении гармонических колебаний, удобно использовать комплексную форму записи. Наряду с волной (4.18), рассмотрим волновой процесс вида

$$S_1(\mathbf{r}, t) = a(\mathbf{r}) \sin(\omega t - \varphi(\mathbf{r})). \quad (4.19)$$

Ясно, что если функция (4.18) действительно описывает некоторый волновой процесс (т. е. удовлетворяет волновому уравнению), то функция (4.19) также описывает волну: ведь (4.19) можно получить из (4.18) просто изменением начала отсчета времени: $S_1(\mathbf{r}, t) = S(\mathbf{r}, t + T/4)$, где $T = 2\pi/\omega$ — период колебания. Ясно также, что линейная комбинация функций (4.18) и (4.19) вида

$$V(\mathbf{r}, t) = S(\mathbf{r}, t) - iS_1(\mathbf{r}, t) \quad (4.20)$$

также удовлетворяет волновому уравнению. (Линейная суперпозиция решений волнового уравнения (4.5) также является решением.)

Используя тождество $\cos \alpha - i \sin \alpha = e^{-i\alpha}$, перепишем (4.20) в виде

$$V(\mathbf{r}, t) = a(\mathbf{r})e^{-i[\omega t - \varphi(\mathbf{r})]}. \quad (4.21)$$

Функцию $V(\mathbf{r}, t)$ можно записать в виде произведения двух функций, одна из которых

$$f(\mathbf{r}) = a(\mathbf{r})e^{i\varphi(\mathbf{r})} \quad (4.22)$$

зависит только от координат, а вторая $e^{-i\omega t}$ — только от времени,

$$V(\mathbf{r}, t) = f(\mathbf{r})e^{-i\omega t}. \quad (4.23)$$

Комплексная функция $f(\mathbf{r})$ полностью определяет пространственную структуру волны (т. е. распределение в пространстве амплитуд $a(\mathbf{r})$ и фаз $\varphi(\mathbf{r})$ колебаний) и называется *комплексной амплитудой волны*.

Итак, комплексная функция $V(\mathbf{r}, t)$, определяемая равенством (4.23), должна быть решением волнового уравнения

$$\nabla^2 V - \frac{1}{v^2} \frac{\partial^2 V}{\partial t^2} = 0. \quad (4.24)$$

Дифференцируя (4.23) дважды по координатам, получим

$$\nabla^2 V = e^{-i\omega t} \nabla^2 f.$$

Дифференцируя (4.23) дважды по времени, находим

$$\frac{\partial^2 V}{\partial t^2} = f(\mathbf{r})(-i\omega)^2 e^{-i\omega t}.$$

Подставляя выражения $\nabla^2 V$ и $\partial^2 V / \partial t^2$ в (4.24), приходим (после сокращения на множитель $e^{-i\omega t}$) к уравнению

$$\nabla^2 f + k^2 f = 0, \quad (4.25)$$

где $k = \omega/v$ — волновое число. Уравнение (4.25) для комплексных амплитуд называется *уравнением Гельмгольца*.

Пример 1. *Комплексная амплитуда плоской волны.* В этом случае, согласно (4.12), $a(\mathbf{r}) = a = \text{const}$ и $\varphi(\mathbf{r}) = \mathbf{k} \cdot \mathbf{r} + \varphi_0$, поэтому

$$f(\mathbf{r}) = ae^{i\varphi_0} e^{i\mathbf{k}\mathbf{r}} = ce^{i(k_x x + k_y y + k_z z)}, \quad (4.26)$$

где $c = ae^{i\varphi_0}$ — комплексный множитель, определяющий амплитуду волны a и ее начальную фазу φ_0 .

Пример 2. *Комплексная амплитуда сферической волны.* Используя (4.16), имеем $a(\mathbf{r}) = a_0/r$ и $\varphi(\mathbf{r}) = kr + \varphi_0$. Напомним, в отличие от (4.12) здесь $r = \sqrt{x^2 + y^2 + z^2}$ — расстояние от источника волны до точки наблюдения (x, y, z) . Соответствующая комплексная амплитуда, согласно определению (4.23), есть

$$f(r) = \frac{a_0}{r} e^{ikr}. \quad (4.27)$$

Используя параболическое приближение (4.17), имеем

$$f(r) = \frac{a_0}{z} e^{ikz} e^{i\frac{k}{2z}(x^2+y^2)}. \quad (4.28)$$

(полагаем $\varphi_0 = 0$).

4.3. Векторные волны

До сих пор речь шла о скалярных волнах, т. е. о волновом возмущении $S(\mathbf{r}, t)$, которое описывается скалярной величиной S . Например, таким образом можно представить звуковые волны, распространяющиеся в воздухе. В этом случае $S(\mathbf{r}, t)$ — изменения плотности или давления — скалярные величины.

Теперь мы рассмотрим кинематику простейших векторных волн: речь будет идти об изменяющейся в пространстве и во времени векторной величине \mathbf{S} (например, напряженности электрического или магнитного полей в электромагнитной волне).

В общем случае вектор \mathbf{S} может иметь произвольное направление по отношению к направлению распространения (т. е. по отношению к волновому вектору \mathbf{k}). Если вектор \mathbf{S} параллелен направлению распространения, волна называется *продольной*. Если вектор \mathbf{S} перпендикулярен направлению распространения, волна называется *поперечной*. Именно особенности структуры поперечных волн мы здесь рассмотрим.

Наиболее простой тип поперечной волны — так называемая *плоскополяризованная волна* (употребляется также термин *линейно-поляризованная волна*). В плоскополяризованной волне колебания вектора \mathbf{S} происходят в фиксированной плоскости.

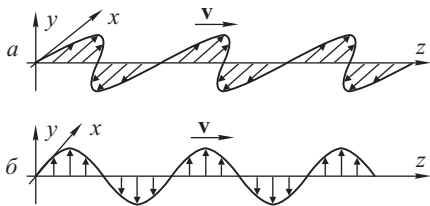


Рис. 4.6

Например, на рис. 4.6 а показана плоскополяризованная волна, бегущая вдоль оси z , при этом колебания вектора \mathbf{S} происходят в плоскости xz параллельно оси x : вектор \mathbf{S} не имеет проекции на ось y . На рис. 4.6 б — линейно-поляризованная волна, также бегущая вдоль оси z . В этой

волне колебания \mathbf{S} происходят в плоскости yz параллельно оси y , т. е. вектор \mathbf{S} не имеет x -компоненты (и, конечно, не имеет проекции на направление распространения z).

В общем случае поперечная волна, бегущая вдоль оси z , может иметь как x -проекцию S_x , так и y -проекцию S_y . Рассмотрим ситуацию, когда каждая из проекций представляет собой плоскую гармоническую волну одной и той же частоты:

$$S_x = a_1 \cos(\omega t - kz - \alpha_1), \quad S_y = a_2 \cos(\omega t - kz - \alpha_2), \quad (4.29)$$

где α_1 и α_2 — константы, начальные фазы каждой из волн S_x и S_y . Из последних равенств следует (см. § 1.5), что в любой фиксированной плоскости $z = \text{const}$ (перпендикулярной направлению распространения) конец вектора \mathbf{S} описывает эллипс. Поскольку разность фаз колебаний S_x и S_y не зависит от z (и равна $\alpha_1 - \alpha_2$), то и эллипсы во всех плоскостях $z = \text{const}$ одинаковы (они описываются одним и тем же уравнением). Таким образом, гармоническая поперечная волна является в общем случае *эллиптически поляризованной*.

В частности, при $\alpha_2 - \alpha_1 = \pm\pi/2$ легко получить (упражнение для читателя):

$$\left(\frac{S_x}{a_1}\right)^2 + \left(\frac{S_y}{a_2}\right)^2 = 1,$$

т. е. главные оси эллипсов совпадают с осями координат. Если, кроме того, $a_1 = a_2 = a$, то имеем уравнение окружности

$$S_x^2 + S_y^2 = a^2.$$

В этом случае волна называется *поляризованной по кругу*: в любой плоскости $z = \text{const}$ конец вектора \mathbf{S} описывает окружность.

При $\alpha_2 - \alpha_1 = 0$ из (4.29) получаем уравнение прямой

$$S_y = \frac{a_2}{a_1} S_x,$$

т. е. в любой момент времени конец вектора \mathbf{S} находится на прямой, составляющей угол φ ($\text{tg } \varphi = a_2/a_1$) с осью x — колебания \mathbf{S} происходят вдоль этой прямой. Мы видим, что в любой фиксированной плоскости $z = \text{const}$ колебания \mathbf{S} происходят по одному и тому же направлению — вдоль прямой, лежащей в первом и третьем квадрантах (рис. 4.7 а), т. е. мы имеем линейно-поляризованную волну.

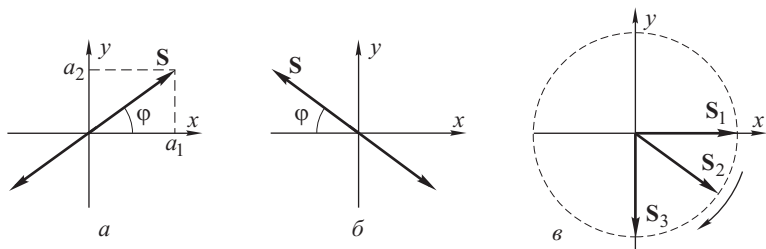


Рис. 4.7

При $\alpha_2 - \alpha_1 = \pm\pi$ уравнение прямой есть

$$S_y = -\frac{a_2}{a_1} S_x$$

— вновь имеем линейно-поляризованную волну (рис. 4.7б), однако колебания \mathbf{S} происходят вдоль прямой, лежащей во втором и четвертом квадрантах.

В общем случае, как отмечалось, конец вектора \mathbf{S} описывает эллипс, совершая полный оборот за один период колебаний. Как определить направление вращения вектора? Вернемся к эллиптически-поляризованной волне, в которой $\alpha_2 - \alpha_1 = \pi/2$. Фиксируем плоскость наблюдения $z = \text{const}$ и проследим за ориентацией вектора \mathbf{S} в этой плоскости в различные моменты времени. Полагая $kz + \alpha_1 = \beta_1$ и $kz + \alpha_2 = \beta_2$, перепишем (4.29) в виде

$$S_x = a_1 \cos(\omega t - \beta_1), \quad S_y = a_2 \cos(\omega t - \beta_2),$$

и поскольку $\beta_2 - \beta_1 = \pi/2$, то

$$S_x = a_1 \cos(\omega t - \beta_1), \quad S_y = -a_2 \sin(\omega t - \beta_1).$$

Из последних равенств следует, что в момент времени, когда $\omega t - \beta_1 = 0$, $S_x = a_1$, $S_y = 0$, т. е. вектор \mathbf{S} направлен вдоль оси x . На рис. 4.7в его положение в этот момент изображено вектором \mathbf{S}_1 . В последующие моменты времени (когда $0 < \omega t - \beta_1 < \pi/2$) $0 < S_x < a_1$, а $S_y < 0$, т. е. вектор \mathbf{S} , повернувшись по часовой стрелке, занял положение, показанное на рисунке вектором \mathbf{S}_2 . Далее, при $\omega t - \beta_1 = \pi/2$ имеем $S_x = 0$, $S_y = -a_2$ — вектор \mathbf{S} занял положение, показанное вектором \mathbf{S}_3 .

Итак, мы видим, что вектор \mathbf{S} вращается по часовой стрелке (с точки зрения читателя, к которому направлен волновой вектор). Это — так называемая *правая эллиптическая поляризация*. Читатель может самостоятельно убедиться в том, что при $\alpha_2 - \alpha_1 = -\pi/2$ направление вращения меняется на противоположное (*левая эллиптическая поляризация*).

Задачи

1. Напишите выражение для комплексной амплитуды плоской волны, имеющей амплитуду a и распространяющуюся под углом $\alpha = \pi/6$ к оси z (вектор \mathbf{k} лежит в плоскости xz).

2. То же для плоской волны, распространяющейся вдоль оси z .

3. Напишите выражение для комплексной амплитуды сферической волны, источник которой находится в начале координат.

4. Докажите, что комплексные амплитуды плоской и сферической волн удовлетворяют уравнению Гельмгольца.

5. Докажите, что функция $S(z - vt)$ описывает волну, бегущую в направлении оси z , при этом функция $S(z + vt)$ описывает волну, бегущую в противоположном направлении — навстречу оси z .

6. Покажите, что расстояние между узлами стоячей волны равно $\lambda/2$.

7. Напишите уравнение монохроматической эллиптически поляризованной волны с правой поляризацией.

8. Каково достаточное условие применимости параболического приближения для сферической волны?

Глава 5

УПРУГИЕ ВОЛНЫ

Мы погружены в океан волн, создающих особый таинственный и чарующий мир звуков: рокот морского прибоя, звучание скрипичной струны, голоса людей — все это волны, рожденные в упругих средах.

В этой главе мы рассмотрим конкретные физические механизмы, приводящие к появлению упругих (акустических) волн в твердых телах, жидкостях и газах.

5.1. Продольные упругие волны в твердом теле

Рассмотрим механизм распространения продольных упругих взаимодействий в стержне. Пусть ось z направлена вдоль оси стержня и пусть частицы, составляющие стержень, могут смещаться лишь вдоль оси z . Рассмотрим частицы, находящиеся в момент времени t в некоторой плоскости $z = \text{const}$ и пусть в момент времени t' эти частицы сместились, оказавшись в плоскости $z' = z + \zeta$, т. е. величина ζ представляет собой *смещение* сечения z (рис. 5.1). Разные сечения смещаются, вообще говоря, на разные расстояния ζ , поэтому ζ есть функция рассматриваемого сечения $\zeta = \zeta(z)$.

Рассмотрим участок стержня между двумя сечениями z и $z + \Delta z$ (длина этого участка равна Δz). Пусть в результате распространения упругих взаимодействий в стержне сечение z сместилось в положение z' , а сечение $z + \Delta z$ — в положение $z' + \Delta z'$. Мы имеем

$$z' = z + \zeta(z), \quad z' + \Delta z' = z + \Delta z + \zeta(z + \Delta z)$$

здесь $\zeta(z + \Delta z)$ — смещение сечения $z + \Delta z$.

Длина рассматриваемого участка изменилась и стала равной $(z' + \Delta z') - z' = \Delta z'$, а изменение его длины (удлинение) равно

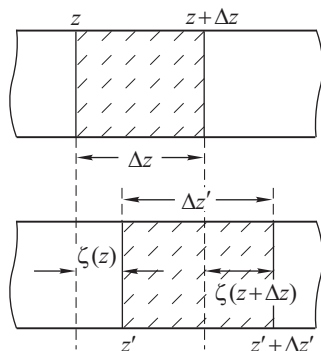


Рис. 5.1

$\Delta z' - \Delta z = \zeta(z + \Delta z) - \zeta(z)$. Отношение удлинения к первоначальной длине (относительное удлинение) равно

$$\frac{\zeta(z + \Delta z) - \zeta(z)}{\Delta z}.$$

В пределе, при $\Delta z \rightarrow 0$, эта величина называется *продольной деформацией*:

$$\varepsilon = \lim_{\Delta z \rightarrow 0} \frac{\zeta(z + \Delta z) - \zeta(z)}{\Delta z} = \frac{d\zeta}{dz}. \quad (5.1)$$

Если в данном сечении $\varepsilon > 0$ (как на рис. 5.1), то произошло удлинение данного участка; если $\varepsilon < 0$, то его сжатие. Итак, согласно определению (5.1), *деформация $\varepsilon(z)$ есть производная от смещения $\zeta(z)$* , т. е. на графике функции $\zeta(z)$ тангенс угла наклона касательной в каждой точке дает величину $\varepsilon(z)$. Рассмотрим некоторое сечение стержня $z = \text{const}$ (рис. 5.2).

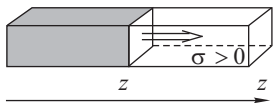


Рис. 5.2

Если стержень деформируется (растягивается или сжимается), то левая (заштрихованная) часть стержня действует на правую часть с некоторой силой. Отношение этой силы к площади сечения стержня называется *напряжением σ* .

Условимся считать величину σ положительной, если правая часть стержня *тянет* левую часть вправо, в положительном направлении оси z (как на рис. 5.2).

Далее мы будем полагать, что в стержне отсутствуют статические напряжения, обусловленные механическими нагрузками, электрическими, магнитными полями и т. д.

Основной закон упругих взаимодействий (закон Гука см. раздел «Механика»), справедливый при малых деформациях, гласит: *напряжение прямо пропорционально величине деформации*:

$$\sigma = E\varepsilon, \quad (5.2)$$

где E — константа, называемая *модулем Юнга*.

Покажем теперь, что распространение упругих взаимодействий, т. е. передача деформации ε от одного сечения стержня к другому, действительно представляет собой волновой процесс, т. е. величина ε , зависящая от времени t и координаты z , подчиняется волновому уравнению

$$\frac{\partial^2 \varepsilon}{\partial z^2} - \frac{1}{v^2} \frac{\partial^2 \varepsilon}{\partial t^2} = 0.$$

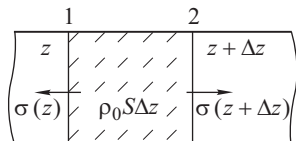


Рис. 5.3

Применим второй закон Ньютона к участку стержня, заключенному между двумя сечениями z и $z + \Delta z$ (рис. 5.3). Масса рассматриваемого участка $m = \rho_0 S \Delta z$ (S — площадь сечения, ρ_0 —

плотность). Пусть ζ — смещение центра масс участка стержня, $d^2\zeta/dt^2$ — его ускорение. Мы получаем

$$\rho_0 S \Delta z \frac{d^2\zeta}{dt^2} = S\sigma(z + \Delta z) - S\sigma(z)$$

(в правой части равенства — сумма внешних сил, действующих на участок стержня Δz со стороны прилегающих к нему частей стержня справа и слева). Разделив обе части равенства на $S \cdot \Delta z$ и перейдя к пределу при $\Delta z \rightarrow 0$, находим:

$$\rho_0 \frac{d^2\zeta}{dt^2} = \frac{d\sigma}{dz},$$

или, используя закон Гука:

$$\rho_0 \frac{d^2\zeta}{dt^2} = E \frac{d\varepsilon}{dz}. \quad (5.3)$$

Согласно (5.1) деформация ε есть производная от смещения, т. е. $d\varepsilon/dz = d^2\zeta/dz^2$, поэтому окончательно получаем

$$\frac{\partial^2\zeta}{\partial z^2} = \frac{\rho_0}{E} \frac{\partial^2\zeta}{\partial t^2}. \quad (5.4)$$

Мы пришли к волновому уравнению, а это означает, что смещение $\zeta(z, t)$ распространяется по стержню в виде волн:

$$\zeta = \zeta(z \pm vt), \quad (5.5)$$

причем скорость распространения волн (скорость звука в стержне) равна

$$v = \sqrt{\frac{E}{\rho_0}}. \quad (5.6)$$

Из (5.6) следует, что скорость звука тем больше, чем жестче (больше E) и легче (меньше ρ_0) среда. Например, в стали ($E = 1,96 \cdot 10^{10}$ Н/м², $\rho_0 = 7,7 \cdot 10^3$ кг/м³) $v \approx 5000$ м/с, примерно в пять раз больше, чем в свинце.

Если смещение ζ найдено (т. е. найдено решение (5.5) волнового уравнения), то скорость, с которой движутся различные сечения z стержня, найдем, дифференцируя (5.5) по t (при фиксированном z):

$$u = \frac{\partial\zeta}{\partial t} = \pm v \zeta'(z \pm vt), \quad (5.7)$$

где ζ' означает производную по аргументу $z \pm vt$.

Волну деформации $\varepsilon(z, t)$ найдем, используя (5.1):

$$\varepsilon(z, t) = \frac{\partial\zeta}{\partial z} = \zeta'(z \pm vt). \quad (5.8)$$

Следует подчеркнуть, что волна смещений (как и волна деформаций) распространяется вдоль стержня на сколь угодно большие расстояния (в соответствии с (5.5) и (5.8)). Однако сами частицы стержня, находящиеся в любом сечении, смещаются при этом лишь незначительно, приводя в движение частицы соседних сечений, те в свою очередь передают свой импульс соседним частицам и т. д. — именно так малые смещения частиц стержня приводят к распространению волнового возмущения на большие расстояния.

5.2. Упругие волны в жидкостях и газах

Рассмотрим жидкость или газ, заключенный в гладкой цилиндрической трубе. Быстрое (не квазистатическое!) изменение давления в прилегающем к поршню слое нарушают равновесное состояние: давление и плотность в различных сечениях трубы оказываются различными. Является ли процесс распространения возмущений вдоль оси трубы волной?

Рассмотрим слой среды между плоскостями z и $z + \Delta z$ (см. рис. 5.1). Его ускорение обусловлено разностью давлений в этих сечениях. Пусть давление в плоскости z равно $p_0 + \Delta p(z)$, а в плоскости $z + \Delta z$ давление равно $p_0 + \Delta p(z + \Delta z)$. Используя второй закон Ньютона, получаем

$$\rho S \Delta z \frac{d^2 \xi}{dt^2} = S (\Delta p(z) - \Delta p(z + \Delta z)),$$

где ξ — смещение центра масс участка среды, $d^2 \xi / dt^2$ — его ускорение. Разделив обе части равенства на $S \Delta z$ и переходя к пределу при $\Delta z \rightarrow 0$, получаем

$$\rho \frac{d^2 \xi}{dt^2} = - \frac{\partial}{\partial z} \Delta p. \quad (5.9)$$

Перепад давлений приводит не только к ускоренному движению участка среды Δz , но и к его деформации — сжатию либо расширению, а следовательно, к изменению плотности.

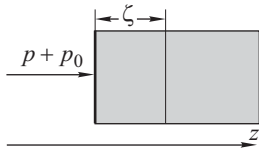


Рис. 5.4

Пусть в невозмущенном состоянии давление в газе (или жидкости) равно p_0 . При быстром смещении подвижного поршня, находящегося на левом конце трубы, на расстояние ζ (рис. 5.4) давление в тонком прилегающем слое изменится на величину Δp . Будем предполагать, что $\Delta p \ll p_0$.

В общем случае давление является функцией плотности ρ и температуры T . В силу малой теплопроводности газов и жид-

костей можно полагать, что процесс сжатия и разряжения данной массы жидкости или газа (заключенного на участке трубы Δz) происходит без обмена теплом с соседними участками, т. е. *адиабатически*. В этом случае плотность и температура связаны *уравнением адиабаты* (в частности, для идеального газа $T = T_0(\rho/\rho_0)^{\gamma-1}$, где $\gamma = c_p/c_v$) и поэтому давление можно считать функцией лишь одного параметра — плотности, т. е. $p = p(\rho)$. Тогда, *при малых изменениях плотности*, можно записать

$$\Delta p = \frac{dp}{d\rho} \Delta \rho. \quad (5.10)$$

Производная $dp/d\rho$ берется в точке $\rho = \rho_0$, $p_0 = p(\rho_0)$. Величина Δp называется *акустическим давлением*, это отклонение давления от равновесного значения при распространении акустической волны.

Свяжем теперь изменение плотности с изменением давления, используя соотношение $(\rho_0 + \Delta\rho)(V_0 + \Delta V) = \rho_0 V_0 = m$ (m — фиксированная масса газа между сечениями z и $z + \Delta z$). Получаем

$$\frac{\Delta V}{V_0} = \varepsilon = -\frac{\Delta \rho}{\rho_0 + \Delta \rho} \approx -\frac{\Delta \rho}{\rho_0}.$$

Таким образом, формула (5.10) приобретает вид

$$\Delta p = -\rho_0 \frac{dp}{d\rho} \varepsilon. \quad (5.11)$$

Как видно из (5.11), роль модуля упругости в жидкости и газе играет величина $E = \rho (dp/d\rho)$, причем $\Delta p = -\sigma$ (знак «-» указывает, что увеличение давления ($\Delta p > 0$) приводит к сжатию ($\varepsilon > 0$)). Подставляя (5.11) в (5.9), приходим к волновому уравнению

$$\frac{\partial^2 \zeta}{\partial t^2} = \frac{1}{dp/d\rho} \frac{\partial^2 \zeta}{\partial z^2},$$

в котором величина $dp/d\rho$ представляет собой квадрат скорости распространения звуковой волны

$$v = \sqrt{\frac{dp}{d\rho}}. \quad (5.12)$$

В качестве примера рассмотрим идеальный газ: $p = \rho RT$. Как показано Лапласом, скорость звука в таком газе

$$v = \sqrt{\frac{\gamma p_0}{\rho_0}} = \sqrt{\gamma RT_0}, \quad (5.13)$$

где $\gamma = c_p/c_v$.

Формула (5.13) имеет ясный физический смысл: передача возмущений в звуковой волне в газе осуществляется за счет теп-

лового движения молекул, поэтому не удивительно, что скорость звука равна по порядку величины скорости теплового движения (равной $\sqrt{v^2} = \sqrt{3RT/\mu}$). Сжимаемость жидкостей существенно меньше, чем газов, т. е. одинаковый перепад давлений dp приводит к меньшему изменению плотности жидкости dp в сравнении с газом. Поэтому, в согласии с формулой (5.12), скорость звука в жидкостях больше, чем в газах. Приведем характерные значения скоростей звуковых волн в некоторых средах при атмосферном давлении и комнатной температуре: воздух $v = 331$ м/с, кислород $v = 316$ м/с, водород $v = 1\,265$ м/с, вода $v = 1\,480$ м/с, ртуть $v = 1\,460$ м/с, медь $v = 3\,680$ м/с, сталь $v = 4\,980$ м/с.

Согласно принятой терминологии акустические колебания с частотами $\nu < 16$ Гц называют *инфразвуком*, а с частотами $\nu > 20\,000$ Гц и вплоть до $\nu = 10^9$ Гц — наиболее высоких полученных в настоящее время частот акустических волн — *ультразвуком*. Собственно диапазон звуковых колебаний (т. е. слышимых человеком) чрезвычайно широк и простирается от 16 Гц до 20 кГц. Наиболее чувствительно ухо человека к диапазону частот от 1 до 5 кГц. Соответствующие длины акустических волн в воздухе от 20 м до 2 см (звук) и от 2 см до $0,5 \cdot 10^{-4}$ см (ультразвук). Гармоническое колебание называют в акустике *чистым тоном*, *тембр звука* определяется результатом суперпозиции колебаний кратных частот: *основной частоты* ν_1 , определяющей *высоту звука* и *обертонов* с частотами $2\nu_1, 3\nu_1, \dots$, амплитуды которых, вообще говоря, меньше амплитуды основного тона. Различие в составе обертонов и отличает различные звуки по тембру.

5.3. Плотность и поток энергии в упругой волне. Вектор Умова

Рассмотрим малый участок стержня Δz . Его масса $m = \rho S \Delta z$ (S — площадь поперечного сечения). При распространении в стержне звуковой волны происходит смещение вдоль оси z различных участков стержня $\zeta = \zeta(z \pm vt)$; рассматриваемый участок имеет скорость $u = \partial \zeta / \partial t = \pm v \zeta'$ (напомним, ζ' означает дифференцирование по аргументу $z \pm vt$). Кинетическая энергия участка стержня есть $mu^2/2 = (1/2)\rho S \Delta z \cdot u^2$. На единицу объема приходится энергия

$$W_k = \frac{\rho u^2}{2}, \quad (5.14)$$

называемая *объемной плотностью кинетической энергии*. Распространение упругой волны сопровождается, как мы видели, деформацией различных участков стержня. Пусть для простоты

левая граница участка Δz (сечение 1 на рис. 5.5) неподвижна, а правая смещается на расстояние $\zeta = \varepsilon \Delta z$. Работа внешней силы \mathbf{F} (упругой силы, действующей на выделенный участок со стороны прилегающего к нему участка справа) при растяжении участка Δz на величину ζ есть

$$A = \int_0^{\zeta} F d\zeta.$$

Величина деформации ε и, соответственно, упругая сила $F = E\varepsilon S$ меняются в процессе деформации. При смещении сечения 2 (на рис. 5.5) на величину $d\zeta$ деформация участка стержня меняется на величину $d\varepsilon$, причем $d\zeta = \Delta z d\varepsilon$. Упругая сила F при заданной величине деформации, согласно закону Гука, есть $F = SE\varepsilon$, а ее работа при смещении сечения 2 на $d\zeta$ равна

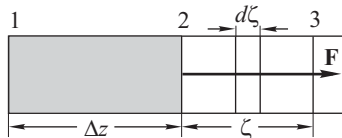


Рис. 5.5

$$dA = F d\zeta = S\Delta z \varepsilon d\varepsilon.$$

Полная работа (при смещении правой границы на величину ζ , которой соответствует конечная величина деформации ε), есть

$$A = \int_0^{\zeta} F d\zeta = \int_0^{\varepsilon} S\Delta z E\varepsilon' d\varepsilon' = S\Delta z \frac{E\varepsilon^2}{2} = V \frac{E\varepsilon^2}{2},$$

где V — объем кусочка стержня $V = S\Delta z$.

Работа внешних сил равна запасенной участком стержня потенциальной (упругой) энергии. На единицу объема приходится энергия

$$W_{\text{п}} = \frac{E\varepsilon^2}{2}, \quad (5.15)$$

которая называется *объемной плотностью потенциальной энергии*. Общая плотность энергии (кинетическая плюс потенциальная) есть

$$W = \frac{1}{2} (\rho u^2 + E\varepsilon^2). \quad (5.16)$$

В бегущей волне $\zeta(z \pm vt)$ плотность энергии в данный момент времени равна нулю там, где максимально смещение ζ , так как при этом $u = \pm v\zeta' = 0$ и $\varepsilon = E\zeta' = 0$ (см. (5.7) и (5.8)). Плотность энергии максимальна там, где и деформация ε , и скорость u максимальны (причем ε и u максимальны в каждый момент времени в одном и том же сечении z , что ясно из тех же соотношений (5.7) и (5.8)). В бегущей гармонической волне $\zeta = a \cos(\omega t \pm kz)$ и $\zeta' = -a \sin(\omega t \pm kz)$ нулевым значениям

смещения ζ соответствует максимум ζ' и, следовательно, максимумы u и ε , а значит и максимум плотности энергии.

Поток энергии. Вектор Умова. Найдем изменение энергии в куске стержня между сечениями 1 и 2 при распространении звуковой волны (рис. 5.3). Слева (в сечении 1) действует сила $F_1 = -\sigma_1 S$. Напомним, что σ положительно, если правая от сечения часть стержня тянет левую вправо, в положительном направлении оси z , при этом часть стержня, находящаяся левее сечения 1, действует на выделенный участок с силой, направленной влево, т. е. должна быть отрицательной. Справа (в сечении 2) действует внешняя сила $F_2 = \sigma_2 S$ (при σ_2 положительном $F_2 > 0$). Работа сил F_1 и F_2 за время dt равна

$$F_1 u_1 dt + F_2 u_2 dt = -(\sigma_1 u_1 - \sigma_2 u_2) S dt.$$

Эта работа равна изменению энергии dW куска стержня за время dt

$$\frac{dW}{dt} = -(\sigma_1 u_1 S - \sigma_2 u_2 S). \quad (5.17)$$

Введем *плотность потока энергии*, $q = -\sigma u$. Тогда $-qS$ — поток энергии через площадку S и формула (5.17) выглядит так:

$$\frac{dW}{dt} = (q_1 - q_2) S. \quad (5.18)$$

Равенству (5.18) можно дать следующее толкование: увеличение энергии dW в куске стержня равно разности между потоком энергии $q_1 = -\sigma_1 u_1 S$, проникшем в кусок стержня через сечение 1 и потоком $q_2 = \sigma_2 u_2 S$, вытекшим из куска стержня через сечение 2. Векторная величина

$$\mathbf{q} = -\sigma u \mathbf{n}, \quad (5.19)$$

где \mathbf{n} — единичный вектор, направление которого совпадает с направлением распространения бегущей волны (осью z), называется *вектором потока энергии*. Этот вектор впервые введен русским ученым Н.А. Умовым (1846–1915), и его называют также *вектором Умова*. Это аналог введенного в предыдущем разделе вектора Пойтинга — потока электромагнитной энергии.

Аналогичным образом, поток энергии в акустической волне в газе или жидкости

$$\mathbf{q} = \Delta p u \mathbf{n}. \quad (5.20)$$

Среднюю величину потока энергии, переносимого акустической волной в единицу времени через единичную площадку (в направлении нормали) называют *интенсивностью* или *силой звука*. Для гармонической волны ($\zeta(z, t) = a \cos(\omega t - kz)$, a — амплитуда смещений частиц) находим

$$u = \frac{d\zeta}{dt} = -a\omega \sin(\omega t - kz); \quad \varepsilon = \frac{d\zeta}{dz} = ak \sin(\omega t - kz),$$

где $k = \omega/v$ — волновое число. Соответствующие волны напряжений $\sigma(z, t)$ или акустических давлений Δp можно найти с помощью (5.2) и (5.11):

$$\sigma = E\varepsilon = \frac{Ea\omega}{v} \sin(\omega t - kz); \quad \Delta p = \rho \left(\frac{dp}{d\rho} \right) \frac{a\omega}{v} \sin(\omega t - kz).$$

Для интенсивности (силы звука) в твердых телах получаем

$$I = \overline{\sigma u} = \frac{a^2 \omega^2 E}{v} \overline{\sin^2(\omega t - kz)} = \frac{1}{2} a^2 \omega^2 \sqrt{\rho E} = \frac{1}{2} a^2 \omega^2 \rho v.$$

Учитывая, что $v = \sqrt{dp/d\rho}$, интенсивность звука в жидкостях и газах есть

$$I = \overline{\Delta p u} = \frac{1}{2} a^2 \omega^2 \rho \sqrt{\frac{dp}{d\rho}} = \frac{1}{2} a^2 \omega^2 \rho v.$$

Важно обратить внимание, что интенсивность пропорциональна *квадрату амплитуды смещений*.

Из полученных соотношений легко найти связь между амплитудой акустического давления и интенсивностью

$$\Delta p_0 = \sqrt{2\rho v I}. \quad (5.21)$$

Интересно отметить, что *порог слышимости* (минимальная слышимая интенсивность звука) $I_0 \approx 10^{-12}$ Вт/м² близок к предельно возможному: кинетическая энергия барабанной перепонки, колеблющейся под действием теплового движения молекул воздуха примерно того же порядка, что и под действием пороговой акустической волны. Диапазон регистрируемых ухом человека интенсивностей огромен: «болевой порог» составляет примерно $10^{12} I_0$ (такова сила звука на концерте рок-музыки в первых рядах). В силу столь большого диапазона удобно выражать интенсивность звука в *децибеллах* (обозначается через β):

$$\beta = 10 \lg \frac{I}{I_0} [\text{дБ}].$$

Величина $10^{10} I_0$ составляет $\beta = 100$ дБ, порог слышимости $\beta_0 = 0$, а болевой порог $\beta = 120$ дБ.

Акустические волны большой интенсивности создают в среде большие перепады давлений. Амплитуда акустического давления Δp_0 в волне с $I \approx 10^5$ Вт/м² составляет в воде, согласно (5.21), $\Delta p_0 \approx 6$ атм, а большие градиенты акустического давления, возникающие при распространении ультразвуковых волн, могут приводить к разрывам жидкости (т. е. к образованию пустот). Это явление называется *кавитацией*. В приведенном примере при частоте ультразвука $\nu = 10^6$ Гц ($\lambda = 5 \cdot 10^{-4}$ м) градиент акустического давления равен $2,40 \cdot 10^4$ атм/м.

5.4. Стержень, закрепленный на концах. Собственные моды колебаний

Рассмотрим бесконечно длинный стержень, по которому бегут две встречные гармонические волны одинаковой частоты и амплитуды:

$$\zeta_1 = a \cos(\omega t - kz); \quad \zeta_2 = a \cos(\omega t + kz + \pi).$$

Мы попрежнему говорим об упругих продольных волнах. Начальная фаза в волне ζ_2 выбрана так, чтобы в сечении $z = 0$ колебания встречных волн происходили в противофазе, т. е. так, чтобы смещение $\zeta = \zeta_1 + \zeta_2$ частиц стержня в этом сечении в любой момент времени равнялось нулю. Суммарная стоячая волна $\zeta(z, t) = \zeta_1(z, t) + \zeta_2(z, t)$ имеет вид

$$\zeta(z, t) = 2a \sin kz \sin \omega t. \quad (5.22)$$

Выделим мысленно кусок стержня длины L между сечениями $z = 0$ и $z = L$ и подберем частоту ω встречных волн так, чтобы смещение ζ в сечении $z = L$ (также, как в сечении $z = 0$) было равным нулю в любой момент времени. Как ясно из (5.22), для этого необходимо, чтобы $\sin kL = 0$, т. е. $kL = n\pi$ (n — целое число). Так как $k = 2\pi/\lambda$, то последнее условие можно записать в виде

$$L = n \frac{\lambda}{2}. \quad (5.23)$$

Или, поскольку частота ω и длина волны λ связаны соотношением $\omega\lambda = 2\pi v$, то

$$\omega_n = \frac{\pi n v}{L} \quad (n = 1, 2, \dots) \quad \text{или} \quad \nu_n = \frac{nv}{2L}. \quad (5.24)$$

Используя (5.6), можно последнее соотношение записать в виде

$$\omega_n = \frac{n\pi}{L} \sqrt{\frac{E}{\rho}} \quad (n = 1, 2, \dots). \quad (5.25)$$

Итак, имеется дискретный набор частот ω_n , при которых сечения $z = 0$ и $z = L$ остаются неподвижными в любой момент времени (узлы стоячей волны (5.22)). Соответствующие частотам ω_n стоячие волны имеют вид (поскольку $k = 2\pi/\lambda = n\pi/L$)

$$\zeta_n(z, t) = 2a \sin \left(n \frac{\pi}{L} z \right) \sin \omega_n t. \quad (5.26)$$

Мы выяснили ранее, что поток энергии в стоячей волне равен тождественно нулю (в любой момент времени) в тех сечениях z , где смещение равно нулю. Следовательно, выделенный нами кусок стержня длины L не обменивается энергией с внешними частями, расположенными левее (при $z < 0$) и правее (при $z > L$).

Если теперь в некоторый момент времени закрепить «намертво» сечения $z = 0$ и $z = L$, то волновое движение, описываемое формулой (5.26), не изменится, поскольку и до закрепления эти сечения были неподвижны. Можно теперь обрезать внешние по отношению к этому куску участки стержня, стоячая волна на оставшемся куске при этом никак не изменится, кусок стержня будет колебаться в согласии с уравнением (5.26), не отдавая и не получая энергии извне. Разумеется, для возбуждения этого колебания никто не берет бесконечно длинный стержень, возбуждая в нем бегущие встречные волны, а затем обрезаая в нужном месте кусок длины L . На самом деле поступают наоборот: берут стержень нужной длины L , закрепленный жестко на концах, а затем возбуждают в нем колебания одной из возможных частот ω_n .

Функции (5.26) определяют возможные типы продольных упругих волн в стержне с закрепленными концами. Соответствующие частоты этих волн ω_n даются формулами (5.24) или (5.25). Гармонические стоячие волны $\zeta_n(z, t)$ называются *собственными типами* или *собственными модами* (или просто *модами*). Не отдавая энергию в местах закрепления (поскольку поток энергии в этих местах равен нулю), такие колебания могут самоподдерживаться как свободные колебания гармонического осциллятора без затухания. Мода, отвечающая индексу $n = 1$, называется *основной*.

Распределение амплитуды смещений по координате z :

$$A_n(z) = 2a \sin n \frac{\pi}{L} z$$

определяет пространственную структуру разных типов колебаний. Эти распределения для $n = 1, 2, 3$ показаны на рис. 5.6 а, б, в.

Еще раз обратим внимание на важное обстоятельство: если в бесконечном стержне возможно создать стоячую волну *любой частоты* ω , то в стержне конечной длины L с закрепленными концами возможны лишь определенные, дискретные конфигурации стоячих волн с дискретным набором частот ω_n , поскольку только такие типы стоячих волн могут *удовлетворять граничным условиям* $\zeta(0) = \zeta(L) \equiv 0$.

Вывод, к которому мы пришли, носит достаточно общий характер: как только мы пытаемся «запереть» волновое движение в конечной области пространства, тут же возникает *дискретность* возможных частот, дискретность возможных конфигураций волн и дискретность возможных значений энергии колебаний.

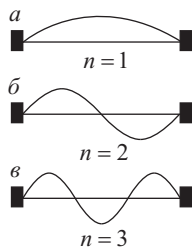


Рис. 5.6

Возможным типам стоячих волн смещений $\zeta_n(z, t)$ (5.26) соответствуют стоячие волны скорости смещений $u_n(z, t) = \partial\zeta_n/\partial t$ и деформаций $\varepsilon_n(z, t) = \partial\zeta_n/\partial z$. Мы получаем

$$\begin{aligned} u_n(z, t) &= 2a\omega_n \sin \frac{n\pi z}{L} \cos \omega_n t, \\ \varepsilon_n(z, t) &= 2 \frac{\pi n}{L} a \cos \frac{n\pi z}{L} \sin \omega_n t. \end{aligned} \quad (5.27)$$

Рассмотрим энергетические превращения при колебаниях стержня.

В процессе колебаний происходит обмен между кинетической и потенциальной частями полной энергии (последняя, как мы выяснили, остается неизменной). Как видно из формул (5.27), дважды за период скорость смещений $u_n(z, t)$ и деформация стержня $\varepsilon_n(z, t)$ становятся равными нулю в любом сечении стержня.

Следовательно, дважды за период становится равной нулю кинетическая энергия ($u_n(z, t) = 0$ при $t = T/4$ и $t = 3T/4$) — и в эти моменты времени полная энергия колебаний — это потенциальная энергия упругой деформации:

$$W_{\text{п}} = \frac{1}{2} \int_0^L E \varepsilon^2(z) dz, \quad (5.28)$$

где $\varepsilon(z) = \pm(2\pi n/L) a \cos(n\pi z/L)$ ($\sin \omega_n t = 1$ при $t = T/4$ и $t = 3T/4$).

Аналогично, дважды за период обращается в нуль потенциальная энергия ($\varepsilon_n(z, t) = 0$ при $t = 0$ и $t = T/2$). В эти моменты времени энергия колебаний — это кинетическая энергия, равная

$$W_{\text{к}} = \frac{1}{2} \int_0^L \rho u_n^2 dz, \quad (5.29)$$

где $u_n(z) = \pm 2a\omega_n \sin(n\pi z/L)$. Поскольку полная энергия колебаний постоянна, то любую из формул (5.28) или (5.29) можно использовать для подсчета ее конкретного значения. В частности, из (5.29) легко получить

$$W_{\text{к}} = a^2 \omega_n^2 V, \quad (5.30)$$

где V — объем стержня, ω_n определяется формулой (5.25). Основной тип колебаний имеет наименьшую энергию, поскольку ему соответствует наименьшая частота. Энергия n -го типа пропорциональна n^2 . Формула (5.30) определяет возможные дискретные значения энергии колебаний.

5.5. Поведение звука на границе раздела двух сред

Рассмотрим плоскую звуковую волну

$$\zeta_1 = a_1 \cos(\omega t - k_1 z), \quad (5.31)$$

распространяющуюся в однородной среде I в области $z < 0$ (рис. 5.7). Волна распространяется слева направо, в положительном направлении оси z , $k_1 = \omega/v_1$ — волновое число, v_1 — скорость звука в среде I. Плоскость $z = 0$ является *границей раздела двух сред*. Область $z > 0$ занята средой II, скорость звука в которой $v_2 \neq v_1$ (напомним, что скорости v_1 и v_2 зависят от упругих свойств и плотности сред (см. (5.6))).

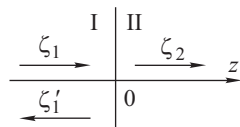


Рис. 5.7

Что происходит с волной $\zeta_1(z, t)$, падающей на границу раздела? В соответствии с тем, что подсказывает нам опыт, предположим, что возникает волна, прошедшая во вторую среду

$$\zeta_2(z, t) = a_2 \cos(\omega t - k_2 z), \quad (5.32)$$

где $k_2 = \omega/v_2$, а в первой среде, кроме волны $\zeta_1(z, t)$, бегущей слева направо и падающей на границу раздела, возникает *отраженная волна*

$$\zeta_1'(z, t) = a_1' \cos(\omega t + k_1 z). \quad (5.33)$$

Обратите внимание на знак «+» в аргументе косинуса: отраженная волна бежит справа налево; кроме того, волновое число этой волны то же, что и в падающей волне, так как обе они распространяются в среде I. Суммарная волна в среде I есть

$$\zeta_1(z, t) + \zeta_1'(z, t) = a_1 \cos(\omega t - k_1 z) + a_1' \cos(\omega t + k_1 z). \quad (5.34)$$

Наша задача: найти амплитуды отраженной волны a_1' и волны, прошедшей во вторую среду a_2 , зная амплитуду падающей волны a_1 .

Воспользуемся условиями на границе $z = 0$. Во-первых, одно и то же смещение границы $\zeta(t)$ можно выразить как через волновое возмущение (5.34) в первой среде, так и через волновое возмущение (5.32) во второй среде; равенство волновых возмущений при $z = 0$ есть:

$$\zeta_1(0, t) + \zeta_1'(0, t) = \zeta_2(0, t). \quad (5.35)$$

Во-вторых, упругая сила (на единицу площади), действующая со стороны первой среды на вторую, равна по величине силе, действующей со стороны второй среды на первую:

$$\sigma_1(0, t) + \sigma_1'(0, t) = \sigma_2(0, t), \quad (5.36)$$

где $\sigma = E(\partial\zeta/\partial z)$. Подставляя выражения (5.33) и (5.34) (при $t = 0$) в граничное условие (5.35), получаем (сокращая на общий множитель $\cos \omega t$)

$$a_1 + a'_1 = a_2. \quad (5.37)$$

Аналогичным образом, дифференцируя функции (5.33) и (5.34) по z и используя граничное условие (5.36), находим (сокращая на $\sin \omega t$)

$$E_1 k_1 (a_1 - a'_1) = E_2 k_2 a_2. \quad (5.38)$$

Из (5.37) и (5.38) имеем выражения, которые связывают амплитуды отраженной и прошедшей волны с амплитудой падающей волны

$$a'_1 = \frac{1 - \gamma}{1 + \gamma} a_1, \quad a_2 = \frac{2}{1 + \gamma} a_1. \quad (5.39)$$

Мы использовали обозначение $\gamma = \frac{E_2 k_2}{E_1 k_1} = \frac{\sqrt{\rho_2 E_2}}{\sqrt{\rho_1 E_1}}$.

Используя выражения для среднего потока энергии \bar{q} в упругой среде за период колебаний (см. (5.19)), можно найти величины $R = \bar{q}'_1 / \bar{q}_1$ (коэффициент отражения) и $T = \bar{q}_2 / \bar{q}'_1$ (коэффициент прозрачности):

$$R = \left(\frac{1 - \gamma}{1 + \gamma} \right)^2, \quad T = \frac{4\gamma}{(1 + \gamma)^2}. \quad (5.40)$$

5.6. Поперечные волны в струне

Рассмотрим натянутую струну, ось которой совпадает в равновесии с осью z . Рассмотрим распространение поперечных упругих волн в струне, причем будем рассматривать смещения, происходящие в плоскости yz .

Пусть струна выведена из положения равновесия; смещение куска струны Δz (между точками z и $z + \Delta z$) в данный момент времени t есть $\eta(z, t)$ (рис 5.8). Будем полагать далее, что натяжение струны (равное при равновесии σ) столь велико, что при *малом смещении* из положения равновесия оно остается практически неизменным и равным по-прежнему σ в любом сечении:

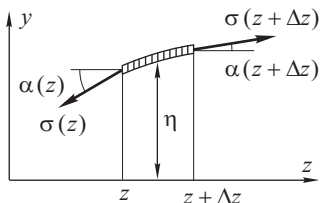


Рис. 5.8

$$\sigma(z) = \sigma(z + \Delta z) = \sigma.$$

Напомним, что σ — сила, действующая на единицу площади выделенного участка струны со стороны участков, прилегающих слева и справа.

Уравнение движения (второй закон Ньютона) для куска струны Δz , масса которого равна $\rho S \Delta z$, имеет (в проекции на ось y) вид

$$\rho S \Delta z \frac{d^2 \eta}{dt^2} = \sigma S [\sin \alpha(z + \Delta z) - \sin \alpha(z)],$$

где $\alpha(z)$ — угол между направлением касательной к струне в точке z и осью z : $\operatorname{tg} \alpha = d\eta/dz$. Полагая (для малых отклонений), что $\sin \alpha \approx \operatorname{tg} \alpha = d\eta/dz$, получаем $\rho \Delta z (d^2 \eta / dt^2) = \sigma ((d\eta/dz)_{z+\Delta z} - (d\eta/dz)_z)$. Разделив обе части последнего равенства на Δz и перейдя к пределу при $\Delta z \rightarrow 0$, получаем волновое уравнение

$$\frac{\partial^2 \eta}{\partial z^2} = \frac{\rho}{\sigma} \frac{\partial^2 \eta}{\partial t^2}. \quad (5.41)$$

Согласно (5.41) скорость распространения упругих волн в струне

$$v = \sqrt{\frac{\sigma}{\rho}}. \quad (5.42)$$

Возможными решениями уравнения являются волны вида

$$\eta(z, t) = A \cos(\omega t \pm kz), \quad (5.43)$$

бегущие слева направо (знак « $-$ » в (5.43)) и справа налево (знак « $+$ » в (5.43)), а также их суперпозиция — *стоячая волна*

$$\eta(z, t) = A \cos kz \cos \omega t.$$

Понятно, что мы имеем дело с *поперечными волнами*, поскольку, согласно (5.43), волны распространяются вдоль оси z , а смещения частиц струны происходят в направлении y , т. е. перпендикулярно направлению распространения (напомним, что в случае стержня речь шла о продольных волнах). Волновое число k в (5.43) есть: $k = \omega/v = \omega \sqrt{\rho/\sigma}$.

Рассуждения, которые мы провели, исследуя продольные упругие волны в стержне, закрепленном на концах, полностью применимы к натянутой струне. В частности, возможные типы колебаний описываются теми же формулами: возможны лишь такие стоячие волны, для которых на струне укладывается целое число полуволн:

$$L = n \frac{\lambda}{2}.$$

Соответствующие частоты колебаний

$$\omega_n = \frac{n\pi v}{L} = \frac{n\pi}{L} \sqrt{\frac{\sigma}{\rho}} \quad \text{или} \quad \nu_n = \frac{nv}{2L} = \frac{n\sqrt{\sigma/\rho}}{L}.$$

Возможные типы стоячих волн определяются формулой

$$\eta_n(z, t) = A \cos n \left(\frac{\pi}{L} \right) z \cdot \cos \omega_n t. \quad (5.44)$$

5.7. Общие выводы

Мы рассмотрели несколько физических систем, в которых возбуждаются упругие волны: продольные упругие волны в твердых телах, жидкостях и газах, а также поперечные волны в натянутой струне. Во всех случаях мы пришли к выводу, что возникшее в среде возмущение распространяется в виде волны, т. е. удовлетворяет одному и тому же уравнению (4.4), в котором константа v , имеющая смысл скорости распространения волны, определяется физическими параметрами *невозмущенной системы*: модулем упругости E и плотностью в случае твердого тела, величиной $dp/d\rho$ в равновесном состоянии в случае газов, натяжением при равновесии σ и плотностью в случае поперечных колебаний струны. Этот результат обусловлен тем, что при выводе волнового уравнения мы каждый раз делали *предположение о малости возмущений*: исследуя продольные волны в стержне, мы предполагали, что справедлив закон Гука — линейная связь между напряжением σ и величиной деформации ε — закон, справедливый лишь при *малых деформациях*. При изучении волн в газах мы полагали, что имеет место линейная связь между изменением давления и изменением плотности газа, а это справедливо лишь при малых изменениях давления (малых в сравнении с давлением в газе в стационарных условиях). При выводе волнового уравнения для натянутой струны мы также предположили, что струна мало отклоняется от положения равновесия, поэтому неизменным (и равным стационарному значению) остается величина натяжения струны при колебаниях.

Именно благодаря предположению о малости возникающих возмущений, мы каждый раз приходили к *линейному уравнению*, для которого справедлив принцип суперпозиции: если в системе возможны волны $S_1(z, t)$ и $S_2(z, t)$, то возможна и волна вида

$$S(z, t) = a_1 S_1(z, t) + a_2 S_2(z, t),$$

где a_1 и a_2 — произвольные постоянные. Другими словами, если S_1 и S_2 есть решения уравнения (4.4), то и функция $S(z, t)$, являющаяся линейной комбинацией решений S_1 и S_2 , также есть решение. В частности, мы выяснили, что стоячие волны $\zeta_n(z, t)$, описываемые соотношением (5.26), являются собственными типами продольных упругих волн (модами) для стержня длины L , закрепленного на концах. Следовательно, в таком стержне возможно и волновое возмущение вида

$$\zeta(z, t) = \sum_n a_n \zeta_n(z, t) \quad (5.45)$$

при произвольных постоянных a_n . Ясно, что, во-первых, функ-

ция (5.45) является решением волнового уравнения (5.4) и, во-вторых, подчиняется граничным условиям на концах

$$\zeta(z, t) \equiv 0 \quad \text{при} \quad z = 0 \quad \text{и} \quad z = L.$$

Последнее очевидно, поскольку каждое слагаемое в (5.45) обращается в нуль на закрепленных концах стержня.

Более того, как доказывают математики, любая возможная волна в стержне с закрепленными концами может быть представлена (при правильном подборе коэффициентов a_n) линейной суперпозицией собственных типов волн, т. е. в виде (5.45). Конкретный вид волны, т. е. конкретный набор коэффициентов зависит от начальных условий — от способа создания волнового возмущения.

Уравнение, описывающее волновой процесс, перестает быть линейным, если возмущение не является малым: изменение давления при распространении волны в газе не мало по сравнению со стационарным значением p_0 ; деформация твердого тела при распространении упругой волны велика настолько, что закон Гука нарушается; отклонение струны при колебаниях столь велико, что натяжение изменяется. При этом нарушается принцип суперпозиции. Физически это связано с тем, что волна большой амплитуды изменяет свойства среды, в которой она распространяется. В конечном счете это приводит к тому, что параметр v , входящий в волновое уравнение, уже не определяется только характеристиками невозмущенной среды, он начинает зависеть от амплитуды волны, если амплитуда не мала. Ясно, что волновое уравнение (4.4), в котором v есть функция возмущения S ($v = v(S)$), нелинейно, что и приводит к нарушению принципа суперпозиции.

Одно из проявлений нелинейности — изменение формы волнового возмущения по мере распространения волны. Дело в том, что при звуковых возмущениях волновые поверхности, соответствующие местам сильных (нелинейных) сжатий упругой среды, бегут с большей скоростью, чем поверхности, в которых среда разрежена, т. е. горбы деформаций бегут быстрее «впадин», что и приводит к изменению формы волны, к увеличению крутизны волновых фронтов, отвечающих участкам сильных сжатий.

Первоначально синусоидальная упругая волна большой амплитуды (рис. 5.9а) становится по мере распространения в среде несинусоидальной (рис. 5.9б и в), а следовательно, немонохроматической: возникают волны различных частот из первоначального гармонического колебания.

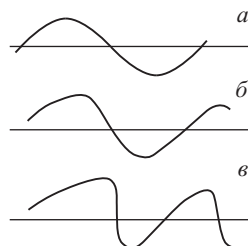


Рис. 5.9

Примером нарушения линейности является *ударная волна*, возникающая в упругой среде при взрыве или при движении в ней тела со сверхзвуковой скоростью. Летящий со сверхзвуковой скоростью самолет приводит в движение слой газа перед собой, который также движется со сверхзвуковой скоростью. Возникший поток газа сжимает впереди лежащие слои и, поскольку это сжатие не успевает столь же быстро передаваться впереди лежащим слоям «естественным» образом за счет теплового движения (такая «естественная» передача осуществляется со звуковой скоростью), то возникает слой с большим перепадом давления, плотности и температуры. Условие $\Delta p \ll p_0$ нарушается. Тонкий слой, ширина которого порядка длины свободного пробега молекул, называется *скачком уплотнения*, а распространяющаяся волна — *ударной волной*. Наконец, для некоторых нелинейных систем деформации поверхности согласуются со скоростью волнового возбуждения таким образом, что возникает уединенная нелинейная волна — *солитон*.

5.8. Эффект Доплера

Пусть в однородной и изотропной упругой среде находится маленький пульсирующий шарик, радиус которого меняется по гармоническому закону с частотой ω_0 . Такой шарик является источником акустической (упругой) волны. Сферические волновые поверхности (поверхности сжатий и разрежений) бегут от источника со скоростью звука c . Если источник не перемещается, то центры всех сфер, всех волновых поверхностей, бегущих от источника одна за другой, находятся в одном месте, в точке S , где находится источник.

Пусть теперь источник движется вдоль оси z со скоростью v (рис. 5.10). В момент времени $t = 0$, когда источник находится в точке S_1 , от источника начинает бежать волновая поверхность,

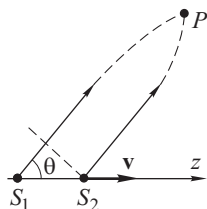


Рис. 5.10

например, поверхность максимального сжатия, которая достигает приемника, находящегося в удаленной точке наблюдения P , в момент времени $\tau_1 = S_1 P/c$. Следующая волновая поверхность максимального сжатия излучается через время T_0 (через период колебания), когда источник уже находится в точке S_2 , пройдя за это время расстояние

$$S_1 S_2 = v T_0.$$

Положение приемника определяется углом θ — направлением на точку наблюдения, которая, как мы полагаем, находится столь

далеко, что прямые S_1P и S_2P можно считать параллельными, поэтому $S_1P - S_2P = S_1S_2 \cos \theta = vT_0 \cos \theta$.

Когда очередная поверхность максимального сжатия (излученная из точки S_2) достигнет приемника? Очевидно, в момент времени

$$\tau_2 = T_0 + \frac{S_2P}{c},$$

где S_2P/c — время распространения возмущения из точки S_2 до точки P .

Интервал времени между моментами прихода двух последовательных волновых поверхностей одинаковой фазы колебания — это и есть период колебаний T , регистрируемых приемником

$$T = \tau_2 - \tau_1 = T_0 + \frac{S_2P - S_1P}{c} = T_0 \left(1 - \frac{v}{c} \cos \theta \right).$$

Из последнего равенства получаем окончательное выражение, связывающее частоту излучения, регистрируемого приемником $\omega = 2\pi/T$, с частотой колебаний источника $\omega_0 = 2\pi/T_0$:

$$\omega = \frac{\omega_0}{1 - (v/c) \cos \theta}. \quad (5.46)$$

При $v/c \ll 1$ из (5.46) следует приближенное выражение

$$\omega \approx \omega_0 \left(1 + \frac{v}{c} \cos \theta \right). \quad (5.47)$$

Пусть теперь источник неподвижен, а приемник движется в направлении, образующем угол θ с прямой, соединяющей источник и приемник. Пусть в момент времени $t = 0$ приемник находится в точке P_1 на одной из волновых поверхностей, бегущих от источника со скоростью c (пусть, например, это поверхность максимального сжатия A_0A_0 , рис. 5.11). Через отрезок времени t приемник, двигаясь со скоростью v из точки P_1 , встретит следующую поверхность максимального сжатия — поверхность BB в точке P_2 . (В момент $t = 0$ поверхность BB находилась в положении B_0B_0 на расстоянии $\lambda_0 = cT_0$ от поверхности A_0A_0 , и за время t переместилась в положение BB , пройдя расстояние ct .)

Время t между моментами регистрации приемником двух последовательных поверхностей максимального сжатия — это и есть период колебаний, регистрируемых движущимся приемником.

Мы имеем

$$P_1P_2 = vT; \quad DP_1 = c(T_0 - T)$$

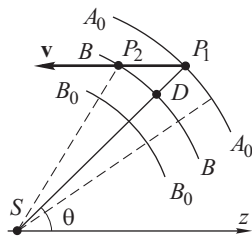


Рис. 5.11

и из $\Delta P_1 P_2 D$

$$vT \cos \theta = c(T_0 - T).$$

Окончательно получаем (так как $T = 2\pi/\omega$, $T_0 = 2\pi/\omega_0$):

$$\omega = \omega_0 \left(1 + \frac{v}{c} \cos \theta\right). \quad (5.48)$$

Изменение частоты колебаний, регистрируемых неподвижным приемником при движении источника (5.46), либо изменение частоты, регистрируемой движущимся приемником при неподвижном источнике (5.48), называется *эффектом Доплера* (по имени австрийского физика и астронома К. Доплера (1803–1853)).

Как следует из сравнения этих формул, не безразлично, что движется относительно среды — источник или приемник, однако при малых скоростях $v \ll c$ обе формулы дают одинаковый результат.

Задачи

1. По заданному графику мгновенного распределения смещений $\zeta(z)$ по длине стержня (рис. 5.12а) постройте график скоростей $u(z)$ и деформаций $\varepsilon(z)$ для волны, бегущей слева направо и справа налево.

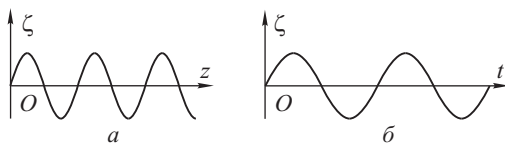


Рис. 5.12

2. По заданной осциллограмме смещений $\zeta(t)$ (в фиксированной плоскости z , рис. 5.12б) постройте осциллограммы скорости $u(t)$ и деформации (напряжения) в волне, бегущей слева направо.

3. По стержню распространяются две встречные волны одинаковой частоты и амплитуды. Найдите результирующую волну смещений $\zeta(z, t)$, скоростей $u(z, t)$ и деформаций $\varepsilon(z, t)$. Изобразите графики $u(z)$ и $\varepsilon(z)$ при фиксированном t .

4. Изобразите моментальные снимки распределения деформации $\varepsilon(z)$ и скорости $u(z)$ в волне сжатия, бегущей по стержню, если соответствующий снимок смещений $\zeta(z)$ имеет вид, показанный на рис. 5.13а.

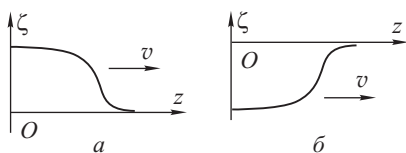


Рис. 5.13

5. Решите ту же задачу для волны разрежения (рис. 5.13б).

6. Найдите изменение скорости звука в воздухе при изменении температуры от -10°C до $+20^\circ\text{C}$.

7. Вычислите скорость звука в стальном стержне.

8. Выведите формулу (5.13) для скорости звука в идеальном газе.

9. Найдите мгновенное распределение потока энергии в бегущей волне $\zeta(z, t) = a \cos(\omega t - kz)$ как функции координаты z ; постройте график зави-

симости $u(z)$, $\sigma(z)$ и $P(z)$. Найдите точки максимумов и минимумов потока энергии.

10. Покажите, что средний за период колебания поток энергии в бегущей волне $\zeta = a \cos(\omega t - kz)$ одинаков через любое сечение z и равен

$$\bar{q} = \frac{1}{T} \int_0^T q(z, t) dt = \frac{1}{2} E k \omega a^2. \quad (5.49)$$

11. Найдите поток энергии, протекающий через единичное сечение за один период колебания.

Ответ: $\bar{q}T = \pi k a^2 E$.

12. Найдите скорость смещения u и деформацию ε как функции z и t в стоячей упругой волне.

13. Найдите поток энергии в стоячей волне как функции координаты z и времени t ; найдите расстояние между сечениями z , поток энергии через которые равен нулю в любой момент времени. Найдите полную энергию стоячей волны между двумя ближайшими сечениями, поток энергии через которые равен нулю. Каков закон изменения во времени кинетической и потенциальной энергии, заключенной между ближайшими сечениями, поток энергии через которые равен нулю.

14. Постройте графики мгновенных распределений скоростей $u_n(z, t)$ вдоль оси стержня в моменты времени $t = 0$, $t = T/4$, $t = 3T/4$ для типов колебаний с индексами $n = 1$, $n = 2$, $n = 3$.

15. Постройте графики мгновенных распределений деформаций $\varepsilon_n(z, t)$ вдоль оси стержня в моменты времени $t = 0$, $t = T/4$ и $t = 3T/4$ для тех же типов колебаний.

16. Каковы осциллограммы (зависимости от времени) скоростей $u_n(t)$ и деформаций $\varepsilon_n(t)$ в фиксированном сечении $z = L/4$ для типов колебаний $n = 1, 2, 3$.

17. Постройте осциллограммы скоростей и деформаций в сечениях $z = L/4$ и $z = 3L/4$ для типов колебаний $n = 1, 2, 3$.

18. Показать, что сумма потоков энергии в отраженной и прошедшей волнах равна потоку энергии в падающей волне.

19. Показать, что при $\rho_1 E_1 = \rho_2 E_2$, $\gamma = 1$ и $R = 0$, $T = 1$, т. е. вся энергия проходит во вторую среду — отражения нет.

20. Показать, что при $\gamma = 0$ и $\gamma = \infty$, $R = 1$, $T = 0$, т. е. нет прошедшей волны — вся энергия возвращается в первую среду.

21. Показать, что коэффициенты отражения и прохождения не зависят от того, с какой стороны волна падает на границу раздела двух сред.

22. Используя (5.39), покажите, что a_2 всегда имеет тот же знак, что и a_1 , т. е. колебания в прошедшей волне на границе $z = 0$ происходят *синфазно* с колебаниями в падающей волне.

23. Показать, что на границе $z = 0$ колебания в *отраженной волне* происходят *синфазно* с колебаниями в падающей волне (a'_1 и a_1 имеют одинаковый знак), если $\gamma < 1$. Если же $\gamma = v_1/v_2 > 1$, то a'_1 и a_1 имеют разные знаки, т. е. колебания в отраженной волне происходят в *противофазе* с колебаниями в падающей волне.

24. Как зависит константа γ от плотности ρ и модуля упругости E первой и второй среды?

25. Рассмотрите предельные случаи $\gamma = 0$ и $\gamma = \infty$ (когда нет прошедшей во вторую среду волны и $a'_1 = \pm a_1$). Покажите, что при этом в первой среде возникает стоячая волна с *пучностью* смещений на границе $z = 0$, если $\gamma = 0$, и узлом на той же границе, если $\gamma = \infty$. Какой из этих случаев реализуется, когда звуковая волна в воздухе падает на границу с металлом?

Глава 6

ЭЛЕКТРОМАГНИТНЫЕ ВОЛНЫ

Мы живем в мире электромагнитных волн. Радиоволны и волны видимого диапазона, инфракрасные и ультрафиолетовые волны, рентгеновское и γ -излучение — все это *электромагнитные волны* различного диапазона частот. Особое место в этом многообразии занимают волны видимого света. Хотя на шкале электромагнитных волн они занимают не слишком много места (рис. 6.1), именно с помощью света мы получаем подавляющее

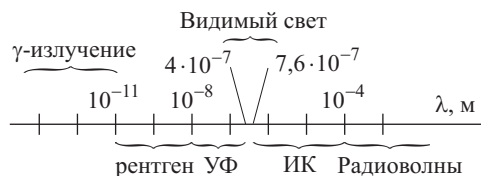


Рис. 6.1

количество информации об окружающем мире, о многообразии форм, красок, оттенков, воспринимаемой с помощью тончайшего органа зрения, которым наградила нас природа.

Что же такое электромагнитные волны, какие законы природы ими управляют?

6.1. Уравнения Максвелла и волновое уравнение

Все многообразие электрических и магнитных явлений описывается четырьмя универсальными соотношениями — уравнениями Максвелла (по имени английского физика Дж.К. Максвелла (1831–1879)), которые в электромагнетизме играют столь же фундаментальную (и даже более фундаментальную) роль, как и законы Ньютона в классической механике. Два из этих четырех соотношений связывают между собой электрическое и магнитное поля.

Во-первых, меняющееся во времени магнитное поле \mathbf{B} порождает электрическое поле \mathbf{E} (закон электромагнитной индукции):

$$\oint_{(l)} E_l dl = - \int_{(S)} \dot{B}_n dS. \quad (6.1)$$

В левой части равенства (6.1) — циркуляция вектора \mathbf{E} по произвольному замкнутому контуру l . В правой части — скорость изменения потока вектора \mathbf{B} через поверхность S , на-

тянутую на контур (рис. 6.2), $\dot{\mathbf{B}}$ — производная по времени поля \mathbf{B} .

Во-вторых, переменное электрическое поле \mathbf{D} является источником магнитного поля \mathbf{H} :

$$\oint_{(l)} \mathbf{H}_l dl = \int_{(S)} \dot{\mathbf{D}}_n dS. \quad (6.2)$$

В левой части последнего равенства — циркуляция магнитного поля \mathbf{H} по произвольному замкнутому контуру, в правой части — скорость изменения потока вектора \mathbf{D} через поверхность S ($\dot{\mathbf{D}}$ — производная по времени поля \mathbf{D}).

Так и рождаются электромагнитные волны: если в какой-либо точке пространства возникло меняющееся электрическое поле ($\dot{\mathbf{D}}_n \neq 0$), то в соответствии с (6.2) возникает магнитное поле, которое, изменяясь во времени, порождает в соответствии с (6.1) электрическое поле. Так, взаимно индуцируя друг друга, возникают меняющиеся во времени электрическое и магнитное поля — возникает электромагнитная волна. Почему именно волна? Как это следует из законов (6.1) и (6.2)?

Мы рассмотрим наиболее простую ситуацию: однородную изотропную среду, свойства которой характеризуются двумя параметрами: константами ε (диэлектрическая проницаемость среды) и μ (магнитная проницаемость). Как известно, эти величины связывают между собой поля \mathbf{E} и \mathbf{D} ; \mathbf{B} и \mathbf{H} (так называемые материальные уравнения):

$$\mathbf{D} = \varepsilon_0 \varepsilon \mathbf{E}, \quad \mathbf{B} = \mu_0 \mu \mathbf{H}. \quad (6.3)$$

Напомним, что равенства (6.3) не являются универсальными и справедливы лишь для *линейной среды*, свойства которой не зависят от того, имеются ли в среде поля или их нет, велики эти поля или малы. В рассматриваемой нами среде нет токов, поэтому в правой части уравнения (6.2) отсутствует слагаемое, содержащее токи проводимости. Полагаем также, что среда является изотропной, т.е. свойства среды (ε и μ) не зависят от ориентации в пространстве электрического и магнитного поля \mathbf{E} и \mathbf{B} .

Будем полагать далее, что поля \mathbf{E} и \mathbf{B} (а значит, и \mathbf{D} и \mathbf{H}) зависят лишь от одной координаты, например, координаты z и, разумеется, от времени:

$$\mathbf{E} = \mathbf{E}(z, t), \quad \mathbf{B} = \mathbf{B}(z, t).$$



Рис. 6.2

Это означает, что в любой плоскости $z = \text{const}$ величина поля \mathbf{B} (как и \mathbf{E}) одна и та же в любой фиксированный момент времени во всех точках этой плоскости. Используя (6.3), перепишем уравнение (6.2), введя обозначение $c = 1/\sqrt{\varepsilon_0\mu_0}$, в виде

$$\oint_{(l)} B_l dl = \frac{\varepsilon\mu}{c^2} \int_{(S)} \dot{E}_n dS, \quad (6.4)$$

константы ε и μ вынесены из-под знака интегралов. Применим уравнения (6.1) и (6.2) к контуру $ABCD$, в котором $AB = CD = l$ и $AD = BC = \Delta z$. Контур $ABCD$ лежит в плоскости xz (рис. 6.3; стрелками показано положительное направление обхода). Мы получаем

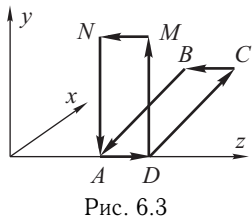


Рис. 6.3

$$\begin{aligned} E_x(z + \Delta z)l - E_x(z)l &= -\dot{B}_y \Delta z l, \\ B_x(z + \Delta z)l - B_x(z)l &= \frac{\varepsilon\mu}{c^2} \dot{E}_y \Delta z l. \end{aligned} \quad (6.5)$$

Вклад в контурный интеграл участков AD и CB бесконечно мал при $\Delta z \rightarrow 0$. Разделив левые и правые части равенств (6.5) на Δz и переходя к пределу при $\Delta z \rightarrow 0$, находим:

$$\frac{\partial E_x}{\partial z} = -\frac{\partial B_y}{\partial t}, \quad \frac{\partial B_x}{\partial z} = \frac{\varepsilon\mu}{c^2} \frac{\partial E_y}{\partial t}. \quad (6.6)$$

Аналогичным образом, применив уравнения (6.1) и (6.2) к контуру $ADMN$ (этот контур лежит в плоскости yz , $AN = DM = l$), получаем

$$\frac{\partial E_y}{\partial z} = \frac{\partial B_x}{\partial t}, \quad \frac{\partial B_y}{\partial z} = -\frac{\varepsilon\mu}{c^2} \frac{\partial E_x}{\partial t}. \quad (6.7)$$

Выпишем отдельно пару уравнений из (6.6) и (6.7), содержащих только поля E_x и B_y :

$$\frac{\partial E_x}{\partial z} = -\frac{\partial B_y}{\partial t}, \quad \frac{\partial B_y}{\partial z} = -\frac{\varepsilon\mu}{c^2} \frac{\partial E_x}{\partial t}. \quad (6.8)$$

Другая пара содержит поля E_y и B_x :

$$\frac{\partial E_y}{\partial z} = \frac{\partial B_x}{\partial t}, \quad \frac{\partial B_x}{\partial z} = \frac{\varepsilon\mu}{c^2} \frac{\partial E_y}{\partial t}. \quad (6.9)$$

Мы видим, что пара полей (E_x, B_y) совершенно не связана с другой парой (E_y, B_x) , т. е. изменяющиеся во времени и пространстве поля E_x и B_y могут существовать совершенно независимо от другой пары E_y и B_x . Рассмотрим одну из этих

пар, например E_x и B_y . Продифференцируем первое из уравнений (6.8) по z , а второе — по t . Мы получим

$$\frac{\partial^2 E_x}{\partial z^2} = -\frac{\partial^2 B_y}{\partial t \partial z}, \quad \frac{\partial^2 B_y}{\partial z \partial t} = -\frac{\varepsilon \mu}{c^2} \frac{\partial^2 E_x}{\partial t^2}.$$

Исключая смешанную производную $\partial^2 B_y / \partial t \partial z = \partial^2 B_y / \partial z \partial t$ из этой пары уравнений, получаем

$$\frac{\partial^2 E_x}{\partial z^2} = \frac{\varepsilon \mu}{c^2} \frac{\partial^2 E_x}{\partial t^2}. \quad (6.10)$$

Мы пришли к важному результату: поле $E_x(z, t)$ подчиняется волновому уравнению, т. е. изменение во времени и пространстве электрического поля E_x происходит по законам распространения волн, в частности, описывается функцией вида

$$E_x(z, t) = E_x(z \pm vt).$$

Эта функция, как мы знаем, описывает волну, распространяющуюся вдоль оси z (в положительном (знак «−»), либо в отрицательном (знак «+») направлении оси z), причем скорость волны v , как следует из (6.10), есть

$$v = \frac{c}{\sqrt{\varepsilon \mu}}. \quad (6.11)$$

Константа c определяет скорость электромагнитной волны в вакууме ($\varepsilon = \mu = 1$) и равна

$$c \approx 2,998 \cdot 10^8 \text{ м/с} \approx 3 \cdot 10^8 \text{ м/с}.$$

Это одна из самых замечательных мировых констант в физике.

Мы могли бы, дифференцируя первое из уравнений (6.8) по t , а второе — по z , получить (исключив из системы смешанную производную $\partial^2 E_x / \partial z \partial t = \partial^2 E_x / \partial t \partial z$) волновое уравнение для магнитного поля B_y :

$$\frac{\partial^2 B_y}{\partial z^2} = \frac{\varepsilon \mu}{c^2} \frac{\partial^2 B_y}{\partial t^2},$$

т. е. точно такое же волновое уравнение, что и (6.10). Этому уравнению подчиняется волновой процесс вида

$$B_y(z, t) = B_y(z \pm vt).$$

Волна B_y бежит вдоль оси z с той же скоростью, что и волна E_x . Эта пара волн (E_x, B_y) представляет собой бегущую вдоль оси z *плоскую электромагнитную волну*, в которой поле \mathbf{E} имеет только x -компоненту, а поле \mathbf{B} имеет только y -компоненту, т. е. колебания полей \mathbf{E} и \mathbf{B} происходят в плоскости, перпендикулярной направлению распространения — оси z . Итак, *плоская*

электромагнитная волна поперечна (и в этой волне колебания полей \mathbf{E} и \mathbf{B} — взаимно перпендикулярны).

Рассмотрев таким же образом пару уравнений (6.9), мы пришли бы к аналогичному выводу: волны $E_y(z \pm vt)$ и $B_x(z \pm vt)$ образуют плоскую электромагнитную волну, бегущую вдоль оси z , в которой неразрывно связаны между собой (взаимно индуцируют друг друга) поля E_y и B_x , эта электромагнитная волна также поперечна.

Вернемся к электромагнитной волне (E_x, B_y) . Пусть изменения E_x в волне описываются некоторой функцией $f(z \pm vt)$: $E_x = f(z \pm vt)$, а изменения B_y — функцией $g(z \pm vt)$: $B_y = g(z \pm vt)$. Используя правило дифференцирования сложной функции, найдем:

$$\frac{\partial E_x}{\partial z} = f', \quad \frac{\partial B_y}{\partial t} = (\pm v)g',$$

где штрих означает дифференцирование по аргументу $(z \pm vt)$. Первое из уравнений (6.8) как раз и связывает между собой производные $\partial E_x / \partial z$ и $\partial B_y / \partial t$:

$$f' = (\pm v)g'.$$

Интегрируя последнее равенство, получаем

$$f = \pm v g + \text{const.}$$

Знак «+» в последнем равенстве относится к волне, бегущей слева направо (т. е. к аргументу $z - vt$), а знак «−» к волне, бегущей справа налево (к аргументу $z + vt$). Произвольная постоянная в последнем равенстве представляет собой некоторое статическое поле, которое нас не интересует, поскольку оно не связано с переменными полями, образующими волну. Полагая поэтому $\text{const} = 0$, находим

$$E_x = \pm v B_y \quad \text{или} \quad E_x = \pm \frac{c}{\sqrt{\epsilon \mu}} B_y = \pm \sqrt{\frac{\mu'}{\epsilon'}} H_y, \quad (6.12)$$

где $\mu' = \mu \mu_0$, $\epsilon' = \epsilon \epsilon_0$.

Итак, *в бегущей плоской электромагнитной волне поля \mathbf{E} и \mathbf{B} в каждый момент времени в каждой точке пространства пропорциональны друг другу.*

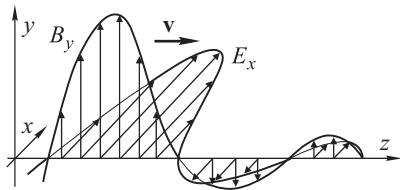


Рис. 6.4

На рис. 6.4 показана «мгновенная фотография» — распределение полей \mathbf{E} и \mathbf{B} как функции координаты z в некоторый момент времени t для волны, бегущей слева направо. На рисунке показано, что E_x максимально в данный момент вре-

мени в той же точке, что и B_y . Нулевому значению E_x соответствует и нулевое значение B_y , причем при $E_x > 0$, $B_y > 0$ (знак «+» в формуле (6.12)). Волна, бегущая справа налево показана на рис. 6.5. Отметим, что положительному полю E_x отвечает отрицательное значение B_y .

До сих пор мы не конкретизировали вид функций, описывающих электромагнитную волну. Важным конкретным видом плоской электромагнитной волны является монохроматическая волна

$$E_x = E_0 \cos(\omega t - kz), \quad B_y = B_0 \cos(\omega t - kz), \quad (6.13)$$

причем в согласии с (6.12) $E_0 = \pm c/\sqrt{\epsilon\mu} B_0$, $k = (\omega/c)\sqrt{\epsilon\mu}$ — волновое число. Мгновенная фотография этой волны (бегущей

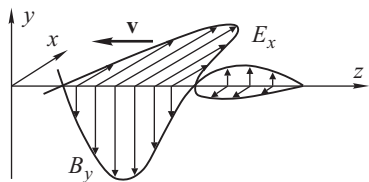


Рис. 6.5

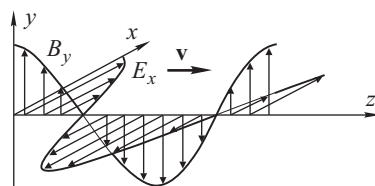


Рис. 6.6

слева направо) показана на рис. 6.6. Можно убедиться (упражнение для читателя), что функции E_x и B_y , описываемые соотношениями (6.13), удовлетворяют волновому уравнению.

6.2. Поляризация электромагнитных волн

Уравнения (6.8) и (6.9) показывают, что могут независимо существовать две электромагнитные волны, распространяющиеся в положительном направлении оси z . В одной из них вектор \mathbf{E} колеблется в плоскости xz параллельно оси x , а вектор \mathbf{B} — в плоскости yz параллельно оси y . В другой волне, наоборот, колебания поля \mathbf{E} происходят в плоскости yz , а колебания поля \mathbf{B} — в плоскости xz . Каждую из этих волн называют *плоскополяризованной* (или линейно-поляризованной, см. гл. 4.3 — векторные волны), поскольку электрическое и магнитное поля сохраняют неизменной плоскость колебаний. Назовем первую из этих двух независимых волн x -волной (по направлению колебаний вектора \mathbf{E}), а вторую — y -волной.

Если существуют одновременно обе волны, то суммарный вектор \mathbf{E} может менять свою ориентацию в пространстве, оставаясь, разумеется, в плоскости, перпендикулярной оси z . В дальнейшем мы будем говорить об электрическом поле волны, поскольку воздействие на регистрирующие приборы связано, в основном, с электрическим полем.

Рассмотрим суперпозицию двух плоских монохроматических волн, бегущих в положительном направлении оси z (x -волны и y -волны). Колебания в некоторой фиксированной плоскости $z = \text{const}$ можно записать в виде

$$E_x = a_1 \cos(\omega t - \varphi_1); \quad E_y = a_2 \cos(\omega t - \varphi_2),$$

где a_1 и a_2 — амплитуды волн, φ_1 и φ_2 — начальные фазы: $\varphi_1 = kz + \alpha_1$, $\varphi_2 = kz + \alpha_2$. Задача сложения двух взаимно перпендикулярных колебаний одинаковой частоты была рассмотрена в гл. 1.5. Мы выяснили, что в общем случае конец суммарного вектора описывает эллипс, вид которого зависит от амплитуд и разности фаз $\Delta\varphi = \varphi_2 - \varphi_1$ слагаемых колебаний. В частных случаях $\Delta\varphi = 0$ и $\Delta\varphi = \pm\pi$ результирующая волна оказывается *линейно-поляризованной* — эллипс вырождается в прямую. Плоскость колебаний суммарного вектора \mathbf{E} при $\Delta\varphi = 0$ ориентирована так, как показано на рис. 1.29 б, а при $\Delta\varphi = \pm\pi$ — как на рис. 1.29 г. При $\Delta\varphi = \pm\pi/2$ оси эллипса совпадают с осями x и y (см. рис. 1.29 в). Если, кроме того, $a_1 = a_2$, то получаем *круговую поляризацию* — конец суммарного вектора описывает окружность (см. рис. 1.29 д). В гл. 1.5 мы обсуждали вопрос о направлении вращения суммарного вектора в фиксированной плоскости $z = \text{const}$. Рекомендуем читателю еще раз вернуться к этому обсуждению.

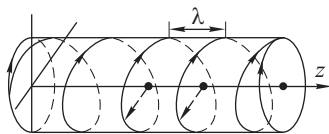


Рис. 6.7

Пространственную структуру эллиптически поляризованной волны поясняет рис. 6.7 (на примере волны с левой круговой поляризацией). Винтовая линия, изображенная на рисунке, есть геометрическое место концов суммарного вектора \mathbf{E} в различных плоскостях в фиксированный момент времени. Шаг винта равен длине волны λ . С течением времени изображенная на рисунке винтовая линия перемещается, не деформируясь, в направлении распространения волны, вдоль оси z . В каждой фиксированной плоскости $z = \text{const}$ вектор \mathbf{E} вращается против часовой стрелки, если смотреть навстречу волне.

6.3. Энергетические характеристики электромагнитных волн. Вектор Пойнтинга

Напомним, что электрическое и магнитное поля обладают энергией. В каждой единице объема, заполненного электрическим и магнитным полями, локализована энергия, равная

$$w_{\text{эм}} = \frac{\varepsilon\varepsilon_0 E^2}{2} + \frac{B^2}{2\mu\mu_0}. \quad (6.14)$$

Первое слагаемое $w_{\varepsilon} = (1/2)\varepsilon\varepsilon_0 E^2$ — объемная плотность энергии электрического поля (ε — диэлектрическая проницаемость среды). Второе слагаемое $w_{\mu} = 1/(2\mu\mu_0) B^2 = (1/2)\mu\mu_0 H^2$ — объемная плотность энергии магнитного поля.

В электромагнитной волне поля \mathbf{E} и \mathbf{B} изменяются во времени, поэтому и объемная плотность энергии $w_{\text{эм}}$ также со временем изменяется.

Рассмотрим плоскую монохроматическую волну, бегущую вдоль оси z . В этой волне поля \mathbf{E} и \mathbf{B} изменяются в данном месте пространства (при фиксированном z) по гармоническому закону. В некоторый момент времени поля \mathbf{E} и \mathbf{B} становятся максимальными (одновременно максимальными!). Максимальна в этот момент времени и объемная плотность энергии в данном месте. Через четверть периода колебаний поля \mathbf{E} и \mathbf{B} одновременно обращаются в нуль, при этом и объемная плотность энергии становится равной нулю. Спрашивается, куда она делась? При распространении электромагнитная волна переносит энергию из одного объема в другой, подобно текущей жидкости, которая из одного объема в другой переносит свою массу. Поясним введенное ранее (см. раздел 2) понятие *потока энергии* электромагнитного поля на примере плоской электромагнитной волны. Итак, рассмотрим x -волну, описываемую системой уравнений (6.8).

Умножим первое из уравнений на H_y , а второе — на $1/(\mu\mu_0)E_x$ и сложим левые и правые части уравнений (предварительно заменив B_y на $\mu\mu_0 H_y$). Мы получаем

$$H_y \frac{\partial E_x}{\partial z} + E_x \frac{\partial H_y}{\partial z} = - \left(\mu\mu_0 H_y \frac{\partial H_y}{\partial t} + \varepsilon\varepsilon_0 E_x \frac{\partial E_x}{\partial t} \right).$$

Последнему равенству можно придать вид

$$-\frac{\partial}{\partial z}(E_x H_y) = \frac{\partial}{\partial t} \left(\frac{1}{2}\mu\mu_0 H^2 + \frac{1}{2}\varepsilon\varepsilon_0 E^2 \right). \quad (6.15)$$

Выражение в скобках в правой части равенства есть объемная плотность энергии электромагнитного поля, поэтому вместо (6.15) имеем

$$\frac{\partial}{\partial t} w_{\text{эм}} = -\frac{\partial}{\partial z} (E_x H_y).$$

Интегрируя последнее равенство по z в пределах от z_1 до z_2 , найдем

$$\frac{\partial}{\partial t} W_{\text{эм}} = E_x H_y \Big|_{z=z_1} - E_x H_y \Big|_{z=z_2},$$

где $W_{\text{эм}}$ — энергия электромагнитного поля, заключенного в столбике единичного сечения между плоскостями z_1 и z_2 .

Введем величину

$$S = E_x H_y. \quad (6.16)$$

Тогда последнее соотношение примет вид

$$\frac{\partial}{\partial t} W_{\text{эм}} = S_1 - S_2. \quad (6.17)$$

Это — частный случай теоремы Пойнтинга (разд. II, гл. 9): изменение энергии электромагнитного поля в единицу времени обусловлено тем, что через сечение z_1 в объем *втекает* энергия, равная S_1 , а через сечение z_2 из объема *вытекает* энергия, равная S_2 . Таким образом, физический смысл величины S , которую мы определили равенством (6.16), — это *плотность потока электромагнитной энергии*, т. е. энергия, протекающая в единицу времени через единичную площадку, перпендикулярную направлению распространения волны. Обратим внимание, что для x -волны, бегущей слева направо, величины E_x и B_y имеют одинаковый знак, поэтому в этой волне в любой момент времени $S \geq 0$ — энергия течет слева направо.

Вектор \mathbf{S} , введенный с помощью равенства

$$\mathbf{S} = [\mathbf{E} \mathbf{H}], \quad (6.18)$$

определяет величину и направление потока электромагнитной энергии и называется вектором плотности потока энергии, или *вектором Пойнтинга*.

Рассмотрим поток энергии в волне, бегущей слева направо вдоль оси z , которая содержит как x -волну, так и y -волну. Суммарное электрическое поле этой волны есть

$$\mathbf{E} = E_x \mathbf{i} + E_y \mathbf{j},$$

где \mathbf{i} и \mathbf{j} — единичные орты в направлении осей x и y соответственно. Магнитное поле волны

$$\mathbf{H} = H_y \mathbf{j} - H_x \mathbf{i},$$

так как в y -волне положительному полю E_y соответствует отрицательное поле H_x . Вектор Пойнтинга в суммарной волне есть

$$\mathbf{S} = [\mathbf{E} \mathbf{H}] = [(E_x \mathbf{i} + E_y \mathbf{j}), (H_y \mathbf{j} - H_x \mathbf{i})].$$

Поскольку $[\mathbf{i} \mathbf{i}] = 0$ и $[\mathbf{j} \mathbf{j}] = 0$, то из последнего равенства получаем

$$S = E_x H_y - E_y H_x = S^{(x)} + S^{(y)},$$

где $S^{(x)}$ и $S^{(y)}$ — потоки энергии, которые существовали бы в отдельно взятых x -волне и y -волне. В частном случае монохроматических волн, используя (6.13), можно получить (упражнение для читателя):

$$\langle S \rangle = \frac{1}{2} \sqrt{\frac{\varepsilon'}{\mu'}} (E_1^2 + E_2^2), \quad (6.19)$$

где $\langle S \rangle$ — среднее значение потока энергии за период колебания.

6.4. Стоячая электромагнитная волна

Рассмотрим суперпозицию двух бегущих навстречу друг другу гармонических волн одинаковой частоты, амплитуды и поляризации (например, две x -волны). В волне, бегущей слева направо:

$$E_1 = a \cos(\omega t - kz), \quad B_1 = (\sqrt{\varepsilon\mu}/c)a \cos(\omega t - kz).$$

В волне, бегущей справа налево:

$$E_2 = a \cos(\omega t + kz), \quad B_2 = -(\sqrt{\varepsilon\mu}/c)a \cos(\omega t + kz).$$

Еще раз подчеркнем, что колебания полей E_1 и E_2 происходят вдоль оси x , а полей B_1 и B_2 — вдоль оси y . Результирующие поля E и B есть

$$\begin{aligned} E &= E_1 + E_2 = 2a \cos kz \cdot \cos \omega t; \\ B &= B_1 + B_2 = 2 \frac{\sqrt{\varepsilon\mu}}{c} a \sin kz \cdot \sin \omega t. \end{aligned} \quad (6.20)$$

Соотношения (6.20) описывают стоячую электромагнитную волну. Связь между полями E и B в этой волне в корне отличается от (6.12) — связи E и B в бегущей волне. Напомним, что, согласно (6.12), в любой фиксированной плоскости $z = \text{const}$ колебания полей E_x и B_y происходят синфазно; там, где в данный момент максимально поле E , там же в этот момент максимально поле B (значит, там же в этот момент максимальна плотность электромагнитной энергии и там же максимален и поток энергии). Это — полная аналогия с тем, что происходит в бегущей упругой волне. В стоячей электромагнитной волне колебания E_x и B_y сдвинуты по фазе на четверть периода: в тот момент времени, когда электрическое поле максимально в данном месте ($\cos \omega t = 1$), магнитное поле обращается в нуль ($\sin \omega t = 0$) и наоборот. Значит, за четверть периода происходит полное превращение электрической энергии в магнитную и наоборот. Кроме того, пучностям электрического поля ($\cos kz = 1$) соответствуют узлы магнитного поля ($\sin kz = 0$), т. е. пучности электрического и магнитного полей сдвинуты друг относительно друга на $\lambda/4$ (разумеется, так же сдвинуты друг относительно друга и узлы). Согласно (6.18) поток энергии равен нулю в тех местах, где обращается в нуль либо электрическое поле волны, либо магнитное поле. Следовательно, промежуток пространства, равный $\lambda/4$ (вдоль оси z) — от узла электрического поля до ближайшего узла магнитного поля, не обменивается энергией с соседними участками (сравните с продольной упругой стоячей волной в стержне с закрепленными концами, либо с поперечной волной в натянутой струне). Распределение электрического

и магнитного полей в стоячей волне вдоль оси z (мгновенная фотография) показано на рис. 6.8.

Стоячие электромагнитные волны возникают в резонаторе Фабри–Перо, образованном двумя параллельными идеально проводящими металлическими стенками. Поскольку на поверхности

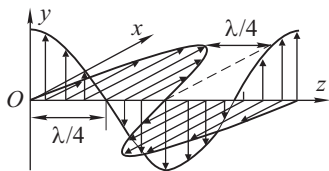


Рис. 6.8

идеального проводника тангенциальная компонента электрического поля должна обращаться в нуль, то ситуация в точности подобна той, которая возникает в стержне с закрепленными концами либо в натянутой и закрепленной на концах струне (поскольку аналогичны граничные условия: $E(0) = E(L) = 0$, где L — рас-

стояние между проводящими стенками). Возможные частоты колебаний определяется соотношением $\nu_n = nc/(2L)$, аналогичным (5.25), а возможные типы колебаний

$$E_n(z, t) = 2a \sin\left(n \frac{\pi}{L} z\right) \cdot \sin \omega_n t$$

полностью аналогичны типам упругих стоячих волн в стержне.

6.5. Излучение колеблющегося диполя

Колеблющийся диполь является простейшей системой, излучающей электромагнитную волну. Модель колеблющегося диполя используется при исследовании излучения многих реальных излучающих систем, таких, как радиоантенна или даже излучения света атомом.

Практической реализацией диполя является элементарный осциллятор, изображенный на рис. 6.9а: два проводящих шарика, заряды которых $\pm q$ меняются по гармоническому закону. Это изменение обеспечивается переменным напряжением $U(t)$, которым питается осциллятор. Дипольный момент такой системы есть, по определению, $\mathbf{p} = q\mathbf{l}$, где \mathbf{l} — вектор, направленный от отрицательного заряда к положительному (см. разд. II).

При гармоническом изменении заряда $q = q_0 \cos \omega t$ имеем

$$\mathbf{p} = \mathbf{p}_0 \cos \omega t, \quad \text{где } \mathbf{p}_0 = q_0 \mathbf{l}. \quad (6.21)$$

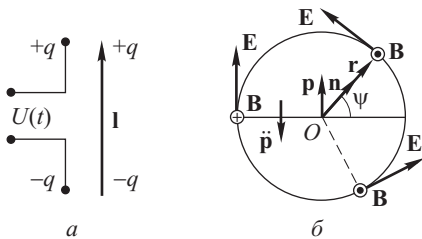


Рис. 6.9

Будем полагать, что размеры осциллятора много меньше длины излучаемой волны:

$$l \ll \lambda = \frac{2\pi c}{\omega}, \quad (6.22)$$

где c — скорость электромагнитных волн: вибратор находится в вакууме. Кроме того, нас будет интересовать излучение диполя на расстояниях от него, больших по сравнению с длиной волны

$$r \gg \lambda.$$

При выполнении указанных условий диполь называется *точечным*.

Согласно классическим представлениям, возбужденный электрон в атоме ведет себя как гармонический осциллятор, совершающий колебания относительно положения равновесия. При этом частота колебаний ω определяется массой электрона и «жесткостью связи» (см., например, модель атома Томсона, гл. 4.1). В результате гармонических колебаний электрона атом приобретает переменный дипольный момент (6.21).

Приведем без вывода формулы, определяющие электрическое и магнитное поля в электромагнитной волне, излучаемой колеблющимся диполем:

$$\mathbf{B}(t) = -\frac{1}{4\pi\epsilon_0} \frac{1}{c^3 r} \left[\mathbf{n} \times \ddot{\mathbf{p}} \left(t - \frac{r}{c} \right) \right], \quad \mathbf{E}(t) = c [\mathbf{B}(t) \times \mathbf{n}]. \quad (6.23)$$

Формулы (6.23) определяют поля \mathbf{B} и \mathbf{E} в точке наблюдения, отстоящей от середины диполя на расстояние r . Единичный вектор \mathbf{n} направлен от центра диполя к точке наблюдения (рис. 6.9б). Обсудим физическое содержание формул (6.23).

1. Поля \mathbf{B} и \mathbf{E} определяются *второй производной по времени дипольного момента \mathbf{p}* (вектором $\ddot{\mathbf{p}}$), т. е. *ускорением*, с которым меняется заряд диполя q (либо расстояние между зарядами l): волна излучается лишь ускоренно движущимися зарядами.

2. Поля \mathbf{B} и \mathbf{E} на расстоянии r от колеблющегося диполя в момент времени t зависят от значения $\ddot{\mathbf{p}}$ в более ранний момент времени $(t - r/c)$ — запаздывание зависит от времени распространения волны от осциллятора до точки наблюдения $\tau = r/c$.

3. Электромагнитная волна, излучаемая диполем, является *плоскополяризованной*: вектор $\ddot{\mathbf{p}}$ параллелен вектору \mathbf{p} , т. е. оси диполя, поэтому вектор \mathbf{B} перпендикулярен плоскости, в которой лежат векторы \mathbf{p} и \mathbf{n} (плоскости рис. 6.9б); вектор \mathbf{E} лежит в плоскости рисунка и перпендикулярен вектору \mathbf{n} . Пространственная картина полей \mathbf{E} и \mathbf{B} показана на рис. 6.10: линии вектора \mathbf{E} расположены по

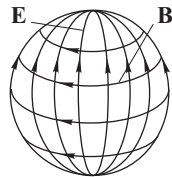


Рис. 6.10

меридианам, а линии \mathbf{B} — по параллелям сферы с центром в точке, совпадающей с центром диполя.

4. Поля E и B убывают при удалении от диполя обратно пропорционально *первой степени* расстояния $1/r$ (в отличие от поля статического диполя, которое убывает по закону $1/r^3$ — т. е. гораздо быстрее).

5. Излучение диполя обладает определенной анизотропией: амплитуды полей зависят от широтного угла ψ : E и B максимальны в направлении $\psi = 0$ (т. е. на экваторе) и обращаются в нуль на полюсах ($\psi = \pm \pi/2$).

Найдем, используя (6.21), $\ddot{\mathbf{p}}(t - r/c)$. Имеем $\ddot{\mathbf{p}}(t) = -\omega^2 p_0 \cos \omega t$; $\ddot{\mathbf{p}}(t - r/c) = -\omega^2 p_0 \cos(\omega t - kr)$. Подставляя последнее выражение в (6.23), найдем

$$\begin{aligned} B(t) &= \frac{\omega^2 p_0}{4\pi\epsilon_0 c^3 r} \cos \psi \cos(\omega t - kr), \\ E(t) &= \frac{\omega^2 p_0}{4\pi\epsilon_0 c^2 r} \cos \psi \cos(\omega t - kr). \end{aligned} \quad (6.24)$$

Важно подчеркнуть, что на больших расстояниях ($r \gg \lambda$) волновые поверхности, излучаемые колеблющимся диполем, имеют сферическую форму, а амплитуда колебаний убывает по закону $1/r$ — так же, как в сферической волне, излучаемой точечным источником.

Электромагнитная энергия, излучаемая колеблющимся диполем. Используя (6.24), вычислим поток энергии в точке

r, ψ в волне, излучаемой диполем. Имеем (используя формулы $H = (1/\mu_0)B$ и $c^2 = 1/(\epsilon_0\mu_0)$):

$$\begin{aligned} S = EH &= \frac{1}{(4\pi)^2 \epsilon_0} \frac{\omega^4 p_0^2}{c^3 r^2} \cos^2 \psi \times \\ &\times \cos^2(\omega t - kr). \end{aligned} \quad (6.25)$$

Направление потока совпадает с нормалью \mathbf{n} к волновой поверхности. Среднее за период значение S есть

$$\overline{S} = \frac{\omega^4 p_0^2}{32\pi^2 \epsilon_0 c^3 r^2} \cos^2 \psi. \quad (6.26)$$

На рис. 6.11а показана *полярная диаграмма направленности* излучающего диполя — зависимость S от ψ (при фиксированном r) в полярных координатах.

Поток монотонно (по закону $\cos^2 \psi$) уменьшается с ростом угла ψ от максимального значения, равного $S_{\max} = \omega^4 p_0^2 / (32\pi^2 \epsilon_0 c^3 r^2)$ при $\psi = 0$ до нулевого при $\psi = \pi/2$ —

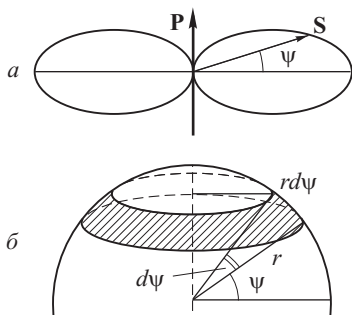


Рис. 6.11

в направлении своей оси колеблющийся диполь не излучает!

Найдем общую мощность, излучаемую диполем по всем направлениям. Рассмотрим на волновой поверхности — сфере радиуса r — элементарное колечко, заштрихованное на рис. 6.11 б. Его площадь

$$d\sigma = 2\pi r \cos \psi \cdot r d\psi = 2\pi r^2 \cos \psi d\psi. \quad (6.27)$$

Средний поток энергии через это кольцо $dP = \bar{S} d\sigma$, а полная энергия, излучаемая в 1 с по всем направлениям $P = \int \bar{S} d\sigma$. Используя (6.26), находим

$$P = \frac{\omega^4 p_0^2}{16\pi \varepsilon_0 c^3} \int_{-\pi/2}^{\pi/2} \cos^3 \psi d\psi = \frac{\omega^4 p_0^2}{12\pi \varepsilon_0 c^3}. \quad (6.28)$$

Важное следствие полученного выражения: энергия, излучаемая диполем за 1 с, быстро (пропорционально ω^4) растет с ростом частоты и именно поэтому для передачи радиосигналов используются высокие частоты (в радиовещании от 10^5 до 10^8 Гц, что соответствует длинам волн от 3 км до 3 м).

Поскольку колеблющийся диполь излучает энергию, то для поддержания незатухающих гармонических колебаний необходим источник, восполняющий потерю энергии, равную P . Выражение (6.28) можно представить в виде, аналогичном формуле $P = RI^2/2$, определяющей мощность джоулевых потерь в колебательном контуре с сопротивлением R . Поскольку заряд колеблющегося диполя меняется по закону $q = q_0 \cos \omega t$, то ток $I = \dot{q} = -q_0 \omega \sin \omega t = -I_0 \sin \omega t$, где $I_0 = \omega q_0 = \omega p_0/l$. Используя последнее выражение, получим из (6.28):

$$P = \frac{\omega^2 l^2}{12\pi \varepsilon_0 c^3} I_0^2. \quad (6.29)$$

Роль сопротивления играет величина

$$R_{\text{изл}} = \frac{\omega^2 l^2}{6\pi \varepsilon_0 c^3}.$$

При таком определении формула (6.29) принимает вид

$$P = \frac{R_{\text{изл}} I_0^2}{2}.$$

Введенная величина $R_{\text{изл}}$ играет для потерь на излучение ту же роль, что и сопротивление контура для потерь на нагревание. Величина $R_{\text{изл}}$ называется *сопротивлением излучения*.

Вернемся к классической модели атома, согласно которой электрон в атоме совершает гармонические колебания (как ко-

леблющийся осциллятор): $l = l_0 \cos \omega t$. Энергия, запасенная осциллирующим электроном, равна $W = m\dot{l}^2/2 + kl^2/2$. Вычисляя ее в момент, когда электрон проходит положение равновесия ($l = 0$ и $\dot{l} = \omega l_0$), найдем $W = m\omega^2 l_0^2/2 = m\omega^2 p_0^2/(2q^2)$. Потери энергии за период колебаний, связанные с излучением, равны $\Delta W = PT = P \cdot 2\pi/\omega$, где P определяется формулой (6.29). Добротность нашей колебательной системы найдем, используя (2.42):

$$Q = 2\pi \frac{W}{\Delta W} = \frac{3mc^3}{2\omega q^2}.$$

Для электрона, колеблющегося с частотой $\omega/2\pi = 6 \cdot 10^{14}$ Гц (излучение зеленого света), используя числовые значения $m = 9,1 \cdot 10^{-31}$ кг, $q = 1,6 \cdot 10^{-19}$ Кл, получим

$$Q \approx 1,4 \cdot 10^7.$$

Столь высокой добротностью не обладает ни одна «рукотворная» колебательная система. Время затухания (в течение которого амплитуда осцилляций уменьшается в e раз) равно

$$\tau = \frac{1}{\delta} = \frac{2Q}{\omega} \approx 7,7 \cdot 10^{-9} \text{ с.}$$

Затухание осциллятора, обусловленное излучением электромагнитной волны, называют *радиационным затуханием*.

6.6. Отражение электромагнитной волны от идеального проводника

Напомним основные результаты решения задачи, рассмотренной ранее в § 2.5. Пусть плоская электромагнитная волна, бегущая вдоль оси z , падает на плоскую поверхность идеального проводника (поверхность $z = 0$, рис. 6.12)



Рис. 6.12

$$E_1 = E_{1x}(z, t) = a \cos(\omega t - kz), \quad (6.30)$$

$$B_1 = B_{1y}(z, t) = \frac{\sqrt{\epsilon\mu}}{c} a \cos(\omega t - kz).$$

Волна (6.30) линейно-поляризована: вектор \mathbf{E} имеет только x -компоненту, а вектор \mathbf{B} только y -компоненту (ось y перпендикулярна плоскости рисунка и направлена к читателю).

Как известно, на поверхности проводника (при $z = 0$) должно выполняться граничное условие

$$E_x(0, t) = 0 \quad (6.31)$$

(равенство нулю тангенциальной компоненты электрического поля). Это условие окажется выполненным, если мы предположим, что наряду с падающей на поверхность проводника волной (6.30) возникает отраженная волна

$$\begin{aligned} E'_{1x}(z, t) &= a' \cos(\omega t + kz), \\ B'_{1y}(z, t) &= -\frac{\sqrt{\varepsilon\mu}}{c} a' \cos(\omega t + kz). \end{aligned} \quad (6.32)$$

Направление отраженной волны противоположно положительному направлению оси z , поэтому аргумент косинуса в (6.32) есть $\omega t + kz$. Кроме того, векторы \mathbf{E}' , \mathbf{B}' и \mathbf{k}' должны составлять «правую» тройку (как в любой бегущей плоской электромагнитной волне), поэтому при изменении направления волны на противоположное изменяется и направление одного из векторов — либо \mathbf{E} , либо \mathbf{B} . Граничное условие (6.31) принимает вид

$$E_x(0, t) = E_{1x}(0, t) + E'_{1x}(0, t) = 0. \quad (6.33)$$

Подставляя (6.30) и (6.32) в граничное условие (6.33), получаем $a = -a'$, т. е. амплитуда отраженной волны равна амплитуде падающей волны — коэффициент отражения равен единице, а знак «—» в последнем равенстве означает, что на поверхности проводника (при $z = 0$) колебания поля E'_{1x} в отраженной волне происходят в противофазе с колебаниями поля E_1 в падающей волне:

$$\begin{aligned} E'_{1x}(z, t) &= -a \cos(\omega t + kz), \\ B'_{1y}(z, t) &= \frac{\sqrt{\varepsilon\mu}}{c} a \cos(\omega t + kz). \end{aligned} \quad (6.34)$$

Тогда результирующая волна в диэлектрике (расположенном над проводником), т. е. сумма волн (6.34) и (6.30), есть

$$\begin{aligned} E_x(z, t) &= E_{1x}(z, t) + E'_{1x}(z, t) = 2a \sin kz \sin \omega t, \\ B_y(z, t) &= B_{1y}(z, t) + B'_{1y}(z, t) = 2 \frac{\sqrt{\varepsilon\mu}}{c} a \cos kz \cos \omega t. \end{aligned} \quad (6.35)$$

Последние формулы описывают стоячую электромагнитную волну, в которой на поверхности проводника (при $z = 0$) находится узел электрического поля $E_x(0, t) = 0$ и пучность магнитного поля $B_y(0, t) = 2(\sqrt{\varepsilon\mu}/c) a \cos \omega t$.

Заметим, что разрыв тангенциальных компонент магнитного поля (внутри идеального проводника $B = B_y = 0$) обусловлен токами проводимости, которые индуцируются падающей электромагнитной волной.

6.7. Волноводы

Пусть плоская электромагнитная волна падает на поверхность идеального проводника, как показано на рис. 6.13: вектор \mathbf{k} лежит в плоскости yz — на этот раз мы выбрали оси

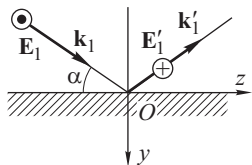


Рис. 6.13

координат так, что плоскость yz совпадает с плоскостью рисунка, причем ось y перпендикулярна плоской поверхности проводника, а ось z параллельна поверхности т. е. плоская поверхность проводника — плоскость xz .

Будем полагать, что падающая волна линейно-поляризована, причем вектор \mathbf{E} перпендикулярен плоскости рисунка yz , т. е. имеет только x -компоненту и, следовательно, параллелен поверхности проводника. Уравнение падающей волны имеет вид

$$E_{1x} = a \cos(\omega t - k_y y - k_z z), \quad (6.36)$$

x -компонента вектора \mathbf{k} равна нулю, $k_y = k \sin \alpha$ и $k_z = k \cos \alpha$ — соответственно y - и z -компоненты вектора \mathbf{k}_1 , α — угол между вектором \mathbf{k}_1 и осью z .

Вновь полагаем, что наряду с волной E_{1x} возникает отраженная волна

$$E'_{1x} = a' \cos(\omega t - k'_y y - k'_z z + \varphi),$$

причем ее амплитуда a' , направление распространения α' и начальная фаза φ должны быть выбраны так, чтобы выполнялось граничное условие — равенство нулю тангенциальной компоненты вектора \mathbf{E} в суммарной волне на поверхности (при $y = 0$). Легко видеть, что это условие выполняется, если амплитуда отраженной волны равна амплитуде падающей волны ($a' = a$), угол падения равен углу отражения ($\alpha = \alpha'$), а колебания E'_{1x} на поверхности (при $y = 0$) противофазны с колебаниями E_{1x} ($\varphi = \pi$), т. е.

$$E'_{1x} = a \cos(\omega t + k_y y - k_z z + \pi), \quad (6.37)$$

вектор \mathbf{k}' отраженной волны имеет компоненты $k'_y = -k_y$, $k'_z = k_z$ поэтому аргумент косинуса содержит слагаемое $+k_y y$ вместо слагаемого $-k_y y$ в падающей волне.

Действительно, суммарная волна над проводящей стенкой (сумма волн (6.36) и (6.37)) имеет вид

$$E_x(y, z, t) = -2a \sin(k_y y) \sin(\omega t - k_z z). \quad (6.38)$$

На проводящей поверхности (при $y = 0$)

$$E_x(0, z, t) = 0,$$

т. е. требуемое граничное условие выполняется.

Обратим теперь внимание, что если на расстоянии d от проводящей стенки (т. е. при $y = d$) установить вторую проводящую стенку (параллельно первой), причем выбрать расстояние между стенками d так, чтобы $d = y = n\pi/k_y$, т. е.

$$k_y = \frac{n\pi}{d} \quad (n = 1, 2, 3 \dots), \quad (6.39)$$

то волна (6.38) в области между стенками $0 < y < d$ не изменится, поскольку необходимые граничные условия — равенство нулю тангенциальной компоненты поля на проводящих поверхностях (при $y = 0$ и $y = d$), автоматически выполняются. При условии (6.39) имеем из (6.38)

$$E_x(y, z, t) = 2a \sin\left(n\pi \frac{y}{d}\right) \sin(\omega t - k_z z), \quad (6.40)$$

где

$$k_z = \sqrt{k^2 - \left(n\frac{\pi}{d}\right)^2} = \sqrt{\left(\frac{\omega}{c}\right)^2 - n^2 \left(\frac{\pi}{d}\right)^2}. \quad (6.41)$$

Итак, между двумя параллельными проводящими стенками, расположенными на расстоянии d друг от друга (коридор между стенками представляет собой простейший волновод) могут распространяться волны вида (6.40): каждому значению n отвечает свой тип волны — своя «мода» волновода, разным модам отвечают различные конфигурации поля — различные распределения амплитуды колебаний в фиксированном сечении $z = \text{const}$

$$A_n(y) = 2a \sin\left(n\pi \frac{y}{d}\right). \quad (6.42)$$

На рис. 6.14 показаны распределения амплитуды $A_n(y)$, отвечающие индексам $n = 1$ и $n = 2$. Как и должно быть в соответствии с граничными условиями, $A_n(y) = 0$ при $y = 0$ и $y = d$ (ось волновода — ось z — перпендикулярна плоскости рис. 6.14, а вектор \mathbf{E} в волне параллелен оси x).

Каждый тип волны, отвечающий фиксированному значению n , можно рассматривать как суперпозицию плоской волны, падающей на проводящую стенку, и волны, отраженной стенкой, т. е. каждый тип волны образован суммой двух плоских волн, волновые векторы которых \mathbf{k} и \mathbf{k}' составляют угол α_n с осью волновода, причем разным n отвечают различные углы α :

$$\sin \alpha_n = \frac{n\pi}{kd} = \frac{n\lambda_0}{2d}, \quad (6.43)$$

где $\lambda_0 = 2\pi/k = 2\pi c/\omega$ — длина волны в вакууме.

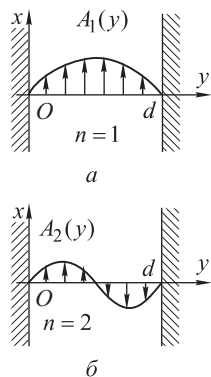


Рис. 6.14

Поверхности одинаковой фазы колебаний, как видно из (6.40) — это поверхности $z = \text{const}$. Это утверждение в данном случае означает, что в плоскости $z = \text{const}$ колебания либо синфазны как, например, в моде $n = 1$, либо отличаются по фазе на π как, например, в моде $n = 2$, где колебания в области $0 < y < d/2$ сдвинуты по фазе на π по отношению к колебаниям в области $d/2 < y < d$.

Волновые поверхности перемещаются вдоль оси волновода — оси z , причем фазовая скорость может быть найдена из (6.40) и (6.41)

$$\omega t - k_z z = \text{const},$$

откуда

$$v_\phi = \frac{dz}{dt} = \frac{\omega}{k_z} = \frac{\omega}{\sqrt{(\omega/c)^2 - n^2(\pi/d)^2}} = \frac{c}{\sqrt{1 - n^2(\lambda_0/2d)^2}}. \quad (6.44)$$

Как ясно из (6.44), фазовая скорость волн в волноводе больше скорости света в вакууме.

Расстояние между волновыми поверхностями, на которых фаза колебаний отличается на 2π (т. е. колебания синфазны) — это длина волны в волноводе, равная

$$\lambda = \frac{2\pi}{k_z} = \frac{\lambda_0}{\sqrt{1 - n^2(\lambda_0/2d)^2}} > \lambda_0. \quad (6.45)$$

Итак, длина волны в волноводе больше длины волны той же частоты в вакууме.

Наконец, из (6.40) ясно, что если k_z — действительное число, то $E_x(y, z, t)$ действительно представляет собой бегущую вдоль оси z волну с перемещающимися волновыми поверхностями. Из (6.41) следует, что для этого необходимо выполнение неравенства

$$\left(\frac{\omega}{c}\right)^2 \geq n^2 \left(\frac{\pi}{d}\right)^2.$$

Таким образом, существует *критическая частота* $\omega_{\text{кр}}$ — *минимальная частота* волны, которая может распространяться (бежать) по волноводу. Эта частота отвечает значению $n = 1$

$$\omega_{\text{кр}} = \frac{\pi c}{d}. \quad (6.46)$$

Соответствующая длина волны в вакууме $\lambda_{\text{кр}} = 2\pi c/\omega_{\text{кр}} = 2d$ — это максимальное значение длины волны в вакууме, которая может бежать по волноводу. Критическому условию отвечает угол α (направление пары плоских волн, составляющих моду волновода), равный, согласно (6.43):

$$\sin \alpha = \frac{\lambda_{\text{кр}}}{2d} = 1, \quad \text{т. е.} \quad \alpha = \frac{\pi}{2}.$$

Такая пара волн образует, как мы знаем, стоячую волну между стенками, которая никуда не распространяется. Узлы стоячей волны должны находиться на проводящих стенках (чтобы выполнялись требуемые граничные условия). Ясно, что при $\lambda_0 > \lambda_{\text{кр}} = 2d$ половина длины волны (расстояние между узлами) «не уместается» между стенками, и поэтому для таких волн необходимые граничные условия не могут быть выполнены.

Разумеется, мы описали здесь лишь простейшие типы волн в волноводе с проводящими параллельными стенками, типы, в которых поле **E** параллельно стенкам (оси x) и перпендикулярно оси волновода — оси z (рис. 6.15а). В таких типах волн магнитное поле **B** имеет проекцию на ось z — направление распространения, т. е. электромагнитная волна в волноводе не является поперечной (в данном случае поперечно поле **E**, и не поперечно поле **B**). Рисунок 6.15б иллюстрирует это обстоятельство: линии магнитного поля **B** (перпендикулярного оси x , т. е. полю **E**) «охватывают» линии поля **E**.

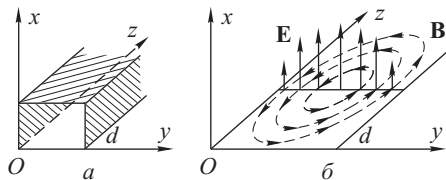


Рис. 6.15

6.8. Электромагнитная волна на границе раздела двух диэлектриков

Полностью задача падения электромагнитной волны на границу раздела двух диэлектриков решена в § 11.3. Здесь мы приведем лишь результаты, относящиеся к нормальному падению (рис. 6.16), чтобы обратить внимание на аналогию с поведением

звуковых волн на границе раздела двух сред с различными акустическими свойствами.

Для того чтобы необходимые граничные условия:

$$E_{\tau 1} = E_{\tau 2}, \quad H_{\tau 1} = H_{\tau 2} \quad (6.47)$$

— непрерывность тангенциальных компонент векторов **E** и **H** на границе раздела

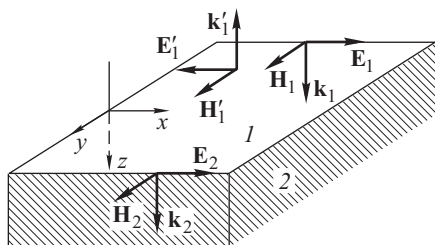


Рис. 6.16

двух диэлектриков — оказались выполненными, вновь предположим, что в диэлектрике 1 с диэлектрической проницаемостью ϵ_1 (расположенном над границей раздела $z = 0$ на рис 6.16) наряду

с падающей волной

$$E_1(z, t) = a_1 \cos(\omega t - k_1 z), \quad H_1(z, t) = \sqrt{\varepsilon'/\mu'} a_1 \cos(\omega t - k_1 z) \quad (6.48)$$

возникает отраженная волна

$$E_1'(z, t) = a_1' \cos(\omega t + k_1 z), \quad H_1'(z, t) = -\sqrt{\varepsilon'/\mu'} a_1' \cos(\omega t + k_1 z), \quad (6.49)$$

а в диэлектрике 2 (ε_2), расположенном под границей $z = 0$ — прошедшая волна:

$$E_2(z, t) = E_{2x} = a_2 \cos(\omega t - k_2 z), \quad (6.50)$$

$$H_2(z, t) = H_{2y} = \sqrt{\varepsilon'/\mu'} a_2 \cos(\omega t - k_2 z).$$

Для волн ($\mathbf{E}_1, \mathbf{H}_1$) и ($\mathbf{E}_1', \mathbf{H}_1'$) в диэлектрике 1 $k_1 = \omega/v_1 = (\omega/c)n_1$; для волны ($\mathbf{E}_2, \mathbf{H}_2$) в диэлектрике 2 $k_2 = \omega/v_2 = (\omega/c)n_2$, где величины $n_1 = \sqrt{\varepsilon_1\mu_1}$ и $n_2 = \sqrt{\varepsilon_2\mu_2}$ называются *показателями преломления*. Граничные условия (6.47) можно теперь записать в виде

$$E_1(0, t) + E_1'(0, t) = E_2(0, t), \quad H_1(0, t) + H_1'(0, t) = H_2(0, t).$$

Используя выражения (6.48), (6.49) и (6.50) при $z = 0$, после несложных вычислений получим формулы, связывающие амплитуды отраженной и прошедшей волны с амплитудой падающей волны

$$a_1' = \frac{n_1 - n_2}{n_1 + n_2} a_1, \quad a_2 = \frac{2n_1}{n_1 + n_2} a_1. \quad (6.51)$$

(При $n_1 < n_2$ a_1' и a_2' имеют разные знаки. Это означает, что в отраженной волне колебания вектора \mathbf{E}_1' противофазны колебаниям вектора \mathbf{E}_1 в падающей волне при $z = 0$. При этом колебания магнитных полей, как видно из (6.49), оказываются синфазными. Рисунок 6.16 соответствует именно этому случаю.)

Для коэффициентов отражения и прохождения, равных соответственно отношению средних за период потоков энергии в отраженной и прошедшей волнах к среднему потоку энергии в падающей волне, находим

$$R = \left(\frac{n_1 - n_2}{n_1 + n_2} \right)^2, \quad T = \frac{4n_1 n_2}{(n_1 + n_2)^2}. \quad (6.52)$$

Обратите внимание на замечательное сходство полученных выражений с формулами (5.40), определяющими коэффициенты отражения и прохождения звуковой волны, падающей на границу раздела двух сред с различными значениями $\sqrt{E\rho}$ (E — модуль упругости, ρ — плотность).

Эта величина для акустических волн играет ту же роль, что и показатель преломления $n = \sqrt{\varepsilon\mu}$ для электромагнитной волны.

Глава 7

ИНТЕРФЕРЕНЦИЯ ВОЛН

7.1. Принцип суперпозиции волн

Пусть в некоторой среде один источник создает волновой процесс $\mathbf{E}_1(\mathbf{r}, t)$, а другой — волновой процесс $\mathbf{E}_2(\mathbf{r}, t)$. Согласно принципу суперпозиции, если излучают одновременно два источника, то результирующий волновой процесс $\mathbf{E}(\mathbf{r}, t)$ есть

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_1(\mathbf{r}, t) + \mathbf{E}_2(\mathbf{r}, t). \quad (7.1)$$

Смысл этого равенства состоит в том, что распространяющиеся одновременно волны не влияют друг на друга, поэтому результирующее возмущение есть просто сумма неискажающих друг друга возмущений. Другими словами, в равенстве (7.1) $\mathbf{E}_1(\mathbf{r}, t)$ — это волна, которую создавал бы первый источник (если бы второй волны не было), так же как $\mathbf{E}_2(\mathbf{r}, t)$ — это волна, которую создавал бы второй источник в отсутствие первого.

Опыт подтверждает справедливость принципа суперпозиции в широких пределах: свету далекой звезды, идущему к нам из космоса, не мешает распространяться свет других звезд или свет горящей поблизости лампочки; радиоволны, заполняющие все окружающее пространство, не влияют друг на друга — вращая ручку настройки приемника, мы ловим передачу определенной радиостанции, несмотря на то, что приемник буквально погружен в океан радиоволн; звуки голосов многих людей, разговаривающих в комнате, не отскакивают друг от друга, подобно бильярдным шарам — они беспрепятственно проходят друг сквозь друга, не искажая друг друга.

Итак, для всех линейных волн независимо от их природы, справедлив принцип суперпозиции: можно сказать, что линейность волнового уравнения является математическим отражением (следствием) принципа суперпозиции: если возможны решения $\mathbf{E}_1(\mathbf{r}, t)$ и $\mathbf{E}_2(\mathbf{r}, t)$, то возможно также и решение (7.1).

Рассматривая в двух последующих главах явления интерференции и дифракции, мы будем иметь в виду либо скалярные, либо векторные, но одинаково поляризованные волны. Под E понимается, таким образом, либо колебания плотности среды при распространении в ней звуковой волны, либо колебания проекции электрического поля в электромагнитной волне (полагаем при этом, что все волны имеют, скажем, только x -компоненту),

либо какую-либо другую величину, изменяющуюся по волновым законам. При этом векторная сумма (7.1) заменяется скалярной суммой

$$E(\mathbf{r}, t) = E_1(\mathbf{r}, t) + E_2(\mathbf{r}, t). \quad (7.2)$$

Каков же результат наложения двух волн? Рассмотрим вначале наиболее простой, и в то же время исключительно важный случай, а именно, монохроматические волны.

7.2. Интерференция монохроматических волн

Итак, пусть в пространстве распространяются две *монохроматические* волны одинаковой частоты ω :

$$E_1(\mathbf{r}, t) = a_1(\mathbf{r}) \cos(\omega t - \varphi_1(\mathbf{r})), \quad E_2(\mathbf{r}, t) = a_2(\mathbf{r}) \cos(\omega t - \varphi_2(\mathbf{r})).$$

Согласно принципу суперпозиции колебательный процесс в любой точке наблюдения (радиус-вектор \mathbf{r} фиксирован) есть сумма гармонических колебаний. Квадрат амплитуды (т. е. интенсивность) результирующего колебания найдем, используя (1.7) (см. § 1.2):

$$a^2(\mathbf{r}) = a_1^2(\mathbf{r}) + a_2^2(\mathbf{r}) + 2a_1a_2 \cos \Delta\varphi(\mathbf{r}), \quad (7.3)$$

где $\Delta\varphi(\mathbf{r}) = \varphi_2(\mathbf{r}) - \varphi_1(\mathbf{r})$ — разность фаз слагаемых колебаний в точке наблюдения. Напомним, что интенсивность $I = a^2$ — это величина, равная (с точностью до константы) плотности потока энергии в монохроматической волне. Равенство (7.3) показывает, что *плотность потока энергии в результирующей волне не равна в общем случае сумме плотностей потоков энергии в слагаемых волнах*. В пространстве, где налагаются два волновых процесса, происходит *перераспределение потоков энергии*: в некоторых точках пространства результирующая интенсивность оказывается больше суммы интенсивностей $I_1 = a_1^2$ и $I_2 = a_2^2$ слагаемых волн; в других точках, наоборот, результирующая интенсивность меньше суммы интенсивностей. Результат зависит от *интерференционного слагаемого* $2a_1a_2 \cos \Delta\varphi(\mathbf{r})$, т. е. от разности фаз $\Delta\varphi(\mathbf{r})$ колебаний в данной точке наблюдения.

Явление наложения волн, при котором результирующая интенсивность оказывается не равной в общем случае сумме интенсивностей слагаемых волн, называется *интерференцией*.

Соотношение (7.3) является основным для описания интерференции двух монохроматических волн. Оно может быть записано в виде

$$I(\mathbf{r}) = I_1(\mathbf{r}) + I_2(\mathbf{r}) + 2\sqrt{I_1I_2} \cos \Delta\varphi(\mathbf{r}). \quad (7.4)$$

Чередование максимумов и минимумов интенсивности, характерное для интерференции, связано с изменением знака $\cos \Delta\varphi(\mathbf{r})$.

В тех точках, где $\Delta\varphi = 2\pi m$, интенсивность максимальна и равна $a_{\max}^2 = (a_1 + a_2)^2$. Геометрическое место точек, удовлетворяющих этому условию, образует максимум (полосу) m -го порядка. Там, где $\Delta\varphi = (2m + 1)\pi$, возникают минимумы интенсивности (темные интерференционные полосы), для которых $a_{\min}^2 = (a_1 - a_2)^2$. Контраст интерференционной картины принято характеризовать величиной *видности* V , определяемой равенством

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = \frac{2a_1 a_2}{a_1^2 + a_2^2}. \quad (7.5)$$

Видность максимальна и равна единице при равных амплитудах $a_1 = a_2$ (при этом $I_{\min} = 0$). Видность близка к нулю, когда интенсивность одной из интерферирующих волн существенно больше другой.

Принцип суперпозиции справедлив, в силу линейности уравнения Гельмгольца, и для комплексных амплитуд (см. § 4.2): комплексная амплитуда результирующей волны $f(\mathbf{r}) = a(\mathbf{r})e^{i\varphi(\mathbf{r})}$ равна сумме комплексных амплитуд $f_1(\mathbf{r}) = a_1(\mathbf{r})e^{i\varphi_1(\mathbf{r})}$ и $f_2(\mathbf{r}) = a_2(\mathbf{r})e^{i\varphi_2(\mathbf{r})}$ слагаемых волн:

$$f(\mathbf{r}) = f_1(\mathbf{r}) + f_2(\mathbf{r}).$$

Соответственно, интенсивность результирующей волны $I(\mathbf{r}) = a^2(\mathbf{r}) = |f(\mathbf{r})|^2$ есть

$$I(\mathbf{r}) = |f_1(\mathbf{r}) + f_2(\mathbf{r})|^2.$$

Последнее равенство можно записать в виде

$$I(\mathbf{r}) = |f_1(\mathbf{r})|^2 + |f_2(\mathbf{r})|^2 + f_1(\mathbf{r})f_2^*(\mathbf{r}) + f_1^*(\mathbf{r})f_2(\mathbf{r}), \quad (7.6)$$

где знак $*$ означает комплексно-сопряженную величину, $I_1 = |f_1(\mathbf{r})|^2$ и $I_2 = |f_2(\mathbf{r})|^2$ — интенсивности слагаемых волн.

Интерференционный эффект определяется двумя последними слагаемыми в (7.6): $f_1 f_2^*$ и $f_1^* f_2$ — это два комплексно-сопряженных слагаемых, поэтому $f_1 f_2^* + f_1^* f_2 = 2 \operatorname{Re} f_1^* f_2$.

Таким образом

$$I(\mathbf{r}) = |f_1(\mathbf{r})|^2 + |f_2(\mathbf{r})|^2 + 2 \operatorname{Re} f_1^* f_2. \quad (7.7)$$

Ясно, что интерференционное слагаемое в (7.3) имеет тот же вид, что и в (7.7):

$$\begin{aligned} 2 \operatorname{Re} f_1^* f_2 &= 2 \operatorname{Re} (a_1 e^{-i\varphi_1} \cdot a_2 e^{+i\varphi_2}) = 2 \operatorname{Re} (a_1 a_2 e^{i\Delta\varphi}) = \\ &= 2 a_1 a_2 \cos \Delta\varphi. \end{aligned}$$

Большинство эффектов интерференции и дифракции, которые будут рассматриваться в данной и последующей главах, наилучшим образом иллюстрируются на примере видимого света, т. е. электромагнитных волн с $\lambda \sim 400\text{--}700\text{ нм}$. При этом эффект интерференции проявляется в виде чередующихся светлых и темных полос на экране наблюдения.

Рассмотрим несколько характерных случаев интерференции монохроматических волн.

Интерференция сферических волн¹⁾. Две сферические волны излучаются точечными источниками S_1 и S_2 (рис. 7.1). Комплексные амплитуды волн в точке наблюдения P есть

$$f_1 = \frac{a_0}{r_1} e^{ikr_1}, \quad f_2 = \frac{a_0}{r_2} e^{ikr_2}$$

(мы полагаем, что источники излучают синфазно, с одинаковой амплитудой), r_1 и r_2 — расстояния от источников S_1 и S_1 до точки наблюдения.

Разность фаз колебаний в точке наблюдения $\Delta\varphi = k \cdot \Delta$, где $\Delta = r_2 - r_1$ — разность хода волн, приходящих в точку наблюдения. При $\Delta = m\lambda$ (разность хода равна целому числу длин волн) разность фаз $\Delta\varphi = k\Delta = 2\pi m$ и, следовательно, получаем интерференционный максимум m -го порядка.

При $\Delta = (2m + 1)\lambda/2$ получаем интерференционный минимум.

Геометрическое место точек, для которых $r_2 - r_1 = \Delta = \text{const}$ — это гиперболоиды вращения. Таким образом поверхности интерференционных максимумов и минимумов — гиперболоиды, сечение которых плоскостью, в которой лежат источники S_1 и S_2 (плоскость рисунка) показаны на рис. 7.1.

Ясно, что форма полос, наблюдаемых на каком-либо плоском экране, зависит от положения этого экрана. Например, если плоскость наблюдения перпендикулярна линии S_1S_2 , соединяющей источники (плоскость Π_1 на рис. 7.1), то интерференционные полосы имеют вид чередующихся светлых и темных колец.

Если экран расположен параллельно линии S_1S_2 (экран Π_2), то интерференционные полосы представляют собой семейство гипербол — почти параллельных прямых, если ограничиться

¹⁾ Нелишне напомнить, что поперечные электромагнитные волны в принципе не могут быть сферическими; их элементарный источник — не точка, а колеблющийся диполь. Тем не менее, в оптике нередко пользуются приближением сферической волны (точечного источника и точечного фокуса). Это означает, что интересующий нас участок волнового фронта имеет сферическую форму с точностью до малых поправок. Если же кривизна фронта существенна лишь в одной плоскости, а в ортогональной плоскости пренебрежимо мала, используется приближение цилиндрической волны.

достаточно малой областью вблизи центра экрана (точка O на рис. 7.2).

Целое число m называется *порядком интерференции*. Нулевой порядок интерференции соответствует разности хода $\Delta = 0$ (геометрическое место точек нулевой разности хода — плоскость симметрии OO на рис. 7.1).

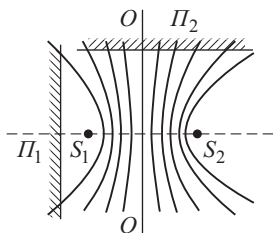


Рис. 7.1

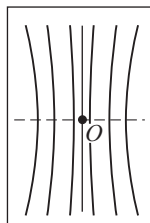


Рис. 7.2

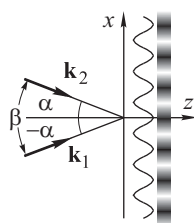


Рис. 7.3

Если рассматривать небольшую область наблюдения, в которой амплитуды колебаний двух слагаемых волн примерно одинаковы: $a_0/r_1 \approx a_0/r_2 \approx a$, то получаем

$$I = 2I_0 \left[1 + \cos \left(\frac{\omega}{c} \right) \Delta \right], \quad (7.8)$$

где c — скорость волны, I_0 — интенсивность каждой из волн в плоскости наблюдения.

Формула (7.8) имеет универсальный характер, она справедлива при интерференции *любых* монохроматических волн одинаковой частоты и интенсивности.

Интерференция плоских волн. Рассмотрим результат интерференции двух плоских волн, волновые векторы \mathbf{k}_1 и \mathbf{k}_2 которых составляют углы $\pm\alpha$ с осью z и лежат в плоскости рис. 7.3.

Проекции векторов \mathbf{k}_1 и \mathbf{k}_2 на ось y равны нулю. Комплексные амплитуды волн есть соответственно, $f_1 = a_1 e^{i(k_{1x}x + k_{1z}z)}$, $f_2 = a_2 e^{i(k_{2x}x + k_{2z}z)}$ (см. § 1.2). Поскольку $k_{1x} = k \sin \alpha$, $k_{2x} = -k \sin \alpha$, а z -компоненты векторов k_1 и k_2 одинаковы и равны $k_{1z} = k_{2z} = k \cos \alpha$, то имеем

$$f_1(x, z) = a_1 e^{i(kx \sin \alpha + kz \cos \alpha)}, \quad f_2(x, z) = a_2 e^{i(-kx \sin \alpha + kz \cos \alpha)}.$$

Результирующую картину интенсивности найдем, используя общее соотношение (7.4):

$$I(x) = a_1^2 + a_2^2 + 2a_1 a_2 \cos(2kx \sin \alpha). \quad (7.9)$$

Она не зависит от z , т. е. одинакова в любой плоскости наблюдения $z = \text{const}$ и поскольку она не зависит от y , то в любой

плоскости $z = \text{const}$ картина имеет вид чередующихся светлых и темных полос, параллельных оси y (т. е. перпендикулярных плоскости рисунка). Функция $I(x)$ показана на рис. 7.4.

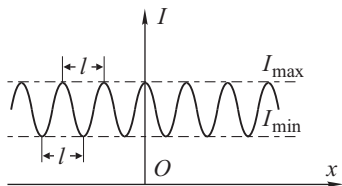


Рис. 7.4

Расстояние между двумя соседними максимумами интенсивности (интерференционными полосами) или между двумя соседними минимумами называют *шириной интерференционной полосы* (расстояние l на рис. 7.4). Переход от одной светлой полосы к другой соответствует изменению аргумента косинуса в (7.9) на 2π

$$2k \sin \alpha \cdot \Delta x = 2\pi,$$

откуда

$$\Delta x = l = \frac{\pi}{k \sin \alpha} = \frac{\lambda}{2 \sin \alpha}.$$

Обозначая через β угол схождения волн ($\beta = 2\alpha$), получаем

$$l = \frac{\lambda}{2 \sin (\beta/2)} \quad (7.10)$$

или для малых углов схождения: $\sin (\beta/2) \approx \beta/2$

$$l \approx \frac{\lambda}{\beta}. \quad (7.11)$$

В оптике, в силу малости $\lambda (\sim 5 \cdot 10^{-5} \text{ см})$, для получения достаточно широких полос ($l \sim 0,1\text{--}1 \text{ см}$) угол схождения волн β приходится выбирать малым ($\sim 10^{-4}\text{--}10^{-5}$ радиан), поэтому практически всегда можно пользоваться приближенным выражением (7.11).

Для оценки минимально возможной ширины полосы необходимо, разумеется, использовать точное соотношение (7.10). Из него следует, что минимальная ширина интерференционных полос достигается при $\beta = \pi$, т. е. при интерференции встречных волн. Мы имеем при этом $\sin (\beta/2) = 1$ и

$$l_{\min} = \frac{\lambda}{2}. \quad (7.12)$$

Это, как мы знаем, есть расстояние между соседними пучностями (или узлами) в стоячей волне.

Из формулы (7.12) можно сделать достаточно общий вывод: в монохроматическом волновом поле характерный масштаб пространственных неоднородностей не может быть существенно меньшим длины волны λ .

При интерференции плоских волн одинаковой амплитуды $a_1 = a_2 = a$ из (7.9) имеем

$$I = 2I_0(1 + \cos(2kx \sin \alpha)), \quad (7.13)$$

где $I_0 = a^2$ — интенсивности слагаемых волн.

7.3. Квазимонохроматические волны

Элементарные представления о статистической природе излучения. В предыдущем параграфе мы говорили о волновых процессах, имеющих гармоническую зависимость от времени. Однако ни один реальный источник света не излучает волну, напряженность поля в которой меняется по гармоническому закону. Это подтверждается множеством экспериментальных фактов. Рассмотрим один из них: результат суперпозиции волн, излучаемых двумя независимыми источниками (например двумя электрическими лампочками).

Казалось бы, мы должны получить характерное чередование светлых и темных полос. Однако ничего подобного в действительности не происходит: если в комнате, где горит электрическая лампочка, зажечь еще одну точно такую же, то освещенность стен просто удваивается. Другими словами, при наложении света от двух лампочек складываются интенсивности и никаких интерференционных явлений не наблюдается.

Для того чтобы правильно ответить на возникающие вопросы и объяснить результат рассмотренного эксперимента (а также результаты множества других подобных фактов), необходимо отказаться от принятой ранее синусоидальной (гармонической) идеализации и более тщательно рассмотреть структуру оптического излучения.

Если бы мы имели столь малоинерционный прибор, что он мог бы следить за изменением напряженности электрического поля в световой волне, испущенной каким-либо источником, то мы получили бы запись *случайного процесса*, а вовсе не гармонический сигнал. Что мы понимаем под словами *случайный процесс*?

Возьмем большое количество (ансамбль) одинаковых источников света: одинаковых электрических лампочек или одинаковых кювет, наполненных нагретым газом (одинаковые температура, давление, вообще все макроскопические параметры).

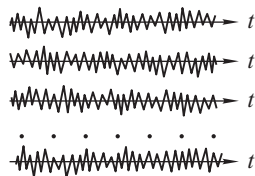


Рис. 7.5

Используя различные источники из ансамбля, мы будем получать *различные* записи колебательного процесса $E_n(t)$ (рис. 7.5).

Полученная совокупность записей представляет собой *ансамбль случайных функций*. Случайный характер процесса проявляется в том, что каждый раз мы получаем новую функцию $E_n(t)$, не повторяющую в точности ни одну из других функций; при этом невозможно предсказать заранее, какую точно из *реализаций ансамбля* мы получим в данном эксперименте. Любую из записей $E_n(t)$ можно характеризовать лишь *вероятностно* (определяя, например, вероятность, с которой то или иное значение принимает функция в любой момент времени). В этом состоит отличие *случайного* процесса, определяемого вероятностными законами, от процесса *детерминированного*, в котором каждое значение измеряемой величины абсолютно точно предсказуемо.

Статистический характер светового излучения обусловлен фундаментальными физическими свойствами источника. Всякий реальный источник — это совокупность огромного числа атомов. Каждый из атомов может стать элементарным источником излучения, если благодаря какому-либо механизму (столкновение с другим атомом, электронный удар и т. д.) он будет переведен в возбужденное состояние. Случайный, статистический характер механизма возбуждения определяет статистическую природу излучения. *Любой из атомов может начать высвечиваться в произвольный, случайный момент времени t_i* . Другое важное обстоятельство связано с характером излучения отдельного атома. *Возбужденный атом излучает конечное время*. С классической точки зрения атом можно рассматривать как колебательную

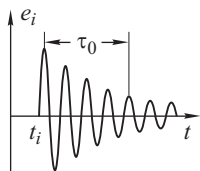


Рис. 7.6

систему (осциллятор) с определенной добротностью. Излучение такого осциллятора представляет собой процесс свободных затухающих колебаний. Мы приходим, таким образом, к картине, изображенной на рис. 7.6 (под $e_i(t)$ можно понимать изменение напряженности электрического поля в цуге излучения). Нужно отметить, что атомные системы обладают чрезвычайно малым затуханием (высокой добротностью), поэтому *длительность цуга τ_0* существенно больше периода колебания: $\tau_0 \gg T$ (характерные длительности цугов для оптического диапазона излучения $\tau_0 \sim 10^{-8} - 10^{-10}$ с, что на много порядков превышает период светового колебания $T \approx 10^{-15}$ с). Возбужденный атом, излучая цуг, постепенно теряет энергию и по прошествии времени порядка τ_0 переходит в основное, невозбужденное состояние; говорят, что *длительность цуга есть время жизни атома в возбужденном состоянии*. Последовательность цугов, излучаемых атомом какого-либо источника света, показана на рис. 7.7. Статистический характер

механизма возбуждения приводит к тому, что моменты возникновения излучаемых цугов t_1, t_2, \dots, t_n являются случайными независимыми величинами.

Наконец, важно иметь в виду, что в излучении одновременно принимает участие большое число атомов. Обозначая последовательность цугов, излучаемых каким-либо n -м атомом через $e_n(t)$, запишем суммарное излучение в виде

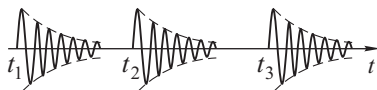


Рис. 7.7

$$E(t) = \sum_n e_n(t), \quad (7.14)$$

причем сумма (7.14) содержит в каждый момент времени огромное число «живых» (отличных от нуля) слагаемых.

Попробуем качественно описать свойства суммарного колебания (7.14), исходя из представления о цугах конечной продолжительности τ_0 и случайном характере их возникновения. Ясно, что вклад в суммарное колебание в некоторый момент времени t вносят лишь те цуги, начало которых t_n лежит в интервале от $t - \tau_0$ до t , поскольку все цуги, начавшиеся до момента $t - \tau_0$, успевают затухнуть к моменту t . Для уяснения качественной картины будем полагать, что число «живых» слагаемых в сумме (7.14) (т. е. число излучающих в каждый момент времени t атомов) равно N . Поскольку любой атом с *равной вероятностью* может начать излучать в любой момент времени t , то среднее число цугов, начало которых приходится на любой малый интервал времени Δt равно $n = N\Delta t/\tau_0$.

Таким образом, доля цугов, заканчивающихся за время Δt и заменяющихся новыми, равна $\Delta t/\tau_0$, причем через время $\Delta t \geq \tau_0$ в сумме (7.14) все N излучающих цугов заменяются новым набором.

Все сказанное выше можно пояснить, используя векторную диаграмму. Вектор, изображающий суммарное колебание (7.14), представляется на векторной диаграмме суммой большого числа

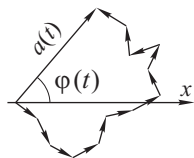


Рис. 7.8

элементарных векторов, соответствующих излучению отдельных атомов (рис. 7.8). Ориентация каждого элементарного вектора на этой диаграмме случайна. За время, малое по сравнению с длительностью цуга τ_0 , ориентация суммарного вектора и его длина изменятся незначительно, так как лишь небольшая часть элементарных векторов заменится другими. Через время порядка τ_0 возникает совершенно

новая совокупность элементарных векторов, что приводит к новым значениям длины a и угла наклона φ суммарного вектора, которые не связаны со старыми значениями этих величин.

Причем качественно ясно, что случайные функции $a(t)$ и $\varphi(t)$ могут изменяться *независимо* друг от друга. Действительно, при другой реализации направлений элементарных векторов может образоваться суммарный вектор той же длины $a(t)$, но имеющий произвольное направление $\varphi(t)$: угол наклона суммарного вектора никак не связан с его длиной. Можно сказать, что случайные функции $a(t)$ и $\varphi(t)$ *статистически независимы*.

Как следует из изложенного выше, амплитуда и фаза суммарного колебания испытывают случайные изменения с характерным временным масштабом τ_0 . Мы приходим, таким образом, к представлению о *квазимонохроматическом процессе* и можем записать колебание (7.14) в виде

$$E(t) = a(t) \cos(\omega_0 t - \varphi(t)), \quad (7.15)$$

где $a(t)$ и $\varphi(t)$ — медленно и хаотически меняющиеся функции. Характерный временной масштаб их изменения τ_0 носит название *времени корреляции* или *времени когерентности*. Принято говорить, что значения $a(t)$ и $a(t + \tau)$ *коррелированы* при $\tau < \tau_0$ и *некоррелированы* при $\tau > \tau_0$. Это же относится и к значениям фазы $\varphi(t)$ и $\varphi(t + \tau)$. Еще раз напомним: мы полагаем, что функции $a(t)$ и $\varphi(t)$ остаются практически неизменными на интервалах времени Δt , содержащих большое число периодов светового колебания $2\pi/\omega_0$. Именно в этом смысле понимается медленность их изменения и именно поэтому процесс (7.15) является квазимонохроматическим.

Итак, квазимонохроматическое колебание (7.15) изображается на векторной диаграмме в виде вектора с *медленно и хаотически* изменяющимися длиной и направлением. Этому вектору соответствует комплексная функция

$$f(t) = a(t)e^{i\varphi(t)}, \quad (7.16)$$

которая является аналогом комплексной амплитуды гармонического процесса и которую можно назвать комплексной амплитудой квазимонохроматического процесса (7.15). Мы будем пользоваться комплексным представлением квазимонохроматических колебаний, ставя в соответствие реальному процессу (7.15) комплексную функцию

$$V(t) = a(t)e^{-i(\omega t - \varphi(t))} = f(t)e^{-i\omega t}. \quad (7.17)$$

Подчеркнем, что в выражении (7.17) статистический (случайный) характер имеет первый сомножитель — комплексная амплитуда $f(t)$, второй же сомножитель $e^{-i\omega t}$ описывает гармоническое колебание с частотой ω .

Можно сказать, что $f(t)$ является медленно и хаотически меняющейся *огibaющей* в случайном процессе $V(t)$, т. е. функция

$f(t)$ описывает закон случайной амплитудно-фазовой модуляции гармонического процесса. Реальное возмущение $E(t)$ связано с $V(t)$ формулой

$$E(t) = \operatorname{Re} V(t) = \operatorname{Re} (f(t)e^{-i\omega t}). \quad (7.18)$$

Поскольку в дальнейшем мы будем использовать комплексное представление (7.17), необходимо установить соотношение между комплексной функцией $V(t)$ и результатами экспериментальных наблюдений.

Сделаем несколько замечаний о свойствах фотоприемника, с помощью которого проводится анализ светового излучения. Таким фотоприемником может быть глаз, фотопластинка, фотоэлемент и т. д. Все эти устройства регистрируют *поток энергии, усредненный за некоторое, характерное для данного фотоприемника, время ΔT , которое можно назвать временем регистрации (или постоянной времени фотоприемника)*. Поскольку световой поток (поток энергии) пропорционален квадрату напряженности поля $E^2(t)$ в волне, то регистрируется величина

$$I = \frac{1}{\Delta T} \int_{t-\Delta T/2}^{t+\Delta T/2} E^2(t) dt, \quad (7.19)$$

называемая *интенсивностью* света.

Постоянные времени разных приемников сильно различаются: у сетчатки глаза время регистрации $\Delta T \approx 10^{-1}$ с (глаз не успевает замечать мелькания кадров на экране кино и телевизора). Время экспозиции фотоматериалов обычно $\Delta T \approx \approx 10^{-2}$ – 10^{-4} с, хотя может быть и минуты, и даже часы. У некоторых типов фотоэлементов время регистрации $\Delta T \approx 10^{-6}$ – 10^{-8} с. Результат измерения существенно зависит от соотношения между временем экспозиции и характерным временным масштабом регистрируемых процессов.

Квазимонохроматические колебания характеризуются двумя временными масштабами: периодом «несущего колебания» $T_0 = 2\pi/\omega_0$ и временем когерентности τ_0 . В диапазоне видимого света, как отмечалось выше, $T_0 \approx 10^{-15}$ с, а время когерентности даже для самых узких спектральных линий не превышает значений $\tau_0 \approx 10^{-7}$ – 10^{-8} с (излучение когерентных источников света — лазеров — имеет время когерентности на несколько порядков больше). Таким образом, в оптике всегда $\Delta T \gg T_0$ — не существует фоторегистрирующих устройств, способных следить за отдельным периодом светового колебания. Однако при сравнении ΔT и τ_0 можно выделить два случая:

1. $\Delta T \lesssim \tau_0$. Такой фотоприемник может следить за случайными изменениями светового потока, обусловленными конечной длительностью цугов излучения атомов.

2. $\Delta T \gg \tau_0$. Фотоприемник регистрирует средний световой поток (среднюю интенсивность). Этот способ регистрации можно назвать инерционным. Именно этот случай реализуется в большинстве оптических экспериментов и именно его мы будем иметь в виду в дальнейшем.

Используя (7.15), перепишем (7.19) в виде

$$I = \frac{1}{\Delta T} \int_{t-\Delta T/2}^{t+\Delta T/2} a^2(t) \cos^2(\omega t - \varphi(t)) dt.$$

Выражая квадрат косинуса с помощью равенства $\cos^2 \alpha = (1/2)(1 + \cos 2\alpha)$, получим

$$I = \frac{1}{2\Delta T} \int_{t-\Delta T/2}^{t+\Delta T/2} a^2(t) dt + \frac{1}{2\Delta T} \int_{t-\Delta T/2}^{t+\Delta T/2} a^2(t) \cos(2\omega_0 t - 2\varphi(t)) dt. \quad (7.20)$$

За время наблюдения ΔT подынтегральная функция во втором слагаемом в формуле (7.20), осциллируя с удвоенной световой частотой, многократно меняет знак, поэтому второй интеграл пренебрежимо мал в сравнении с первым, тогда

$$I = \frac{1}{2} \overline{a^2(t)} = \frac{1}{2\Delta T} \int_{t-\Delta T/2}^{t+\Delta T/2} a^2(t) dt$$

(черта означает усреднение по времени регистрации). С другой стороны, используя (7.16), имеем

$$\overline{f(t)f^*(t)} = \overline{a(t)e^{i\varphi(t)}a(t)e^{-i\varphi(t)}} = \overline{a^2(t)}.$$

Таким образом, наблюдаемая интенсивность выражается через комплексную амплитуду $f(t)$ равенством

$$I = \frac{1}{2} \overline{f(t)f^*(t)}.$$

Очевидно теперь, что интенсивность можно выразить аналогичным образом и через комплексную функцию $V(t)$ (используя (7.17)):

$$I = \frac{1}{2} \overline{V(t)V^*(t)}.$$

Будем в дальнейшем определять интенсивность, как и раньше, равенством $I = \overline{a^2(t)}$ (без множителя $1/2$); тогда

$$I = \overline{a^2(t)} = \overline{f(t)f^*(t)} = \overline{V(t)V^*(t)}. \quad (7.21)$$

7.4. Функция когерентности

Интенсивность — одна из характеристик светового излучения в фиксированной точке наблюдения — не является исчерпывающей. Действительно, рассмотрим два случайных процесса $V_1(t)$ и $V_2(t)$, показанных на рис. 7.9. Пусть средняя интенсивность за время регистрации ΔT у этих двух процессов одинакова. Однако качественно ясно различие между ними: первый процесс $V_1(t)$ — это «медленный» процесс, второй, $V_2(t)$, — «быстрый» процесс: изменения функции $V_2(t)$ происходят существенно быстрее, чем функции $V_1(t)$. Итак, кроме величины средней интенсивности, определяемой равенством (7.21), нужна также и характеристика, которая была бы связана с быстротой изменений случайной функции $f(t)$. Такой характеристикой является функция $\Gamma(\tau)$, определяемая равенством:

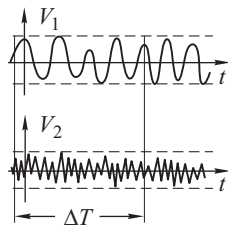


Рис. 7.9

$$\Gamma(\tau) = \overline{V(t)V^*(t+\tau)} = \overline{f(t)f^*(t+\tau)} e^{i\omega\tau}. \quad (7.22)$$

В теории случайных процессов эта функция называется *функцией корреляции*, а в оптике — *временной функцией когерентности* случайного процесса $V(t)$, описывающего хаотический характер световых колебаний (хаотические изменения амплитуды $a(t)$ и фазы $\varphi(t)$).

Как мы увидим в дальнейшем, она определяет результат интерференции квазимонохроматических волн и является *количественной мерой корреляции (когерентности)*, качественные представления о которой были введены при обсуждении статистической природы светового излучения.

Используя определение (7.22), найдем

$$\Gamma(\tau) = e^{i\omega\tau} \frac{1}{\Delta T} \int_{t-\Delta T/2}^{t+\Delta T/2} f(t)f^*(t+\tau) dt. \quad (7.23)$$

Определение $\Gamma(\tau)$ включает в себя следующую последовательность операций: нужно взять значение $f(t)$ в момент времени t_i и умножить на значение функции $f^*(t)$ в момент времени $t_i + \tau$. Перебираем всевозможные значения t_i (при фиксиро-

ванном τ) на интервале времени измерения от $t - \Delta T/2$ до $t + \Delta T/2$. Найдя их среднее, получим значение функции $\Gamma(\tau)$ при заданном фиксированном значении τ .

Мы полагаем, что от положения интервала усреднения ΔT на оси времени функция Γ не зависит, т. е. световое излучение представляет собой *стационарный случайный процесс*. Это означает, что физические (макроскопические) параметры, определяющие свойства источника излучения (температура, давление и т. д.), а также условия на пути распространения света от источника к точке наблюдения остаются неизменными во времени.

Исследуем характер поведения функции $\Gamma(\tau)$. При $\tau = 0$ функция $\Gamma(\tau)$ равна интенсивности света

$$\Gamma(0) = \frac{1}{\Delta T} \int_{t-\Delta T/2}^{t+\Delta T/2} f(t)f^*(t) dt = I. \quad (7.24)$$

Отметим, что (7.24) — это сумма действительных положительных чисел $f(t_i)f^*(t_i)\Delta t_i = a^2(t_i)\Delta t_i$, т. е. изображающие отдельные колебания вектора на векторной диаграмме выстроены коллинеарно вдоль действительной оси. Ясно, что суммарный вектор $\Gamma(0)$ имеет при этом максимально возможную длину. Таким образом, функция $\Gamma(\tau)$ имеет при $\tau = 0$ максимально возможное значение, равное интенсивности света. Как изменяется это значение с ростом τ ? Ясно, что пока $\tau \ll \tau_0$ — времени когерентности (но может быть значительно больше $T_0 = 2\pi/\omega$), аргумент комплексной функции $f(t)f^*(t+\tau) = a(t)a(t+\tau)e^{i(\varphi(t)-\varphi(t+\tau))}$, равный $\Delta\varphi = \varphi(t) - \varphi(t+\tau)$, близок к нулю при любом t на интервале усреднения ΔT , так как $\varphi(t+\tau)$ хорошо «помнит» значение фазы $\varphi(t)$ в предшествующий момент времени и повторяет эти изменения почти в точности: суммарное колебание $V(t)$ состоит почти из той же совокупности цугов, что и $V(t+\tau)$.

Итак, $\varphi(t+\tau) - \varphi(t) \approx 0$, аналогичным образом $a(t+\tau) \approx a(t)$, поэтому $f(t)f^*(t+\tau) \approx a^2(t) = I_0$ и $\Gamma(\tau) \approx I_0 e^{i\omega\tau}$ при $\tau \ll \tau_0$.

Мы пришли к выводу, что *при $\tau \ll \tau_0$ функция $\Gamma(\tau)$ имеет модуль, мало отличающийся от интенсивности, и аргумент, близкий к $\omega\tau$* .

При $\tau \geq \tau_0$ значения фаз $\varphi(t)$ и $\varphi(t+\tau)$ некоррелированы, т. е. при заданном $\varphi(t)$ фаза $\varphi(t+\tau)$ может с равной вероятностью принимать любые значения от нуля до 2π , следовательно и разность фаз $\Delta\varphi = \varphi(t) - \varphi(t+\tau)$ может быть любой, принимая с равной вероятностью (за время наблюдения $\Delta T \gg \tau_0$) все возможные значения. Величина $e^{i\Delta\varphi} = \cos \Delta\varphi + i \sin \Delta\varphi$ оказывается близкой к нулю.

Таким образом,

$$\Gamma(\tau) = \overline{f(t)f^*(t+\tau)} e^{i\omega\tau} \approx 0 \text{ при } \tau \gtrsim \tau_0.$$

Итак, введенная нами функция $\Gamma(\tau)$ имеет «правильное поведение», которое соответствует нашим качественным представлениям о сильной корреляции при $\tau \ll \tau_0$ (при этом $\Gamma(\tau) \approx I_0 \times \times e^{i\omega\tau}$) и об отсутствии корреляции при $\tau \gtrsim \tau_0$ (при этом $\Gamma(\tau) \approx 0$). Функция $\Gamma(\tau)$ и является количественной мерой коррелированности значений случайного процесса $V(t)$ в два момента времени разделенных промежутком τ .

Подчеркнем, что в результате усреднения мы получаем *детерминированную* функцию $\Gamma(\tau)$, которая является одной из важнейших характеристик стационарного случайного процесса $f(t)$. Эта функция определяет закон, по которому постепенно уменьшается когерентность с ростом τ .

Удобно, наряду с (7.22), ввести нормированную функцию

$$\gamma(\tau) = \frac{\Gamma(\tau)}{\Gamma(0)} = \frac{\overline{f(t)f^*(t+\tau)}}{I_0} e^{i\omega\tau}, \quad (7.25)$$

которая изменяется от единицы при $\tau = 0$ до нуля при $\tau \gtrsim \tau_0$. Функцию $\gamma(\tau)$ называют *комплексной степенью когерентности случайного процесса* $V(t)$.

Понятие временной когерентности, введенное выше для случайного колебательного процесса $V(t)$, можно обобщить, введя *функцию взаимной когерентности* двух случайных процессов $V_1(t)$ и $V_2(t)$:

$$\Gamma_{12}(\tau) = \overline{V_1(t)V_2^*(t+\tau)},$$

которая является количественной мерой когерентности двух процессов, разделенных промежутком времени τ . Процессы $V_1(t)$ и $V_2(t+\tau)$ называются некогерентными, если $\Gamma_{12}(\tau) \equiv 0$ при любом τ .

Рассмотрим вопрос о когерентности колебаний, созданных в некоторой точке наблюдения P волнами, приходящими от двух независимых источников S_1 и S_2 (рис. 7.10).

Каждая из волн образована своей совокупностью цугов: волна V_1 образована цугами, излученными атомами источника S_1 , а волна V_2 состоит из цугов, излученных атомами источника S_2 . Так как атомы S_1 излучают независимо от атомов S_2 , то моменты возникновения цугов, составляющих волну V_1 , никак не связаны с моментами возникновения цугов в волне V_2 . Это означает, что положение каждого элементарного вектора в совокупности, образующей суммарный вектор V_1 , не связано с положением каждого элементарного вектора в совокупности V_2 (рис. 7.11).

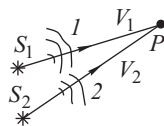


Рис. 7.10

Следовательно, и положения суммарных векторов V_1 и V_2 никак не связаны между собой: разность фаз $\varphi_1 - \varphi_2$ может быть любой; длины векторов (амплитуды колебаний V_1 и V_2) также независимы.

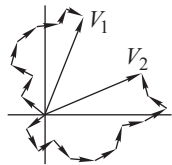


Рис. 7.11

Выше мы говорили, что корреляция (когерентность) в каждой из волн сохраняется в течение времени порядка τ_0 , значит, через время $\sim \tau_0$ каждое из колебаний V_1 и V_2 образуется уже новой совокупностью не связанных между собой цугов, так что через время порядка τ_0 разность фаз $\varphi_1 - \varphi_2$ изменяется случайным образом.

Таким образом, при любом τ (в том числе и при $\tau = 0$)

$$\overline{V_1(t)V_2^*(t+\tau)} \equiv 0,$$

т. е. волны, излучаемые двумя *независимыми* источниками света, *некогерентны*.

7.5. Интерференция квазимонохроматических волн

Пусть колебание в некоторой точке наблюдения создано суперпозицией двух квазимонохроматических волн E_1 и E_2 вида (7.15). Тогда суммарный колебательный процесс в этой точке имеет вид

$$E(t) = E_1(t) + E_2(t).$$

Регистрируемая приемником света интенсивность, согласно (7.21), есть

$$I = \overline{E^2(t)} = \overline{a^2(t)},$$

где $a(t)$ — амплитуда результирующего колебания $E(t)$. Результирующий колебательный процесс представляет собой хаотически модулированное колебание с медленно меняющимися (в масштабе периода $T_0 = 2\pi/\omega_0$) амплитудой $a(t)$ и фазой $\varphi(t)$

$$E(t) = a(t) \cos(\omega_0 t + \varphi(t)).$$

Ясно, что поскольку в слагаемых колебаниях амплитуды $a_1(t)$ и $a_2(t)$, а также фазы $\varphi_1(t)$ и $\varphi_2(t)$ сохраняются неизменными на протяжении большого числа периодов светового колебания, то и амплитуда, и фаза суммарного колебания также практически неизменны на протяжении большого числа периодов (их характерное время изменения во всяком случае не меньше времени когерентности). Мы уже познакомились в § 1.3

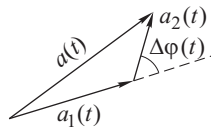


Рис. 7.12

с тем, что квазигармонические колебания можно складывать по правилу сложения векторов так же, как строго гармонические колебания (рис. 7.12).

Используя теорему косинусов, получаем

$$a^2(t) = a_1^2(t) + a_2^2(t) + 2a_1a_2 \cos \Delta\varphi(t), \quad (7.26)$$

где $\Delta\varphi(t) = \varphi_1(t) - \varphi_2(t)$.

Проводя усреднение последнего равенства в интервале времени регистрации (которое, напомним, много больше времени когерентности), находим

$$I = I_1 + I_2 + 2\overline{a_1(t)a_2(t) \cos \Delta\varphi(t)}, \quad (7.27)$$

где $I_1 = \overline{a_1^2(t)}$ и $I_2 = \overline{a_2^2(t)}$ — интенсивности слагаемых волн, а $I = \overline{a^2(t)}$ — интенсивность результирующего колебания.

Используя комплексные представления (7.16) и (7.17), легко получить

$$\overline{V_1(t)V_2^*(t)} = \overline{a_1(t)a_2(t)e^{i\Delta\varphi(t)}}. \quad (7.28)$$

Третье слагаемое в (7.26) можно теперь представить, используя (7.28) в виде

$$\overline{2a_1(t)a_2(t) \cos \Delta\varphi(t)} = 2 \operatorname{Re} \overline{V_1(t)V_2^*(t)}.$$

Итак,

$$I = I_1 + I_2 + 2 \operatorname{Re} \overline{V_1(t)V_2^*(t)}. \quad (7.29)$$

Последняя формула представляет собой наиболее общее выражение, справедливое для любой двухлучевой интерференционной схемы. Наибольший интерес представляет интерференционное слагаемое $2 \operatorname{Re} \overline{V_1(t)V_2^*(t)}$. Именно с ним связано отличие суммарной интенсивности I от суммы интенсивностей $I_1 + I_2$, т. е. интерференционный эффект.

Первый важный вывод, который можно сделать из (7.29), состоит в том, что *некогерентные волны не могут интерферировать*. Действительно, при $\overline{V_1(t)V_2^*(t)} = 0$ мы имеем $I = I_1 + I_2$, т. е. при наложении некогерентных волн имеет место закон сложения интенсивностей. В частности, если волны V_1 и V_2 испущены независимыми источниками, то они некогерентны и не могут давать интерференционных эффектов. Интерференция может наблюдаться лишь в том случае, если волны по крайней мере частично когерентны. Поэтому в интерференционных схемах всегда используют один источник света, а две (или несколько) интерферирующих волн получают путем деления (расщепления) одной волны, излучаемой источником. На рис. 7.13 приведены примеры нескольких интерференционных схем, широко применяемых в оптике: *а* — опыт Юнга, *б* — схема Ллойда, *в* — бипризма Френеля, *г* — билинза Бийе, *д* — интерферометр Майкельсона, *е* — звездный интерферометр Майкельсона: лучи 1 и 2 идут от удаленного источника. Во всех схемах излучение источника *S*

попадает на экран, где наблюдается интерференция, по двум различным путям (по двум «плечам» интерференционной схемы), отмеченным на рисунках индексами 1 и 2.

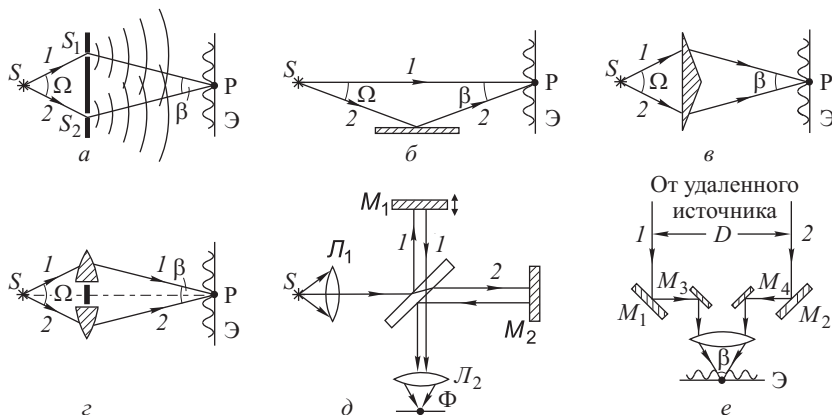


Рис. 7.13

Наиболее важными характеристиками любой интерференционной схемы являются: угол схождения волн β в точке наблюдения и угол Ω между лучами, исходящими из источника, которые каким-либо способом сводятся далее в точку наблюдения. Этот угол Ω называется *апертурой интерференции*.

Пусть $V(t)$ — это колебательный процесс вблизи источника S и пусть $V_1(t)$ — это колебательный процесс в точке наблюдения P , созданный волной, пришедшей от источника по пути 1. Соответственно $V_2(t)$ — колебание в точке наблюдения, созданная второй волной (пришедшей по пути 2). Тогда:

$$V_1(t) = \alpha_1 V(t - \tau_1); \quad V_2(t) = \alpha_2 V(t - \tau_2), \quad (7.30)$$

т. е. колебательный процесс $V_1(t)$ в точности повторяет колебание $V(t)$ вблизи источника, лишь с запаздыванием на время τ_1 , и с уменьшенной амплитудой, то же касается и процесса $V_2(t)$; $\tau_1 = r_1/c$ и $\tau_2 = r_2/c$ — времена распространения волны от источника до точки наблюдения (r_1 — путь, проходимый волной вдоль первого плеча интерференционной схемы, а r_2 — путь волны V_2 по второму «плечу»). Коэффициенты пропорциональности α_1 и α_2 учитывают ослабления волн из-за сферической расходимости или поглощения. Используя (7.30), получаем

$$\overline{V_1(t)V_2^*(t)} = \alpha_1 \alpha_2 \overline{V(t - \tau_1)V^*(t - \tau_2)}.$$

Мы полагаем, что условия стационарности выполнены, поэтому интервал усреднения можно сместить, т. е. середину ин-

тервала $[t - \Delta T/2, t + \Delta T/2]$ — момент t , — сместить в точку $t' = t - \tau_1$. Тогда

$$\overline{V_1(t)V_2^*(t)} = \alpha_1\alpha_2\overline{V(t')V^*(t' + \tau)} = \alpha_1\alpha_2\Gamma(\tau),$$

где $\tau = \tau_1 - \tau_2$ — относительное запаздывание, зависящее от разности длин плеч $\Delta = r_1 - r_2$ интерференционной схемы: $\tau = \Delta/c$.

Таким образом, интерференционный эффект определяется функцией *временной когерентности* колебательного процесса $V(t)$.

Используя (7.29), запишем

$$I = I_1 + I_2 + 2\alpha_1\alpha_2 \operatorname{Re} \Gamma(\tau). \quad (7.31)$$

Так как $I_1 = \alpha_1^2 I_0$ и $I_2 = \alpha_2^2 I_0$, где I_0 — интенсивность света вблизи источника, то $\alpha_1\alpha_2 = \sqrt{I_1 I_2} / I_0$ и вместо (7.31) получаем

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} \operatorname{Re} \gamma(\tau), \quad (7.32)$$

где $\gamma(\tau)$ — комплексная степень когерентности — нормированная функция (см. (7.25)), которая изменяется от 1 при $\tau = 0$ до нуля при $\tau \gtrsim \tau_0$.

При равных интенсивностях обеих волн $I_1 = I_2 = I_0$ формула (7.32) принимает вид

$$I = 2I_0 (1 + \operatorname{Re} \gamma(\tau)). \quad (7.33)$$

Комплексную функцию $\gamma(\tau)$ запишем в виде $\gamma(\tau) = \frac{|\gamma(\tau)|e^{i(\omega\tau + \theta(\tau))}}{f(t) \cdot f^*(t + \tau)}$, где θ — аргумент комплексной функции $f(t) \cdot f^*(t + \tau)$, т. е. $\theta = \varphi(t) - \varphi(t + \tau)$. Тогда

$$I(\tau) = 2I_0 [1 + |\gamma(\tau)| \cos(\omega\tau + \theta(\tau))], \quad (7.34)$$

причем $\theta(0) = 0$. Значит $I_{\max} = 2I_0(1 + |\gamma(\tau)|)$, $I_{\min} = 2I_0(1 - |\gamma(\tau)|)$ и для *видности* находим:

$$V = |\gamma(\tau)|.$$

Общие выводы, которые можно сделать из (7.34), состоят в следующем.

1. Пусть временное запаздывание одной волны относительно другой превышает время когерентности $\tau \gtrsim \tau_0$, соответственно разность хода волн $\Delta \gtrsim c\tau_0$. В этом случае $\gamma(\tau) \approx 0$ и $I = I_1 + I_2$. Таким образом, при большом временном запаздывании ($\tau \gtrsim \tau_0$) или, другими словами, при большой разности хода $\Delta \gtrsim c\tau_0$ волны перестают быть когерентными и не интерferируют (хотя эти волны и созданы одним источником света).

2. Пусть $\tau < \tau_0$. Тогда $\gamma(\tau) \neq 0$ и $I \neq I_1 + I_2$: волны частично когерентны, наблюдается интерференция.

3. Если $\tau \ll \tau_0$, то $\gamma(\tau) \approx 1$ и вместо (7.34) получаем

$$I(\tau) = 2I_0(1 + \cos \omega\tau),$$

где $\tau = \Delta/c$ — временное запаздывание, Δ — разность хода. Точно такое же выражение получается, если рассматривать строго гармонические волны. Мы получили, таким образом, ответ на вопрос о применимости синусоидальной идеализации: при сложении когерентных волн все происходит так, как если бы волны были строго гармоническими.

Общий характер зависимости интенсивности в интерференционной картине от $\tau = \Delta/c$ (т. е. при изменении положения точки наблюдения), показан на рис. 7.14.

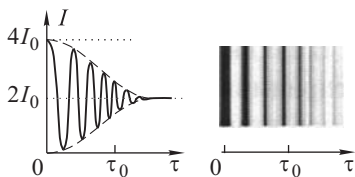


Рис. 7.14

С ростом τ постепенно уменьшается видность (контраст) полос. Наконец, при $\tau \gtrsim \tau_0$ полосы вовсе исчезают и мы приходим к закону сложения интенсивностей (при равных интенсивностях $I = I_0 + I_0 = 2I_0$).

Подчеркнем, что поскольку в квазимонохроматической волне $\tau \gg T_0$, то быстро меняющийся при изменении τ множитель $\cos(\omega_0\tau + \theta(\tau))$ (его значение меняется от 1 до 0 при изменении τ на $T_0/2$) «промодулирован» медленно спадающей функцией $|\gamma(\tau)|$ ($|\gamma(\tau)|$ спадает от единицы до нуля при изменении τ от нуля до τ_0), т. е. мы наблюдаем множество переходов от светлой полосы к темной (т. е. множество интерференционных полос) прежде, чем они окончательно исчезнут.

Разумеется, точный ход зависимости $I(\tau)$ определяется конкретным видом функции $\gamma(\tau)$, т. е. в конечном счете свойствами источника, от которого зависит форма цугов излучения. Рисунок 7.14 дает качественное представление об этой зависимости, достаточное, впрочем, для проведения некоторых оценок. Можно заключить, что максимальная разность хода, при которой еще видна интерференция $\Delta_{\max} \approx c\tau_0$, а максимальный порядок интерференции

$$m_{\max} = \frac{\Delta_{\max}}{\lambda} = \frac{c\tau_0}{\lambda} = \frac{\tau_0}{T_0}.$$

Величину $l = c\tau_0$ принято называть *длиной цуга* (это пространственная протяженность цуга излучения длительности τ_0).

Таким образом, максимально допустимая разность хода ограничена длиной цуга, максимальный порядок интерференции — числом периодов колебаний, содержащихся в цуге излучения: $m_{\max} \approx \tau_0/T_0$.

7.6. Связь функции когерентности и распределения энергии по спектру. Соотношение неопределенностей

Изучая распределение интенсивности в интерференционной картине, можно определить функцию когерентности квазимонохроматической волны, излучаемой тепловым источником света — таков вывод из выражения (7.34).

С другой стороны, о свойствах излучения можно судить по спектральным измерениям, исследуя с помощью спектрального прибора распределение энергии по спектру квазимонохроматической волны в фиксированной точке наблюдения. Важно понять, в каком соотношении находятся эти способы измерения, какова связь между функцией когерентности и энергетическим спектром?

Распределение энергии по спектру квазимонохроматического колебания $\mathcal{I}(\omega)$ можно определить например с помощью призменного спектрографа (рис. 7.15).

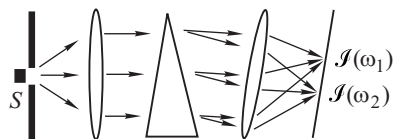


Рис. 7.15

Параллельный пучок квазимонохроматического света падает на призму и, благодаря дисперсии — зависимости показателя преломления от частоты, разделяется на множество монохроматических пучков, каждый с определенной частотой ω . Интенсивность каждого из этих пучков $\mathcal{I}(\omega)$, определяющую распределение энергии по частоте, называют *спектральной интенсивностью*. Сумма интенсивностей всех монохроматических пучков равна, очевидно, интенсивности квазимонохроматического света, попадающего на входную щель спектрографа:

$$I = \int \mathcal{I}(\omega) d\omega. \quad (7.35)$$

Смысл соотношения (7.35) состоит в том, что полная интенсивность света может быть получена путем аддитивного сложения интенсивностей отдельных монохроматических волн, составляющих суммарную квазимонохроматическую волну: другими словами, *спектральные гармоники разных частот некогерентны между собой*. Следовательно, интерференционная картина, получаемая с помощью источника квазимонохроматического излучения, может быть рассчитана следующим образом: нужно взять отдельную монохроматическую компоненту частоты ω (и интенсивности $\mathcal{I}(\omega) d\omega$) и найти интерференционную картину, создаваемую этой компонентой, используя формулу (7.8)

$$dI = 2\mathcal{I}(\omega) d\omega \left[1 + \cos \left(\frac{\omega}{c} \Delta \right) \right], \quad (7.36)$$

после чего некогерентно сложить вклады:

$$I(\Delta) = 2 \int \mathcal{J}(\omega) \left[1 + \cos \left(\frac{\omega}{c} \Delta \right) \right] d\omega. \quad (7.37)$$

Очевидно, результаты расчетов, основанные на спектральном подходе (формула (7.37)) и временном подходе (формула (7.33)) должны быть тождественными. Исходя из этого, мы можем связать спектральную интенсивность $\mathcal{J}(\omega)$ с функцией временной когерентности $\Gamma(\tau)$.

Наиболее важной является связь между шириной спектра $\Delta\omega$ (интервалом частот, в котором спектральная интенсивность $\Delta\omega$ заметно отличается от нуля) и временем когерентности τ_0 (интервал значений τ , в котором отлична от нуля функция когерентности $\Gamma(\tau)$).

Рассмотрим простой пример. Пусть функция $\mathcal{J}(\omega)$ имеет вид (рис. 7.16)

$$\mathcal{J}(\omega) = \begin{cases} \mathcal{J}_0 & \text{при } |\omega - \omega_0| < \Delta\omega/2 \\ 0 & \text{при } |\omega - \omega_0| > \Delta\omega/2 \end{cases}.$$

Тогда, используя (7.37), найдем

$$I = 2\mathcal{J}_0 \int_{\omega_0 - \Delta\omega/2}^{\omega_0 + \Delta\omega/2} \left(1 + \cos \left(\frac{\omega}{c} \Delta \right) \right) d\omega.$$

В результате интегрирования получим

$$I = 2I_0 \left\{ 1 + \frac{\sin [(\Delta\omega/2c) \Delta]}{(\Delta\omega/2c) \Delta} \cos \left(\frac{\omega_0}{c} \Delta \right) \right\}. \quad (7.38)$$

где $I_0 = \mathcal{J}_0 \cdot \Delta\omega$ — интенсивность интерферирующих волн (т. е. суммарная интенсивность всех спектральных компонент).

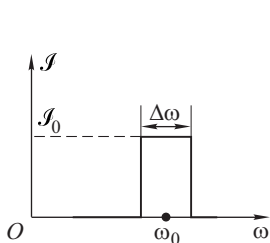


Рис. 7.16

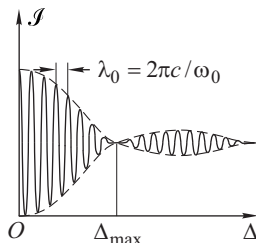


Рис. 7.17

Функция $\mathcal{J}(\Delta)$ изображена на рис. 7.17. Здесь быстро изменяющийся множитель $\cos [(\omega_0/c) \Delta]$ промодулирован медленным множителем $\frac{\sin [(\Delta\omega/2c) \Delta]}{(\Delta\omega/2c) \Delta}$ (так как $\Delta\omega \ll \omega_0$). Максимально до-

пустимую (для наблюдения интерференционной картины) разность хода Δ_{\max} можно оценить из условия: $(\Delta\omega/2c)\Delta_{\max} = \pi$, здесь Δ_{\max} — значение Δ , при котором функция «огibaющей» (выражение в квадратных скобках) обращается в нуль:

$$\Delta_{\max} \approx \frac{2\pi c}{\Delta\omega}. \quad (7.39)$$

Максимальный порядок интерференции m_{\max} есть

$$m_{\max} \approx \frac{\Delta_{\max}}{\lambda_0} = \frac{2\pi c}{\lambda_0 \Delta\omega}. \quad (7.40)$$

Сопоставляя выражения для максимально допустимой разности хода, полученные при временном ($\Delta_{\max} \approx c\tau_0$) и спектральном (формула (7.39)) подходах, находим связь между шириной спектра $\Delta\omega$ и временем корреляции τ_0

$$\tau_0 \Delta\omega \approx 2\pi. \quad (7.41)$$

Сужение спектра $\Delta\omega$ приводит, согласно (7.41), к увеличению времени когерентности τ_0 , т. е. к увеличению длительности цугов; наоборот, чем короче цуги, тем шире спектр — таков смысл соотношения (7.41), называемого *соотношением неопределенностей*.

Выражая ширину спектрального интервала в длинах волн $\Delta\lambda$, имеем для m_{\max} и Δ_{\max} следующие соотношения:

$$m_{\max} = \frac{\lambda_0}{\Delta\lambda}; \quad \Delta_{\max} = \frac{\lambda_0^2}{\Delta\lambda}. \quad (7.42)$$

Приведем числовые оценки m_{\max} и Δ_{\max} для некоторых реальных источников.

1. Дуговой фонарь без фильтра, визуальное наблюдение

$$\Delta\lambda = 2,5\text{--}3 \cdot 10^{-5} \text{ см}, \quad \lambda_0 = 5 \cdot 10^{-5} \text{ см};$$

$$m_{\max} \approx 2; \quad \Delta_{\max} \approx 10^{-4} \text{ см}.$$

2. То же со стеклянным светофильтром

$$\Delta\lambda \approx 300 \text{ Å} = 3 \cdot 10^{-6} \text{ см}, \quad \lambda_0 \approx 5 \cdot 10^{-5} \text{ см};$$

$$m_{\max} \approx 20, \quad \Delta_{\max} \approx 10^{-3} \text{ см}.$$

3. Узкая спектральная линия

$$\Delta\lambda \approx 10^{-3} \text{ Å}, \quad m_{\max} \approx 5 \cdot 10^6; \quad \Delta_{\max} \approx 2,5 \text{ м}.$$

4. Излучение стабилизированного газового лазера

$$m_{\max} \approx 10^{10}; \quad \Delta_{\max} \approx 5 \text{ км}.$$

Соотношение неопределенностей (7.41) получилось не случайно. Оно является следствием общей связи, существующей

между функцией когерентности $\Gamma(\tau)$ и спектральной интенсивностью $\mathcal{I}(\omega)$. К нахождению этой связи мы и перейдем.

Вернемся к соотношению (7.37), представив правую часть равенства в виде

$$I(\Delta) = 2 \int \mathcal{I}(\omega) d\omega + 2 \int \mathcal{I}(\omega) \cos\left(\frac{\omega}{c} \Delta\right) d\omega.$$

Интеграл в первом слагаемом представляет собой суммарную интенсивность света в каждой из интерферирующих волн:

$$\int \mathcal{I}(\omega) d\omega = I_0,$$

(т. е. первое слагаемое $2I_0$ — это сумма интенсивностей слагаемых волн). Второе слагаемое перепишем в виде

$$2 \int \mathcal{I}(\omega) \cos\left(\frac{\omega}{c} \Delta\right) d\omega = 2 \operatorname{Re} \int \mathcal{I}(\omega) e^{i(\omega/c)\Delta} d\omega.$$

Таким образом, вместо (7.37) имеем

$$I(\Delta) = 2I_0 \left(1 + \frac{\operatorname{Re} \int \mathcal{I}(\omega) e^{i(\omega/c)\Delta} d\omega}{I_0} \right).$$

Сравнивая полученное выражение с (7.33), найдем

$$\int \mathcal{I}(\omega) e^{i(\omega/c)\Delta} d\omega = I_0 \gamma(\tau) = \Gamma(\tau).$$

Итак, функция $\mathcal{I}(\omega)$ — распределение энергии по спектру — является преобразованием Фурье функции когерентности $\Gamma(\tau)$

$$\mathcal{I}(\omega) = \int \Gamma(\tau) e^{-i\omega\tau} d\tau. \quad (7.43)$$

Понятно, что соотношение неопределенностей (7.41), связывающее ширину спектра $\Delta\omega$ со временем когерентности, является следствием общей связи (7.43).

Замечание. Качественно ясно, почему с ростом ширины спектра излучения контраст интерференционных полос уменьшается. Это связано с тем, что полосы, отвечающие разным частотам (разным длинам волн), имеют разную ширину ($l \approx \lambda/\beta$). Полосы, отвечающие бóльшим длинам волн, более широкие, поэтому с ростом порядка полосы m (т. е. с ростом разности хода $\Delta = m\lambda$) смещение полос одинакового порядка m , отвечающее разным λ , постепенно нарастает. Наконец, наступает ситуация, когда на $(m+1)$ -ю полосу длины волны λ налагается m -я полоса длины волны $\lambda + \Delta\lambda$, т. е.

$$\Delta = (m+1)\lambda = m(\lambda + \Delta\lambda).$$

При дальнейшем росте Δ полосы практически исчезают: на минимум интенсивности одной длины волны налагается максимум какой-либо другой длины волны из спектрального диапазона излучения. Суммарная картина интенсивности усредняется. Из последнего равенства получаем уже знакомую нам оценку максимального порядка интерференции $m_{\max} = \lambda/\Delta\lambda$.

7.7. Пространственная когерентность и интерференционные явления при использовании протяженных источников

До сих пор речь шла о том, как на интерференционных явлениях сказывается отличие реальной волны от строго монохроматической. При этом источники света считались точечными (содержащими однако огромное число излучающих атомов).

Понятие «точечный источник» — идеализация, ничуть не меньшая, чем понятие «монохроматический источник». Все источники света имеют определенную протяженность, а это приводит к возникновению интерференционной картины меньшей контрастности, чем это следует из приведенных ранее расчетов. Как показывает опыт, при увеличении размеров источника видность интерференционной картины постепенно уменьшается и при достаточно больших размерах интерференционные полосы исчезают вовсе. Легко качественно понять характер изменения видности полос при изменении размеров источника, если исходить из представления о том, что протяженный источник света — это совокупность большого количества *независимых* источников (рис. 7.18) (действительно, каждая малая площадка источника ΔS_i состоит из атомов, излучающих независимо от атомов другой площадки, ΔS_j ($i \neq j$)), и таким образом волны, излучаемые этими площадками, V_i и V_j некогерентны:

$$\overline{V_i(t)V_j^*(t)} \equiv 0 \text{ при } i \neq j.$$

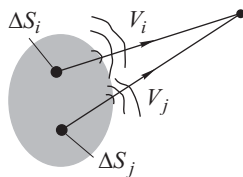


Рис. 7.18

Некогерентность излучения разных точек (малых элементов ΔS) протяженного источника означает, что при расчете интерференционной картины необходимо найти распределение интенсивности, создаваемое каждой точкой, а затем сложить интенсивности.

Пусть источник S и плоскость наблюдения в опыте английского физика Т. Юнга (1773–1829) удалены от экрана с отверстиями S_1 и S_2 настолько, что волны, приходящие от каждого малого элемента источника к этим отверстиям, а также волны, приходящие к плоскости наблюдения от этих отверстий,

можно считать плоскими. Тогда, очевидно, интерференционная картина, создаваемая крайней точкой B источника (штриховая кривая I_B на рис. 7.19), смещена относительно интерференционной картины от центральной

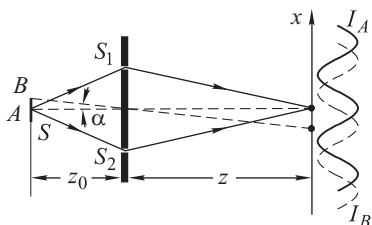


Рис. 7.19

точки A (сплошная кривая I_A) на расстояние, равное αz . Если это смещение меньше ширины интерференционной полосы $l \approx \lambda/\beta$ ($\beta = d/z$ — угол схождения интерферирующих волн, d — расстояние между отверстиями S_1 и S_2), то интерференционная картина наблюдается.

В противном случае (при $\alpha > \lambda/d$) интерференция не наблюдается: суммарная интенсивность уже не имеет чередующихся максимумов и минимумов. Это связано с тем, что на темную полосу (минимум интенсивности от одной точки источника) накладывается светлая полоса (максимум интенсивности от другой точки), и суммарная интенсивность оказывается константой — получается равномерная освещенность плоскости наблюдения. Итак, имеем оценку для максимально допустимой ширины источника b :

$$b \lesssim \frac{\lambda}{d} z_0 = \frac{\lambda}{\Omega}, \quad (7.44)$$

или при заданной ширине источника, максимально допустимое расстояние между отверстиями d (которое определяет «апертуру интерференции» $\Omega \approx d/z_0$)

$$d \lesssim \frac{\lambda}{b} z_0 \quad \text{или} \quad \Omega \lesssim \frac{\lambda}{b}. \quad (7.45)$$

Мы привели лишь качественные рассуждения, которые не дают количественной оценки видности интерференционной картины. Такая количественная оценка тесно связана с важным свойством волны, излучаемой протяженным источником света, с *пространственной когерентностью*.

Если в опыте Юнга используется точечный источник света S , находящийся на одинаковом расстоянии от отверстий S_1 и S_2 , то волны от этого источника создают на отверстиях одинаковые колебания: амплитуда и фаза колебаний $V_1(t)$ в точке S_1 в любой момент времени такие же, как и в колебании $V_2(t)$ (в точке S_2):

$$V_1(t) = V_2(t) = V(t).$$

Отметим, что при этом

$$\overline{V_1(t)V_2^*(t)} = \overline{V(t)V^*(t)} = I.$$

Таким образом, колебания $V_1(t)$ и $V_2(t)$ полностью когерентны.

В случае протяженного источника колебания в точках S_1 и S_2 создаются всеми точками источника, причем лишь центральная точка источника создает на отверстиях одинаковые по амплитуде и фазе колебания. Любая другая точка посылает волну, создающую на отверстиях разные по фазе колебания (просто из-за того, что она находится на разных расстояниях от S_1 и S_2). Ясно поэтому, что и суммарное колебание $V_1(t)$ в точке S_1 отличается по фазе от колебания $V_2(t)$ в точке S_2 . Это приводит, как мы увидим, к уменьшению величины

$$\Gamma_{12}(0) = \overline{V_1(t)V_2^*(t)}.$$

Как мы вскоре убедимся, $\Gamma_{12}(0)$ является функцией расстояния между точками S_1 и S_2 . Она называется *функцией пространственной когерентности*.

Пусть $V_1(t)$ и $V_2(t)$ — колебания, созданные протяженным квазимонохроматическим источником S в точках S_1 и S_2 плоскости Π_1 (рис. 7.20). Чтобы экспериментально определить, когерентны ли колебания $V_1(t)$ и $V_2(t)$ установим в плоскости Π_1 непрозрачный экран, проделав два малых отверстия в точках S_1 и S_2 .

Волны света, проходя через отверстия, идут (из-за дифракции) в разных направлениях и, перекрываясь, создают (или не создают) интерференционную картину в плоскости наблюдения Π_2 . Эта картина, как мы уви-

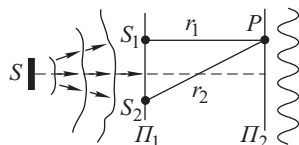


Рис. 7.20

дим, даст нам информацию о функции когерентности $\overline{V_1(t)V_2^*(t)}$.

Колебание в точке наблюдения P есть результат наложения волн, пришедших в эту точку от S_1 и S_2 .

Колебания, созданные волной, пришедшей от отверстия S_1 , запишем в виде $\alpha_1 V_1(t - \tau_1)$, где константа α_1 учитывает ослабление волны из-за расходимости, а $\tau_1 = r_1/c$ есть время распространения волны от S_1 до P , т. е. колебание в точке P в момент t такое же, как на отверстии S_1 в момент $t - \tau_1$, лишь уменьшенной амплитуды.

Аналогично волна от отверстия S_2 создает колебание $\alpha_2 V_2(t - \tau_2)$, где $\tau_2 = r_2/c$. Результирующее колебание, согласно принципу суперпозиции, есть

$$V(t) = \alpha_1 V_1(t - \tau_1) + \alpha_2 V_2(t - \tau_2).$$

Как и ранее, находим

$$I = |V(t)|^2 = I_1 + I_2 + 2 \operatorname{Re} \overline{V_1(t)V_2^*(t + \tau)}.$$

Будем считать, что для всей интересующей нас области наблюдения запаздывание одной волны относительно другой τ ма-

ло по сравнению со временем когерентности τ_0 (но конечно, может быть много больше периода светового колебания $2\pi/\omega_0$), $\tau \ll \tau_0$.

Это означает, что за время $\tau = \Delta/c$ ($\Delta = r_1 - r_2$) амплитуда и фаза колебания $V_2(t)$ не успевает заметно измениться, т. е. колебание в момент времени $t + \tau$ состоит из той же совокупности цугов, что и в момент времени t .

Тогда

$$f_2(t + \tau) \approx f_2(t) \quad \text{и} \quad V_2(t + \tau) = f_2(t + \tau)e^{i\omega(t+\tau)} \approx f_2(t)e^{i\omega(t+\tau)}.$$

Вспомнив определение функции пространственной когерентности, получаем

$$I = I_1 + I_2 + 2 \operatorname{Re} (\Gamma_{12}(0)e^{i\omega_0\tau}). \quad (7.46)$$

Из (7.46) следует, что интенсивность света во всех максимумах интерференционной картины одна и та же (независимо от порядка интерференции), так как величина $\Gamma_{12}(0)$ не зависит от положения точки наблюдения (не зависит от τ при условии что $\tau \ll \tau_0$), а определяется лишь апертурой интерференции (в опыте Юнга — угловым расстоянием между отверстиями S_1 и S_2) или, при заданном расстоянии, размерами источника. Таким образом, видность интерференционной картины одинакова во всех точках.

Вводя, как и ранее, нормированную функцию

$$\gamma_{12}(0) = \frac{\Gamma_{12}(0)}{\sqrt{I_1 I_2}},$$

называемую *степенью пространственной когерентности*, получим из (7.46)

$$I(\tau) = I_1 + I_2 + 2\sqrt{I_1 I_2} \operatorname{Re} (\gamma_{12}(0)e^{i\omega_0\tau}). \quad (7.47)$$

Или, при равных интенсивностях $I_1 = I_2 = I_0$

$$I(\tau) = 2I_0 [1 + \operatorname{Re} (\gamma_{12}(0)e^{i\omega_0\tau})]. \quad (7.48)$$

Полученные выражения (7.47) и (7.48), определяющие интерференционную картину при использовании протяженного источника, справедливы для любой двухлучевой интерференционной схемы. Формулу (7.48) можно записать в виде

$$I(\tau) = 2I_0 [1 + |\gamma_{12}(0)| \cos(\omega_0\tau + \theta)], \quad (7.49)$$

где θ — аргумент комплексной функции γ_{12} .

Видность V выражается через $\gamma_{12}(0)$ равенством

$$V = |\gamma_{12}(0)|.$$

Пока мы имеем лишь качественную оценку (7.45) влияния размеров источника на видность интерференционной картины. Теперь необходимо найти количественную связь функции $\gamma_{12}(0)$ с размером источника.

Для этого рассмотрим интерференционную картину, создаваемую протяженным источником, как сумму интерференционных картин от отдельных точек источника. В качестве примера вновь обратимся к опыту Юнга (рис. 7.21).

Протяженный источник — светящаяся полоска шириной b — и плоскость наблюдения Π находятся на расстояниях z_0 и z от экрана с отверстиями S_1 и S_2 , расстояние между отверстиями d .

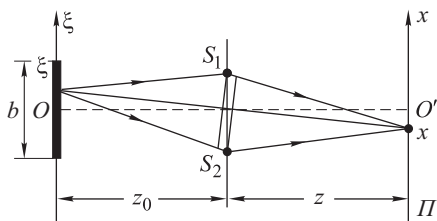


Рис. 7.21

Еще раз подчеркнем, что степень монохроматичности источника достаточно велика (т. е. для всей интересующей нас области наблюдения в плоскости Π разность хода волн много меньше длины цуга $\Delta \ll c\tau_0$). Поэтому можно при расчете картины интерференции, созданной какой-либо точкой источника, пользоваться приближением строго монохроматических волн (формула (7.8)) заданной частоты.

Картина интерференции, созданная в плоскости наблюдения центральной точкой источника « O » (точнее малым элементом источника) имеет вид

$$dI_0 = 2\mathcal{I}_0 d\xi \left[1 + \cos \left(\frac{2\pi}{l} x \right) \right].$$

Здесь $\mathcal{I}_0 d\xi$ — интенсивность волны, прошедшей через одно из отверстий (и созданной в плоскости наблюдения малым элементом источника $d\xi$), $l = (\lambda/d)z$ — ширина интерференционной полосы. Ясно, что нулевая полоса (соответствующая разности хода $\Delta = 0$) находится на оси симметрии, в точке O' .

Из рис. 7.21 также ясно, что элемент $d\xi$ источника, имеющий координату ξ , создает картину интерференции, смещенную на расстояние $x = \xi z/z_0$ (в точку x смещается нулевая полоса, соответствующая разности хода $\Delta = 0$). Мы полагаем, что расстояния z_0 и z велики по сравнению с размером источника b и расстоянием между отверстиями d , т. е.

$$dI = 2\mathcal{I}_0 d\xi \left\{ 1 + \cos \left[\frac{2\pi}{l} \left(x - \frac{z}{z_0} \xi \right) \right] \right\}. \quad (7.50)$$

Результирующую картину интерференции найдем суммированием картин (7.50) по всем точкам источника. Получаем

$$I(x) = 2\mathcal{I}_0 \int_{-b/2}^{b/2} \left\{ 1 + \cos \left[\frac{2\pi}{l} \left(x - \frac{z}{z_0} \xi \right) \right] \right\} d\xi. \quad (7.51)$$

Производя интегрирование (7.51), найдем

$$I(x) = 2I_0 \left[1 + \frac{\sin\left(\pi \frac{d}{\lambda z_0/b}\right)}{\pi \frac{d}{\lambda z_0/b}} \cos\left(\frac{2\pi}{l} x\right) \right]. \quad (7.52)$$

Сравнивая (7.52) и (7.49), находим степень пространственной когерентности, которая определяет видность интерференционной картины:

$$V = |\gamma_{12}| = \left| \frac{\sin\left(\pi \frac{d}{\lambda z_0/b}\right)}{\pi \frac{d}{\lambda z_0/b}} \right|. \quad (7.53)$$

Функция $|\gamma_{12}|$ изображена на рис. 7.22 (аргументом является расстояние d между точками S_1 и S_2). Как следует из (7.53), степень когерентности колебаний в двух точках, разнесенных на расстояние d уменьшается от 1 при $d = 0$ до нуля при $d = \rho_0 = 2\pi z/(kb) = \lambda z/b$.

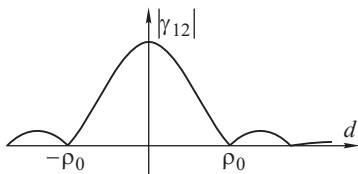


Рис. 7.22

Таким образом, если расстояние между точками S_1 и S_2 меньше $\lambda z/b$, то колебания в этих точках частично когерентны: при этом $|\gamma_{12}| \neq 0$. При $d \gtrsim \lambda z/b$ $|\gamma_{12}| \approx 0$, т. е. колебания в точках S_1 и S_2 некогерентны.

Величину $\rho_0 = \lambda z/b$ называют *радиусом пространственной когерентности*, а площадку с линейным размером $\lambda z/b$ называют *площадкой когерентности*. Вводя угловой размер источника $\psi = b/z_0$, можно записать

$$\rho_0 = \frac{\lambda}{\psi}. \quad (7.54)$$

Итак, колебания в любых двух точках, лежащих внутри площадки размером ρ_0 частично когерентны, а в точках, расстояние между которыми больше ρ_0 — некогерентны.

Возвращаясь к опыту Юнга, можно иначе, чем это сделано выше, сформулировать условия, при которых можно наблюдать интерференционную картину. Если расстояние между отверстиями S_1 и S_2 меньше радиуса когерентности, то интерференция наблюдается. Аналогичное условие можно сформулировать для любой интерференционной схемы: апертура интерференции должна быть меньше углового размера площадки когерентности.

Если зафиксировать расстояние между точками S_1 и S_2 и приближать плоскость xy , в которой они находятся, к источнику, то, как следует из (7.53), степень когерентности колебаний

в этих точках уменьшается. При стремлении z_0 к нулю (плоскость xu придвинута вплотную к источнику) колебания в любых двух точках, как угодно близко расположенных, некогерентны (ведь любые две точки источника излучают некогерентно). Наоборот, с удалением плоскости xu колебания в точках S_1 и S_2 становятся все более когерентными и при достаточном удалении когерентность колебаний становится полной. Размер площадок когерентности по мере удаления от источника постепенно (пропорционально z) увеличивается. Мы видим, таким образом, что первоначально некогерентное поле излучения в процессе распространения приобретает частичную когерентность. В этом и состоит смысл соотношения (7.53).

7.8. Интерферометры и интерферометрия

Явление интерференции находит широкое применение при решении самых разнообразных физических и технических проблем. Используемые для этого оптические устройства называются *интерферометрами* (например, интерферометры Майкельсона на рис. 7.13), а область оптики, исследующую возможности использования интерферометров называют *интерферометрией*.

Мы убедились ранее, что интерференционная картина позволяет получить информацию о временной степени когерентности $\gamma(\tau)$, а следовательно и о спектральном составе излучения $\mathcal{I}(\omega)$. Это, в свою очередь, дает информацию о физических процессах в источнике (энергетических уровнях атомов, процессах их взаимодействия, температуре и т. д.).

С другой стороны, измерения видности интерференционной картины, созданной каким-либо протяженным источником, дают возможность определить его угловой размер. Интерференционный метод определения угловых размеров звездных объектов (либо измерения углового расстояния между компонентами двойных звезд) был предложен и реализован американским физиком А. Майкельсоном (1852–1931).

Видность интерференционной картины в фокальной плоскости объектива «О» в схеме звездного интерферометра Майкельсона (рис. 7.13 е) зависит от степени когерентности колебаний на зеркалах M_1 и M_2 (эти колебания созданы излучением какой-либо удаленной звезды — протяженного источника света).

Расстояние D между зеркалами M_1 и M_2 называется *базой интерферометра* и может изменяться. Пока это расстояние мало (меньше радиуса когерентности) степень когерентности колебаний в точках M_1 и M_2 достаточно велика и интерференционная картина в фокальной плоскости Э объектива достаточно четкая. По мере увеличения D видность картины постепенно

уменьшается и при $D = \rho_0$ (ρ_0 — радиус когерентности, определяемый формулой (7.54)) интерференционная картина полностью исчезает. Определяемый угловой размер звезды равен

$$\psi = \frac{\lambda}{\rho_0} = \frac{\lambda}{D}. \quad (7.55)$$

Таким образом Майкельсону удалось определить угловой размер красного гиганта в созвездии Ориона — звезды Бетельгейзе. Измеренный экспериментально радиус когерентности оказался равным 6 м, а значение ψ , рассчитанное по формуле (7.55), равным $0,0047''$. Для сравнения, угловой размер Солнца $\psi \approx 30' \approx 10^{-2}$ рад, а соответствующий радиус когерентности $\rho_0 \approx 0,06$ мм.

Вообще подавляющее большинство звезд имеет диаметр d , мало отличающийся от диаметра Солнца, поэтому их угловой размер $\psi = d/r$ (в силу огромности расстояния r) чрезвычайно мал. Поэтому для измерений угловых размеров большинства звезд требуются интерферометры с базой D в несколько десятков метров. Создание оптических устройств с такой базой чрезвычайно сложная техническая проблема: ведь нужно обеспечить высокую стабильность расположения зеркал M_1 и M_2 в процессе измерения («дрожания» одного зеркала относительно другого порядка длины световой волны ($\lambda \approx 5 \cdot 10^{-7}$ м) приводят к полному исчезновению картины интерференции). Серьезной помехой является также атмосферная турбулентность.

Проблему позволяют решить *радиоинтерферометры*. Роль зеркал M_1 и M_2 играют в данном случае радиотелескопы (две радиоантенны), принимающие сигнал с измеряемого источника в диапазоне радиоволн ($\lambda \approx 10$ см). Эти две антенны могут располагаться даже на разных континентах, так что база интерферометра D может быть тысячи и десятки тысяч километров. Сигналы с двух радиотелескопов подают на общий детектор, измеряющий интенсивность суммарного сигнала. При использовании радиоинтерферометров переход от коротких (оптических) длин волн к длинным (радиоволнам) компенсируется увеличением базы интерферометра, поэтому угловой размер измеряемого объекта $\psi = \lambda/D$ может быть чрезвычайно малым (примерно в 100 раз меньше, чем при использовании «оптического» интерферометра Майкельсона).

Широкое применение находят интерферометры для высокоточных измерений расстояний, углов, показателей преломления сред или их изменений и т. д. На рис. 7.23 показана схема интерферометра Рэлея, предназначенного для измерения показателей преломления жид-

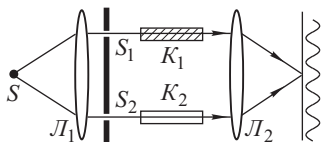


Рис. 7.23

кий и т. д. На рис. 7.23 показана схема интерферометра Рэлея, предназначенного для измерения показателей преломления жид-

костей или газов. Пучки света, прошедшие через щели S_1 и S_2 и кюветы K_1 и K_2 , заполненные исследуемыми жидкостями или газами, собираются объективом L_2 и образуют в его фокальной плоскости интерференционную картину. От разности хода волн, прошедших через кюветы $\Delta = l(n_2 - n_1)$ зависит смещение интерференционных полос (l — длина кювет, n_1 и n_2 — показатели преломления в кюветах K_2 и K_1). Измеряя смещение полос, можно определить разность показателей преломления Δn .

Широко распространенный в настоящее время метод изучения спектрального состава излучения основан на использовании интерферометра Майкельсона (см. рис. 7.13 д), в котором одно из зеркал перемещается в процессе измерения.

При использовании монохроматического источника интенсивность света, регистрируемого фотоприемником Φ (расположенном в фокальной плоскости объектива L_2), равна

$$I = 2I_0 \left[1 + \cos \left(\frac{\omega}{c} \Delta \right) \right]$$

(см. формулу (7.8)). Пусть одно из зеркал перемещается со скоростью v , разность хода Δ интерферирующих волн изменяется: $\Delta = 2vt$. Тогда

$$I(t) = 2I_0 \left[1 + \cos \left(2 \frac{\omega}{c} vt \right) \right].$$

Частота изменения тока фотоприемника $I(t)$ равна $\Omega = 2(\omega/c)v$, т. е. оптической частоте излучения ω соответствует измеряемая частота Ω фототока.

Если источник содержит несколько монохроматических компонент различных частот ω_n , то ток фотоприемника также будет содержать несколько соответствующих компонент Ω_n .

Таким образом, с помощью подвижного зеркала в интерферометре Майкельсона спектр, расположенный в области оптических частот, содержащихся в источнике света, трансформируется в фотоприемнике в низкочастотный спектр (в область радиочастот).

Например, при $\nu = 4 \cdot 10^{15}$ Гц ($\omega = 2\pi\nu$), что соответствует зеленой линии оптического спектра и скорости перемещения зеркала $v = 1$ см/с, частота фототока $f \approx 2,4 \cdot 10^5$ Гц. Изучая спектральный состав фототока, можно судить о спектре излучения источника.

Описанный метод исследования оптических спектров называется *фурье-спектрометрией*, поскольку на последней стадии измерений определяется спектр (преобразование Фурье) фототока.

Вопросы и задачи

1. Проанализируйте работу интерференционных схем, показанных на рис. 7.13. Какие параметры схем определяют ширину интерференционной полосы и апертуру интерференции?

Чем определяется минимальный порядок интерференции в схеме Ллойда и в интерферометре Майкельсона?

2. В какой из схем на рис. 7.13 возникают наиболее сильные ограничения на степень монохроматичности и на размеры источника.

3. Радиоизлучение космического источника длины волны λ , имеющего угловой размер ψ , принимается антенной, расположенной на отвесном берегу на высоте h над уровнем моря. Как будет меняться интенсивность принимаемого сигнала в зависимости от угла возвышения источника над горизонтом α . При каких значениях ψ интенсивность не зависит от λ ?

Поверхность воды можно рассматривать как плоское зеркало, считать углы α и ψ малыми.

4. Три плоские монохроматические волны с амплитудами 1, a и a ($a \ll 1$) падают на плоскость $z = 0$ под углами 0 , α и $-\alpha$ так, что колебания в точке $x = 0$ оказываются синфазными. При смещении плоскости наблюдения в область $z > 0$ происходят периодические изменения контраста интерференционной картины. Объясните явление. Найдите положение плоскости наблюдения, в которых контраст картины максимален и минимален. Чему он равен?

5. Излучающая система состоит из ряда равноотстоящих параллельных вибраторов с линейно меняющейся вдоль ряда разностью фаз излучения. Как должен меняться со временем сдвиг фаз между двумя соседними вибраторами, чтобы главный лепесток диаграммы направленности всей системы (т. е. главный дифракционный максимум) совершал круговой обзор местности с постоянной угловой скоростью Ω (при отсутствии вращения самой решетки вибраторов).

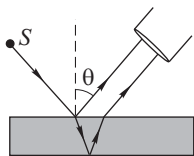


Рис. 7.24

6. С помощью зрительной трубы, установленной на бесконечность, наблюдают полосы равного наклона в тонкой плоскопараллельной пластинке толщиной $h = 0,2$ мм с показателем преломления $n = 1,41$. При этом угол наблюдения θ может изменяться от 0 до 90° (рис. 7.24). Найти максимальный и минимальный порядок интерференционных полос. Оценить допустимую степень монохроматичности $\Delta\lambda$, при которой будут четко наблюдаться все интерференционные полосы. Каков допустимый размер источника в этом эксперименте для $\lambda = 500$ нм?

7. Каково минимальное угловое разрешение радиointерферометра, установленного на Земле, при работе на длине волны $\lambda = 10$ м?

Глава 8

ДИФРАКЦИЯ

Дифракцией называют совокупность волновых явлений, в которых проявляются отклонения от законов геометрической оптики. Пусть, например, непрозрачный экран с круглым отверстием освещается параллельным пучком света. В плоскости наблюдения, расположенной недалеко от экрана, можно наблюдать картину, показанную на рис. 8.1 а: равномерно освещенное круглое пятно с резкой границей света и тени, в точности повторяющее размеры и форму отверстия, т. е. изображение получается геометрическим проецированием отверстия на плоскость наблюдения. Однако стоит отнести плоскость наблюдения подальше (рис. 8.1 б), и наблюдаемая картина существенно изменяется: граница света и тени становится размытой, а внутри пятна возникают чередующиеся светлые и темные кольца. Наконец, если плоскость наблюдения отнести еще дальше и (или) уменьшить размеры отверстия, через которое прошел свет, то размер освещенного пятна в плоскости наблюдения становится существенно больше размеров отверстия — мы видим явное нарушение геометрического закона прямолинейного распространения света.

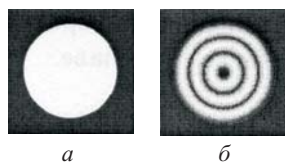


Рис. 8.1

Итак, нам нужен рецепт, который позволил бы описывать законы распространения волн, отличные от законов геометрической оптики.

В этой главе мы будем рассматривать строго монохроматические волны с заданной частотой ω . Как и при изучении интерференционных явлений, будем описывать волновой процесс скалярной функцией

$$E(\mathbf{r}, t) = a(\mathbf{r}) \cos(\omega t - \varphi(\mathbf{r})), \quad (8.1)$$

в которой $a(\mathbf{r})$ и $\varphi(\mathbf{r})$ — функции координат, определяющие пространственную структуру волны. Напомним, что монохроматической волне можно поставить в соответствие комплексную амплитуду $f(\mathbf{r}) = a(\mathbf{r})e^{i\varphi(\mathbf{r})}$, которая (при заданной частоте) содержит всю необходимую информацию о волне. Эту функцию координат мы будем называть далее *волновым полем* (или просто полем).

Постановка простейшей дифракционной задачи такова. Пусть непрозрачный экран (маска) с отверстием, расположенный в плоскости $z = 0$, освещается волной, созданной источниками, расположенными слева от экрана, в области $z < 0$. Требуется определить волновое поле справа от экрана, в области $z > 0$ (рис. 8.2), в частности, в плоскости наблюдения Π , расположенной на расстоянии z от экрана.

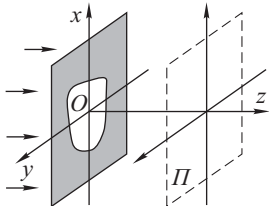


Рис. 8.2

Для решения этой задачи необходимо найти ответ на два основных вопроса. Первый вопрос — определение граничных условий — ставится следующим образом: известно поле $f_s(x, y)$, созданное в плоскости $z = 0$ (в отсутствие маски) какими-либо «сторонними» источниками, расположенными в области $z < 0$. Это может быть поле плоской волны, нормально либо косо падающей на плоскость $z = 0$, или поле точечного источника — сферическая волна. Совершенно не важно, с помощью каких «сторонних» источников и каким образом это поле создано — мы его считаем заданным. Затем в плоскости $z = 0$ помещается непрозрачный экран с отверстием. Требуется ответить на вопрос, какое поле $f_0(x, y)$ возникает в плоскости, непосредственно примыкающей к экрану с отверстием справа от него (эту плоскость далее будем называть плоскостью $z = 0_+$).

Это и есть то *граничное условие*, которое необходимо знать для решения дифракционной задачи.

Второй вопрос: каким образом, зная граничное поле $f_0(x, y)$, определить искомое поле $f(x, y, z)$ в области $z > 0$ (т. е. в пространстве справа от экрана)? В частности, каково поле в фиксированной плоскости наблюдения $z = \text{const} > 0$?

Основными параметрами, существенно определяющими характер дифракционных явлений, являются: длина волны λ , размер отверстия b , расстояние до плоскости (или до точки) наблюдения z . Как показывает дальнейший анализ (см. § 8.3–8.7), тот или иной характер дифракционных явлений существенно зависит от значения *волнового параметра*

$$p = \frac{\sqrt{\lambda z}}{b}.$$

Область значений волнового параметра $p \ll 1$ называется областью *геометрической оптики*. Область значений $p \gg 1$ — областью *дифракции Фраунгофера* (или дальней волновой зоной). Промежуточная область $p \approx 1$ — областью *дифракции Френеля* (или ближней волновой зоной).

8.1. Граничные условия Кирхгофа

Определение граничных условий представляет собой исключительно трудную проблему. Действительно, пусть речь идет об электромагнитной волне, падающей на проводящий экран с отверстием. Переменное электрическое поле волны порождает на проводящем экране переменные токи, которые в свою очередь индуцируют «вторичную» электромагнитную волну. Эта волна зависит как от физических свойств экрана, так и от геометрии отверстия. Суммарное волновое поле в плоскости $z = 0_+$ представляет собой суперпозицию первичной волны (поля сторонних источников) и вторичной волны, индуцированной токами проводимости, причем электрическое и магнитное поля суммарной волны должны удовлетворять известным в электродинамике граничным условиям. Имеется лишь небольшое число задач, для которых найдено точное решение (например, задача дифракции на идеально проводящем шаре).

Мы, следуя немецкому физику Г. Кирхгофу (1824–1887), определим *приближенные граничные условия* следующим образом: в той части плоскости $z = 0_+$, которая затенена непрозрачным экраном, граничное поле полагается равным нулю. На открытой части волнового фронта (т. е. в области отверстия) граничное поле полагается равным полю сторонних источников. Другими словами, в этой области установка экрана с отверстием совершенно не влияет на падающую волну. Итак, граничное поле определяется следующим равенством:

$$f_0(x, y) = \begin{cases} f_s(x, y) & \text{в точках } (x, y), \text{ принадлежащих} \\ & \text{отверстию,} \\ 0 & \text{в точках, затененных маской.} \end{cases} \quad (8.2)$$

Еще раз подчеркнем, что граничные условия (8.2) являются приближенными: наличие экрана в действительности приводит к тому, что поле на отверстии отлично от падающего поля $f_s(x, y)$, и особенно сильно оно искажено вблизи краев отверстия; да и в затененной области поле отлично от нуля вблизи краев.

Однако, как показывает сравнение теоретических выводов, основанных на приближенных граничных условиях Кирхгофа, с экспериментом, согласие оказывается удовлетворительным при условии, что, во-первых, линейные размеры отверстия b велики по сравнению с длиной волны λ , во-вторых, плоскость наблюдения находится на расстоянии, много большем длины волны:

$$b > \lambda, \quad z \gg \lambda,$$

причем первое неравенство не обязательно должно быть сильным.

Равенству (8.2) можно придать другую форму, если ввести *функцию пропускания* непрозрачного экрана с отверстием $t(x, y)$ согласно формуле

$$t(x, y) = \begin{cases} 1 & \text{в точках } (x, y), \text{ принадлежащих отверстию } \sigma, \\ 0 & \text{в точках } (x, y), \text{ лежащих вне отверстия } \sigma. \end{cases}$$

Тогда можно записать

$$f_0(x, y) = f_s(x, y)t(x, y). \quad (8.3)$$

Равенство (8.3) можно обобщить на случай произвольного *тонкого экрана*. Представим себе достаточно тонкую пластинку-транспарант, пропускаемость которой меняется от точки к точке по закону $a(x, y)$, а показатель преломления (или толщина) по закону $n(x, y)$ (или $\Delta(x, y)$).

Пусть такая пластинка освещается плоской нормально падающей волной. Что происходит с волной на выходе из пластинки в плоскости $z = 0_+$? Переменная поглощательная способность транспаранта приводит к тому, что амплитуда колебаний на выходе будет не константой, а некоторой функцией координат $a(x, y)$. Переменный показатель преломления n или толщина Δ приводят к различному в разных точках набегу фазы $\varphi(x, y) = kn\Delta$, так что фронт прошедшей волны будет не плоским: распределение фаз колебаний на выходе будет описываться функцией $\varphi(x, y)$.

Если ввести комплексную пропускаемость транспаранта согласно формуле

$$t(x, y) = a(x, y)e^{i\varphi(x, y)},$$

то комплексная амплитуда волны на выходе из пластинки может быть найдена по той же формуле (8.3), которая и определяет граничное поле $f_0(x, y)$ для произвольного *тонкого* экрана.

Принципиально важным оказывается то обстоятельство, что влияние маски затрагивает лишь пространственные характеристики волнового поля, не меняя его частоту (а в случае немонохроматических волн — частотный спектр). Это — общее свойство линейных и стационарных оптических систем.

Приведем примеры амплитудных и фазовых транспарантов, часто применяемых в оптике.

1. Дифракционная решетка имеет функцию пропускания

$$t(x) = \sum_{-N}^N p_b(x - nd), \quad (8.4)$$

изображенную на рис. 8.3 (d — период решетки, $2b$ — ширина щели, $2N + 1$ — число штрихов).

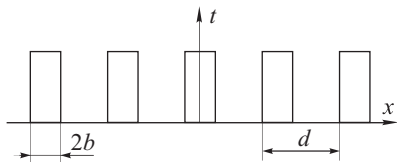


Рис. 8.3

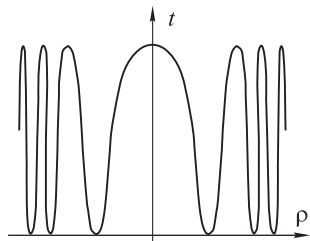


Рис. 8.4

2. Амплитудная синусоидальная решетка с периодом $d = 2\pi/\Omega$ и глубиной модуляции t

$$t(x) = 1 + t \cos \Omega x. \quad (8.5)$$

3. Фазовая синусоидальная решетка —

$$t(x) = e^{im \cos \Omega x}. \quad (8.6)$$

(Это пример чисто фазовой структуры: амплитуда волны прошедшей через решетку не меняется, а фаза изменяется по синусоидальному закону.)

4. Зонная решетка Габора (рис. 8.4) —

$$t(\rho) = 1 + \cos k \sqrt{z_0^2 + \rho^2}, \quad (8.7)$$

где $\rho^2 = x^2 + y^2$, $z_0 = \text{const}$.

5. Функция пропускания линзы. Основным элементом любой оптической системы (фотоаппарат, микроскоп, телескоп) является линза. Рассмотрим тонкую прозрачную пластинку, одна из поверхностей которой плоская, а другая — сферическая радиуса R (рис. 8.5). Считая линзу тонкой, мы полагаем, что распространение волн через линзу происходит по законам геометрической оптики, т. е. чтобы связать световое поле в точке x на входе в линзу (в плоскости $z = 0_-$) с полем в той же точке x на выходе (в плоскости $z = 0_+$, примыкающей к линзе справа), достаточно найти набег фазы вдоль луча S , проходящего через линзу на расстоянии x от оптической оси.

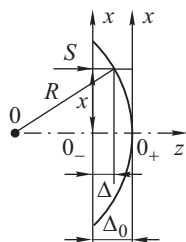


Рис. 8.5

Луч, проходящий вдоль оптической оси ($x = 0$) приобретает набег фазы, равный $\varphi_0 = kn\Delta_0$ (n — показатель преломления, Δ_0 — толщина линзы на оптической оси — расстояние между плоскостями, примыкающими к линзе слева и справа). Для луча S набег фазы равен

$\varphi = kn\Delta + k(\Delta_0 - \Delta)$ — мы учитываем при этом, что часть оптического пути луч проходит внутри линзы, т. е. в среде с показателем преломления n (Δ — толщина линзы на расстоянии x от оптической оси), а другую часть пути, равную $(\Delta_0 - \Delta)$, — вне линзы, т. е. в среде с показателем преломления $n = 1$. Относительный набег фазы, который только и представляет интерес, есть

$$\Delta\varphi = \varphi - \varphi_0 = k(n - 1)(\Delta - \Delta_0),$$

где $\Delta_0 - \Delta = R - \sqrt{R^2 - x^2} \approx x^2/2R$.

Это выражение получено с помощью разложения радикала в ряд по степеням $(x/R)^2$ и означает фактически, что сферическая поверхность линзы заменяется параболической.

Таким образом,

$$\Delta\varphi = -\frac{k(n-1)}{2R}x^2 = -\frac{k}{2f}x^2, \quad (8.8)$$

где через f обозначена величина, определяемая равенством

$$\frac{1}{f} = \frac{n-1}{R}.$$

Она называется *фокусным расстоянием* линзы. В общем случае, если радиусы кривизны поверхностей R_1 и R_2 , то вместо последнего равенства легко получить

$$\frac{1}{f} = (n-1) \left(\frac{1}{R_1} + \frac{1}{R_2} \right). \quad (8.9)$$

Эта формула хорошо известна в геометрической оптике. Будем называть линзу с описанными выше характеристиками *идеальной*. Проходя через идеальную линзу, волна приобретает различный в разных точках набег фазы, при этом амплитуда волны не меняется (отражением света от поверхности линзы, а также поглощением, мы пренебрегаем).

Таким образом, функция пропускания идеальной линзы определяется равенством

$$t(x) = \exp\left(-\frac{ik}{2f}x^2\right). \quad (8.10)$$

Разумеется, если мы говорим в действительности о сферических поверхностях (а не цилиндрических), то вместо (8.8) в двумерном случае следует написать

$$\Delta\varphi = -\frac{k}{2f}(x^2 + y^2), \quad (8.11)$$

а вместо (8.10)

$$t(x, y) = \exp\left[-\frac{ik}{2f}(x^2 + y^2)\right]. \quad (8.12)$$

Функция пропускания связывает между собой комплексную амплитуду волны $f_-(x, y)$ на входе в линзу с комплексной амплитудой волны $f_+(x, y)$ на выходе:

$$f_+(x, y) = f_-(x, y) t(x, y). \quad (8.13)$$

Любая реальная линза имеет конечные размеры. Можно учесть этот факт, накрыв бесконечную линзу с функцией пропускания (8.10) непрозрачным экраном с отверстием (диафрагмой) диаметра $2a$, центр которого совпадает с центром линзы (рис. 8.6). Очевидно функция пропускания задиафрагмированной линзы имеет вид

$$t(x, y) = p_a(x, y) \exp \left[-\frac{ik}{2f} (x^2 + y^2) \right], \quad (8.14)$$

где $p_a(x, y)$ — единично-нулевая функция,

$$p_a(x, y) = \begin{cases} 1 & \text{при } x^2 + y^2 \leq a^2, \\ 0 & \text{при } x^2 + y^2 > a^2. \end{cases}$$

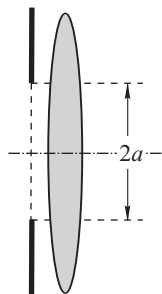


Рис. 8.6

8.2. Спектральный метод решения задачи дифракции (метод Рэлея)

Перейдем к решению второй проблемы: как, зная граничное поле (8.3), найти волновое поле в области $z > 0$. Математически проблема сводится к решению волнового уравнения, удовлетворяющего заданному граничному условию.

Рассмотрим вначале спектральный подход к решению проблемы, основанный на представлении дифрагированной волны в виде *суперпозиции плоских волн разных направлений* (метод Рэлея). Напомним, что комплексная амплитуда плоской волны, волновой вектор \mathbf{k} которой составляет угол α с осью z , имеет вид (4.27)

$$f(x, z) = a e^{i(kx \sin \alpha + kz \cos \alpha + \varphi)},$$

где a — амплитуда волны, $k \sin \alpha = k_x$ и $k \cos \alpha = k_z$ — проекции вектора \mathbf{k} на оси x и z соответственно (рис. 8.7). Мы рассматриваем для простоты случай $k_y = 0$, φ — начальная фаза, т. е. фаза колебания в точке $x = 0$, $z = 0$. Обратим внимание на чрезвычайно простое правило, согласно которому можно найти волновое поле плоской волны в любой плоскости $z > 0$. Действительно, в плоскости $z = 0$ имеем

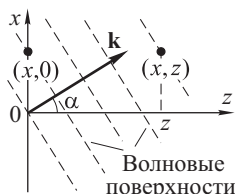


Рис. 8.7

$$f(x, 0) = a e^{i\varphi} e^{ikx \sin \alpha},$$

поэтому

$$f(x, z) = f(x, 0)e^{ikz \cos \alpha},$$

т. е. комплексные амплитуды $f(x, 0)$ и $f(x, z)$ отличаются множителем $e^{ikz \cos \alpha}$, определяющим набег фазы плоской волны при распространении между двумя плоскостями, разделенными промежутком z (набег фазы, равный $k_z z = kz \cos \alpha$).

Запишем, наконец, поле плоской волны в плоскости $z = 0$ в виде

$$f(x, 0) = ce^{iu x}, \quad (8.15)$$

где множитель $c = ae^{i\varphi}$ определяет как амплитуду a , так и начальную фазу φ волны, а через u обозначена x -компонента вектора \mathbf{k} : $k \sin \alpha = k_x = u$ (при этом z -компонента вектора \mathbf{k} есть $k \cos \alpha = \sqrt{k^2 - u^2}$).

Обратим внимание на аналогию формулы (8.15) с выражением $f(t) = ce^{i\omega t}$, которое, как мы знаем, есть не что иное, как комплексная форма записи гармонического колебания частоты ω , причем комплексный множитель $c = ae^{i\varphi}$ определяет амплитуду колебания и его начальную фазу. На основании этой аналогии величина $u = k \sin \alpha$ может быть названа *пространственной частотой*.

Можно сказать, что волны разных направлений α — это волны разных пространственных частот.

Итак, представим граничное поле $f_0(x)$ в виде

$$f_0(x) = \sum c_n e^{iu_n x}. \quad (8.16)$$

Математически выражение (8.16) полностью аналогично формуле (1.37) — представлению колебательного процесса в виде суперпозиции гармонических колебаний разных частот ω_n ; гармоническому колебанию частоты ω_n в разложении (1.37) соответствует плоская волна определенного направления α_n ($\sin \alpha_n = u_n/k$) в разложении (8.16).

Подчеркнем, что модуль $|\mathbf{k}| = \omega/c$ всех гармоник в пространственном фурье-разложении (8.16) одинаков.

Все, что было сказано ранее по поводу возможности представления функции времени в виде суммы гармонических колебаний разных частот, можно повторить по отношению к (8.16): представлению волнового поля $f_0(x)$ (функции координат) в виде суммы плоских волн разных направлений — ведь математику не очень заботит различие в физическом смысле переменных (времени t в (1.37) или координаты x в (8.16)).

В общем случае, как мы знаем, произвольный колебательный процесс представляется в виде интеграла Фурье — непрерывной

суммы гармонических колебаний различных частот:

$$f(t) = \frac{1}{2\pi} \int C(\omega) e^{i\omega t} d\omega. \quad (8.17)$$

Точно так же произвольное граничное поле $f_0(x)$ представляется в виде интеграла

$$f_0(x) = \frac{1}{2\pi} \int C_0(u) e^{iux} du, \quad (8.18)$$

т. е. в виде непрерывной суммы плоских волн различных пространственных частот.

Ранее мы говорили о замечательном соотношении, связывающем длительность сигнала Δt с шириной его спектра $\Delta\omega$ (соотношении неопределенностей (1.52)): $\Delta t \cdot \Delta\omega \approx 2\pi$. Из математической тождественности выражений (8.17) и (8.18) следует, что аналогичное соотношение неопределенностей должно связывать протяженность Δx граничного поля с шириной Δu его пространственного спектра:

$$\Delta x \cdot \Delta u \approx 2\pi. \quad (8.19)$$

Пространственная протяженность граничного поля, согласно граничным условиям Кирхгофа (8.2), определяется характерным размером препятствия, в нашем примере — размером b отверстия в непрозрачном экране. Поэтому ширина спектра плоских волн (область значений пространственных частот Δu , в которой пространственный спектр $C_0(u)$ заметно отличен от нуля) можно оценить так:

$$\Delta u \approx \frac{2\pi}{b}.$$

Разброс пространственных частот определяет разброс направлений слагаемых плоских волн за препятствием:

$$\Delta u = k\Delta(\sin \alpha),$$

откуда $\Delta(\sin \alpha) \approx \lambda/b$ или, для малых углов,

$$\Delta\alpha \approx \frac{\lambda}{b}.$$

Это и есть *дифракционная расходимость пучка света за отверстием размера b* .

Важно подчеркнуть, что речь идет о линейной задаче: распространение волны от плоскости $z = 0_+$ до плоскости наблюдения $z = \text{const} > 0$ описывается линейным волновым уравнением (4.5) или, поскольку мы пользуемся комплексным представлением, линейным уравнением Гельмгольца (4.26). Основное свойство линейного уравнения: сумма решений является решением, или,

более подробно, если мы имеем два решения $f_1(x, z)$ и $f_2(x, z)$, удовлетворяющих на границе $z = 0_+$ условиям

$$f_1(x, 0) = f_{10}(x) \quad \text{и} \quad f_2(x, 0) = f_{20}(x)$$

($f_{10}(x)$ и $f_{20}(x)$ — заданные функции координаты x), то линейная суперпозиция решений $f_1(x, z)$ и $f_2(x, z)$:

$$f(x, z) = c_1 f_1(x, z) + c_2 f_2(x, z)$$

есть решение уравнения Гельмгольца, удовлетворяющее граничному условию:

$$f_0(x) = c_1 f_{10}(x) + c_2 f_{20}(x).$$

Вспомним теперь, что плоская волна

$$f_n(x, z) = c_n e^{i(u_n x + \sqrt{k^2 - u_n^2} z)}$$

есть решение уравнения Гельмгольца, удовлетворяющее на плоскости $z = 0_+$ граничному условию

$$f_n(x, 0) = c_n e^{i u_n x},$$

поэтому сумма плоских волн

$$f(x, z) = \sum_n c_n e^{i(u_n x + \sqrt{k^2 - u_n^2} z)} \quad (8.20)$$

есть решение, удовлетворяющее граничному условию (8.16).

Это и есть решение дифракционной задачи. Мы нашли волновое поле в любой плоскости наблюдения, отстоящей от препятствия (непрозрачного экрана с отверстием) на расстояние z .

В более общем случае, если граничное поле $f_0(x)$ представляется непрерывной суперпозицией плоских волн (8.18), искомое решение имеет вид

$$f(x, z) = \int C_0(u) e^{i(ux + \sqrt{k^2 - u^2} z)} du, \quad (8.21)$$

где $C_0(u)$ — преобразование Фурье граничного поля $f_0(x)$ (т. е. спектр плоских волн граничного поля).

В выражениях (8.20) или (8.21) существенным является следующее обстоятельство: каждая слагаемая плоская волна при распространении до плоскости наблюдения $z = \text{const} > 0$ приобретает свой фазовый набег $\varphi_n = \sqrt{k^2 - u_n^2} z$, зависящий от ее пространственной частоты u_n . Поэтому фазовые соотношения между слагаемыми плоскими волнами на границе $z = 0_+$ и в плоскости наблюдения, отстоящей на расстоянии z , различны. Изменение фазовых соотношений между слагаемыми плоскими волнами приводит к тому, что изменяется результат интерференции этих плоских волн. Поэтому результирующее поле $f(x, z)$

в плоскости наблюдения $z = \text{const} > 0$ может кардинально отличаться от граничного поля $f_0(x)$ (хотя и то, и другое составлено из суперпозиции тех же бегущих плоских волн).

Перепишав выражение (8.21) в виде

$$f(x, z) = \int C_0(u) e^{i\sqrt{k^2 - u^2} z} e^{iux} du = \int C(u) e^{iux} du,$$

мы видим, что функция

$$C(u) = C_0(u) e^{i\sqrt{k^2 - u^2} z} \quad (8.22)$$

представляет собой не что иное, как преобразование Фурье (пространственный спектр) монохроматического волнового поля $f(x, z)$ в плоскости наблюдения (это поле рассматривается как функция координаты x в плоскости наблюдения *при фиксированном z*).

Равенство (8.22) связывает между собой пространственные спектры $C_0(u)$ и $C(u)$ (преобразования Фурье) световых полей $f_0(x)$ и $f(x)$ в двух плоскостях, разделенных промежутком свободного пространства z . Оно совершенно аналогично спектральному равенству (2.99): $B(\omega) = C(\omega)H(\omega)$, которое связывает между собой спектры сигналов на входе и выходе линейного фильтра с частотной характеристикой $H(\omega)$ (см. гл. 2, $C(\omega)$ и $B(\omega)$ — преобразования Фурье входного $f(t)$ и выходного $S(t)$ сигналов соответственно). Используя эту аналогию, множитель

$$H(\omega) = e^{i\sqrt{k^2 - u^2} z} \quad (8.23)$$

можно назвать *пространственной частотной характеристикой* (или кратко просто частотной характеристикой) *пространственного линейного фильтра* — промежутка z свободного пространства между входной плоскостью $z = 0_+$ и выходной плоскостью $z = \text{const} > 0$. Соответственно световые поля — функции координат $f_0(x)$ и $f(x)$ — входной и выходной сигналы пространственного фильтра, показанного на рис. 8.8 а; на рис. 8.8 б — соответствующая блок-схема.

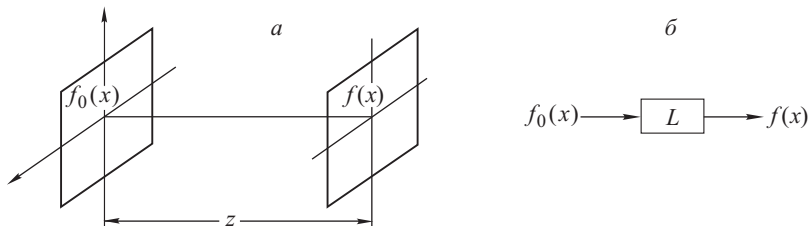


Рис. 8.8

В большинстве оптических дифракционных задач речь идет о препятствиях, размеры которых заметно превышают длину световой волны, поэтому пространственные частоты плоских волн, возникающих при дифракции на препятствиях размера $b \gg \lambda$, согласно (8.19), существенно меньше волнового числа $k = 2\pi/\lambda$:

$$\frac{2\pi}{b} \ll \frac{2\pi}{\lambda}.$$

Используя условие $u \ll k$, получаем, разлагая радикал $\sqrt{k^2 - u^2}$ в ряд по степеням малого параметра $(u/k)^2$ и ограничиваясь двумя членами разложения,

$$\sqrt{k^2 - u^2} \approx kz - \frac{z}{2k} u^2. \quad (8.24)$$

Тогда выражение для пространственной частотной характеристики свободного пространства $H(u)$ принимает вид

$$H(u) \approx e^{ikz} \exp\left(-i \frac{z}{2k} u^2\right),$$

который используют обычно при решении дифракционных задач в оптике. В частности вместо формулы (8.21) получаем

$$f(x, z) \approx e^{ikz} \int C_0(u) \exp\left(-i \frac{z}{2k} u^2 + ux\right) du. \quad (8.25)$$

Напомним еще раз, что функция $f(x, z)$ описывает картину дифракции (комплексную амплитуду волны) в фиксированной плоскости наблюдения, находящейся на расстоянии z за препятствием, а функция $C_0(u)$ — пространственный спектр (преобразование Фурье) светового поля $f_0(x)$ в плоскости, непосредственно примыкающей к препятствию (это поле находится с помощью граничных условий Кирхгофа (8.3)).

Все соотношения легко обобщаются на случай, когда световое поле во «входной плоскости» $z = 0_+$ является функцией двух переменных. Тогда вместо (8.16) имеем

$$f_0(x, y) = \sum c_n e^{i(u_n x + v_n y)},$$

где u_n, v_n — пространственные частоты — проекции вектора \mathbf{k}_n n -й плоской волны на оси x, y : $u_n = k_x, v_n = k_y$. В случае непрерывного спектра плоских волн вместо (8.18) получаем

$$f_0(x, y) = \frac{1}{4\pi^2} \iint C_0(u, v) e^{i(u x + v y)} du dv,$$

где $C_0(u, v)$ — двумерное преобразование Фурье (пространственный спектр светового поля $f_0(x, y)$):

$$C_0(u, v) = \frac{1}{4\pi^2} \iint f_0(x, y) e^{-i(u x + v y)} dx dy.$$

Решение дифракционной задачи (искомое световое поле в «выходной плоскости» $z = \text{const} > 0$) принимает вид

$$f(x, y, z) = \sum c_n e^{i(u_n x + v_n y + \sqrt{k^2 - u_n^2 - v_n^2} \cdot z)}$$

или, в случае непрерывного спектра,

$$f(x, y, z) = \iint C_0(u, v) e^{i(u x + v y + \sqrt{k^2 - u^2 - v^2} \cdot z)} du dv.$$

Частотная характеристика свободного пространства является в общем случае также функцией двух переменных u, v

$$H(u, v) = e^{i\sqrt{k^2 - u^2 - v^2} \cdot z},$$

где $\sqrt{k^2 - u^2 - v^2}$ — z -компонента вектора \mathbf{k} плоской волны с пространственными частотами $u = k_x$, $v = k_y$ ($k_x^2 + k_y^2 + k_z^2 = k^2 = (\omega/c)^2$). Эта функция связывает между собой пространственные спектры световых полей во «входной» и «выходной» плоскостях

$$C(u, v, z) = C_0(u, v) e^{i\sqrt{k^2 - u^2 - v^2} \cdot z}.$$

Проиллюстрируем спектральный метод, рассмотрев задачу дифракции плоской волны на амплитудной синусоидальной решетке. Итак, пусть плоская волна единичной амплитуды, бегущая вдоль оси z , $f_s = 1 \cdot e^{ikz}$ падает на решетку с функцией пропускания (8.5) (решетка установлена в плоскости $z = 0$). Тогда на выходе из решетки в плоскости $z = 0_+$, непосредственно примыкающей к ней, имеем, используя (8.3):

$$f_0(x) = 1 + m \cos \Omega x.$$

Представим это граничное поле в виде (8.16). В данном случае с помощью формулы Эйлера находим

$$f_0(x) = 1 + \frac{m}{2} e^{i\Omega x} + \frac{m}{2} e^{-i\Omega x}.$$

Мы получили разложение (8.16): каждое слагаемое в граничном поле $f_0(x)$ «ответственно» за появление «своей» волны в области $z > 0$. Первое слагаемое дает волну, бегущую вдоль оси z : $f_1(x, z) = 1 \cdot e^{ikz}$. (Действительно, $f_1(x, z)$ есть решение уравнения Гельмгольца, удовлетворяющее граничному условию $f_1(x, z)|_{z=0} = 1$). Аналогичным образом, два других слагаемых ответственны за появление волн $f_2(x, z) = (m/2) e^{i(\Omega x + \sqrt{k^2 - \Omega^2} \cdot z)}$ и $f_3(x, z) = (m/2) e^{i(-\Omega x + \sqrt{k^2 - \Omega^2} \cdot z)}$, имеющих пространственные частоты $\pm \Omega$ и бегущих в направлениях $\sin \alpha = \pm \Omega/k$. Полагая, что период решетки $d = 2\pi/\Omega$ существенно больше длины

волны λ и, следовательно, пространственная частота Ω много меньше k .

Тогда, используя приближение (8.24), находим

$$f(x, z) = 1 \cdot e^{ikz} + \frac{m}{2} e^{ikz} e^{i[\Omega x - (z/2k)\Omega^2]} + \frac{m}{2} e^{ikz} e^{i[-\Omega x - (z/2k)\Omega^2]}.$$

После простых преобразований окончательно получаем

$$f(x, z) = e^{ikz} \left[1 + m \cos \Omega x \cdot \exp \left(-\frac{iz}{2k} \Omega^2 \right) \right].$$

Последнюю формулу можно использовать для анализа картины дифракции, возникающей на различных расстояниях z от решетки.

На расстояниях z_n , для которых $(z_n/2k)\Omega^2 = 2\pi n$ (т.е. $z_n = (4\pi k/\Omega^2)n = (2d^2/\lambda)n$), имеем: $e^{-i(z_n/2k)\Omega^2} = 1$, поэтому $f(x, z) = e^{ikz_n} f_0(x)$. Следовательно, с точностью до постоянного множителя e^{ikz_n} воспроизводится граничное световое поле. Воспроизводится, разумеется, наблюдаемая картина интенсивности: $I(x, z_n) = I_0(x)$. Найдем ее видность: $V = (I_{\max} - I_{\min})/(I_{\max} + I_{\min})$, полагая глубину модуляции m малой и пренебрегая членами порядка m^2 . Тогда

$$I(x, z_n) = |f(x, z_n)|^2 \approx 1 + 2m \cos \Omega x \quad \text{и} \quad V \approx 2m.$$

На расстояниях z_n , для которых $e^{i(z_n/2k)\Omega^2} = i$, т.е. $(z_n/2k)\Omega^2 = \pi/2 + 2\pi n$, имеем

$$f(x, z_n) = 1 + im \cos \Omega x.$$

Картина интенсивности: $I(x, z_n) \approx 1 = \text{const}$ — с точностью до членов порядка m^2 получаем равномерную (не зависящую от x) засветку плоскости наблюдения: $V = 0$. Таким образом, периодически по z изменяется видность наблюдаемой дифракционной картины. Причина этих изменений, разумеется, в различии фазовых набегов трех плоских волн, бегущих в области $z > 0$: осевой волны, бегущей вдоль оси z , и двух боковых волн, бегущих в направлениях $\sin \alpha = \pm \Omega/k$.

8.3. Дифракция на периодических структурах (эффект Талбота)

Рассмотрим теперь произвольную плоскую периодическую структуру. Такая структура может быть образована одинаковыми объектами, расположенными на равных расстояниях d друг от друга в плоскости $z = 0$, либо представлять собой экран с периодически расположенными отверстиями одинаковой формы. Пусть такая структура (будем называть ее решеткой) освещается слева

плоской нормально падающей волной. Тогда, согласно граничным условиям Кирхгофа (8.3), в плоскости $z = 0_+$, примыкающей к структуре справа, возникает пространственно-периодическое с периодом d световое поле $f_0(x)$ (рассматриваем для простоты функцию одной переменной), которое можно представить в виде ряда Фурье:

$$f_0(x) = \sum c_n e^{in(2\pi/d)x}. \quad (8.26)$$

Каждое слагаемое ряда ответственно за появление плоской волны в области $z > 0$, пространственная частота которой $u_n = n \cdot 2\pi/d$ определяет направление волнового вектора \mathbf{k}_n ($u_n = k \sin \alpha_n$) этой волны, соответственно, $\sin \alpha_n = u_n/k = n \lambda/d$. Комплексная амплитуда волны в плоскости наблюдения, отстоящей на расстоянии z от решетки, имеет вид (8.20). Напомним, что величина $\varphi_n = \sqrt{k^2 - u_n^2} z$ представляет собой набег фазы плоской волны, имеющей пространственную частоту u_n .

Будем полагать, что в сумме плоских волн (8.26), образующих периодическую структуру, существенно отличны от нуля лишь значения c_n , для которых $u_n = n \cdot 2\pi/d$ много меньше $|\mathbf{k}| = 2\pi/\lambda$. В этом случае, используя приближение (8.24), получаем

$$\varphi_n \approx kz - \frac{z}{2k} u_n^2. \quad (8.27)$$

Или, поскольку $u_n = n \cdot 2\pi/d$, то $\varphi_n = kz - (z/2k)(2\pi/d)^2 n^2$. Конечно, существенную роль играет лишь разность фазовых набегов, поэтому будем сравнивать набег фазы n -й плоской волны с набегом фазы φ_0 плоской волны, бегущей вдоль оси z (перпендикулярной плоскости решетки): $\varphi_0 = kz$. Мы получаем

$$\Delta\varphi_n = \varphi_0 - \varphi_n = \frac{z}{2k} \left(\frac{2\pi}{d}\right)^2 n^2 = \frac{n^2 \pi \lambda z}{d^2}. \quad (8.28)$$

Рассмотрим плоскость наблюдения, отстоящую от решетки на расстояние

$$z_1 = \frac{2d^2}{\lambda}. \quad (8.29)$$

В этой плоскости имеем $\Delta\varphi_n = 2\pi n^2$, т. е. относительный набег фазы всех плоских волн кратен величине 2π . Очевидно, что и разность фазовых набегов любых двух волн (с пространственными частотами $n_1 \cdot 2\pi/d$ и $n_2 \cdot 2\pi/d$), равная $2\pi(n_1^2 - n_2^2)$, также кратна 2π . Но изменение разности фаз колебаний на величину, кратную 2π , ничего не меняет в суммарном колебании. Мы пришли к замечательному результату: фазовые соотношения между слагаемыми плоскими волнами оказались одинаковыми как в плоскости, примыкающей к решетке (где сумма плоских

волн (8.26) образовала граничное периодическое поле $f_0(x)$, так и в плоскости (8.29). Одинаковость (с точностью до величины, кратной 2π) фазовых соотношений слагаемых плоских волн приводит к тому, что одинаков и результат интерференции этих плоских волн, т. е. световое поле в плоскости $z_1 = 2d^2/\lambda$ отличается от граничного поля $f_0(x)$ лишь постоянным фазовым множителем e^{ikz} :

$$f(x, z_1) = e^{ikz_1} f_0(x).$$

Мы наблюдаем в плоскости z_1 периодическую структуру, тождественно повторяющую граничное поле $f_0(x)$. Очевидно также, что такое восстановление (воспроизведение) изображения периодической структуры повторяется на расстояниях, кратных z_1 :

$$z_m = m \frac{2d^2}{\lambda} \quad (m = 1, 2, \dots). \quad (8.30)$$

Описанный эффект называют *эффектом самовоспроизведения*, или по имени первооткрывателя У. Талбота (1800–1877) *эффектом Талбота*.

Заметим, что периодически по z воспроизведение граничной периодической структуры происходит в области френелевской дифракции $p^2 = (\lambda z_m/d^2) = 2m$. Реально любая структура имеет конечные размеры (содержит конечное число периодов d), что не учитывалось нами при изучении эффекта. Рано или поздно с ростом m конечные размеры решетки начинают приводить к ухудшению эффекта самовоспроизведения.

8.4. Область геометрической оптики

Опыт показывает, что в некоторых случаях (когда плоскость наблюдения отстоит недалеко от препятствия или когда размеры препятствия достаточно велики) наблюдаемая картина оказывается подобной картине поля на границе — ситуация, показанная на рис. 8.1 а: наблюдаемая картина получается *геометрическим проецированием* отверстия на плоскость наблюдения. Итак, выясним условия, когда выводы, основанные на представлениях геометрической оптики, справедливы.

Пусть щель в непрозрачном экране ширины b освещается плоской нормально падающей волной. Тогда граничное поле $f_0(x)$ имеет вид, показанный на рис. 8.9 а. Его пространственный спектр $C_0(u)$ (фурье-преобразование функции $f_0(x)$) показан на рис. 8.9 б; он имеет точно такой же вид, что и спектр прямоугольного импульса, который мы изучали в гл. 1.4.

Функция $C_0(u)$ заметно отлична от нуля в пределах *главного максимума*, в области $|u| < 2\pi/b$. Поскольку мы полагаем, что

$b > \lambda$, то $|u| < k = 2\pi/\lambda$. Итак, интервал пространственных частот плоских волн имеющих заметную амплитуду, ограничен величиной $|\Delta u| < 2\pi/b$.

Эта оценка справедлива по порядку величины и в общем случае для произвольного препятствия с характерным размером b , что следует из соотношения неопределенностей.

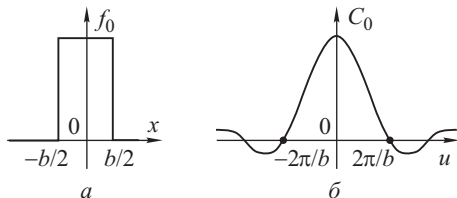


Рис. 8.9

Для плоской волны, бегущей вдоль оси z ($u = 0$), набег фазы при распространении до плоскости наблюдения равен kz . Для волны с пространственной частотой u он равен $\varphi(u) = \sqrt{k^2 - u^2} z$, или приближенно: $\varphi(u) \approx kz - (z/2k)u^2$. Разность фазовых набегов $\Delta\varphi \approx (z/2k)u^2$.

Пространственные частоты плоских волн, бегущих от щели (интерес представляют лишь волны, амплитуды которых имеют заметную величину) не превышают величины $|u| < 2\pi/b$, поэтому разность фазовых набегов не превышает величины

$$\Delta\varphi < \frac{z}{2k} \left(\frac{2\pi}{b} \right)^2 = \frac{\pi \lambda z}{b^2}. \quad (8.31)$$

Если эта разность много меньше величины π , то можно считать, что все плоские волны из интересующего нас интервала $|u| < 2\pi/b$ приобретают один и тот же фазовый набег, равный kz . Следовательно, фазовые соотношения между плоскими волнами, образующими граничное поле $f_0(x)$, и теми же плоскими волнами в плоскости наблюдения $z = \text{const} > 0$ практически одинаковы. Значит, интерференция этих плоских волн дает в плоскости наблюдения суммарное поле, которое (с точностью до фазового множителя e^{ikz}) тождественно повторяет граничное поле $f_0(x)$:

$$f(x, z) = e^{ikz} f_0(x).$$

При этом картина интенсивности $I(x, z)$ тождественно повторяет картину на границе $I(x, z) = I_0(x) = |f_0(x)|^2$.

Как следует из (8.31), условие $\Delta\varphi \ll \pi$ выполнено, если

$$p^2 = \frac{\lambda z}{b^2} \ll 1. \quad (8.32)$$

Это и есть *условие геометрической оптики*. При любых фиксированных z и b это условие выполняется, если $\lambda \rightarrow 0$, а при любом фиксированном λ законы геометрической оптики работают, как следует из неравенства (8.32), при достаточно малых z и (или) при достаточно больших b .

8.5. Дифракция Фраунгофера

Используем общее соотношение (8.21) для изучения дифракционной картины на большом удалении от препятствия, в области значений волнового параметра $p = \sqrt{\lambda z}/b \gg 1$ (область дифракции Фраунгофера (по имени немецкого физика Й. Фраунгофера (1787–1826))). Для оценки интеграла (8.21) воспользуемся методом, называемым *методом стационарной фазы*. Показатель экспоненты в (8.21), имеющий вид

$$\varphi(u) = \sqrt{k^2 - u^2} \cdot z + ux, \quad (8.33)$$

разложим в ряд Тейлора в окрестности некоторой точки $u = u_0$:

$$\begin{aligned} \varphi(u) = \varphi(u_0) + \varphi'(u_0)(u - u_0) + \frac{1}{2} \varphi''(u_0)(u - u_0)^2 + \\ + \frac{1}{6} \varphi'''(u_0)(u - u_0)^3 + \dots \end{aligned} \quad (8.34)$$

Выберем точку u_0 так, чтобы

$$\varphi'(u_0) = 0.$$

Используя (8.33), находим $\varphi'(u) = -\frac{zu}{\sqrt{k^2 - u^2}} + x$ и, приравнявая производную $\varphi'(u)$ нулю, получаем

$$u = u_0 = \frac{kx}{\sqrt{z^2 + x^2}} = \frac{kx}{r}, \quad (8.35)$$

где $r = \sqrt{z^2 + x^2}$ — расстояние от начала координат граничной плоскости $z = 0_+$ до точки наблюдения (x, z) . Значение u_0 является точкой экстремума функции $\varphi(u)$. Ее называют *стационарной точкой*: в ее окрестности фаза $\varphi(u)$ изменяется наиболее медленно.

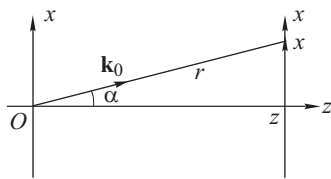


Рис. 8.10

Из всего спектра плоских волн, бегущих от границы $z = 0_+$ в области $z > 0$, стационарная точка u_0 выделяет плоскую волну с волновым вектором \mathbf{k}_0 (рис. 8.10): $u_0 = k_{0x} = k_0 \sin \alpha$, где α — направление на точку наблюдения ($\sin \alpha = x/r$).

Дважды дифференцируя функцию (8.33), найдем $\varphi''(u)$:

$$\varphi''(u) = -\frac{k^2 z}{(k^2 - u^2)^{3/2}}.$$

Используя далее значение u_0 , имеем

$$\varphi(u_0) = kr, \quad \varphi''(u_0) = -\frac{r^3}{kz^2}. \quad (8.36)$$

Интерес, как правило, представляет область значений x в плоскости наблюдения $z = \text{const}$, размер которой мал в сравнении с z , поэтому $r \approx z$ и

$$\varphi''(u_0) \approx \frac{r}{k}.$$

Используя найденные значения $\varphi(u_0)$ и $\varphi''(u_0)$, получаем следующее приближение: $\varphi(u) \approx kr - (r/2k)(u - u_0)^2$. Теперь решение дифракционной задачи выражается формулой:

$$f(x, z) = \frac{1}{2\pi} e^{ikr} \int_{-\infty}^{\infty} C_0(u) \exp \left[-i \frac{r}{2k} (u - u_0)^2 \right] du. \quad (8.37)$$

Далее мы убедимся, что остаточный член ряда Тейлора $(1/6) \varphi'''(u_0)(u - u_0)^3$ вносит пренебрежимо малую поправку.

После замены переменных $u - u_0 = v$ получаем

$$f(x, z) = \frac{1}{2\pi} e^{ikr} \int_{-\infty}^{\infty} C_0(u_0 + v) \exp \left(-i \frac{r}{2k} v^2 \right) dv \quad (8.38)$$

(стационарной точке соответствует значение $v = 0$). Напомним, что интеграл $\int_{-\infty}^{\infty} e^{-i\alpha v^2} dv$ мы исследовали ранее с помощью векторной диаграммы, называемой *спиралью Корню*. Значение интеграла определяется вектором, соединяющим фокусы спирали. Мы имели

$$\int_{-\infty}^{\infty} e^{-i\alpha v^2} dv = \sqrt{\frac{\pi}{\alpha}} e^{-i\pi/4}. \quad (8.39)$$

В нашем случае $\alpha = r/(2k)$, причем основной вклад в интеграл определяется длиной одного полувитка спирали: вектор, проведенный из начала спирали (точка O) в конец первого полувитка (точка B_1), мало отличается по длине и направлению от вектора, проведенного в фокус спирали (рис. 8.11). Точке B_1 соответствует область интегрирования, определяемая условием

$$\alpha v_1^2 = \pi, \quad \text{т. е.} \quad v_1 = \sqrt{\frac{\pi}{\alpha}} = \sqrt{\frac{2\pi k}{r}}.$$

Фокусам спирали соответствуют бесконечные пределы интегрирования, откуда следует оценка

$$\int_{-\infty}^{\infty} e^{-i\alpha v^2} dv \approx \int_{-v_1}^{v_1} e^{-i\alpha v^2} dv.$$

Потребуем теперь, чтобы в области интегрирования (окрестности стационарной точки) $|v| < v_1$ функция $C_0(u_0 + v)$

оставалась постоянной величиной (равной $C_0(u_0)$). Тогда

$$\begin{aligned} \int_{-v_1}^{v_1} C_0(u_0 + v) e^{-i\alpha v^2} dv &\approx C_0(u_0) \int_{-v_1}^{v_1} e^{-i\alpha v^2} dv \approx \\ &\approx C_0(u_0) \int_{-\infty}^{\infty} e^{-i\alpha v^2} dv. \end{aligned} \quad (8.40)$$

Это требование будет выполнено, если область значений Δv , в которой функция $C_0(u_0 + v)$ может заметно измениться, окажется много больше v_1 . Эту область можно оценить с помощью

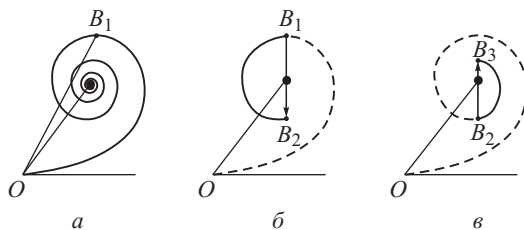


Рис. 8.11

соотношения неопределенностей $\Delta v \cdot b \approx 2\pi$, где b — характерный масштаб изменения граничного поля $f_0(x)$ (например, размер отверстия в непрозрачном экране).

Итак, достаточно потребовать, чтобы

$$\frac{2\pi}{b} \gg v_1 = \sqrt{\frac{\pi}{\alpha}} = \sqrt{\frac{2\pi k}{r}}, \quad (8.41)$$

т. е.

$$b \ll \sqrt{\lambda r} \approx \sqrt{\lambda z}. \quad (8.42)$$

Последнее неравенство и есть условие фраунгоферовой дифракции. Далее следует принять во внимание, что на интервалах изменения v , соответствующих второму и третьему, четвертому и пятому и т. д. виткам спирали, функция $C_0(u_0 + v)$ также остается неизменной в силу условия (8.41). Например, на интервале изменения v от v_1 (соответствующей точке B_1) до v_3 (соответствующей точке B_3) можно принять $C_0(u_0 + v) \approx C_0(u_0 + v_2)$ при $v_1 < |v| < v_3$ (v_2 — примерно середина интервала (v_1, v_3)). На этом интервале можно приближенно считать, что

$$\int_{v_1}^{v_3} C_0(u_0 + v) e^{-i\alpha v^2} dv \approx C_0(u_0 + v_2) \int_{v_1}^{v_3} e^{-i\alpha v^2} dv.$$

Последний интеграл представляет собой сумму антиколлинеарных векторов $\overrightarrow{B_1 B_2}$ и $\overrightarrow{B_2 B_3}$, пренебрежимо малую в сравнении с вектором, проведенным из начала O в фокус спирали. Понятно, что если спектр $C_0(u)$ описывается достаточно плавно меняющейся функцией, то вкладом области интегрирования $|v| > v_1$ можно пренебречь и окончательно, с учетом (8.39), (8.40) и имея ввиду (8.35), получаем

$$f(x, z) = \frac{e^{ikr}}{\sqrt{i\lambda r}} C_0\left(\frac{kx}{r}\right). \quad (8.43)$$

Остается оценить величину остаточного члена ряда Тейлора в области $|v| \leq v_1$.

Найдем: $\varphi'''(u) = -3k^2 z(k^2 - u^2)^{-5/2} u$, а также значение $\varphi'''(u_0)$:

$$\varphi'''(u_0) = -\frac{3x}{k^2 z^4} (z^2 + x^2)^2 = -\frac{3x r^4}{k^2 z^4} \approx -\frac{3x}{k^2},$$

так как $z \approx r$. Остаточный член ряда

$$\frac{1}{6} \varphi'''(u_0)(u - u_0)^3 = -\frac{x}{2k^2} v^3 \leq \frac{x}{2k^2} \left(\sqrt{\frac{2\pi k}{r}} \right)^3 = \pi \frac{x}{r} \sqrt{\frac{\lambda}{r}} \ll \pi.$$

Напомним, что область применимости граничных условий Кирхгофа: $r \gg \lambda$ и, так как $x < r$, то ясно, что полученное нами приближение справедливо.

Итак, мы получили чрезвычайно простой способ определения дифракционной картины Фраунгофера. Поле в точке наблюдения (x, z) , удаленной в область $z \gg \frac{b^2}{\lambda}$, определяется значением спектра $C_0(u)$ граничного поля в стационарной точке $u_0 = kx/r$. Более детально дифракция Фраунгофера будет исследована в § 8.9.

8.6. Принцип Гюйгенса–Френеля

Принцип Гюйгенса–Френеля является рецептом для решения дифракционных задач, отличным от изложенного выше спектрального подхода.

Пусть волна света, созданная источниками, расположенными в области $z < 0$, достигла плоскости $z = 0$ (каких-либо препятствий на пути света пока еще нет; рис. 8.12а). Световое поле в этой плоскости нам известно. Пусть его комплексная амплитуда есть $f_0(x, y) = a_0(x, y)e^{i\varphi_0(x, y)}$, где функции $a_0(x, y)$ и $\varphi_0(x, y)$ описывают распределение амплитуд и фаз колебаний в плоскости $z = 0$. Что происходит с волной далее, по мере ее распространения в области $z > 0$, как изменяется структура

волны — распределение амплитуд и фаз колебаний, что собой представляет световое колебание в некоторой точке наблюдения P ?

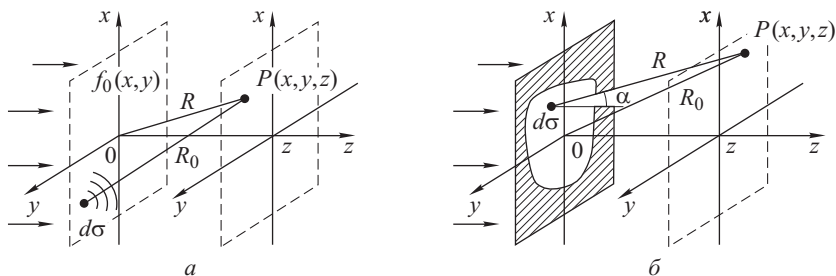


Рис. 8.12

Согласно принципу Гюйгенса каждую точку (x, y) плоскости $z = 0$, куда пришла волна (каждую малую площадку $d\sigma$), можно рассматривать как источник вторичной волны. То есть можно представить себе, что волна возбуждает колебания некоторого фиктивного источника, который и переизлучает вторичную волну. Френель дополнил принцип Гюйгенса, предложив рассматривать световое колебание в любой точке наблюдения P в области $z > 0$ как результат интерференции этих вторичных волн.

Какие свойства приписывает принцип Гюйгенса–Френеля фиктивным источникам и вторичным волнам, которые переизлучаются ими?

Поясним принцип Гюйгенса–Френеля, вернувшись к рассмотренной ранее задаче — дифракции на непрозрачном экране с отверстием (рис. 8.12б). Маленькая площадка $d\sigma$, расположенная в точке (x, y) на открытой части волнового фронта (т. е. в области отверстия), рассматривается как источник сферической волны

$$\frac{a(x, y)}{R} e^{i(kR + \varphi(x, y))},$$

где R — расстояние от источника до точки наблюдения P , т. е. для вычисления вклада, который дает эта площадка в суммарное колебание, нужно учесть ослабление амплитуды (множитель $1/R$) и набег фазы (множитель e^{ikR}) как в сферической волне. При этом предполагается, что амплитуда излучения $a(x, y)$ площадки $d\sigma$ и ее начальная фаза $\varphi(x, y)$ задаются амплитудой $a_0(x, y)$ и фазой $\varphi_0(x, y)$ колебания, созданного в точке (x, y) освещающей волной. Напомним, что, согласно принятым граничным условиям (8.2), на открытой части волнового фронта, т. е. в области отверстия, волна не искажается препятствием, причем работают лишь вторичные источники, находящиеся на открытой части, не затененной непрозрачным экраном.

Разумеется, амплитуда $a(x, y)$ пропорциональна также размеру переизлучающей площадки $d\sigma$, т. е. «числу» вторичных когерентно переизлучающих фиктивных источников внутри площадки. Наконец, предполагается, что амплитуда колебания в точке наблюдения пропорциональна видимой из этой точки величине площадки $d\sigma$, т. е. пропорциональна $d\sigma \cos \alpha$ (см. рис. 8.12; $\cos \alpha$ — множитель наклона). Итак, $a(x, y) \sim a_0(x, y) d\sigma \cos \alpha$, а $\varphi(x, y) = \varphi_0(x, y)$. Таким образом, вклад элемента $d\sigma$ можно записать в виде

$$\frac{a_0(x, y)}{R} e^{i(kR + \varphi_0(x, y))} d\sigma \cos \alpha = f_0(x, y) \frac{e^{ikR}}{R} \cos \alpha \cdot d\sigma.$$

Полное световое колебание есть результат интерференции всех вторичных волн, т. е. волн, посылаемых всеми площадками $d\sigma$, расположенными в области отверстия:

$$f(P) = K_0 \int_S f_0(x, y) \frac{e^{ikR}}{R} \cos \alpha d\sigma, \quad (8.44)$$

где K_0 — некоторый коэффициент пропорциональности, который еще подлежит определению, $f(P)$ — комплексная амплитуда в точке наблюдения P . Формула (8.44) представляет собой количественную формулировку принципа Гюйгенса–Френеля, которую можно использовать для решения конкретных дифракционных задач.

8.7. Дифракция Френеля. Дифракционные задачи с осевой симметрией

Круглое отверстие в непрозрачном экране. Применим принцип Гюйгенса–Френеля для нахождения светового колебания в точке наблюдения P , находящейся на оси круглого отверстия в непрозрачном экране (рис. 8.13). Отверстие освещается плоской нормально падающей волной амплитуды A_0 .

Полагаем, что радиус отверстия r_0 много меньше расстояния z до точки наблюдения, поэтому множитель наклона в (8.44) можно считать близким к единице: $\cos \alpha \approx 1$. Кроме того, при том же условии $r_0 \ll z$ амплитудный множитель $1/R$, учитывающий сферическую расходямость волн от вторичных источников, расположенных в области отверстия, можно считать приближенно одинаковым для всех вторичных источников и равным $1/R \approx 1/z$. Тогда из (8.44), находим

$$f(P) = K_0 \frac{A_0}{z} \int_{(\sigma)} e^{ikR} d\sigma. \quad (8.45)$$

Разумеется, расстояние R в фазовом множителе e^{ikR} необходимо вычислить аккуратнее: ведь ошибка в определении R на величину $\lambda/2$ приводит к ошибке в фазе колебания в точке наблюдения на величину π , что недопустимо (ошибки в фазах складываемых колебаний должны быть малы в сравнении с π). Расстояние R от некоторого вторичного источника $d\sigma$, находящегося на расстоянии ρ от центра отверстия до точки P , оценим так (рис. 8.13):

$$R = (z^2 + \rho^2)^{1/2} = z \left(1 + \frac{\rho^2}{z^2} \right)^{1/2} \approx z + \frac{\rho^2}{2z}. \quad (8.46)$$

Тогда из (8.45) получим

$$f(P) = K_0 \frac{A_0 e^{ikz}}{z} \int_{(\sigma)} \exp \left(\frac{ik}{2z} \rho^2 \right) d\sigma. \quad (8.47)$$

Рассмотрим все элементарные площадки, т. е. все вторичные источники, находящиеся на фиксированном расстоянии ρ от центра отверстия в пределах тоненького колечка радиуса ρ малой толщины $d\rho$ (рис. 8.13). Площадь $d\sigma$ такого колечка $d\sigma = 2\pi\rho d\rho$. Интеграл в последнем выражении принимает вид

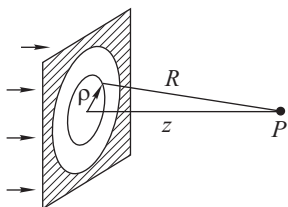


Рис. 8.13

$$\begin{aligned} & \int_0^{r_0} \exp \left(\frac{ik}{2z} \rho^2 \right) \cdot 2\pi\rho d\rho = \\ & = \pi \int_0^{r_0^2} \exp \left(\frac{ik}{2z} \xi \right) d\xi, \quad (8.48) \end{aligned}$$

где введено обозначение $\rho^2 = \xi$ ($d\xi = 2\rho d\rho$).

Нахождение суммарного колебания в точке наблюдения сводится к суммированию вкладов элементарных колечек. Разбив все отверстие на такие элементарные колечки одинаковой площади $\Delta\sigma = 2\pi\rho\Delta\rho$, мы приходим к задаче сложения колебаний одинаковой амплитуды $\Delta\xi$, фазы которых растут пропорционально номеру кольца: $\varphi_n = (nk/2z) \Delta\xi$. Сдвиг по фазе обусловлен тем, что каждое последующее колечко (большого радиуса) отстоит от точки наблюдения на большее расстояние и сдвиг по фазе связан с разностью расстояний ΔR : $\Delta\varphi = k\Delta R$. Суммарное колебание $\sum_n \Delta\xi e^{(ink/2z)\Delta\xi}$ в пределе, при $\Delta\xi \rightarrow 0$ дает значение интеграла. Векторная диаграмма, которая и позволит нам вычислить интеграл (8.48) — это спираль Френеля (см. § 1.3).

Итак, колебание, созданное элементом $d\sigma$, находящимся в центре отверстия ($\xi = 0$) изобразим вектором, длина которого равна амплитуде колебания, а его направление горизонтально (фаза колебания $\varphi = (k/2z)\xi = 0$). Вклад n -го колечка,

($\xi_n = n\Delta\xi$) есть $\Delta\xi e^{(ink/2z)\Delta\xi}$; он изображается вектором той же длины $\Delta\xi$ с углом наклона $\varphi_n = (nk/2z)\Delta\xi$.

Мы получаем цепочку векторов, в которой каждый последующий вектор повернут относительно предыдущего на один и тот же угол $\Delta\varphi = (k/2z)\Delta\xi$ (рис. 8.14). Последний из изображенных на рис. 8.14 векторов s_N повернут относительно первого вектора s_1 на угол π , т. е. последнему колечку соответствует фаза колебания, равная π . Получаем

$$\frac{k}{2z}\xi_N = \pi, \quad \xi_N = \frac{2\pi z}{k} = \lambda z. \quad (8.49)$$

Соответствующий радиус кольца есть:

$$\rho_1 = \sqrt{\xi_N} = \sqrt{\lambda z}. \quad (8.50) \quad \text{Рис. 8.14}$$

Он называется радиусом *первой зоны Френеля* (т. е. первая зона Френеля — круг радиуса ρ_1). Разность фаз колебаний s_N и s_1 , равная π , означает, что расстояние R_1 от края первой зоны Френеля до точки наблюдения P больше, чем расстояние от центра отверстия до точки P на $\lambda/2$: $R_1 = z + \lambda/2$. Суммарный вклад первой зоны Френеля изображается вектором A_1 — суммой всей цепочки векторов на рис. 8.14.



Рис. 8.15

Продолжая далее такое разбиение на тонкие колечки, мы получим векторную диаграмму, показанную на рис. 8.15. Последний вектор в этой цепочке векторов антиколлинеарен первому, т. е. последний вектор изображает вклад колечка, которое отстоит от точки наблюдения на расстояние R_2 , которое больше расстояния R_1 на $\lambda/2$: $R_2 = R_1 + \lambda/2$ или $R_2 = z + \lambda$.

Ясно, что колебание, созданное последним колечком, сдвинуто по фазе на 2π , т. е. синфазно с колебанием, созданным малой площадкой, расположенной в центре отверстия (вектор s_1 на рис. 8.14). Разности расстояний λ соответствует разность фаз 2π , т. е. $(k/2z)\xi = 2\pi$, откуда $\xi = 2\lambda z$, а радиус последнего колечка (которому отвечает последний вектор на диаграмме рис. 8.15) равен $\rho_2 = \sqrt{2\lambda z}$.

Кольцевая зона на отверстии между ρ_1 и ρ_2 называется *второй зоной Френеля*, ρ_2 — это внешний радиус второй зоны Френеля. Суммарный вклад второй зоны изображается вектором A_2 (рис. 8.15), который антиколлинеарен вектору A_1 на рис. 8.14, т. е. колебание, созданное в точке наблюдения второй зоной Френеля, противофазно колебанию от первой зоны. Если бы вклады всех элементарных колечек имели равную амплитуду, то амплитуды колебаний $|A_1|$ и $|A_2|$ были бы одинаковы и колебания первой и второй зон погасили бы друг друга полностью: $A_1 + A_2 = 0$. В действительности с ростом радиуса колечка амплитуда его вклада постепенно уменьшается: роль

при этом играет как постепенное, очень медленное уменьшение множителя наклона $\cos \alpha$, так и уменьшение амплитуды с ростом расстояния $\sim 1/R$. Итак, каждый элементарный вектор в цепочке векторов на рис. 8.15 чуть меньше каждого вектора в цепочке на рис. 8.14. Поэтому и суммарный вектор \mathbf{A}_2 чуть меньше вектора \mathbf{A}_1 . Векторная диаграмма, изображающая вклад двух зон Френеля показана на рис. 8.16. Суммарный вклад двух зон хотя и чрезвычайно мал, но все же отличен от нуля. Итак, интенсивность света в точке P близка к нулю, если отверстие содержит две зоны Френеля: $r_0 = \sqrt{2\lambda z}$.

Пусть радиус отверстия постепенно увеличивается и, наконец, становится равным $\rho_3 = \sqrt{3\lambda z}$. Можно легко проверить, что край отверстия при этом отстоит от точки P на расстояние $R_3 = z + 3\lambda/2$, а колебание, созданное тонким колечком, примыкающим к краю, сдвинуто по фазе на величину 3π по отношению к колебанию от центрального элемента отверстия (т. е. эти колебания противофазны). ρ_3 — это внешний радиус *третьей зоны Френеля* (кольцевой зоны между ρ_2 и ρ_3). Очевидно, суммарный вклад третьей зоны \mathbf{A}_3 лишь чуть меньше по амплитуде вклада второй зоны. Векторная диаграмма, изображающая вклад трех зон Френеля представлена на рис. 8.17. Амплитуда суммарного колебания лишь немного меньше амплитуды \mathbf{A}_1 колебания, созданного отверстием в одну зону Френеля.

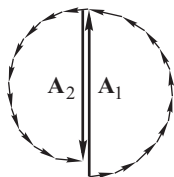


Рис. 8.16



Рис. 8.17

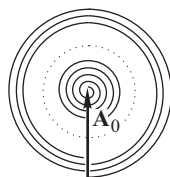


Рис. 8.18

При дальнейшем увеличении радиуса отверстия «открываются» все новые и новые зоны Френеля: начинают «работать» все новые и новые витки очень медленно скручивающейся спирали, которая показана на рис. 8.18 и называется *спиралью Френеля*. Радиус m -й зоны Френеля равен

$$\rho_m = \sqrt{m\lambda z}. \quad (8.51)$$

Ясно, что если отверстие содержит четное число зон Френеля, то интенсивность света в точке наблюдения близка к нулю, а если нечетное, то интенсивность почти такая же, как и при отверстии в одну зону Френеля.

Конечно, с ростом радиуса отверстия «скручивание» спирали проявляется все сильнее и сильнее, множитель наклона $\cos \alpha$

уменьшается с ростом α , сказывается также увеличение расстояния между вторичными источниками, лежащими вблизи края отверстия и точкой P — все это приводит к уменьшению амплитуд колебаний от этих источников в точке наблюдения. Понятно, что если работают все зоны Френеля (т. е. непрозрачный экран отсутствует), то все витки спирали на рис. 8.18 дают свой вклад, а вектор A_0 , проведенный из начала спирали в ее фокус есть ни что иное, как амплитуда колебаний в точке P при отсутствии какого-либо препятствия на пути волны. Обратите внимание на удивительное обстоятельство: амплитуда колебаний в точке наблюдения при отсутствии препятствия на пути волны A_0 вдвое меньше, чем амплитуда A_1 колебания, созданного отверстием в одну зону Френеля: $A_0 \approx A_1/2$, а интенсивность вчетверо меньше: $I_0/I_1 \approx 1/4$.

Построенная нами векторная диаграмма позволяет попутно определить константу K_0 в формуле (8.44). Длину цепочки векторов, образующих дугу полуокружности на рис. 8.14, можно найти с помощью (8.49). Действительно, каждый элементарный вектор в цепочке имеет длину $\Delta\xi$, а длина всей цепочки $\xi_N = N\Delta\xi$. Согласно (8.49) она равна: $\xi_N = 2\pi z/k = \lambda z$. Радиус этой полуокружности представляет собой не что иное, как значение интеграла (8.48) при $r_0 \rightarrow \infty$ (что соответствует ситуации, когда препятствие на пути волны отсутствует). Мы получаем

$$\pi \int_0^\infty \exp\left(\frac{ik}{2z}\xi\right) d\xi = \lambda z i.$$

Мнимая единица учитывает, что суммарный вектор на рис. 8.18 повернут на $\pi/2$ ($i = e^{i\pi/2}$) относительно первого вектора, фазу которого мы приняли за нуль. Подставляя полученное значение в правую часть (8.47), а также учитывая, что при отсутствии препятствия $f(P) = A_0 e^{ikz}$, находим $A_0 e^{ikz} = K_0(A_0 e^{ikz}/z) \lambda z i$, откуда получаем $K_0 = 1/(i\lambda)$.

Количественная формулировка принципа Гюйгенса–Френеля (8.44) принимает вид

$$f(P) = \frac{1}{i\lambda} \int_S f_0(x, y) \frac{e^{ikR}}{R} \cos \alpha dS. \quad (8.52)$$

Зонная пластинка Френеля. Поставим на пути плоской, нормально падающей волны прозрачную пластинку и фиксируем положение точки наблюдения P на расстоянии z за пластинкой (рис. 8.19). Нарисуем на пластинке кольцевые зоны Френеля согласно формуле (8.51). Далее, все нечетные зоны Френеля (первую, третью, пятую и т. д.) оставим прозрачными, а все четные зоны (вторую, четвертую, шестую и т. д.) зачерним, сделав

их непрозрачными для света. Такая пластинка (с чередующимися прозрачными и непрозрачными зонами Френеля) называется *зонной пластинкой Френеля*.

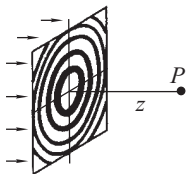


Рис. 8.19

В векторной диаграмме на рис. 8.18 остаются «работающими» только полувитки, отвечающие нечетным зонам; полувитки четных зон «выбывают из игры», поскольку заполняющие их вторичные источники оказались затененными. Оставим на векторной диаграмме лишь работающие полувитки нечетных зон, пристроив их последовательно друг к другу (без изменения ориентации всех элементарных векторов, рис. 8.20). Мы видим, что вклады всех оставшихся нечетных зон A_1, A_3, A_5, \dots изображаются коллинеарными векторами, т. е. колебания, созданные в точке наблюдения всеми нечетными зонами, оказываются синфазными. Амплитуда результирующего колебания A равна при этом сумме амплитуд слагаемых колебаний: $A = A_1 + A_3 + A_5 + \dots$, а интенсивность $I = A^2 = (A_1 + A_3 + A_5 + \dots)^2$. Мы видим, что зонная пластинка Френеля обладает фокусирующими свойствами, существенно увеличивая интенсивность света в точке P , которая является точкой фокусировки зонной пластинки. Заметим, что размеры пластинки (ее радиус R_0) обычно существенно меньше, чем расстояние z до точки фокусировки, поэтому вклады всех зон Френеля примерно равны по амплитуде и равны A_1 . Найдем выигрыш от фокусировки — отношение интенсивности света I в фокусе зонной пластины к интенсивности света I_0 , падающего на пластинку (при отсутствии пластинки такая же интенсивность I_0 была бы и в точке P). Итак, число зон Френеля, укладывающихся на пластинке радиуса R_0 найдем из условия: $R_0 = \sqrt{m\lambda z}$, откуда $m = R_0^2/(\lambda z)$ (из них половина зон — нечетные). Каждая нечетная зона дает вклад, равный примерно A_1 , поэтому суммарная амплитуда колебаний в точке P равна $A = A_1 m/2 = A_1 R_0^2/(2\lambda z) = 2A_0 R_0^2/(2\lambda z) = A_0 R_0^2/(\lambda z)$.

Мы находим:

$$\frac{I}{I_0} = \left(\frac{A}{A_0} \right)^2 = \frac{R_0^4}{(\lambda z)^2}.$$

Это и есть выигрыш от фокусировки. Например, при $R_0 = 1$ см, $z = 10$ см и $\lambda = 5 \cdot 10^{-5}$ см (зеленый цвет) получаем $I/I_0 \approx 4 \cdot 10^6$!

Подумайте, каким образом можно еще увеличить выигрыш от фокусировки, используя вклады четных зон Френеля? Какова максимально возможная интенсивность в точке наблюдения, если

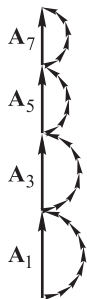


Рис. 8.20

размер отверстия в непрозрачном экране равен радиусу первой зоны Френеля (отверстие освещается плоской волной интенсивности I_0)?

Пятно Пуассона. Пусть на пути плоской волны расположен непрозрачный круглый диск (либо круглый шарик). Какова интенсивность света в точке P , расположенной за диском на оси, проходящей через его центр (рис. 8.21)? Если расстояние z столь мало, что выполняются условия геометрической оптики ($\sqrt{\lambda z}/R_0 \ll 1$, R_0 — радиус диска), то в плоскости наблюдения мы получим резкую границу света и тени, точка P находится как раз посередине круговой области тени и интенсивность равна нулю. Будем постепенно удалять точку наблюдения от диска. Число зон Френеля, перекрываемых диском, будет постепенно уменьшаться. К чему это приводит?

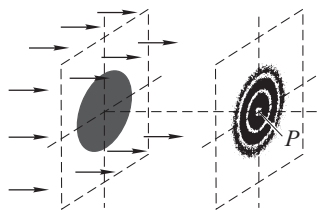


Рис. 8.21

Прежде, чем ответить на этот вопрос, напомним, что при отсутствии препятствия на пути волны колебание в точке наблюдения изображается вектором \mathbf{A}_0 (т. е. амплитуда колебаний определяется длиной вектора \mathbf{A}_0), проведенным из начала спирали в ее фокус (см. рис. 8.18). Понятно, что это колебание можно

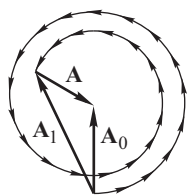


Рис. 8.22

представить как сумму колебаний, которые создаются сравнительно небольшим количеством нескольких первых зон Френеля (вектор \mathbf{A}_1 на рис. 8.22 изображает вклад трех целых зон плюс примерно треть четвертой зоны — три и одна треть полувитков спирали Френеля) и колебания, созданного всеми остальными зонами — их вклад изображается вектором \mathbf{A} на рис. 8.22. Действительно, мы видим, что сумма векторов \mathbf{A}_1 и \mathbf{A} дает нам вектор \mathbf{A}_0 .

Если теперь на пути волны установить непрозрачный диск, перекрывающий вклад нескольких первых зон Френеля (например, тех самых трех и одной трети зон), то колебание в точке наблюдения создается теперь лишь оставшимися периферийными зонами, вклад которых изображается вектором \mathbf{A} . Мы видим, что если число перекрытых диском зон не слишком велико, так что скручиванием спирали можно пренебречь, то длина вектора \mathbf{A} , т. е. амплитуда колебаний в центре тени от диска практически такая же, какая была бы в точке P при отсутствии диска.

Из сказанного выше ясно, что совершенно неважно, каков конкретный радиус диска (или шарика). Длина вектора \mathbf{A} на рис. 8.22 практически не изменилась бы, если бы диск перекрыл не три с небольшим зоны Френеля, а, скажем, семь с небольшим

Понятно, что и суммарное колебание, созданное в точке P полоской площадок $d\sigma$, отстает по фазе на величину $\Delta\varphi$ от колебания, созданного полоской площадок $d\sigma_0$, находящихся на оси η . Итак, колебание от интересующей нас полоски вторичных источников в точке наблюдения P можно записать в виде

$$a_0 d\xi \exp\left(\frac{ik}{2z} \xi^2\right). \quad (8.54)$$

Разбив всю щель на такие элементарные полоски, параллельные краям щели, и сложив колебания от всех элементарных полосок, заполняющих щель (от b_1 до b_2), найдем

$$f(P) = a_0 \int_{b_1}^{b_2} \exp\left(\frac{ik}{2z} \xi^2\right) d\xi. \quad (8.55)$$

Мы приходим к спирали Корню (см. § 1.3), которая служит для оценки интеграла (8.55).

Последний из изображенных на рис. 8.24 векторов \mathbf{a}_1 антиколлинеарен первому вектору \mathbf{a}_0 . Он изображает вклад полоски вторичных источников, который противофазен вкладу первой полоски (находящейся на оси η). Мы получаем $(k/2z)\xi_n^2 = \pi$, откуда находим расстояние от края этой (n -й) полоски до оси η : $\xi_n = \sqrt{2\pi z/k} = \sqrt{\lambda z}$. Суммарный вклад всей полосы (от оси η до n -й полоски) изображается вектором \mathbf{A}_1 , введенным из начала первого вектора \mathbf{a}_0 в конец последнего вектора \mathbf{a}_1 . Будем продолжать наше построение, пристраивая на векторной диаграмме вклады последовательно расположенных полосок, все далее отстоящих от оси η . Мы получим еще один полувиток (рис. 8.25), причем вектор \mathbf{a}_2 изображает вклад полоски, отстоящей от точки наблюдения на расстояние $z + \lambda$ (оно больше, чем расстояние от оси η до точки наблюдения, на длину волны λ), поэтому вектора \mathbf{a}_0 и \mathbf{a}_2 коллинеарны — разность фаз колебаний $\Delta\varphi = 2\pi$. Мы имеем $(k/2z)\xi^2 = 2\pi$, откуда $\xi = \sqrt{2\lambda z}$. Вектор \mathbf{A}_2 изображает суммарный вклад полосы, лежащей между двумя прямыми $\xi = \xi_1 = \sqrt{\lambda z}$ и $\xi = \xi_2 = \sqrt{2\lambda z}$. В данном случае вместо кольцевых зон Френеля мы имеем зоны в виде полос (их называют *зонами Шустера* по имени английского физика А. Шустера (1851–1934)).

Продолжая дальше построение векторной диаграммы, мы придем к *спирали Корню*, изображенной на рис. 8.26. Спираль, очень быстро скручиваясь (благодаря квадратичной зависимости фазы колебаний $\varphi = (k/2z)\xi^2$ от координаты ξ), все ближе при-

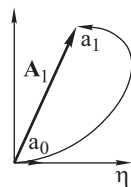


Рис. 8.24

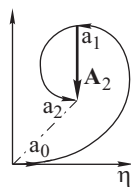


Рис. 8.25

ближается к точке S_1 , называемой *фокусом спирали*. Каждый последующий полувиток спирали изображает вклад в суммарное колебание последовательно расположенных зон Шустера, причем n -я зона Шустера — это полоса, внешний край которой отстоит от точки наблюдения на расстояние $z + n\lambda/2$ и на расстояние $\xi_n = \sqrt{n\lambda z}$ от оси η . Нижняя половина спирали Корню (от нуля до точки S_2) описывает вклад вторичных источников (элементарных полосок ξ_n), расположенных ниже оси η ($\xi_n < 0$). С ростом ширины щели ($b_1 \rightarrow \infty$, $b_2 \rightarrow -\infty$ на рис. 8.23) работают все более мелкие витки спирали Корню, которые все ближе прилегают к фокусам S_1 и S_2 . Вектор \mathbf{A}_0 , соединяющий фокусы спирали, определяет амплитуду колебаний в точке наблюдения при отсутствии препятствия.

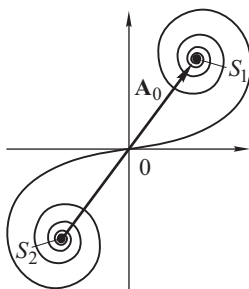


Рис. 8.26

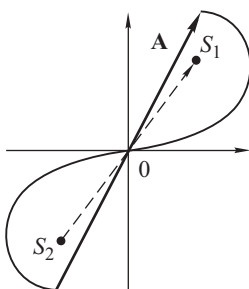


Рис. 8.27

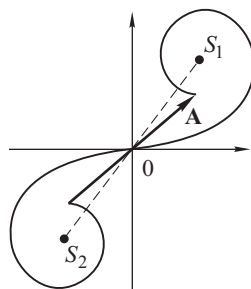


Рис. 8.28

Очевидно, амплитуда колебаний в точке P близка к максимальной, если открыты две зоны Шустера (т. е. «работает» один полувиток верхней половины спирали и один полувиток нижней половины, рис. 8.27). Амплитуда колебаний (длина вектора \mathbf{A}) при этом больше, чем амплитуда колебаний A_0 при отсутствии препятствия. Соответствующая ширина щели равна при этом $b = 2\sqrt{\lambda z}$. Если размер щели увеличить таким образом, что $b/2 = \sqrt{2\lambda z}$ (так, что на верхней и нижней полуспирали начинают «работать» полувитки вторых зон Шустера, рис. 8.28), то амплитуда колебаний в точке наблюдения уменьшается: длина вектора \mathbf{A} на рис. 8.28 меньше, чем A_0 .

Дифракция на краю экрана. С помощью спирали Корню можно качественно исследовать картину дифракции плоской волны на полубесконечном экране (непрозрачный экран с прямолинейным краем перегораживает нижнюю полуплоскость $\xi < 0$, рис. 8.29). Согласно законам геометрической оптики в плоскости наблюдения мы должны были бы иметь распределение интенсивности, показанное на рис. 8.29 штриховой линией: интенсивность равна нулю в области тени ($x < 0$) и равна $I_0 = \text{const}$ при $x > 0$.

Реальная картина $I(x)$, полученная с помощью спирали Корню, показана на рисунке. (Непрозрачный экран перекрывает вклад нижней полуспирали Корню.)

Упражнение. Найдите с помощью спирали, изображенной на рис. 8.26, какова интенсивность I_1 на границе света и тени, в точке 0, если при отсутствии препятствия она равна I_0 . Исследуйте с помощью векторной диаграммы рис. 8.26 характер осциллиций интенсивности в области $x > 0$, а также характер изменения интенсивности в области тени $x < 0$.

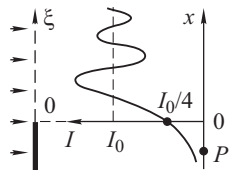


Рис. 8.29

8.9. Дифракция Фраунгофера ¹⁾

Удаляясь от препятствия все дальше, мы оказываемся в *дальней волновой зоне* (или в зоне фаунгоферовой дифракции) $p = \sqrt{\lambda z} / b \gg 1$.

Вновь обратимся к принципу Гюйгенса–Френеля и его математической формулировке — равенству (8.52). Используем приближения, о которых уже неоднократно говорили ранее: при условии, что размер препятствия (например, размер отверстия в непрозрачном экране) мал по сравнению с расстоянием R_0 до точки наблюдения, амплитудный множитель $1/R$, учитывающий уменьшение амплитуды в сферической волне по мере удаления от вторичного источника $d\sigma$, можно заменить постоянной величиной $1/R_0$ (рис. 8.30).

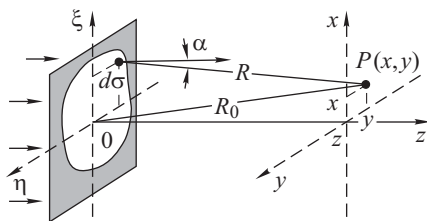


Рис. 8.30

Множитель наклона $\cos \alpha$ также считаем приблизительно одинаковым (и равным единице) для всех вторичных источников $d\sigma$, расположенных в области отверстия. Тогда из (8.52) находим

$$f(P) = \frac{1}{i\lambda R_0} \iint f_0(\xi, \eta) e^{ikR} d\xi d\eta. \quad (8.56)$$

Точное выражение для расстояния R от вторичного источника (малой площадки $d\sigma = d\xi d\eta$, расположенной в точке (ξ, η)) до точки наблюдения $P(x, y, z)$ равно

$$R = \sqrt{z^2 + (x - \xi)^2 + (y - \eta)^2}. \quad (8.57)$$

¹⁾ См. § 8.5.

Вычисляя величину R , входящую в фазовый множитель e^{ikR} , мы не можем довольствоваться грубой оценкой $R \approx R_0$: ошибка при вычислении фазы колебаний kR должна быть мала по сравнению с π и, следовательно, ошибка в вычислении R мала по сравнению с длиной волны λ . Выражение (8.57) запишем в виде

$$R = \sqrt{(z^2 + x^2 + y^2) - 2x\xi - 2y\eta + (\xi^2 + \eta^2)}$$

и далее, поскольку $z^2 + x^2 + y^2 = R_0$, то

$$R = R_0 \sqrt{1 - 2 \frac{x\xi + y\eta}{R_0^2} + \frac{\xi^2 + \eta^2}{R_0^2}}.$$

Считая поправку к единице под знаком радикала малой, получим

$$R \approx R_0 - \frac{x\xi + y\eta}{R_0} + \frac{\xi^2 + \eta^2}{2R_0}. \quad (8.58)$$

Разумеется, мы полагаем при этом, что отброшенные члены разложения радикала в степенной ряд малы по сравнению с длиной волны λ . Обратим внимание, что, согласно граничным условиям Кирхгофа, граничное поле $f_0(\xi, \eta)$ отлично от нуля лишь на открытой части волнового фронта, поэтому при оценке выражения (8.58) интерес представляют лишь значения (ξ, η) , лежащие в области отверстия. Очевидно, что последнее слагаемое в (8.58) ограничено величиной $b^2/(2R_0)$:

$$\frac{\xi^2 + \eta^2}{2R_0} \lesssim \frac{b^2}{2R_0},$$

где b — максимальный линейный размер отверстия. Потребуем теперь, чтобы точка наблюдения находилась столь далеко от отверстия и (или) размер отверстия был столь мал, чтобы выполнялось неравенство $b^2/(2R_0) \ll \lambda$ или, что то же,

$$p = \frac{\sqrt{\lambda R_0}}{b} \gg 1. \quad (8.59)$$

Последнее неравенство — условие применимости полученных соотношений — называется *условием дифракции Фраунгофера*.

Обсуждая ранее условия, при которых справедливы выводы геометрической оптики (в частности, закон прямолинейного распространения света), мы пришли к неравенству (8.32): $\lambda z/b^2 \ll 1$, которое, очевидно, противоположно неравенству (8.59). Если геометрическое приближение справедливо при малых расстояниях z и (или) больших размерах препятствий b , то фраунгоферовы дифракционные картины возникают, напротив, на больших расстояниях z и (или) при малых размерах препятствий,

расположенных на пути волны. Условие $p \gg 1$ эквивалентно следующему: размер отверстия b много меньше размера первой зоны Френеля. Промежуточная область $p \sim 1$ называется обычно *областью дифракции Френеля*. В этой ситуации размер отверстия b и радиус первой зоны Френеля $\sqrt{\lambda z}$ сравнимы между собой.

Итак, при условии (8.59) получим следующее приближенное выражение для R :

$$R \approx R_0 - \frac{x\xi + y\eta}{R_0} \quad (8.60)$$

и вместо (8.56) имеем

$$f(P) = \frac{e^{ikR_0}}{i\lambda R_0} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_0(\xi, \eta) e^{-i((kx/R_0)\xi + (ky/R_0)\eta)} d\xi d\eta. \quad (8.61)$$

Введем обозначения $kx/R_0 = u$, $ky/R_0 = v$; тогда формула (8.61) приобретает вид

$$f(P) \sim \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_0(\xi, \eta) e^{-i(u\xi + v\eta)} d\xi d\eta. \quad (8.62)$$

Последнее соотношение означает, что для нахождения светового поля в точке наблюдения достаточно найти преобразование Фурье граничного поля $f_0(\xi, \eta)$: $f(x, y, z) \sim C_0(u, v)$, взяв в качестве аргументов преобразования $u = kx/R_0$, $v = ky/R_0$. В частном случае, если граничное поле является функцией одной переменной, $f_0(\xi)$, то (8.62) приобретает вид

$$f(P) \sim \int_{-\infty}^{\infty} f_0(\xi) e^{-iu\xi} d\xi. \quad (8.63)$$

(Сравните (8.63) с полученной ранее формулой (8.43).)

Еще раз отметим, что в (8.62) (или одномерном аналоге (8.63)) положение точки наблюдения полностью определяется параметрами u , v (или параметром u — в случае функции одной переменной).

Обратимся для простоты к одномерной задаче (рис. 8.31). Рассмотрим колебание в точке P как сумму колебаний от всех вторичных источников $d\sigma$, заполняющих отверстие. Принимая фазу колебания от вторичного источника, расположенного в начале координат, нулевой, найдем фазу колебания от источника $d\sigma$, находящегося на

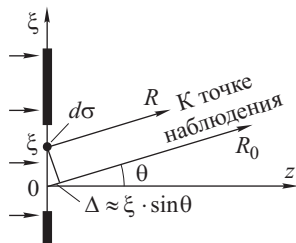


Рис. 8.31

расстоянии ξ от начала координат. При достаточном удалении точки P (неравенство (8.59)) разность расстояний $\Delta = R - R_0$ можно оценить так: $\Delta \approx -\xi \sin \theta$ (угол θ определяет направление на точку наблюдения: $\sin \theta = x/R_0$). Разность фаз колебаний $k\Delta \approx -k\xi \sin \theta = (-kx/R_0)\xi$.

Если падающая на щель волна имеет в точке ξ комплексную амплитуду $f_0(\xi)$, то колебание в точке наблюдения от этого вторичного источника есть $df \sim f_0(\xi) d\xi e^{-i(kx/R_0)\xi}$, а суммарное колебание

$$f(P) \sim \int f_0(\xi) e^{-i(kx/R_0)\xi} d\xi,$$

что совпадает с выражением (8.63), где $u = kx/R_0 = k \sin \theta$.

Еще раз подчеркнем, что при выполнении неравенства (8.59) поле в точке наблюдения $f(x, z)$ определяется направлением на эту точку — углом θ .

Если точка наблюдения смещается вдоль фиксированного направления, приближаясь или удаляясь, так что угол θ остается неизменным: $\sin \theta = x/R_0 = \text{const}$ (и при этом остается справедливым неравенство (8.59)), то значение интеграла (8.63) остается неизменным:

$$f(\theta) \sim \int_{-\infty}^{\infty} f_0(\xi) e^{-ik\xi \sin \theta} d\xi. \quad (8.64)$$

В частности, картину дифракции можно характеризовать распределением интенсивности света (потока энергии) по углам $I = I(\theta)$

$$I(\theta) \sim \left| \int_{-\infty}^{\infty} f_0(\xi) e^{-ik\xi \sin \theta} d\xi \right|^2. \quad (8.65)$$

Это распределение $I(\theta)$, изображенное в полярных координатах, называют *диаграммой направленности*.

Продолжая изучение дифракции Фраунгофера, рассмотрим три конкретных примера.

Дифракция Фраунгофера на щели (ширина щели b , щель освещается слева плоской нормально падающей волной). Граничное поле $f_0(\xi)$, возникающие в плоскости $z = 0$, примыкающей к непрозрачному экрану со щелью справа, имеет вид изображенный на рис. 8.9а. Согласно (8.64) для нахождения картины фраунгоферовой дифракции, необходимо найти преобразование Фурье этой функции. Задача полностью аналогична нахождению спектра прямоугольного импульса (см. § 1.4). Здесь роль длительности импульса τ играет ширина щели b , а роль частоты ω — величина $u = k \sin \theta$ (которую мы называли ранее

пространственной частотой). Используя эту аналогию, находим

$$f(\theta) \sim \frac{\sin\left(\frac{kb}{2} \sin \theta\right)}{\frac{kb}{2} \sin \theta}. \quad (8.66)$$

Распределение интенсивности $I(\theta) = |f(\theta)|^2$ показано на рис. 8.32. Ближайшие к направлению $\theta = 0$ направления, в которых $I(\theta)$ обращается в нуль, определяются условием $(kb/2) \sin \theta = \pm\pi$, откуда находим

$$\sin \theta = \pm \frac{\lambda}{b}. \quad (8.67)$$

Как показывает анализ формулы (8.66), в интервале углов

$$|\sin \theta| \leq \frac{\lambda}{b}$$

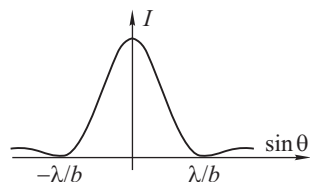


Рис. 8.32

сосредоточена подавляющая величина потока энергии (более 80%). Этот интервал углов (от $-\lambda/b$ до λ/b) называют *главным максимумом* дифракционной картины (легко найти с помощью (8.66), что максимальная интенсивность в ближайшем боковом максимуме, в направлении $\sin \theta = (3/2)\lambda/b$, составляет примерно 0,05 от интенсивности в направлении $\sin \theta = 0$). Мы видим, что сужение щели (уменьшение b) приводит к уширению угла θ : свет расходится от щели во все более широком интервале углов. Можно сказать и по-другому: с уменьшением ширины щели уширяется пространственный спектр — спектр плоских волн, бегущих от щели.

Дифракция Фраунгофера на круглом отверстии (рис. 8.33). При дифракции плоской волны на круглом отверстии в непрозрачном экране, на удаленной плоскости наблюдения (т. е. при условии $\sqrt{\lambda z}/d \gg 1$, d — диаметр отверстия)

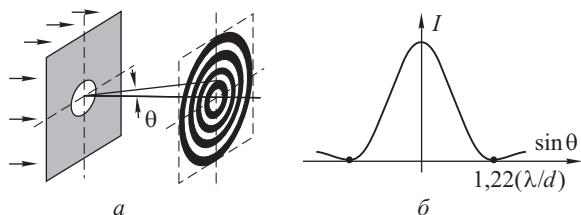


Рис. 8.33

наблюдается картина дифракции, показанная на рис. 8.33 а: центральное яркое дифракционное пятно (которое называется *пятном Эйри*) окружено чередующимися светлыми и темными кольцами. Соответствующий график $I(\theta)$ показан на рис. 8.33 б.

Угловая полуширина пятна Эйри (главного максимума в картине $I(\theta)$) определяется условием

$$\sin \theta = 1,22 \frac{\lambda}{d}. \quad (8.68)$$

В телесном угле θ сосредоточена подавляющая доля потока энергии ($\sim 80\%$) дифрагированной волны.

Дифракция Фраунгофера на решетке. Рядом со щелью, дифракцию на которой мы рассмотрели выше (эта картина дифракции описывается функцией $f(\theta)$, см. формулу (8.66)), расположим параллельно еще одну такую же щель, центр которой находится в точке $\xi = d$, рис. 8.34. Расстояние от второй щели до точки наблюдения (от любого элемента этой щели) на величину $\Delta = d \sin \theta$ меньше расстояния между соответствующим элементом первой щели и точкой наблюдения (рис. 8.34). Соответственно, фаза колебаний отличается на величину

$$\alpha = -k\Delta = -kd \sin \theta, \quad (8.69)$$

поэтому колебательный процесс, созданный второй щелью в точке наблюдения, описывается функцией $f(\theta)e^{i\alpha}$. Волны, посылаемые в точку наблюдения двумя щелями, интерferируют, суммарный ко-

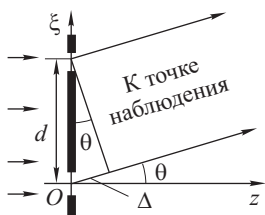


Рис. 8.34

лебательный процесс в точке наблюдения есть $f(\theta) + f(\theta)e^{i\alpha}$. Если расположить рядом еще одну щель, сдвинутую относительно первой на расстояние $2d$ (и на расстояние d от второй), то получим суммарное колебание $f(\theta) + f(\theta)e^{i\alpha} + f(\theta)e^{i2\alpha}$. Наконец, если мы имеем решетку, состоящую из N параллельно расположенных щелей (период решетки d), то суммарное колебание имеет вид

$$g(\theta) = f(\theta) \sum_{n=0}^{N-1} e^{in\alpha}. \quad (8.70)$$

Сумму, входящую в (8.70), мы уже находили ранее (1.16) (сумма колебаний, фазы которых составляют арифметическую прогрессию) с помощью векторной диаграммы, показанной на рис. 1.7. В нашем случае $\alpha = kd \sin \theta$ и мы получаем

$$g(\theta) = f(\theta) \frac{\sin (Nkd \sin \theta / 2)}{\sin (kd \sin \theta / 2)}. \quad (8.71)$$

Распределение интенсивности $I(\theta)$ по углам описывается формулой

$$I(\theta) = |f(\theta)|^2 \left| \frac{\sin (Nkd \sin \theta / 2)}{\sin (kd \sin \theta / 2)} \right|^2. \quad (8.72)$$

Первый сомножитель $|f(\theta)|^2$ описывает картину дифракции на щели (формула (8.66)). Второй сомножитель связан с интерференцией волн, приходящих от разных щелей к точке наблюдения.

Наиболее интересным в картине дифракции на решетке является наличие узких дифракционных максимумов, в которые идет подавляющая доля общего потока энергии.

Направления на эти максимумы θ_m определяются условием

$$\Delta_m = d \sin \theta_m = m\lambda \quad (m = 1, 2, \dots), \quad (8.73)$$

— волны от всех щелей решетки приходят в точку наблюдения с разностью хода Δ_m , равной целому числу длин волн и, следовательно, создают синфазные колебания. Итак, разность фаз $\alpha_m = k\Delta_m = 2\pi m$. Амплитуда суммарного колебания A оказывается при этом в N раз больше амплитуды колебаний, созданных одной щелью: $A(\theta) = Nf(\theta)$, а интенсивность $I(\theta_m) = N^2|f(\theta)|^2$ в N^2 раз превышает интенсивность волны от одной щели, так как при $\alpha = 2\pi m$ все векторы, составляющие цепочку векторов на рис. 1.7, оказываются коллинеарными, поэтому длина суммарного вектора в N раз больше длины слагаемых векторов). Читатель может самостоятельно убедиться, что второй сомножитель в (8.72) равен N при условии (8.73):

$$\lim_{\alpha \rightarrow 2\pi m} \frac{\sin(N\alpha/2)}{\sin(\alpha/2)} = N. \quad (8.74)$$

График функции (8.72) показан на рис. 8.35.

Штриховой линией показана «огibaющая» — зависимость от θ первого сомножителя $|f(\theta)|^2$, описывающего картину фраунгоферовой дифракции на щели ширины b .

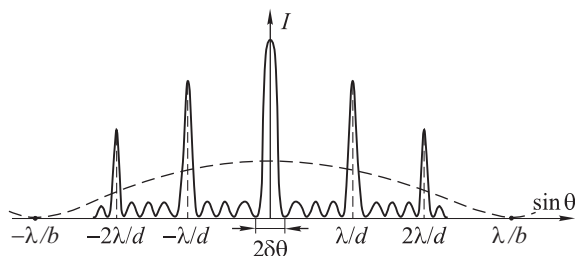


Рис. 8.35

Для картины, изображенной на рис. 8.35, характерно, что уже при небольшом отклонении от направления θ_m (т. е. от направления, определяемого условием (8.73)) интенсивность резко уменьшается, обращаясь в нуль при $\theta = \theta_m \pm \delta\theta$. Определить величину этого отклонения можно из следующих соображений.

В направлении θ_m векторная диаграмма имеет вид, показанный на рис. 8.36 а — цепочка, состоящая из N коллинеарных векторов: $\alpha_m = kd \sin \theta_m = 2\pi m$. Спрашивается, при каком отклонении $\delta\theta$ от направления θ_m векторная диаграмма будет иметь вид, показанный на рис. 8.36 б: замкнутая цепочка векторов, в которой конец последнего вектора совпадает с началом первого вектора и, следовательно, длина суммарного вектора будет равна нулю. Чтобы векторная диаграмма, показанная на рис. 8.36 а, преобразовалась в замкнутый многоугольник векторов рис. 8.36 б, необходимо, чтобы разность фаз колебаний от двух соседних щелей решетки в точке наблюдения изменилась на величину $\delta\alpha = 2\pi/N$. Тогда (при $N \gg 1$) последний N -й вектор коллинеарен с первым: $N\delta\alpha = 2\pi$, т. е.

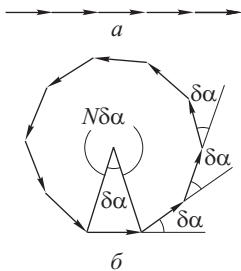


Рис. 8.36

нилась на величину $\delta\alpha = 2\pi/N$. Тогда (при $N \gg 1$) последний N -й вектор коллинеарен с первым: $N\delta\alpha = 2\pi$, т. е.

$$\delta(kd \sin \theta) = \frac{2\pi}{N} \text{ или } \delta(\sin \theta) = \frac{\lambda}{Nd}. \quad (8.75)$$

Для сравнительно небольших углов θ можно приближенно написать

$$\delta\theta = \frac{\lambda}{Nd}. \quad (8.76)$$

Это и есть оценка *полуширины главных дифракционных максимумов*. Целое число m в (8.73) называется *порядком главного максимума* (или порядком дифракции). Например, при $m = 1$ имеем первый порядок дифракции; значению $m = -1$ отвечает минус первый порядок дифракции. Максимальное значение m , как ясно из (8.73), ограничено величиной

$$m_{\max} \leq \frac{d}{\lambda}. \quad (8.77)$$

Реально же заметными являются лишь те дифракционные максимумы, которые лежат в пределах углов

$$|\sin \theta| \leq \frac{\lambda}{b}, \quad (8.78)$$

поскольку, как мы выяснили ранее, только в пределах этих углов в основном распространяется поток энергии от каждой из щелей решетки.

Поэтому (при $b > \lambda$) максимальный порядок m можно оценить из условия (см. рис. 8.35)

$$\frac{m\lambda}{d} \leq \frac{\lambda}{b} \text{ или } m_{\max} \leq \frac{d}{b}. \quad (8.79)$$

При этом общее число главных дифракционных максимумов, в которые попадает подавляющая часть потока энергии, равно приблизительно $1 + 2d/b$.

8.10. Разрешающая способность спектральных приборов

Ранее мы говорили о том, что излучение любого источника света можно представить как сумму волн различных частот (спектральное разложение).

Распределение энергии по спектру излучения (зависимость интенсивности от частоты $I(\omega)$) является чрезвычайно важной характеристикой, которая исследуется с помощью спектральных приборов. Мы рассмотрим несколько возможных способов.

Дифракционная решетка. Два обстоятельства позволяют использовать дифракционную решетку для спектральных измерений. Во-первых, как следует из (8.73), положение главных дифракционных максимумов зависит от длины волны, и во-вторых, ширина этих максимумов, как видно из (8.76), чрезвычайно мала, а поток энергии (интенсивность) велик при большом числе щелей решетки.

Схема спектрального прибора, основанного на использовании дифракционной решетки, показана на рис. 8.37 а.

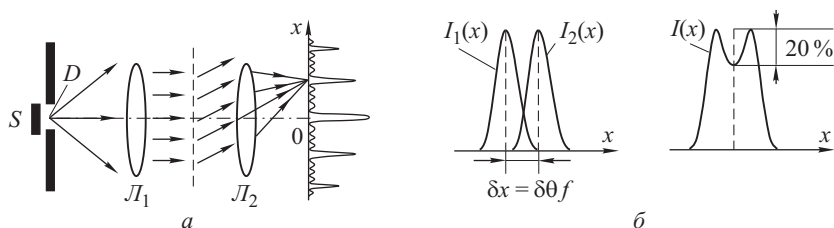


Рис. 8.37

Излучение исследуемого источника S проходит через малую диафрагму (отверстие в непрозрачном экране) D , расположенную в фокальной плоскости объектива L_1 . Сколлимированное объективом излучение (параллельный пучок света) падает на дифракционную решетку. При дифракции на решетке возникает ряд пучков, распространяющихся в направлениях θ_m , определяемых равенством (8.73). Угловая расходимость каждого из них $\delta\theta$ дается формулой (8.76). Следовательно, в фокальной плоскости объектива L_2 возникает картина дифракционных максимумов, положение которых (расстояние от оптической оси) есть

$$x_m = f \operatorname{tg} \theta_m \approx f \sin \theta_m = \frac{m \lambda f}{d},$$

а полуширина максимумов $\delta x = f \delta \theta = \lambda f / Nd$; f — фокусное расстояние объектива L_2 .

Пусть излучение содержит две близкие спектральные линии λ и $\lambda + \delta \lambda$ одинаковой интенсивности. Положение m -го дифракционного максимума, отвечающего длине волны λ , определяется условием

$$x_1 = f \sin \theta_1 = \frac{m \lambda f}{d}.$$

Соответственно, положение m -го максимума для длины волны $\lambda + \delta \lambda$ есть

$$x_2 = f \sin \theta_2 = m \frac{\lambda + \delta \lambda}{d} f.$$

Какова минимальная величина $\delta \lambda$, при которой мы можем по наблюдаемой картине дифракции в фокальной плоскости линзы L_2 , зафиксировать наличие двух спектральных линий?

Согласно *критерию Рэлея* предельно разрешимыми считаются спектральные линии, для которых смещение дифракционных максимумов $\Delta x = x_2 - x_1 = m f \delta \lambda / d$ в точности равно их полуширине $\delta x = \lambda f / (Nd)$ (в данном случае полуширина линии определяется по первому нулю интенсивности).

Эта ситуация показана на рис. 8.37 б. Слева изображены кривые интенсивности $I_1(x)$ и $I_2(x)$, отвечающие m -м максимумам для длин волн λ и $\lambda + \delta \lambda$ соответственно. Справа наблюдаемая в предельном случае суммарная картина интенсивности $I(x) = I_1(x) + I_2(x)$. Как показывает расчет, основанный на формуле (8.72), в наблюдаемой картине $I(x)$ имеется примерно 20%-ный провал, по наличию которого легко определить, что излучение источника содержит две спектральных линии. Разумеется, критерий Рэлея является условным: например, если расстояние Δx между центрами картин $I_1(x)$ и $I_2(x)$ будет немного меньше величины δx , то провал в суммарной картине будет несколько меньшим (например 15 % или 10 %), что визуально также можно определить. При достаточно малом смещении Δx или существенно разной интенсивности линий провал в суммарной картине вообще может отсутствовать, однако и в этом случае наблюдаемая картина отличается от картины дифракции, создаваемой строго монохроматическим источником. Тем не менее, критерий Рэлея широко применяется для оценки возможностей спектральных приборов. Итак, используя критерий Рэлея $\Delta x = \delta x$, находим

$$\frac{m \delta \lambda}{d} = \frac{\lambda}{Nd},$$

откуда получаем

$$\frac{\lambda}{\delta \lambda} = m N. \quad (8.80)$$

Отношение длины волны λ к предельно допустимому (согласно критерию Рэлея) интервалу длин волн $\delta\lambda$ называется *разрешающей способностью спектрального прибора*.

Высокая разрешающая способность дифракционных решеток достигается за счет большого числа N (до 10^5 и более). При этом общий рабочий размер решеток $D = Nd$ не может быть слишком большим (обычно $D \lesssim 10$ см), так как в противном случае возникают трудно выполнимые требования как к качеству объективов \mathcal{L}_1 и \mathcal{L}_2 , так и к качеству самих решеток (равномерности нанесения штрихов). Расстояние между штрихами (период решеток d) при этом сильно ограничено: у хороших решеток величина d составляет всего несколько длин волн (в оптическом диапазоне). Соответственно максимальная величина порядка дифракции m_{\max} , согласно (8.77), ограничена несколькими единицами. Из (8.80) и (8.77) следует, что максимальная величина разрешающей способности дифракционной решетки есть

$$\left(\frac{\lambda}{\delta\lambda}\right)_{\max} = m_{\max} N \approx \frac{D}{\lambda}. \quad (8.81)$$

Из (8.81) получаем оценку минимально разрешимого интервала длин волн

$$(\delta\lambda)_{\min} \approx \frac{\lambda^2}{D}. \quad (8.82)$$

Имеется еще одна важная характеристика: максимальный спектральный интервал излучения, который может быть проанализирован с помощью спектрального прибора. Итак, пусть спектральный состав излучения содержит ряд компонент, лежащих в интервале от λ до $\lambda + \Delta\lambda$ (рис. 8.38). При этом разрешающая способность решетки достаточно высока, так что даже ближайшие линии в спектре излучения (с расстоянием между ними $\delta\lambda$) разрешаются нашей решеткой. Однако есть еще одна проблема.

Каждая спектральная компонента создается в фокальной плоскости объектива \mathcal{L}_2 (рис. 8.37 а) свою картину дифракционных максимумов. Пусть положение $(m+1)$ -го максимума спектральной компоненты λ (находящейся на левом краю спектрального диапазона излучения источника) совпадает с положением m -го дифракционного максимума спектральной компоненты $\lambda + \Delta\lambda$ (находящейся на правом краю спектрального диапазона излучения), т. е. выполняется условие:

$$d \sin \theta = (m+1)\lambda = m(\lambda + \Delta\lambda). \quad (8.83)$$

Ясно, что при этом наблюдаемая картина дифракции в фокальной плоскости объектива \mathcal{L}_2 не дает возможности определить, наблюдается ли m -й максимум для длины волны $\lambda + \Delta\lambda$

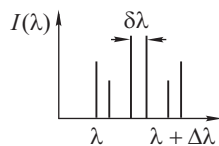


Рис. 8.38

или мы видим $(m + 1)$ -й максимум для длины волны λ . Равенство (8.83) является предельным условием, при котором спектральные максимумы разных порядков начинают перепутываться; спектральный диапазон, превышающий предельное значение $\Delta\lambda$, определяемое условием (8.83), не может быть проанализирован. Этот максимально допустимый диапазон $\Delta\lambda$ называется *областью дисперсии*. Из (8.83) находим

$$\Delta\lambda = \frac{\lambda}{m}. \quad (8.84)$$

Чем больше порядок дифракции m , в котором анализируется излучение, тем меньше допустимая область дисперсии. В частности, в первом порядке дифракции ($m = 1$) может быть проанализирован весь видимый диапазон излучения, однако разрешающая способность при этом (8.80) минимальна.

Интерферометр Фабри–Перо. Этот прибор, изобретенный французскими физиками Ш. Фабри (1867–1945) и А. Перо (1863–1925) представляет собой плоскопараллельную прозрачную пластинку, на поверхность которой нанесены высокоотражающие покрытия, либо интерферометр состоит из двух пластин с параллельными отражающими покрытиями, разделенных воздушным промежутком. Многолучевая интерференция в данном случае возникает за счет многократного переотражения падающей волны от высокоотражающих покрытий (зеркал) интерферометра. Пусть плоская монохроматическая волна падает на интерферометр под углом θ (угол между волновым вектором и нормалью к отражающим поверхностям — осью z на рис. 8.39). Введем τ и ρ — амплитудные коэффициенты пропускания и отражения зеркал.

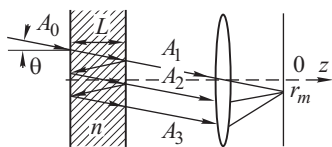


Рис. 8.39

Коэффициенты τ и ρ связывают между собой амплитуду падающей на зеркало волны с амплитудами прошедшей и отраженной волн. Согласно определению

$$A_{\text{прош}} = A_{\text{пад}} \tau, \quad A_{\text{отр}} = A_{\text{пад}} \rho.$$

Соответственно интенсивности волн связаны равенствами: $I_{\text{прош}} = I_{\text{пад}} \tau^2$, $I_{\text{отр}} = I_{\text{пад}} \rho^2$. Будем полагать, что поглощение отсутствует. Тогда, очевидно $I_{\text{пад}} = I_{\text{прош}} + I_{\text{отр}}$, откуда следует $\tau^2 + \rho^2 = 1$. Если ввести энергетические коэффициенты отражения и пропускания R и T , то $R + T = 1$.

Пусть падающая на интерферометр волна имеет амплитуду A_0 (рис. 8.40). Пройдя внутри интерферометра, волна ослабится, ее амплитуда станет равной $A_0\tau$. Пройдя затем через второе зеркало, волна выйдет из интерферометра вправо, имея амплитуду $A_1 = A_0\tau^2$. Назовем эту волну *первой волной*.

Однако часть волны (с амплитудой $A_0\tau$) отразится от зеркала 2, а затем и от зеркала 1. После первого отражения ее амплитуда станет равной $A_0\tau\rho$, а после второго — $A_0\tau\rho^2$. После выхода из интерферометра вправо (после прохождения сквозь зеркальное покрытие 2), ее амплитуда станет равной $A_2 = A_0\tau\rho^2 \cdot \tau = A_0\tau^2\rho^2$. Назовем эту волну *второй волной*. Легко сообразить, что *третья волна* будет иметь амплитуду $A_3 = A_0\tau^2\rho^4$. Итак, на выходе из интерферометра мы имеем суперпозицию плоских волн, все они распространяются под углом θ к оси z , а амплитуды этих волн отличаются множителем ρ^2 : $A_{n+1} = A_n\rho^2$.

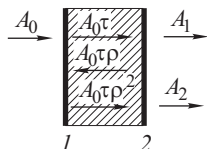


Рис. 8.40

Каждая последующая волна в этой суперпозиции проходит в интерферометре дополнительный путь $2L$ (L — расстояние между зеркалами), приобретая набег фазы α , равный $k_z \cdot 2L$:

$$\alpha = 2L k \cos \theta \quad (8.85)$$

(для интерферометра, состоящего из двух зеркал, разделенных воздушным промежутком).

Таким образом, суммарная волна на выходе из интерферометра имеет амплитуду

$$A = A_0\tau^2 + A_0\tau^2\rho^2 e^{i\alpha} + A_0\tau^2\rho^4 e^{i2\alpha} + \dots = \frac{A_0\tau^2}{1 - \rho^2 e^{i\alpha}} \quad (8.86)$$

(сумма бесконечной геометрической прогрессии со знаменателем $\rho^2 e^{i\alpha}$). Амплитуда суммарной волны, прошедшей сквозь интерферометр, максимальна, если происходит синфазное сложение волн, т. е.

$$\alpha_m = 2kL \cos \theta_m = 2\pi m \quad (m — \text{целое число}). \quad (8.87)$$

При этом векторная диаграмма, изображающая сумму (8.86), представляет собой цепочку коллинеарных векторов, длины которых уменьшаются в геометрической прогрессии, а амплитуда прошедшей волны равна амплитуде падающей волны:

$$A = A_0 \frac{T}{1 - R} = A_0, \quad (8.88)$$

т. е. волна полностью проходит сквозь интерферометр! Поучительно рассчитать, какова при этом амплитуда отраженной волны (разумеется, мы должны получить в результате расчета $A_{\text{отр}} = 0$). Мы предоставляем проделать этот расчет читателю в качестве упражнения.

Синфазное сложение волн, возникающих при многократных переотражениях от зеркал (при условии (8.87)) приводит к тому, что амплитуда колебаний поля внутри резонатора может суще-

ственно превосходить амплитуду падающей волны (так же, как амплитуда вынужденных колебаний в высокодобротном колебательном контуре существенно превосходит амплитуду внешней ЭДС, вызывающей процесс колебаний). Действительно, поскольку амплитуда прошедшей волны равна $A_0 T / (1 - R)$, то амплитуда волны, падающей на зеркало 2 внутри резонатора (бегущей слева направо), есть

$$A_{\rightarrow} = A_0 \frac{\tau}{1 - R} = \frac{A_0}{\sqrt{1 - R}},$$

а амплитуда волны, бегущей справа налево внутри резонатора, как следует из последнего равенства, равна

$$A_{\leftarrow} = A_0 \frac{\sqrt{R}}{\sqrt{1 - R}}. \quad (8.89)$$

Если коэффициент отражения R близок к единице, то амплитуда суммарной волны (в пучностях) близка к значению $2A_{\rightarrow} = 2A_0 / \sqrt{1 - R}$. При значении $R = 0,96$ (что легко достигается при использовании многослойных диэлектрических зеркал) получаем $A \approx 10A_0$, а интенсивность волны внутри резонатора почти в 100 раз больше интенсивности падающей волны.

Вернемся к вопросу о разрешающей способности интерферометра Фабри–Перо. Условие резонанса (8.87) выделяет длины волн λ_m , которые проходят сквозь интерферометр, имея максимально возможную амплитуду. Это условие можно переписать в виде

$$\Delta_m = 2L \cos \theta = m\lambda_m. \quad (8.90)$$

При изменении λ (при фиксированном угле падения θ) нарушается синфазность волн, возникающих при переотражениях от зеркал. Разность фаз становится отличной от $2\pi m$: $\alpha = 2\pi m + \varepsilon$, а разность хода Δ — от целого числа длин волн $m\lambda$:

$$\Delta = m\lambda + \lambda \frac{\varepsilon}{2\pi}.$$

Нарушение синфазности интерферирующих волн приводит к уменьшению амплитуды суммарной волны, прошедшей сквозь интерферометр.

Так же, как и у дифракционной решетки, спектральная избирательность интерферометра Фабри–Перо обусловлена многолучевой интерференцией. Изучая дифракционную решетку, мы выяснили, что ее разрешающая способность определяется формулой

$$\frac{\lambda}{\delta\lambda} = mN, \quad (8.91)$$

где m — порядок интерференции — отношение разности хода между волнами, прошедшими через соседние щели решетки, к длине волны λ (для решетки эта величина равна

$m = d \sin \theta / \lambda$), N — число интерферирующих волн (в решетке оно просто равно числу щелей). Формулу (8.91) можно использовать и для оценки разрешающей способности интерферометра Фабри–Перо. В этом случае порядок интерференции, как следует из (8.90), равен

$$m = \frac{\Delta_m}{\lambda} = \frac{2L \cos \theta}{\lambda} \approx \frac{2L}{\lambda}$$

(условие малости угла θ всегда используется в спектральных измерениях).

Число интерферирующих волн в принципе бесконечно, а амплитуда каждой последующей волны A_{n+1} отличается от предыдущей множителем R : $A_n = A_{n-1} R$. Однако ясно, что реально результат интерференции не зависит от волн, амплитуда которых слишком мала. Мы можем говорить о некотором эффективном числе интерферирующих волн $N_{\text{эф}}$, вклад которых существен. Качественную оценку $N_{\text{эф}}$ можно сделать следующим образом. Представим себе, что синфазно складываются волны одинаковой амплитуды, равной амплитуде первой прошедшей через интерферометр волны $A_1 = A_0 \tau^2 = A_0 T$. Сколько таких волн дадут в результате интерференции волну, амплитуда которой равна суммарной амплитуде прошедшей через интерферометр волны? Положим $A_1 N_{\text{эф}} = A_0$, откуда получаем $N_{\text{эф}} \approx 1/T = 1/(1 - R)$; для разрешающей способности получаем оценку

$$\frac{\lambda}{\delta \lambda} \approx \frac{2L}{\lambda(1 - R)}. \quad (8.92)$$

Одной из важных характеристик любого резонансного устройства (маятник, колебательный контур) является *добротность* Q — величина, характеризующая скорость потерь энергии при колебаниях:

$$Q = 2\pi \frac{W}{\delta W}, \quad (8.93)$$

где W — энергия колебаний, δW — потери энергии за время, равное одному периоду колебаний. Напомним также, что добротность определяет относительную ширину резонансной кривой колебательной системы, т. е. ее спектральную избирательность: при отклонении частоты возбуждающей силы ω от резонансной частоты ω_m амплитуда вынужденных колебаний заметно уменьшается (например, в $\sqrt{2}$ раз):

$$Q = \frac{\omega_m}{\delta \omega} = \frac{\lambda_m}{\Delta \lambda}. \quad (8.94)$$

Интерферометр Фабри–Перо также является колебательной системой, для которой резонансными являются определенные частоты (длины волн) колебаний. К этой системе полностью применимо понятие добротности, определив которую, мы определим и

спектральную избирательность, т. е. разрешающую способность системы.

Итак, пусть I — интенсивность волны, бегущей в резонаторе слева направо, к зеркалу 2. После отражения от зеркала интенсивность уменьшается на величину $\Delta I_1 = I(1 - R)$. При значениях R , близких к единице, примерно таково же уменьшение интенсивности при втором отражении (от зеркала 1). Уменьшение интенсивности при полном обходе волны через резонатор (туда-обратно) равно, следовательно, $\Delta I = 2I(1 - R)$. Соответственно, связь между полной энергией волны в резонаторе W и потерями энергии при полном обходе, т. е. за время $\tau = 2L/c$, есть

$$\Delta W = 2W(1 - R).$$

За время, равное одному периоду колебания $T_0 = \lambda/c$, потери энергии составляют

$$\delta W = \frac{T_0}{\tau} \Delta W = \frac{\lambda}{L} W(1 - R),$$

откуда

$$\frac{W}{\delta W} = \frac{L}{\lambda(1 - R)} \quad \text{и} \quad Q \approx \frac{2\pi L}{\lambda(1 - R)}. \quad (8.95)$$

Формулы (8.92) и (8.95) дают качественные оценки разрешающей способности интерферометра Фабри–Перо и справедливы по порядку величины. Точный расчет основан на формуле (8.86), из которой можно найти связь между интенсивностью прошедшей волны $I = |A|^2$ и интенсивностью падающей волны $I = A_0^2$ и получить

$$I(\alpha) = I_0 \frac{(1 - R)^2}{1 + R^2 - 2R \cos \alpha}. \quad (8.96)$$

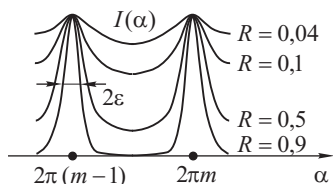


Рис. 8.41

Кривые $I(\alpha)$ (зависимость интенсивности прошедшей волны от разности фаз α) при различных значениях R показаны на рис. 8.41. На графике показана величина ε — отклонение разности фаз двух соседних интерферирующих волн от резонансного значения $2\pi m$, при котором интенсивность прошедшей волны уменьшается вдвое (т. е. $I = I_0/2$).

Эта величина может быть найдена с помощью (8.96) при подстановке $\cos \alpha = \cos(2\pi m \pm \varepsilon) = \cos \varepsilon$. При R , близких к единице, можно получить

$$\varepsilon = \frac{2(1 - R)}{\sqrt{R}}. \quad (8.97)$$

С другой стороны, изменение разности фаз $\varepsilon = \alpha - 2\pi m$ при отклонении длины волны $\lambda_1 = \lambda_m + \delta\lambda$ от резонансного

значения λ_m можно найти из равенств

$$\alpha_m = 2 \frac{2\pi}{\lambda_m} L \cos \theta = 2\pi m; \quad \alpha = 2 \frac{2\pi L \cos \theta}{\lambda_m + \delta\lambda}.$$

При малых $\delta\lambda$ находим

$$\begin{aligned} \alpha &= 2 \frac{2\pi}{\lambda_m} \left(\frac{1}{1 + \delta\lambda/\lambda_m} \right) L \cos \theta \approx 2 \frac{2\pi}{\lambda_m} L \cos \theta \left(1 - \frac{\delta\lambda}{\lambda_m} \right) = \\ &= 2\pi m \left(1 - \frac{\delta\lambda}{\lambda_m} \right), \end{aligned}$$

откуда

$$\varepsilon = \alpha - \alpha_m = 2\pi m \frac{\delta\lambda}{\lambda}. \quad (8.98)$$

Из (8.97) и (8.98) получаем

$$\frac{\lambda}{\delta\lambda} = m \frac{\pi\sqrt{R}}{(1-R)} = \frac{2\pi L\sqrt{R}}{\lambda(1-R)}. \quad (8.99)$$

Это и есть правильное выражение для разрешающей способности, справедливое при значениях R , близких к единице.

Расстояние между зеркалами (отражающими поверхностями) в интерферометре Фабри–Перо составляет обычно величину порядка нескольких сантиметров. Резонаторы лазеров могут иметь длину порядка метра, поэтому порядки интерференции $m = 2L/\lambda$ огромны: в оптическом диапазоне $m \approx 10^5$ и более. Соответственно, чрезвычайно малой оказывается рабочая область дисперсии $\Delta\lambda = \lambda/m \approx \lambda^2/2L$. В этом существенная особенность интерферометра Фабри–Перо как спектрального прибора.

8.11. Оптическое изображение и пространственная фильтрация. Разрешающая способность оптических систем

Идеальная линза. Построение изображений с помощью лучей (т. е. основанное только на законах геометрической оптики) не позволяет ответить на многие важные вопросы: действительно ли изображение подобно предмету? Каковы предельные возможности оптических систем? Как управлять их характеристиками? Ответ можно найти, используя волновые представления.

Поле в фокальной плоскости линзы. Рассмотрим точечный источник — светящуюся точку S , находящуюся в фокальной плоскости идеальной линзы, на расстоянии ξ от оптической оси (рис. 8.42). Каковы колебания поля в плоскости $z = 0$, примыкающей к линзе слева?

В области значений x , малых по сравнению с расстоянием R_0 от источника до центра линзы O , амплитуду колебаний в сферической волне можно считать постоянной величиной

$a_0/R \approx a_0/R_0 = a = \text{const.}$ Распределение фаз колебаний есть $\varphi(x) = kR = k\sqrt{(x - \xi)^2 + f^2}$. Заменяя, как и ранее, сферический волновой фронт параболическим, найдем

$$\varphi(x) = kR_0 - k \frac{\xi}{R_0} x + \frac{kx^2}{2R_0},$$

где $R_0^2 = f^2 + \xi^2$.

Играет роль, конечно, относительная фаза колебаний, поэтому, принимая фазу колебаний на оптической оси на входе в линзу (т. е. при $x = 0$) нулевой, запишем

$$\varphi(x) = -k \left(\frac{\xi}{R_0} \right) x + \frac{kx^2}{2R_0}.$$

Вводя угол α ($\sin \alpha = \xi/R_0$) и полагая его малым, так что в последнем слагаемом можно положить $R_0 \approx f$, получим

$$\varphi(x) = -kx \sin \alpha + \frac{kx^2}{2f}. \quad (8.100)$$

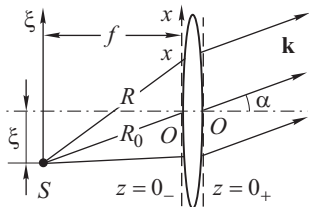


Рис. 8.42

При распространении через линзу возникает дополнительный набег фазы $\Delta\varphi$, определяемый формулой (8.8). Таким образом, на выходе из линзы, т. е. в плоскости, примыкающей к линзе справа, получаем

$$\varphi_+(x) = \varphi(x) + \Delta\varphi(x) = -k \sin \alpha \cdot x. \quad (8.101)$$

Формула (8.101) выражает следующий важный результат: волна от точечного источника, расположенного в фокальной плоскости (квазисферическая волна), преобразуется линзой в волну с плоским волновым фронтом, причем направление волны (угол α) определяется положением источника в фокальной плоскости $\alpha \approx \xi/f$. Разумеется, верно и обратное: плоская волна, падающая на линзу под углом α к оптической оси, т. е. имеющая пространственную частоту $u = k \sin \alpha$, преобразуется идеальной линзой в волну квазисферическую, которая фокусируется (сходится) в точку ξ , расположенную на расстоянии $\xi = f\alpha$ от оптической оси (рис. 8.43)¹⁾.

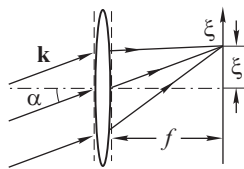


Рис. 8.43

Что собой представляет картина поля в фокальной плоскости, если идеальная линза освещается произвольной монохроматической волной с комплексной амплитудой $f(x)$?

¹⁾ Напомним, что понятия точечного источника, а равно и сферической волны, для поперечной электромагнитной волны суть некоторые приближения при малом параметре λ/R .

Целесообразно представить эту волну в виде суперпозиции плоских волн разных направлений α_n , т. е. разных пространственных частот $u_n = k \sin \alpha_n$ (см. § 8.2, равенство (8.16)). Каждая плоская волна, фокусируясь идеальной линзой в свою точку $\xi_n \approx f \sin \alpha_n = f u_n / k$, создает в этой точке колебания, амплитуда и фаза которых определяется амплитудой и фазой той плоской волны, которая в эту точку фокусируется. Имеется, как мы видим, взаимно однозначное соответствие между точками фокальной плоскости и пространственными частотами плоских волн, суперпозиция которых образует падающую на линзу волну: $u_n = k \xi_n / f$. Таким образом, поле в точке ξ_n фокальной плоскости пропорционально величине c_n в сумме (8.16). Обозначив комплексную амплитуду поля в фокальной плоскости через $g(\xi)$, мы можем записать

$$g(\xi_n) \sim c_n.$$

Разумеется, наше утверждение справедливо и в общем случае, когда вместо дискретной суммы (8.16) волна, падающая на линзу, представляется непрерывной суперпозицией плоских волн:

$$g(\xi) \sim C \left(\frac{k\xi}{f} \right), \quad (8.102)$$

где $C(u) = \int f(x) e^{-iux} dx$ — преобразование Фурье поля $f(x)$.

Итак, картина поля в фокальной плоскости линзы является преобразованием Фурье поля, падающего на линзу.

Заметим, что картина фраунгоферовой дифракции также связана преобразованием Фурье с граничным полем (т. е. с полем, возникающим непосредственно за препятствием), причем аргументом преобразования является величина $k\xi/z$. Таким образом, для наблюдения дифракции Фраунгофера нет необходимости удаляться от препятствия на большое расстояние $z \gg b^2/\lambda$ — достаточно установить за препятствием линзу и наблюдать картину в ее фокальной плоскости, которая лишь масштабом отличается от картины дифракции Фраунгофера.

Принцип двойной дифракции и формирование оптического изображения. Пусть точечный источник света находится на расстоянии z_1 от линзы на ее оптической оси (рис. 8.44). Тогда распределение фаз колебаний в плоскости, примыкающей к линзе слева, определяется формулой $\varphi_-(x, y) = (k/2z_1)(x^2 + y^2)$. После прохождения линзы (в плоскости, примыкающей к ней справа) получим

$$\varphi_+(x, y) = \varphi_-(x, y) + \Delta\varphi(x, y) = \frac{k}{2} \left(\frac{1}{z_1} - \frac{1}{f} \right) (x^2 + y^2),$$

где $\Delta\varphi(x, y)$ — набег фазы волны при прохождении линзы (см. (8.11)).

Сравнив полученную формулу с выражением

$$\varphi(x, y) = -\frac{k}{2z_2} (x^2 + y^2), \quad (8.103)$$

описывающим распределение фаз колебаний в сходящейся сферической волне (см. § 4.1), находим

$$\frac{1}{z_1} + \frac{1}{z_2} = \frac{1}{f}. \quad (8.104)$$

Функция $\varphi(x, y)$ описывает форму волнового фронта в сходящейся волне, если $z_2 > 0$, что соответствует действительному изображению (рис. 8.44). При $z_2 < 0$ формула (8.103) описывает расходящуюся волну, при этом изображение оказывается мнимым (рис. 8.45).

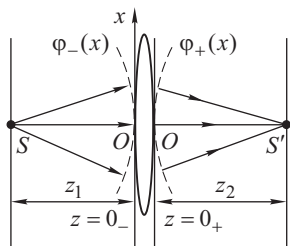


Рис. 8.44

Аналогичным образом можно показать, что формула линзы (8.104) справедлива для любой точки S , расположенной недалеко от оптической оси, т. е. для предметов малого размера.

Формирование изображения с помощью линзы можно рассматривать, основываясь на идее пространственного спектрального разложения. Такой подход аналогичен спектральному анализу процесса вынужденных колебаний в линейной колебательной системе (например, в колебательном контуре). Напомним, что спектральный анализ основан на представлении внешнего воздействия, возбуждающего вынужденные колебания, в виде суперпозиции гармонических колебаний разных частот.

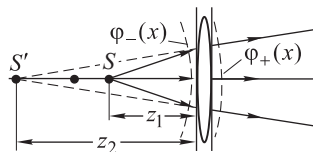


Рис. 8.45

Представим монохроматическую волну, идущую от предмета, в виде суперпозиции плоских волн разных направлений α , т. е. разных пространственных частот $u = k \sin \alpha$. Каждая гармоника — плоская волна определенного направления — фокусируется линзой в «свою» точку фокальной плоскости, в которой возникает, таким образом, картина пространственного спектра: амплитуда и фаза колебаний в точке ξ фокальной плоскости однозначно определяется амплитудой и фазой колебаний той плоской волны, которая в эту точку фокусируется. По этой причине фокальную плоскость линзы называют *фурье-плоскостью*. По терминологии немецкого инженера и оптика Э. Аббе, впервые предложившего такой подход, поле в фокальной плоскости называют *первичным изображением*. На рис. 8.46 показана ситуация, когда предметом является решетка, освещаемая плоской нормально падающей

волной. При этом в фурье-плоскости, как мы знаем, возникает картина фраунгоферовой дифракции: набор ярких точек — дифракционных максимумов. Итак, в процессе распространения света от предмета до фурье-плоскости осуществляется преобразование Фурье светового поля (по терминологии Аббе, *первая дифракция*).

Далее каждая точка фурье-плоскости (каждый дифракционный максимум) рассматривается как источник сферической волны. Все сферические волны, исходящие из разных точек фурье-плоскости, интерферируя, образуют в плоскости, находящейся на расстоянии z_2 за линзой, собственное изображение объекта. Это изображение Аббе назвал *вторичным*, а процесс распространения света от фурье-плоскости до плоскости изображения — *второй дифракцией*.

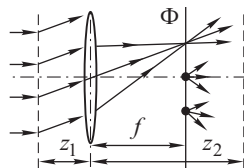


Рис. 8.46

Особенно наглядно принцип двойной дифракции проявляется в оптической схеме, показанной на рис. 8.47. Схема состоит из

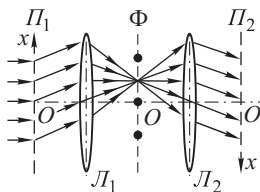


Рис. 8.47

двух линз с общей фокальной плоскостью Φ . Задняя фокальная плоскость линзы L_1 совпадает с передней фокальной плоскостью линзы L_2 . В этом случае первая дифракция — это распространение света от передней фокальной плоскости линзы L_1 , где расположен предмет (плоскость Π_1 на рис. 8.47), к плоскости Φ , где возникает картина пространственного спектра — первичное изображение. Далее, сферическая волна идущая из любой точки фурье-плоскости, преобразуется линзой L_2 в плоскую волну. Таким образом, каждая плоская волна идущая от предмета, преобразуется системой двух линз в плоскую волну, приходящую к плоскости изображения (задней фокальной плоскости Π_2 линзы L_2). Причем, как видно из рис. 8.47, если фокусные расстояния линз одинаковы, то волна с пространственной частотой $u = k \sin \alpha$ преобразуется в волну с пространственной частотой « $-u$ » (x -компонента вектора \mathbf{k} изменяет знак). Это приводит к инверсии — изображение оказывается перевернутым. Можно сказать, что в процессе образования изображения происходит два последовательных преобразования Фурье: от входной плоскости Π_1 к фурье-плоскости Φ — первая дифракция, и затем от фурье-плоскости с помощью линзы L_2 к выходной плоскости Π_2 — вторая дифракция.

Особая роль фурье-плоскости обусловлена тем, что именно в этой плоскости возможно избирательное воздействие на разные пространственные гармоники: установив в любой точке ξ фурье-плоскости маленькую пластинку, вносящую определенное

поглощение и (или) фазовую задержку, мы изменим амплитуду и (или) фазу плоской волны с пространственной частотой $u = k\xi/f$, не изменяя амплитуд и фаз других плоских волн. Устанавливая в фурье-плоскости различные амплитудно-фазовые маски, можно направленно изменять пространственный спектр изображения, влияя таким образом на его характеристики. Этим путем можно решать самые разнообразные задачи: улучшение качества изображений, разрешающей способности оптических систем, визуализация фазовых объектов, выполнение самых разнообразных преобразований пространственной структуры световых полей и т. д., т. е. решать широкий круг задач оптической обработки информации.

Визуализация фазовых объектов. В качестве примера рассмотрим проблему визуализации фазовых объектов, которую можно решить, используя *метод фазового контраста*, предложенный Цернике. Пусть фазовый объект — тонкая прозрачная пластинка, имеющая разный в разных точках показатель преломления (или толщину), но не изменяющая амплитуду прошедшей волны, находится во входной плоскости Π_1 оптической системы, показанной на рис. 8.47. Функция пропускания такой пластинки $t(x) = e^{i\varphi(x)}$, где $\varphi(x) = kn(x)d$ (d — толщина, $n(x)$ — распределение показателя преломления). При освещении пластинки (фазового объекта) плоской, нормально падающей волной, комплексная амплитуда волны в плоскости, примыкающей к пластинке справа, согласно (8.3), есть $f_0(x) = e^{i\varphi(x)}$. Если оптическая система идеальна, то комплексная амплитуда в выходной плоскости Π_2 тождественно повторяет (с точностью до инверсии) входное поле $f_0(x)$, а наблюдаемая картина интенсивности $I_0(x) = |f_0(x)|^2 \equiv 1$, т. е. в плоскости Π_2 мы наблюдаем равномерную засветку: информация о фазовой структуре предмета потеряна, фазовый объект не видим.

Для визуализации фазового объекта Цернике предложил установить в фурье-плоскости, на оптической оси, маленькую фильтрующую пластинку, которая, не изменяя амплитуды прошедшей волны, вносит фазовую задержку, равную $\pi/2$. Проанализируем структуру светового поля в выходной плоскости Π_2 , рассмотрев в качестве примера объект — фазовую синусоидальную решетку с малой глубиной модуляции ($m \ll 1$). В этом случае

$$f_0(x) = e^{im \cos \Omega x} \approx 1 + im \cos \Omega x = 1 + \frac{im}{2} e^{i\Omega x} + \frac{im}{2} e^{-i\Omega x}. \quad (8.105)$$

Итак, входное поле представляется в виде суммы трех слагаемых, в соответствии с этим, от входной плоскости Π_1 вправо распространяются три плоские волны. Первое слагаемое ответственно за появление плоской волны единичной амплитуды

$f_1(x, z) = e^{ikz}$, бегущей вдоль оси оптической системы (оси z). Второе и третье слагаемые — плоские волны с амплитудой $m/2$, направления распространения которых составляют углы $\pm\alpha$ с оптической осью, где $\sin \alpha = \pm\Omega/k$ (т. е. Ω и $-\Omega$ — пространственные частоты этих волн). Обратим внимание, что в точке $x = 0$ входной плоскости колебание первой волны отличается по фазе на $\pi/2$ от колебаний двух наклонных волн (обратите внимание на множитель $i \equiv e^{i\pi/2}$).

Три слагаемых в (8.105) — это три гармонических колебания или, если угодно, три комплексных числа, которые можно изобразить в виде векторов. На рис. 8.48 *а* показано положение этих векторов в начале координат ($x = 0$). Горизонтальный вектор единичной длины изображает колебание, созданное осевой волной, два вектора длины $m/2 \ll 1$, повернутые на угол $\pi/2$ — колебания боковых волн. При смещении из точки $x = 0$

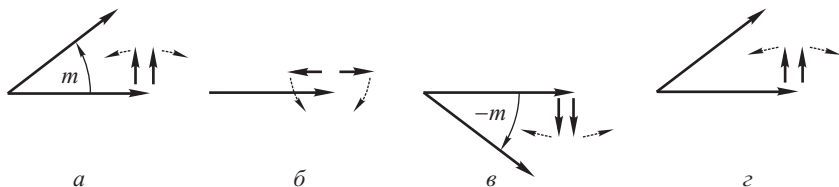


Рис. 8.48

перпендикулярно оси z , фаза колебания осевой волны не меняется, поэтому соответствующий вектор остается горизонтальным. Фазы боковых (наклонных) волн изменяются: $\Delta\varphi = \pm\Omega x$, поэтому изображающие векторы поворачиваются (по и против часовой стрелки). На рис. 8.48 *б* показано положение векторов в точке, где $\Omega x = \pi/2$: векторы повернулись на угол $\pm\pi/2$, а на рис. 8.48 *в* в точке x , где $\Omega x = \pi$, векторы повернулись на угол $\pm\pi$. При смещении на расстояние, равное периоду решетки $x = 2\pi/\Omega$, восстанавливается исходное расположение векторов (рис. 8.45 *г*). Легко видеть, что суммарный вектор, не меняя своей длины (с точностью до величины порядка m^2) изменяет угол наклона φ от $-m$ до $+m$, что и соответствует фазовой структуре волнового поля. Обратите внимание на полную аналогию с векторным изображением колебания, модулированного по фазе (см. § 1.5).

Осевая плоская волна, фокусируясь линзой в начало координат фурье-плоскости ($\xi = 0$) проходит через фазовую фильтрующую пластинку, а две наклонные волны, фокусируясь в точки $\xi_{1,2} = f\alpha = f\Omega/k$, не «задевают» пластинку. Далее линза L_2 преобразует сферические волны, исходящие из точек $\xi = 0$ и $\xi_{1,2} = f\Omega/k$, в плоские волны, которые, интерферируя, образуют изображение.

Наличие маленькой фазовой пластинки в фурье-плоскости на оптической оси приводит к относительной фазовой задержке в $\pi/2$ осевой волны (относительно боковых наклонных волн), поэтому поле выходной плоскости можно записать в виде

$$f(x) = e^{i(\pi/2)} + \frac{im}{2} e^{i\Omega x} + \frac{im}{2} e^{-i\Omega x},$$

или

$$f(x) = i \left(1 + \frac{m}{2} e^{i\Omega x} + \frac{m}{2} e^{-i\Omega x} \right) = i(1 + m \cos \Omega x).$$

Изменение фазовых соотношений между осевой и наклонными волнами иллюстрируют векторные диаграммы, показанные на рис. 8.49. Поворот вектора единичной длины (изображающего колебания осевой волны) на $\pi/2$ приводит к тому, что в точке $x = 0$ все

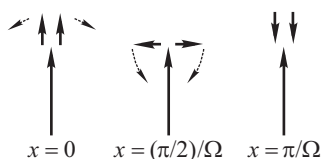


Рис. 8.49

три изображающих вектора оказываются коллинеарными. При смещении из начала координат изменяется фаза боковых волн.

Как видно из векторных диаграмм, длина суммарного вектора при смещении по координате x изменяется, а угол наклона (ориентации) суммарного вектора остается неизменным, что и соответствует чисто амплитудной структуре, т. е. полю с плоским волновым фронтом и меняющейся от точки к точке амплитудой. Таким образом, метод фазового контраста позволяет преобразовать исходную фазовую решетку в амплитудную решетку в плоскости изображения.

Наблюдаемая картина интенсивности имеет вид

$$I(x) = |f(x)|^2 = (1 + m \cos \Omega x)^2 \approx 1 + 2m \cos \Omega x$$

(с самого начала, благодаря условию $m \ll 1$, членами порядка m^2 и выше мы пренебрегаем). Итак, фазовые изменения входного поля $f_0(x)$ оказались визуализированы: мы наблюдаем изменения интенсивности, повторяющие изменения фазы входного поля. Обратите внимание на аналогию описанного здесь метода фазового контраста с методом преобразования колебания, модулированного по фазе, в амплитудно-модулированное колебание (так называемый прием *с изменением фазы несущей*, см. § 1.5). Читатель может самостоятельно проанализировать «метод темного поля». В этом методе вместо фазовой пластинки в $\pi/2$ в фурье-плоскости на оптической оси устанавливается непрозрачный маленький экран. Осевая плоская волна, фокусируясь линзой в начало координат фурье-плоскости, поглощается непрозрачным экраном и не участвует в формировании изображения. Боковые же волны остаются без изменения. Какова в этом случае картина интенсивности в выходной плоскости?

Метод темного поля аналогичен методу, который в радиотехнике используется для преобразования фазовой модуляции в амплитудную и называется «приемом без несущей».

Разрешающая способность оптических систем. Всякая реальная оптическая система отличается от идеальной по крайней мере одним важным обстоятельством: ее линзы имеют конечные размеры. Учет конечных размеров используемых объективов чрезвычайно важен для оценки предельных возможностей оптических систем. Необходимо иметь в виду, что функция пропускания линзы (8.12), о которой мы говорили ранее, правильно описывает преобразования волны линзой только в той области значений $xу$, которые находятся в пределах «зрчка» линзы. Если диаметр линзы D , то функция (8.12) справедлива при $|x^2 + y^2| \leq (D/2)^2$. Действительно, свет, проходящий «мимо» линзы, не представляет интереса, поскольку не участвует в формировании изображения. Все происходит так, как если бы идеальная бесконечная линза (функция пропускания которой описывается формулой (8.12)) была задиафрагмирована непрозрачным экраном с отверстием диаметра D , оставляющим открытой центральную часть линзы. Именно такая модель используется обычно для расчета дифракционных эффектов, связанных с конечным размером объектива.

Итак, пусть на линзу падает плоская волна, распространяющаяся вдоль оптической оси. Волна излучается удаленным источником, настолько удаленным, что в пределах площади зрчка, т. е. в пределах диафрагмы диаметра D (рис. 8.50), волновой фронт неотличим от плоского.

Пройдя через диафрагму, волна падает на идеальную (бесконечную) линзу. Изображенная на рисунке оптическая схема есть не что иное, как схема наблюдения дифракции Фраунгофера плоской волны на круглом отверстии диаметра D . Мы уже отмечали, что для наблюдения дифракции Фраунгофера на каком-либо препятствии достаточно установить за препятствием линзу — при этом картина в фокальной плоскости линзы подобна картине, которая возникает на большом расстоянии за препятствием ($z \gg D^2/\lambda$).

Еще раз поясним, в чем суть дела. Плоская волна, пройдя через отверстие-диафрагму, уже не является плоской волной: за отверстием мы имеем непрерывный спектр плоских волн (т. е. набор плоских волн разных направлений). Каждая плоская волна из этого спектра фокусируется идеальной линзой в свою точку фокальной плоскости, в результате мы и получаем распределе-

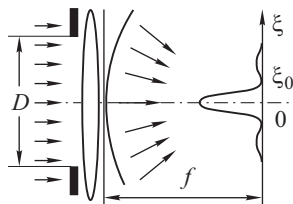


Рис. 8.50

ние интенсивности, показанное на рис. 8.51 — дифракционное пятно (*пятно Эйри*), в котором концентрируется подавляющая доля светового потока, окруженное чередующимися светлыми (очень слабыми) и темными кольцами. Полуширина пятна Эйри (см. (8.68))

$$\xi_0 = 1,22 \frac{\lambda f}{D}. \quad (8.106)$$

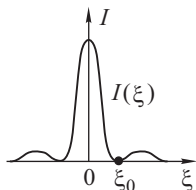


Рис. 8.51

Это и есть изображение бесконечно удаленного точечного источника: изображение не является точкой, как следует из законов геометрической оптики, а представляет собой дифракционное пятно, причем чем больше размер объектива D , тем меньше дифракционное пятно, тем больше это пятно похоже на точку.

Следует подчеркнуть, что дифракционная формула (8.106) не приводит к правильному результату в пределе при $D \rightarrow \infty$: при увеличении D пятно не становится как угодно малым. Мы уже подчеркивали, что бегущие волны не могут образовать в сумме световое поле, которое бы резко изменялось на расстояниях порядка длины волны. Пятно, размер которого существенно меньше длины волны, получить невозможно!

Поток энергии, падающий на линзу и равный $I_0 S_0$, где I_0 — интенсивность волны, $S_0 = \pi D^2/4$ — площадь линзы, концентрируется в маленькое пятнышко, площадь которого равна $S = \pi \xi_0^2 \approx \pi (\lambda f/D)^2$. Средняя интенсивность волны в пятне Эйри оказывается при этом равной

$$I \approx I_0 \frac{S_0}{S} \approx \frac{I_0}{4} \frac{D^4}{(\lambda f)^2}. \quad (8.107)$$

Для линзы диаметром 1 см с фокусным расстоянием $f = 10$ см выигрыш от фокусировки для длины волны $\lambda = 500$ нм (зеленый цвет) равен $I/I_0 \approx 10^6$! В реальных линзах такой выигрыш достигается, если удастся устранить aberrации, возникающие в линзах, поверхности которых имеют сферическую (а не параболическую) форму.

Дифракционная формула (8.106) с тем же успехом применима для оценки размера дифракционного пятна в фокусе параболического зеркала диаметром D (рис. 8.52), поскольку распределение фаз колебаний в плоскости, примыкающей к зеркалу (после отражения плоской волны от зеркала), аналогично фазовому распределению (8.11), которое обеспечивается линзой.

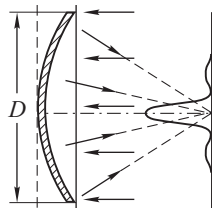


Рис. 8.52

При освещении линзы плоской нормально падающей волной на выходе из линзы в плоскости, примыкающей к ней справа, распределение фаз колебаний описывается формулой (8.8). Эта формула означает, что колебания в точке, отстоящей на расстояние x от оптической оси, опережают по фазе колебания в точке, расположенной на оптической оси, на величину $\Delta\varphi = kx^2/(2f)$.

Вспомним теперь о принципе Гюйгенса–Френеля: представим себе, что плоскость, примыкающая к линзе справа, заполнена вторичными источниками, излучающими сферические волны. Легко сообразить, что колебания, созданные всеми этими вторичными источниками в точке F — главном фокусе линзы — оказываются синфазными, поскольку опережение по фазе $\Delta\varphi(x)$ компенсируется отставанием, возникающим при распространении света от вторичного источника до точки F . Действительно, разность расстояний равна: $R - f = \sqrt{x^2 + f^2} - f \approx x^2/(2f)$, а соответствующая разность фаз $k(R - f) = kx^2/(2f)$ в точности компенсирует начальную разность фаз $\Delta\varphi$ для любой точки x . Это свойство линзы называется *таутохронизмом*: все оптические пути (для любого луча, проходящего от источника через линзу в точку, являющуюся изображением) занимают одно и то же время.

Синфазность колебаний в точке F от всех вторичных источников, заполняющих плоскость $z = 0_+$ (примыкающую к линзе справа) означает, что с помощью линзы мы выстраиваем все элементарные векторы, составляющие спираль Френеля, в коллинеарную цепочку векторов: часть спирали Френеля, образованная зонами, уложившимися в диафрагму, как бы вытягивается в прямую линию (сравните с векторной диаграммой для зонной пластинки Френеля, см. рис. 8.19).

Длина полувитка спирали равна πA_0 , где A_0 — амплитуда волны, падающей на линзу. Тогда суммарная амплитуда колебаний есть $A = m\pi A_0$, где m — число зон Френеля, укладывающихся на линзе диаметром D : $D/2 = \sqrt{m\lambda f}$, откуда $m = D^2/(4\lambda f)$ и $A = \pi D^2/(4\lambda f) A_0$. Отношение интенсивности света в фокусе к интенсивности падающей на линзу волны равно

$$\frac{I}{I_0} = \left(\frac{A}{A_0}\right)^2 = \frac{\pi^2}{16} \frac{D^4}{(\lambda f)^2}. \quad (8.108)$$

Разумеется, интенсивность в фокусе больше, чем средняя интенсивность в пятне Эйри.

Если точечный источник света находится на конечном расстоянии z_1 от линзы, то дифракционное пятно Эйри возникает не в фокальной плоскости, а на расстоянии z_2 за линзой, где z_1 и z_2 связаны формулой линзы (8.104). При этом полуширина

пятна определяется формулой

$$\xi_0 = 1,22 \frac{\lambda z_2}{D}. \quad (8.109)$$

Это и есть изображение точечного источника.

Обратимся теперь непосредственно к вопросу о разрешающей способности оптических систем. Пусть во входной плоскости (на расстоянии z_1 от линзы) находятся два близко расположенных некогерентных источника света S_1 и S_2 . В плоскости изображения (на расстоянии z_2) возникают два пятна Эйри. В силу некогерентности источников картина интенсивности в плоскости изображения представляет собой сумму интенсивностей, созданных каждым из источников. Если расстояние между центрами пятен Эйри мало (меньше их размера), то пятна накладываются друг на друга. Возникает вопрос: можно ли по наблюдаемой в плоскости изображения картине интенсивности сделать вывод о том, что во входной плоскости действительно имеются два источника света, а не один (вдвое более яркий)?

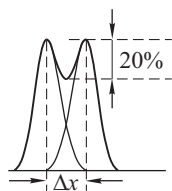


Рис. 8.53

Согласно критерию Рэлея предельно допустимой является ситуация, показанная на рис. 8.53: расстояние между центрами пятен Δx в точности равно полуширине пятна Эйри:

$$\Delta x = \xi_0 = 1,22 \frac{\lambda z_2}{D}.$$

При этом, как показывает расчет, в суммарной картине интенсивности между двумя максимумами возникает небольшой $\sim 20\%$ провал, по наличию которого можно визуальнo разрешить два источника.

Смещению пятен Δx в плоскости изображения соответствует расстояние между точечными источниками во входной плоскости (на расстоянии z_1 от линзы), равное $l = (z_1/z_2) \Delta x$. Мы получаем оценку минимального разрешимого расстояния между точечными источниками

$$l_{\min} = 1,22 \frac{\lambda z_1}{D}. \quad (8.110)$$

Если вы рассматриваете с расстояния наилучшего зрения $z_1 \approx 25$ см две светящиеся точки, то, принимая диаметр зрачка равным $D \approx 3$ мм и $\lambda = 500$ нм (наиболее чувствительная для глаза область спектра), получим $l_{\min} \approx 5 \cdot 10^{-2}$ мм. Это и есть оценка минимально разрешимого расстояния для человека с нормальным зрением.

При оценке возможностей любой оптической системы (в том числе глаза человека) следует принимать во внимание и характеристики непосредственно регистрирующей среды: фотоэмуль-

сии, если мы используем фотопленку или фотопластинку, или чувствительных к свету элементов сетчатки человеческого глаза (палочки и колбочки), если речь идет о визуальном наблюдении. Любая регистрирующая среда обладает определенной дискретностью; чувствительные элементы — зерна фотоэмульсии или палочки в сетчатке глаза имеют конечные размеры и, что существенно, реагируют на свет (поток энергии) как целое. Это означает, в частности, что свет, фокусируясь оптической системой в малое пятнышко, которое, быть может, много меньше размеров зерна, вызывает химическую реакцию в зерне целиком. Другими словами, минимальный размер площадки в фоточувствительной среде, реагирующий на свет, не может быть меньше размера зерна фотоэмульсии.

Интересно, что размер палочек в сетчатке человеческого глаза удивительным образом согласован с размером зрачка, т. е. как раз равен размеру пятна Эйри на сетчатке.

Если рассматриваются удаленные объекты, то обычно говорят об угловом расстоянии $\alpha = l/z_1$. Изображение рассматривается при этом в фокальной плоскости линзы (при $z_1 \rightarrow \infty$, $z_2 \rightarrow f$). Согласно (8.109) имеем

$$\alpha_{\min} \approx 1,22 \frac{\lambda}{D}.$$

Рассматривая удаленные источники невооруженным глазом ($D \approx 3$ мм), получаем $\alpha_{\min} \approx 2 \cdot 10^{-4}$ рад. При наблюдении удаленных источников (например двух звезд) в телескоп с диаметром объектива $D = 5$ м, находим $\alpha_{\min} \approx 1 \cdot 10^{-7}$ рад. Зеркальный объектив такого размера, т. е. параболическое зеркало, используемое для фокусировки света, находится в лаборатории Маунт-Паломар, в Кордильерах, США; на Кавказе имеется телескоп с диаметром зеркала $D \approx 6$ м (телескоп-рефлектор).

8.12. Принципы голографии

Мы уже отмечали ранее, что любой приемник света (глаз человека, фотопластинка, фотоэлемент) реагирует на энергию световой волны, т. е. величину, пропорциональную квадрату амплитуды волны. По этой причине все эти приемники объединяют общим названием — *квадратичные детекторы*. При таком способе регистрации волнового поля информация о фазовой структуре волны (т. е. о форме волновых поверхностей) оказывается как будто бы утраченной. Проблема, которую называют в оптике «фазовой проблемой», состоит в том, чтобы по картине интенсивности (квадрата амплитуды) восстановить полную информацию о волновом поле: не только распределение амплитуд колебаний, но также и фазовую структуру волны, т. е. форму волновых поверхностей.

Пусть некоторый предмет (например, пешка, изображенная на рис. 8.54) освещается когерентным светом лазера и пусть волна, отраженная предметом (будем называть ее далее *предметной волной*), создает в плоскости $z = 0$ световое поле:

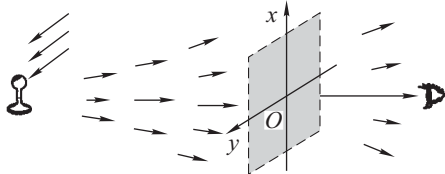


Рис. 8.54

$$f_{\text{п}}(x, y) = a(x, y)e^{i\varphi(x, y)}.$$

Сейчас мы хотим подчеркнуть наиболее важное для понимания идеи голографии обстоятельство: задание распределения амплитуд $a(x, y)$ и фаз $\varphi(x, y)$ колебаний в плоскости $z = 0$ однозначно определяет ход волнового процесса в области $z > 0$. Это утверждение означает, в частности, следующее: представим себе, что нам удалось воссоздать в плоскости $z = 0$ то же самое распределение амплитуд и фаз колебаний в отсутствие предмета.

Тогда в области $z > 0$ возникает волновое поле, тождественное полю, созданному предметом: ни один наблюдатель, находящийся в области $z > 0$, не отличит это поле от поля реально присутствующего предмета. Наблюдатель увидит этот предмет именно там, где он находился, создавая на плоскости $z = 0$ поле $f_{\text{п}}(x, y)$. Вопрос состоит в том, как реализовать «правильное» граничное условие? Попробуем сделать это, установив в плоскости $z = 0$ в схеме эксперимента, изображенного на рис. 8.54, фотопластинку. Предположим, что функция пропускания фотопластинки после необходимой обработки (проявление, закрепление) пропорциональна интенсивности волны, освещающей пластину во время экспозиции, т. е. пропорциональна величине $I(x, y) = |a(x, y)|^2$. Попробуем теперь, используя полученную пластинку, восстановить изображение предмета. Установим пластинку в плоскости $z = 0$ и осветим ее слева плоской, нормально падающей волной (предмет при этом отсутствует). Пластинка, изменяя амплитуду прошедшей волны, не влияет на форму волнового фронта, поэтому в плоскости, примыкающей к ней справа, получим световое поле с плоским волновым фронтом (и разной в разных точках (x, y) амплитудой колебаний). Ясно, что полученное на выходе из пластинки (в плоскости $z = 0_+$) граничное поле отличается от поля, которое создавалось предметной волной. Поэтому и волна в области $z > 0$, будет иметь мало общего с волной, возникающей при отражении от предмета — изображение не восстановится.

Попробуем поступить по другому. Представим себе, что нам удалось каким-то образом изготовить тонкую пластинку-

транспарант толщины d , амплитудная прозрачность которой меняется от точки к точке по закону $a(x, y)$, а набег фазы волны при прохождении пластинки есть $\varphi(x, y)$; например, за счет правильно подобранного распределения показателя преломления $n(x, y)$: $\varphi(x, y) = kn(x, y)d$, т. е. функция пропускания пластинки есть $t(x, y) = a(x, y)e^{i\varphi(x, y)}$. Установим эту пластинку в плоскости $z = 0$ и осветим ее слева плоской нормально падающей волной (никакого предмета при этом нет). Тогда на выходе из пластинки, в плоскости, примыкающей к ней справа, мы и получим поле с требуемой комплексной амплитудой $f_{\text{п}}(x, y)$, а в области $z > 0$ возникнет волна, неотличимая от волны, созданной предметом. Наблюдатель, находящийся в области $z > 0$, который смотрит на пластинку (как в окно) увидит предмет, именно в том месте и в том положении, где он находился, создавая граничное поле. Однако технологические трудности, возникающие на таком пути получения изображений, трудно преодолимы.

Голография — интерференционный метод регистрации и последующего восстановления изображений, предложенный Габором в 1947 г., позволяет обойти эти трудности. Следуя идее Габора, изменим схему эксперимента, изображенного на рис. 8.54. На фотопластинку, установленную в плоскости $z = 0$ (рис. 8.55), теперь падают две волны: предметная волна, создающая на фотопластинке поле $f_{\text{п}}(x, y)$, и опорная волна с известным распределением амплитуд и фаз колебаний $f_{\text{о}}(x, y)$. При этом необходимо обеспечить когерентность предметной и опорной волны — волны должны интерферировать. Напри-

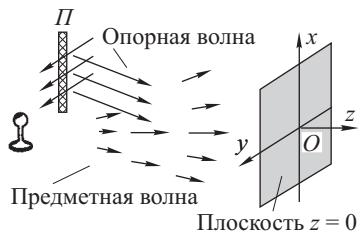


Рис. 8.55

мер, как показано на рис. 8.55: излучение лазера частично проходит через полупрозрачную пластинку Π , освещая предмет, а частично отражается от нее, создавая опорный пучок. Когерентность предметной и опорной волны обеспечивается высокой степенью монохроматичности лазерного излучения (разность хода между предметной и опорной волной меньше длины когерентности для любой точки фотопластинки). Суммарное поле на фотопластинке

$$f(x, y) = f_{\text{п}}(x, y) + f_{\text{о}}(x, y),$$

а интенсивность (а следовательно, и функция пропускания фотопластинки после необходимой обработки) есть

$$t(x, y) \sim I(x, y) = |f_{\text{п}}(x, y) + f_{\text{о}}(x, y)|^2.$$

Фотопластинка с зарегистрированным на ней результатом интерференции предметной и опорной волны называется *голограммой*.

Для уяснения сути метода рассмотрим в качестве предмета точечный источник света S , т. е. создадим сферическую предметную волну. В качестве опорной волны возьмем плоскую волну, бегущую вдоль оси z и падающую нормально на фотопластинку, расположенную в плоскости $z = 0$ (рис. 8.56):

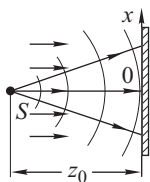


Рис. 8.56

$$f_0 = ae^{ikz},$$

т. е. опорная волна создает во всех точках фотопластинки поле одинаковой амплитуды. Принимая фазу колебаний в плоскости $z = 0$ равной нулю, запишем

$$f_0 = a.$$

Поле предметной волны есть $a_0 e^{ikr}/r$, где $r = \sqrt{z_0^2 + x^2 + y^2}$ — расстояние от источника S до точки (x, y) фотопластинки. Для упрощения формул будем далее полагать, что амплитуда сферической волны во всех точках фотопластинки равна амплитуде плоской волны. Тогда $f_{\text{п}} \approx ae^{ikr}$, суммарное поле есть

$$f = ae^{ikr} + a.$$

После необходимой фотообработки получаем голограмму с функцией пропускания

$$t(x, y) \sim I(x, y) = |a + ae^{ikr}|^2. \quad (8.111)$$

Мы описали первую стадию голографического процесса — процесс записи голограммы. Теперь необходимо использовать полученную голограмму для восстановления (реконструкции) изображения.

Устанавливаем голограмму в плоскости $z = 0$ и освещаем ее плоской нормально падающей волной (рис. 8.57). Для упрощения формул будем считать амплитуду волны равной единице, а фазу без ограничения общности равной нулю, т. е. комплексная амплитуда этой волны (ее называют *восстанавливающей волной*) есть $f_-(x, y) = 1$.

Тогда на выходе голограммы, в плоскости, примыкающей к ней справа, получим

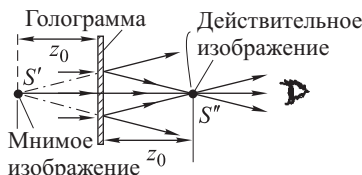


Рис. 8.57

$$f_+(x, y) = f_-(x, y)t(x, y) = |a + ae^{ikr}|^2. \quad (8.112)$$

Итак, суммарное поле за голограммой описывается действительно положительной функцией. Исследуем его более детально. Из (8.112) находим

$$f_+(x, y) = 2a^2 + a^2 e^{ikr} + a^2 e^{-ikr}. \quad (8.113)$$

Поле в плоскости $z = 0_+$ представляется в виде суммы трех слагаемых. Соответственно, волна в области $z > 0$ есть сумма трех волн: каждое слагаемое на границе $z = 0_+$ ответственно за появление «своей» волны в области $z > 0$.

Первое слагаемое $2a^2$ «отвечает» за появление волны $f_1 = 2a^2 e^{ikz}$ — плоской волны, бегущей в области $z > 0$ вдоль оси z . Второе слагаемое $f_2 = a^2 e^{ikr}$ — волновое поле со сферическим волновым фронтом: именно такое поле создавал в каждой точке плоскости $z = 0$ точечный источник света, который находился на расстоянии z_0 слева от голограммы в процессе записи. Ясно, что и в области $z > 0$ поле f_2 порождает расходящуюся сферическую волну, неотличимую от предметной волны (сферической волны, исходящей из точки S в процессе записи голограммы). Поэтому наблюдатель, который находится справа (в области $z > 0$) и смотрит на голограмму как в окошко, «увидит в окне», на расстоянии z_0 слева от голограммы, светящуюся точку S' , т. е. расходящаяся сферически волна кажется наблюдателю исходящей из точки S' , хотя никакой светящейся точки на стадии реконструкции изображения нет — есть только фотопластинка-голограмма, освещаемая плоской волной. Это — мнимое изображение предмета. Рассмотрим теперь последнее, третье слагаемое в (8.113): $f_3 = a^2 e^{-ikr}$.

Волновой фронт в этой волне «вывернут наизнанку» по отношению к волне f_2 : если в волне f_2 колебание в точке (x, y) плоскости $z = 0$ отстает по фазе от колебаний в начале координат, то в волне f_3 колебание в точке (x, y) опережает по фазе колебание в начале координат на ту же величину. Это слагаемое описывает поле, которое создавала бы в плоскости $z = 0$ сферическая волна, сходящаяся в точку S'' на расстоянии z_0 справа (рис. 8.57). Ясно поэтому, что граничное поле f_3 , возникающее в процессе реконструкции, порождает сферическую волну, сходящуюся в точку S'' . Это — действительное изображение предмета. Его можно наблюдать, расположив на расстоянии z_0 за голограммой экран наблюдения — мы увидим яркую точку, в которую сфокусирована волна.

Описанные закономерности сохраняются в общих чертах при голографировании произвольного предмета: в процессе реконструкции возникают три волны: волна создающая мнимое изображение, которое находится слева от голограммы, там же, где находился предмет при записи; волна, создающая действитель-

ное изображение, которое располагается симметрично справа от голограммы; а также волна, бегущая вдоль оси z и не несущая информации о форме волнового фронта предметной волны.

Вернемся к интерференционной картине (8.112), которая записана на голограмме точечного источника. Как она выглядит?

Функцию (8.113) можно переписать в виде

$$I(x, y) = a^2 |1 + e^{ikr}|^2 = 2a^2(1 + \cos kr).$$

Используем приближенное выражение $r \approx \sqrt{z_0^2 + x^2 + y^2} \approx z_0 + \frac{x^2 + y^2}{2z_0} = z_0 + \rho^2/2z_0$, где $\rho^2 = \sqrt{x^2 + y^2}$ — расстояние от начала координат до точки (x, y) в плоскости фотопластинки. Соответственно фаза колебаний в сферической волне в точке (x, y) есть $\varphi = kr = kz_0 + k\rho^2/(2z_0)$. Принимая начальную фазу на оси z в плоскости $z = 0$ нулевой, запишем: $\varphi = k/(2z_0)\rho^2$. Мы полагаем таким образом, что колебания опорной (плоской) волны и сферической волны в точке $(x = 0, y = 0)$ синфазны. В результате получим

$$I(\rho) = 2a^2 \left(1 + \cos \frac{k\rho^2}{2z_0} \right). \quad (8.114)$$

Мы видим, что интерференционная картина имеет вид колец, центр которых находится в начале координат. Функция $I(\rho)$ показана на рис. 8.58. Радиус первого (темного) кольца ρ_1 находим из условия $k\rho_1^2/(2z_0) = \pi$ (при этом $\cos(k\rho_1^2/2z_0) = -1$ и $I(\rho_1) = 0$) откуда получаем: $\rho_1 = \sqrt{\lambda z_0}$. Вообще радиусы светлых и темных колец находятся по формуле

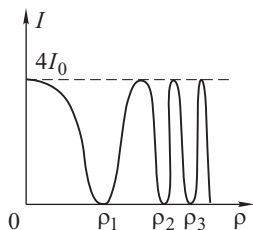


Рис. 8.58

$$\rho_m = \sqrt{m\lambda z_0}$$

(при m нечетном — темные кольца, при m четном — светлые), которая совпадает с выражением, определяющим радиусы темных и светлых колец в зонной пластинке Френеля. Отличие состоит в том, что в нашем случае переход от светлых колец к темным происходит плавно. Интерференционная картина (8.114) (голограмма точечного источника) показана на рис. 8.59 и называется *зонной решеткой Габора* (по имени Д. Габора (1900–1979)). Как следует из проведенного выше анализа, зонная пластинка Габора работает одновременно и как собирающая линза (фокусируя параллельный пучок света в точку S'' — действительный фокус зонной пластинки), и как рассеивающая линза,



Рис. 8.59

которая преобразует параллельный пучок, падающий на пластинку, в расходящуюся сферическую волну, исходящую из мнимого фокуса S' , и как плоскопараллельная пластинка, поскольку часть света за пластинкой сохраняет структуру освещающего параллельного пучка.

В случае произвольного предмета на голограмме записывается сложный интерференционный узор, который можно рассматривать как совокупность зонных кольцевых решеток Габора, причем каждая из них ответственна за восстановление «своей» точки предмета на стадии реконструкции изображения.

8.13. Разрешающая способность голограммы

Рассмотрим более детально волну, ответственную за создание действительного изображения. Конечно, сходящаяся сферическая волна не фокусируется строго в точку S'' : изображением является не точка, а маленькое дифракционное пятнышко. Действительно, плоская восстанавливающая волна, пройдя сквозь голограмму, преобразуется в сферическую волну (мы говорим сейчас о той части волны, которая ответственна за создание действительного изображения), причем сферический волновой фронт ограничен размерами голограммы. Эта сходящаяся сферическая волна ничем не отличается от сферической волны, возникающей за собирающейся линзой, размер D которой равен размеру голограммы. При фокусировке света линзой возникает дифракционное пятно (пятно Эйри), размер которого определяется формулой (8.109). Ясно, что и голограмма создает изображение — дифракционное пятно, размер которого определяется аналогичной формулой:

$$\Delta x \approx \frac{\lambda z_0}{D}, \quad (8.115)$$

где z_0 — расстояние от точечного источника до голограммы в процессе записи, D — размер голограммы. Если записывать на голограмму изображения двух точек, расстояние между которыми меньше Δx , то при восстановлении изображений возникают два пятна, налагающиеся друг на друга так, что, согласно критерию Рэлея, они оказываются неразрешимыми. Сказанное касается, разумеется, и мнимого изображения точечного источника: наблюдателю оно кажется маленьким пятнышком размера Δx , находящимся на расстоянии z_0 за голограммой.

Ограничения на разрешающую способность голограммы могут быть связаны также с зернистостью регистрирующей среды. Для того чтобы оценка (8.115) была справедлива, необходимо, чтобы размер зерна фоторегистрирующей среды был меньше мельчайшего размера интерференционных полос на голограмме, возникающих при интерференции предметной и опорной волн.

Вернемся к примеру, рассмотренному выше — голограмме точечного источника. Оценим минимальное расстояние между интерференционными кольцами в зонной пластинке Габора диаметра D (это расстояние уменьшается от центра голограммы к краю). Пусть кольцо, примыкающее к краю голограммы, имеет порядок m . Тогда $r_m = D/2 = \sqrt{m\lambda z}$, откуда $m = D^2/(4\lambda z)$. Минимальное расстояние между кольцами

$$\begin{aligned}\Delta r = r_m - r_{m-1} &= \sqrt{m\lambda z} \left(1 - \sqrt{\frac{m-1}{m}} \right) = \\ &= \sqrt{m\lambda z} \left(1 - \sqrt{1 - \frac{1}{m}} \right) \approx \frac{D}{2} \left[1 - \left(1 - \frac{1}{2m} \right) \right] = \frac{D}{4m},\end{aligned}$$

мы полагаем $m \gg 1$. Отсюда имеем

$$\Delta r \approx \frac{D}{4m} = \frac{D \cdot 4\lambda z}{4D^2} = \frac{\lambda z}{D}.$$

Итак, чтобы полностью использовать разрешающую способность голограммы размера D , необходимо использовать фоторегистрирующую среду с размером зерна δ_3 , определяемым условием $\delta_3 \leq \lambda z/D$.

8.14. Схема с наклонным опорным пучком

При наблюдении голографических изображений, полученных по схеме Габора, действительное и мнимое изображения создают взаимные помехи: изображения S' и S'' находятся на одной прямой с точкой, в которой расположен глаз наблюдателя (рис. 8.57). Помехи создаются также возникающей при восстановлении плоской волной, бегущей вдоль оси z и не несущей информации об объекте. В 1962 г. американские исследователи Лейт и Упатниекс предложили использовать для записи голограммы наклонный опорный пучок.

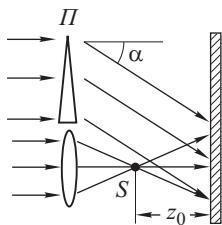


Рис. 8.60

Проследим, каким образом введение наклонного опорного пучка обеспечивает возможность независимого (без взаимных помех) наблюдения действительного и мнимого изображений. На рис. 8.60 «предметом» является «точечный источник» S . Наклон опорного пучка обеспечивается призмой P . Поле предметной волны на голограмме есть $f_{\Pi}(x, y) = ae^{ik/(2z_0)(x^2+y^2)}$ (мы вновь используем параболическое приближение). Наклонная опорная волна имеет вид $f_0 = ae^{iu_0 x}$, где $u_0 = k \sin \alpha$ — пространственная частота опорной волны. Мы полагаем, что волны имеют равные

амплитуды и создают в начале координат плоскости xy (плоскость голограммы) синфазные колебания. Функция пропускания проявленной голограммы, пропорциональная интенсивности суммарной волны при записи, имеет вид

$$t(x, y) \sim \left| a e^{i u_0 x} + a \exp \left[\frac{ik}{2z_0} (x^2 + y^2) \right] \right|^2.$$

Предоставляем читателю убедиться, что слагаемые, ответственные за появление действительного и мнимого изображения при восстановлении, можно представить в виде

$$f_3(x, y) \sim b \exp \left[\frac{ik}{2z_0} \left((x + z_0 \sin \alpha)^2 + y^2 \right) \right],$$

$$f_4(x, y) \sim b \exp \left[-\frac{ik}{2z_0} \left((x + z_0 \sin \alpha)^2 + y^2 \right) \right],$$

где $b = a_0^2 e^{(iz_0/2k)u_0^2}$ — комплексная константа. Выражение $f_3(x, y)$ представляет собой поле, которое создавал бы в плоскости $z = 0$ точечный источник, находящийся в точке с координатами $x_0 = -z_0 \sin \alpha$, $y = 0$, на расстоянии z_0 слева от голограммы. Это мнимое изображение предмета S' (рис. 8.61). Соответственно $f_4(x, y)$ описывает поле сходящейся в точку S'' сферической волны. Эта точка находится справа от голограммы на расстоянии z_0 и смещена относительно оси z на расстояние $x_0 = -z_0 \sin \alpha$. Это — действительное изображение. Мы видим, что введение наклонного опорного пучка приводит к сдвигу мнимого

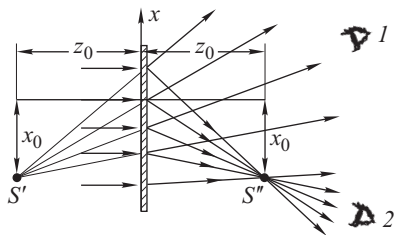


Рис. 8.61

и действительного изображения таким образом, что наблюдатель может рассматривать их из разных положений (1 и 2 на рис. 8.61), в которых они не создают взаимных помех. Разумеется, угол наклона опорного пучка α (т. е. его пространственная частота) должен быть выбран правильно: чем шире пространственный спектр предметной волны, тем больше должен быть угол наклона опорного пучка, при котором волны, ответственные за создание действительного и мнимого изображений, не создают взаимных помех.

8.15. Объемная голограмма

То, что мы рассматривали до сих пор, называется *тонкой голограммой*. Тонкой потому, что интерференция предметной и опорной волны регистрируется в тонком слое регистрирующей

среды-фотоэмульсии. Рассмотрим теперь, какие возможности открывает объемная регистрирующая среда. Чтобы понять, когда и почему начинает играть роль третье измерение — толщина фоторегистрирующей среды, — посмотрим, каков результат интерференции двух плоских волн, если рассматривать его не только на плоскости $z = \text{const}$, но и по толщине, вдоль оси z . Итак, пусть в фотоэмульсии распространяются две волны. Заметим, что до фотообработки (проявления) фотоэмульсия прозрачна. Одну из волн, бегущую вдоль оси z , $f_o = ae^{ikz}$ будем считать опорной, а вторую $f_{\Pi} = ae^{i(kx \sin \alpha + kz \cos \alpha)}$, волновой вектор которой составляет угол α с осью z , — предметной волной (при этом мы имеем в виду, что любую предметную волну можно представить в виде суперпозиции плоских волн разных направлений α). Легко найти интенсивность суммарной волны $I(x, z) = |f_o + f_{\Pi}|^2 = 4a^2 \cos^2 [kx \sin \alpha/2 - kz(1 - \cos \alpha)/2]$. Максимумы интенсивности располагаются вдоль плоскостей

$$x \sin \alpha - z(1 - \cos \alpha) = \pi m \quad (m - \text{целое число}).$$

При этом $I(x, z) = 4a^2$. Эти плоскости составляют с осью z угол β , равный $\text{tg } \beta = (1 - \cos \alpha) / \sin \alpha = \text{tg}(\alpha/2)$, т. е. $\beta = \alpha/2$: плоскости интерференционных максимумов располагаются по биссектрисе угла между направлениями опорной волны f_o и

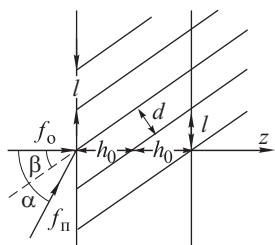


Рис. 8.62

предметной f_{Π} (рис. 8.62). В любой плоскости $z = \text{const}$ волны сходятся под одним и тем же углом α , поэтому и ширина полосы l в любой фиксированной плоскости $z = \text{const}$ одна и та же и равна $l = \lambda / \sin \alpha$. Расстояние между плоскостями интерференционных максимумов

$$d = l \cos \beta = \frac{\lambda}{2 \sin(\alpha/2)}. \quad (8.116)$$

При переходе от одной плоскости $z = \text{const}$ к другой происходит сдвиг интерференционных полос. Сдвиг на ширину полосы происходит при расстоянии Δz между сечениями $z = \text{const}$, равном

$$\Delta z = h_0 = \frac{l}{\text{tg } \beta} = \frac{\lambda}{1 - \cos \alpha}.$$

Перемещаясь вдоль оси z на расстояние $h_0/2$, мы перейдем из максимума в минимум.

В максимумах, где интенсивность суммарной волны велика, велико химическое воздействие света на среду, в минимумах воздействие света мало. Будем полагать, что плоскости, где интенсивность света велика, становятся после обработки подобными частично отражающим свет зеркалам. Плоскости, где интенсивность близка к нулю, после обработки полностью прозрачны.

Итак, в результате обработки в фотоэмульсии образуется набор «зеркал», частично отражающих, а частично прозрачных для света. Период d этой объемной решетки зеркал определяется формулой (8.116).

Фотообработка проэкспонированной голограммы может производиться и таким образом, что от интенсивности света при экспозиции будет зависеть не отражающая способность слоев фотоэмульсии, а их оптическая плотность (коэффициент преломления n). В этом случае объемная решетка оказывается фазовой.

Толщина h фоторегистрирующей среды никак не будет проявляться, если она заметно меньше величины $h_0/2$: ведь в этом случае тонкий луч света, проходя через среду, не будет пересекать слои различной прозрачности. Такая голограмма аналогична тонкому экрану. Наоборот, при толщине, много большей h_0 , луч света, бегущий вдоль оси z , много раз пересекает слои разной прозрачности, частично проходя сквозь каждый слой — зеркало, а частично отражаясь от него. Итак, эффект объемности фоторегистрирующей среды проявляется, если ее толщина h превышает величину h_0 :

$$h > h_0 = \frac{\lambda}{1 - \cos \alpha}. \quad (8.117)$$

Из последнего условия ясно, что один и тот же слой фотоэмульсии можно считать тонким или толстым, в зависимости от того, как организован эксперимент. Если волны сходятся под малым углом α , то $h_0 \approx 2\lambda/\alpha^2$ велико, толщина данного фотослоя оказывается много меньше h_0 и голограмма получается тонкой. Именно так записывается голограмма по методу Габора или Лейта–Упатниекса.

Голографические фотослои имеют толщину от нескольких микрометров ($\sim 10^{-4}$ см) до нескольких десятков микрометров ($\sim 10^{-3}$ см) (длина волны зеленого цвета $\lambda \approx 0,5$ мкм). В схеме Лейта–Упатниекса углы схождения волн порядка и меньше 0,1 рад, поэтому параметр h_0 имеет величину порядка и более 100 мкм, т. е. $h < h_0$ — условие (8.117) не выполняется. Как уменьшить параметр h_0 ? Кардинальный способ — осветить слой фотоэмульсии с разных сторон. Тогда угол α приближается к π , а величина h_0 — к $\lambda/2$. При этом на той же толщине фотоэмульсии укладывается много зеркальных слоев, возникающих после обработки. То есть восстанавливающая волна, бегущая на стадии реконструкции (восстановления) изображения вдоль оси z , испытывает много отражений от этих слоев, образующих объемную решетку. Схема получения объемной голограммы, предложенная советским ученым Ю. Н. Денисюком (р. 1927), показана на рис. 8.63. Излучение направляется на фотопластинку, за которой расположен голографируемый объект. Свет свобод-

но проходит сквозь прозрачную фотоэмульсию, отражается от предмета и вновь проходит через слой фотоэмульсии. Прямой лазерный пучок, проходящий через фотоэмульсию слева направо — это опорная волна, а отраженный от предмета и идущий ему навстречу — это предметная волна. Схема Денисюка так и называется — *схемой во встречных пучках*. Углы схождения волн в этой схеме близки к π , что и обеспечивает выполнение условия (8.117).

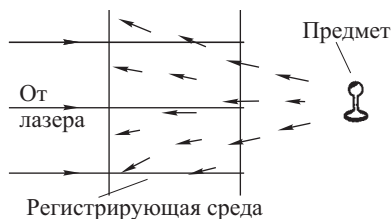


Рис. 8.63

Рассмотрим подробнее процесс восстановления изображения объемной голограммой, вер-

нувшись к простому примеру, когда предметной волной является плоская волна, составляющая угол α с осью z , а опорной — плоская волна, бегущая вдоль оси z . Итак, пусть восстанавливающая волна — плоская волна, бегущая вдоль оси z , падает на голограмму, которая (напомним) представляет собой объемную решетку, состоящую из большого числа частично отражающих зеркальных слоев, составляющих угол β с осью z (рис. 8.64). Восстанавливающая волна частично отражается от зеркального слоя, а частично проходит сквозь него, встречает следующий слой и снова частично отражаясь, проходит далее и т. д. Так возникает большое количество отраженных волн $1, 2, 3, \dots$ и т. д. (три зеркальных слоя AA', BB', CC' показаны на рис. 8.64). Обратим внимание на то, что отражается восстанавливающая волна в направлении, точно совпадающим с направлением предметной волны, которая падала на голограмму в процессе записи (ее интерференция с опорной волной и образовала эту систему зеркальных слоев), поскольку угол падения на зеркало равен углу отражения. Существенно при этом, что волны $1, 2, 3, \dots$, отраженные разными «зеркалами», интерферируют. Каков результат интерференции?

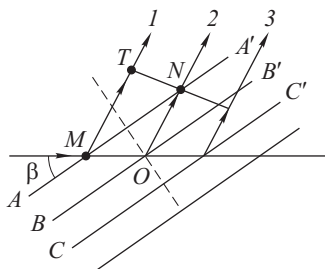


Рис. 8.64

Найдем разность хода волн, отраженных от двух соседних зеркал AA' и BB' (расстояние d между слоями — период объемной решетки — определяется формулой (8.116)). Волна 2, отраженная от зеркала BB' , проходит оптический путь $MO + ON$, в то время как волна 1, отраженная от зеркала AA' , проходит путь MT . Если точки N и T лежат в одной плоскости, перпендикулярной направлениям отраженных волн

(т. е. на одной волновой поверхности), то это означает, что колебания в плоскости NT' , созданные волнами 1 и 2, синфазны, а значит разность оптических путей $(MO + NT) - MT$ равна длине волны λ' , используемой при восстановлении изображения: $(MO + NT) - MT = \lambda'$. Легко убедиться, что последнее условие можно записать в виде

$$2d \sin \beta = \lambda'. \quad (8.118)$$

Это — условие интерференционного отражения волны объемной решеткой с периодом d , где β — угол скольжения (угол между падающей волной и плоскостями решетки). Оно называется *условием Брэгга-Вульфа* (по имени английского физика Л. Брэгга (1890–1971) и российского физика Г.В. Вульфа (1863–1925)). Если длина восстанавливающей волны λ' такова, что выполняется условие Брэгга-Вульфа, то волны, отраженные от разных зеркальных слоев решетки, оказываются синфазными — возникает сильное отражение — восстанавливается предметная волна. Вспомнив, что период d решетки определяется формулой (8.116), а $\beta = \alpha/2$, находим: сильное интерференционное отражение возникает лишь на длине волны λ' , совпадающей с длиной волны λ , используемой при записи голограммы. Мы видим, что объемная голограмма обладает *спектральной избирательностью* — она «запоминает» цвет — длину волны света, с помощью которого записывается голограмма. Если восстанавливать голограмму белым светом, то объемная решетка «выбирает» из белого света лишь ту спектральную компоненту, которая использована при записи. Это свойство объемной голограммы используется для записи голограмм художественных произведений: при этом добиваются великолепной цветопередачи при восстановлении изображения белым светом. Отметим еще одно важное свойство голограммы: при большом числе отражающих зеркальных слоев, составляющих объемную решетку, интенсивность восстановленной волны оказывается близкой к интенсивности волны, падающей на голограмму при реконструкции изображения; другими словами, практически весь световой поток преобразуется объемной голограммой в волну, образующую единственное мнимое изображение.

8.16. Чем отличается голографическое изображение от фотографии

При фотографировании каждая точка предмета рассеивает сферическую волну, которая с помощью объектива фотоаппарата фокусируется в небольшое пятнышко на фотопластине. Так как отражательная способность разных точек объекта различна, то различна и интенсивность света, падающего на разные участ-

ки фотопластинки. Почернение разных участков фотопластинки после фотообработки (которое пропорционально интенсивности), повторяет распределение яркости на предмете — мы получаем негативное изображение.

Отметим три особенности фотографического изображения. Во-первых, на фотопластинке фиксируется лишь распределение интенсивности и, следовательно, фотоизображение содержит неполную информацию о предмете (нет распределения фаз и в частности, нет рельефа предмета). Во-вторых, на каждом небольшом участке фотопластинки получается изображение лишь небольшого («сопряженного») участка предмета. В-третьих, трехмерные объекты регистрируются в виде плоских двумерных изображений.

Голограмма, как мы видели, сохраняет полную информацию об объекте: возникающий на ней интерференционный узор зависит как от амплитуды, так и от фазы предметной волны.

Наблюдатель, который при восстановлении смотрит сквозь голограмму (как в окно), видит как бы реальный предмет; так, изменяя положение, наблюдатель может увидеть те детали объекта, которые не были видны (затенены другими деталями) ранее — голографическое изображение трехмерно. Важнейшее свойство голограммы состоит в том, что любой ее малый участок содержит информацию обо всем объекте: ведь поле в каждой точке голограммы является суперпозицией полей, посылаемых всеми точками предмета (и опорной волной).

Другими словами, интерференционная картина на каждом небольшом участке голограммы содержит информацию об амплитуде и фазе колебаний всех точек предмета, поэтому изображение может быть восстановлено с помощью небольшого осколка голограммы, полученной при записи. Разумеется, разрешающая способность определяется размером осколка.

Вопросы и задачи

1. Найти спектр плоских волн за синусоидальной решеткой с амплитудным коэффициентом пропускания $t(x) = 1 + a \cos \Omega x$ ($a < 1$), освещаемой нормально падающей плоской волной.

2. Плоский монохроматический пучок света с длиной волны λ дифрагирует на двух последовательно расположенных синусоидальных решетках с амплитудным коэффициентом пропускания $t_1(x) = t_2(x) = 1/2(1 + \cos \Omega x)$. При смещении одной из решеток вдоль оси z со скоростью v интенсивность нулевого дифракционного максимума периодически изменяется. Определить частоту ω этих изменений, а также отношение максимальной и минимальной интенсивностей.

3. Диск из стекла с показателем преломления n закрывает полторы зоны Френеля для точки наблюдения P . При какой толщине диска освещенность в точке P максимальна?

Ответ: $h = (2m + 5/4)/(2(n - 1)) \lambda$ ($m = 0, 1, \dots$).

4. Какова интенсивность волны в центре картины дифракции на круглом непрозрачном экране, если он закрывает 1) одну зону Френеля, 2) две зоны Френеля, 3) 4,5 зоны Френеля?

5. Показать, что если точечный источник света находится на расстоянии a от непрозрачного экрана с отверстием, а точка наблюдения — на расстоянии b от него, то радиус m -й зоны Френеля определяется формулой $\rho_m = \sqrt{m \lambda ab / (a + b)}$.

6. Точечный источник света и точка наблюдения P расположены симметрично на расстоянии $2L$ друг от друга на оси круглого отверстия в непрозрачном экране. Отверстие оставляет открытой одну зону Френеля. Во сколько раз изменится интенсивность света в точке P , если к отверстию без нарушения осевой симметрии приложить тонкую линзу с фокусным расстоянием $f = L$.

7. Оценить, с какого расстояния L можно увидеть отдельно свет от двух фар автомобиля.

8. Коллиматорная щель S , освещаемая источником света, находится в фокусе линзы с фокусным расстоянием $f = 20$ см. Пройдя через линзу, свет падает на дифракционную решетку с числом штрихов $N = 1000$ и периодом $d = 0,001$ см. Какова должна быть ширина коллиматорной щели, чтобы была полностью использована разрешающая способность решетки в окрестности длины волны $\lambda = 500$ нм?

9. С помощью интерферометра Фабри–Перо исследуется выделенный системой фильтров участок спектра шириной $\Delta\lambda = 1,2$ нм. Минимальная разность длин волн соседних спектральных линий $\Delta\lambda = 0,001$ нм. Оценить максимальное значение коэффициента пропускания зеркал, при котором можно исследовать этот участок спектра.

10. Два пучка белого света, полученные от одного точечного источника, сходятся на входной щели регистрирующего спектрального прибора. Разность хода равна $\Delta = 300$ м. Оценить разрешающую способность R спектрального прибора, который может обнаружить интерференцию этих пучков.

Ответ: $R \geq \Delta / \lambda \approx 5 \cdot 10^8$.

11. Линза с фокусным расстоянием $f = 50$ см и диаметром $D = 5$ см освещается параллельным пучком света ($\lambda = 630$ нм). Во сколько раз интенсивность волны I в фокусе линзы больше интенсивности I_0 падающей на линзу волны? Оценить размер пятна в фокальной плоскости линзы.

12. Камера с малым отверстием глубиной $L = 10$ см предназначена для фотографирования удаленных объектов. Оценить диаметр отверстия, при котором она имеет максимальную разрешающую способность, $\lambda = 500$ нм.

13. При аэрофотосъемке местности используется объектив с фокусным расстоянием $f = 10$ см и диаметром $D = 5$ см. Съемка производится на фотопленку, имеющую разрешающую способность $R = 100$ мм⁻¹. Определить, какие детали местности могут быть разрешены на фотографиях, если съемка производится с высоты $h = 10$ км.

14. Оценить расстояние L , с которого можно увидеть невооруженным глазом свет лазера, генерирующего в непрерывном режиме мощность $P = 10$ Вт на частоте $\nu = 4 \cdot 10^{15}$ Гц, если для формирования луча используется параболическое зеркало диаметром $D = 50$ см. Глаз видит источник в зеленой части спектра, если в зрачок (диаметра $d = 5$ мм) попадает $n = 60$ квантов в секунду.

Глава 9

ДИСПЕРСИЯ ВОЛН

9.1. Фазовая и групповая скорость. Формула Рэлея

Приступая к изучению волновых процессов, мы ввели понятие *фазовой скорости волны* — скорости, с которой перемещаются волновые поверхности (поверхности с фиксированным значением фазы колебаний). Это понятие применимо к монохроматической волне с заданной частотой ω , бесконечно протяженной в пространстве и во времени. Реальный волновой процесс всегда ограничен во времени, а это означает, что он представляется суперпозицией волн различных частот ω . При этом различные гармонические составляющие суммарного волнового процесса могут иметь разные фазовые скорости.

Наблюдаемая зависимость фазовой скорости от частоты называется *дисперсией*. Среда, в которой возникает дисперсия, называется *диспергирующей средой*. Зависимость $v = v(\omega)$ называют *законом дисперсии*. Часто закон дисперсии среды задают в виде зависимости волнового числа k от частоты: $k(\omega) = \omega/v(\omega)$ или зависимости показателя преломления от частоты: $n(\omega) = c/v(\omega)$. Для описания процесса распространения волн в диспергирующей среде возникает необходимость ввести новую характеристику, называемую *групповой скоростью*.

Поясним физический смысл этого понятия на следующем примере. Пусть некоторый излучатель является источником сигнала $f(t) = a(t)e^{i\omega_0 t}$ (пусть, например, эта функция описывает изменение электрического поля вблизи излучателя), где ω_0 — частота несущего колебания, $a(t)$ — закон модуляции, который и содержит передаваемое сообщение. Возникшая при этом электромагнитная волна начинает распространяться в диспергирующей среде и через некоторое время достигает приемной антенны, находящейся на расстоянии z от излучателя. Будем считать, что сигнал имеет конечную длительность τ , причем на протяжении большого числа периодов несущего колебания $2\pi/\omega_0$ функция $a(t)$ остается практически неизменной, т. е. характерное время ее изменения τ много больше периода несущей: $\tau \gg 2\pi/\omega_0$.

На рис. 9.1 в показан пример сигнала $f(t)$, который представляет собой произведение медленно меняющейся «огибающей» $a(t)$ длительности τ — функции, равной тождественно нулю

вне интервала времени τ , большим по сравнению с периодом несущего колебания $\tau \gg 2\pi/\omega_0$ (рис. 9.1 а), и быстро осциллирующей гармонической функции $e^{i\omega_0 t}$. Разумеется, на графике

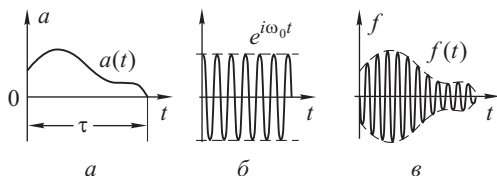


Рис. 9.1

рис. 9.1 б показана $\text{Re } e^{i\omega_0 t}$, т. е. косинусоида. Представим закон модуляции $a(t)$ в виде суммы гармонических колебаний различных частот Ω

$$a(t) = \int A(\Omega) e^{i\Omega t} d\Omega, \quad (9.1)$$

где $A(\Omega)$ — спектр (преобразование Фурье) функции $a(t)$, показанный на рис. 9.2 а.

Поскольку изменения функции $a(t)$ медленные в масштабе периода несущей, ширина ее спектра $A(\Omega)$, равная согласно соотношению неопределенности $\Delta\Omega \approx 2\pi/\tau$, много меньше частоты несущей: $\Delta\Omega \ll \omega_0$. Используя (9.1), можно записать:

$$f(t) = \int A(\Omega) e^{i(\omega_0 + \Omega)t} d\Omega \quad (9.2)$$

или, после замены переменной $\omega = \omega_0 + \Omega$:

$$f(t) = \int A(\omega - \omega_0) e^{i\omega t} d\omega. \quad (9.3)$$

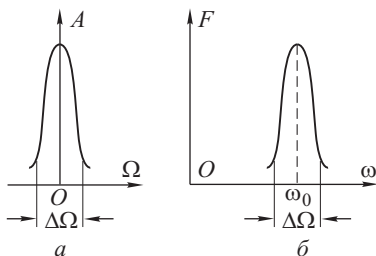


Рис. 9.2

Сравнивая последнее равенство с выражением, связывающим функцию $f(t)$ с ее спектром $C(\omega)$:

$$f(t) = \int C(\omega) e^{i\omega t} d\omega,$$

находим

$$C(\omega) = A(\omega - \omega_0),$$

т. е. спектр сигнала $f(t)$ можно получить из спектра «огибающей» $a(t)$, перенося последний по оси частот на величину ω_0 , как показано на рис. 9.2 б.

Каждой гармонической составляющей «огибающей» $a(t)$, имеющей низкую частоту Ω , соответствует гармоническая составляющая сигнала $f(t)$ высокой частоты $\omega = \omega_0 + \Omega$, причем $\omega_0 \gg \Omega$, поскольку все низкие частоты Ω , составляющие

огibaющую $a(t)$, лежат внутри полосы частот $\Delta\Omega$: $|\Omega| < \Delta\Omega$. Условие $\Delta\Omega \ll \omega_0$ есть условие квазимонохроматичности сигнала $f(t)$.

Каждая гармоническая составляющая сигнала $f(t)$, имеющая согласно (9.2) вид $A(\Omega)e^{i(\omega_0+\Omega)t}$ порождает в среде монохроматический волновой процесс вида

$$A(\Omega)e^{i[(\omega_0+\Omega)t-kz]},$$

где волновое число k соответствует частоте $\omega_0 + \Omega$, т. е. $k = k(\omega_0 + \Omega)$. Мы полагаем для простоты, что возникает плоское волновое возмущение, бегущее вдоль оси z по направлению к приемной антенне.

Суммарный волновой процесс, созданный в среде всеми гармоническими составляющими, частоты которых лежат в пределах узкой полосы частот $\Delta\Omega$ (вблизи несущей частоты ω_0), есть

$$f(t, z) = \int A(\Omega)e^{i[(\omega_0+\Omega)t-k(\omega_0+\Omega)z]} d\Omega. \quad (9.4)$$

Такую группу волн принято называть *волновым пакетом*. Что происходит с волновым пакетом по мере распространения в диспергирующей среде, через какое время этот пакет достигнет приемной антенны, каковы колебания поля вблизи приемной антенны?

Пусть закон дисперсии среды задан в виде $k = k(\omega)$ (рис. 9.3). Рассмотрим простейшую ситуацию. Пусть для всех частот ω , лежащих внутри спектральной полосы сигнала

$|\omega - \omega_0| \lesssim \Delta\Omega$, кривая дисперсии мало отклоняется от касательной, проведенной в точке ω_0 , т. е. дисперсионную зависимость на этом интервале частот можно заменить приближенно линейной зависимостью

$$k(\omega) = k(\omega_0 + \Omega) \approx k(\omega_0) + \left(\frac{dk}{d\omega}\right)_{\omega_0} \Omega. \quad (9.5)$$

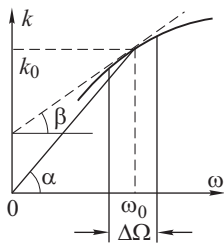


Рис. 9.3

На графике рис. 9.3 производная $(dk/d\omega)_{\omega_0}$ определяет тангенс угла наклона касательной: $\operatorname{tg} \beta = (dk/d\omega)_{\omega_0}$. Введем обозначение $k(\omega_0) = k_0$, т. е. k_0 — волновое число, соответствующее монохроматической волне частоты ω_0 . Напомним, что фазовая скорость этой волны есть $v = \omega_0/k_0$.

Используя (9.5), запишем показатель экспоненты в (9.4) в виде

$$(\omega_0 + \Omega)t - k(\omega_0 + \Omega)z \approx (\omega_0 t - k_0 z) + \Omega \left(t - z \frac{dk}{d\omega} \right).$$

Далее, экспоненциальный множитель $e^{i(\omega_0 t - k_0 z)}$, не зависящий от переменной Ω , можно вынести из под знака интеграла.

Мы получаем

$$f(t, z) = \left\{ \int A(\Omega) \exp \left[i\Omega \left(t - z \frac{dk}{d\omega} \right) \right] d\Omega \right\} e^{i(\omega_0 t - k_0 z)}.$$

Запишем окончательно, учитывая соотношение (9.1):

$$f(t, z) = a \left(t - z \frac{dk}{d\omega} \right) e^{i(\omega_0 t - k_0 z)}.$$

Полученный нами результат означает следующее: волновой пакет, форма которого задается «огibaющей» $a(t)$, распространяется в среде без искажения и на расстоянии z от излучателя воспроизводит закон модуляции $a(t)$ с запаздыванием по времени Δt , равным

$$\Delta t = z \frac{dk}{d\omega}.$$

Из последнего равенства находим скорость волнового пакета

$$u = \frac{z}{\Delta t} = \frac{d\omega}{dk}.$$

Иллюстрацией является рис. 9.4, на котором показано положение волнового пакета вблизи излучателя (к моменту времени, когда излучение закончилось) и его положение через время Δt : волновой пакет сместился на расстояние $z = u\Delta t$ почти без изменения своей формы. Скорость движения волнового пакета как целого называется *групповой скоростью*.

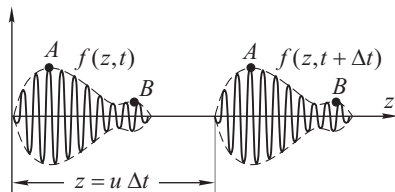


Рис. 9.4

Заметим, что угол наклона касательной β на рис. 9.3 $\operatorname{ctg} \beta = 1/(dk/d\omega) = d\omega/dk = u$ отличается от угла α , определяющего фазовую скорость монохроматической волны частоты ω_0 : $\operatorname{ctg} \alpha = \omega_0/k_0 = v$, т. е. групповая скорость — скорость, с которой переносится сигнал, отличается, вообще говоря, от фазовой скорости «несущей волны» $e^{i(\omega_0 t - k_0 z)}$. Поскольку волновой пакет не искажается (не «деформируется») в процессе движения, для определения групповой скорости можно следить за движением любой точки волнового пакета, например, за движением максимума огибающей (точкой A на рис. 9.4), т. е. максимума амплитуды, либо точкой B — локального максимума амплитуды, либо любой другой точкой волнового пакета.

Согласно современным представлениям универсальная физическая константа c , равная скорости света в вакууме, определяет предельную скорость распространения любых сигналов. В частности, *групповая скорость* любого пакета, какова бы ни была его форма, в какой бы среде он ни распространялся, какова

бы ни была физическая природа волны, не может превосходить скорость света в вакууме. Что касается *фазовой скорости*, то она может быть любой, как больше, так и меньше скорости света в вакууме, поскольку скорость перемещения волновой поверхности идеальной, бесконечно длящейся синусоидальной волны не связана с переносом какой-либо энергии или передачей информации.

Найдем связь между фазовой и групповой скоростью. Пусть закон дисперсии задан в виде зависимости фазовой скорости от длины волны $v = v(\lambda)$. Используя выражение для фазовой скорости $v = \omega/k$, запишем $\omega = vk$. Тогда находим $u = \frac{d\omega}{dk} = \frac{d(vk)}{dk} = v + k \frac{dv}{dk}$.

Далее, поскольку $\lambda = 2\pi/k$ и, следовательно $k/dk = -\lambda/d\lambda$, имеем

$$u = v - \lambda \frac{dv}{d\lambda}. \quad (9.6)$$

Формула (9.6), связывающая фазовую и групповую скорость, называется *формулой Рэлея*.

Если $\frac{dv}{d\lambda} > 0$ (фазовая скорость растет с ростом длины волны), то $u < v$ — групповая скорость меньше фазовой: огибающая $a(t - z/u)$ отстает при движении волнового пакета от несущей гармонической волны $e^{i(\omega_0 t - k_0 z)}$ (например, максимум амплитуды движется медленнее, чем волновые поверхности). Такого рода дисперсия называется *нормальной*. В обратном случае $\frac{dv}{d\lambda} < 0$ имеет место *аномальная* дисперсия.

Если фазовая скорость всех волн, составляющих волновой пакет, одинакова ($v = \text{const}$, независимая от λ), то $\frac{dv}{d\lambda} = 0$ и дисперсия отсутствует — групповая скорость равна фазовой.

Часто свойства оптически прозрачной среды задаются зависимостью показателя преломления среды от частоты $n = n(\omega)$. Напомним, что согласно определению $n = c/v$ и следовательно $dv/v = -dn/n$. Кроме того, из соотношения $\omega\lambda = 2\pi c/n$ имеем $d\lambda/\lambda + d\omega/\omega = -dn/n$. Используя последние соотношения, можно получить из формулы Рэлея:

$$u = \frac{c}{n + \omega \frac{dn}{d\omega}}. \quad (9.7)$$

Поскольку фазовая скорость $v = c/n$, то из (9.7) следует, что при $\frac{dn}{d\omega} > 0$ $u < v$ — имеет место нормальный закон дисперсии, а при $\frac{dn}{d\omega} < 0$, $u > v$ — аномальная дисперсия.

Выше мы рассмотрели ситуацию, когда можно использовать линейный закон дисперсии (9.5). Именно в этом случае волновой

пакет распространяется в среде не деформируясь и можно ввести понятие *групповой скорости*. Если же форма волнового пакета изменяется по мере перемещения (например так, как показано на рис. 9.5, то становится неясным, с движением какой точки волнового пакета следует отождествлять групповую скорость. Однако и в этих случаях нередко оказывается удобным формально использовать величину $\frac{\partial \omega}{\partial k}$.

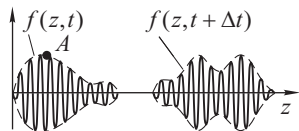


Рис. 9.5

Чтобы выяснить условия, при которых деформацией волнового пакета при его распространении можно пренебречь, воспользуемся разложением функции $k(\omega_0 + \Omega)$ в степенной ряд, учтя следующий после линейного член разложения. Мы имеем

$$k(\omega + \Omega) \approx k_0 + \left(\frac{dk}{d\omega}\right)_{\omega_0} \Omega + \frac{1}{2} \left(\frac{d^2k}{d\omega^2}\right)_{\omega_0} \Omega^2.$$

Поправка в фазе колебания, обусловленная последним слагаемым, есть $\frac{1}{2} \left(\frac{d^2k}{d\omega^2}\right)_{\omega_0} \Omega^2 z$.

Если эта поправка мала ($\ll \pi$) для всех спектральных компонент, составляющих волновой пакет, т. е. для всех $|\Omega| \lesssim \Delta\Omega$, то ею можно пренебречь; тогда остается справедливой линейная зависимость (9.5) — волновой пакет, не деформируясь, движется с групповой скоростью $u = \left(\frac{d\omega}{dk}\right)_{\omega_0}$. Итак, достаточное условие справедливости (9.5) имеет вид

$$\frac{1}{2} \left(\frac{d^2k}{d\omega^2}\right)_{\omega_0} (\Delta\Omega)^2 z \ll \pi.$$

Последнее неравенство налагает ограничения на ширину спектра пакета $\Delta\Omega$, дистанцию z , которую проходит пакет в среде, и закон дисперсии $k(\omega)$ (именно величина второй производной определяет, как быстро кривая $k(\omega)$ отклоняется от касательной). Заметим, что, если справедлива линейная связь между k и ω (9.5), которую можно записать в виде $k = a + b\omega$, то имеет место также линейная связь между v и λ : $v = a' + b'\lambda$. В этом легко убедиться, используя связи: $k = 2\pi/\lambda$ и $\omega = 2\pi v/\lambda$.

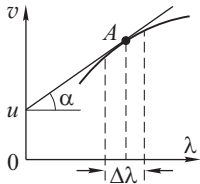


Рис. 9.6

Если закон дисперсии задан в виде $v = v(\lambda)$ (рис. 9.6), то лишь в области длин волн $\Delta\lambda$ вблизи длины λ_0 (соответствующей частоте ω_0), где касательная мало отклоняется от кривой

$v(\lambda)$, дисперсия не приводит к деформации волнового пакета при распространении в среде. Из рис. 9.6 ясно, что точка пересечения касательной с осью ординат как раз и определяет величину групповой скорости $u = v - \lambda \operatorname{tg} \alpha = v - \lambda \frac{dv}{d\lambda}$.

9.2. Дисперсия электромагнитных волн

Классическая теория, созданная Лоренцем, основана на идеализированной модели среды, в которой распространяется электромагнитная волна. Согласно этой модели среда представляет собой набор невзаимодействующих между собой атомов (наилучшим приближением является газообразная среда). В простейшем случае предполагается, что в атоме имеется один «оптический» электрон, квазиупруго связанный с ядром: сила, действующая на этот электрон со стороны поля ядра (экранированного частично электронами внутренних оболочек) линейно связана со смещением \mathbf{r} из положения равновесия: $\mathbf{f}_{\text{упр}} = -k\mathbf{r}$. Таким образом, электрон в атоме рассматривается как линейный осциллятор (см. § 2.1). Вынужденные колебания осциллятора обусловлены полем электромагнитной волны: внешняя сила, действующая на электрон, есть $eE(t)$, где e — заряд электрона, $\mathbf{E}(t) = \mathbf{E}_0 e^{i\omega t}$ — электрическое поле монохроматической волны частоты ω , бегущей в среде.

Колеблющийся электрон сам становится источником электромагнитной волны, благодаря чему его свободные колебания затухают; это обстоятельство учитывается феноменологически введением «силы трения», пропорциональной скорости и равной $\beta \dot{\mathbf{r}}$ (так называемое радиационное затухание, см. § 6.5).

Уравнение движения электрона имеет вид

$$m\ddot{\mathbf{r}} = -k\mathbf{r} - \beta \dot{\mathbf{r}} + e\mathbf{E}_0 e^{i\omega t}$$

или, вводя как обычно константы $\omega_0^2 = k/m$ и $\gamma = \beta/2m$:

$$\ddot{\mathbf{r}} + 2\gamma \dot{\mathbf{r}} + \omega_0^2 \mathbf{r} = \frac{e}{m} \mathbf{E}_0 e^{i\omega t}.$$

Уравнение вынужденных колебаний гармонического осциллятора имеет хорошо известное решение:

$$\mathbf{r} = \frac{e/m}{\omega_0^2 - \omega^2 + 2i\omega\gamma} \mathbf{E}_0 e^{i\omega t}. \quad (9.8)$$

В этом решении опущены две константы интегрирования \mathbf{v}_0 и \mathbf{r}_0 . Тем самым подразумевается, что в среде нет ни макроскопического заряда, ни макроскопических токов, которые были бы обусловлены положениями и скоростями элементарных зарядов в среде.

Смещение $\mathbf{r}(t)$ из положения равновесия приводит к появлению дипольного момента атома $\mathbf{p} = e\mathbf{r}$, который, как ясно из (9.8), изменяется по гармоническому закону, пропорционально меняющемуся по гармоническому закону электрическому полю волны. Важно обратить внимание на то, что коэффициент пропорциональности, связывающий \mathbf{p} и \mathbf{E} (и называемый *атомной поляризуемостью*)

$$\mathbf{p} = \alpha \mathbf{E}, \quad \alpha = \frac{e^2/m}{\omega_0^2 - \omega^2 + 2i\gamma\omega}, \quad (9.9)$$

является, в общем случае, комплексным, а это означает, что колебания \mathbf{p} происходят со сдвигом фазы по отношению к колебаниям \mathbf{E} , причем фазовый сдвиг зависит от частоты. Колеблющиеся диполи вещества сами становятся источниками излучения. Это переизлучение происходит с той же частотой ω . Переизлученные колеблющимися диполями волны, складываясь (интерferируя) с первичной волной, образуют волну, распространяющуюся в среде, причем отставание по фазе вторичных волн по отношению к первичной волне (зависящее от частоты), приводит к тому, что фазовая скорость результирующей волны также становится зависящей от частоты — в этом суть явления дисперсии.

Рассматриваемая нами среда достаточно разрежена, она состоит, как было отмечено, из невзаимодействующих, т. е. достаточно далеко отстоящих друг от друга, атомов. Тем не менее, в пределах элемента объема, малого по сравнению с длиной волны, в котором поле $\mathbf{E}(t)$ можно считать однородным (одинаковым в каждый момент времени во всех его точках), находится достаточно большое число атомов. Например, для видимого света $\lambda \approx 5 \cdot 10^{-7}$ м, и в объеме, равном λ^3 , содержится около $4 \cdot 10^5$ атомов газообразного вещества при нормальных условиях. Итак, мы можем считать, что большое число осцилляторов в пределах малого элемента объема в каждый момент времени имеют один и тот же дипольный момент. Тогда поляризация среды (дипольный момент единицы объема среды) равен $\mathbf{P} = N\mathbf{p}$, где N — число осцилляторов в единице объема.

Используя (9.9), находим

$$\mathbf{P}(t) = \frac{Ne^2/m}{\omega_0^2 - \omega^2 + 2i\omega\gamma} \mathbf{E}_0 e^{i\omega t}. \quad (9.10)$$

Итак, поляризация среды \mathbf{P} колеблется с частотой ω , равной частоте волны, бегущей в среде. Согласно определению *диэлектрическая восприимчивость* среды χ — это коэффициент пропорциональности, связывающий \mathbf{P} и \mathbf{E} : $\mathbf{P} = \epsilon_0 \chi \mathbf{E}$. Мы получаем

$$\chi(\omega) = \frac{Ne^2/(\epsilon_0 m)}{\omega_0^2 - \omega^2 + 2i\gamma\omega}. \quad (9.11)$$

Соответствующее выражение для *диэлектрической проницаемости* $\varepsilon = 1 + \chi$ есть:

$$\varepsilon(\omega) = 1 + \frac{Ne^2/(\varepsilon_0 m)}{\omega_0^2 - \omega^2 + 2i\gamma\omega}. \quad (9.12)$$

Зависимость диэлектрической проницаемости среды от частоты определяет закон дисперсии среды $n(\omega) = \sqrt{\varepsilon(\omega)}$ или $k(\omega) = (\omega/c)n(\omega)$.

Обратим внимание на еще одно важное обстоятельство: поскольку диэлектрическая проницаемость $\varepsilon(\omega)$ является в общем случае комплексной величиной $\varepsilon(\omega) = \varepsilon'(\omega) + i\varepsilon''(\omega)$, то и показатель преломления $n(\omega)$, и следовательно, волновое число — также комплексные величины:

$$n(\omega) = n'(\omega) + in''(\omega); \quad k(\omega) = k'(\omega) + ik''(\omega).$$

Что собой представляет плоская электромагнитная волна при комплексном k ? Подставляя в уравнение плоской волны $f(z, t) = ae^{-i(\omega t - kz)}$ комплексную величину $k = k' + ik''$, находим $f(z, t) = ae^{-k''z}e^{-i(\omega t - k'z)}$. Мы видим, что волна $f(z, t)$ не является плоской волной в обычном понимании, поскольку по мере роста z амплитуда волны $ae^{-k''z}$ экспоненциально убывает (показатель экспоненты $-k''z$ — действительное отрицательное число), волна затухает, причем затухание связано с мнимой частью k'' волнового числа (и, следовательно, с мнимой частью показателя преломления n''). Действительная часть k' и n' определяют фазовую скорость волны $v = \omega/k'$.

Итак, рассмотрим различные ситуации, основываясь на общем соотношении (9.12). Введя обозначение:

$$\frac{Ne^2}{\varepsilon_0 m} = \omega_p^2 \quad (9.13)$$

(величина ω_p называется *плазменной частотой*), запишем:

$$\varepsilon(\omega) = 1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2 + 2i\gamma\omega}. \quad (9.14)$$

Нормальная дисперсия. Сравним величины $\omega_0^2 - \omega^2$ и $2\gamma\omega$, входящие в знаменатель второго слагаемого в (9.14). Пусть $\Delta\omega = \omega_0 - \omega$, тогда $\omega_0^2 - \omega^2 = (\omega_0 + \omega)\Delta\omega$. Величина γ , как известно, определяет ширину резонансной кривой при вынужденных колебаниях гармонического осциллятора. Пусть частота волны далека от резонансной частоты осциллятора ω_0 , тогда $\Delta\omega \gg \gamma$ и следовательно $(\omega_0 + \omega)\Delta\omega \gg 2\gamma\omega$, т. е. мнимое слагаемое в знаменателе (9.9) и (9.14) мало по сравнению с действительным и им можно пренебречь. Тогда

$$\alpha = \frac{e^2/m}{\omega_0^2 - \omega^2}, \quad \varepsilon(\omega) = 1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2}, \quad (9.15)$$

т. е. диэлектрическая проницаемость — действительное число. Действительным является также и волновое число: $k'' \approx 0$, а это означает отсутствие затухания — среда оказывается прозрачной для электромагнитной волны. Физически этот результат понятен: при частотах внешней силы, далеких от резонансной частоты осцилляторов, амплитуда вынужденных колебаний мала; волна, распространяясь в среде, практически не отдает энергию осцилляторам, т. е. не поглощается.

Собственные частоты «оптических» электронов многих веществ лежат в далекой ультрафиолетовой области спектра, поэтому для волн видимой области $\omega \ll \omega_0$ и условие $\Delta\omega \gg \gamma$ выполняется.

Кроме того, если концентрация осцилляторов достаточно мала (что имеет место в газах), то $\omega_p^2 \ll |\omega_0^2 - \omega^2|$. В этом случае имеем

$$n(\omega) = \sqrt{1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2}} \approx 1 + \frac{1}{2} \frac{\omega_p^2}{\omega_0^2 - \omega^2}. \quad (9.16)$$

Зависимость (9.16) показана на рис. 9.7 (штриховой линией обозначена область графика $n(\omega)$, в котором перестает выполняться неравенство $\Delta\omega \gg \gamma$ и, следовательно, указанная зависимость уже не справедлива).

Эмпирическая формула

$$n^2(\lambda) = 1 + A \left(1 + \frac{B^2}{\lambda^2} \right), \quad (9.17)$$

где A и B — константы, была известна задолго до создания электронной теории дисперсии. Она может быть получена из (9.16), причем константы A и B связаны с характеристиками среды равенствами

$$A = \left(\frac{\omega_p}{\omega_0} \right)^2, \quad B = \frac{4\pi^2 c^2}{\omega_0^2}. \quad (9.18)$$

Исследовав экспериментально зависимость $n(\lambda)$, можно определить константы A и B и с помощью равенств (9.18) найти величины ω_0 и ω_p .

В области частот, где формула (9.16) справедлива (сплошная кривая), имеет место нормальная дисперсия: n растет с ростом частоты. Левая ветвь соответствует частотам, меньшим ω_0 (в частности, видимый свет). Для этих частот показатель преломления больше единицы и, следовательно, фазовая скорость меньше скорости света в вакууме $v = c/n < c$. Для частот $\omega > \omega_0$ (например, рентгеновский диапазон спектра) $n < 1$ и $v > c$, что, как мы

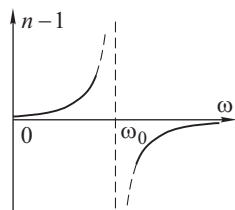


Рис. 9.7

отмечали ранее, не противоречит специальной теории относительности, согласно которой скорость любых тел или сигналов ограничена величиной c , что не имеет отношения к фазовой скорости монохроматической волны.

Если выполнено сильное неравенство $\omega \gg \omega_0$ (и тем более $\omega \gg \gamma_0$), то формула (9.15) дает:

$$\varepsilon(\omega) = n^2(\omega) \approx 1 - \frac{\omega_p^2}{\omega^2}. \quad (9.19)$$

Эта формула справедлива приближенно как для волн рентгеновского диапазона, так и для радиоволн в ионосфере, которая представляет собой плазму, т.е. полностью ионизованный газ, дисперсия волн в которой обусловлена свободными электронами. В последнем случае формула (9.19) справедлива, поскольку на свободные электроны не действует квазиупругая сила, связывающая электроны в атоме, поэтому $\omega_0 = 0$ и сильное неравенство $\omega \gg \omega_0$ выполняется даже для низкочастотных радиоволн.

Соотношение (9.19) позволяет объяснить особенности характера распространения радиоволн в ионосфере.

При $\omega < \omega_p$ (что соответствует длинам волн $\lambda \gtrsim 10$ м) показатель преломления

$$n = \sqrt{1 - \frac{\omega_p^2}{\omega^2}} \quad (9.20)$$

становится чисто мнимым, амплитуда волны экспоненциально спадает и волна не проникает вглубь ионосферного слоя; при этом коэффициент отражения (определяемый, согласно формулам Френеля, действительной частью n), равен единице. Напомним, что, в частности, при нормальном падении $R = ((n' - 1)/(n' + 1))^2 = 1$, поскольку $n' = 0$.

Таким образом, полное внутреннее отражение происходит при любом угле падения.

Отражение длинных радиоволн от ионосферного слоя используется для дальней радиосвязи (рис. 9.8), когда передатчик A

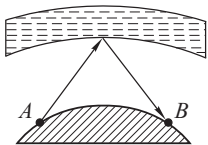


Рис. 9.8

и приемник B находятся вне прямой видимости. Наоборот, ультракороткие радиоволны ($\lambda < 10$ м) свободно проходят сквозь ионосферу: при $\omega > \omega_p$ показатель преломления — действительное число, $n'' = 0$ и волна проходит ионосферный слой без затухания. Такие волны используются для связи со спутниками, локации планет и Луны.

В случае если имеет место равенство (9.20): $n = c/v = \sqrt{1 - \omega_p^2/\omega^2}$, легко получить простую связь между фазовой и групповой скоростью. Возводя обе части последнего равенства в

квадрат и используя $v = \omega/k$, находим $c^2 k^2 = \omega^2 - \omega_p^2$, откуда:
 $2c^2 k dk = 2\omega d\omega$. Мы получаем

$$\frac{\omega}{k} \frac{d\omega}{dk} = c^2.$$

Таким образом,

$$vu = c^2. \quad (9.21)$$

Последнюю связь удобно использовать вместо формулы Рэлея, если речь идет о законе дисперсии (9.20).

Закономерности поведения *электромагнитных волн в ионосфере* во многом подобны их поведению *в металлах*. И в том, и в другом случае дисперсия обусловлена взаимодействием волны со свободными электронами, на которые не действует квазиупругая сила: $\omega_0 = 0$, поэтому справедлива формула

$$\varepsilon(\omega) = n^2(\omega) = 1 - \frac{\omega_p^2}{\omega^2 - i \cdot 2\gamma\omega}, \quad (9.22)$$

которая получается из общего соотношения (9.14) при $\omega_0 = 0$. При низких частотах, таких, что $\omega \ll \gamma$ и, следовательно, $\omega^2 \ll 2\gamma\omega$ получаем из (9.22)

$$n^2(\omega) = 1 - i \frac{\omega_p^2}{2\gamma\omega}.$$

Концентрация свободных электронов в металлах такова, что справедливо неравенство $\omega_p \gg \gamma$, поэтому $\omega_p^2/(2\gamma\omega) \gg \omega_p^2/(2\gamma^2) \gg 1$, т. е. мнимое слагаемое в (9.22) много больше по абсолютной величине первого (действительного) слагаемого. Мы получаем

$$n(\omega) \approx \sqrt{\frac{i \omega_p^2}{2\gamma\omega}} = \exp\left(-i \frac{\pi}{4}\right) \frac{\omega_p}{\sqrt{2\gamma\omega}}. \quad (9.23)$$

Поскольку $\exp(i \pi/4) = \cos(\pi/4) + i \sin(\pi/4) = (1 + i)/\sqrt{2}$, то из (9.23) следует, что показатель преломления является комплексным числом с равными действительной и мнимой частью

$$n' = n'' = \frac{\omega_p}{2\sqrt{\gamma\omega}}.$$

Из неравенства $n'' \gg 1$ следует, что амплитуда волны быстро уменьшается по мере проникновения в металл (глубина проникновения $z \approx 1/k''$ много меньше длины волны), при этом коэффициент отражения, определяемый действительной частью показателя преломления, близок к единице: $R = ((n' - 1)/(n' + 1))^2 = 1$ при $n' \gg 1$ (при нормальном падении) — волна практически полностью отражается.

В случае высоких частот $\omega \gg \gamma$ приходим к формуле (9.20), из которой можно сделать те же выводы, которые относились

ранее к дисперсии радиоволн в ионосфере. В частности, при $\omega > \omega_p$ показатель преломления становится действительным, а металл прозрачен для волны (обычно условие $\omega > \omega_p$ начинается выполняться для металлов в рентгеновской области частот). В оптическом диапазоне частот ситуация не столь проста, о чем говорит хотя бы эффект избирательного отражения — желтый цвет золота и меди. Это значит, что помимо электродинамики, здесь важна атомная физика.

Аномальная дисперсия. Выше мы рассматривали дисперсию в области частот, далеких от собственной частоты осцилляторов: $|\omega - \omega_0| \gg \gamma$. В этом случае диэлектрическая проницаемость $\varepsilon(\omega)$ является действительной функцией, среда оказывается прозрачной для волны (отсутствует поглощение) и показатель преломления растет с ростом частоты, т. е. имеет место нормальная дисперсия.

Обратимся теперь к случаю, когда частота волны близка к собственной частоте атомных осцилляторов. Одну важную особенность мы уже отмечали: если $|\omega - \omega_0| \approx \gamma$, то функция $\varepsilon(\omega)$ оказывается комплексной. Поглощение волны средой в этом случае обусловлено передачей энергии волны колеблющимся осцилляторам, амплитуда вынужденных колебаний которых становится особенно большой, когда частота силы, раскачивающей колебания (т. е. частота колебания электрического поля волны), близка к резонансной.

Детальное исследование общего соотношения (9.14) показывает, что в области, где волна испытывает сильное поглощение, характер дисперсии является аномальным: в области частот $|\omega - \omega_0| \lesssim \gamma$ показатель преломления уменьшается с ростом частоты.

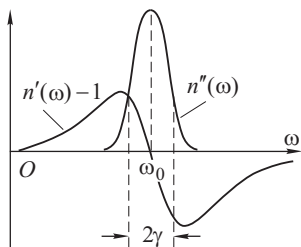


Рис. 9.9

Графики зависимости действительной и мнимой частей показателя преломления от частоты, следующие из общей формулы (9.14), показаны на рис. 9.9. Следует обратить внимание, что зависимость $n''(\omega)$ мнимой части показателя преломления от частоты (ответственного за поглощение волны в среде) подобна резонансной кривой атомных осцилляторов: чем больше амплитуда вынужденных колебаний, тем сильнее поглощение волны в среде; и именно в пределах ширины полосы резонансной кривой имеет место аномальный характер дисперсии. В области частот, далеких от резонансной, кривая $n'(\omega)$ на рис. 9.9 совпадает с кривой, показанной на рис. 9.7.

Эффектная демонстрация аномальной дисперсии была предложена Кундтом. Использованный им метод скрещенных призм показан на рис. 9.10.

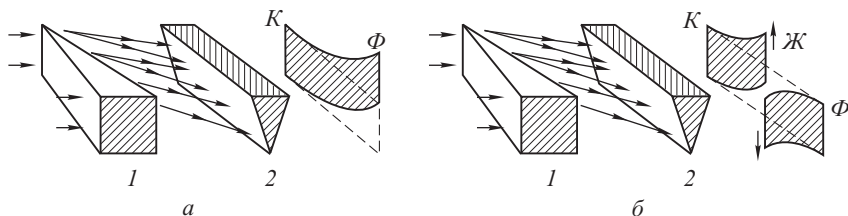


Рис. 9.10

Первая призма (стеклянная) с вертикальным основанием обладает нормальной дисперсией, поэтому разлагает пучок белого света в горизонтальную радужную полоску: от красного до фиолетового цвета (фиолетовая, более коротковолновая волна, преломляется сильнее).

Вторая (скрещенная) призма имеет горизонтальное основание, поэтому смещает полоску по вертикали. Если она также обладает нормальной дисперсией, то фиолетовая (коротковолновая) часть полоски смещается по вертикали сильнее, чем красная (как показано на рис. 9.10а).

На рис. 9.10б вторая призма обладает аномальной дисперсией. В опытах Кундта вторая призма образована парами натрия, обладающими сильным поглощением в желтой части спектра. Именно там, где должна была бы располагаться желтая часть цветной полоски, образуется разрыв (желтая часть спектра поглотилась парами натрия). При этом примыкающая к желтому цвету более длинноволновая часть полоски (для которой $n > 1$) отклоняется вверх (к основанию призмы), а более коротковолновая часть (для нее $n < 1$) отклоняется вниз, т. е. форма полоски воспроизводит кривую дисперсии $n(\lambda)$.

Вопросы и задачи

1. Вычислить групповую скорость u для различных законов дисперсии (v — фазовая скорость):

- $v = v_0 = \text{const}$ (недиспергирующая среда);
- $v = a\sqrt{\lambda}$ (гравитационные волны на поверхности воды);
- $v = a/\sqrt{\lambda}$ (капиллярные волны);
- $v = a/\lambda$ (поперечные колебания натянутой струны);
- $v = \sqrt{c^2 + b^2 + \lambda^2}$ — электромагнитные волны в ионосфере (c — скорость света в вакууме, λ — длина волны в среде).

2. Найти групповую скорость u рентгеновского излучения в среде, если предельный угол полного внутреннего отражения при падении этих волн в среду из воздуха равен α .

Указание: закон дисперсии рентгеновских волн определяется выражением $n(\omega) = 1 - \omega_p^2/\omega^2$, где ω_p — константа.

3. Плоская волна распространяется в среде с законом дисперсии $v = a + b\lambda$, где a и b — постоянные. Показать, что каково бы ни было возмущение, его форма периодически восстанавливается по истечении времени $\tau = d\lambda/dv = 1/b$. Показать, что отношение пути S , пройденного возмущением за время τ , к продолжительности этого промежутка, равно групповой скорости.

Указание: достаточно показать, что по истечении времени τ восстанавливается взаимное расположение любых двух монохроматических компонент суммарной волны.

4. Показатель преломления ионосферы для радиоволн с частотой $\nu = 10$ МГц равен $n = 0,9$. Найти концентрацию N электронов в ионосфере, а также фазовую и групповую скорость для этих радиоволн.

Ответ: $N \approx 2,4 \cdot 10^{11} \text{ м}^{-3}$.

5. Для оценки интегральных и средних характеристик межзвездной плазмы можно использовать экспериментальный факт, установленный сразу же после открытия пульсаров. Оказалось, что из-за дисперсии плазмы импульсы радиоизлучения пульсаров на более низких частотах всегда запаздывают по отношению к импульсам более высоких частот. Определить полное число n свободных электронов на пути сигналов (т. е. их полное число в цилиндре площадью 1 м^2 и высотой, равной расстоянию источник–приемник), если испущенные одновременно импульсы с длинами несущих волн $\lambda_1 = 3 \text{ см}$ и $\lambda_2 = 5 \text{ см}$ запаздывают друг относительно друга на время $\Delta t = 10^{-5} \text{ с}$. Определить также среднюю концентрацию свободных электронов на пути сигналов, если их относительное запаздывание $\Delta t/t_0 = 10^{-15}$. Концентрация электронов на пути сигналов не постоянна, хотя показатель преломления мало отличается от единицы.

Указание: воспользуйтесь формулой, связывающей фазовую и групповую скорость волн в плазме: $uv = c^2$; время распространения импульсов определяется групповой скоростью.

Ответ: $n = \frac{2\pi c}{\lambda_2^2 - \lambda_1^2} \Delta t / r_0 \approx 4 \cdot 10^{17}$, где $r_0 = \frac{e^2}{m_e c^2} \approx 2,8 \cdot 10^{-13} \text{ см}$,
 $N = \frac{n}{ct_0} = \frac{2\pi}{r_0(\lambda_2^2 - \lambda_1^2)} \frac{\Delta t}{t_0} \approx 4 \cdot 10^{-3} \text{ см}^{-3}$.

Глава 10

ВОЛНЫ В АНИЗОТРОПНЫХ СРЕДАХ. ЭЛЕМЕНТЫ КРИСТАЛЛООПТИКИ

10.1. Модель анизотропной среды

Принятая в предыдущей главе модель среды основана на представлении о том, что атомные осцилляторы, заполняющие среду, являются изотропными: ни жесткость квазиупругой связи (и, следовательно, собственная частота осцилляторов), ни коэффициент затухания γ не зависят от ориентации вектора \mathbf{E} в электромагнитной волне, распространяющейся в среде и вызывающей процесс вынужденных колебаний осцилляторов. Согласно (9.9) это означает, что атомная поляризуемость α не зависит от ориентации вектора \mathbf{E} , а направление дипольного момента атома \mathbf{p} совпадает с направлением поля \mathbf{E} : $\mathbf{p} = \alpha \mathbf{E}$. Не зависят от ориентации \mathbf{E} также и диэлектрическая восприимчивость вещества $\chi = \varepsilon_0 N \alpha$, и диэлектрическая проницаемость $\varepsilon = 1 + \chi$ и показатель преломления $n = \sqrt{\varepsilon}$. В каком бы направлении не распространялась волна и как бы не был ориентирован вектор \mathbf{E} этой волны (в плоскости, перпендикулярной направлению распространения), фазовая скорость волны $v = c/n$ одна и та же. Суммарное электрическое поле вторичных волн, переизлучаемых колеблющимися осцилляторами, совпадает по направлению с полем первичной волны, поэтому линейно-поляризованная волна (с фиксированной плоскостью колебаний вектора \mathbf{E}) остается линейно-поляризованной, и в ней сохраняется неизменной плоскость колебаний суммарного вектора \mathbf{E} .

В этой главе мы рассмотрим особенности оптических явлений в *анизотропных средах*. Физическая природа анизотропии вещества, главным образом монокристаллов, связана с тем, что квазиупругая сила, возвращающая оптический электрон в положение равновесия, обусловлена не только сферически симметричным полем ядра, но и экранирующим полем внутренних оболочек, которые не обязательно сферически симметричны; полем соседних атомов и ионов; а равно и электронов, принадлежащих не конкретному атому, а кристаллу как целому.

На рис. 10.1 показана двумерная механическая модель анизотропного осциллятора. Если внешняя сила, вызывающая процесс малых вы-

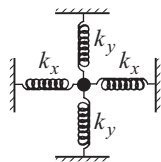


Рис. 10.1

нужденных колебаний, горизонтальна, то квазиупругая сила обусловлена пружинами k_x . При вертикальной внешней силе работают пружины жесткости $k_y \neq k_x$.

Соответственно различными оказываются и собственные частоты осциллятора при горизонтальных и вертикальных колебаниях $\omega_{0x} \neq \omega_{0y}$. Различны, как следует из (9.11), также, поляризуемость $\alpha_x \neq \alpha_y$ и дипольные моменты $p_x \neq p_y$.

При произвольной ориентации вектора \mathbf{E} (с проекциями E_x и E_y) проекции дипольного момента осциллятора есть $p_x = \alpha_x E_x$ и $p_y = \alpha_y E_y$. Ясно, что при $\alpha_x \neq \alpha_y$ вектор \mathbf{p} дипольного момента осциллятора не совпадает по направлению с вектором \mathbf{E} .

Итак, в анизотропной среде дипольные моменты осцилляторов не совпадают, вообще говоря, по направлению с электрическим полем волны, которая индуцирует эти дипольные моменты. Соответственно не совпадают по направлению с вектором \mathbf{E} и вектор поляризации среды $\mathbf{P} = N\mathbf{p}$ и вектор $\mathbf{D} = \varepsilon_0\mathbf{E} + \mathbf{P}$. В отличие от изотропной среды, в которой справедлива векторная связь $\mathbf{D} = \varepsilon_0\varepsilon\mathbf{E}$ и свойства которой характеризуются единственным параметром ε , в анизотропной среде связь между векторами \mathbf{D} и \mathbf{E} имеет вид

$$\begin{aligned} D_x &= (\varepsilon_{11} E_x + \varepsilon_{12} E_y + \varepsilon_{13} E_z) \varepsilon_0, \\ D_y &= (\varepsilon_{21} E_x + \varepsilon_{22} E_y + \varepsilon_{23} E_z) \varepsilon_0, \\ D_z &= (\varepsilon_{31} E_x + \varepsilon_{32} E_y + \varepsilon_{33} E_z) \varepsilon_0, \end{aligned} \quad (10.1)$$

т. е. x -компонента вектора \mathbf{D} зависит в общем случае как от x -компоненты вектора \mathbf{E} , так и от его y - и z -компонент. То же самое касается и других проекций вектора \mathbf{D} . Такая связь между векторами \mathbf{D} и \mathbf{E} называется тензорной, а совокупность девяти чисел

$$\varepsilon_{ij} = \begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} & \varepsilon_{13} \\ \varepsilon_{21} & \varepsilon_{22} & \varepsilon_{23} \\ \varepsilon_{31} & \varepsilon_{32} & \varepsilon_{33} \end{pmatrix}$$

образует тензор диэлектрических проницаемостей. Заметим, что тензором называется не любая матрица, но лишь такая, которая связывает два истинных вектора согласно правилу

$$D_j = \varepsilon_0 \sum_{j=1}^3 \varepsilon_{ij} E_j.$$

Компоненты тензора ε_{ij} зависят от выбора системы координат. Можно выбрать направления осей декартовой системы координат так, что симметричный тензор ε_{ij} становится диагональным:

$$\varepsilon_{ij} = \begin{pmatrix} \varepsilon_{11} & 0 & 0 \\ 0 & \varepsilon_{22} & 0 \\ 0 & 0 & \varepsilon_{33} \end{pmatrix}.$$

В этой системе координат, вместо (10.1), имеем

$$D_x = \varepsilon_0 \varepsilon_x E_x, \quad D_y = \varepsilon_0 \varepsilon_y E_y, \quad D_z = \varepsilon_0 \varepsilon_z E_z. \quad (10.2)$$

Это означает, что существуют три направления в кристалле, такие, что, если вектор \mathbf{E} в волне совпадает с одним из этих направлений, то векторы \mathbf{D} и \mathbf{E} оказываются коллинеарными. Очевидно, что коллинеарны при этом и векторы \mathbf{E} и \mathbf{P} : направление дипольного момента атомных осцилляторов совпадает с направлением электрического поля волны. Направления x , y , z в кристалле, для которых справедливы равенства (10.2), называются *главными направлениями*.

Трем параметрам ε_x , ε_y и ε_z соответствуют три *главных показателя преломления* $n_x = \sqrt{\varepsilon_x}$, $n_y = \sqrt{\varepsilon_y}$ и $n_z = \sqrt{\varepsilon_z}$. Это означает, что волна, в которой вектор \mathbf{E} параллелен оси x , бежит с фазовой скоростью $v_x = c/n_x$; соответственно волны, в которых колебание поля происходит параллельно осям y и z , имеют фазовые скорости $v_y = c/n_y$, $v_z = c/n_z$. Кристаллы, для которых все три главных показателя преломления различны: $n_x \neq n_y \neq n_z$, называются *двуосными*. Мы далее будем говорить об *одноосных* кристаллах, для которых $n_z \neq n_x = n_y$. В этом случае ось z называется *оптической осью* кристалла.

Показатель преломления n_z обозначают обычно через n_e , показатели преломления $n_x = n_y$ обозначают через n_o . Подчеркнем, что оптическая ось — это не фиксированная прямая в кристалле, а направление, т. е. любая прямая, параллельная оси z , является оптической осью кристалла.

Кристаллы, у которых $n_e > n_o$ называются *положительными*. Примером является кварц, для которого $n_e - n_o \approx 0,01$. У *отрицательных* кристаллов $n_e < n_o$. Примером отрицательного кристалла является исландский шпат, обладающий сильной анизотропией: $n_o - n_e \approx 0,2$.

10.2. Волны в одноосных кристаллах

Итак, рассмотрим, как проявляется анизотропия кристалла при распространении в нем световой волны. Пусть x , y , z — главные направления в кристалле, а z — оптическая ось.

Прежде всего, введем терминологию, которая используется для описания волновых явлений в одноосных кристаллах.

Плоскость, в которой лежат луч света, т. е. нормаль к волновой поверхности, и оптическая ось, называется *главной плоскостью*, или *главным сечением* кристалла. Подчеркнем, что эта плоскость лишь одной из прямых, ее образующих, привязана к кристаллической структуре, другую же задает световая волна. Если колебания вектора \mathbf{E} в волне происходят в главной плос-

кости, то такая волна называется *необыкновенной* (смысл этих терминов будет ясен из дальнейшего).

Пусть, например, луч света S_1 лежит в плоскости zy (плоскость рис. 10.2, где z — оптическая ось), тогда плоскость zy — главная плоскость, а волна S_1 — необыкновенная, поскольку колебания \mathbf{E} происходят в главной плоскости (что показано поперечными стрелками на рисунке). В волне S_2 (луч S_2 также лежит в плоскости zy) колебания вектора \mathbf{E} перпендикулярны главной плоскости. Эта волна является обыкновенной. Луч S_3 также, как S_1 и S_2 , лежит в плоскости рисунка — плоскости zy . Эта волна содержит как обыкновенную компоненту (с колебаниями поля, перпендикулярными главной плоскости — плоскости zy), так и необыкновенную компоненту, в которой колебания поля происходят в плоскости рисунка.

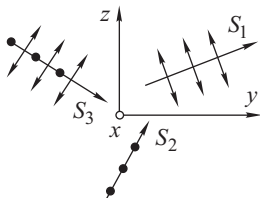


Рис. 10.2

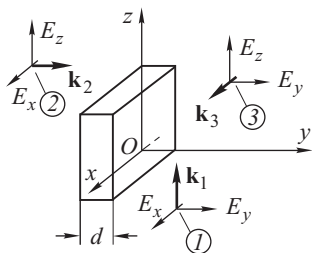


Рис. 10.3

Рассмотрим вначале, что происходит при распространении света вдоль одного из главных направлений. Пусть монохроматическая волна бежит вдоль оптической оси (волна 1 на рис. 10.3). В поперечной электромагнитной волне, бегущей вдоль оси z , вектор \mathbf{E} может иметь только x - и y -проекции E_x и E_y , т. е. волну можно рассматривать как суперпозицию двух линейно-поляризованных волн, в одной из которых вектор \mathbf{E} колеблется вдоль оси x , а во второй — вдоль оси y . Фазовые скорости этих волн одинаковы: $v = c/n_x = c/n_y = c/n_o$. В суммарной волне вектор \mathbf{E} не имеет проекции на оптическую ось (ось z) — она является *обыкновенной* (ее фазовая скорость определяется показателем преломления n_o). Волна, бегущая вдоль оптической оси — всегда обыкновенная. Распространяясь в кристалле, каждая из составляющих волн E_x и E_y приобретает один и тот же фазовый набег $k_x z = k_y z = (\omega/v_o)z$, поэтому разность фаз колебаний E_x и E_y не меняется по мере распространения волны в глубь кристалла. Это означает, что не изменяется состояние поляризации волны. Если, скажем, на входе в кристалл волна была линейно-поляризована (разность фаз колебаний на входе равна нулю или π , см. § 4.3), то и в любом месте внутри

кристалла она остается линейно-поляризованной с неизменной ориентацией плоскости колебаний вектора \mathbf{E} .

Рассмотрим теперь волну 2, бегущую вдоль другого главного направления, например, вдоль оси y . Две компоненты волны E_x и E_z имеют вид

$$E_z = A_1 \cos(\omega t - k_z y), \quad E_x = A_2 \cos(\omega t - k_x y).$$

В первой из этих волн вектор \mathbf{E} колеблется в главной плоскости (параллельно оптической оси z). Эта волна необыкновенная, ее фазовая скорость $v_e = c/n_e$ отличается от фазовой скорости второй, обыкновенной волны $v_o = c/n_o$, соответственно, отличаются волновые числа k_z и k_x : $k_z = (\omega/c)n_e = (2\pi/\lambda)n_e$, $k_x = (2\pi/\lambda)n_o$, где λ — длина волны в вакууме. В соответствии с принятой терминологией, *показатели преломления* n_e и n_o называются соответственно *необыкновенным* и *обыкновенным*. Проникая в кристалл на глубину y , волны приобретают разность фаз $\Delta\varphi = (2\pi/\lambda)(n_e - n_o)y$. Пусть пластинка вырезана из одноосного кристалла параллельно оптической оси (как показано на рис. 10.3), толщина пластинки d . Пучок света падает на нее по нормали, т. е. бежит вдоль оси y . Тогда на выходе из пластинки возникает дополнительная разность фаз, равная

$$\Delta\varphi = \frac{2\pi}{\lambda} (n_e - n_o)d. \quad (10.3)$$

Если толщина d такова, что

$$\Delta\varphi = \frac{\pi}{2} + 2\pi m, \text{ где } m — \text{целое число}, \quad (10.4)$$

то такая пластинка называется пластинкой в $\lambda/4$, или *четверть-волновой пластинкой*.

Пусть на входе в пластинку (при $y = 0$) колебания E_x и E_z синфазны. Так происходит, если на пластинку падает линейно-поляризованная волна, в которой плоскость колебаний лежит в первом и третьем квадрантах (рис. 10.4 а). Тогда на выходе (при $y = d$) разность фаз равна $\pi/2$. Мы получаем эллиптически поляризованную волну, причем главные оси эллипса совпадают с осями x и z . Если, кроме того, плоскость колебаний вектора \mathbf{E} на входе составляет угол 45° с осями x и z , то амплитуды волн E_x и E_z одинаковы и на выходе получаем свет, поляризованный по кругу.

Если

$$\Delta\varphi = \pi + 2\pi m, \quad (10.5)$$

то пластинку называют пластинкой в $\lambda/2$.

Линейно-поляризованная на входе в пластинку волна с синфазно колеблющимися составляющими E_x и E_y останется на выходе линейно-поляризованной, причем колебания E_x и E_y ста-

нут противофазными, а это означает, что положение плоскости колебаний изменится, как показано на рис. 10.4б.

Если разность фазовых набегов обыкновенной E_x и необыкновенной E_z компонент волны кратна целому числу 2π :

$$\Delta\varphi = 2\pi m, \quad (10.6)$$

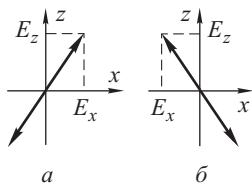


Рис. 10.4

то пластинку называют *пластинкой в λ* . Такая пластинка не изменяет характера поляризации падающей волны.

В общем случае, если толщина d пластинки не удовлетворяет равенствам (10.4)–(10.6) (разность фаз колебаний E_x и E_y на выходе отлична от нуля, π или $\pi/2$), линейно-поляризованная на входе в пластинку волна становится эллиптически-поляризованной, причем главные оси эллипса поляризации не совпадают с главными направлениями пластинки.

Следует подчеркнуть, что разность фазовых набегов, определяемая формулой (10.3), зависит от частоты (длины волны) как благодаря множителю $2\pi/\lambda = \omega/c$, так и вследствие возможной дисперсии, т. е. зависимости $n_e(\omega)$ и $n_o(\omega)$. Поэтому, если мы говорим, что данная пластинка является, например, четвертьволновой, то это относится к волне определенной частоты (или длины волны), а именно той длины волны, для которой $\Delta\varphi = \pi/2 + 2\pi m$.

Пусть в одноосном кристалле, где ось z — оптическая ось кристалла, в начале координат имеется точечный источник света S (рис. 10.5а). Проведем из начала координат во всевозможных направлениях лучи, лежащие в плоскости рисунка. Эта плоскость является главной плоскостью. Пусть вектор \mathbf{E} во всех этих лучах перпендикулярен главной плоскости (так будет например, если излучателем является точечный диполь, ось которого перпендикулярна главной плоскости — плоскости рисунка), т. е. источник излучает обыкновенную волну.

Отложим на каждом луче отрезок, равный фазовой скорости волны. Все эти отрезки одинаковы, поскольку, каким бы ни было направление луча, вектор \mathbf{E} в этом луче не имеет проекции на ось z . В нашем случае он имеет проекцию лишь на другое главное направление — ось x , поэтому фазовая скорость равна $v_o = c/n_o$ для лучей любого направления. Концы всех отрезков находятся на одинаковом расстоянии от источника — на окружности, радиус которой равен v_o . Разумеется, если мы рассмотрим множество лучей, лежащих в любой другой плоскости, повернутой вокруг оси z вместе с излучающим диполем так, что колебания поля \mathbf{E} происходят перпендикулярно этой плоскости, а значит и оптической оси z , то результат не изменится: концы

всех отрезков будут находиться на расстоянии v_o от источника. Итак, множество концов отложенных отрезков лежит на сфере радиуса v_o — эта сфера является волновой поверхностью обыкновенной волны.

Обыкновенные волны в кристаллах ведут себя так же, как волны в изотропных средах, в которых фазовая скорость волны не зависит от направления, поэтому волновые поверхности, исходящие из точечного источника, являются сферами. Отсюда смысл названия — обыкновенная волна.

Рассмотрим теперь лучи, в которых колебания вектора **Е** лежат в главной плоскости (в плоскости рис. 10.5 б), т. е. излучение источника является необыкновенной волной в кристалле. В луче 2, бегущим перпендикулярно оптической оси (вдоль оси y на рис. 10.5 б) вектор **Е** параллелен оптической оси, он не имеет проекции на главные направления y и x : в этом направлении волна бежит со скоростью $v_e = c/n_e$.

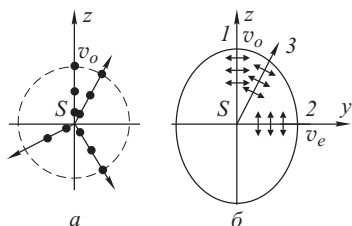


Рис. 10.5

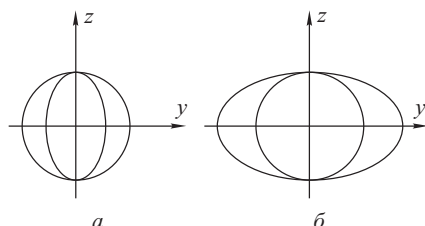


Рис. 10.6

В луче 1, бегущим вдоль оптической оси z , вектор **Е** не имеет проекции на ось z (в нашем случае он имеет отличную от нуля проекцию лишь на другое главное направление — ось y), поэтому в направлении оптической оси необыкновенная волна бежит со скоростью, равной $v_o = c/n_y = c/n_o$.

Если луч бежит в каком-либо промежуточном направлении (луч 3), то реализуется промежуточное значение скорости между значениями v_o и v_e .

Если, как и ранее, отложить на каждом луче отрезки, равные фазовой скорости волны, их концы образуют эллипсоид вращения с полуосями c/n_o и c/n_e . Заметим еще раз, что скорость волны, бегущей вдоль оптической оси z , равна c/n_o , т. е. совпадает со скоростью обыкновенной волны: на оптической оси эллипсоид лучевых скоростей необыкновенной волны касается сферической поверхности лучевых скоростей обыкновенной волны.

На рис. 10.6 показаны поверхности лучевых скоростей обыкновенной и необыкновенной волны для положительного кристалла (для которого $n_e > n_o$, поэтому $v_e < v_o$ и эллипсоид лежит внутри сферы, рис. 10.6 а) и для отрицательного кри-

сталла ($v_e > v_o$ — сфера лежит внутри эллипсоида, рис. 10.6 б). В любом случае точки касания сферы и эллипсоида лежат на оптической оси.

10.3. Преломление на границе анизотропной среды, двойное лучепреломление

Ранее мы рассмотрели, как ведет себя луч света, падающий нормально на кристаллическую пластинку, вырезанную параллельно оптической оси (т. е. оптическая ось параллельна поверхности, на которую падает луч света): направление распространения луча не меняется, а анизотропия проявляется в том, что скорость волны оказывается различной для волн разной

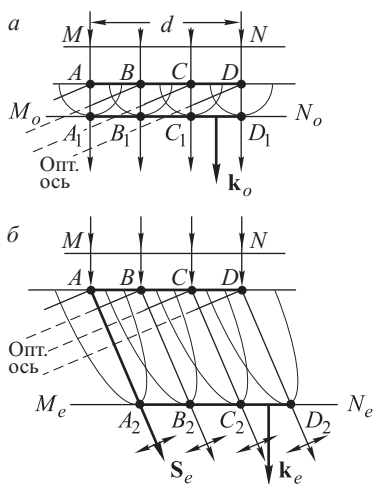


Рис. 10.7

ориентации вектора \mathbf{E} (обыкновенная волна, в которой вектор \mathbf{E} перпендикулярен оптической оси, бежит со скоростью c/n_o , и необыкновенная, в которой вектор \mathbf{E} параллелен оптической оси, бежит со скоростью c/n_e).

Проиллюстрируем особенности возможных более сложных ситуаций на следующем примере. Пусть на поверхность одноосного кристалла падает нормально луч света с поперечным сечением d (рис. 10.7). Оптическая ось кристалла z составляет угол α с поверхностью и лежит в плоскости рисунка, поэтому эта плоскость является главной плоско-

стью кристалла. Будем считать падающий на кристалл свет естественным, в котором содержится как компонента вектора \mathbf{E} , лежащая в плоскости падения, так и перпендикулярная плоскости падения.

Определим направление волн в кристалле, используя построение Гюйгенса.

Волновая поверхность MN в падающей волне (перпендикулярная направлению распространения и следовательно, параллельная поверхности кристалла) одновременно достигает всех точек поверхности (от A до D) возбуждая в диэлектрике вторичные волны. Источниками этих вторичных волн являются возбуждаемые падающей волной осцилляторы среды, находящиеся в точках поверхности от A до D . Рассмотрим сначала обычно-

венную волну в кристалле, в которой вектор \mathbf{E} перпендикулярен главной плоскости (плоскости рис. 10.7 а). В этом случае каждая точка является вторичным источником сферической волны (так как скорость обыкновенной волны одинакова во всех направлениях). Огибающая этих сферических поверхностей одинакового радиуса — плоская поверхность, параллельная поверхности кристалла — является волновой поверхностью обыкновенной волны; мы видим, что вектор \mathbf{k}_o обыкновенной волны, направленный по нормали к волновому фронту, совпадает по направлению с вектором \mathbf{k} в падающей волне (как и должно быть в соответствии с законом преломления).

А каково направление потока энергии, т. е. вектора Пойнтинга, в обыкновенной волне? Обратим внимание, что к моменту времени, когда волновая поверхность обыкновенной волны, удаляясь от поверхности кристалла со скоростью v_o , достигает положения $M_o N_o$, колебания поля на фронте имеются лишь на участке от A_1 до D_1 .

Действительно, колебание в точке A_1 в этот момент времени создается только источником, находящемся в точке A (волны от других вторичных источников B , C , D не успевают достичь точки A_1 к этому моменту). Колебание в точке B_1 создано только источником в точке B и т. д. Итак, возмущение из участка AD переместилось к участку $A_1 D_1$, сместившись вертикально вниз. Можно сделать вывод: вектор потока энергии \mathbf{S}_o направлен так же, как и волновой вектор \mathbf{k}_o на рис. 10.7 а — вертикально вниз.

Рассмотрим теперь необыкновенную волну (рис. 10.7 б). Волновые поверхности необыкновенной волны, переизлучаемой каждой из точек A , B , C , D — эллипсоиды. (На рисунке показан отрицательный кристалл: для него $n_e < n_o$ и следовательно $v_e > v_o$). Огибающая этих эллипсоидов — плоская поверхность $M_e N_e$ — является волновой поверхностью необыкновенной волны. Показанная на рисунке волновая поверхность $M_e N_e$, удаляясь от поверхности кристалла со скоростью, большей, чем скорость v_o обыкновенной волны, оказалась дальше от поверхности кристалла. Нормаль к волновой поверхности $M_e N_e$ направлена вертикально вниз, именно так направлен волновой вектор \mathbf{k}_e необыкновенной волны, что также находится в согласии с законом преломления.

А какое направление имеет поток энергии в необыкновенной волне? Очевидно, что возмущение на волновом фронте локализовано на участке от A_2 до D_2 . Действительно, к моменту времени, зафиксированному на рисунке, возмущение в точке A_2 создано лишь излучателем, расположенным в точке A поверхности кристалла (возмущения от всех прочих излучателей B , C , D не успели дойти до точки A_2). То же самое касается

всех прочих точек волнового фронта $M_e N_e$: возмущение в точке B_2 создано излучателем B , в точке C_2 — излучателем C и т. д. До точек волнового фронта, лежащих вне участка $A_2 D_2$ возмущение не успело дойти ни от одного из излучателей от A до D , возбужденных падающим на кристалл лучом света. Таким образом, результирующее возмущение в необыкновенной волне, локализованное первоначально на участке AD , оказалось локализованным через некоторое время на участке $A_2 D_2$. Направление, в котором распространяется луч, определяется направлением прямой, соединяющей точку A с точкой касания эллипсоида (волновой поверхности, удаляющейся от точки A) с волновым фронтом $M_e N_e$, т. е. с точкой A_2 . То же направление имеют и прямые BB_2 , CC_2 или DD_2 .

Итак, падающий на кристалл луч света возбуждает в кристалле *необыкновенный луч*, направление которого не совпадает в общем случае с направлением падающей волны даже при нормальном падении. Такое необычное поведение и явилось причиной названия — необыкновенный луч. Важно обратить внимание на тот факт, что в необыкновенной волне направление скорости перемещения волновых поверхностей (волнового вектора \mathbf{k}_e) не совпадает в общем случае с направлением потока энергии (вектора Пойнтинга \mathbf{S}_e), т. е. с направлением луча.

Еще раз подчеркнем, что закон преломления определяет направление волновых поверхностей (вектора \mathbf{k}) в преломленной волне, а не направление луча — потока энергии.

Итак, при падении на одноосный кристалл световой волны, в нем возникают две волны, распространяющиеся от поверхности в общем случае в разных направлениях с разными скоростями и по-разному поляризованных (в главной плоскости необыкновенный луч и перпендикулярно главной плоскости — обыкновенный луч).

Это явление называется *двойным лучепреломлением* и было открыто в 1670 г. датским физиком Э. Бартолином (1625–1698) в кристалле исландского шпата.

Опыт, демонстрирующий эффект двойного лучепреломления, показан на рис. 10.8: узкий пучок света нормально падает на естественную грань кристалла исландского шпата, имеющего форму ромбоэдра. Из противоположной грани выходят два параллельных пучка света: несмещенный (обыкновенный) и смещенный (необыкновенный). При повороте кристалла вокруг оси падающего пучка, обыкновенный луч остается неподвижным,

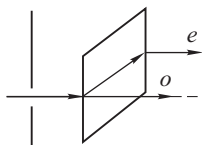


Рис. 10.8

а необыкновенный поворачивается вокруг оси обыкновенного луча. Лучи o и e оказываются поляризованными во взаимно перпендикулярных направлениях.

10.4. Дихроизм. Поляризаторы

Еще одним проявлением анизотропии является *дихроизм*. Это явление состоит в том, что в некоторых кристаллах световая волна с определенной ориентацией вектора **Е** поглощается сильнее, чем волны с другой ориентацией электрического поля. Например, кристалл турмалина толщиной около 1 мм поглощает практически полностью обыкновенный луч и почти не поглощает необыкновенный луч. Аналогичным свойством обладает синтетический дихроичный материал, называемый *поляроидом*. Сильный дихроизм поляроидной пленки проявляется уже при толщине порядка 0,1 мм.

Дихроичные материалы используются для получения и анализа поляризованного света, т.е. в качестве поляризаторов и анализаторов. При нормальном падении естественного света, который можно представить как суперпозицию двух взаимно перпендикулярных колебаний (например, E_x и E_y) с хаотически изменяющейся разностью фаз и равной интенсивностью, одно из направлений колебаний совпадающее с «запрещенным» направлением поляризатора (например, x -составляющая) практически полностью поглощается, а составляющая света, колебания в которой совпадают с «разрешенным» направлением, проходит поляризатор без ослабления. Таким образом, волна на выходе поляризатора оказывается практически полностью линейно-поляризованной, в ней ориентация вектора **Е** совпадает с «разрешенным» направлением поляризатора, а ее интенсивность равна половине интенсивности падающей на поляризатор волны $I = I_0/2$. Если теперь прошедшую через поляризатор световую волну направить на второй поляризатор, разрешенное направление которого составляет угол φ с разрешенным направлением первого (рис. 10.9), а значит и с плоскостью колебаний вектора **Е** в падающей на него волне, то он пропустит без поглощения лишь составляющую волны, в которой колебания поля совпадают по направлению с разрешенным направлением P ; составляющая с колебаниями, перпендикулярными P , полностью погасится. Амплитуда волны на выходе $A_{\text{вых}}$ будет равна $A_{\text{вых}} = A_0 \cos \varphi$, а интенсивность (квадрат амплитуды)

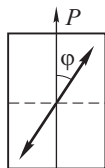


Рис. 10.9

$$I_{\text{вых}} = I \cos^2 \varphi. \quad (10.7)$$

Это соотношение называется *законом Малюса*. Его открыл в 1808 г. французский физик Э. Малюс (1775–1812).

Для получения поляризованного света используют также поляризационные призмы. Первая поляризационная призма

(рис. 10.10), изготовленная в 1828 г. шотландским ученым У. Николем (1768–1851), так и называется — *призмой Николя*.

Разрезанная по диагонали AB призма из исландского шпата склеивается канадским бальзамом, показатель преломления которого $n = 1,549$ меньше, чем показатель преломления n_o и больше величины n_e для исландского шпата.

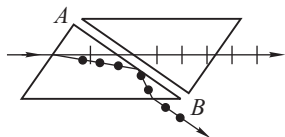


Рис. 10.10

При определенных углах падения на границу раздела исландский шпат–канадский бальзам обыкновенный луч (для которого $n_o > n$) испытывает полное внутренне отражение, а необыкновенный луч, частично отражаясь, проходит через слой бальзама, так как для него исландский шпат является оптически менее плотной средой, чем бальзам ($n_e < n$). Таким образом, луч, выходящий из призмы параллельно падающему лучу, оказывается полностью поляризованным (колебания электрического поля в нем происходят в главной плоскости).

10.5. Электрооптические и магнитооптические эффекты

Эффект Керра. Многие вещества состоят из анизотропных молекул, не обладающих собственным дипольным моментом. Анизотропия этих молекул проявляется в том, что дипольный момент, который приобретает молекула во внешнем электрическом поле, зависит от направления поля. В некотором направлении молекула поляризуется внешним полем легче, чем в других (*ось наибольшей поляризации*).

Пусть, например, электрическое поле \mathbf{E} направлено вдоль оси z , а молекула ориентирована в пространстве таким образом, что индуцированный полем дипольный момент молекулы оказывается равным $\mathbf{p} = \alpha_1 \mathbf{E}$ (α_1 — поляризуемость вдоль оси z). Если поле направлено вдоль оси x или вдоль оси y , то $\mathbf{p} = \alpha_2 \mathbf{E}$. При $\alpha_1 > \alpha_2$ ось z является осью наибольшей поляризуемости молекулы. Ясно, что при произвольном направлении внешнего поля (относительно оси наибольшей поляризуемости) дипольный момент, который приобретает молекула, не совпадает, вообще говоря, с направлением поля. Хаотическая ориентация в пространстве анизотропных молекул (в жидкостях и газах) приводит к тому, что оптическая анизотропия тем не менее не проявляется: ни одна из ориентаций молекул не является преимущественной, поэтому макроскопически (в объеме, содержащем большое число молекул) среда оказывается оптически изотропной — характер

распространения света не зависит от его направления и поляризации.

Если такое вещество поместить во внешнее поле, то возникающий при этом момент силы, действующей на индуцированный полем диполь, заставляет молекулу повернуться таким образом, чтобы ось наибольшей поляризуемости сориентировалась преимущественно вдоль поля — хаотическая ориентация осей наибольшей поляризуемости молекул сменяется частично упорядоченной ориентацией и среда становится *оптически анизотропной*, подобной одноосному кристаллу, в котором направление оптической оси совпадает с направлением поля. Световой волне, в которой вектор \mathbf{E} совпадает с направлением ориентирующего внешнего поля (т. е. необыкновенной волне), соответствует показатель преломления n_e и фазовая скорость $v_e = c/n_e$, а волне, в которой вектор \mathbf{E} не имеет проекции на направление внешнего поля (обыкновенной волне), соответствует показатель преломления $n_o < n_e$, т. е. среда во внешнем поле приобретает свойства положительного кристалла. Анизотропия, возникающая во внешнем электрическом поле, была открыта в 1875 г. шотландским физиком Дж. Керром (1824–1907). Опыт показывает, что разность фаз обыкновенной и необыкновенной волн в среде равна

$$\varphi = \frac{2\pi}{\lambda} (n_e - n_o)d = 2\pi B d E^2, \quad (10.8)$$

где d — путь, проходимый лучом света в среде. Квадратичная зависимость разности фаз от напряженности поляризующего поля указывает на независимость знака $n_e - n_o$ от направления поля (вверх или вниз на рис. 10.11). Константа B называется *постоянной Керра*. Особенно заметно эффект Керра проявляется в нитробензоле, который обладает большим значением B : на пути $d = 10$ см и напряженности поля $E = 1,5 \cdot 10^6$ В/м разность фаз близка к значению $\varphi \approx \pi/2$ (луч света распространяется в среде перпендикулярно приложенному полю).

На рис. 10.11 показана схема наблюдения эффекта Керра. Между двумя скрещенными поляризаторами P_1 и P_2 (их разрешенные направления взаимно перпендикулярны) помещается «ячейка Керра» — конденсатор, заполненный исследуемым веществом. При отсутствии напряжения на конденсаторе, луч света не проходит через систему, что указывает на то, что среда при отсутствии поля изотропна. При наложении внешнего поля (приложении разности потенциалов между пластинами конденсатора) среда становится оптически анизотропной, поэтому линейно-поляризованная первым поляризатором P_1 волна, пройдя через среду, становится эллиптически

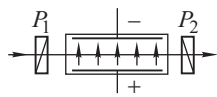


Рис. 10.11

поляризованной и частично проходит через второй поляризатор P_2 .

Эффект Керра проявляется также в некоторых веществах, имеющих полярные молекулы, дипольный момент которых не совпадает с направлением оси наибольшей поляризуемости. Пусть, например, эти направления взаимно перпендикулярны, а собственный дипольный момент существенно превосходит дипольный момент, индуцированный внешним полем. Тогда, оказавшись во внешнем поле, молекулы ориентируются преимущественно таким образом, что с направлением поля будет совпадать собственный дипольный момент, а ось наибольшей поляризуемости окажется перпендикулярной полю. Такое вещество в электрическом поле подобно отрицательному кристаллу.

Важно отметить исключительно малую инерционность эффекта Керра: при включении поля анизотропия вещества проявляется уже через время порядка 10^{-12} с. Подавая на обкладки конденсатора переменное напряжение, можно использовать устройство на рис. 10.11 для прерывания (модуляции) лазерного пучка вплоть до частот порядка 10^9 Гц. На этом принципе основаны высокочастотные оптические затворы.

Эффект Поккельса. Электрическое поле может изменять не только оптические характеристики жидкостей и газов, оптически изотропных в отсутствие поля, но также и оптические свойства кристаллов. Если поместить одноосный кристалл в электрическое поле и направить вдоль оптической оси z луч света, в котором поле \mathbf{E} может иметь только x - и y -компоненты, то фазовые скорости волн E_x и E_y становятся различными (в отсутствие поля обе компоненты имеют равные скорости $v_x = v_y = c/n_o$). Это явление называется *эффектом Поккельса* (по имени немецкого физика Ф.К. Поккельса (1865–1913)). Малая инерционность позволяет использовать эффект для высокочастотной модуляции света. В отличие от эффекта Керра, разность фаз волн E_x и E_y в кристалле оказывается пропорциональной первой степени напряженности поля. Изменение показателя преломления (Δn) при эффекте Поккельса происходит за счет смещения электронов в сложном кристалле в сторону того или иного иона (например, в случае ниобата лития LiNbO_3 — это ион ниобия), при этом за счет изменения электронной конфигурации меняется поляризуемость, а с ней — и показатель преломления. В силу линейности эффект Поккельса может наблюдаться лишь в кристаллах, не обладающих центром симметрии. Дело в том, что при изменении направления ориентирующего поля на противоположное, должен измениться на противоположный и знак Δn . Это невозможно в кристаллах с центром симметрии, поскольку для них оба

взаимнопротивоположных направления поля физически эквивалентны.

Оба эффекта — Керра и Поккельса — могут проявляться как в изменении свойств оптически анизотропных сред, так и в наведении искусственной анизотропии в среде, изотропной в отсутствие внешнего воздействия.

Анизотропия в магнитном поле проявляется у веществ, молекулы которых обладают постоянными магнитными моментами. (*Эффект Коттона–Мутона*, открытый в 1905 г. французскими физиками Э. Коттоном (1881–1967) и А. Мутоном.)

Эффект Фарадея. При прохождении света через вещество, помещенное в магнитное поле, наблюдается поворот плоскости поляризации (плоскости колебаний вектора \mathbf{E}) линейно-поляризованной световой волны. Этот эффект открыт Фарадеем в 1846 г. Чтобы понять физический механизм явления, следует обратиться к модели среды, которая используется в классической теории дисперсии. Необходимо лишь учесть, что, кроме квазиупругой силы, действующей на оптический электрон в атоме, и силы электрического поля световой волны $e\mathbf{E}$, действует также сила Лоренца, обусловленная магнитным полем $\mathbf{F}_\text{л} = e[\dot{\mathbf{r}} \mathbf{B}]$, где $\dot{\mathbf{r}}$ — скорость электрона. С учетом этого обстоятельства уравнение движения электрона (при пренебрежении затуханием) имеет вид

$$\ddot{\mathbf{r}} + \omega_0^2 \mathbf{r} = \frac{e}{m} (\mathbf{E} + [\dot{\mathbf{r}} \mathbf{B}]).$$

Вводя далее вектор циклотронной частоты $\boldsymbol{\Omega}$ согласно равенству

$$\boldsymbol{\Omega} = \frac{e}{m} \mathbf{B}, \quad (10.9)$$

запишем

$$\ddot{\mathbf{r}} - [\dot{\mathbf{r}} \boldsymbol{\Omega}] + \omega_0^2 \mathbf{r} = e\mathbf{E}. \quad (10.10)$$

При наблюдении эффекта Фарадея пучок линейно-поляризованного света направляют вдоль направления магнитного поля (оси z). В этом пучке поле \mathbf{E} может иметь x - и y -компоненты: E_x и E_y . Для дальнейшего удобно представить линейно-поляризованную волну как суперпозицию двух волн с круговой поляризацией, в одной из которых вектор \mathbf{E} вращается против часовой стрелки (волна E_+), а в другой — по часовой стрелке (волна E_-). Тогда можно записать

$$E_x = E_0 \cos \omega t; \quad E_y = \pm E_0 \sin \omega t, \quad E_z = 0.$$

Знак «+» в E_y соответствует волне E_+ , а знак «-» волне E_- . Согласно (10.9) вектор $\boldsymbol{\Omega}$ имеет проекции на оси x , y , z соответственно $(0, 0, \Omega)$, поэтому проекции вектора $[\dot{\mathbf{r}} \boldsymbol{\Omega}]$ есть соответственно $(\dot{y}\Omega, -\dot{x}\Omega, 0)$. Проекция векторного уравнения

(10.10) на оси x , y , z имеют вид

$$\begin{aligned}\ddot{x} - \Omega \dot{y} + \omega_0^2 x &= \frac{e}{m} E_0 \cos \omega t, \\ \ddot{y} + \Omega \dot{x} + \omega_0^2 y &= \frac{e}{m} E_0 \sin \omega t, \\ \ddot{z} + \omega_0^2 z &= 0\end{aligned}\quad (10.11)$$

(ясно, что магнитное поле, направленное вдоль оси z , не влияет на движение электрона вдоль этой оси). Можно убедиться непосредственной подстановкой в уравнения (10.11), что вынужденные колебания $x(t)$ и $y(t)$ (решения уравнений) имеют вид

$$x(t) = \frac{(e/m) E_x(t)}{\omega_0^2 - \omega^2 \pm \Omega \omega}, \quad y(t) = \frac{(e/m) E_y(t)}{\omega_0^2 - \omega^2 \pm \Omega \omega}. \quad (10.12)$$

Из уравнений (10.12) ясно, что волна круговой поляризации вынуждает оптический электрон двигаться по окружности с частотой ω , причем радиус окружности $R = \sqrt{x^2 + y^2}$ различен для волн со встречным направлением вращения вектора \mathbf{E} (т. е. для волн E_+ и E_-).

Возводя левые и правые части равенств (10.12) в квадрат и складывая их, находим траекторию движения электрона

$$R_{\pm}^2 = x^2 + y^2 = \frac{(e/m)^2 E_0^2}{(\omega_0^2 - \omega^2 \pm \Omega \omega)^2},$$

т. е. вращающийся вектор \mathbf{E} световой волны заставляет вращаться дипольный момент атома, равный по модулю $p_{\pm} = e R_{\pm}$. Таким образом, величина индуцированного дипольного момента оказывается различной для волн E_+ и E_- ; соответственно различна поляризация среды, диэлектрическая проницаемость и, следовательно, показатель преломления.

Проекция индуцированного дипольного момента есть $p_x = ex(t)$ и $p_y = ey(t)$. Напомним, что атомная поляризуемость определяется равенством $\mathbf{p} = \varepsilon_0 \alpha \mathbf{E}$. Используя (10.12), находим атомную поляризуемость для волн E_+ и E_-

$$\alpha_{\pm} = \frac{e^2/(m\varepsilon_0)}{\omega_0^2 - \omega^2 \pm \Omega \omega}. \quad (10.13)$$

Заметим, что циклотронная частота Ω даже в очень сильных магнитных полях существенно меньше частоты света оптического диапазона ($\omega \approx 10^{15} \text{ с}^{-1}$, $\Omega \approx 10^{10} \text{ с}^{-1}$ при $B \approx 10 \text{ Тл}$), поэтому знаменатель в (10.13) удобно записать приближенно: $\omega_0^2 - \omega^2 \pm \Omega \omega \approx \omega_0^2 - (\omega \pm \Omega/2)^2$, тогда получим

$$\alpha_{\pm} \approx \frac{e^2/(m\varepsilon_0)}{\omega_0^2 - (\omega \pm \Omega/2)^2}. \quad (10.14)$$

Сравним полученное выражение с формулой (9.15), определяющей атомную поляризуемость α в отсутствие магнитного поля (при $\Omega = 0$). Формула (10.14) может быть получена из (9.15) заменой ω на $\omega \pm \Omega/2$ т. е. $\alpha_{\pm}(\omega) = \alpha(\omega \pm \Omega/2)$. Поскольку показатель преломления среды связан с α равенством $n = \sqrt{\epsilon} = \sqrt{1 + N\alpha}$, то соответственно

$$n_{\pm}(\omega) = n\left(\omega \pm \frac{\Omega}{2}\right) \approx n(\omega) \pm \frac{dn}{d\omega} \frac{\Omega}{2}. \quad (10.15)$$

Мы пришли к следующему результату: показатель преломления для волн круговой поляризации с различным направлением вращения вектора \mathbf{E} различен. Следовательно, различны и фазовые скорости волн E_+ и E_- : $v_{\pm} = c/n_{\pm}$ — это и приводит к повороту плоскости колебаний вектора \mathbf{E} линейно-поляризованной волны. Поскольку величина Ω отрицательна (заряд электрона $e < 0$), то $n_- > n_+$, а $v_+ > v_-$ при $dn/d\omega > 0$, т. е. в области нормальной дисперсии. Связать угол поворота плоскости поляризации с разностью показателей преломления ($n_+ - n_-$) можно, используя следующие соображения.

Пусть на входе в среду (в плоскости $z = 0$) вектор \mathbf{E} в волне колеблется в вертикальной плоскости — вдоль линии OO на рис. 10.12а. Это эквивалентно тому, что векторы волн \mathbf{E}_+ и \mathbf{E}_- , поляризованных по кругу (сумма которых и дает волну линейной поляризации), вращаются в противоположных направлениях с одинаковой угловой скоростью ω , занимая в любой момент времени t симметричное расположение относительно вертикальной оси OO : $\varphi_+ = \varphi_- = \omega t$ (считаем для простоты, что при $t = 0$ $\varphi_+ = \varphi_- = 0$, т. е. векторы \mathbf{E}_+ и \mathbf{E}_- в этот момент направлены вертикально вверх и, значит, суммарный вектор \mathbf{E} в этот момент имеет максимальную длину и направлен вертикально вверх).

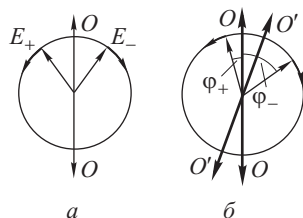


Рис. 10.12

Пройдя в среде путь, равный z , волны \mathbf{E}_+ и \mathbf{E}_- приобретают набег фазы, равный $\varphi_{\pm} = kz = (\omega/c)n_{\pm}z$ — он и определяет положение векторов \mathbf{E}_+ и \mathbf{E}_- в момент времени t . Как ясно из рис. 10.12б, суммарный вектор \mathbf{E} повернут относительно первоначальной вертикальной ориентации на угол

$$\varphi = \frac{1}{2}(\varphi_- - \varphi_+) = \frac{\omega}{2c}(n_- - n_+)z, \quad (10.16)$$

колебания происходят вдоль направления $O'O'$. Используя (10.15) и (10.16), находим

$$\varphi = \frac{\omega}{c} \left(\frac{dn}{d\omega} \right) \Omega z = \frac{e\omega}{2mc} \frac{dn}{d\omega} z B. \quad (10.17)$$

Схема эксперимента, демонстрирующая эффект Фарадея, показана на рис. 10.13. Исследуемое вещество помещается между полюсами электромагнита. В сердечнике магнита просверлен канал, через который пропускается пучок света, линейно-поляризованный поляризатором P_1 . При отсутствии магнитного поля (ток в обмотке электромагнита выключен) свет не проходит через скрещенный поляризатор P_2 . При включении магнитного поля происходит поворот плоскости поляризации света на угол φ , который пропорционален пути света в веществе z и индукции магнитного поля B :

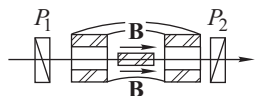


Рис. 10.13

$$\varphi = \rho z B, \quad (10.18)$$

где константа ρ называется *постоянной Верде*. Угол поворота плоскости колебаний можно определить поворотом анализатора, определив тем самым константу ρ экспериментально, и сравнить полученное значение с теоретическим

$$\rho = \frac{e\omega}{2mc} \frac{dn}{d\omega}, \quad (10.19)$$

если закон дисперсии среды $n(\omega)$ известен. Данные эксперимента находятся в хорошем согласии с формулой (10.19), полученной на основе приведенной здесь теории. Отметим, что в соответствии с теорией, направление поворота плоскости поляризации не зависит от направления распространения света (по полю или против поля). Поместив исследуемое вещество между двумя зеркалами и заставив таким образом луч света многократно проходить путь туда-обратно, можно существенно увеличить угол поворота.

Время установления значений n_+ и n_- после включения поля менее 10^{-9} с, что позволяет использовать эффект Фарадея для высокочастотной модуляции света.

Вопросы и задачи

1. В интерференционном опыте Юнга между щелью S и щелями S_1 и S_2 введен поляризатор P (рис. 10.14), разрешенное направление которого параллельно либо перпендикулярно щелям S_1 и S_2 . Как изменится интерференционная картина на экране \mathcal{E} , если и S_1 и S_2 прикрыть пластинками в $\lambda/2$, ориентированными взаимно перпендикулярно друг к другу (параллельно и перпендикулярно к щелям)?

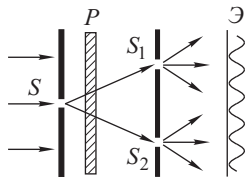


Рис. 10.14

Что произойдет, если поляризатор повернуть на 90° ? Какая картина будет наблюдаться, если убрать поляризатор?

2. Ответить на те же вопросы, если вместо пластинок в $\lambda/2$ использовать пластинки в $\lambda/4$.

3. Некогерентная смесь линейно-поляризованного света и света, поляризованного по кругу, рассматривается через поляририд. Найдено положение поляриоида, соответствующее максимальной интенсивности прошедшего света. При повороте поляриоида из этого положения на угол $\lambda = 30^\circ$ интенсивность света уменьшилось на 20%. Найти отношение интенсивности света I_k , поляризованного по кругу, к интенсивности линейно поляризованного света.

4. Как будет окрашена в скрещенных поляроидах кристаллическая пластинка, дающая разность хода $d(n_e - n_o) = 650$ нм.

Ответ: в синий свет (дополнительный к оранжевому).

5. Один поляририд пропускает 30% света, если на него падает естественный свет. После прохождения света через два таких поляриоида интенсивность падает до 9%. Найти угол между разрешенными направлениями поляриридов.

6. Плоская монохроматическая волна, поляризованная по кругу, создает в точке P интенсивность I_0 . На пути волны ставят две большие пластинки 1 и 2 в $\lambda/4$, как показано на рис. 10.15. Главные направления пластинок ориентированы взаимноперпендикулярно. Найти интенсивность I в точке P .

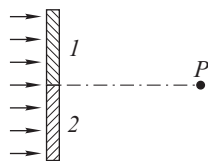


Рис. 10.15

7. Расположив пластинку из исландского шпата, вырезанную параллельно его оптической оси, между скрещенными поляроидами, можно осуществить монохроматор, позволяющий задержать одну из спектральных линий дублета натрия и пропустить другую. Какой должна быть минимальная толщина d_{\min} пластинки и как ее следует ориентировать? Показатели преломления n_e и n_o для исландского шпата

$$n_{e1} = 1,48654, n_{o1} = 1,65846 \text{ (для линии } \lambda_1 = 589 \text{ нм),}$$

$$n_{e2} = 1,48652, n_{o2} = 1,65843 \text{ (для линии } \lambda_2 = 589,6 \text{ нм).}$$

8. Показатель преломления кристаллического кварца для длины волны $\lambda = 589$ нм равен $n_o = 1,544$ для обыкновенной волны и $n_e = 1,553$ для необыкновенной волны. На пластинку из кварца, вырезанную параллельно оптической оси, нормально падает линейно-поляризованный свет, занимающий спектральный интервал $\Delta\lambda = 40$ нм. Найти толщину пластинки d и направление поляризации падающего света, если свет на выходе из пластики оказался неполяризованным.

ЭЛЕМЕНТЫ НЕЛИНЕЙНОЙ ОПТИКИ

11.1. Модель нелинейной среды

В явлениях, которые мы изучали до сих пор (интерференция и дифракция, дисперсия и распространение волн в анизотропных средах), молчаливо предполагалось, что характеристики среды, в которой распространяются волны (в частности ее показатель преломления $n = \sqrt{\varepsilon}$), не зависят от того, велика или мала напряженность электрического поля волны (хотя, конечно, показатель преломления может зависеть от координат, если среда неоднородна и от ориентации вектора \mathbf{E} , если она не изотропна). Следствием является линейность волнового уравнения и справедливость принципа суперпозиции: распространение волны в среде не изменяет ее свойства, поэтому одна волна проходит в среде так, как будто другой волны нет — волны распространяются в среде, не «замечая» друг друга.

Независимость показателя преломления от напряженности поля для световых волн, излучаемых обычными тепловыми источниками (нить накаливания электрической лампы, нагретый газ и т. д.), обусловлена пренебрежимой малостью поля световых волн по сравнению с внутренним микроскопическим полем в веществе. Напряженность поля в излучении обычных источников не превосходит величины порядка 10^4 В/м. Сравните эту величину например, с полем, действующим на электрон в атоме водорода: $e/(4\pi\varepsilon_0 R^2) \approx 1,5 \cdot 10^{11}$ В/м.

Можно ожидать, что световая волна изменит характеристики среды, если ее поле окажется сравнимо с внутриатомными полями в веществе. Возможность получения интенсивных пучков света появилась после создания квантовых генераторов советскими физиками А.М. Прохоровым (1916–2002) и Н.Г. Басовым (1922–2001) и независимо от них американским физиком Ч.Х. Таунсом (р. 1915) в 1954 г. В сфокусированном лазерном пучке напряженность поля достигает значений порядка 10^8 – 10^{10} В/м. Именно лазеры открыли по-настоящему эру нелинейной оптики (хотя и до их создания отдельные нелинейные эффекты в оптике наблюдались).

Чтобы понять механизм влияния интенсивных световых волн на свойства среды, полезно вернуться к модели, которую мы использовали ранее: материальная среда представляет собой набор

осцилляторов (не взаимодействующих между собой в простейшем случае). Осциллятор — это оптический электрон в атоме. Исследуя ранее явление дисперсии, мы молчаливо предполагали, что напряженность поля световой волны достаточно мала, поэтому малым оказывается смещение электрона из положения равновесия, вызванное этим полем. При малых смещениях \mathbf{r} , сила, действующая на электрон со стороны поля ядра и стремящаяся вернуть электрон в положение равновесия, предполагалась пропорциональной смещению (квазиупругая сила):

$$\mathbf{f} = -k\mathbf{r}, \quad (11.1)$$

подобно тому, как малые деформации пружины при ее растяжении или сжатии приводят к линейной зависимости упругой силы от величины деформации — закон Гука. Исходя из линейной связи (11.1) мы и получали ранее линейную связь между поляризацией среды и напряженностью поля световой волны $\mathbf{P} = \varepsilon_0\chi\mathbf{E}$, где χ — поляризуемость среды — определяется формулой (9.11) и не зависит от \mathbf{E} .

В сильных полях смещение уже нельзя считать малым, при этом связь между силой и смещением усложняется: в линейное соотношение (11.1) следует ввести нелинейные поправки. Для того чтобы понять, какова их роль и к каким эффектам они приводят, будем далее полагать что поправки, возникающие в сильных полях, являются все же малыми по сравнению с первым (линейным) слагаемым.

$$\mathbf{f} = -k\mathbf{r} + l\mathbf{r}^2 + q\mathbf{r}^3, \quad (11.2)$$

здесь l и q — константы, и именно они, как мы увидим далее, приводят к новым оптическим эффектам.

В общем случае в анизотропной среде направление смещения \mathbf{r} не совпадает с направлением «упругой силы», возвращающей электрон в положение равновесия, однако мы далее будем пользоваться упрощенной моделью, считая, что векторы \mathbf{f} и \mathbf{r} параллельны (т. е. в равенстве (11.2) $l\mathbf{r}^2 = lr\mathbf{r}$ и $q\mathbf{r}^3 = qr^2\mathbf{r}$).

11.2. Эффект удвоения частоты. Оптическое выпрямление

Рассмотрим вначале эффекты, к которым приводит наличие лишь первой нелинейной поправки, т. е.

$$\mathbf{f} = -k\mathbf{r} + l\mathbf{r}^2. \quad (11.3)$$

Отметим, что такого рода зависимость невозможна в изотропных средах или в кристаллах, обладающих центром симметрии, например, в кубических кристаллах, поскольку, согласно (11.3),

изменение знака смещения ($-r$ вместо r) приводит к изменению величины силы f (при замене r на $-r$ второе слагаемое не меняет знака, а первое изменяет, поэтому суммарная сила f изменяет величину).

Уравнение движения электрона, осциллирующего под действием поля $\mathbf{E}(t)$ световой волны, имеет при этом вид

$$m\ddot{\mathbf{r}} = -k\mathbf{r} + l\mathbf{r}^2 + e\mathbf{E}(t)$$

(вновь пренебрегаем затуханием, полагая, что частота поля световой волны далека от собственной частоты осциллятора).

В силу сделанных выше предположений можно перейти к скалярному уравнению

$$\ddot{r} + \omega_0^2 r - \frac{l}{m} r^2 = \frac{e}{m} E(t),$$

где $\omega_0 = \sqrt{k/m}$ — собственная частота малых колебаний осциллятора. Умножая левую и правую части последнего равенства на Nq , перейдем от уравнения для смещения r к уравнению для поляризации среды $P = Ner = Np$ ($p = er$ — дипольный момент осциллятора, N — число осцилляторов в единице объема)

$$\ddot{P} + \omega_0^2 P - \frac{l}{Nem} P^2 = \frac{Ne^2}{m} E(t). \quad (11.4)$$

Колебание поля световой волны $E(t)$ считаем гармоническим $E(t) = E_0 \cos \omega t$.

Малость нелинейной поправки ($l/(Nem)P^2 \ll \omega_0^2 P$) позволяет при решении уравнения (11.4) воспользоваться методом последовательных приближений: решение уравнения будем искать в виде $P(t) = P_0(t) + P_1(t)$, где $P_0(t)$ — нулевое приближение — решение линейного уравнения, которое получается из (11.4) отбрасыванием нелинейного члена $l/(Nem)P^2$:

$$\ddot{P}_0 + \omega_0^2 P_0 = \frac{Ne^2}{m} E_0 \cos \omega t, \quad (11.5)$$

а $P_1(t)$ — малая поправка к нулевому приближению ($P_1(t) \ll P_0(t)$). Решение $P_0(t)$ хорошо известно:

$$P_0(t) = \frac{Ne^2/m}{\omega_0^2 - \omega^2} E_0 \cos \omega t. \quad (11.6)$$

Коэффициент пропорциональности между $P_0(t)$ и $E(t)$ — это поляризуемость линейной среды χ_0 (не зависящая от напряженности поля $E(t)$)

$$\chi_0 = \frac{Ne^2/(m\varepsilon_0)}{\omega_0^2 - \omega^2}. \quad (11.7)$$

Колебания поляризации $P_0(t)$ происходят с частотой световой волны ω и приводят к переизлучению на той же частоте. Поля-

ризуемость $\chi_0(\omega)$ является функцией частоты ω . Эта функция и определяет хорошо известный закон дисперсии в линейном приближении. Далее, подставляя функцию $P(t) = P_0(t) + P_1(t)$ (с уже найденным нулевым приближением $P_0(t)$) в уравнение (11.4), получаем

$$\ddot{P}_0 + \ddot{P}_1 + \omega_0^2(P_0 + P_1) - \frac{l}{Nem} (P_0 + P_1)^2 = \frac{Ne^2}{m} E_0 \cos \omega t$$

или, в силу (11.5)

$$\ddot{P}_1 + \omega_0^2 P_1 = \frac{l}{Nem} (P_0 + P_1)^2. \quad (11.8)$$

Это и есть уравнение для нахождения поправки $P_1(t)$ к найденному решению $P_0(t)$. В правой части уравнения (11.8) члены, содержащие $P_1^2(t)$ и $P_0(t)P_1(t)$, можно отбросить (они малы по сравнению с членом, содержащим $P_0^2(t)$), поэтому

$$\ddot{P}_1 + \omega_0^2 P_1 = \frac{l}{Nem} P_0^2(t), \quad (11.9)$$

где $P_0(t)$ — решение линейного уравнения (11.5) — определяется формулой (11.6), т. е.

$$\ddot{P}_1 + \omega_0^2 P_1 = \frac{l}{Nem} \varepsilon_0^2 \chi_0^2 E_0^2 \cos^2 \omega t$$

(χ_0 — линейная диэлектрическая восприимчивость), или, так как $\cos^2 \omega t = 1/2 + (1/2) \cos 2\omega t$,

$$\ddot{P}_1 + \omega_0^2 P_1 = \beta E_0^2 + \beta E_0^2 \cos 2\omega t,$$

где введена константа β

$$\beta = \frac{l\chi_0^2\varepsilon_0^2}{2Nem}.$$

Итак, для поправки P_1 к нулевому приближению мы получили линейное дифференциальное уравнение. Точно такой же вид имеет уравнение гармонического осциллятора, на который действует внешняя сила, состоящая из двух слагаемых: постоянной силы, равной $m\beta E_0^2$, и гармонической внешней силы, меняющейся с частотой 2ω . Соответственно, и решение P_1 (в силу линейности уравнения) состоит из двух слагаемых. Одно из них (отвечающее постоянной внешней силе) имеет вид $P'_1 = \beta E_0^2 / \omega_0^2$, т. е. при постоянной внешней силе осциллятор отклоняется от положения равновесия на постоянную величину (свободные колебания осциллятора относительно постоянного значения P'_1 рано или поздно прекращаются при сколь угодно малом реально существующем затухании). Второе слагаемое, отвечающее гармонической внешней силе частоты 2ω , как легко

убедиться, имеет вид

$$P_2(t) = \frac{\beta E_0^2}{\omega_0^2 - (2\omega)^2} \cos 2\omega t. \quad (11.10)$$

Итак, к колебанию поляризации среды $P_0(t) = \varepsilon_0 \chi_0 E(t)$, изменяющейся с частотой возбуждающей волны $E(t)$, прибавляются два слагаемых: одно из них — постоянная (не меняющаяся во времени) добавка к поляризации $P = \beta E_0^2 / \omega_0^2$ и второе слагаемое — колебание поляризации $P_2(t)$, происходящее с удвоенной частотой. Итак:

$$P(t) = \varepsilon_0 \chi_0 E_0 \cos \omega t + \frac{\beta E_0^2}{\omega_0^2} + \frac{\beta E_0^2}{\omega_0^2 - 4\omega^2} \cos 2\omega t. \quad (11.11)$$

Согласно (11.11) колебание гармонического светового поля $E(t)$ возбуждает в среде колебание поляризации, которое не является гармоническим. Спектр функции $P(t)$ содержит кроме компоненты частоты ω (совпадающей с частотой колебаний световой волны) также и спектральную компоненту удвоенной частоты, а также постоянную составляющую. Мы хорошо знаем, что в линейных системах внешнее гармоническое воздействие приводит к возникновению вынужденных колебаний, частота которых совпадает с частотой внешней силы. Присутствие двух последних слагаемых в (11.11) свидетельствует о возникновении двух нелинейных эффектов.

Постоянная добавка к поляризации среды ответственна за появление в среде постоянных, не меняющихся во времени поляризационных зарядов. Эти поляризационные заряды создают свое собственное постоянное, не меняющееся во времени, электрическое поле. Итак, в нелинейной среде быстропеременное поле световой волны (меняющееся со световой частотой ω) порождает постоянное не меняющееся во времени поле. По понятным причинам этот нелинейный эффект называют *оптическим выпрямлением*.

Последнее слагаемое в (11.11) — колебание поляризации с удвоенной частотой. Это колебание приводит к переизлучению на частоте 2ω . Таким образом, монохроматическая световая волна частоты ω приводит к возникновению в среде переизлученной

волны удвоенной частоты. Волна удвоенной частоты существует, конечно, наряду с переизлучением на частоте ω , обусловленным первым слагаемым в поляризации $P(t)$.

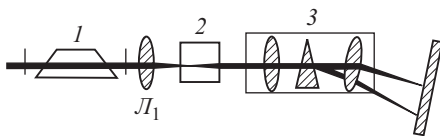


Рис. 11.1

Схема одного из первых экспериментов, демонстрирующая эффект удвоения частоты, показана на рис. 11.1.

Импульс излучения рубинового лазера I ($\lambda_1 = 694,3$ нм) фокусировался с помощью объектива L_1 в кристалле кварца 2. Мощность в импульсе длительности ~ 10 мкс составляла приблизительно 10 кВт. Излучение, выходящее из кристалла, содержит, кроме переизлучения с длиной волны λ_1 , также и вторую гармонику ($\lambda_2 = 347,15$ нм), лежащую в ультрафиолетовой области спектра. Оно анализируется с помощью призменного спектрографа 3.

В первых экспериментах (1961 г.) мощность, преобразованная во вторую гармонику, составляла приблизительно 0,1 мВт, т. е. около 10^{-8} мощности исходного лазерного импульса.

Чтобы понять, почему в первых экспериментах коэффициент преобразования энергии исходной волны в волну второй гармоники оказался столь малым, а также понять, каким образом можно его увеличить, рассмотрим более детально волновые процессы, происходящие в нелинейном кристалле.

Итак, пусть исходная волна, бегущая в кристалле, имеет вид

$$E(z, t) = E_0 \cos(\omega t - k_1 z).$$

Это плоская волна частоты ω , волновое число k_1 этой волны определяет ее фазовую скорость

$$v = \frac{\omega}{k_1} = \frac{c}{n(\omega)},$$

$n(\omega)$ — показатель преломления для волны, имеющей частоту ω .

Эта волна возбуждает в каждой точке среды колебания поляризации с частотой 2ω , определяемые формулой (11.10). Существенно, что волна поляризации $P_2(z, t)$ бежит с той же фазовой скоростью, что и исходная волна $E(z, t)$. Действительно, если принять, что колебания поля (при любом фиксированном z) происходят по закону $E_0 \cos \omega t$, то колебания $P_2(t)$ происходят по закону (11.10), т. е. без сдвига фазы, который зависел бы от z . Ясно поэтому, что волна поляризации $P_2(z, t)$ имеет вид

$$P_2(z, t) = A \cos(2\omega t - 2k_1 z),$$

т. е. волновое число этой волны есть $2k_1$, а ее фазовая скорость $v = 2\omega/2k_1 = \omega/k_1$ как раз оказывается равной фазовой скорости исходной волны. Эту волну поляризации можно рассматривать как бегущую в среде антенну или движущийся и колеблющийся с частотой 2ω электрический диполь. Как следствие (см. раздел «Электричество и магнетизм»), должна излучиться волна второй гармоники $E_2(z, t)$ частоты 2ω

$$E_2(z, t) = B \cos(2\omega t - k_2 z).$$

Из-за дисперсии ее фазовая скорость $v_2 = 2\omega/k_2 = c/n(2\omega)$ не равна, вообще говоря, фазовой скорости исходной волны (так

как $n(\omega) \neq n(2\omega)$), а значит, и фазовой скорости волны поляризации $P_2(z, t)$, т. е. скорости излучающей эту волну антенны. Это означает, что между волнами, излучаемыми «антенной» в разные моменты времени (и в разных местах) возникает фазовый сдвиг, равный $\Delta\varphi = (k_2 - 2k_1)z$.

Пока путь волны в кристалле достаточно мал:

$$\Delta\varphi = (k_2 - 2k_1)z \ll \pi, \quad (11.12)$$

все «вторичные волны», излученные бегущей антенной, складываются синфазно, поэтому амплитуда суммарной волны нарастает.

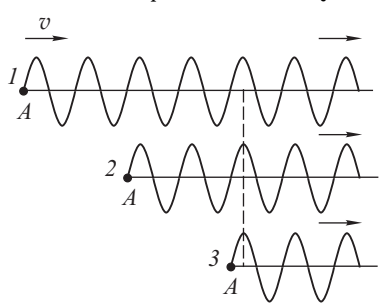


Рис. 11.2

Такое синфазное сложение схематически показано на рис. 11.2: все три волны, излученные бегущей антенной A в положении 1, 2, 3 оказываются почти синфазными, и энергия исходной волны преобразуется в энергию второй гармоники.

Условие (11.12), при котором волны второй гармоники, испущенные разными осцилляторами среды, складываясь синфазно,

образуют в сумме волну, амплитуда которой при распространении в кристалле нарастает, называется *условием фазового синхронизма*. Эффективное расстояние, на котором энергия второй гармоники нарастает, называется *длиной когерентности*. Мы получаем

$$l_{\text{ког}} \approx \frac{\pi}{k_2 - 2k_1}, \quad (11.13)$$

где $k_1 = \frac{\omega}{c} n(\omega)$, $k_2 = \frac{2\omega}{c} n(2\omega)$. Для кварца разность показателей преломления $n(2\omega) - n(\omega) \approx 0,025$ (для длины волны $\lambda = 2\pi c/\omega = 694,3$ нм). Согласно (11.13) длина когерентности оказывается чрезвычайно малой: $l_{\text{ког}} \approx 10\lambda \approx 10^{-2}$ мм. При увеличении пути в кристалле $z > l_{\text{ког}}$ расфазировка вторичных волн, излученных бегущей антенной (волной поляризации $P_2(z, t)$) становится существенной: амплитуда суммарной волны на частоте 2ω начинает уменьшаться, а значит, происходит обратная «перекачка» энергии второй гармоники в исходную волну. Таким образом, по мере распространения в кристалле, волна второй гармоники периодически (с периодом $2l_{\text{ког}}$) нарастает, то уменьшается. Качественно зависимость интенсивности второй гармоники от глубины проникновения в кристалл показана на рис. 11.3. Проблема, следовательно, состоит в том, чтобы увеличить длину когерентности и таким образом увеличить размер области, в которой амплитуда второй гармоники нарастает.

Заметим, что в изотропной среде условие (11.12) невозможно выполнить на всем пути распространения волны в кристалле: из-за дисперсии ($n(\omega) \neq n(2\omega)$) оно рано или поздно нарушается. Однако если волна распространяется в анизотропном кристалле, можно, при определенном направлении распространения относительно оптической оси, добиться выполнения условия $n_o(\omega) = n_e(2\omega)$ — равенства показателей преломления обыкновенной волны частоты ω и необыкновенной волны частоты 2ω .

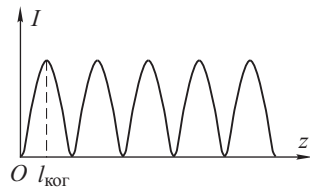


Рис. 11.3

Напомним, что в линейных средах (т. е. в слабых полях) векторная связь $\mathbf{P} = \chi \mathbf{E}$, справедливая для изотропных сред, заменяется тензорной связью в анизотропной среде:

$$P_i = \sum_j \chi_{ij} E_j$$

(каждый из индексов i, j пробегает значения 1, 2, 3, соответствующие проекциям на оси z, x, y).

Соответственно, если речь идет о нелинейной среде, то вместо связи

$$\mathbf{P} = \chi \mathbf{E} + b \mathbf{E} \cdot \mathbf{E},$$

справедливой для изотропной среды, в анизотропных кристаллах имеет место равенство:

$$P_i = \sum_j \chi_{ij} E_j + \sum_{j,k} b_{ijk} E_j E_k. \quad (11.14)$$

Будем считать для определенности, что индекс 1 соответствует оптической оси z , а индексы 2 и 3 — двум другим главным направлениям x и y .

Если исходная волна в кристалле является обыкновенной, вектор \mathbf{E} в этой волне перпендикулярен оптической оси (проекции E_j и E_k равны нулю при $j, k = 1$), но благодаря нелинейным слагаемым в (11.14), содержащим произведения $E_j E_k$ (т. е. величины $E_2 E_3$, E_2^2 и E_3^2), проекция вектора поляризации \mathbf{P} на оптическую ось z (ось 1) оказывается отличной от нуля: $P_1^{(2\omega)}$ содержит слагаемые, пропорциональные величинам $E_2^{(\omega)} E_3^{(\omega)}$, $E_2^{(\omega)} E_2^{(\omega)}$ и $E_3^{(\omega)} E_3^{(\omega)}$. Другими словами, благодаря нелинейности обыкновенная волна частоты ω порождает колебания поляризации, которые происходят с удвоенной частотой вдоль оптической оси (индексы сверху подчеркивают, что колебания E_2 и E_3 происходят с частотой ω , а проекции вектора поляризации на оптическую ось P_1 — с частотой 2ω). Колебания поляризации

$P_1^{(2\omega)}$, параллельные оптической оси, порождают световую волну удвоенной частоты (вторую гармонику), в которой колебания вектора \mathbf{E} также параллельны оптической оси, т. е. вторая гармоника в кристалле является необыкновенной волной. Далее, в некоторых кристаллах можно выбрать направление распространения, в котором фазовая скорость необыкновенной волны частоты 2ω оказывается равной фазовой скорости обыкновенной волны частоты ω . Рисунок 11.4 иллюстрирует такую возможность.

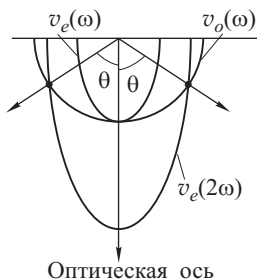


Рис. 11.4

Например, в отрицательном одноосном кристалле дигидрофосфата калия KN_2PO_4 (краткое название KDP), имеется направление распространения (целый конус направлений), где волновая поверхность

обыкновенной волны частоты ω (сфера $v_o(\omega)$) пересекается с волновой поверхностью частоты 2ω необыкновенной волны (эллипсоид $v_e(2\omega)$), т. е. в этом направлении, составляющем угол θ с оптической осью (в кристалле KDP $\theta = 41,5^\circ$) фазовая скорость исходной (обыкновенной) волны в кристалле совпадает со скоростью необыкновенной волны частоты 2ω , излучаемой волной поляризации. Для этого направления условие фазового синхронизма $k_2 = 2k_1$ выполняется при любом z , а это означает, что на всем пути распространения волны второй гармоники в кристалле ее амплитуда нарастает (происходит перекачка энергии из исходной волны). Таким образом удастся существенно увеличить коэффициент преобразования исходной волны во вторую гармонику. Так, в опытах с кристаллом KDP, в которых используется лазер на неодимовом стекле ($\lambda_1 = 1,06$ мкм — инфракрасное излучение) во вторую гармонику ($\lambda_2 = 0,53$ мкм — зеленый свет) преобразуется до 30% энергии исходной волны.

11.3. Генерация третьей гармоники. Самофокусировка и самоканализация

Рассмотрим теперь эффекты, возникающие благодаря кубической нелинейной поправке в (11.2):

$$\mathbf{f} = -k\mathbf{r} + \beta\mathbf{r}^3. \quad (11.15)$$

Такого рода зависимость проявляется в изотропных средах, где квадратичная нелинейность $-l\mathbf{r}^2$ отсутствует (обратим внимание, что изменение знака r приводит к изменению знака силы, но не меняет ее абсолютную величину, что характерно для изотропной среды).

Уравнение движения осциллятора имеет в этом случае вид

$$m\ddot{r} + kr - \beta r^3 = E_0 \cos \omega t.$$

Мы по-прежнему полагаем, что затухание отсутствует, а поле световой волны меняется по гармоническому закону с частотой ω .

Переходя, как и ранее, к уравнению для поляризации $P = Ner$, получаем

$$\ddot{P} + \omega_0^2 P - \frac{\beta}{mN^2 e^2} P^3 = \frac{Ne^2}{m} E_0 \cos \omega t. \quad (11.16)$$

Вновь используем метод последовательных приближений: будем искать решение (11.16) в виде

$$P(t) = P_0(t) + P_3(t), \quad (11.17)$$

где $P_3 \ll P_0$ — малая поправка к решению $P_0(t)$ линейного уравнения (11.5), которое, как мы знаем, описывается функцией (11.6). Подставляя решение (11.17) в уравнение (11.16) и оставляя в нелинейном слагаемом $(P_0 + P_3)^3$ лишь член P_0^3 (поскольку остальные слагаемые малы в сравнении с P_0^3), получим для нелинейной поправки P_3 линейное дифференциальное уравнение

$$\ddot{P}_3 + \omega_0^2 P_3 = \frac{\beta}{mN^2 e^2} \chi_0^3 E_0^3 \cos^3 \omega t. \quad (11.18)$$

Точно такой же вид имеет уравнение линейного осциллятора, вынужденные колебания которого возбуждаются внешней силой, меняющейся со временем пропорционально $\cos^3 \omega t$.

Поскольку $\cos^3 \omega t = (3 \cos \omega t + \cos 3\omega t)/4$, то правая часть (11.18) — «внешняя сила» — представляется суммой двух сил, одна из которых меняется с частотой ω , а вторая — с утроенной частотой 3ω . Соответственно, и вынужденные колебания $P_3(\omega)$ также являются суммой гармонических колебаний с частотами ω и 3ω . Найдя решение, отвечающее каждому из двух слагаемых правой части (11.18), получим в результате полное выражение для поляризации среды $P(t)$, которое с учетом линейного слагаемого $P_0(t) = \varepsilon_0 \chi_0 E_0 \cos \omega t$ имеет вид

$$P(t) = \varepsilon_0 \chi_0 \left(1 + b_1 E_0^2 \right) E_0 \cos \omega t + b_2 E_0^3 \cos 3\omega t, \quad (11.19)$$

где b_1 и b_2 — константы, определяемые равенствами

$$b_1 = \frac{3\beta\chi_0^2}{4\varepsilon_0 m N^2 e^2 (\omega_0^2 - \omega^2)}; \quad b_2 = \frac{\beta\chi_0^3}{4m N^2 e^2 (\omega_0^2 - 9\omega^2)}.$$

Таким образом, кубичная ангармоничность (т. е. поправка к линейной зависимости $f = -kr$, пропорциональная кубу смещения r) приводит к двум эффектам.

Во-первых, изменяется поляризуемость среды на частоте ω световой волны:

$$\chi = \chi_0 + \chi_1 E_0^2, \quad \text{где } \chi_1 = \chi_0 b_1, \quad (11.20)$$

причем это изменение пропорционально квадрату напряженности поля E_0 (т. е. пропорционально интенсивности волны); χ_0 — линейная поляризуемость среды, т. е. поляризуемость, не зависящая от напряженности поля волны в среде.

Во-вторых, возникают колебания поляризации с частотой 3ω . Эти колебания генерируют в среде волну утроенной частоты — нелинейный эффект, называемый *генерацией третьей гармоники*.

Остановимся более подробно на первом из упомянутых эффектов.

Будем рассматривать лишь волну, частота которой равна частоте ω падающей на среду волны. Поляризуемость среды на этой частоте определяется формулой (11.20), соответственно диэлектрическая проницаемость и показатель преломления среды на частоте ω есть

$$\varepsilon = 1 + \chi = (1 + \chi_0) + \chi_1 E_0^2, \quad n = \sqrt{(1 + \chi_0) + \chi_1 E_0^2}.$$

Считая поправку к линейному показателю преломления $n_0 = \sqrt{1 + \chi_0}$ малой, получим

$$n = n_0 + n_1 E_0^2, \quad (11.21)$$

где $n_1 = \chi_1 / (2n_0)$.

Рассмотрим пучок света ограниченного сечения частоты ω распространяющийся в нелинейной среде вдоль оси z (рис. 11.5). Согласно (11.21) в пределах сечения пучка показатель преломления под действием поля в пучке увеличивается ($n > n_0$); вне

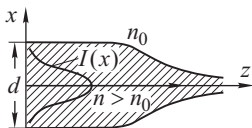


Рис. 11.5

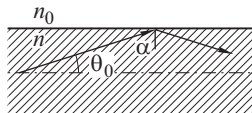


Рис. 11.6

пучка при $|x| > d/2$ показатель преломления остается равным n_0 (рис. 11.6). В лазерных пучках света интенсивность обычно максимальна у оси пучка (на оси z) и спадает к краям. Соответственно и показатель преломления n максимален на оси, постепенно уменьшаясь к краям, т. е. среда под действием све-

тового пучка становится подобной собирающей линзе. Диаметр пучка по мере распространения в среде постепенно уменьшается так, как если бы на его пути находилась фокусирующая линза. Этот нелинейный эффект называется *самофокусировкой*.

Разумеется, фокусирующему воздействию нелинейной среды препятствует дифракционная расходимость пучка: в обычной (линейной) среде с показателем преломления n пучок сечения d с плоским волновым фронтом не может распространяться, оставаясь параллельным пучком неизменного сечения. Угловая расходимость пучка, определяемая дифракцией, равна $\sin \theta_0 \approx \lambda/d = \lambda_0/(nd)$, где λ_0 — длина волны в вакууме. Смысл последнего равенства состоит в следующем: пучок света ограниченного сечения может быть представлен суммой плоских волн разных направлений, причем в этой сумме заметную амплитуду имеют лишь плоские волны, направления которых (углы с осью z) ограничены углом θ_0 . Если для волны, распространяющейся под углом θ_0 , выполнено условие полного внутреннего отражения на границе пучка и окружающей среды: $\sin \alpha_{\text{пред}} = n_0/n$ (рис. 11.6), то и для всех волн, направления которых $\theta < \theta_0$, это условие также будет выполнено, т. е. дифракционная расходимость пучка оказывается полностью скомпенсированной фокусирующим эффектом полного отражения. Получаем

$$\sin \alpha_{\text{пред}} = \cos \theta_0 = \frac{n_0}{n_0 + n_1 E_0^2}. \quad (11.22)$$

Поскольку разность показателей преломления $n - n_0 = n_1 E_0^2$ мала, то $\cos \theta_0$ близок к единице. Тогда $\cos \theta_0 \approx 1 - \theta_0^2/2$. Правую часть равенства (11.22) также приближенно представим в виде

$$\frac{n_0}{n_0 + n_1 E_0^2} = \frac{1}{1 + (n_1/n_0) E_0^2} \approx 1 - \frac{n_1}{n_0} E_0^2,$$

откуда получаем $\theta_0^2 \approx (2n_1/n_0) E_0^2$. Используя далее выражение для угла дифракционной расходимости $\sin \theta_0 \approx \theta_0 = \lambda_0/n_0 d$, находим

$$E_0^2 \approx \frac{\lambda_0^2}{2n_0 n_1 d^2}. \quad (11.23)$$

Последняя формула определяет пороговое значение напряженности поля в световой волне, при которой пучок распространяется в нелинейной среде без расходимости (т. е. дифракционная расходимость компенсируется полным внутренним отражением) и сечение пучка остается неизменным. Этот нелинейный эффект называется *самоканализацией*.

Свяжем напряженность поля E_0 с мощностью пучка, т. е. с потоком энергии, проходящим через сечение пучка в единицу времени. Мы имеем $S = \varepsilon_0 c^2 E B$ — абсолютная величина век-

тора Пойнтинга и, поскольку $B = En/c$, то $\langle S \rangle = \varepsilon_0 cn \langle E^2 \rangle = cn E_0^2 \varepsilon_0 / 2$. Эта формула определяет поток через единичную площадку. Полный поток через сечение пучка диаметра d равен $P = \langle S \rangle \pi d^2 / 4 = \pi \varepsilon_0 cn E_0^2 d^2 / 8$, откуда $E_0^2 = 8P / (\pi \varepsilon_0 c n d^2)$ и, используя (11.23), находим окончательно

$$P_{\text{пор}} \approx \frac{\pi \varepsilon_0 c \lambda_0}{16 n_1}. \quad (11.24)$$

Например, пороговая мощность излучения рубинового лазера ($\lambda = 694,3$ нм) в сероуглероде равна приблизительно $P_{\text{пор}} \approx 20$ кВт.

11.4. Параметрические процессы в нелинейных средах

Пусть в среде с квадратичной нелинейностью (11.3) распространяются две монохроматические волны с частотами ω_1 и ω_2 : $E_1(t) = E_1 \cos \omega_1 t$ и $E_2(t) = E_2 \cos \omega_2 t$. Повторяя выкладки проведенные в гл. 11.2, вновь придем к уравнению (11.9), определяющему поправку $P_1(t)$ к линейному приближению $P_0(t)$. Последнее в нашем случае является решением уравнения

$$\ddot{P}_0 + \omega_0^2 P_0 = \frac{Ne^2}{m} (E_1 \cos \omega_1 t + E_2 \cos \omega_2 t).$$

Решение этого (линейного) уравнения является, очевидно, суммой двух решений, каждое из которых отвечает «своей» правой части

$$P_0(t) = \varepsilon_0 \chi_1 E_1 \cos \omega_1 t + \varepsilon_0 \chi_2 E_2 \cos \omega_2 t,$$

где χ_1 и χ_2 — диэлектрические восприимчивости на частотах ω_1 и ω_2

$$\chi_1 = \frac{Ne^2}{\varepsilon_0 m (\omega_0^2 - \omega_1^2)}, \quad \chi_2 = \frac{Ne^2}{\varepsilon_0 m (\omega_0^2 - \omega_2^2)}.$$

Подставляя это решение в (11.9), находим

$$\ddot{P}_1 + \omega_0^2 P_1 = \frac{l \varepsilon_0}{N e m} (\chi_1 E_1 \cos \omega_1 t + \chi_2 E_2 \cos \omega_2 t)^2.$$

Наличие в правой части слагаемых, содержащих квадраты полей $E_1(t)$, и $E_2(t)$ приводит к уже рассмотренным нами эффектам: генерации вторых гармоник с частотами $2\omega_1$ и $2\omega_2$ и оптическому выпрямлению (появлению постоянного электрического поля).

Новый эффект, который здесь возникает, связан с наличием в правой части слагаемого, содержащего произведение $\cos \omega_1 t \cos \omega_2 t$, что эквивалентно сумме двух колебаний с частотами $\omega_1 + \omega_2$ и $\omega_1 - \omega_2$. Соответственно нелинейная поправка к поляризации $P_1(t)$ содержит колебания на частотах $\omega_1 + \omega_2$

и $\omega_1 - \omega_2$, что приводит к переизлучению на этих частотах, причем амплитуды волн пропорциональны произведению $E_1 E_2$ амплитуд исходных волн с частотами ω_1 и ω_2 . Генерация волн с суммарной и разностной частотой позволяет получать когерентное излучение в более широком диапазоне частот. Кроме того, появляется возможность, смешав в нелинейном кристалле инфракрасное излучение (ИК) с видимым светом, преобразовать частоту ИК-диапазона в видимый спектр, где шумовые характеристики существующих приемников света существенно лучше приемников ИК-области спектра.

Качественно эффект возникновения комбинационных частот в излучении, прошедшем нелинейную среду, можно понять следующим образом. Пусть волна частоты ω_1 имеет достаточно большую амплитуду, так что ее распространение в среде приводит к изменению оптических свойств среды с частотой ω_1 , т. е. под действием этой волны возникает модуляция показателя преломления среды, а следовательно, модуляция фазы второй волны. Мы знаем, что спектр колебания, имеющего несущую частоту ω_2 , фаза которого промодулирована с частотой ω_1 , содержит компоненты с боковыми частотами $\omega_1 \pm \omega_2$.

Сказанное выше позволяет понять идею параметрического усиления света, предложенную С. А. Ахмановым и Р. В. Хохловым в 1962 году.

Пусть в среде распространяются волны с частотами ω_1 и ω_2 и пусть волна частоты ω_1 имеет большую амплитуду, приводящую к нелинейному изменению свойств среды (возникновению квадратичной нелинейности в кристалле). Назовем мощную волну E_1 частоты ω_1 *волной накачки*, а слабую волну E_2 частоты ω_2 *сигнальной волной*. Это приводит, как мы видели, к появлению нелинейной добавки $P_1(t)$ к поляризации среды $P_0(t)$, которая осциллирует с частотой $\omega_1 - \omega_2 = \omega_3$. Колебания поляризации с частотой ω_3 индуцируют в среде вторичные волны — переизлучение на этой частоте. Волна поляризации $P_1(\mathbf{r}, t) = A_1 \cos(\omega_3 t - (\mathbf{k}_1 - \mathbf{k}_2)\mathbf{r})$ распространяется в среде с фазовой скоростью $v_\phi = \omega_3/k$, где $\mathbf{k} = \mathbf{k}_1 - \mathbf{k}_2$. Направим теперь на нелинейный кристалл, в котором распространяются волны с частотами ω_1 и ω_2 , третью волну малой интенсивности, частота которой равна $\omega_1 - \omega_2 = \omega_3$. Эту волну называют *холостой*. Выберем направление холостой волны (волновой вектор \mathbf{k}_3) так, чтобы выполнялось условие $\mathbf{k}_3 = \mathbf{k}_1 - \mathbf{k}_2$. Тогда ее уравнение имеет вид

$$E_3(t, \mathbf{r}) = A_3 \cos(\omega_3 t - \mathbf{k}_3 \mathbf{r}),$$

а ее фазовая скорость $v_\phi = \omega_3/k_3$ совпадает с фазовой скоростью волны поляризации.

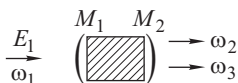
Переизлученные волной поляризации $P_1(\mathbf{r}, t)$ вторичные волны оказываются при этом точно сфазированными с волной $E_3(\mathbf{r}, t)$ (выполнено условие фазового синхронизма):

$$\mathbf{k}_3 = \mathbf{k}_1 - \mathbf{k}_2. \quad (11.25)$$

Амплитуда колебаний волны поляризации пропорциональна амплитуде волны накачки (частоты ω_1), поэтому и вторичные волны частоты ω_3 также пропорциональны ее амплитуде. В результате исходная слабая волна E_3 , благодаря фазовому синхронизму со вторичными волнами той же частоты, усиливается по мере прохождения в нелинейной среде, черпая свою энергию из исходной сильной волны — явление называется *параметрическим усилением света*.

Поместим нелинейный кристалл, например, кристалл ниобата лития LiNbO_3 (KDP) между зеркалами M_1 и M_2 оптического резонатора (рис. 11.7). Кристалл ориентируем таким образом, чтобы для волны E_1 (волны накачки частоты ω_1), а также некоторых волн E_2 и E_3 , распространяющихся вдоль оси резонатора перпендикулярно поверхности зеркал M_1 и M_2 , частоты которых ω_2 и ω_3 удовлетворяют условию резонанса, выполнялось условие фазового синхронизма (11.25). Волна накачки E_1 от внешнего лазера попадает на кристалл через зеркало M_1 , прозрачного для волны E_1 . Слабое излучение E_2 и E_3 (с частотами ω_2 и ω_3) существует в резонаторе благодаря колебаниям осцилляторов, возбужденных тепловым движением. Для волн E_2 и E_3 , частоты которых удовлетворяют условию $\omega_2 + \omega_3 = \omega_1$, а волновые векторы — условию фазового синхронизма $\mathbf{k}_2 + \mathbf{k}_3 = \mathbf{k}_1$, возникает усиление первоначально слабого спонтанного излучения на частотах ω_2 и ω_3 . Если при достаточно мощной волне накачки усиление превосходит потери за проход (включая потери при отражении от зеркал, поглощение и рассеяние), происходит самовозбуждение генератора. Изменение ориентации кристалла приводит к тому, что условие фазового синхронизма выполняется для другой пары частот ω_2 и ω_3 , так как благодаря изменению показателя преломления необыкновенной волны, распространяющейся вдоль оси резонатора, изменяются волновые числа. Таким образом, может осуществляться плавная перестройка частоты генерации. Генерация волн с частотами ω_2 и ω_3 происходит благодаря модуляции параметров нелинейной среды под действием поля волны накачки, что и обуславливает название — *параметрическая генерация*.

Рис. 11.7



В связи со сказанным выше имеет смысл вспомнить о параметрической расщепке механических или электрических колебаний (см. гл. 3).

11.5. Комбинационное рассеяние света

В линейном приближении квазиупругая сила, действующая на оптический электрон со стороны поля ядра, может быть представлена как производная от потенциальной энергии $U(r) = kr^2/2$: $f = -\partial U/\partial r = -kr$.

Если смещение r электрона не мало, то в выражении для потенциальной энергии следует учитывать следующий член разложения в степенной ряд по r потенциальной энергии:

$$U(r) = \frac{kr^2}{2} - \frac{1}{3}lr^3, \quad (11.26)$$

откуда получается квадратичная поправка к квазиупругой силе

$$f = -\frac{\partial U}{\partial r} = -kr + lr^2.$$

При этом следует учесть, что потенциальная энергия в положении равновесия $r = 0$ принимается равной нулю и, кроме того, линейный по r член в выражении $U(r)$ отсутствует, поскольку положение равновесия соответствует минимуму потенциальной энергии, т. е. $(\partial U/\partial r)_0 = 0$.

Изучая поведение оптических электронов в поле световой волны, мы полагали, что из-за большой массы ядра атомов не смещаются. Рассмотрим теперь эффекты, возникающие при учете этого смещения. Пусть смещение ядра относительно узла кристаллической решетки, где оно находится в отсутствие поля, либо относительно положения равновесия в молекуле жидкости или газа есть R . При учете этого смещения потенциальная энергия (как функция двух переменных r и R) может быть записана в виде

$$U(r, R) = \frac{kr^2}{2} + \frac{KR^2}{2} + \frac{1}{3}lr^3 + \frac{1}{3}LR^3 + \beta r^2 R + \alpha r R^2. \quad (11.27)$$

Используя это выражение, получаем силу, действующую на электрон. Она имеет вид

$$f = -\frac{\partial U(r, R)}{\partial r} = -kr - lr^2 - 2\beta r R - \alpha R^2.$$

Эффект, связанный с появлением квадратичной поправки по r мы уже обсуждали. Он отличен от нуля лишь в анизотропной среде и приводит к генерации второй гармоники. Здесь мы его учитывать не будем. Последнее слагаемое зависит только от закона колебания ядра $R(t)$, которое из-за большой массы имеет низкие частоты, далеко выходящие за рамки оптического диапазона (соответствующее излучение осцилляторов приходится на далекую инфракрасную область). Это слагаемое мы также рассматривать не будем.

В результате остаются два слагаемых

$$f = -kr - 2\beta rR,$$

причем второе (нелинейное) слагаемое является малой поправкой к квазиупругой силе $-kr$.

Используя полученное выражение, имеем уравнение движения электрона

$$m\ddot{r} + (k + 2\beta R)r = eE_0 \cos \omega t.$$

Вводя эквивалентную жесткость квазиупругой силы $k(R) = k + 2\beta R$, имеем

$$\ddot{r} + \omega_0^2(R)r = \frac{e}{m} E_0 \cos \omega t,$$

где зависящая от движения ядра собственная частота осциллятора равна

$$\omega_0^2(R) = \frac{k}{m} + \frac{2\beta R}{m} = \omega_0^2 + \frac{2\beta R(t)}{m}. \quad (11.28)$$

Переходя от уравнения для r к уравнению для поляризации $P = eNr$, получаем

$$\ddot{P} + \omega_0^2(R)P = \frac{Ne^2}{m} E_0 \cos \omega t. \quad (11.29)$$

Обратимся теперь к уравнению, описывающему колебания ядра массы $M \gg m$. Упругую силу, действующую на ядро, найдем, дифференцируя выражение для энергии (11.27) по координате R :

$$F = -\frac{\partial U}{\partial R} = -KR - LR^2 - 2\alpha rR - \beta r^2.$$

Априори ясно, что смещение ядра мало, поэтому членом $-LR^2$ можно пренебречь. Третье слагаемое не существенно, так как произведение rR меняется с частотой, близкой к частоте колебаний электрона, которая велика по сравнению с частотой колебаний тяжелого ядра (другими словами, за время, равное периоду колебаний ядра, это слагаемое в силе $F(t)$ множество раз изменяет знак, поэтому его усредненное влияние ничтожно). Остаются лишь два слагаемых:

$$F = -KR - \beta r^2. \quad (11.30)$$

Как следует из последнего равенства, влияние движения электрона $r(t)$ на колебания ядра существенно лишь в случае сильного поля, когда необходимо учитывать нелинейное слагаемое. В слабых полях (обычных тепловых источников света) им можно также пренебречь. В этом случае уравнение движения ядра имеет вид

$$m\ddot{R} = -KR + eE_0 \cos \omega t.$$

Наконец, влиянием поля световой волны высокой частоты ω пренебрежем (так же, как мы пренебрегли быстро меняющимся

слагаемым $-2\alpha r R$), и мы получаем уравнение свободных колебаний

$$\ddot{R} + \Omega^2 R = 0,$$

где $\Omega^2 = K/M$. Возбуждаются эти колебания благодаря тепловому движению атомов и молекул, их средняя энергия пропорциональна температуре.

$$R(t) = R_0 \cos \Omega t.$$

Для молекул частота Ω лежит в инфракрасном диапазоне спектра.

Вернувшись теперь к уравнениям (11.28) и (11.29), мы видим, что колебания ядра (или совокупности ядер, образующих молекулу) приводят к модуляции собственной частоты $\omega_0(R)$ осциллятора, частота модуляции равна Ω , а глубина модуляции — отношение амплитуды переменного слагаемого к постоянному слагаемому $-\beta R_0/(m\omega_0^2)$, обычно много меньше единицы. Поэтому в (11.29) в слагаемом $\omega_0^2(R)P$, которое содержит малую поправку к ω_0^2 , заменим поляризацию P «невозмущенным» решением $P_0 = \varepsilon_0 \chi(\omega) E_0 \cos \omega t$.

В результате искомое уравнение для поляризации примет вид

$$\ddot{P} + \omega_0^2 P = \frac{Nq^2}{m} E_0 \cos \omega t - \frac{2\beta\varepsilon_0}{m} \chi E_0 \cos \omega t \cdot R_0 \cos \Omega t. \quad (11.31)$$

Правая часть уравнения (11.31), которая играет роль внешней силы, возбуждающей вынужденные колебания поляризации, содержит три слагаемых с частотами ω , $\omega + \Omega$ и $\omega - \Omega$, поскольку $\cos \omega t \cos \Omega t = (1/2)(\cos(\omega + \Omega)t + \cos(\omega - \Omega)t)$. Соответственно, вынужденные колебания поляризации содержат слагаемые на частотах ω и $\omega \pm \Omega$. Возбуждаемые этими осцилляторами вторичные волны имеют соответственно частоты ω и $\omega \pm \Omega$. Рассеяние первичной волны, содержащее, кроме основной частоты ω , также и комбинационные частоты $\omega \pm \Omega$, называется *комбинационным рассеянием*.

Излучение на частоте $\omega - \Omega$ называется *красным сателлитом* или *стоксовой* компонентой (в честь английского физика Дж. Стокса (1819–1903), впервые наблюдавшего это излучение). Излучение на частоте $\omega + \Omega$ называется *фиолетовым сателлитом* или *антистоксовой* компонентой.

11.6. Вынужденное комбинационное рассеяние

Выше мы рассмотрели комбинационное излучение, обусловленное воздействием на колебания молекул колебаний ядер атомов, образующих молекулу. При этом в выражении (11.30) для

упругой силы мы пренебрегли слагаемым

$$\beta r^2 = \frac{\beta}{\epsilon N^2} P^2, \quad (11.32)$$

где $P = Ner$ — поляризация, обусловленная смещением r оптического электрона.

Пусть теперь источником возбуждения является сильное световое поле (мощный импульс излучения лазера), а нелинейная среда, в которой возникает комбинационное излучение, помещена в резонатор. Две причины заставляют нас учесть теперь влияние этого слагаемого.

Учтя ранее влияние колебания ядра на колебание электрона, мы выяснили, что поляризация среды состоит из слагаемых на частотах ω и $\omega \pm \Omega$. Учтем далее лишь стоксову компоненту, т. е. будем полагать

$$P(t) = P^{(\omega)}(t) + P^{(\omega-\Omega)}(t).$$

Ясно, что при этом P^2 содержит слагаемое, осциллирующее на разностной частоте $\omega - (\omega - \Omega) = \Omega$. Эта частота совпадает с резонансной частотой колебаний молекулы и может вызвать сильную раскачку этих колебаний. Разумеется, воздействие на этой частоте имело место и при тепловом возбуждении, однако комбинационное излучение, возбужденное тепловым движением, имеет случайную фазу, т. е. фазы колебаний разных молекул не согласованы между собой, поэтому и не возникает эффективного накопления излучения на комбинационной частоте. При помещении нелинейной среды в резонатор, комбинационное излучение, которое распространяется вдоль оси резонатора, отражается от зеркал и, следовательно, вновь возвращается в нелинейную среду, оказывая обратное влияние на колебания молекул: их фазы колебаний оказываются согласованными, а комбинационное излучение когерентно. Такое комбинационное излучение называется *вынужденным* или *стимулированным*.

Ситуация отличается от предыдущей в двух отношениях. Во-первых, увеличивается амплитуда колебаний внешнего воздействия на молекулу на резонансной частоте Ω (она пропорциональна амплитуде колебаний поляризации на частоте ω и, следовательно, амплитуде колебания светового поля в импульсе излучения лазера).

Во-вторых, комбинационное излучение, распространяющееся вдоль оси резонатора, отражаясь от зеркал, накапливается в резонаторе и «синхронизирует» колебания молекул: их комбинационное излучение оказывается когерентным (при когерентном сложении интенсивность пропорциональна N^2 , в отличие от некогерентного, при котором интенсивность пропорциональна N).

Схема на рис. 11.8 представляет собой генератор излучения на комбинационной частоте Ω и может быть названа комбинационным лазером. Между зеркалами резонатора (полностью отражающего M_1 , и частично прозрачного M_2) помещается кристалл рубина 1, являющийся источником импульса на частоте ω , сосуд с нелинейной средой (бензолом) 2 и ячейка Керра 3, служащая для получения гигантских импульсов излучения (напряженности 10^7 – 10^8 В/см). В комбинационное излучение преобразуется до 20% энергии импульса на основной частоте ω .

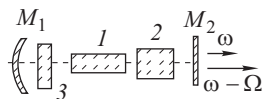


Рис. 11.8

11.7. Вынужденное рассеяние Мандельштама–Бриллюэна (ВРМБ)

В 1964 г. был экспериментально обнаружен эффект, предсказанный ранее советским физиком Л.И. Мандельштамом (1879–1944) и французским физиком Л. Бриллюэном (1889–1969): мощное лазерное излучение частоты ω , попадая на кристалл, индуцирует в нем появление упругой (акустической) волны частоты Ω с одновременным рассеянием света на частоте $\omega - \Omega$.

Качественное объяснение этого эффекта с классической точки зрения таково. Из-за сильной связи атомов друг с другом в твердых телах и жидкостях тепловое колебательное движение одного атома неизбежно передается другому. Такое движение можно рассматривать как совокупность упругих (акустических) волн, распространяющихся во всевозможных направлениях и имеющих всевозможные частоты, спектр которых чрезвычайно широк: от звуковых ($\sim 10^2$ Гц) до гиперзвуковых ($\sim 10^{10}$ Гц). Акустические волны в среде — это волны плотности (или давления), появление которых влияет и на оптические свойства среды: изменения плотности, с «оптической» точки зрения, это изменения числа осцилляторов в единице объема что, согласно классической теории дисперсии, приводит к изменениям (флуктуациям) показателя преломления $\delta n(\mathbf{r}, t)$, имеющих волновой характер.

Итак, тепловое движение приводит к появлению в среде волн показателя преломления, которые можно представить в виде совокупности бегущих плоских волн различных частот Ω и направлений \mathbf{K} . Любая из волн в этой совокупности может быть записана в виде

$$\delta n(\mathbf{r}, t) = \delta n_0 \cos(\Omega t - \mathbf{K}\mathbf{r} - \theta), \quad (11.33)$$

где волновое число есть, по определению, $K = \Omega/v_{\text{зв}}$ ($v_{\text{зв}}$ — скорость звука в среде), т. е. каждая волна показателя преломления

$\delta n(\mathbf{r}, t)$ представляет собой бегущую фазовую синусоидальную решетку; θ — начальная фаза.

Что же происходит, когда световая волна большой интенсивности $E_0(\mathbf{r}, t) = a_0 \cos(\omega t - \mathbf{k}\mathbf{r})$ падает на среду, состоящую из набора таких бегущих решеток? Чтобы в этом разобраться, вернемся к проблемам, которые мы обсуждали, рассматривая процесс записи и восстановления объемных голограмм (голограмм Денисюка).

Прежде всего заметим, что бегущую интерференционную картину $I(\mathbf{r}, t)$ в объемной среде можно создавать с помощью интерференции опорной волны $E_0(r, t)$ с «предметной» волной

$$E_1(\mathbf{r}, t) = a_1 \cos[(\omega - \Omega)t - (\mathbf{k} - \mathbf{K})\mathbf{r} - \theta]. \quad (11.34)$$

Суммарная картина интенсивности

$$I(\mathbf{r}, t) = \overline{|E_0(\mathbf{r}, t) + E_1(\mathbf{r}, t)|^2}$$

(где черта означает усреднение за время, существенно большее периода светового колебания и много меньшее периода «биений» $2\pi/\Omega$) имеет вид

$$I(r, t) = a_0^2 + a_1^2 + 2a_0a_1 \cos(\Omega t - \mathbf{K}\mathbf{r}). \quad (11.35)$$

Напомним, что можно, находя сумму гармоничных колебаний с близкими частотами, воспользоваться векторной диаграммой: интенсивность суммарного колебания (квадрат длины суммарного вектора) найти с помощью теоремы косинусов.

Обратите внимание, что последнее слагаемое в (11.35) описывается точно таким же соотношением, что и (11.33). Ясно поэтому, что если фотообработку полученной голограммы осуществить таким образом, чтобы голограмма стала «фазовой», т. е. чтобы оптическая плотность (т. е. показатель преломления) объемной среды после обработки стала пропорциональна интенсивности суммарной волны при записи голограммы, то наша голограмма при восстановлении изображения ничем не будет отличаться от среды, в которой возбуждена бегущая решетка (11.33). Разумеется, мы говорим сейчас о некоторой гипотетической голограмме с изменяющейся во времени функцией пропускания. Мы знаем, что голограмма в процессе восстановления изображения, при освещении ее волной, совпадающей с опорной волной, которая была использована при записи, восстанавливает предметную волну.

Ясно поэтому, что волна $E_0(r, t)$, дифрагируя на бегущей фазовой решетке вида (11.35), дает «рассеянную» волну, которая отличается от (11.34) лишь встречным направлением (встречным перемещением волновых фронтов). Поскольку фазовая решетка, созданная тепловыми флуктуациями, чрезвычайно слаба

($\delta n_0 \ll n_0$), то и дифрагировавшая волна имеет чрезвычайно малую амплитуду, если амплитуда волны $E_0(\mathbf{r}, t)$ не достаточно велика.

Поскольку акустические волны (11.33) разных частот и направлений имеют случайную начальную фазу θ (они созданы тепловым движением, имеющим случайный характер), то и рассеянные разными решетками волны некогерентны — это так называемое *спонтанное рассеяние*, аналогично излучению обычных тепловых источников, которое образовано спонтанным, несогласованным излучением разных атомов.

Чрезвычайно важно, далее, обратное воздействие на среду электрического поля волны $E_0(\mathbf{r}, t)$ большой интенсивности, которая способна влиять на оптические свойства среды (изменять ее показатель преломления). Интерферируя с первоначально слабой, дифрагировавшей на «тепловой» решетке волной $E_1(\mathbf{r}, t)$, эта волна создает в среде бегущую решетку, которая ничем не отличается от уже имеющейся в среде решетки (которая и «породила» волну $E_1(\mathbf{r}, t)$). Дополнительное давление в среде и, следовательно, дополнительное изменение ее показателя преломления пропорционально квадрату напряженности суммарного поля $E^2 = (E_0 + E_1)^2$ — нелинейный эффект, который мы уже обсуждали, рассматривая явление самофокусировки и самоканализации. Таким образом, «затравочная», первоначально слабая волна показателя преломления когерентно усиливается бегущими волнами, образованными интерференцией интенсивной волны E_0 и рассеянной волны E_1 . Последняя, благодаря увеличению амплитуды δn_0 волны показателя преломления, также когерентно усиливается (ведь она возникает в результате дифракции на фазовой решетке со все большей глубиной модуляции фазы $\delta n_0/n_0$). Таким образом, как амплитуда акустической волны, так и амплитуда рассеянной волны частоты $\omega - \Omega$ будут постепенно нарастать по мере распространения в нелинейной среде — их энергия черпается из исходной волны $E_0(\mathbf{r}, t)$ — волны накачки. В отличие от некогерентного рассеяния на тепловых флуктуациях показателя преломления, вынужденное мощной волной накачки рассеянное излучение является когерентным.

11.8. Обращение волнового фронта

Смысл термина волна с «обращенным» волновым фронтом состоит в том, что в обращенной волне волновые поверхности (поверхности одинаковой фазы колебаний), имея ту же пространственную форму, перемещаются во встречном направлении, т. е. в каждой точке волнового фронта волновой вектор \mathbf{k} имеет противоположное направление. Можно сказать, что обращенная

волна повторяет ход прямой волны, но в обратной последовательности — «обращает ход событий во времени».

Математически это выражается в том, что комплексная амплитуда обращенной волны $f_{\text{обр}}(\mathbf{r}, t)$ в каждой точке $\mathbf{r}(x, y, z)$ связана с комплексной амплитудой прямой волны $f(\mathbf{r})$ равенством $f_{\text{обр}}(\mathbf{r}) = f^*(\mathbf{r})$. Например, для плоской волны с комплексной амплитудой $ae^{i(\mathbf{k}\mathbf{r}+\alpha)}$ (α — начальная фаза), обращенной является волна $ae^{-i(\mathbf{k}\mathbf{r}+\alpha)}$, что подразумевает изменение знака всех трех компонент вектора \mathbf{k} , а также изменение знака начальной фазы α . Расходящаяся сферическая волна $(a_0/r)e^{i(kr+\alpha)}$ после обращения становится сходящейся волной $(a_0/r)e^{-i(kr+\alpha)}$.

Вспомним, что волне с комплексной амплитудой $f(\mathbf{r}) = a(\mathbf{r})e^{i\varphi(\mathbf{r})}$ соответствует реальный волновой процесс $E(\mathbf{r}, t) = a(\mathbf{r})\cos(\omega t - \varphi(\mathbf{r}))$, т. е. обращение волнового фронта означает, что реальная волна $E(\mathbf{r}, t)$ превращается в волну $E_{\text{обр}}(\mathbf{r}, t) = a(\mathbf{r})\cos(\omega t + \varphi(\mathbf{r}))$. Тот же самый результат получается, если в реальном волновом процессе изменить знак времени t (заменить t на $-t$). Мы получим $E(\mathbf{r}, -t) = a(\mathbf{r})\cos(-\omega t - \varphi(\mathbf{r}))$, что (в силу четности функции $\cos \alpha$) равно $E(\mathbf{r}, -t) = a(\mathbf{r})\cos(\omega t + \varphi(\mathbf{r}))$, а это и есть обращенная волна. Поэтому-то обращенная волна воспроизводит все изменения, которые происходят с прямой волной по мере ее распространения (быть может, в неоднородной среде), но в обратном порядке.

Рисунок 11.9 напоминает нам схему записи и восстановления объемной голограммы (запись во встречных пучках). Рисунок 11.9а — схема записи: слева на голограмму падает опорная волна I , справа — предметная волна. В нашем случае предмет — точечный источник света, и соответственно, предметная волна — это «сферическая» волна, расходящаяся из источника — точки S . Направление штрихов условно показывает положение слоев интерференционных максимумов и минимумов, которые после фотообработки образовали объемную фазовую решетку (т. е. чередующиеся слои большего и меньшего показателя преломления в объемной регистрирующей среде 2).

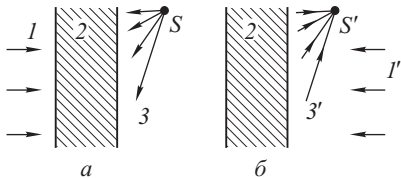


Рис. 11.9

Рисунок 11.9б — схема восстановления изображения. Восстанавливающая волна I' падает на голограмму справа, т. е.

навстречу опорной волне I при записи голограммы. Дифрагируя на объемной фазовой решетке 2 эта волна образует восстановленную волну 3 — сферическую волну, сходящуюся в точку S' .

Эта сходящаяся сферическая волна имеет «обращенный» по отношению к расходящейся предметной волне волновой фронт.

Рисунок 11.10 иллюстрирует замечательное свойство обращенной волны. Пусть на слой неоднородной среды направлен параллельный пучок света I (с плоским волновым фронтом). Тогда на выходе из слоя волновые поверхности 3 окажутся сильно искаженными, деформированными (рис. 11.10а). Если нам затем удалось обратить волновой фронт с помощью некоего «волшебного зеркала» и заставить обращенную волну $3'$ пройти тот же слой неоднородной среды 2 в обратном направлении, то на выходе из слоя мы получим хороший параллельный пучок света I' — ту же форму волнового фронта, который имела волна до того, как ее волновой фронт был «испорчен» (рис. 11.10б).

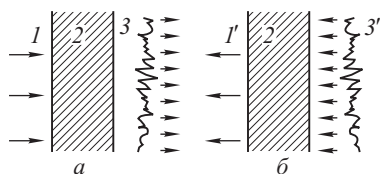


Рис. 11.10

Голография, как мы видели, дает возможность получить обращенную волну в двухэтапном процессе (сначала этап записи голограммы, а затем стадия реконструкции — восстановления изображения, восстановления обращенной волны).

При вынужденном комбинационном рассеянии обращенная волна возникает в реальном времени, непосредственно при освещении нелинейной среды волной большой интенсивности. Схема

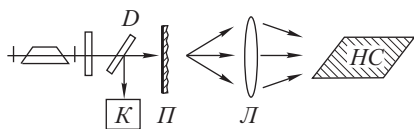


Рис. 11.11

эксперимента, подтверждающего, что при ВРМБ возникает обращенная волна, показана на рис. 11.11. Лазерный пучок большой интенсивности, имеющий хороший волновой фронт, искажается затем

матовой пластинкой Π , которая сильно деформирует форму волнового фронта. Затем с помощью линзы L , которая в основном перехватывает весь рассеянный пластинкой свет, излучение направляется на нелинейную среду HC (кювета, заполненная метаном или сероуглеродом). После обратного рассеяния (ВРМБ) свет вновь проходит через линзу и матовую пластинку. При этом искажения, которые были внесены в пучок при его распространении слева направо, компенсируются при обратном ходе волны через ту же искажающую среду, и на выходе из пластинки Π слева от нее волновой фронт снова становится «хорошим», почти плоским, что и контролируется с помощью полупрозрачного зеркала D , отводящего часть отраженного обратно излучения в систему контроля K . Понятно, что такая компенсация внесенных в пучок искажений возможна лишь по-

тому, что рассеянное назад нелинейной средой излучение точно воспроизводит форму сложного волнового фронта падающей на среду волны, лишь изменяя направление перемещения волновых фронтов на противоположное. Кювета с нелинейной средой и является тем «волшебным зеркалом», которое обращает ход событий во времени. (Если бы вы вдруг решили, что то же самое можно сделать, установив вместо кюветы с нелинейной средой самое обыкновенное зеркало, то вы убедились бы, что при обратном ходе волны через матовую пластинку искажения не скомпенсировались бы, а лишь увеличились, добавившись к уже имевшимся искажениям: обыкновенное зеркало не превращает расходящуюся сферическую волну в сходящуюся — расходящаяся волна отразившись от зеркала, остается расходящейся.)

Именной указатель

- Ампер А.М., 247, 248, 260, 262, 266, 289, 290, 292, 300
Аристотель, 14–16, 58
Архимед, 135, 153, 154, 289
Бартолин Э., 662
Басов Н.Г., 672
Бернулли Д., 153, 156, 157, 163, 164
Био Ж.Б., 247, 249, 251, 252, 255, 265, 285, 323, 339
Бор Н., 25, 270, 271, 291, 348
Браге Тихо, 123
Бриллюэн Л., 691
Брэгг Л., 635
Вейсс П.Э., 266
Вольт А., 204
Вульф Г.В., 635
Габор Д., 383, 567, 625, 628–630, 633
Галилей Г., 20, 22, 25, 38, 39, 100, 165–167, 175, 178–182, 185, 362–364
Гаусс К.Ф., 199–202, 205, 209, 211, 213, 217, 228, 234, 239, 245, 254, 285, 297, 303, 321, 326, 328, 331, 339, 346, 374, 418
Гейзенберг В., 25
Гук Р., 43, 46, 144–146, 148, 422, 488, 489, 493, 502, 503, 673
Гюйгенс Х., 132, 583–585, 589, 621, 660
Де Бройль Л., 26
Денисюк Ю.Н., 633, 634, 692
Дирак П., 245
Доплер К., 504, 506
Ирншоу С., 220
Кёниг А., 101, 105, 142
Кавендиш Ш., 196
Кеплер И., 97, 107, 123, 124, 126
Керр Дж., 664–667, 691
Кирхгоф Г., 239, 241, 565
Кобеко П.П., 223
Кориолис Г.Г., 172, 176
Корню М.А., 392, 394, 395, 581, 593–595
Коттон Э., 667
Кулон Ш., 45, 115, 195–197, 199, 201–203, 206, 216, 233, 255, 285, 323, 338, 339, 377
Курчатов И.В., 223
Кюри Ж., 222
Кюри П., 222, 266, 271
Лаплас П.С., 205, 206, 247, 347, 476, 491
Лармор Дж., 267, 268
Лейбниц Г., 28
Лейт Э., 630, 633
Лоренц Х., 84, 177, 181, 182, 184–186, 248, 261, 277, 284, 299, 343, 350, 363, 644, 667
Майкельсон А., 177, 361, 545, 559–562
Максвелл Дж.К., 192, 238, 250, 321, 323–326, 338, 341, 346, 354–356, 359, 361–364, 373, 508
Малюс Э., 663
Мандельштам Л.И., 691
Минковский Г., 185, 186
Мутон А., 667
Николь У., 664
Ньютон И., 15, 16, 24–26, 28, 41–46, 49, 50, 52–54, 56, 57, 59, 62, 65–67, 75, 76, 85, 86, 90, 91, 96, 97, 106–108, 110, 120–124, 127, 129, 132, 136, 141, 165–170, 174, 175, 185, 187, 191, 193, 247, 323, 361–363, 416–418, 425, 430, 438, 488, 490, 501, 508
Ом Г., 212, 235–239, 241, 265, 276, 285, 298, 304, 306, 311, 312, 316, 324, 367, 368, 418, 422, 426, 431, 438
Остроградский М.В., 201
Паскаль Б., 153–155, 157, 163, 289
Перо А., 518, 606, 608–611, 637
Поккельс Ф.К., 666, 667
Прохоров А.М., 672
Пуазейль Ж., 157, 160, 161
Пуассон С., 147, 150, 205, 206, 217, 226, 254, 255, 258, 381, 591, 592
Резерфорд Э., 418
Рейнольдс О., 162, 163
Савар Ф., 247, 249, 251, 252, 255, 265, 285, 323, 339
Стокс Дж., 689
Талбот У., 576, 578
Таунс Ч.Х., 672
Тейлор Б., 580, 581, 583
Тесла Н., 246
Томсон Дж.Дж., 418, 424, 425
Торричелли Э., 156, 157
Умов Н.А., 492, 494
Упатниекс Ю., 630, 633
Фабри Ш., 518, 606, 608–611, 637
Фраунгофер Й., 564, 580, 583, 594–600, 613, 619
Холл Э.Г., 277, 278, 332
Циолковский К.Э., 64, 88
Штейнер Я., 131, 141
Шустер А., 593
Эйнштейн А., 25, 177–180, 185, 189, 362
Юнг Т., 146, 150, 164, 488, 545, 553, 554, 556–558, 670

Предметный указатель

- Абсолютно неупругий удар, 116
 - твердое тело, 126
 - упругий удар, 115
- Автогенератор Ван-дер-Поля, 466
- Автоколебания, 464
- Амплитуда волны, 477
 - колебания, 385
 - — комплексная, 483
- Амплитудная модуляция, 396
 - решетка, 567
- Анизотропная среда, 653
- Атом Томсона, 418, 424
- Атомная поляризуемость, 645
- Биения, 398, 399
- Бинормаль, 33
- Брюстера угол, 360
- Вектор, 22
 - , модуль, 23
 - Пойнтинга, 353, 514, 516, 661, 662, 684
 - Умова, 494
 - плотности потока энергии, 494
 - поля скоростей, 34
 - угловой скорости, 33
- Векторное произведение, 33
- Вес, 44, 108, 109
- Видность, 531, 547, 553–556, 558, 559
- Волна, 474, 476
 - бегущая плоская, 479
 - в анизотропных средах, 653
 - векторная, 484
 - гармоническая, 477
 - линейно-поляризованная, 484, 485, 513, 514
 - монохроматическая, 482
 - накачки, 685
 - необыкновенная, 656, 659, 661
 - обыкновенная, 656, 659
 - плоская, 476, 477
 - плоскополяризованная, 484, 513, 519
 - поляризации, 685
 - поляризованная по кругу, 485
 - поперечная, 484, 501
 - предметная, 624
 - продольная, 484
 - сигнальная, 685
 - , скорость в волноводе, 526
 - стоячая, 479, 496, 501
 - — , пучности, 480, 517
 - — , узлы, 479, 517
 - — электромагнитная, 517
 - сферическая, 480, 481
 - ударная, 504
 - упругая, 487
 - — в жидкостях и газах, 490
 - — в стержне, 497
- Волна холостая, 685
 - электромагнитная, 509, 512
 - — в ионосфере, 649
 - — в металлах, 649
 - — , отражение от идеального проводника, 522
 - — , поляризация, 513
 - — , энергетические характеристики, 514
 - эллиптически поляризованная, 485
- Волноводы, 524
- Волновое поле, 563
 - число, 477
- Волновой вектор, 478
 - пакет, 640
 - параметр, 564
- Волновые поверхности, 477
- Волны в анизотропных средах, 653
 - — кристаллах, 655
 - — нелинейных средах, 672, 676
 - — плазме, 377
 - — среде, 351
 - , генерация третьей гармоники, 680, 682
 - квазимонохроматические, 535, 544
 - , параметрическая генерация, 686
 - электромагнитные, 361, 508
 - — на границе двух сред, 353, 357, 359
 - — , поляризация, 359
 - , эффект возникновения комбинационных частот, 685
 - — , удвоения частоты, 673, 676
- Время корреляции, 538, 551
 - постоянная, 428
 - регистрации, 539
- Вынужденное рассеяние Мандельштама–Бриллюэна, 691
- Вязкость, 157
 - , коэффициент, 158
- Газовый разряд, 366
 - — дуговой, 371
 - — искровой, 371
 - — коронный, 371
 - — несамостоятельный, 370
 - — самостоятельный, 370
 - — тлеющий, 370
- Гармоники боковые, 402
- Гармонический осциллятор, 419
 - — , колебания вынужденные, 438
 - — , — затухающие, 425
 - — , — свободные, 420
 - — , уравнение, 419
- Гиромагнитное отношение, 259
- Гироскоп, 136, 137
- Гистерезис, 223, 268
- Главные оси инерции, 137

- Голограмма, 626, 630
— во встречных пучках, 634
— , восстанавливающая волна, 626
— объемная, 631, 633
— , предметная волна, 624
— , разрешающая способность, 629
— , спектральная избирательность, 635
Голография, 623, 635
Гравитационная постоянная, 194
Гравитационное взаимодействие, 107
Градиент, 84, 205
Граница раздела двух диэлектриков, 527
— — — сред, 499
Граничное поле, 571
Давление, 147
— акустическое, 491
— магнитного поля, 289, 292
Движение вращательное, 37
— , интеграл, 65, 88
— относительно двух тел, 97
— под действием постоянной силы, 57
— поступательное, 37
— равномерное, 30
— — по окружности, 31, 32
— реактивное, 62
Двойное лучепреломление, 660, 662
Декремент затухания, 429
Деформация, 146, 488
— продольная, 488
Диамагнетизм, 266
Диамагнетик, 264
Дивергенция, 152, 201
Диполь, 207, 518
— в магнитном поле, 290
— взаимодействия, 210
— магнитный, 255, 258
— точечный, 519
— , энергия в электрическом поле, 209
Дипольное приближение, 207
Дипольный момент, 207
— — магнитный, 291
Дисперсия аномальная, 642, 650, 651
— волн, 638
— нормальная, 642, 646
— электромагнитных волн, 644, 646, 649
Дифракционная расходимость, 571
— решетка, 566, 603
Дифракция, 529, 563, 584
— , граничные условия, 564
— , диаграмма направленности, 598
— на краю экрана, 594
— , спектральный метод, 569
— Фраунгофера, 564, 595–597, 619
— — , главный максимум, 599
— — на круглом отверстии, 599
— — — решетке, 600
— — — щели, 598
— Френеля, 564, 585, 592
— — на щели, 592
Дихроизм, 663
Диэлектрик, 213, 214, 216, 219, 221
Диэлектрическая восприимчивость, 645
Диэлектрическая проницаемость, 216,
220, 646, 647, 650, 682
— — тензорная, 654
Длина волны, 478
— — в волноводе, 526
— когерентности, 678
— цуга, 548
Домены, 271
Емкость, 202, 204
— конденсатора, 204
— проводника, 204
— шара, 204
Жидкость идеальная, 150
— несжимаемая, 150, 158
Закон Ампера, 247
— Био–Савара, 247, 249, 265
— всемирного тяготения, 106, 107
— Гаука, 46, 144, 145, 148, 488
— Джоуля–Ленца, 235, 237
— дисперсии, 638
— Кулона, 196, 197
— Лапласа, 247
— Малюса, 663
— Ома, 235
— Паскаля, 153
— сохранения заряда, 233, 234
— — импульса, 51, 52
— — массы, 100
— — момента импульса, 122
— — энергии, 78, 83, 84, 89, 104, 105
— — — замкнутой системы, 101
— Фарадея, 282
Законы Кеплера, 123
— Кирхгофа, 276
— Ньютона, 41
— — второй, 43, 165, 168
— — первый, 42
— — третий, 49, 52, 174
— подобия, 163
Заряд поверхностный, 215
— связанный, 215
— электрона, 195
Звук, граница раздела двух сред, 499
Зонная пластинка Габора, 630
— — Френеля, 589, 590
— — Габора, 628
Зоны Френеля, 587–590, 637
— Шустера, 593
Изменение фазы несущей, 403,
404, 618
Измерение, 17, 46
Изохронность, 461
Импульс, 49, 50, 188
— , момент, 120
— обобщенный, 136
— релятивистский, 186
— силы, 50
Индуктивность, 280, 281, 295
— коаксиальных проводников, 282
— соленоида, 282
Индукция, 216, 218, 223

- Индукция взаимная, 280
— магнитного поля, 245, 246
— электрическая, 213
— электромагнитная, 280
Инертность, 52
Инерциальная система отсчета, 42
Инерция, 41
Интенсивность, 387, 494, 539, 540, 542, 545, 547, 549
— звука, 494
— спектральная, 549
Интерференция, 529, 530, 532, 533, 544, 545, 547, 550, 553, 558
— , апертура, 546
— , порядок, 533
— , — максимальный, 548, 551
— , ширина полосы, 534
Интерферометр, 559
— Майкельсона, 545, 559, 561
— Рэлея, 560
— Фабри–Перо, 606
— — , добротность, 609
— , база, 559
Интерферометрия, 559
Инфразвук, 492
Кавитация, 495
Качение, условие непроскальзывания, 142
Кинематика, 28
Классическая механика, граница, 25
Когерентность взаимная, функция, 543
— , время, 538, 550
— пространственная, 553, 554, 558
— — , площадка, 558
— — , радиус, 558
— , степень, 543
— , функция, 541
Колебание амплитудно-модулированное, 618
— двух тел, 97
— квазигармоническое, 396
— , комплексная амплитуда, 390
— модулированное по амплитуде, 396
— — по фазе, 396, 618
— несущее, 402
— под действием внешней силы, 69
Колебания, амплитуда, 68
— в струне, 500
— векторная интерпретация, 386
— векторные, 411
— вынужденные, 309, 438
— — , спектр, 452
— гармонические, 66, 68, 89
— — , комплексная форма записи, 388
— диполя, 518, 520
— затухающие, 425
— квазимонохроматические, 538, 547, 549
— , квантование энергии, 497
— линейно-поляризованные, 413
— модулированные, 395
— нормальные, 437
Колебания параметрические, 458
— , период, 68, 385
— плазменные, 376
— противофазные, 387
— синфазные, 387
— , спектральный анализ, 445, 448
— , фаза, 68
— — начальная, 68
— , частота, 69
— — круговая, 69
— — собственная, 70
— — циклическая, 69
— эллиптически поляризованные, 412
— , энергия превращения, 422–425
Коллапс контура, 305, 417, 423, 431, 458, 464
— — , релаксационные процессы, 303
Комбинационное рассеяние, 689
— — , антистоксова компонента, 689
— — вынужденное, 689
— — света, 687
— — , стоксова компонента, 689
Комплексная плоскость, 389
Комплексно-сопряженные числа, 390
Конденсатор, 204, 216, 229
— плоский, 205
— сферический, 205
— , энергия, 225
Координата обобщенная, 136
Космическая скорость вторая, 110, 111
— первая, 110
— третья, 111
Коэффициент жесткости, 45
— затухания, 426
— отражения, 357, 360, 606
— преломления, 351
— прозрачности, 357, 360
— пропускания, 606
— Пуассона, 147
— самоиндукции, 281
— трения, 86
Кристалл, 222, 265
— анизотропный, 679
— , главная плоскость, 655
— , главные направления, 655
— — показатели преломления, 655
— двуосный, 655
— одноосный, 655
— , оптическая ось, 655
— , оптические свойства, 666
— отрицательный, 655
— положительный, 655
Кристаллооптика, 653
Круговая поляризация, 514
— частота, 385
Кручение, 148
Ламерея лестница, 466
Ламинарное течение, 161
Лапласа оператор, 206, 476
— уравнение, 206
Линейная среда, 509
Линейные системы, 445, 449

- Линейный фильтр, 573
Линза, 611, 613, 619, 621
— идеальная, 568
— , фокальная плоскость, 611, 614
— , фокусное расстояние, 568
Линия тока, 151
Луч необыкновенный, 660, 662
— обыкновенный, 660, 662
Магнетон Бора, 270
Магнитная восприимчивость, 264
— проницаемость, 264
— , дифференциальная, 269
Магнитные цепи, 275
Максвелловская релаксация, 238
Максимум ширины источника, 554
Малость возмущений, 502
Масса, 43, 52, 188, 193
— обобщенная, 136
— приведенная, 96
Математический маятник, 416, 461
Материальная точка, 24
Маятник пружинный, 417, 463
— физический, 417
Мгновенная ось вращения, 140
Метод изображений, 209, 274
— квадратичного детектирования, 453
— Рэлея, 569
— фазового контраста, 616
Методы подобия и размерности, 162
Мода, 437, 497
— основная, 497
— собственная, 497
Модулированные колебания, 395
— глубина модуляции, 402
Модуль кручения, 149
— сдвига, 148
— Юнга, 146, 488
Молекулярные токи, 262
Момент импульса, 120, 122, 259
— , закон сохранения, 122
— твердого тела, 129
— инерции, 128, 130
— , перенос оси, 131
— тонкого диска, 130, 132
— цилиндра, 134
— магнитный, 255, 257–259
— электрона, 270
Мощность, 75
— , единица, 75
Напряжение, 146, 204, 488
Напряженность магнитного поля, 246, 250
— электрического поля, 197
Невесомость, 109, 174
Неинерциальные системы отсчета, 165
Нелинейная оптика, 672
— среда, 672
Нормальные моды, 437
Обертон, 492
Область дисперсии спектрального прибора, 606
Обменное взаимодействие, 270
Обратимые деформации, 144
Обратная связь, 467, 468
Обращение волнового фронта, 693
Оптика геометрическая, 564, 578, 579
Оптическое выпрямление, 673, 676
Опыт Юнга, 557, 558
Остаточная намагниченность, 269
Осциллятор, амплитудная характеристика, 442
— ангармонический, 461
— , добротность, 429
— затухающий, фазовая плоскость, 433
— , постоянная времени, 428
— с большим затуханием, 432
— малым затуханием, 427
— , частотная характеристика, 442
Ось инерции, 137
— наибольшей поляризации, 664
Пара сил, 123
Параболическое приближение, 482
Парамагнетизм, 266
Парамагнетик, 264, 271
Параметрическая раскачка, 460
Параметрическое усиление света, 686
Период колебаний, 385, 478
Пирозлектрик, 223
Плазма, 366, 371
— , квазинейтральность, 374
— , ленгмюровские колебания, 375
— , температура, 372
— , удержание магнитным полем, 373
— , экранирование, 375
— , электромагнитные волны, 377
Плазменная частота, 646
Пластика четвертьволновая, 657
Плечо силы, 134
Плотность кинетической энергии, 492
— потенциальной энергии, 493
— потока энергии, 494
— энергии, 156, 515
Поверхностный заряд, 219
Показатель преломления, 351, 528, 642, 657, 682
— главный, 655
— , дисперсия, 352
— относительный, 353
Поле, магнитная индукция, 246
— , напряженность, 246
— магнитное, 244, 245
— в веществе, 262
— , векторный потенциал, 253, 255
— , граничные условия, 272, 273
— , индукция, 274
— , напряженность, 250, 252, 255, 274
— , потенциал, 257
— , поток, 253
— тока, 244, 248, 251, 252
— , электрический потенциал, 255
— , энергия, 294
— потенциальное, 82, 255
— соленоидальное, 255

- Поле электрическое в веществе, 214
 — — , граничные условия, 218
 — — , напряженность, 197
 — — , потенциальность, 206
 — — проводника, 206
 — — , энергия, 224, 226–228
 — электромагнитное глубина проникновения, 364
 — — , импульс, 329
 — — квазистационарное, 302, 303
 — — , проводники, 288
 — — , энергия, 288
 Положение равновесия, устойчивое, 416
 Поля, электрические и магнитные, 26
 Поляризаторы, 663
 Поляризационные эффекты, 359
 Поляризация, 214, 215, 219–221, 486, 676
 — ось наибольшей поляризации, 664
 Поляризуемость, 216, 654, 668, 682
 Порог нелинейной неустойчивости, 93
 — слышимости, 495
 Постоянная Верде, 670
 — времени фотоприемника, 539
 — всемирного тяготения, 107
 — гравитационная, 107
 — Керра, 665
 — магнитная, 246
 — электрическая, 196
 Потенциал, 202–206, 226
 — , аддитивность, 203
 Поток энергии, 500, 515, 516
 Правила Кирхгофа, 239, 241
 Предел упругости, 145
 Преобразования Галилея, 39, 166
 — Лоренца, 177, 181, 182
 — Фурье, 407
 Прещесия, 138
 Приближение абсолютно твердого тела, 35
 — материальной точки, 35
 — параболическое, 482
 Приведенная масса, 96
 Призма Николя, 664
 Принцип Гюйгенса–Френеля, 583, 589, 621
 — двойной дифракции, 613, 615
 — Ленца, 282
 — относительности Галилея, 165, 167
 — — Эйнштейна, 177, 178
 — постоянства скорости света, 178
 — соответствия, 187
 — суперпозиции, 81, 200, 502, 529
 — — полей, 198, 203
 Пробный заряд, 202
 Проводимость, 235, 236
 Проводник, 211, 273
 Пространственная частота, 570, 579
 Пространство, 18, 42
 Псевдовектор, 34
 Псевдоскаляр, 34
 Пьезоэлектрик, 221, 222
 Пятно Пуассона, 591
 — Эйри, 599, 620–623, 629
 Работа, 73, 74, 81
 — , единица, 75
 Равновесие, 91
 Радиационное затухание, 522
 Радиоинтерферометр, 560
 Радиус дебаевский, 376
 Радиус-вектор, 23
 Размерность, 46, 48
 Разрешающая способность голограммы, 629
 — — оптических систем, 611, 619, 622
 — — спектральных приборов, 603, 608
 Рассеяние, 114
 Резонанс, 71, 312, 443
 — параметрический, 458
 Резонансная кривая, 444
 — характеристика, 443
 Релятивистская механика, 177
 — длина тел, 182
 — длительность процессов, 182
 — энергия, 186
 — — кинетическая, 188
 Релятивистский импульс, 186
 — интервал, 185
 Решетка, 618
 Ротор, 250
 Рычаг, 134
 Рэля критерий, 604
 Ряд Тейлора, 583
 — Фурье, 451
 Самоканализация, 680, 683
 Самофокусировка, 680, 683
 Связанные осцилляторы, 434, 437
 Сдвиг, 148
 Сегнетоэлектрик, 221–223
 Сердечник, 275
 Сила, 193
 — Ампера, 248, 292
 — Архимеда, 153
 — гироскопическая, 84
 — гравитационного притяжения однородного шара, 111
 — диссипативная, 84
 — звука, 494
 — инерции, 167
 — — во вращающейся системе отсчета, 170
 — — поступательной, 169
 — — центробежная, 172
 — консервативная, 79, 82
 — Кориолиса, 172
 — — инерции, 172
 — коэрцитивная, 269
 — Лоренца, 248, 261
 — , момент, 120, 129
 — непотенциальная, 84
 — нормального давления, 85
 — обобщенная, 136
 — , объемная плотность, 154
 — , принцип суперпозиции, 81

- Сила притяжения, 107
— реакции опоры, 86
— трения, 45, 84, 85
— — покоя, 85
— — скольжения, 85, 86
— тяжести, 108
— упругости, 45
— центральная, 79, 81
Силовое поле, 78
— постоянное, 90
Силовые линии, 198
Силы гироскопические, 248
Система единиц СГС, 48
— — СИ, 46
— замкнутая, 50, 95
— изолированная, 50
— отсчета, 22
— — лабораторная, 100
— — центра масс, 100
Системы отсчета неинерциальные, 167
Скаляр, 23
Скачок уплотнения, 504
Скин-эффект, 315
Скин-овая длина бесстолкновительная, 380
Скорость, 28
— абсолютная, 169
— групповая, 638, 641, 643
— звука, 489, 491
— относительная, 168
— первая космическая, 64
— переносная, 168
— света, 177, 179, 338
— свойства, 30
— угловая, 32, 33
— — , вектор, 136
— — , мгновенное значение, 140
— — , плоское движение, 141
— фазовая, 479, 638, 642
Скэйлинг, 163
Случайный процесс, 535
Собственные моды колебаний, 496
Соленоид, 251, 294
Солитон, 504
Соотношение неопределенностей, 549, 571
Сопротивление, 235
— излучения, 521
Состояние покоя, 91
Соударения неупругие, 114
— упругие, 114
Спектр, 549–551, 571
— непериодического сигнала, 406
Спектральное разложение, 399
— , комплексная форма, 400
Спектральный анализ, 445, 448
— метод, 569
Специальная теория относительности, 178, 362
Спираль Корню, 392, 394, 581, 593
— Френеля, 391, 392, 588
Среднее значение, 423
Статистический характер светового излучения, 536
Стационарное течение, 155
Степени свободы, 36, 37, 126
Стопа Столетова, 360
Сумма гармонических колебаний, 386, 400
Суперпозиция плоских волн, 569
Таутохронизм, 621
Твердое тело, 145
— — , вращение вокруг оси, 126, 129
— — , динамика, 120
— — , плоское движение, 140
— — , трехмерное движение, 136
Тело поликристаллическое, 145
Тембр звука, 492
Теорема Гаусса, 199, 200, 202
— Гаусса–Остроградского, 201
— Кёнига, 101
— Лармора, 268
— о циркуляции, 249, 250, 265, 321
— Пойнтинга, 325
— Стокса, 250
— Штейнера, 131
Течение Пуазейля, 157
Ток намагничивания, 265
— переменный, 309
— смещения, 321
— электрический, 231
— — , плотность тока, 232
— — постоянный, 230
Траектория, 24
Трубка тока, 151
Турбулентное течение, 161
Турбулентность, 161
— гидродинамическая, 161
Угловой размер, 560
Ультразвук, 492
Упругие деформации, 144
Уравнение адиабаты, 491
— Бернулли, 153, 156
— волновое, 476, 479, 489, 491, 501, 508, 511
— — линейное, 502
— — нелинейное, 503
— гармонических колебаний, 384
— Гельмгольца, 483, 571
— Лапласа, 205
— моментов, 120, 122
— непрерывности, 152
— неразрывности струи, 152
— однородное, 439
— Пуассона, 205, 206, 254
— характеристическое, 432
Уравнения интеграл, 65
Уравнения Максвелла, 321, 323, 508
Ускорение, 28, 31, 193
— переносное, 172
— свободного падения, 108
— центростремительное, 32, 171
Условие Беннета, 300
— Брэгга–Вульфа, 635

- Условие Кирхгофа граничное, 565
Устойчивое положение равновесия, 416
Устойчивость, 91, 92
— линейная, 93
— нелинейная, 93
Фаза волны, 477
— колебания, 385
— модулированного колебания, 397
— начальная, 385
— стационарная, 580
Фазовая модуляция, 396
— плоскость, 388
— — затухающего осциллятора, 433
— решетка, 567
— скорость, 478
— траектория, 388
— характеристика, 442
Фазовый синхронизм, условие, 678
Фактор размагничивания, 265
Ферриты, 272
Ферромагнетизм, 268
Ферромагнетик, 264
— магнитотвердый, 269
— магнитомягкий, 269
Фигуры Лиссажу, 411, 413, 414
Фильтры, 445
Фокус, 568
Формула Пуазейля, 160
— Рэлея, 638, 642
— Торричелли, 156
— Френеля, 356, 358
— Циолковского, 64
Функции гармонические, 450
Функция когерентности, 541, 549, 555
— пропускания, 566, 567
Фурье-анализ, 399
Фурье-плоскость, 614
Фурье-спектрометрия, 561
Характеристика, крутизна, 469
— жесткая, 464
— мягкая, 463
— резонансная, 443
— фазовая, 442
— частотная, 442, 450, 451, 573
Центр инерции, 98
— масс, 98, 99, 101
Цуг, 551
— , длина, 548
Частица, 26
Частота критическая, 526
— круговая, 385
— Лармора, 267
— пространственная, 570
— собственная, 70
— — осциллятора, 426
Частота циклическая, 385
— циклотронная, 248
— электронная ленгмюровская, 376
Число волновое, 495
— дебаевское, 377
— Рейнольдса, 162
— — критическое, 163
Чистый тон, 492
Электрические цепи, 280, 302
— резонанс, 312
— релаксационные процессы, 303
Электрический заряд, 195
— дрейф, 261
Электрическое поле, 205
— — в веществе, 211
— — заряженной нити, 200
— — — плоскости, 200
— — , потенциальная энергия, 203
Электродвижущая сила, 239, 240
Электромагнит, 298
Электромагнитные волны, 333, 338, 508, 649
— — , импульс, 341
— — , энергия, 341, 516
Энергия, 188
— взаимодействия, 95, 103
— — гравитационного, 109
— внутренняя, 95, 105
— , закон сохранения, 78, 89
— кинетическая, 73, 76
— — вращения, 135, 136
— — , объемная плотность, 492
— магнитного поля, 292, 294, 297
— , плотность общая, 493
— — потока, 494
— — упругой деформации, 147
— покоя, 188
— полная, 82
— потенциальная, 78, 79, 82, 101
— — взаимодействия, 102
— — , объемная плотность, 493
— , релятивистская механика, 186, 188
Эрмитовость, 407
Эталон, 17
— времени, 21
Эффект Доплера, 506
— Керра, 664, 665, 667
— Коттона–Мутона, 667
— Поккельса, 666
— самовоспроизведения, 578
— Талбота, 578
— удвоения частоты, 673, 676, 679
— Фарадея, 667
— Холла, 277
Эффекты магнитооптические, 664
— электрооптические, 664