Компания «Ваш репетитор»

MathUs.ru

И. В. Яковлев

Физика

Перед вами — электронное пособие по физике, охватывающее всю школьную программу и, соответственно, все темы кодификатора ЕГЭ по физике.

Цель данного пособия — обеспечить школьникам исчерпывающую теоретическую подготовку по физике. Такую подготовку, которая позволит успешно выступить на олимпиадах, набрать высокие баллы на ЕГЭ, поступить в желаемый вуз и впоследствии органично перейти к изучению вузовского курса общей физики.

Пособие состоит из семи глав:

1. Механика
2. Молекулярная физика и термодинамика
3. Электродинамика
4. Оптика
5. Теория относительности
6. Квантовая физика
7. Приложение. Векторы в физике

Содержание пособия выходит за рамки кодификатора ЕГЭ и в отдельных местах за рамки школьной программы. Дополнительный материал позволяет лучше понять рассматриваемые темы и служит «мостиком» к дальнейшему вузовскому курсу. Этот «мостик» совершенно необходим, поскольку разрыв между уровнями изложения физики в школе и в вузе весьма велик. Если заранее не преодолеть этот разрыв хотя бы частично, то есть риск получить проблемы с физикой в первую же сессию.

Автор пособия — профессиональный преподаватель с 25-летним стажем, имеющий красный диплом МФТИ (факультет общей и прикладной физики) и несколько сотен учеников, ставших успешными студентами ведущих московских вузов.

Все главы пособия обсуждались на форуме преподавателей компании «Ваш репетитор», которая является крупнейшим в России репетиторским сообществом. Я выражаю глубокую благодарность Дмитрию Андреевичу Сухоручкину, Антону Марковичу Безбородову, Михаилу Владимировичу Солину, Борису Семёновичу Семёнову, Юрию Анатольевичу Боравлёву, а также другим ведущим преподавателям компании, чьи квалифицированные замечания способствовали улучшению изложения материала.

Текст пособия время от времени обновляется (по мере устранения замеченных ошибок). Последняя версия всегда находится по адресу: http://mathus.ru/phys/book.pdf.

По всем вопросам пишите мне: pandominator@gmail.com

Оглавление

1 Механика **12**

1.1 Производная . . . 14
1.1.1 Предел . . . 14
1.1.2 Мгновенная скорость . . . 16
1.1.3 Определение производной . . . 18
1.1.4 Табличные производные . . . 19
1.1.5 Правила дифференцирования . . . 20
1.1.6 Обозначения производной в физике . . . 22
1.1.7 Предел векторной величины . . . 23
1.1.8 Дифференцирование векторов . . . 24
1.2 Механическое движение . . . 27
1.2.1 Относительность движения . . . 27
1.2.2 Основная задача механики . . . 27
1.2.3 Материальная точка . . . 28
1.2.4 Траектория, путь, перемещение . . . 28
1.2.5 Скорость . . . 29
1.2.6 Ускорение . . . 30
1.2.7 Примеры вычисления скорости и ускорения . . . 30
1.2.8 Закон сложения скоростей . . . 31
1.2.9 Виды механического движения . . . 33
1.3 Равномерное прямолинейное движение . . . 35
1.3.1 Закон движения . . . 35
1.3.2 Интегрирование . . . 36
1.4 Равноускоренное движение . . . 37
1.4.1 Зависимость скорости от времени . . . 37
1.4.2 Закон движения . . . 37
1.4.3 Прямолинейное равноускоренное движение . . . 38
1.4.4 Свободное падение . . . 39
1.4.5 Горизонтальный бросок . . . 40
1.4.6 Бросок под углом к горизонту . . . 41
1.5 Равномерное движение по окружности . . . 42
1.5.1 Угловая скорость . . . 42
1.5.2 Закон движения . . . 43
1.5.3 Центростремительное ускорение . . . 43
1.6 Путь при неравномерном движении . . . 44
1.7 Первый закон Ньютона . . . 47
1.7.1 Инерциальные системы отсчёта . . . 47
1.7.2 Принцип относительности . . . 48
1.8 Масса и плотность . . . 49
1.9 Второй и третий законы Ньютона . . . 50

1.9.1 Принцип суперпозиции . . . 50
1.9.2 Второй закон Ньютона . . . 50
1.9.3 Третий закон Ньютона . . . 50
1.9.4 Как найти закон движения? . . . 51
1.10 Сила упругости . . . 53
1.10.1 Деформация . . . 53
1.10.2 Закон Гука . . . 53
1.10.3 Модуль Юнга . . . 54
1.11 Сила тяготения . . . 55
1.11.1 Закон всемирного тяготения . . . 55
1.11.2 Сила тяжести . . . 55
1.11.3 Вес тела. Невесомость . . . 56
1.11.4 Искусственные спутники . . . 58
1.12 Сила трения . . . 59
1.12.1 Сухое трение . . . 59
1.12.2 Вязкое трение . . . 61
1.13 Статика твёрдого тела . . . 62
1.13.1 Момент силы . . . 63
1.13.2 Условия равновесия . . . 63
1.14 Статика жидкостей и газов . . . 66
1.14.1 Гидростатическое давление . . . 66
1.14.2 Закон Паскаля . . . 67
1.14.3 Гидравлический пресс . . . 68
1.14.4 Закон Архимеда . . . 68
1.14.5 Плавание тел . . . 70
1.15 Импульс . . . 72
1.15.1 Второй закон Ньютона в импульсной форме . . . 72
1.15.2 Пример вычисления силы . . . 74
1.15.3 Импульс системы тел . . . 75
1.15.4 Закон сохранения импульса . . . 76
1.15.5 Закон сохранения проекции импульса . . . 77
1.16 Энергия . . . 79
1.16.1 Работа . . . 79
1.16.2 Мощность . . . 80
1.16.3 Механическая энергия . . . 81
1.16.4 Кинетическая энергия . . . 81
1.16.5 Потенциальная энергия тела вблизи поверхности Земли . . . 82
1.16.6 Потенциальная энергия деформированной пружины . . . 83
1.16.7 Закон сохранения механической энергии . . . 84
1.16.8 Закон изменения механической энергии . . . 84
1.17 Простые механизмы . . . 86
1.17.1 Рычаг . . . 86
1.17.2 Неподвижный блок . . . 86
1.17.3 Подвижный блок . . . 87
1.17.4 Наклонная плоскость . . . 87
1.17.5 Золотое правило механики . . . 88
1.17.6 КПД механизма . . . 89
1.18 Механические колебания . . . 91
1.18.1 Гармонические колебания . . . 91
1.18.2 Уравнение гармонических колебаний . . . 93

1.18.3 Пружинный маятник . . . 94
1.18.4 Математический маятник . . . 95
1.18.5 Свободные и вынужденные колебания . . . 96
1.19 Механические волны . . . 98
1.19.1 Продольные и поперечные волны . . . 98
1.19.2 Звук . . . 99

2 Молекулярная физика и термодинамика 101
2.1 Основные положения МКТ . . . 102
2.1.1 Атомы и молекулы . . . 102
2.1.2 Тепловое движение атомов и молекул . . . 103
2.1.3 Взаимодействие частиц вещества . . . 104
2.2 Газы, жидкости и твёрдые тела . . . 106
2.2.1 Газы . . . 106
2.2.2 Твёрдые тела . . . 106
2.2.3 Жидкости . . . 108
2.3 Основные формулы молекулярной физики . . . 109
2.4 Температура . . . 111
2.4.1 Термодинамическая система . . . 111
2.4.2 Тепловое равновесие . . . 111
2.4.3 Температурная шкала. Абсолютная температура . . . 112
2.5 Уравнение состояния идеального газа . . . 113
2.5.1 Средняя кинетическая энергия частиц газа . . . 113
2.5.2 Основное уравнение МКТ идеального газа . . . 114
2.5.3 Энергия частиц и температура газа . . . 114
2.5.4 Уравнение Менделеева — Клапейрона . . . 115
2.6 Изопроцессы . . . 116
2.6.1 Термодинамический процесс . . . 116
2.6.2 Изотермический процесс . . . 116
2.6.3 Графики изотермического процесса . . . 117
2.6.4 Изобарный процесс . . . 118
2.6.5 Графики изобарного процесса . . . 119
2.6.6 Изохорный процесс . . . 120
2.6.7 Графики изохорного процесса . . . 121
2.7 Насыщенный пар . . . 123
2.7.1 Испарение и конденсация . . . 123
2.7.2 Динамическое равновесие . . . 123
2.7.3 Свойства насыщенного пара . . . 124
2.7.4 Влажность воздуха . . . 125
2.8 Внутренняя энергия . . . 127
2.8.1 Внутренняя энергия одноатомного идеального газа . . . 127
2.8.2 Функция состояния . . . 128
2.8.3 Изменение внутренней энергии: совершение работы . . . 128
2.8.4 Изменение внутренней энергии: теплопередача . . . 128
2.8.5 Теплопроводность . . . 129
2.8.6 Конвекция . . . 129
2.8.7 Тепловое излучение . . . 130
2.9 Количество теплоты . . . 132
2.9.1 Удельная теплоёмкость вещества . . . 132
2.9.2 Уравнение теплового баланса . . . 133

2.10 Фазовые переходы . . . 134
2.10.1 Плавление и кристаллизация . . . 134
2.10.2 График плавления . . . 135
2.10.3 Удельная теплота плавления . . . 136
2.10.4 График кристаллизации . . . 137
2.10.5 Парообразование и конденсация . . . 138
2.10.6 Кипение . . . 139
2.10.7 График кипения . . . 140
2.10.8 График конденсации . . . 141
2.11 Первый закон термодинамики . . . 142
2.11.1 Работа газа в изобарном процессе . . . 142
2.11.2 Работа газа в произвольном процессе . . . 143
2.11.3 Работа, совершаемая над газом . . . 143
2.11.4 Первый закон термодинамики . . . 144
2.11.5 Применение первого закона термодинамики к изопроцессам . . . 144
2.11.6 Адиабатный процесс . . . 145
2.12 Тепловые машины . . . 146
2.12.1 Тепловые двигатели . . . 146
2.12.2 Холодильные машины . . . 148
2.12.3 Тепловая машина Карно . . . 149
2.12.4 Тепловые двигатели и охрана окружающей среды . . . 151
2.13 Второй закон термодинамики . . . 152
2.13.1 Необратимость процессов в природе . . . 152
2.13.2 Постулаты Клаузиуса и Кельвина . . . 152
2.13.3 Эквивалентность постулатов Клаузиуса и Кельвина . . . 153
2.13.4 Обратимые процессы . . . 153
2.13.5 Обратимость машины Карно . . . 154

3 Электродинамика **156**
3.1 Электрический заряд . . . 158
3.1.1 Два вида заряда . . . 158
3.1.2 Электризация тел . . . 159
3.1.3 Закон сохранения заряда . . . 161
3.2 Закон Кулона . . . 162
3.2.1 Принцип суперпозиции . . . 163
3.2.2 Закон Кулона в диэлектрике . . . 164
3.3 Напряжённость электрического поля . . . 165
3.3.1 Дальнодействие и близкодействие . . . 165
3.3.2 Электрическое поле . . . 165
3.3.3 Напряжённость поля точечного заряда . . . 166
3.3.4 Принцип суперпозиции электрических полей . . . 168
3.3.5 Поле равномерно заряженной плоскости . . . 169
3.3.6 Линии напряжённости электрического поля . . . 170
3.4 Потенциал электрического поля . . . 171
3.4.1 Консервативные силы . . . 171
3.4.2 Потенциальность электростатического поля . . . 172
3.4.3 Потенциальная энергия заряда в однородном поле . . . 172
3.4.4 Потенциальная энергия взаимодействия точечных зарядов . . . 174
3.4.5 Потенциал . . . 174
3.4.6 Разность потенциалов . . . 175

3.4.7 Принцип суперпозиции для потенциалов . . . 176
3.4.8 Однородное поле: связь напряжения и напряжённости . . . 176
3.4.9 Эквипотенциальные поверхности . . . 177
3.5 Проводники в электрическом поле . . . 179
3.5.1 Поле внутри проводника . . . 179
3.5.2 Заряд внутри проводника . . . 181
3.5.3 Поле вне проводника . . . 181
3.5.4 Потенциал проводника . . . 181
3.5.5 Напряжённость и потенциал поля проводящей сферы . . . 182
3.6 Диэлектрики в электрическом поле . . . 184
3.6.1 Диэлектрическая проницаемость . . . 184
3.6.2 Полярные диэлектрики . . . 185
3.6.3 Неполярные диэлектрики . . . 186
3.7 Конденсатор. Энергия электрического поля . . . 187
3.7.1 Ёмкость уединённого проводника . . . 187
3.7.2 Ёмкость плоского конденсатора . . . 188
3.7.3 Энергия заряженного конденсатора . . . 190
3.7.4 Энергия электрического поля . . . 192
3.8 Постоянный электрический ток . . . 193
3.8.1 Направление электрического тока . . . 193
3.8.2 Действие электрического тока . . . 194
3.8.3 Сила и плотность тока . . . 195
3.8.4 Скорость направленного движения зарядов . . . 195
3.8.5 Стационарное электрическое поле . . . 197
3.9 Закон Ома . . . 199
3.9.1 Закон Ома для участка цепи . . . 199
3.9.2 Электрическое сопротивление . . . 200
3.9.3 Удельное сопротивление . . . 200
3.10 Соединения проводников . . . 202
3.10.1 Резисторы и подводящие провода . . . 202
3.10.2 Последовательное соединение . . . 203
3.10.3 Параллельное соединение . . . 204
3.10.4 Смешанное соединение . . . 206
3.11 Работа и мощность тока . . . 208
3.11.1 Работа тока . . . 208
3.11.2 Мощность тока . . . 208
3.11.3 Закон Джоуля–Ленца . . . 209
3.12 ЭДС. Закон Ома для полной цепи . . . 210
3.12.1 Сторонняя сила . . . 210
3.12.2 Закон Ома для полной цепи . . . 211
3.12.3 КПД электрической цепи . . . 212
3.12.4 Закон Ома для неоднородного участка . . . 213
3.13 Электрический ток в металлах . . . 215
3.13.1 Свободные электроны . . . 215
3.13.2 Опыт Рикке . . . 216
3.13.3 Опыт Стюарта–Толмена . . . 216
3.13.4 Зависимость сопротивления от температуры . . . 217
3.14 Электрический ток в электролитах . . . 219
3.14.1 Электролитическая диссоциация . . . 219
3.14.2 Ионная проводимость . . . 221

3.14.3 Электролиз . . . 222
3.15 Электрический ток в газах . . . 224
3.15.1 Свободные заряды в газе . . . 225
3.15.2 Несамостоятельный разряд . . . 226
3.15.3 Вольт-амперная характеристика газового разряда . . . 227
3.15.4 Самостоятельный разряд . . . 228
3.16 Полупроводники . . . 229
3.16.1 Ковалентная связь . . . 230
3.16.2 Кристаллическая структура кремния . . . 230
3.16.3 Собственная проводимость . . . 231
3.16.4 Примесная проводимость . . . 233
3.16.5 p–n-переход . . . 236
3.17 Магнитное поле. Линии . . . 239
3.17.1 Взаимодействие магнитов . . . 239
3.17.2 Линии магнитного поля . . . 240
3.17.3 Опыт Эрстеда . . . 240
3.17.4 Магнитное поле прямого провода с током . . . 241
3.17.5 Магнитное поле витка с током . . . 242
3.17.6 Магнитное поле катушки с током . . . 243
3.17.7 Гипотеза Ампера. Элементарные токи . . . 244
3.18 Магнитное поле. Силы . . . 245
3.18.1 Сила Лоренца . . . 245
3.18.2 Сила Ампера . . . 246
3.18.3 Рамка с током в магнитном поле . . . 247
3.19 Электромагнитная индукция . . . 250
3.19.1 Магнитный поток . . . 251
3.19.2 ЭДС индукции . . . 252
3.19.3 Закон электромагнитной индукции Фарадея . . . 252
3.19.4 Правило Ленца . . . 253
3.19.5 Взаимодействие магнита с контуром . . . 254
3.19.6 Закон Фарадея + Правило Ленца = Снятие модуля . . . 255
3.19.7 Вихревое электрическое поле . . . 256
3.19.8 ЭДС индукции в движущемся проводнике . . . 257
3.20 Самоиндукция . . . 260
3.20.1 Индуктивность . . . 261
3.20.2 Электромеханическая аналогия . . . 262
3.20.3 Энергия магнитного поля . . . 263
3.21 Электромагнитные колебания . . . 265
3.21.1 Колебательный контур . . . 265
3.21.2 Энергетические превращения в колебательном контуре . . . 268
3.21.3 Электромеханические аналогии . . . 268
3.21.4 Гармонический закон колебаний в контуре . . . 269
3.21.5 Вынужденные электромагнитные колебания . . . 272
3.22 Переменный ток. 1 . . . 273
3.22.1 Условие квазистационарности . . . 273
3.22.2 Резистор в цепи переменного тока . . . 274
3.22.3 Конденсатор в цепи переменного тока . . . 275
3.22.4 Катушка в цепи переменного тока . . . 277
3.23 Переменный ток. 2 . . . 280
3.23.1 Метод вспомогательного угла . . . 280

3.23.2 Колебательный контур с резистором . . . 280
3.23.3 Резонанс в колебательном контуре . . . 282
3.24 Мощность переменного тока . . . 285
3.24.1 Мощность тока через резистор . . . 285
3.24.2 Мощность тока через конденсатор . . . 287
3.24.3 Мощность тока через катушку . . . 289
3.24.4 Мощность тока на произвольном участке . . . 289
3.25 Электроэнергия . . . 291
3.25.1 Производство электроэнергии . . . 291
3.25.2 Передача электроэнергии . . . 292
3.25.3 Трансформатор . . . 293
3.26 Электромагнитное поле . . . 298
3.26.1 Гипотеза Максвелла . . . 298
3.26.2 Понятие электромагнитного поля . . . 299
3.26.3 Об уравнениях Максвелла . . . 300
3.27 Электромагнитные волны . . . 302
3.27.1 Открытый колебательный контур . . . 302
3.27.2 Свойства электромагнитных волн . . . 304
3.27.3 Плотность потока излучения . . . 305
3.27.4 Виды электромагнитных излучений . . . 307

4 Оптика **311**
4.1 Световые лучи . . . 313
4.1.1 Законы геометрической оптики . . . 313
4.1.2 Геометрическая тень . . . 314
4.2 Отражение света . . . 315
4.2.1 Закон отражения . . . 315
4.2.2 Плоское зеркало . . . 316
4.3 Преломление света . . . 319
4.3.1 Закон преломления (частный случай) . . . 319
4.3.2 Обратимость световых лучей . . . 320
4.3.3 Закон преломления (общий случай) . . . 321
4.3.4 Полное внутреннее отражение . . . 322
4.4 Линзы. Ход лучей . . . 324
4.4.1 Двояковыпуклая линза . . . 324
4.4.2 Двояковогнутая линза . . . 326
4.4.3 Виды собирающих и рассеивающих линз . . . 327
4.5 Тонкие линзы. Ход лучей . . . 329
4.5.1 Понятие тонкой линзы . . . 329
4.5.2 Оптический центр и фокальная плоскость . . . 330
4.5.3 Ход луча через оптический центр . . . 331
4.5.4 Ход лучей в собирающей линзе . . . 332
4.5.5 Ход лучей в рассеивающей линзе . . . 333
4.6 Тонкие линзы. Построение изображений . . . 335
4.6.1 Собирающая линза: действительное изображение точки . . . 335
4.6.2 Собирающая линза: действительное изображение предмета . . . 338
4.6.3 Собирающая линза: мнимое изображение точки . . . 339
4.6.4 Собирающая линза: мнимое изображение предмета . . . 341
4.6.5 Собирающая линза: предмет в фокальной плоскости . . . 341
4.6.6 Рассеивающая линза: мнимое изображение точки . . . 342

4.6.7 Рассеивающая линза: мнимое изображение предмета . . . 344
4.7 Глаз человека . . . 345
4.7.1 Строение глаза . . . 345
4.7.2 Аккомодация . . . 346
4.7.3 Угол зрения . . . 347
4.7.4 Расстояние наилучшего зрения . . . 348
4.7.5 Близорукость . . . 348
4.7.6 Дальнозоркость . . . 349
4.8 Оптические приборы . . . 351
4.8.1 Невооружённый глаз . . . 351
4.8.2 Лупа . . . 352
4.8.3 Микроскоп . . . 353
4.8.4 Труба Кеплера . . . 355
4.8.5 Труба Галилея . . . 356
4.9 Принцип Гюйгенса . . . 357
4.9.1 Волновые поверхности и лучи . . . 357
4.9.2 Сферическая волна . . . 358
4.9.3 Плоская волна . . . 359
4.9.4 Вторичные волны . . . 359
4.9.5 Вывод закона отражения . . . 361
4.9.6 Вывод закона преломления . . . 362
4.10 Интерференция волн . . . 364
4.10.1 Сложение колебаний . . . 364
4.10.2 Интенсивность волны . . . 365
4.10.3 Когерентные источники . . . 366
4.10.4 Условие максимума и минимума . . . 366
4.10.5 Интерференционная картина . . . 367
4.10.6 Схема Юнга . . . 369
4.11 Интерференция света . . . 371
4.11.1 Усреднение интенсивности . . . 371
4.11.2 Некогерентность независимых источников . . . 371
4.11.3 Зеркала Френеля . . . 373
4.11.4 Интерференция в тонких плёнках . . . 374
4.11.5 Кольца Ньютона . . . 375
4.11.6 Просветление оптики . . . 377
4.12 Дифракция света . . . 379
4.12.1 Принцип Гюйгенса–Френеля . . . 380
4.12.2 Опыт Юнга . . . 381
4.12.3 Дифракционная решётка . . . 382
4.12.4 Дифракционная решётка как спектральный прибор . . . 385
4.13 Дисперсия света . . . 387
4.13.1 Опыт Ньютона . . . 387
4.13.2 Хроматическая аберрация . . . 388

5 Теория относительности **389**
5.1 Принцип относительности Галилея . . . 390
5.1.1 Наблюдатель на корабле . . . 390
5.1.2 Инвариантность законов механики . . . 390
5.2 Принципы СТО . . . 393
5.2.1 Гипотеза о мировом эфире . . . 393

5.2.2 Постулаты Эйнштейна . . . 394
5.3 Релятивистская кинематика . . . 398
5.3.1 Одновременность событий . . . 398
5.3.2 Относительность одновременности . . . 401
5.3.3 Относительность промежутков времени . . . 401
5.3.4 Относительность расстояний . . . 403
5.3.5 Преобразования Лоренца . . . 404
5.3.6 Релятивистский закон сложения скоростей . . . 406
5.4 Релятивистская динамика . . . 408
5.4.1 Релятивистская энергия . . . 408
5.4.2 Релятивистский импульс . . . 410
5.4.3 Связь энергии и импульса . . . 412
5.4.4 Релятивистское уравнение движения . . . 412

6 Квантовая физика **415**
6.1 Фотоэффект . . . 416
6.1.1 Опыты Столетова . . . 416
6.1.2 Зависимость фототока от напряжения . . . 417
6.1.3 Законы фотоэффекта . . . 418
6.1.4 Трудности классического объяснения фотоэффекта . . . 419
6.1.5 Гипотеза Планка о квантах . . . 420
6.1.6 Уравнение Эйнштейна для фотоэффекта . . . 420
6.2 Фотоны . . . 423
6.2.1 Энергия фотона . . . 423
6.2.2 Импульс фотона . . . 423
6.2.3 Давление света . . . 424
6.2.4 Двойственная природа света . . . 425
6.3 Корпускулярно-волновой дуализм . . . 427
6.3.1 Гипотеза де Бройля . . . 427
6.3.2 Дифракция электронов . . . 428
6.3.3 Соотношение неопределённостей . . . 429
6.4 Линейчатые спектры . . . 430
6.4.1 Спектр испускания . . . 430
6.4.2 Спектр поглощения . . . 431
6.4.3 Спектральный анализ . . . 432
6.5 Строение атома . . . 433
6.5.1 Модель Томсона . . . 433
6.5.2 Опыты Резерфорда . . . 433
6.5.3 Планетарная модель атома . . . 434
6.6 Атом Бора . . . 436
6.6.1 Постулаты Бора . . . 436
6.6.2 Атом водорода . . . 438
6.6.3 Достоинства и недостатки теории Бора . . . 441
6.7 Лазер . . . 442
6.7.1 Индуцированное излучение . . . 442
6.7.2 Инверсная населённость . . . 443
6.7.3 Трёхуровневая система рубина . . . 444
6.7.4 Устройство лазера . . . 445
6.8 Строение ядра . . . 446
6.8.1 Нуклонная модель ядра . . . 446

6.8.2 Изотопы . . . 447
6.9 Радиоактивность . . . 448
6.9.1 Виды радиоактивных излучений . . . 449
6.9.2 Радиоактивные превращения . . . 449
6.9.3 Закон радиоактивного распада . . . 451
6.10 Энергия связи ядра . . . 453
6.10.1 Ядерные силы . . . 453
6.10.2 Атомная единица массы . . . 454
6.10.3 Удельная энергия связи . . . 456
6.10.4 Насыщение ядерных сил . . . 457
6.11 Ядерные реакции . . . 458
6.11.1 Энергетический выход ядерной реакции . . . 459
6.11.2 Деление ядер . . . 461
6.11.3 Цепная ядерная реакция . . . 464
6.11.4 Термоядерная реакция . . . 464

7 Приложение. Векторы в физике 466
7.1 Скалярные и векторные величины . . . 467
7.2 Сложение векторов . . . 469
7.2.1 Правило треугольника . . . 469
7.2.2 Правило параллелограмма . . . 470
7.2.3 Свойства сложения векторов . . . 471
7.2.4 Вычитание векторов . . . 473
7.3 Умножение скаляра на вектор . . . 475
7.3.1 Что такое умножение скаляра на вектор? . . . 475
7.3.2 Свойства умножения скаляра на вектор . . . 476
7.4 Угол между векторами . . . 479
7.4.1 Что такое угол между векторами? . . . 479
7.4.2 Угол между вектором и осью . . . 479
7.5 Проекция вектора на ось . . . 480
7.5.1 Что такое проекция вектора на ось? . . . 480
7.5.2 Свойства проектирования вектора на ось . . . 481
7.5.3 Операция проектирования в физике . . . 483
7.6 Векторы и координаты на плоскости . . . 484
7.6.1 Разложение вектора по базису . . . 484
7.6.2 Нахождение модуля вектора по его проекциям . . . 484
7.7 Векторы и координаты в пространстве . . . 486
7.7.1 Разложение вектора по базису . . . 486
7.7.2 Нахождение модуля вектора по его проекциям . . . 486
7.8 Скалярное произведение векторов . . . 487
7.8.1 Что такое скалярное произведение? . . . 487
7.8.2 Свойства скалярного произведения . . . 488
7.8.3 Скалярное произведение в физике . . . 489
7.8.4 Вычисление скалярного произведения в координатах . . . 489

Глава 1

Механика

Механика изучает механическое движение тел. Полёт камня и движение автомобиля, суточное и орбитальное вращение Земли, колебания маятника и распространение звука — всё это примеры механического движения.

Не каждое движение является механическим. Скажем, распространение электромагнитных волн не описывается механикой и подчиняется совсем другим законам. Тут работает другой раздел физики — электродинамика.

В механике принято выделять три основных части.

1. Кинематика. Кинематика рассматривает движение тела как таковое и не интересуется тем, почему это движение возникло. Тело каким-то образом движется — вот давайте и будем исследовать характеристики его движения. Траектория, путь, перемещение, скорость, ускорение — примеры физических величин, с которыми имеет дело кинематика.

Поскольку причины возникновения движения не выясняются, из поля зрения кинематики выпадают такие величины, как масса и сила.

Кинематике посвящены следующие разделы данной главы:

→ Механическое движение
→ Равномерное прямолинейное движение
→ Равноускоренное движение
→ Равномерное движение по окружности
→ Путь при неравномерном движении

2. Динамика. Динамика изучает причины возникновения механического движения. В динамике рассматриваются взаимодействия тел, в результате чего появляются новые понятия: масса, сила, импульс, работа, энергия.

Динамика излагается в следующих разделах:

→ Первый закон Ньютона
→ Масса и плотность
→ Второй и третий законы Ньютона
→ Сила упругости
→ Сила тяготения
→ Сила трения
→ Импульс
→ Энергия
→ Простые механизмы
→ Механические колебания
→ Механические волны

Динамика стоит на «трёх китах» — трёх законах Ньютона. Законы Ньютона являются первичными утверждениями, или постулатами: они основаны на многочисленных опытных фактах

и не являются логическим следствием каких-то других утверждений. Попросту говоря, законы Ньютона ни откуда не выводятся[1]; они просто констатируют факт — вот по таким правилам живёт природа.

3. Статика. Статика — сравнительно небольшая часть механики, изучающая условия равновесия тела.

В статике твёрдого тела появляется понятие момента силы, а необходимым условием равновесия служит так называемое правило моментов. Статика жидкостей и газов изучает равновесие тел в этих средах; основную роль тут играют законы Паскаля и Архимеда.

Статике посвящены следующие два раздела:

→ Статика твёрдого тела
→ Статика жидкостей и газов

* * *

Значительно больше внимания (чем это принято в школьных учебниках) уделено использованию производной. Автор не считает нужным «скрывать» от школьников, что производная является естественным инструментом физики. Наоборот, чем скорее и лучше школьник освоится с этим аппаратом, тем проще будет ему впоследствии перейти к вузовским курсам общей физики и теоретической механики.

Поэтому первый раздел «Производная» настоящей главы посвящён дифференцированию. Изложение математических вопросов ведётся на физическом уровне строгости: опуская значительную долю формализма, мы стараемся вывести на первый план основные идеи, связанные с понятием производной. В частности, мы рассказываем о дифференцировании векторов (чего в школе обычно не делают). В вузе, как показывает опыт, никто уже не будет заниматься «разжёвыванием» этого материала.

[1] В продвинутых курсах теоретической физики законы Ньютона выводятся из более общих принципов. Но тогда уже эти новые принципы становятся постулатами — то есть первичными утверждениями, ни откуда не вытекающими.

1.1 Производная

Производная скалярной или векторной функции есть скорость изменения этой функции. В физике мы постоянно интересуемся быстротой изменения каких-либо величин. Вот почему использование производной пронизывает всю физику.

Строгое математическое определение производной опирается на понятие предела, которое в школе не проходят. Но определение предела нам сейчас и незачем. Самое главное — уловить основную идею, которая лежит в основе понятия предела.

1.1.1 Предел

Рассмотрим последовательность:

$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots, \frac{1}{n}, \ldots$$

Изобразим члены данной последовательности на числовой оси (рис. 1.1).

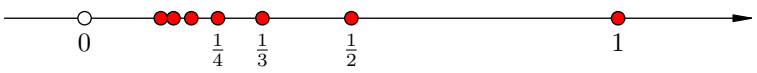


Рис. 1.1. Последовательность чисел $1/n$ ($n \in \mathbb{N}$)

Мы видим, что наши числа неограниченно приближаются к нулю (но никогда его не достигают). Начиная с $n = 10$ все члены последовательности окажутся на расстоянии не более $1/10$ от нуля; начиная с $n = 100$ все они будут на расстоянии не более $1/100$ от нуля; начиная с $n = 1000$ все они будут на расстоянии не более $1/1000$ от нуля и т. д.

Говорят, что последовательность $1/n$ *стремится к нулю*, или *сходится к нулю*, или что *предел* этой последовательности равен нулю. Записывают это так:

$$\lim_{n\to\infty} \frac{1}{n} = 0.$$

Образно говоря, наша последовательность «втекает» в точку 0. Понятие предела как раз и отражает факт этого «втекания».

Точно так же последовательность

$$a_n = 3 + \frac{1}{n} \quad (n \in \mathbb{N})$$

будет «втекать» в точку 3. Поэтому

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} \left(3 + \frac{1}{n}\right) = 3.$$

Подчеркнём, что «втекание последовательности в точку a» означает, что вблизи числа a находятся *все* члены данной последовательности, начиная с некоторого номера. Более точно, смысл выражения «предел последовательности a_n равен a» таков: какое бы расстояние ε мы наперёд ни задали, *все* числа a_n, начиная с некоторого номера, будут находиться от числа a на расстоянии меньше ε.

Например, знакопеременная последовательность $1, -1, 1, -1, \ldots$ не имеет предела: она не «втекает» ни в какую точку. Почему, например, число 1 не является пределом данной последовательности? Потому что найдётся бесконечно много членов последовательности (а именно, все члены с чётными номерами, равные -1), удалённых от точки 1 на расстояние 2. Иными

словами, не найдётся такого номера, начиная с которого все члены данной последовательности окажутся достаточно близко к точке 1.

Можно говорить не только о пределе последовательности, но и о пределе функции. Напомним, что функция $y = f(x)$ — это некоторое правило, которое позволяет для любого допустимого числа x получить единственное соответствующее ему число y. При этом число x называется *аргументом* функции, а число y — *значением* функции.

Нас будет интересовать понятие предела функции в точке. Оно формализует ту же самую идею «втекания». Только на сей раз график функции $y = f(x)$ будет «втекать» в некоторую точку координатной плоскости, когда аргумент x стремится к некоторому значению.

Так, на рис. 1.2 вы видите хорошо известную параболу — график функции $y = x^2$. Возьмём значение $x = 2$ и отметим на графике соответствующую точку $A(2, 4)$.

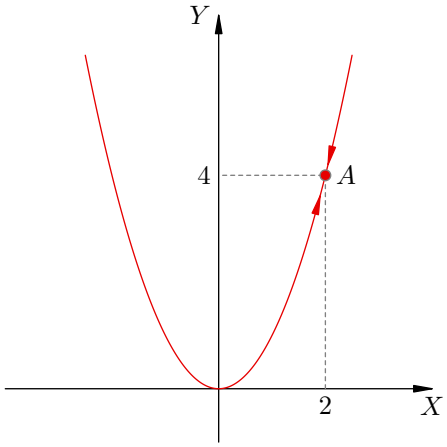


Рис. 1.2. График функции $y = x^2$

Представим себе, что x приближается к 2 (справа или слева — неважно). При этом график «втекает» в точку A, что и показано на рисунке стрелками. Иными словами, значение функции стремится к 4, и данный факт записывается следующим образом:

$$\lim_{x \to 2} x^2 = 4. \tag{1.1}$$

«А что тут такого особенного? — скажете вы. — Ясно же, что если x стремится к 2, то x^2 стремится к $2^2 = 4$. Зачем огород городить, говоря о каких-то пределах?»

Здесь не всё так просто. Взгляните на рис. 1.3.

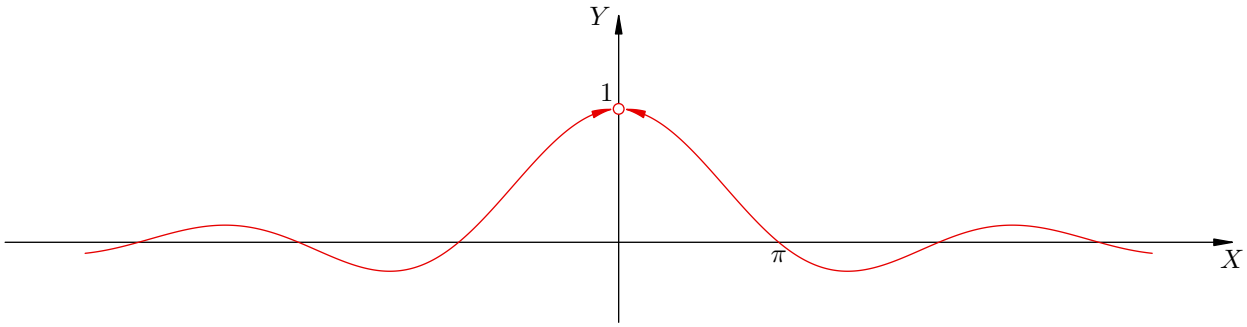


Рис. 1.3. График функции $y = \dfrac{\sin x}{x}$

Перед вами график функции

$$f(x) = \frac{\sin x}{x}\,.$$

И вот что интересно: значение функции при $x = 0$ не определено (при попытке вычислить $f(0)$ мы получаем нуль в знаменателе), но при этом график «втекает» в точку $(0, 1)$. То есть, хотя $f(0)$ не существует, тем не менее при $x \to 0$ значение функции стремится к числу 1. Иными словами, существует предел:

$$\lim_{x \to 0} \frac{\sin x}{x} = 1. \tag{1.2}$$

Он называется *первым замечательным пределом*.

Вы легко можете убедиться в справедливости формулы (1.2), взяв в руки калькулятор. Переведите его в режим «радианы» и вычислите:

$$\frac{\sin 0{,}1}{0{,}1}, \quad \frac{\sin 0{,}01}{0{,}01}, \quad \frac{\sin 0{,}001}{0{,}001}, \quad \dots$$

Вы увидите, что значение дроби становится всё ближе и ближе к единице.

Уяснив, что такое предел, мы теперь обсудим важнейшее физическое понятие *мгновенной скорости*. Оно вплотную подведёт нас к определению производной.

1.1.2 Мгновенная скорость

Спидометр автомобиля показывает 60 км/ч. Что это значит? Ответ простой: если автомобиль будет ехать так в течение часа, то он проедет 60 км.

Допустим, однако, что автомобиль вовсе не собирается ехать так целый час. Например, водитель разгоняет автомобиль с места, давит на газ, в какой-то момент бросает взгляд на спидометр и видит стрелку на отметке 60 км/ч. В следующий момент стрелка уползёт ещё выше. Как же понимать, что *в данный момент времени* скорость равна 60 км/ч?

Давайте выясним это на примере. Предположим, что путь s, пройденный автомобилем, зависит от времени t следующим образом:

$$s(t) = t^2,$$

где путь измеряется в метрах, а время — в секундах. То есть, при $t = 0$ путь равен нулю, к моменту времени $t = 1$ пройденный путь равен $s(1) = 1$, к моменту времени $t = 2$ путь равен $s(2) = 4$, к моменту времени $t = 3$ путь равен $s(3) = 9$, и так далее.

Видно, что идёт разгон, то есть автомобиль набирает скорость с течением времени. Действительно:

- за первую секунду пройдено расстояние 1;
- за вторую секунду пройдено расстояние $s(2) - s(1) = 3$;
- за третью секунду пройдено расстояние $s(3) - s(2) = 5$,

и далее по нарастающей.

А теперь вопрос. Пусть, например, через три секунды после начала движения наш водитель взглянул на спидометр. Что покажет стрелка? Иными словами, какова *мгновенная* скорость автомобиля в момент времени $t = 3$?

Просто поделить путь на время не получится: привычная формула $v = s/t$ работает только для *равномерного* движения (то есть когда стрелка спидометра застыла в некотором фиксированном положении). Но именно эта формула лежит в основе способа, позволяющего найти мгновенную скорость.

Идея способа такова. Отсчитаем от нашего момента $t = 3$ небольшой промежуток времени Δt, найдём путь Δs, пройденный автомобилем за этот промежуток, и поделим Δs на Δt. *Чем меньше будет Δt, тем точнее мы приблизимся к искомой величине мгновенной скорости.*

Давайте посмотрим, как эта идея реализуется. Возьмём для начала $\Delta t = 1$. Тогда

$$\Delta s = s(4) - s(3) = 4^2 - 3^2 = 7,$$

и для скорости получаем:

$$\frac{\Delta s}{\Delta t} = \frac{7}{1} = 7 \tag{1.3}$$

(скорость, разумеется, измеряется в м/с).

Будем уменьшать промежуток Δt. Берём $\Delta t = 0{,}1$:

$$\Delta s = s(3{,}1) - s(3) = 3{,}1^2 - 3^2 = 0{,}61,$$

$$\frac{\Delta s}{\Delta t} = \frac{0{,}61}{0{,}1} = 6{,}1. \tag{1.4}$$

Теперь берём $\Delta t = 0{,}01$:

$$\Delta s = s(3{,}01) - s(3) = 3{,}01^2 - 3^2 = 0{,}0601,$$

$$\frac{\Delta s}{\Delta t} = \frac{0{,}0601}{0{,}01} = 6{,}01. \tag{1.5}$$

Ну и возьмём ещё $\Delta t = 0{,}001$:

$$\Delta s = s(3{,}001) - s(3) = 3{,}001^2 - 3^2 = 0{,}006001,$$

$$\frac{\Delta s}{\Delta t} = \frac{0{,}006001}{0{,}001} = 6{,}001. \tag{1.6}$$

Глядя на значения (1.3)–(1.6), мы понимаем, что величина $\Delta s/\Delta t$ приближается к числу 6. Это означает, что мгновенная скорость автомобиля в момент времени $t = 3$ составляет 6 м/с.

Таким образом, при безграничном уменьшении Δt путь Δs также стремится к нулю, но отношение $\Delta s/\Delta t$ стремится к некоторому пределу v, который и называется *мгновенной скоростью* в данный момент времени t:

$$v = \lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t}\,. \tag{1.7}$$

Можно написать и так:

$$v(t) = \lim_{\Delta t \to 0} \frac{s(t + \Delta t) - s(t)}{\Delta t}\,. \tag{1.8}$$

Давайте вернёмся к нашему примеру с $s(t) = t^2$ и проделаем в общем виде те выкладки, которые выше были выполнены с числами. Итак:

$$\Delta s = s(t + \Delta t) - s(t) = (t + \Delta t)^2 - t^2 = t^2 + 2t\Delta t + \Delta t^2 - t^2 = \Delta t(2t + \Delta t),$$

и для мгновенной скорости имеем:

$$v(t) = \lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t} = \lim_{\Delta t \to 0} \frac{\Delta t(2t + \Delta t)}{\Delta t} = \lim_{\Delta t \to 0} (2t + \Delta t) = 2t. \tag{1.9}$$

В частности, при $t = 3$ формула (1.9) даёт: $v(3) = 2 \cdot 3 = 6$, как и было получено выше.

Теперь мы располагаем всеми необходимыми предварительными сведениями и полностью готовы перейти к обсуждению производной.

1.1.3 Определение производной

Скорость бывает не только у автомобиля. Мы можем говорить о скорости изменения чего угодно — например, физической величины или экономического показателя. Производная как раз и служит обобщением понятия мгновенной скорости на случай абстрактных математических функций.

Рассмотрим функцию $y = f(x)$. Напомним, что x называется *аргументом* данной функции. Отметим на оси X некоторое значение аргумента x, а на оси Y — соответствующее значение функции $f(x)$ (рис. 1.4).

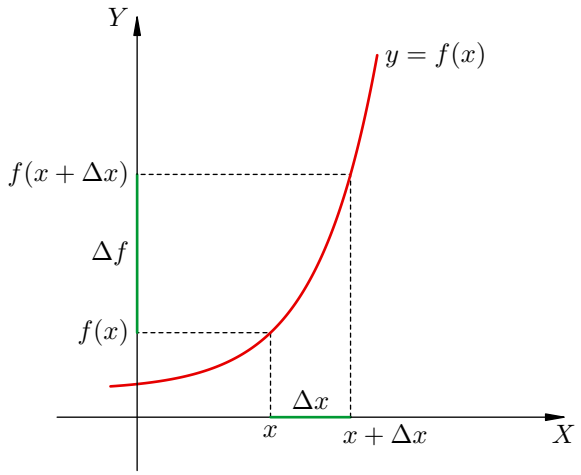


Рис. 1.4. Приращение аргумента и приращение функции

Дадим аргументу x некоторое *приращение*, обозначаемое Δx. Попадём в точку $x + \Delta x$. Обозначим её на рисунке вместе с соответствующим значением функции $f(x + \Delta x)$.

Величина

$$\Delta f = f(x + \Delta x) - f(x) \tag{1.10}$$

называется *приращением функции*, которое отвечает данному приращению аргумента Δx.

Вы видите сходство с предыдущим пунктом? Приращение аргумента Δx есть абстрактный аналог промежутка времени Δt, а соответствующее приращение функции Δf — это аналог пути Δs, пройденного за время Δt. Но на этом аналогия не заканчивается. Производная — это в точности аналог мгновенной скорости.

Определение. Производная $f'(x)$ функции $f(x)$ в точке x — это предел отношения приращения функции к приращению аргумента, когда приращение аргумента стремится к нулю:

$$f'(x) = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}\,. \tag{1.11}$$

Сравните с формулами (1.7) и (1.8). По сути написано одно и то же, не правда ли? Можно сказать, что производная — это мгновенная скорость изменения функции.

Нахождение производной функции называется *дифференцированием*. Нам предстоит научиться дифференцировать различные функции.

Прежде всего нужно знать несколько стандартных производных, которые называются табличными. Самые простые табличные производные вычисляются непосредственно с помощью формулы (1.11).

1.1.4 Табличные производные

Начнём с функции, которая является константой: $f(x) = c$. Приращение этой функции равно нулю:

$$\Delta f = f(x + \Delta x) - f(x) = c - c = 0.$$

Соответственно, обращается в нуль и производная:

$$f'(x) = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} = \lim_{\Delta x \to 0} \frac{0}{\Delta x} = \lim_{\Delta x \to 0} 0 = 0.$$

Итак, имеем первый результат — *производная константы равна нулю*:

$$\boxed{c' = 0.}$$

Теперь будем дифференцировать степенную функцию, то есть функцию вида $f(x) = x^a$. Найдём производную простейшей такой функции: $f(x) = x$. Приращение функции:

$$\Delta f = f(x + \Delta x) - f(x) = x + \Delta x - x = \Delta x.$$

Производная:

$$f'(x) = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta x}{\Delta x} = \lim_{\Delta x \to 0} 1 = 1.$$

Итак,

$$x' = 1.$$

Перейдём к функции $f(x) = x^2$. Это абстрактный аналог рассмотренной выше физической ситуации с $s(t) = t^2$, в которой мы искали мгновенную скорость. Нам остаётся лишь повторить (в других обозначениях) те вычисления, которые привели нас к формуле (1.9).

Приращение функции:

$$\Delta f = f(x + \Delta x) - f(x) = (x + \Delta x)^2 - x^2 = x^2 + 2x\Delta x + \Delta x^2 - x^2 = \Delta x(2x + \Delta x).$$

Производная:

$$f'(x) = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta x(2x + \Delta x)}{\Delta x} = \lim_{\Delta x \to 0} (2x + \Delta x) = 2x.$$

Таким образом,

$$\left(x^2\right)' = 2x.$$

Точно так же можно показать, что:

$$\begin{aligned} \left(x^3\right)' &= 3x^2, \\ \left(x^4\right)' &= 4x^3, \\ &\dots \\ \left(x^n\right)' &= nx^{n-1}. \end{aligned}$$

Оказывается, последняя формула справедлива не только для целого n, но и вообще для любого показателя степени a:

$$\boxed{(x^a)' = ax^{a-1}, \quad a \in \mathbb{R}.} \tag{1.12}$$

Найдём с помощью этой формулы производную функции $f(x) = \sqrt{x}$:

$$\left(\sqrt{x}\right)' = \left(x^{\frac{1}{2}}\right)' = \frac{1}{2} x^{\frac{1}{2}-1} = \frac{1}{2} x^{-\frac{1}{2}} = \frac{1}{2\sqrt{x}}\,.$$

Эта производная встречается очень часто, и её имеет смысл выучить. Запомнить можно так: «производная корня есть один делить на два корня».

Перейдём к тригонометрическим функциям: синусу и косинусу. Вычисления с помощью формулы (1.11) и первого замечательного предела приводят к следующему результату:

$$\boxed{(\sin x)' = \cos x, \quad (\cos x)' = -\sin x.}$$

Четыре производных в рамочке (константа, степенная функция, синус и косинус), как мы сказали выше, называются *табличными*. Эти производные нужно твёрдо знать.

Вычисления производной по определению (то есть как предела) легко проходят для функций, устроенных наиболее просто. А как быть, если нужно продифференцировать функцию наподобие такой: $f(x) = x^7 \sin \sqrt[3]{4x^2 - 5x}$? Здесь вычислять предел (1.11) — занятие не из приятных. В подобных случаях на помощь приходят *правила дифференцирования*, которые позволяют сконструировать производную данной функции из производных более простых функций.

1.1.5 Правила дифференцирования

Как мы уже сказали, правила дифференцирования позволяют находить производные функций достаточно сложного вида. Идея состоит в «расщеплении» исходной функции на более простые функции, производные которых известны и играют роль «кирпичиков» при конструировании искомой производной. Зная небольшое число табличных производных и располагая правилами дифференцирования, мы можем вычислять производные огромного количества функций, не прибегая к определению производной и не вычисляя соответствующий предел (1.11).

Всего имеется пять правил дифференцирования. Мы приводим их здесь без доказательства. Функции $u(x)$ и $v(x)$ являются теми самыми «кирпичиками», из которых строятся функции более сложного вида.

0. Константа выносится за знак производной. Если c — число, то $(cu)' = cu'$.

Данное правило легко получается в качестве следствия правила 2 о дифференцировании произведения. Но применяется оно настолько часто, что мы сделали его «нулевым» правилом, обособленным от остальных.

Согласно этому правилу имеем, например:

$$(5x^2)' = 5(x^2)' = 10x,$$
$$(-3\sin x)' = -3(\sin x)' = -3\cos x.$$

1. Дифференцирование суммы. $(u + v)' = u' + v'$ (*производная суммы равна сумме производных*).

Так, применяя правила 0 и 1, находим:

$$(\sin x + \cos x)' = (\sin x)' + (\cos x)' = \cos x - \sin x,$$
$$(x^3 + 4\cos x - 10)' = (x^3)' + (4\cos x)' + (-10)' = 3x^2 - 4\sin x$$

(производная константы -10 равна нулю!).

2. Дифференцирование произведения. $(uv)' = u'v + uv'$.

Вот пример дифференцирования произведения:

$$(x^2 \sin x)' = (x^2)' \sin x + x^2 (\sin x)' = 2x \sin x + x^2 \cos x.$$

А вот как получается правило 0:

$$(cu)' = c'u + cu' = cu',$$

поскольку $c' = 0$.

3. Дифференцирование частного.

$$\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}\,.$$

Правило дифференцирования частного позволяет найти, например, производную тангенса:

$$(\operatorname{tg} x)' = \left(\frac{\sin x}{\cos x}\right)' = \frac{(\sin x)'\cos x - \sin x(\cos x)'}{\cos^2 x} = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x}\,.$$

Нам осталось обсудить последнее правило — дифференцирование сложной функции. Мы сначала объясним, что такое сложная функция, затем продемонстрируем правило дифференцирования на примерах, и только потом — когда станет ясно, как оно работает — дадим формулировку этого правила.

Пусть, например, $u(x) = \sin x$ и $v(x) = \sqrt{x}$. Давайте сначала извлекать корень из x (то есть применять к x функцию v), а потом брать синус полученного числа (то есть действовать на полученное число $v(x)$ функцией u). Тогда возникает функция:

$$u(v(x)) = \sin\sqrt{x}.$$

Это и есть *сложная* функция, или *композиция* функций u и v. Идея понятна: число x поступает на вход *первой* функции v, а полученное число $v(x)$ поступает на вход *второй* функции u.

Можно, наоборот, сделать u первой функцией, а v — второй. Тогда сначала от x будет вычисляться синус, а потом из синуса извлекаться корень. Получится другая сложная функция:

$$v(u(x)) = \sqrt{\sin x}.$$

Дифференцирование сложной функции — это как снятие листов с кочана капусты. Сначала находим производную второй («внешней») функции и умножаем её на производную первой («внутренней») функции. Применительно к нашим примерам это выглядит так:

$$(\sin\sqrt{x})' = \cos\sqrt{x}\cdot(\sqrt{x})' = \cos\sqrt{x}\cdot\frac{1}{2\sqrt{x}}\,,$$
$$(\sqrt{\sin x})' = \frac{1}{2\sqrt{\sin x}}\cdot(\sin x)' = \frac{1}{2\sqrt{\sin x}}\cdot\cos x.$$

Приведём для ясности ещё один пример:

$$[(4x^2+3x+2)^5]' = 5(4x^2+3x+2)^4\cdot(4x^2+3x+2)' = 5(4x^2+3x+2)^4\cdot(8x+3).$$

И ещё пример (очень важный для физики; здесь A, ω и α — константы):

$$[A\sin(\omega x+\alpha)]' = A\cos(\omega x+\alpha)\cdot(\omega x+\alpha)' = A\omega\cos(\omega x+\alpha).$$

Понятно, как работает правило? Тогда — формулировка.

4. Дифференцирование сложной функции. $[u(v(x))]' = u'(v(x))v'(x)$.

1.1.6 Обозначения производной в физике

Переходя к физическим приложениям производной, мы будем использовать несколько иные обозначения — те, которые приняты в физике.

Во-первых, меняется обозначение функций. В самом деле, какие функции мы собираемся дифференцировать? Этими функциями служат физические величины, зависящие от времени. Например, координата тела $x(t)$ и его скорость $v(t)$ могут быть заданы формулами:

$$x(t) = 1 + 12t - 3t^2, \tag{1.13}$$

$$v(t) = 12 - 6t. \tag{1.14}$$

Таким образом, аргументом функции теперь является время t, а буква x отныне обозначает функцию — координату точки.

Во-вторых, меняется обозначение производной. Штрих в физике зарезервирован для других целей, и вместо него мы используем точку над буквой:

$$\text{производная функции } x(t) \text{ обозначается } \dot{x}(t) \tag{1.15}$$

(читается «икс с точкой»).

Имеется ещё одно обозначение производной, очень распространённое как в математике, так и в физике:

$$\text{производная функции } x(t) \text{ обозначается } \frac{dx}{dt} \tag{1.16}$$

(читается «дэ икс по дэ тэ»).

Остановимся подробнее на смысле обозначения (1.16). Математик понимает его двояко — либо как предел:

$$\frac{dx}{dt} = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = \lim_{\Delta t \to 0} \frac{x(t + \Delta t) - x(t)}{\Delta t}, \tag{1.17}$$

либо как дробь, в знаменателе которой стоит приращение времени dt, а в числителе — так называемый *дифференциал* dx функции $x(t)$. Понятие дифференциала не сложно, но мы не будем его сейчас обсуждать; оно ждёт вас на первом курсе.

Физик, не скованный требованиями математической строгости, понимает обозначение (1.16) более неформально. Пусть dx есть изменение координаты за время dt. Возьмём интервал dt настолько маленьким, что отношение dx/dt близко к своему пределу (1.17) с устраивающей нас точностью.

И тогда, — скажет физик, — *производная координаты по времени есть попросту дробь, в числителе которой стоит достаточно малое изменение координаты dx, а в знаменателе — достаточно малый промежуток времени dt, в течение которого это изменение координаты произошло.*

Такое нестрогое понимание производной характерно для рассуждений в физике. Далее мы будем придерживаться именно этого физического уровня строгости.

Производная $\dot{x}(t)$ физической величины $x(t)$ снова является функцией времени, и эту функцию снова можно продифференцировать — найти производную производной, или *вторую производную* функции $x(t)$. Вот одно обозначение второй производной:

$$\text{вторая производная функции } x(t) \text{ обозначается } \ddot{x}(t)$$

(читается «икс с двумя точками»), а вот другое:

$$\text{вторая производная функции } x(t) \text{ обозначается } \frac{d^2x}{dt^2}$$

(читается «дэ два икс по дэ тэ квадрат» или «дэ два икс по дэ тэ дважды»).

Давайте вернёмся к исходному примеру (1.13) и посчитаем производную координаты, а заодно посмотрим на совместное использование обозначений (1.15) и (1.16):

$$x(t) = 1 + 12t - 3t^2 \Rightarrow$$
$$\dot{x}(t) = \frac{d}{dt}(1 + 12t - 3t^2) = 12 - 6t.$$

(Символ дифференцирования $\frac{d}{dt}$ перед скобкой — это всё равно что штрих сверху за скобкой в прежних обозначениях.)

Обратите внимание, что производная координаты оказалась равна скорости (1.14). Это не случайное совпадение. Связь производной координаты со скоростью тела будет выяснена в следующем разделе «Механическое движение».

1.1.7 Предел векторной величины

Физические величины бывают не только скалярными, но и векторными. Соответственно, часто нас интересует скорость изменения векторной величины — то есть, производная вектора. Однако прежде чем говорить о производной, нужно разобраться с понятием предела векторной величины.

Рассмотрим последовательность векторов $\vec{u}_1, \vec{u}_2, \vec{u}_3, \ldots$ Сделав, если необходимо, параллельный перенос, сведём их начала в одну точку O (рис. 1.5):

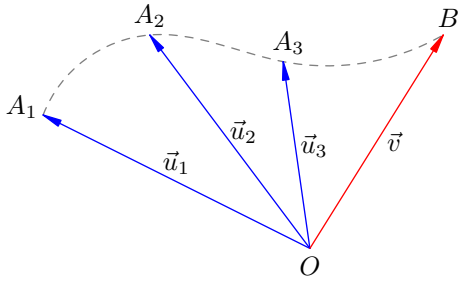


Рис. 1.5. $\lim\limits_{n\to\infty} \vec{u}_n = \vec{u}$

Концы векторов обозначим $A_1, A_2, A_3, \ldots$ Таким образом, имеем:

$$\vec{u}_1 = \overrightarrow{OA_1}, \quad \vec{u}_2 = \overrightarrow{OA_2}, \quad \vec{u}_3 = \overrightarrow{OA_3}, \quad \ldots$$

Предположим, что последовательность точек $A_1, A_2, A_3, \ldots$ «втекает»[2] в точку B:

$$\lim_{n\to\infty} A_n = B.$$

Обозначим $\vec{v} = \overrightarrow{OB}$. Мы скажем тогда, что последовательность синих векторов $\vec{u}_n$ стремится к красному вектору $\vec{v}$, или что вектор $\vec{v}$ является пределом последовательности векторов $\vec{u}_n$:

$$\vec{v} = \lim_{n\to\infty} \vec{u}_n.$$

[2]Вполне достаточно интуитивного понимания этого «втекания», но вас, быть может, интересует более строгое объяснение? Тогда вот оно.

Пусть дело происходит на плоскости. «Втекание» последовательности $A_1, A_2, A_3, \ldots$ в точку B означает следующее: сколь бы малый круг с центром в точке B мы ни взяли, *все* точки последовательности, начиная с некоторой, попадут внутрь этого круга. Иными словами, вне любого круга с центром B имеется лишь конечное число точек нашей последовательности.

А если дело происходит в пространстве? Определение «втекания» модифицируется незначительно: нужно лишь заменить слово «круг» на слово «шар».

Предположим теперь, что концы синих векторов на рис. 1.5 пробегают не дискретный набор значений, а непрерывную кривую (например, указанную пунктирной линией). Таким образом, мы имеем дело не с последовательностью векторов $\vec{u}_n$, а с вектором $\vec{u}(t)$, который меняется со временем. Это как раз то, что нам и нужно в физике!

Дальнейшее объяснение почти такое же. Пусть t стремится к некоторому значению t_0. Если при этом концы векторов $\vec{u}(t)$ «втекают» в некоторую точку B, то мы говорим, что вектор $\vec{v} = \overrightarrow{OB}$ является пределом векторной величины $\vec{u}(t)$:

$$\vec{v} = \lim_{t \to t_0} \vec{u}(t).$$

1.1.8 Дифференцирование векторов

Выяснив, что такое предел векторной величины, мы готовы сделать следующий шаг — ввести понятие производной вектора.

Предположим, что имеется некоторый вектор $\vec{u}(t)$, зависящий от времени. Это означает, что длина данного вектора и его направление могут меняться с течением времени.

По аналогии с обычной (скалярной) функцией вводится понятие изменения (или приращения) вектора. *Изменение вектора* $\vec{u}$ за время Δt есть векторная величина:

$$\Delta\vec{u} = \vec{u}(t + \Delta t) - \vec{u}(t).$$

Обратите внимание, что в правой части данного соотношения стоит *разность векторов*. Изменение вектора $\vec{u}$ показано на рис. 1.6 (напомним, что при вычитании векторов мы сводим их начала в одну точку, соединяем концы и «укалываем» стрелкой тот вектор, из которого производится вычитание).

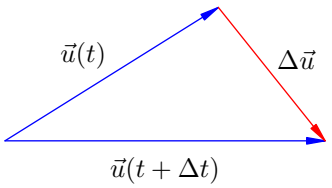


Рис. 1.6. Изменение вектора

Если промежуток времени Δt достаточно мал, то и вектор $\vec{u}$ за это время меняется мало (в физике, по крайней мере, так считается всегда). Соответственно, если при $\Delta t \to 0$ отношение $\Delta\vec{u}/\Delta t$ стремится к некоторому пределу, то этот предел называется *производной вектора* $\vec{u}$:

$$\frac{d\vec{u}}{dt} = \lim_{\Delta t \to 0} \frac{\Delta\vec{u}}{\Delta t} = \lim_{\Delta t \to 0} \frac{\vec{u}(t + \Delta t) - \vec{u}(t)}{\Delta t}. \tag{1.18}$$

При обозначении производной вектора мы не будем использовать точку сверху (так как символ $\dot{\vec{u}}$ не слишком хорошо смотрится) и ограничиваемся обозначением (1.18). Но для производной скаляра мы, разумеется, свободно используем оба обозначения.

Напомним, что $d\vec{u}/dt$ — это *символ* производной. Его можно понимать и как дробь, в числителе которой стоит *дифференциал* вектора $\vec{u}$, соответствующий промежутку времени dt. Выше мы не стали обсуждать понятие дифференциала, так как в школе его не проходят; не будем обсуждать дифференциал и здесь.

Однако на физическом уровне строгости производную $d\vec{u}/dt$ можно считать дробью, в знаменателе которой стоит очень малый интервал времени dt, а в числителе — соответствующее малое изменение $d\vec{u}$ вектора $\vec{u}$. При достаточно малом dt величина данной дроби отличается от

предела в правой части (1.18) столь мало, что с учётом имеющейся точности измерений этим отличием можно пренебречь.

Этого (не вполне строгого) физического понимания производной нам окажется вполне достаточно.

Правила дифференцирования векторных выражений во многом аналогичны правилам дифференцирования скаляров. Нам понадобятся лишь самые простые правила.

1. Постоянный скалярный множитель выносится за знак производной: если $c = \text{const}$, то

$$\frac{d(c\vec{u})}{dt} = c\,\frac{d\vec{u}}{dt}\,.$$

Мы используем это правило в разделе «Импульс», когда второй закон Ньютона

$$m\frac{d\vec{v}}{dt} = \vec{F}$$

будет переписан в виде:

$$\frac{d(m\vec{v})}{dt} = \vec{F}.$$

2. Постоянный векторный множитель выносится за знак производной: если $\vec{c} = \text{const}$, то

$$\frac{d}{dt}\left(x(t)\vec{c}\right) = \dot{x}(t)\vec{c}.$$

3. Производная суммы векторов равна сумме их производных:

$$\frac{d}{dt}\left(\vec{u} + \vec{v}\right) = \frac{d\vec{u}}{dt} + \frac{d\vec{v}}{dt}\,.$$

Последними двумя правилами мы будем пользоваться неоднократно. Посмотрим, как они работают в важнейшей ситуации дифференцирования вектора при наличии в пространстве прямоугольной системы координат $OXYZ$ (рис. 1.7).

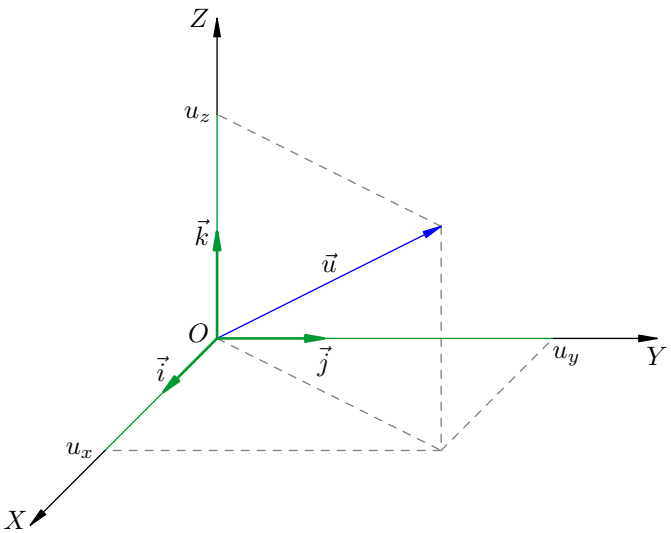


Рис. 1.7. Разложение вектора по базису

Как известно, любой вектор $\vec{u}$ единственным образом раскладывается по базису единичных векторов $\vec{i}$, $\vec{j}$, $\vec{k}$:

$$\vec{u} = u_x\vec{i} + u_y\vec{j} + u_z\vec{k}.$$

Здесь u_x, u_y, u_z — проекции вектора $\vec{u}$ на координатные оси. Они же являются координатами вектора $\vec{u}$ в данном базисе.

Вектор $\vec{u}$ в нашем случае зависит от времени, а это значит, что его координаты u_x, u_y, u_z являются функциями времени:

$$\vec{u}(t) = u_x(t)\,\vec{i} + u_y(t)\,\vec{j} + u_z(t)\vec{k}. \tag{1.19}$$

Дифференцируем это равенство. Сначала пользуемся правилом дифференцирования суммы:

$$\frac{d\vec{u}}{dt} = \frac{d}{dt}\left(u_x(t)\vec{i}\right) + \frac{d}{dt}\left(u_y(t)\vec{j}\right) + \frac{d}{dt}\left(u_z(t)\vec{k}\right).$$

Затем выносим постоянные векторы за знак производной:

$$\frac{d\vec{u}}{dt} = \dot{u}_x(t)\vec{i} + \dot{u}_y(t)\vec{j} + \dot{u}_z(t)\vec{k}. \tag{1.20}$$

Таким образом, если вектор $\vec{u}$ имеет координаты (u_x, u_y, u_z), то координаты производной $d\vec{u}/dt$ являются производными координат вектора $\vec{u}$, а именно — $(\dot{u}_x, \dot{u}_y, \dot{u}_z)$.

Ввиду особой важности формулы (1.20) дадим более непосредственный её вывод.

В момент времени $t + \Delta t$ согласно (1.19) имеем:

$$\vec{u}(t+\Delta t) = u_x(t+\Delta t)\,\vec{i} + u_y(t+\Delta t)\,\vec{j} + u_z(t+\Delta t)\vec{k}.$$

Напишем изменение вектора $\vec{u}$:

$$\begin{aligned}
\Delta\vec{u} &= \vec{u}(t+\Delta t) - \vec{u}(t) = \\
&= \left(u_x(t+\Delta t)\,\vec{i} + u_y(t+\Delta t)\,\vec{j} + u_z(t+\Delta t)\vec{k}\right) - \left(u_x(t)\,\vec{i} + u_y(t)\,\vec{j} + u_z(t)\vec{k}\right) = \\
&= \left(u_x(t+\Delta t) - u_x(t)\right)\vec{i} + \left(u_y(t+\Delta t) - u_y(t)\right)\vec{j} + \left(u_z(t+\Delta t) - u_z(t)\right)\vec{k} = \\
&= \Delta u_x \cdot \vec{i} + \Delta u_y \cdot \vec{j} + \Delta u_z \cdot \vec{k}.
\end{aligned}$$

Делим обе части полученного равенства на Δt:

$$\frac{\Delta\vec{u}}{\Delta t} = \frac{\Delta u_x}{\Delta t}\,\vec{i} + \frac{\Delta u_y}{\Delta t}\,\vec{j} + \frac{\Delta u_z}{\Delta t}\,\vec{k}.$$

В пределе при $\Delta t \to 0$ дроби $\Delta u_x/\Delta t$, $\Delta u_y/\Delta t$, $\Delta u_z/\Delta t$ переходят соответственно в производные $\dot{u}_x$, $\dot{u}_y$, $\dot{u}_z$, и мы снова получаем соотношение (1.20):

$$\frac{d\vec{u}}{dt} = \dot{u}_x\vec{i} + \dot{u}_y\vec{j} + \dot{u}_z\vec{k}.$$

1.2 Механическое движение

Понятие движения является чрезвычайно общим и охватывает самый широкий круг явлений. В физике изучают различные виды движения. Простейшим из них является механическое движение.

Механическое движение — это изменение положение тела (или его частей) в пространстве относительно других тел с течением времени.

1.2.1 Относительность движения

Если тело A меняет своё положение относительно тела B, то и тело B меняет своё положение относительно тела A. Иначе говоря, если тело A движется относительно тела B, то и тело B движется относительно тела A. Механическое движение является *относительным* — для описания движения необходимо указать, относительно какого тела оно рассматривается.

Так, например, можно говорить о движении поезда относительно земли, пассажира относительно поезда, мухи относительно пассажира и т. д. Понятия абсолютного движения и абсолютного покоя не имеют смысла: пассажир, покоящийся относительно поезда, будет двигаться с ним относительно столба на дороге, совершать вместе с Землёй суточное вращение и двигаться по орбите вокруг Солнца.

Тело, относительно которого рассматривается движение, называется *телом отсчёта*.

1.2.2 Основная задача механики

Основной задачей механики является определение положения движущегося тела в любой момент времени. Для решения этой задачи удобно представить движение тела как изменение координат его точек с течением времени.

Чтобы измерить координаты, нужна система координат. Чтобы измерять время, нужны часы. Всё это вместе образует систему отсчёта.

Система отсчёта — это тело отсчёта вместе с жёстко связанной с ним («вмороженной» в него) системой координат и часами.

Система отсчёта показана на рис. 1.8. Движение точки M рассматривается в прямоугольной системе координат $OXYZ$. Начало координат O является телом отсчёта.

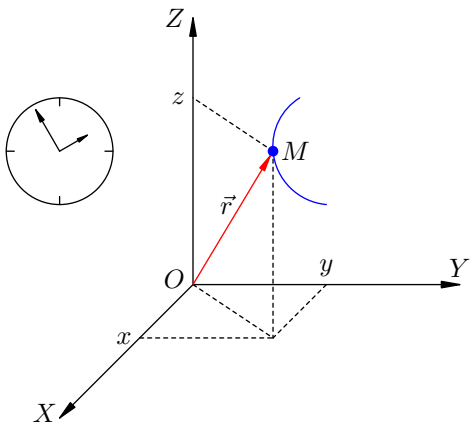


Рис. 1.8. Система отсчёта

Вектор $\vec{r} = \overrightarrow{OM}$ называется *радиус-вектором* точки M. Три координаты x, y, z точки M являются в то же время координатами её радиус-вектора $\vec{r}$.

Решить основную задачу механики для точки M — это значит найти её координаты как функции времени:

$$x = x(t), \quad y = y(t), \quad z = z(t); \tag{1.21}$$

или, что то же самое, — найти зависимость радиус-вектора точки M от времени:

$$\vec{r} = \vec{r}(t). \tag{1.22}$$

Соотношения (1.21) или (1.22) мы будем называть *законом движения*. Таким образом, решение основной задачи механики для точки M состоит в нахождении закона движения этой точки.

1.2.3 Материальная точка

В ряде случаев можно отвлечься от формы и размеров изучаемого объекта и рассматривать его просто как движущуюся точку.

Материальная точка — это тело, размерами которого можно пренебречь в условиях данной задачи.

Так, поезд можно считать материальной точкой при его движении из Москвы в Саратов, но не при посадке в него пассажиров. Землю можно считать материальной точкой при описании её движения вокруг Солнца, но не её суточного вращения вокруг собственной оси.

К характеристикам механического движения материальной точки относятся траектория, путь, перемещение, скорость и ускорение.

1.2.4 Траектория, путь, перемещение

В дальнейшем, говоря о движущемся (или покоящемся) теле, мы всегда полагаем, что тело можно принять за материальную точку. Случаи, когда идеализацией материальной точки пользоваться нельзя, будут специально оговариваться.

- *Траектория* — это линия, вдоль которой движется тело. На рис. 1.8 траекторией точки M является синяя дуга, которую описывает в пространстве конец радиус-вектора $\vec{r}$.
- *Путь* — это длина участка траектории, пройденного телом за данный промежуток времени.
- *Перемещение* — это вектор, соединяющий начальное и конечное положение тела.

Предположим, что тело начало движение в точке A и закончило движение в точке B (рис. 1.9).

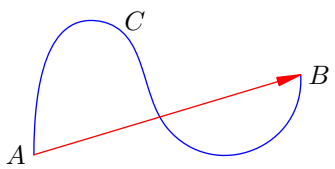


Рис. 1.9. Путь и перемещение

Путь, пройденный телом, есть длина синей дуги — траектории ACB. Перемещение тела — это красный вектор $\overrightarrow{AB}$.

1.2.5 Скорость

В предыдущем разделе «Производная» уже говорилось о мгновенной скорости. Мы определили её как предел:

$$v = \lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t},$$

где Δs — путь, пройденный за малый промежуток времени Δt. При таком определении мгновенная скорость оказывается не чем иным, как производной пути s:

$$v = \frac{ds}{dt}. \tag{1.23}$$

Данное определение, однако, не охватывает всего разнообразия ситуаций, встречающихся в механике. Дело в том, что скорость является *вектором* — она обладает как абсолютной величиной (модулем), так и направлением в пространстве. Между тем, формула (1.23) говорит нам лишь о модуле скорости, но не об её направлении. Стало быть, определение (1.23) нуждается в обобщении.

Рассмотрим движение тела в прямоугольной системе координат с базисом $\vec{i}$, $\vec{j}$, $\vec{k}$ (рис. 1.10).

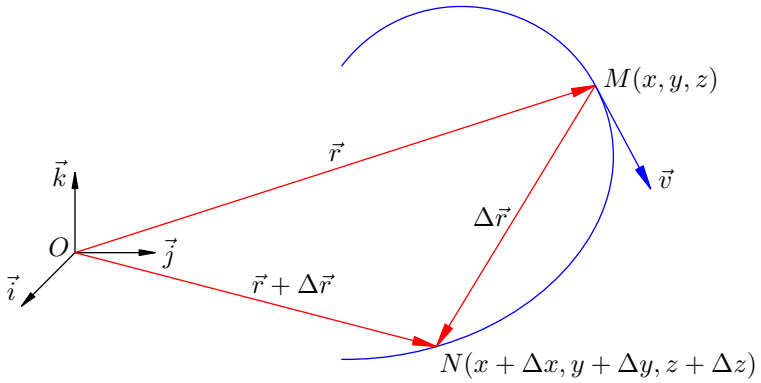


Рис. 1.10. К определению мгновенной скорости

Пусть в момент времени t тело находилось в точке $M(x, y, z)$ с радиус-вектором

$$\overrightarrow{OM} = \vec{r} = x\vec{i} + y\vec{j} + z\vec{k}. \tag{1.24}$$

Спустя малый промежуток времени Δt тело оказалось в точке $N(x + \Delta x, y + \Delta y, z + \Delta z)$ с радиус-вектором

$$\overrightarrow{ON} = \vec{r} + \Delta\vec{r}.$$

Перемещение тела есть вектор $\Delta\vec{r} = \overrightarrow{MN}$. Теперь мы будем говорить не о пределе отношения $\Delta s/\Delta t$ пути ко времени, а о пределе отношения $\Delta\vec{r}/\Delta t$ перемещения ко времени — и тем самым придём к нужному нам обобщению понятия мгновенной скорости на векторный случай.

Итак, *мгновенная скорость* $\vec{v}$ в момент времени t — это предел отношения перемещения $\Delta\vec{r}$ к интервалу времени Δt, когда величина этого интервала стремится к нулю; иными словами, скорость точки — это производная её радиус-вектора:

$$\vec{v} = \lim_{\Delta t \to 0} \frac{\Delta\vec{r}}{\Delta t} = \frac{d\vec{r}}{dt}. \tag{1.25}$$

Отношение $\Delta\vec{r}/\Delta t$ — это вектор, направленный вдоль секущей MN. Когда Δt стремится к нулю, точка N приближается к точке M, а секущая MN превращается в касательную. Соответственно, вектор мгновенной скорости $\vec{v}$ направлен по касательной к траектории в точке M. Это и показано на рис. 1.10.

Продолжаем вычисления в координатах. Пользуясь равенством (1.24) и правилами дифференцирования векторов, получим:

$$\vec{v} = \frac{d}{dt}\left(x\vec{i} + y\vec{j} + z\vec{k}\right) = \dot{x}\vec{i} + \dot{y}\vec{j} + \dot{z}\vec{k}. \tag{1.26}$$

С другой стороны, вектор $\vec{v}$ единственным образом раскладывается по базису $\vec{i}$, $\vec{j}$, $\vec{k}$:

$$\vec{v} = v_x\vec{i} + v_y\vec{j} + v_z\vec{k}. \tag{1.27}$$

Сопоставляя формулы (1.26) и (1.27), мы видим, что проекции вектора скорости на координатные оси являются производными координат точки:

$$v_x = \dot{x}, \quad v_y = \dot{y}, \quad v_z = \dot{z}.$$

1.2.6 Ускорение

Формула (1.25) говорит о том, что вектор скорости характеризует быстроту изменения радиус-вектора тела. Но скорость тела также может меняться — быстрее или медленнее. Ускорение — это характеристика быстроты изменения вектора скорости.

Предположим, что в момент времени t скорость тела равна $\vec{v}$, а спустя малый интервал Δt скорость стала равна $\vec{v} + \Delta\vec{v}$.

Ускорение $\vec{a}$ — это предел отношения изменения скорости $\Delta\vec{v}$ к интервалу Δt, когда этот интервал стремится к нулю; иначе говоря, ускорение — это производная скорости:

$$\vec{a} = \lim_{\Delta t \to 0} \frac{\Delta\vec{v}}{\Delta t} = \frac{d\vec{v}}{dt}.$$

Ускорение, так сказать, есть «скорость изменения скорости». Имеем:

$$\vec{a} = \frac{d}{dt}\left(v_x\vec{i} + v_y\vec{j} + v_z\vec{k}\right) = \dot{v}_x\vec{i} + \dot{v}_y\vec{j} + \dot{v}_z\vec{k}.$$

Следовательно, проекции ускорения являются производными проекций скорости:

$$a_x = \dot{v}_x, \quad a_y = \dot{v}_y, \quad a_z = \dot{v}_z.$$

Скорость, в свою очередь, есть производная радиус-вектора. Поэтому ускорение, будучи производной скорости, оказывается *второй производной* радиус-вектора (то есть результатом двукратного дифференцирования вектора $\vec{r}$):

$$\vec{a} = \frac{d}{dt}\left(\frac{d\vec{r}}{dt}\right) = \frac{d^2\vec{r}}{dt^2}.$$

Соответственно, проекции ускорения являются вторыми производными координат точки:

$$a_x = \ddot{x}, \quad a_y = \ddot{y}, \quad a_z = \ddot{z}.$$

1.2.7 Примеры вычисления скорости и ускорения

Итак, знание закона движения (зависимости координат тела от времени) позволяет находить скорость и ускорение тела — нужно лишь вычислить первые и вторые производные координат. Рассмотрим несколько примеров таких вычислений.

Пример. Вернёмся к примеру (1.13):

$$x = 1 + 12t - 3t^2$$

(координата измеряется в метрах, время — в секундах). Последовательно дифференцируя два раза, получаем:

$$v_x = \dot{x} = 12 - 6t,$$
$$a_x = \dot{v}_x = -6.$$

Как видим, ускорение постоянно по модулю и равно 6 м/с^2. Направлено ускорение в сторону, противоположную оси X.

Приведённый пример есть случай *равноускоренного* движения, при котором модуль и направление ускорения неизменны. Равноускоренное движение — один из важнейших и часто встречающихся видов движения в механике.

Из данного примера нетрудно понять, что при равноускоренном движении проекция скорости является линейной функцией времени, а координата — квадратичной функцией. Мы поговорим об этом более подробно в соответствующем разделе, посвящённом равноускоренному движению.

Пример. Рассмотрим более экзотический случай:

$$x = 2 + 3t - 4t^2 + 5t^3.$$

Дифференцируем:

$$v_x = \dot{x} = 3 - 8t + 15t^2,$$
$$a_x = \dot{v}_x = -8 + 30t.$$

Данное движение не является равноускоренным: ускорение зависит от времени.

Пример. Пусть тело движется вдоль оси X по следующему закону:

$$x = 5\sin 2t.$$

Мы видим, что координата тела периодически изменяется, находясь в пределах от -5 до 5. Данное движение является примером *гармонических колебаний*, когда координата меняется со временем по закону синуса.

Дифференцируем дважды:

$$v_x = \dot{x} = 5\cos 2t \cdot 2 = 10\cos 2t,$$
$$a_x = \dot{v}_x = -20\sin 2t.$$

Проекция скорости меняется по закону косинуса, а проекция ускорения — снова по закону синуса. Величина a_x пропорциональна координате x и противоположна ей по знаку (а именно, $a_x = -4x$); вообще, соотношение вида $a_x = -\omega^2 x$ характерно для гармонических колебаний.

1.2.8 Закон сложения скоростей

Пусть имеются две системы отсчёта. Одна из них связана с неподвижным телом отсчёта O. Эту систему отсчёта обозначим K и будем называть *неподвижной*.

Вторая система отсчёта, обозначаемая K', связана с телом отсчёта O', которое движется относительно тела O со скоростью $\vec{u}$. Эту систему отсчёта называем *движущейся*. Дополнительно

предполагаем, что координатные оси системы K' перемещаются параллельно самим себе (нет вращения системы координат), так что вектор $\vec{u}$ можно считать скоростью движущейся системы относительно неподвижной.

Неподвижная система отсчёта K обычно связана с землёй. Если поезд плавно едет по рельсам со скоростью $\vec{u}$, то система отсчёта, связанная с вагоном поезда, будет движущейся системой отсчёта K'.

Заметим, что скорость *любой* точки вагона[3] равна $\vec{u}$. Если муха неподвижно сидит в некоторой точке вагона, то относительно земли муха движется со скоростью $\vec{u}$. Муха *переносится* вагоном, и потому скорость $\vec{u}$ движущейся системы относительно неподвижной называется *переносной скоростью*.

Предположим теперь, что муха поползла по вагону. Тогда появляются ещё две скорости, которые нужно рассмотреть.

Скорость мухи относительно вагона (то есть в движущейся системе K') обозначается $\vec{v}\,'$ и называется *относительной скоростью*.

Скорость мухи относительно земли (то есть в неподвижной системе K) обозначается $\vec{v}$ и называется *абсолютной скоростью*.

Выясним, как связаны друг с другом эти три скорости — абсолютная, относительная и переносная.

На рис. 1.11 муха обозначена точкой M. Далее:

$\vec{r}$ — радиус-вектор точки M в неподвижной системе K;

$\vec{r}\,'$ — радиус-вектор точки M в движущейся системе K';

$\vec{R}$ — радиус-вектор тела отсчёта O' в неподвижной системе K.

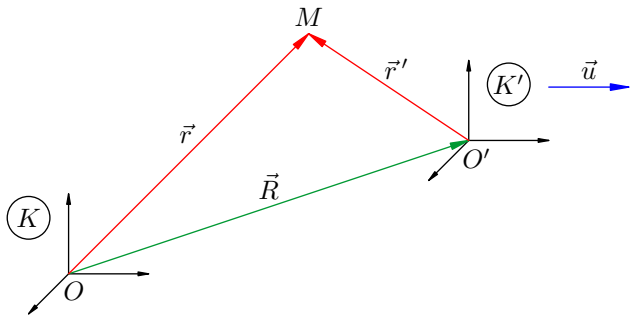


Рис. 1.11. К выводу закона сложения скоростей

Как видно из рисунка,

$$\vec{r} = \vec{R} + \vec{r}\,'.$$

Дифференцируя это равенство, получим:

$$\frac{d\vec{r}}{dt} = \frac{d\vec{R}}{dt} + \frac{d\vec{r}\,'}{dt}\,. \tag{1.28}$$

Производная $d\vec{r}/dt$ есть скорость точки M в системе K, то есть абсолютная скорость:

$$\frac{d\vec{r}}{dt} = \vec{v}.$$

Аналогично, производная $d\vec{r}\,'/dt$ есть скорость точки M в системе K', то есть относительная скорость:

$$\frac{d\vec{r}\,'}{dt} = \vec{v}\,'.$$

[3]Кроме вращающихся колёс, но их мы не берём во внимание.

А что такое $d\vec{R}/dt$? Это скорость точки O' в неподвижной системе, то есть — переносная скорость $\vec{u}$ движущейся системы относительно неподвижной:

$$\frac{d\vec{R}}{dt} = \vec{u}.$$

В результате из (1.28) получаем:

$$\vec{v} = \vec{u} + \vec{v}\,'.$$

Закон сложения скоростей. Скорость точки относительно неподвижной системы отсчёта равна векторной сумме скорости движущейся системы и скорости точки относительно движущейся системы. Иными словами, абсолютная скорость есть сумма переносной и относительной скоростей.

Таким образом, если муха ползёт по движущемуся вагону, то скорость мухи относительно земли равна векторной сумме скорости вагона и скорости мухи относительно вагона. Интуитивно очевидный результат!

1.2.9 Виды механического движения

Простейшими видами механического движения материальной точки являются равномерное и прямолинейное движения.

Движение называется *равномерным*, если модуль вектора скорости остаётся постоянным (направление скорости при этом может меняться).

Движение называется *прямолинейным*, если оно происходит вдоль некоторой прямой (величина скорости при этом может меняться). Иными словами, траекторией прямолинейного движения служит прямая линия.

Например, автомобиль, который едет с постоянной скоростью по извилистой дороге, совершает равномерное (но не прямолинейное) движение. Автомобиль, разгоняющийся на прямом участке шоссе, совершает прямолинейное (но не равномерное) движение.

А вот если при движении тела остаются постоянными как модуль скорости, так и её направление, то движение называется *равномерным прямолинейным*. Итак:

- равномерное движение $\Leftrightarrow |\vec{v}| = \text{const}$;
- равномерное прямолинейное движение $\Leftrightarrow \vec{v} = \text{const}$.

Важнейшим частным случаем неравномерного движения является *равноускоренное* движение, при котором остаются постоянными модуль и направление вектора ускорения:

- равноускоренное движение $\Leftrightarrow \vec{a} = \text{const}$.

Наряду с материальной точкой в механике рассматривается ещё одна идеализация — твёрдое тело.

Твёрдое тело — это система материальных точек, расстояния между которыми не меняются со временем. Модель твёрдого тела применяется в тех случаях, когда мы не можем пренебречь размерами тела, но можем не принимать во внимание *изменение* размеров и формы тела в процессе движения.

Простейшими видами механического движения твёрдого тела являются поступательное и вращательное движения.

Движение тела называется *поступательным*, если всякая прямая, соединяющая две какие-либо точки тела, перемещается параллельно своему первоначальному направлению. При поступательном движении траектории всех точек тела идентичны: они получаются друг из друга параллельным сдвигом.

Так, на рис. 1.12 показано поступательное движение серого квадрата. Произвольно взятый зелёный отрезок этого квадрата перемещается параллельно самому себе. Траектории концов отрезка изображены синими пунктирными линиями.

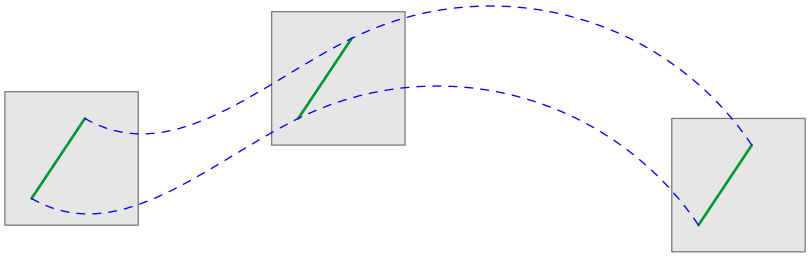


Рис. 1.12. Поступательное движение

Движение тела называется *вращательным*, если все его точки описывают окружности, лежащие в параллельных плоскостях. При этом центры данных окружностей лежат на одной прямой, которая перпендикулярна всем этим плоскостям и называется *осью вращения*.

На рис. 1.13 изображён шар, вращающийся вокруг вертикальной оси. Так обычно рисуют земной шар в соответствующих задачах динамики.

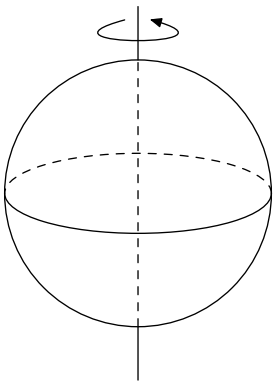


Рис. 1.13. Вращательное движение

1.3 Равномерное прямолинейное движение

Равномерное прямолинейное движение материальной точки — это движение с постоянной скоростью $\vec{v}$. Обратите внимание, что речь идёт о постоянстве *вектора* скорости; это значит, что скорость неизменна как по модулю, так и по направлению.

Траекторией тела при равномерном прямолинейном движении служит прямая (или часть прямой — например, отрезок или луч). Вдоль данной прямой тело движется равномерно, то есть с постоянной по модулю скоростью.

1.3.1 Закон движения

Предположим, что тело, двигаясь равномерно и прямолинейно со скоростью $\vec{v}$, переместилось за время t из точки M_0 в точку M (рис. 1.14). Вектор перемещения есть $\vec{s} = \overrightarrow{M_0M}$.

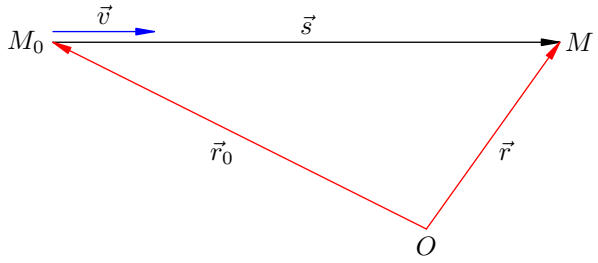


Рис. 1.14. Равномерное прямолинейное движение

Путь, пройденный телом, равен длине s вектора перемещения. Очевидно, что выполнено соотношение:

$$s = vt, \tag{1.29}$$

где v — модуль вектора скорости.

Формула (1.29) справедлива для произвольного равномерного движения (не обязательно прямолинейного). Но в случае прямолинейного равномерного движения эта формула становится соотношением между векторами. В самом деле, поскольку векторы $\vec{s}$ и $\vec{v}$ сонаправлены, формула (1.29) позволяет записать:

$$\vec{s} = \vec{v}t. \tag{1.30}$$

Как обычно, движение тела рассматривается в некоторой системе отсчёта, связанной с телом отсчёта O (рис. 1.14; координатные оси не изображаем). Пусть $\vec{r}_0$ — радиус-вектор начальной точки M_0 и $\vec{r}$ — радиус-вектор конечной точки M. Тогда, очевидно,

$$\vec{s} = \vec{r} - \vec{r}_0.$$

Подставим эту разность в формулу (1.30):

$$\vec{r} - \vec{r}_0 = \vec{v}t.$$

Отсюда получаем *закон движения* (то есть зависимость радиус-вектора тела от времени):

$$\vec{r} = \vec{r}_0 + \vec{v}t. \tag{1.31}$$

Напомним, что нахождение закона движения решает основную задачу механики, которая заключается в определении зависимости координат тела от времени. Переход от векторного соотношения (1.31) к координатам осуществляется элементарно.

Координаты точки M_0 обозначим (x_0, y_0, z_0). Они же являются координатами вектора $\vec{r}_0$. Координаты точки M (и вектора $\vec{r}$) обозначим (x, y, z). Тогда векторная формула (1.31) приводит к трём координатным соотношениям:

$$x = x_0 + v_x t, \tag{1.32}$$

$$y = y_0 + v_y t, \tag{1.33}$$

$$z = z_0 + v_z t. \tag{1.34}$$

Формулы (1.32)—(1.34), представляя координаты тела как функции времени, служат решением основной задачи механики для равномерного прямолинейного движения.

1.3.2 Интегрирование

Ключевая формула (1.31), описывающая равномерное прямолинейное движение, может быть получена из несколько иных соображений. Вспомним, что производная радиус-вектора есть скорость точки:

$$\frac{d\vec{r}}{dt} = \vec{v}. \tag{1.35}$$

В случае равномерного прямолинейного движения имеем $\vec{v} = \text{const}$. Что нужно продифференцировать, чтобы получить постоянный вектор $\vec{v}$? Очевидно, функцию $\vec{v}t$. Но не только: к величине $\vec{v}t$ можно прибавить любой постоянный вектор $\vec{c}$ (это не изменит производную, поскольку производная константы равна нулю). Таким образом:

$$\vec{r} = \vec{c} + \vec{v}t. \tag{1.36}$$

Каков смысл константы $\vec{c}$? Если $t = 0$, то радиус-вектор $\vec{r}$ равен своему начальному значению $\vec{r}_0$. Поэтому, полагая $t = 0$ в формуле (1.36), получим:

$$\vec{r}_0 = \vec{c}.$$

Итак, вектор $\vec{c}$ есть начальное значение радиус-вектора, и теперь из (1.36) мы снова приходим к формуле (1.31):

$$\vec{r} = \vec{r}_0 + \vec{v}t.$$

Мы, таким образом, *проинтегрировали* равенство (1.35) при условии, что $\vec{v} = \text{const}$. Интегрирование — это операция, обратная дифференцированию. Проинтегрировать — значит найти неизвестную функцию, если дана её производная.

Интегрировать в физике приходится очень часто. Например, закон движения определяется с помощью интегрирования. Вы только что убедились в этом; впоследствии у нас возникнут и другие примеры.

1.4 Равноускоренное движение

Равноускоренное движение — это движение с постоянным вектором ускорения $\vec{a}$. Таким образом, при равноускоренном движении остаются неизменными направление и абсолютная величина ускорения.

1.4.1 Зависимость скорости от времени

При изучении равномерного прямолинейного движения вопрос зависимости скорости от времени не возникал: скорость была постоянна в процессе движения. Однако при равноускоренном движении скорость меняется с течением времени, и эту зависимость нам предстоит выяснить.

Давайте ещё раз потренируемся в элементарном интегрировании. Исходим из того, что производная вектора скорости есть вектор ускорения:

$$\frac{d\vec{v}}{dt} = \vec{a}. \tag{1.37}$$

В нашем случае имеем $\vec{a} = \text{const}$. Что надо продифференцировать, чтобы получить постоянный вектор $\vec{a}$? Разумеется, функцию $\vec{a}t$. Но не только: к ней можно добавить ещё произвольный постоянный вектор $\vec{c}$ (ведь производная постоянного вектора равна нулю). Таким образом,

$$\vec{v} = \vec{c} + \vec{a}t. \tag{1.38}$$

Каков смысл константы $\vec{c}$? В начальный момент времени $t = 0$ скорость равна своему начальному значению: $\vec{v} = \vec{v}_0$. Поэтому, полагая $t = 0$ в формуле (1.38), получим:

$$\vec{v}_0 = \vec{c}.$$

Итак, константа $\vec{c}$ — это начальная скорость тела. Теперь соотношение (1.38) принимает свой окончательный вид:

$$\vec{v} = \vec{v}_0 + \vec{a}t. \tag{1.39}$$

В конкретных задачах мы выбираем систему координат и переходим к проекциям на координатные оси. Так, в прямоугольной декартовой системе координат $OXYZ$ векторное соотношение (1.39) даёт три скалярных равенства:

$$\begin{aligned} v_x &= v_{0x} + a_x t, \\ v_y &= v_{0y} + a_y t, \\ v_z &= v_{0z} + a_z t. \end{aligned}$$

1.4.2 Закон движения

Теперь мы можем найти закон движения, то есть зависимость радиус-вектора от времени. Вспоминаем, что производная радиус-вектора есть скорость тела:

$$\frac{d\vec{r}}{dt} = \vec{v}.$$

Подставляем сюда выражение для скорости, даваемое формулой (1.39):

$$\frac{d\vec{r}}{dt} = \vec{v}_0 + \vec{a}t. \tag{1.40}$$

Сейчас нам предстоит проинтегрировать равенство (1.40). Это несложно. Чтобы получить $\vec{v}_0$, надо продифференцировать функцию $\vec{v}_0 t$. Чтобы получить $\vec{a}t$, нужно продифференцировать выражение $\vec{a}t^2/2$. Не забудем добавить и произвольную константу $\vec{c}$:

$$\vec{r} = \vec{c} + \vec{v}_0 t + \frac{\vec{a}t^2}{2}\,.$$

Ясно, что $\vec{c}$ — это начальное значение $\vec{r}_0$ радиус-вектора $\vec{r}$ в момент времени $t = 0$. В результате получаем искомый закон равноускоренного движения:

$$\vec{r} = \vec{r}_0 + \vec{v}_0 t + \frac{\vec{a}t^2}{2}\,. \tag{1.41}$$

Переходя к проекциям на координатные оси, вместо одного векторного равенства (1.41) получаем три скалярных равенства:

$$x = x_0 + v_{0x} t + \frac{a_x t^2}{2}\,, \tag{1.42}$$

$$y = y_0 + v_{0y} t + \frac{a_y t^2}{2}\,, \tag{1.43}$$

$$z = z_0 + v_{0z} t + \frac{a_z t^2}{2}\,. \tag{1.44}$$

Формулы (1.42)—(1.44) дают зависимость координат тела от времени и поэтому служат решением основной задачи механики для равноускоренного движения.

Снова вернёмся к закону движения (1.41). Заметим, что $\vec{r} - \vec{r}_0 = \vec{s}$ — перемещение тела. Тогда получаем зависимость перемещения от времени:

$$\vec{s} = \vec{v}_0 t + \frac{\vec{a}t^2}{2}\,.$$

1.4.3 Прямолинейное равноускоренное движение

Если равноускоренное движение является прямолинейным, то удобно выбрать координатную ось вдоль прямой, по которой движется тело. Пусть, например, это будет ось OX. Тогда для решения задач нам достаточно будет трёх формул:

$$v_x = v_{0x} + a_x t,$$

$$x = x_0 + v_{0x} t + \frac{a_x t^2}{2}\,,$$

$$s_x = v_{0x} t + \frac{a_x t^2}{2}\,,$$

где $s_x = x - x_0$ — проекция перемещения на ось OX.

Но очень часто помогает ещё одна формула, являющаяся их следствием. Выразим из первой формулы время:

$$t = \frac{v_x - v_{0x}}{a_x}$$

и подставим в формулу для перемещения:

$$s_x = v_{0x}\frac{v_x - v_{0x}}{a_x} + \frac{a_x}{2}\left(\frac{v_x - v_{0x}}{a_x}\right)^2.$$

Преобразуем:

$$s_x = \frac{v_{0x}v_x - v_{0x}^2}{a_x} + \frac{v_x^2 - 2v_x v_{0x} + v_{0x}^2}{2a_x}\,,$$

и окончательно получаем:

$$s_x = \frac{v_x^2 - v_{0x}^2}{2a_x}\,.$$

Эта формула не содержит времени t и позволяет быстрее приходить к ответу в тех задачах, где время не фигурирует.

1.4.4 Свободное падение

Важным частным случаем равноускоренного движения является *свободное падение*. Так называется движение тела вблизи поверхности Земли без учёта сопротивления воздуха.

Свободное падение тела, независимо от его массы, происходит с постоянным *ускорением свободного падения* $\vec{g}$, направленным вертикально вниз. Почти во всех задачах при расчётах полагают $g = 10\,\text{м/с}^2$.

Давайте разберём несколько задач и посмотрим, как работают выведенные нами формулы для равноускоренного движения.

Задача. Найти скорость приземления дождевой капли, если высота тучи $h = 2$ км.

Решение. Направим ось OY вертикально вниз, расположив начало отсчёта в точке отрыва капли. Воспользуемся формулой

$$s_y = \frac{v_y^2 - v_{0y}^2}{2a_y}\,.$$

Имеем: $s_y = h$, $v_y = v$ — искомая скорость приземления, $v_{0y} = 0$, $a_y = g$. Получаем: $h = \dfrac{v^2}{2g}$, откуда $v = \sqrt{2gh}$. Вычисляем: $v = \sqrt{2 \cdot 10 \cdot 2000} = 200\,\text{м/с}$. Это 720 км/ч, порядка скорости пули.

На самом деле капли дождя падают со скоростью порядка нескольких метров в секунду. Почему такое расхождение? Сопротивление воздуха!

Задача. Тело брошено вертикально вверх со скоростью $v_0 = 30$ м/с. Найти его скорость через $t = 5$ с.

Решение. Направим ось OY вертикально вверх, поместив начало отсчёта на поверхности Земли. Используем формулу

$$v_y = v_{0y} + a_y t.$$

Здесь $v_{0y} = v_0$, $a_y = -g$, так что $v_y = v_0 - gt$. Вычисляем: $v_y = 30 - 10 \cdot 5 = -20$ м/с. Значит, скорость будет равна 20 м/с. Знак проекции указывает на то, что тело будет лететь вниз.

Задача. С балкона, находящегося на высоте $h = 15$ м, бросили вертикально вверх камень со скоростью $v_0 = 10$ м/с. Через какое время камень упадёт на землю?

Решение. Направим ось OY вертикально вверх, поместив начало отсчёта на поверхности Земли. Берём формулу

$$y = y_0 + v_{0y}t + \frac{a_y t^2}{2}\,.$$

Имеем: $y = 0$, $y_0 = h$, $v_{0y} = v_0$, $a_y = -g$, так что $0 = h + v_0 t - \dfrac{gt^2}{2} = 15 + 10t - 5t^2$, или $t^2 - 2t - 3 = 0$. Решая квадратное уравнение, получим $t = 3$ с.

1.4.5 Горизонтальный бросок

Равноускоренное движение не обязательно является прямолинейным. Рассмотрим движение тела, брошенного горизонтально.

Предположим, что тело брошено горизонтально со скоростью v_0 с высоты h. Найдём время и дальность полёта, а также выясним, по какой траектории происходит движение.

Выберем систему координат OXY так, как показано на рис. 1.15.

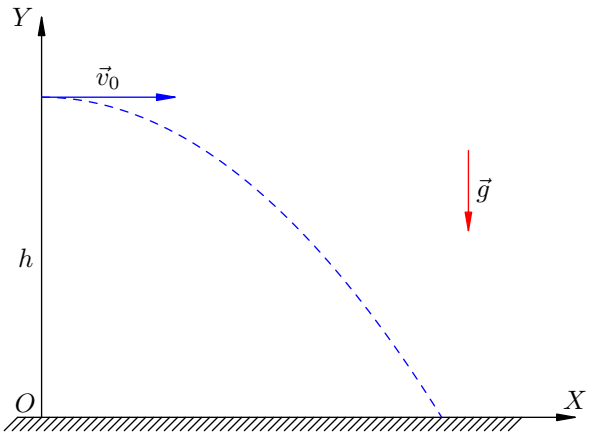


Рис. 1.15. Горизонтальный бросок

Используем формулы:

$$x = x_0 + v_{0x}t + \frac{a_x t^2}{2}\,, \quad y = y_0 + v_{0y}t + \frac{a_y t^2}{2}\,.$$

В нашем случае $x_0 = 0$, $v_{0x} = v_0$, $a_x = 0$, $y_0 = h$, $v_{0y} = 0$, $a_y = -g$. Получаем:

$$x = v_0 t, \quad y = h - \frac{gt^2}{2}\,. \tag{1.45}$$

Время полёта T найдём из условия, что в момент падения координата тела y обращается в нуль:

$$y(T) = 0 \Rightarrow h - \frac{gT^2}{2} = 0 \Rightarrow T = \sqrt{\frac{2h}{g}}\,.$$

Дальность полёта L — это значение координаты x в момент времени T:

$$L = x(T) = v_0 T = v_0\sqrt{\frac{2h}{g}}\,.$$

Уравнение траектории получим, исключая время из уравнений (1.45). Выражаем t из первого уравнения и подставляем во второе:

$$t = \frac{x}{v_0} \Rightarrow$$
$$y = h - \frac{g}{2}\left(\frac{x}{v_0}\right)^2 = h - \frac{gx^2}{2v_0^2}\,.$$

Получили зависимость y от x, которая является уравнением параболы. Следовательно, тело летит по параболе.

1.4.6 Бросок под углом к горизонту

Рассмотрим несколько более сложный случай равноускоренного движения: полёт тела, брошенного под углом к горизонту.

Предположим, что тело брошено с поверхности Земли со скоростью v_0, направленной под углом α к горизонту. Найдём время и дальность полёта, а также выясним, по какой траектории двигается тело.

Выберем систему координат OXY так, как показано на рис. 1.16.

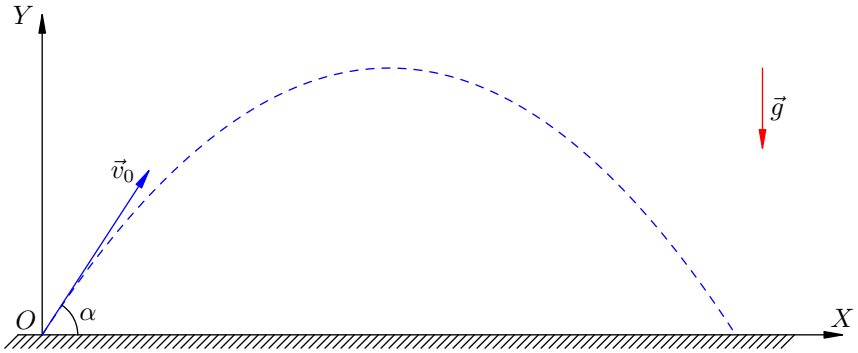


Рис. 1.16. Бросок под углом к горизонту

Начинаем с уравнений:

$$x = x_0 + v_{0x}t + \frac{a_x t^2}{2}\,, \quad y = y_0 + v_{0y}t + \frac{a_y t^2}{2}\,.$$

В нашем случае $x_0 = y_0 = 0$, $v_{0x} = v_0\cos\alpha$, $v_{0y} = v_0\sin\alpha$, $a_x = 0$, $a_y = -g$. Получаем:

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

Дальше действуем так же, как и в случае горизонтального броска. В результате приходим к соотношениям:

$$T = \frac{2v_0\sin\alpha}{g}\,,$$
$$L = \frac{v_0^2\sin 2\alpha}{g}\,,$$
$$y = x\,\mathrm{tg}\,\alpha - \frac{gx^2}{2v_0^2\cos^2\alpha}\,.$$

(Обязательно проделайте эти вычисления самостоятельно!) Как видим, зависимость y от x снова является уравнением параболы.

Попробуйте также показать, что максимальная высота подъёма определяется формулой:

$$H = \frac{v_0^2\sin^2\alpha}{2g}\,.$$

1.5 Равномерное движение по окружности

Равномерное движение по окружности — это достаточно простой пример движения с вектором ускорения, зависящим от времени.

Пусть точка вращается по окружности радиуса r. Скорость точки постоянна по модулю и равна v. Скорость v называется *линейной скоростью* точки.

Период обращения — это время одного полного оборота. Для периода T имеем очевидную формулу:

$$T = \frac{2\pi r}{v}\,. \tag{1.46}$$

Частота обращения — это величина, обратная периоду:

$$\nu = \frac{1}{T}\,.$$

Частота показывает, сколько полных оборотов точка совершает за секунду. Измеряется частота в об/с (обороты в секунду).

Пусть, например, $T = 0{,}1$ с. Это означает, что за время 0,1 с точка совершает один полный оборот. Частота при этом равна: $\nu = 1/0{,}1 = 10$ об/с; за секунду точка совершает 10 полных оборотов.

1.5.1 Угловая скорость

Рассмотрим равномерное вращение точки в декартовой системе координат. Поместим начало координат O в центре окружности (рис. 1.17).

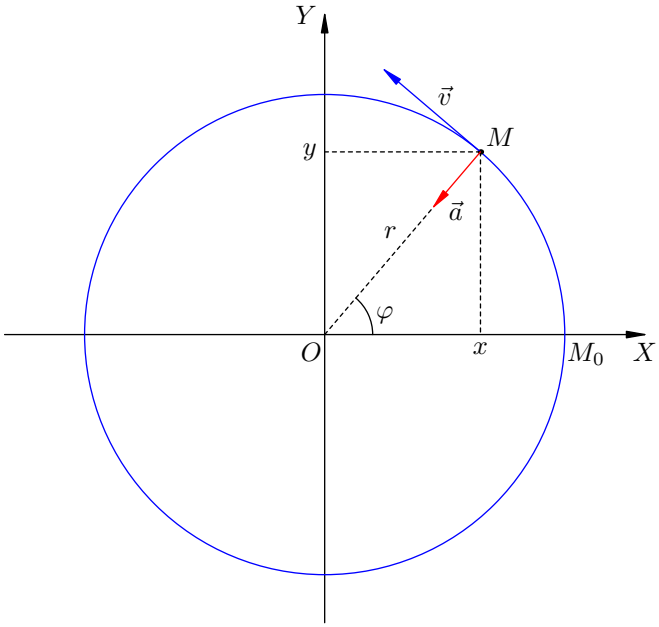


Рис. 1.17. Равномерное движение по окружности

Пусть M_0 — начальное положение точки; иными словами, при $t = 0$ точка имела координаты $(r, 0)$. Пусть за время t точка повернулась на угол φ и заняла положение M.

Отношение угла поворота ко времени называется *угловой скоростью* вращения точки:

$$\omega = \frac{\varphi}{t}\,. \tag{1.47}$$

Угол φ, как правило, измеряется в радианах, поэтому угловая скорость измеряется в рад/с.

За время, равное периоду вращения, точка поворачивается на угол 2π. Поэтому

$$\omega = \frac{2\pi}{T}\,. \tag{1.48}$$

Сопоставляя формулы (1.46) и (1.48), получаем связь линейной и угловой скоростей:

$$v = \omega r. \tag{1.49}$$

1.5.2 Закон движения

Найдём теперь зависимость координат вращающейся точки от времени. Видим из рис. 1.17, что

$$x = r\cos\varphi, \quad y = r\sin\varphi.$$

Но из формулы (1.47) имеем: $\varphi = \omega t$. Следовательно,

$$x = r\cos\omega t, \quad y = r\sin\omega t. \tag{1.50}$$

Формулы (1.50) являются решением основной задачи механики для равномерного движения точки по окружности.

1.5.3 Центростремительное ускорение

Теперь нас интересует ускорение вращающейся точки. Его можно найти, дважды продифференцировав соотношения (1.50):

$$\begin{aligned} v_x = \dot{x} = -\omega r\sin\omega t, \quad v_y = \dot{y} = \omega r\cos\omega t,\\ a_x = \dot{v}_x = -\omega^2 r\cos\omega t, \quad a_y = \dot{v}_y = -\omega^2 r\sin\omega t. \end{aligned}$$

С учётом формул (1.50) имеем:

$$a_x = -\omega^2 x, \quad a_y = -\omega^2 y. \tag{1.51}$$

Полученные формулы (1.51) можно записать в виде одного векторного равенства:

$$\vec{a} = -\omega^2\vec{r}, \tag{1.52}$$

где $\vec{r}$ — радиус-вектор вращающейся точки.

Мы видим, что вектор ускорения направлен противоположно радиус-вектору, т. е. к центру окружности (см. рис. 1.17). Поэтому ускорение точки, равномерно движущейся по окружности, называется *центростремительным*.

Кроме того, из формулы (1.52) мы получаем выражение для модуля центростремительного ускорения:

$$a = \omega^2 r. \tag{1.53}$$

Выразим угловую скорость из (1.49):

$$\omega = \frac{v}{r}$$

и подставим в (1.53). Получим ещё одну формулу для центростремительного ускорения:

$$a = \frac{v^2}{r}\,.$$

1.6 Путь при неравномерном движении

Сейчас мы будем рассматривать *неравномерное* движение — то есть движение, при котором абсолютная величина скорости меняется со временем. Оказывается, существует простая геометрическая интерпретация пути, пройденного телом при произвольном движении.

Начнём с равномерного движения. Пусть скорость тела постоянна и равна v. Возьмём два момента времени: начальный момент t_1 и конечный момент t_2. Длительность рассматриваемого промежутка времени равна $\Delta t = t_2 - t_1$.

Очевидно, что за промежуток времени $[t_1, t_2]$ тело проходит путь:

$$s = v(t_2 - t_1) = v\Delta t. \tag{1.54}$$

Давайте построим график зависимости скорости от времени. В данном случае это будет прямая, параллельная оси абсцисс (рис. 1.18).

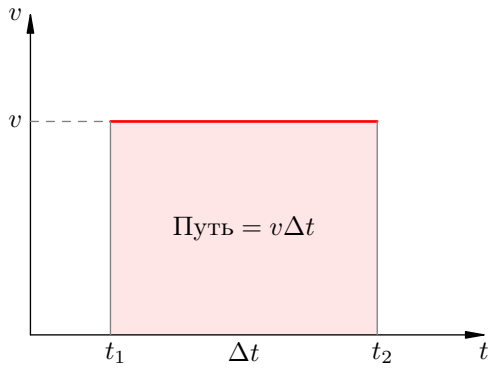


Рис. 1.18. Путь при равномерном движении

Нетрудно видеть, что *пройденный путь равен площади прямоугольника, расположенного под графиком скорости*. В самом деле, первый множитель v в формуле (1.54) есть вертикальная сторона этого прямоугольника, а второй множитель Δt — его горизонтальная сторона.

Теперь нам предстоит обобщить эту геометрическую интерпретацию на случай неравномерного движения.

Пусть скорость тела v зависит от времени, и на рассматриваемом промежутке $[t_1, t_2]$ график скорости выглядит, например, так (рис. 1.19):

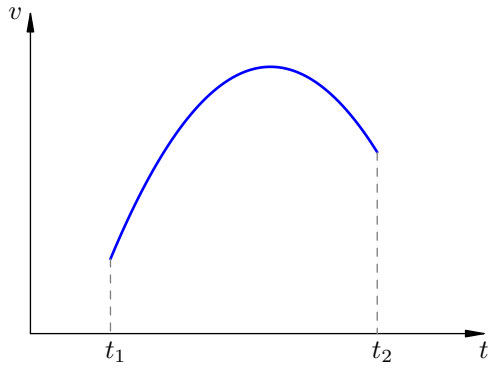


Рис. 1.19. Неравномерное движение

Дальше мы рассуждаем следующим образом.

1. Разобьём наш промежуток времени $[t_1, t_2]$ на небольшие отрезки величиной Δt.

2. Предположим, что на каждом таком отрезке тело движется с *постоянной* скоростью. То есть, плавное изменение скорости заменим ступенчатой аппроксимацией[4]: в течение каждого небольшого отрезка времени тело движется равномерно, а затем скорость тела мгновенно и скачком меняется.

   На рис. 1.20 показаны две ступенчатые аппроксимации. Ширина ступенек Δt на правом рисунке вдвое меньше, чем на левом.

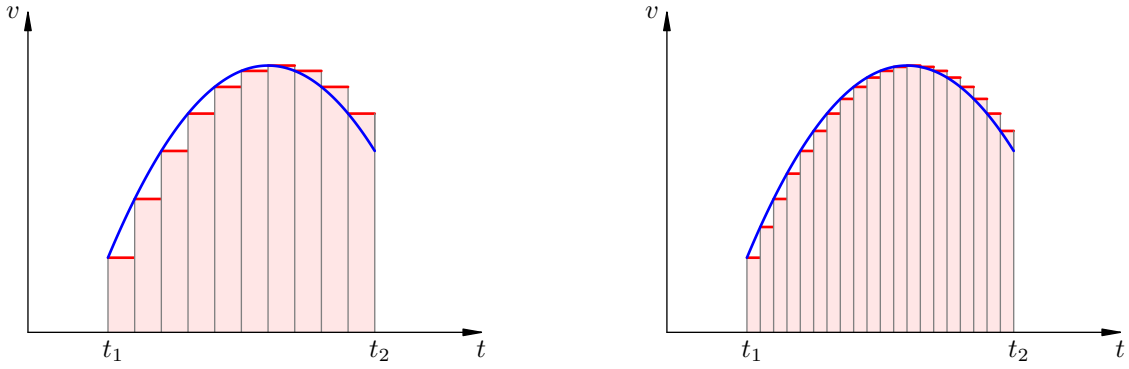


Рис. 1.20. Ступенчатая аппроксимация

   Путь, пройденный за время Δt равномерного движения — это площадь прямоугольника, расположенного под ступенькой. Поэтому путь, пройденный за всё время такого «ступенчатого» движения — это сумма площадей всех прямоугольников на графике.

3. Теперь устремляем Δt к нулю. Ясно, что в пределе наша ступенчатая аппроксимация перейдёт в исходный график скорости на рис. 1.19. Сумма площадей прямоугольников перейдёт в площадь под графиком скорости; следовательно, эта площадь и есть путь, пройденный телом за время от t_1 до t_2 (рис. 1.21).

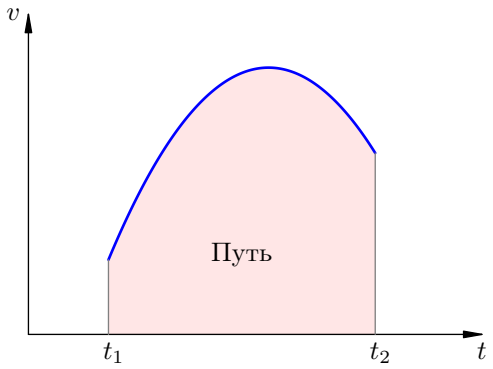


Рис. 1.21. Путь при неравномерном движении

В итоге мы приходим к нужному нам обобщению геометрической интерпретации пути, полученной выше для случая равномерного движения.

Геометрическая интерпретация пути. Путь, пройденный телом при любом движении, равен площади под графиком скорости на заданном промежутке времени.

Посмотрим, как работает эта геометрическая интерпретация в важном частном случае равноускоренного движения.

[4] *Аппроксимация* — это приближённая замена достаточно сложного объекта более простой моделью, которую удобнее изучать.

Задача. Тело, имеющее скорость v_0 в начальный момент $t = 0$, разгоняется с постоянным ускорением a. Найти путь, пройденный телом к моменту времени t.

Решение. Зависимость скорости от времени в данном случае имеет вид:

$$v = v_0 + at. \tag{1.55}$$

График скорости — прямая, изображённая на рис. 1.22. Искомый путь есть площадь трапеции, расположенной под графиком скорости.

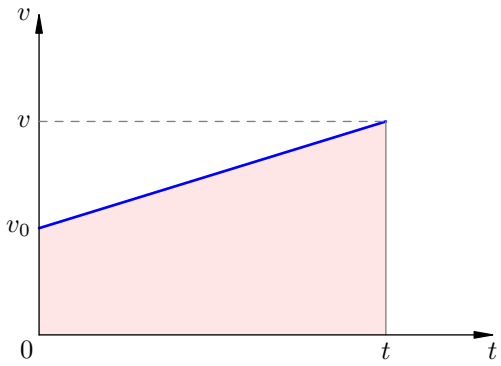


Рис. 1.22. Путь при равноускоренном движении

Меньшее основание трапеции равно v_0. Большее основание равно $v = v_0 + at$. Высота трапеции равна t. Поскольку площадь трапеции есть произведение полусуммы оснований на высоту, имеем:

$$s = \frac{v_0 + v}{2} \cdot t = \frac{v_0 + (v_0 + at)}{2} \cdot t = \frac{2v_0 + at}{2} \cdot t = \frac{2v_0 t + at^2}{2}.$$

Эту формулу можно переписать в более привычном виде:

$$s = v_0 t + \frac{at^2}{2}.$$

Она, разумеется, вам хорошо известна из темы «Равноускоренное движение».

Задача. График скорости тела является полуокружностью диаметра τ (рис. 1.23). Максимальная скорость тела равна v. Найти путь, пройденный телом за время τ.

Решение. Как вы знаете, площадь круга радиуса R равна πR^2. Но в данной задаче необходимо учесть, что радиусы полуокружности имеют разные размерности: горизонтальный радиус есть время $\tau/2$, а вертикальный радиус есть скорость v.

Поэтому пройденный путь, вычисляемый как площадь полукруга, равен половине произведения π на горизонтальный радиус и на вертикальный радиус:

$$s = \frac{1}{2} \cdot \pi \cdot \frac{\tau}{2} \cdot v = \frac{\pi v \tau}{4}.$$

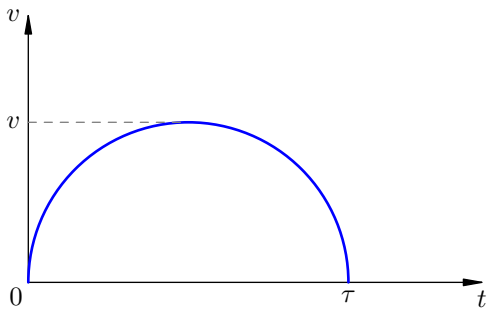


Рис. 1.23. К задаче

1.7 Первый закон Ньютона

Все тела в природе взаимодействуют друг с другом. Однако в некоторых ситуациях воздействия на данное тело со стороны других тел можно не принимать во внимание.

Так, космический корабль в далёком межзвёздном пространстве практически не испытывает гравитационного притяжения объектов Вселенной из-за их колоссальной удалённости[5]. Лежащий на столе карандаш притягивается к Земле, но действие Земли компенсируется упругой реакцией стола, и поэтому карандаш находится в покое, словно никакие силы на него вообще не действуют.

Во всех подобных случаях будем называть тело свободным.

Тело называется свободным, если действия на него со стороны других тел или пренебрежимо малы, или компенсируют друг друга.

1.7.1 Инерциальные системы отсчёта

Повседневный опыт говорит о том, что свободные тела покоятся — как упомянутый карандаш на столе. Поэтому долгое время считалось, что для поддержания какого бы то ни было движения необходимо осуществлять нескомпенсированное внешнее воздействие со стороны других тел.

Но это оказалось неверным. Как установил Галилей, свободное тело может не только находиться в покое, но и двигаться равномерно и прямолинейно! Именно состояние равномерного прямолинейного движения является «естественным» для свободного тела; покой же — частный случай такого движения со скоростью, равной нулю.

Следует учесть, однако, что движение относительно: оно рассматривается не само по себе, а в определённой системе отсчёта. В различных же системах отсчёта движение данного тела будет выглядеть по-разному.

Так, дом с точки зрения неподвижно стоящего наблюдателя будет находиться в покое: сила притяжения дома к Земле компенсируется силой упругости почвы. Если наблюдатель движется относительно земли равномерно и прямолинейно, то и дом относительно наблюдателя будет совершать равномерное прямолинейное движение в полном соответствии с выводами Галилея — ведь дом является свободным телом!

Но если у наблюдателя заплетаются ноги и он бредёт, шатаясь, то ему будет казаться, что дом раскачивается в разные стороны. В этой системе отсчёта дом, будучи свободным телом, совершает отнюдь не равномерное и прямолинейное движение.

Таким образом, утверждение Галилея верно не во всей общности: *не во всякой* системе отсчёта свободное тело движется равномерно и прямолинейно. Но всё же такие системы отсчёта существуют (существуют «хорошие» наблюдатели!), и в этом состоит первый закон Ньютона.

Первый закон Ньютона. Существуют такие системы отсчёта, относительно которых свободное тело движется равномерно и прямолинейно.

Свойство свободного тела сохранять скорость неизменной называется *инерцией*. Поэтому первый закон Ньютона называют ещё *законом инерции*. Равномерное прямолинейное движение свободного тела называется *движением по инерции*.

Система отсчёта, относительно которой свободное тело движется равномерно и прямолинейно, называется *инерциальной*. Первый закон Ньютона — это постулат[6] о *существовании*

[5] Согласно закону всемирного тяготения гравитационные силы обратно пропорциональны квадрату расстояния между телами.

[6] *Постулат* — первичное утверждение физики, не выводимое из каких-либо иных утверждений. Постулат есть констатация факта: *таким вот образом ведёт себя природа*. Сформулировать тот или иной постулат можно лишь на основании наблюдений и физических экспериментов.

инерциальных систем отсчёта. В инерциальных системах отсчёта механические явления описываются наиболее просто.

В действительности инерциальных систем отсчёта существует бесконечно много: всякая система отсчёта, которая движется относительно инерциальной системы равномерно и прямолинейно, сама является инерциальной.

Система отсчёта, которая движется относительно инерциальной системы отсчёта с ускорением, является *неинерциальной*. В такой «плохой» системе отсчёта свободное тело будет двигаться с ускорением, что усложнит описание его движения.

С достаточно высокой точностью можно считать инерциальной *гелиоцентрическую* систему (систему Коперника). Это система отсчёта, начало которой помещено в центре Солнца, а координатные оси направлены на три какие-либо удалённые звезды, которые можно принять за неподвижные.

Инерциальной часто можно считать систему отсчёта, связанную с земной поверхностью. Это, однако, более грубое приближение — ведь мы тем самым отвлекаемся от вращения Земли вокруг собственной оси и вокруг Солнца. Так, звезда, неподвижная в системе Коперника, в земной системе будет совершать сложное движение в виде наложения двух вращений (суточного и годового). Однако в большинстве явлений, происходящих на поверхности Земли, неинерциальность земной системы отсчёта практически никак не сказывается, и ею можно пренебречь.

1.7.2 Принцип относительности

Галилей заметил, что, находясь в трюме корабля, никакими механическими опытами невозможно установить, покоится ли корабль или движется равномерно и прямолинейно. Это означает, что инерциальные системы отсчёта совершенно неотличимы друг от друга с точки зрения законов механики. Иными словами, верен принцип относительности Галилея.

Принцип относительности Галилея. Всякое механическое явление при одних и тех же начальных условиях протекает одинаково в любой инерциальной системе отсчёта.

Впоследствии Эйнштейн распространил этот принцип с механических явлений на вообще все физические явления. Общий принцип относительности Эйнштейна лёг в основу теории относительности.

Принцип относительности Галилея и Эйнштейна мы обсудим подробнее при изучении основ специальной теории относительности.

1.8 Масса и плотность

Масса — одна из самых фундаментальных физических величин. Масса характеризует сразу несколько свойств тела и обладает рядом важных свойств.

1. Масса служит мерой содержащегося в теле вещества.

2. Масса является мерой инертности тела. *Инертностью* называется свойство тела сохранять свою скорость неизменной (в инерциальной системе отсчёта), когда внешние воздействия отсутствуют или компенсируют друг друга. При наличии внешних воздействий инертность тела проявляется в том, что его скорость меняется не мгновенно, а постепенно, и тем медленнее, чем больше инертность (т. е. масса) тела.

3. Массы тел являются причиной их гравитационного притяжения друг к другу (см. раздел «Сила тяготения»).

4. Масса тела равна сумме масс его частей. Это так называемая *аддитивность* массы. Аддитивность позволяет использовать для измерения массы эталон — 1 кг.

5. Масса изолированной системы тел не меняется со временем (*закон сохранения массы*).

6. Масса тела не зависит от скорости его движения. Масса не меняется при переходе от одной системы отсчёта к другой.

Плотностью однородного тела называется отношение массы тела к его объёму:

$$\rho = \frac{m}{V}\,.$$

Плотность не зависит от геометрических свойств тела (формы, объёма) и является характеристикой вещества тела. Плотности веществ представлены в справочных таблицах. Желательно помнить плотность воды: 1000 кг/м^3.

1.9 Второй и третий законы Ньютона

Взаимодействие тел можно описывать с помощью понятия силы. *Сила* — это векторная величина, являющаяся мерой воздействия одного тела на другое.

Будучи вектором, сила характеризуется модулем (абсолютной величиной) и направлением в пространстве. Кроме того, важна точка приложения силы: одна и та же по модулю и направлению сила, приложенная в разных точках тела, может оказывать различное воздействие. Так, если взяться за обод велосипедного колеса и потянуть по касательной к ободу, то колесо начнёт вращаться. Если же тянуть вдоль радиуса, никакого вращения не будет.

1.9.1 Принцип суперпозиции

Опыт показывает, что если на данное тело действуют несколько других тел, то соответствующие силы складываются как векторы. Более точно, справедлив принцип суперпозиции.

Принцип суперпозиции сил. Пусть на тело действуют силы $\vec{F}_1, \vec{F}_2, \ldots, \vec{F}_n$. Если заменить их одной силой $\vec{F} = \vec{F}_1 + \vec{F}_2 + \ldots + \vec{F}_n$, то результат воздействия не изменится.

Сила $\vec{F}$ называется *равнодействующей* сил $\vec{F}_1, \vec{F}_2, \ldots, \vec{F}_n$ или *результирующей* силой.

1.9.2 Второй закон Ньютона

Если равнодействующая сил, приложенных к телу, равна нулю (то есть воздействия других тел компенсируют друг друга), то в силу первого закона Ньютона найдутся такие системы отсчёта (называемые инерциальными), в которых движение тела будет равномерным и прямолинейным. Но если равнодействующая не обращается в нуль, то в инерциальной системе отсчёта у тела появится ускорение.

Количественную связь между ускорением и силой даёт второй закон Ньютона.

Второй закон Ньютона. Произведение массы тела на вектор ускорения есть равнодействующая всех сил, приложенных к телу: $m\vec{a} = \vec{F}$.

Подчеркнём, что второй закон Ньютона связывает *векторы* ускорения и силы. Это означает, что справедливы следующие утверждения.

1. $ma = F$, где a — модуль ускорения, F — модуль равнодействующей силы.
2. Вектор ускорения сонаправлен с вектором равнодействующей силы, так как масса тела положительна.

Например, если тело равномерно движется по окружности, то его ускорение направлено к центру окружности. Стало быть, к центру окружности направлена и равнодействующая всех сил, приложенных к телу.

Второй закон Ньютона справедлив не в любой системе отсчёта. Вспомним шатающегося наблюдателя (раздел «Первый закон Ньютона»): относительно него дом движется с ускорением, хотя равнодействующая всех сил, приложенных к дому, равна нулю. Второй закон Ньютона выполняется лишь в инерциальных системах отсчёта, факт существования которых устанавливается первым законом Ньютона.

1.9.3 Третий закон Ньютона

Опыт показывает, что если тело A действует на тело B, то и тело B действует на тело A. Количественную связь между действиями тел друг на друга даёт третий закон Ньютона («действие равно противодействию»).

Третий закон Ньютона. Два тела действуют друг на друга с силами, равными по модулю и противоположными по направлению. Эти силы имеют одну и ту же физическую природу и направлены вдоль прямой, соединяющей их точки приложения.

Например, если карандаш действует на стол с силой $\vec{P}$, направленной вниз, то стол действует на карандаш с силой $\vec{N}$, направленной вверх (рис. 1.24). Эти силы равны по абсолютной величине.

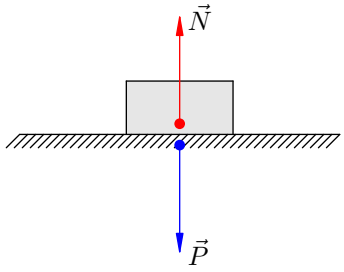


Рис. 1.24. $\vec{P} = -\vec{N}$

Силы $\vec{P}$ и $\vec{N}$, как видим, приложены к разным телам и поэтому не могут уравновешивать друг друга (нет смысла говорить об их равнодействующей).

Третий закон Ньютона, как и второй, справедлив только в инерциальных системах отсчёта.

1.9.4 Как найти закон движения?

Законы Ньютона позволяют решить основную задачу механики — найти закон движения тела. В общих чертах схема действий такова.

1. Записываем второй закон Ньютона: $m\vec{a} = \vec{F}$. С учётом того, что ускорение есть вторая производная радиус-вектора, второй закон Ньютона приобретает вид:

$$m\frac{d^2\vec{r}}{dt^2} = \vec{F}. \tag{1.56}$$

   Необходимо также добавить *начальные условия*: в начальный момент времени $t = 0$ имеем $\vec{r} = \vec{r}_0$ и $\vec{v} = \vec{v}_0$. Начальные значения радиус-вектора и скорости тела предполагаются известными — иначе движение тела нельзя будет описать однозначно.

   Разумеется, должна быть известна и правая часть равенства (1.56) — равнодействующая $\vec{F}$ всех сил, приложенных к телу.

2. Второй закон Ньютона в виде (1.56) является *дифференциальным уравнением*. Это уравнение нужно *проинтегрировать*, то есть найти неизвестную функцию $\vec{r} = \vec{r}(t)$ по известной второй производной этой функции. Выполнив интегрирование, мы и определим закон движения.

Однако легко сказать — «выполнив интегрирование». Сила $\vec{F}$ может зависеть от координат и скорости тела, а также от времени, вследствие чего интегрирование дифференциального уравнения (1.56) окажется весьма сложной задачей. Во многих практических ситуациях такая задача доступна лишь компьютеру.

Вот почему центральное место в школьной механике занимает равноускоренное движение: оно происходит под действием *постоянной* силы, и в этом простейшем случае уравнение (1.56) интегрируется элементарно. Имеем:

$$\frac{d^2\vec{r}}{dt^2} = \frac{\vec{F}}{m} = \vec{a},$$

где $\vec{a}$ — постоянный вектор. Интегрируя один раз, с учётом начальных условий получим:

$$\frac{d\vec{r}}{dt} = \vec{v}_0 + \vec{a}t.$$

Теперь интегрируем второй раз:

$$\vec{r} = \vec{r}_0 + \vec{v}_0 t + \frac{\vec{a}t^2}{2}\,.$$

Получился уже известный вам закон равноускоренного движения.

Механика, основанная на законах Ньютона, называется *классической механикой*. Классическая механика, однако, имеет ограниченную область применимости. В рамках классической механики хорошо описывается движение *не очень маленьких тел с не очень большими скоростями*. При описании атомов и элементарных частиц на замену классической механике приходит *квантовая механика*. Движение объектов со скоростями, близкими к скорости света, происходит по законам *теории относительности*.

1.10 Сила упругости

Как мы знаем, в правой части второго закона Ньютона $m\vec{a} = \vec{F}$ стоит равнодействующая (то есть векторная сумма) всех сил, приложенных к телу. Теперь нам предстоит изучить силы взаимодействия тел в механике. Их три вида: сила упругости, гравитационная сила и сила трения. Начинаем с силы упругости.

1.10.1 Деформация

Силы упругости возникают при деформациях тел. *Деформация* — это изменение формы и размеров тела. К деформациям относятся растяжение, сжатие, кручение, сдвиг и изгиб.

Деформации бывают упругими и пластическими.

Упругая деформация полностью исчезает после снятия внешнего воздействия, которое вызвало деформацию. В результате деформированное поначалу тело восстанавливает свои первоначальные размеры и форму.

Пластическая деформация сохраняется (быть может, частично) после снятия внешней нагрузки, и тело уже не возвращается к прежним размерам и форме.

Частицы тела (молекулы или атомы) взаимодействуют друг с другом силами притяжения и отталкивания, имеющими электромагнитное происхождение (это силы, действующие между ядрами и электронами соседних атомов). Силы взаимодействия зависят от расстояний между частицами. Если деформации нет, то силы притяжения компенсируются силами отталкивания. При деформации изменяются расстояния между частицами, и баланс сил взаимодействия нарушается.

Например, при растяжении стержня расстояния между его частицами увеличиваются, и начинают преобладать силы притяжения. Наоборот, при сжатии стержня расстояния между частицами уменьшаются, и начинают преобладать силы отталкивания. В любом случае возникает сила, которая направлена в сторону, противоположную деформации, и стремится восстановить первоначальную конфигурацию тела.

Сила упругости — это сила, возникающая при упругой деформации тела и направленная в сторону, противоположную смещению частиц тела в процессе деформации. Сила упругости:

1. действует между соседними слоями деформированного тела и приложена к каждому слою;

2. действует со стороны деформированного тела на соприкасающееся с ним тело, вызывающее деформацию, и приложена в месте контакта данных тел перпендикулярно их поверхностям (типичный пример — сила реакции опоры).

Силы, возникающие при пластических деформациях, не относятся к силам упругости. Эти силы зависят не от величины деформации, а от скорости её возникновения. Изучение таких сил выходит далеко за рамки школьной программы.

В школьной физике рассматриваются растяжения нитей и тросов, а также растяжения и сжатия пружин и стержней. Во всех этих случаях силы упругости направлены вдоль осей данных тел.

1.10.2 Закон Гука

Деформация называется *малой*, если изменение размеров тела много меньше его первоначальных размеров. При малых деформациях зависимость силы упругости от величины деформации оказывается линейной.

Закон Гука. Абсолютная величина силы упругости прямо пропорциональна величине деформации. В частности, для пружины, сжатой или растянутой на величину x, сила упругости

даётся формулой:

$$F = kx, \tag{1.57}$$

где k — коэффициент жёсткости пружины.

Коэффициент жёсткости зависит не только от материала пружины, но также от её формы и размеров.

Из формулы (1.57) следует, что график зависимости силы упругости от (малой) деформации является прямой линией (рис. 1.25):

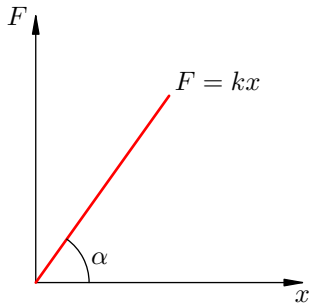


Рис. 1.25. Закон Гука

Коэффициент жёсткости k — это угловой коэффициент в уравнении прямой $F = kx$. Поэтому справедливо равенство:

$$k = \operatorname{tg}\alpha,$$

где α — угол наклона данной прямой к оси абсцисс. Это равенство удобно использовать при экспериментальном нахождении величины k.

Подчеркнём ещё раз, что закон Гука о линейной зависимости силы упругости от величины деформации справедлив лишь при малых деформациях тела. Когда деформации перестают быть малыми, эта зависимость перестаёт быть линейной и приобретает более сложный вид. Соответственно, прямая линия на рис. 1.25 — это лишь небольшой начальный участок криволинейного графика, описывающего зависимость F от x при всех значениях деформации x.

1.10.3 Модуль Юнга

В частном случае малых деформаций *стержней* имеется более детальная формула, уточняющая общий вид (1.57) закона Гука.

Именно, если стержень длиной l и площадью поперечного сечения S растянуть или сжать на величину x, то для силы упругости справедлива формула:

$$F = ES\,\frac{x}{l}\,.$$

Здесь E — *модуль Юнга* материала стержня. Этот коэффициент уже не зависит от геометрических размеров стержня. Модули Юнга различных веществ приведены в справочных таблицах.

1.11 Сила тяготения

Любые два тела притягиваются друг к другу — по той лишь одной причине, что они имеют массу. Эта сила притяжения называется *силой тяготения* или *гравитационной силой*.

1.11.1 Закон всемирного тяготения

Гравитационное взаимодействие любых двух тел во Вселенной подчиняется достаточно простому закону.

Закон всемирного тяготения. Две материальные точки массами m_1 и m_2 притягиваются друг к другу с силой, прямо пропорциональной их массам и обратно пропорциональной квадрату расстояния r между ними:

$$F = G\frac{m_1 m_2}{r^2}\,. \tag{1.58}$$

Коэффициент пропорциональности G называется *гравитационной постоянной*. Это фундаментальная константа, и её численное значение было определено на основе эксперимента Генри Кавендиша:

$$G = 6{,}67 \cdot 10^{-11}\,\frac{\text{Н} \cdot \text{м}^2}{\text{кг}^2}\,.$$

Порядок величины гравитационной постоянной объясняет, почему мы не замечаем взаимного притяжения окружающих нас предметов: гравитационные силы оказываются слишком малыми при небольших массах тел. Мы наблюдаем лишь притяжение предметов к Земле, масса которой грандиозна и равна примерно $6 \cdot 10^{24}$ кг.

Формула (1.58), будучи справедливой для материальных точек, перестаёт быть верной, если размерами тел пренебречь нельзя. Имеются, однако, два важных для практики исключения.

1. Формула (1.58) справедлива, если тела являются однородными шарами. Тогда r — расстояние между их центрами. Сила притяжения направлена вдоль прямой, соединяющей центры шаров.

2. Формула (1.58) справедлива, если одно из тел — однородный шар, а другое — материальная точка, находящаяся вне шара. Тогда r — расстояние от точки до центра шара. Сила притяжения направлена вдоль прямой, соединяющей точку с центром шара.

Второй случай особенно важен, так как позволяет применять формулу (1.58) для силы притяжения тела (например, искусственного спутника) к планете.

1.11.2 Сила тяжести

Предположим, что тело находится вблизи некоторой планеты. *Сила тяжести* — это сила гравитационного притяжения, действующая на тело со стороны планеты. В подавляющем большинстве случаев сила тяжести — это сила притяжения к Земле.

Пусть тело массы m лежит на поверхности Земли. На тело действует сила тяжести mg, где g — ускорение свободного падения вблизи поверхности Земли. С другой стороны, считая Землю однородным шаром, можно выразить силу тяжести по закону всемирного тяготения:

$$mg = G\frac{Mm}{R^2}\,,$$

где M — масса Земли, $R \approx 6400$ км — радиус Земли. Отсюда получаем формулу для ускорения свободного падения на поверхности Земли:

$$g = G\frac{M}{R^2}\,. \tag{1.59}$$

Эта же формула, разумеется, позволяет найти ускорение свободного падения на поверхности любой планеты массы M и радиуса R.

Если тело находится на высоте h над поверхностью планеты, то для силы тяжести получаем:

$$mg(h) = G\frac{Mm}{(R+h)^2}\,.$$

Здесь $g(h)$ — ускорение свободного падения на высоте h:

$$g(h) = G\frac{M}{(R+h)^2} = \frac{gR^2}{(R+h)^2}\,.$$

В последнем равенстве мы воспользовались соотношением

$$GM = gR^2,$$

которое следует из формулы (1.59).

1.11.3 Вес тела. Невесомость

Рассмотрим тело, находящееся в поле силы тяжести. Предположим, что есть опора или подвес, препятствующие свободному падению тела. *Вес* тела — это сила, с которой тело действует на опору или подвес. Подчеркнём, что вес приложен не к телу, а к опоре (подвесу).

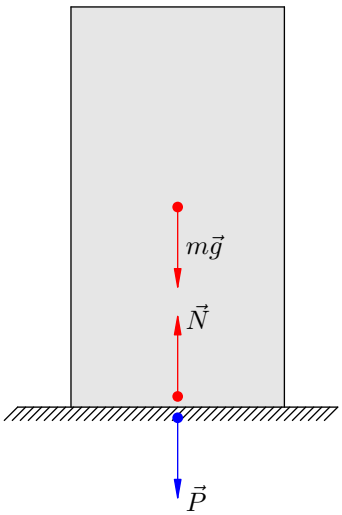


Рис. 1.26. Сила тяжести, реакция опоры и вес тела

На рис. 1.26 изображено тело на опоре. Со стороны Земли на тело действует сила тяжести $m\vec{g}$ (в случае однородного тела простой формы сила тяжести приложена в центре симметрии тела). Со стороны опоры на тело действует сила упругости $\vec{N}$ (так называемая реакция опоры). На опору со стороны тела действует сила $\vec{P}$ — вес тела. По третьему закону Ньютона силы $\vec{P}$ и $\vec{N}$ равны по модулю ($P = N$) и противоположны по направлению.

Предположим, что тело покоится. Тогда равнодействующая сил, приложенных к телу, равна нулю. Имеем:

$$m\vec{g} + \vec{N} = \vec{0} \Rightarrow m\vec{g} = -\vec{N} \Rightarrow mg = N.$$

С учётом равенства $N = P$ получаем $mg = P$. Стало быть, если тело покоится, то его вес равен по модулю силе тяжести.

Рассмотрим две стандартные задачи, которые обязательно нужно уметь решать.

Задача. Тело массы m вместе с опорой движется с ускорением a, направленным вертикально вверх. Найти вес тела.

Решение. Направим ось Y вертикально вверх (рис. 1.27).

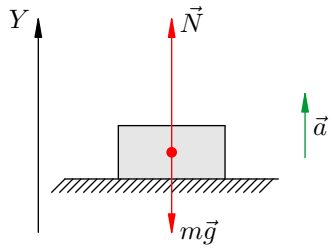


Рис. 1.27. Вес тела больше силы тяжести

Запишем второй закон Ньютона:

$$m\vec{a} = m\vec{g} + \vec{N}.$$

Перейдём к проекциям на ось Y:

$$ma = N - mg.$$

Отсюда $N = mg + ma = m(g + a)$. Следовательно, вес тела

$$P = m(g + a).$$

Как видим, вес тела больше силы тяжести. Такое состояние называется *перегрузкой*.

Задача. Тело массы m вместе с опорой движется с ускорением $a \leqslant g$, направленным вертикально вниз. Найти вес тела.

Решение. Направим ось Y вертикально вниз (рис. 1.28).

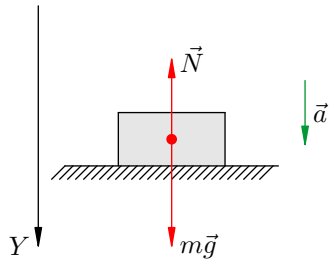


Рис. 1.28. Вес тела меньше силы тяжести

Схема решения та же. Начинаем со второго закона Ньютона:

$$m\vec{a} = m\vec{g} + \vec{N}.$$

Переходим к проекциям на ось Y:

$$ma = mg - N.$$

Отсюда $N = mg - ma = m(g - a)$. Следовательно, вес тела

$$P = m(g - a).$$

В данном случае вес тела меньше силы тяжести. При $a = g$ (свободное падение тела с опорой) вес тела обращается в нуль. Это — состояние *невесомости*, при котором тело вообще не давит на опору.

1.11.4 Искусственные спутники

Для того, чтобы искусственный спутник мог совершать орбитальное движение вокруг планеты, ему нужно сообщить определённую скорость. Найдём скорость кругового движения спутника на высоте h над поверхностью планеты. Масса планеты M, её радиус R (рис. 1.29).

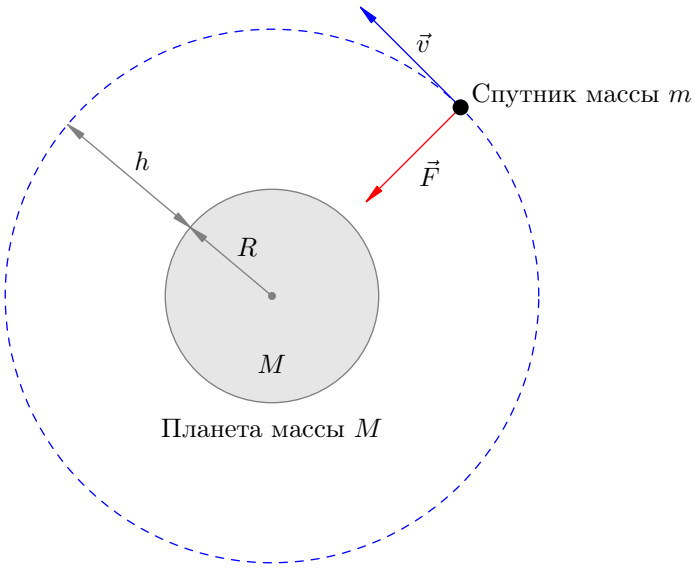


Рис. 1.29. Спутник на круговой орбите

Спутник будет двигаться под действием единственной силы $\vec{F}$ — силы всемирного тяготения, направленной к центру планеты. Туда же направлено и ускорение спутника — центростремительное ускорение

$$a = \frac{v^2}{R+h}\,.$$

Обозначив через m массу спутника, запишем второй закон Ньютона в проекции на ось, направленной к центру планеты: $ma = F$, или

$$m\frac{v^2}{R+h} = G\frac{Mm}{(R+h)^2}\,.$$

Отсюда получаем выражение для скорости:

$$v = \sqrt{\frac{GM}{R+h}}\,.$$

Первая космическая скорость — это максимальная скорость кругового движения спутника, отвечающая высоте $h = 0$. Для первой космической скорости имеем:

$$v_1 = \sqrt{\frac{GM}{R}}\,,$$

или, с учётом формулы (1.59),

$$v_1 = \sqrt{gR}\,.$$

Для Земли приближённо получаем:

$$v_1 = \sqrt{10 \cdot 6400000} = 8000 \text{ м/с} = 8 \text{ км/с}.$$

1.12 Сила трения

Сила трения — это сила взаимодействия между соприкасающимися телами, препятствующая перемещению одного тела относительно другого. Сила трения всегда направлена вдоль поверхностей соприкасающихся тел.

В школьной физике рассматриваются два вида трения.

1. *Сухое трение.* Оно возникает в зоне контакта поверхностей твёрдых тел при отсутствии между ними жидкой или газообразной прослойки.

2. *Вязкое трение.* Оно возникает при движении твёрдого тела в жидкой или газообразной среде или при перемещении одного слоя среды относительно другого.

Сухое и вязкое трение имеют разную природу и отличаются по свойствам. Рассмотрим эти виды трения по отдельности.

1.12.1 Сухое трение

Сухое трение может возникать даже при отсутствии относительного перемещения тел. Так, тяжёлый диван остаётся неподвижным при слабой попытке сдвинуть его с места: наша сила, приложенная к дивану, компенсируется силой трения, возникающей между диваном и полом. Сила трения, которая действует между поверхностями покоящихся тел и препятствует возникновению движения, называется *силой трения покоя*.

Почему вообще появляется сила трения покоя? Соприкасающиеся поверхности дивана и пола являются шероховатыми, они усеяны микроскопическими, незаметными глазу бугорками разных форм и размеров. Эти бугорки зацепляются друг за друга и не дают дивану начать движение. Сила трения покоя, таким образом, вызвана силами электромагнитного отталкивания молекул, возникающими при деформациях бугорков.

Будем плавно увеличивать силу F, приложенную к дивану. Как вам хорошо известно, до некоторого момента диван всё ещё не поддаётся и стоит на месте. Это означает, что *сила трения покоя f возрастает вместе с увеличением внешнего воздействия, оставаясь равной по модулю приложенной силе*: $f = F$ (рис. 1.30, участок OA). Причина возрастания силы трения понятна: увеличиваются деформации бугорков и возрастают силы отталкивания их молекул.

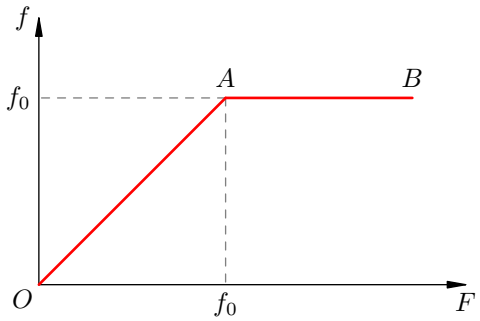


Рис. 1.30. Зависимость силы трения f от внешней силы F

Наконец, при определённой величине внешней силы диван сдвигается с места. Это означает, что *сила трения покоя достигает максимально возможного значения* f_0 (рис. 1.30, точка A). Деформации бугорков оказываются столь велики, что бугорки не выдерживают и начинают разрушаться. Возникает скольжение.

Сила трения, которая действует между проскальзывающими поверхностями, называется *силой трения скольжения*. В процессе скольжения рвутся связи между молекулами в зацепляющихся бугорках поверхностей. При трении покоя таких разрывов нет.

Сила трения скольжения уже не зависит от величины приложенной силы F и остаётся постоянной (рис. 1.30, горизонтальный участок AB). Сила трения скольжения равна максимальной силе трения покоя f_0.

Объяснение сухого трения в терминах бугорков является максимально простым и наглядным. Реальные механизмы трения куда сложнее, и их рассмотрение выходит за рамки элементарной физики.

Сила трения скольжения, приложенная к телу со стороны шероховатой поверхности, направлена противоположно скорости движения тела относительно этой поверхности. При изменении направления скорости меняется и направление силы трения. Зависимость силы трения от скорости — главное отличие силы трения от сил упругости и тяготения (величина которых зависит только от взаимного расположения тел, т. е. от их координат).

В простейшей модели сухого трения выполняются следующие законы. Они являются обобщением опытных фактов и носят приближённый характер.

1. Максимальная величина силы трения покоя равна силе трения скольжения.
2. Абсолютная величина силы трения скольжения прямо пропорциональна силе реакции опоры:
$$f = \mu N.$$
Коэффициент пропорциональности μ называется *коэффициентом трения*.
3. Коэффициент трения не зависит от скорости движения тела по шероховатой поверхности.
4. Коэффициент трения не зависит от площади соприкасающихся поверхностей.

Этих законов достаточно для решения задач.

Задача. На горизонтальной шероховатой поверхности лежит брусок массой $m = 3$ кг. Коэффициент трения $\mu = 0{,}4$. К бруску приложена горизонтальная сила F. Найти силу трения в двух случаях: 1) при $F = 10$ Н; 2) при $F = 15$ Н.

Решение. Сделаем рисунок, расставим силы. Силу трения обозначаем $\vec{f}$ (рис. 1.31).

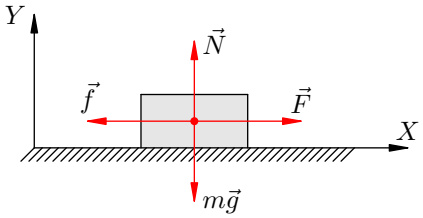


Рис. 1.31. К задаче

Запишем второй закон Ньютона:

$$m\vec{a} = m\vec{g} + \vec{N} + \vec{F} + \vec{f}. \tag{1.60}$$

Вдоль оси Y брусок не совершает движения, $a_y = 0$. Проектируя равенство (1.60) на ось Y, получим: $0 = -mg + N$, откуда $N = mg$.

Максимальная величина f_0 силы трения покоя (она же сила трения скольжения) равна:

$$f_0 = \mu N = \mu mg = 0{,}4 \cdot 3 \cdot 10 = 12 \text{ Н}.$$

1) Сила $F = 10$ Н меньше максимальной силы трения покоя. Брусок остаётся на месте, и сила трения будет силой трения покоя: $f = F = 10$ Н.

2) Сила $F = 15$ Н больше максимальной силы трения покоя. Брусок начнёт скользить, и сила трения будет силой трения скольжения: $f = f_0 = 12$ Н.

1.12.2 Вязкое трение

Сила сопротивления, возникающая при движении тела в вязкой среде (жидкости или газе), обладает совершенно иными свойствами.

Во-первых, отсутствует сила трения покоя. Например, человек может сдвинуть с места плавающий многотонный корабль, просто потянув за канат.

Во-вторых, сила сопротивления зависит от формы движущегося тела. Корпус подводной лодки, самолёта или ракеты имеет обтекаемую сигарообразную форму — для уменьшения силы сопротивления. Наоборот, при движении полусферического тела вогнутой стороной вперёд сила сопротивления очень велика (пример — парашют).

В третьих, абсолютная величина силы сопротивления существенно зависит от скорости. При малых скоростях движения сила сопротивления прямо пропорциональна скорости:

$$f = \alpha v.$$

При больших скоростях сила сопротивления прямо пропорциональна квадрату скорости:

$$f = \beta v^2.$$

Например, при падении в воздухе зависимость силы сопротивления от квадрата скорости имеет место уже при скоростях около нескольких метров в секунду. Коэффициенты α и β зависят от формы и размеров тела, от физических свойств поверхности тела и вязкой среды.

Так, парашютист при затяжном прыжке не набирает скорость безгранично, а с определённого момента начинает падать с установившейся скоростью, при которой сила сопротивления становится равна силе тяжести:

$$\beta v^2 = mg.$$

Отсюда установившаяся скорость:

$$v = \sqrt{\frac{mg}{\beta}}\,. \tag{1.61}$$

Задача. Два металлических шарика, одинаковых по размеру и различных по массе, падают без начальной скорости с одной и той же большой высоты. Какой из шариков быстрее упадёт на землю — лёгкий или тяжёлый?

Решение. Из формулы (1.61) следует, что у тяжёлого шарика установившаяся скорость падения больше. Значит, он дольше будет набирать скорость и потому быстрее достигнет земли.

1.13 Статика твёрдого тела

Статика изучает равновесие тел под действием приложенных к ним сил. *Равновесие* — это состояние тела, при котором каждая его точка остаётся всё время неподвижной в некоторой инерциальной системе отсчёта.

Условием равновесия материальной точки является равенство нулю равнодействующей (т. е. векторной суммы) всех сил, приложенных к точке. В этом случае наша точка будет двигаться равномерно и прямолинейно в произвольной инерциальной системе отсчёта. Значит, система отсчёта, связанная с точкой, также будет инерциальной, и в ней точка будет покоиться.

В случае твёрдого тела ситуация сложнее. Прежде всего, важно учитывать точку приложения каждой силы.

- Сила тяжести приложена в центре тяжести тела. Для тела простой формы центр тяжести совпадает с центром симметрии.
- Силы упругости и трения приложены в точке или в плоскости контакта тела с соприкасающимся телом.

Прямая линия, проходящая через точку приложения вдоль вектора силы, называется *линией действия* силы. Оказывается, точку приложения силы можно переносить вдоль линии её действия — от этого механическое состояние тела не изменится (в частности, равновесие не нарушится).

Для равновесия твёрдого тела недостаточно потребовать равенства нулю векторной суммы всех приложенных к телу сил.

В качестве примера рассмотрим *пару сил* — так называются две равные по модулю противоположно направленные силы, линии действия которых не совпадают. Пусть пара сил $\vec{F}_1$ и $\vec{F}_2$ приложена к твёрдому стержню (рис. 1.32).

Рис. 1.32. Пара сил

Векторная сумма этих сил равна нулю. Но стержень покоиться не будет: он начнёт вращаться. В данном случае не выполнено второе условие равновесия твёрдого тела. Чтобы его сформулировать, нужно ввести понятие момента силы.

Как должна быть направлена линия действия силы, чтобы тело стало вращаться вокруг неподвижной оси? Для начала заметим следующее.

- Если линия действия силы параллельна данной оси, то вращения не будет.
- Если линия действия силы пересекает данную ось, то вращения не будет.

В каждом из этих случаев действие силы вызывает лишь деформацию твёрдого тела.

Чтобы началось вращение, линия действия силы и ось вращения должны быть скрещивающимися прямыми.

Без ограничения общности можно считать эти прямые перпендикулярными друг другу. Мы всегда можем этого добиться, разложив силу на две составляющие — параллельную и перпендикулярную оси вращения — и отбросив параллельную составляющую как не вызывающую вращения. Поэтому везде далее мы считаем, что все силы, действующие на тело, перпендикулярны оси вращения.

1.13.1 Момент силы

Плечо силы — это расстояние от оси вращения до линия действия силы (т. е. длина общего перпендикуляра к двум этим прямым).

В качестве примера на рис. 1.33 изображён диск, к которому приложена сила $\vec{F}$. Ось вращения перпендикулярна плоскости чертежа и проходит через точку O. Плечом силы является величина $l = OH$, где H — основание перпендикуляра, опущенного из точки O на линию действия силы.

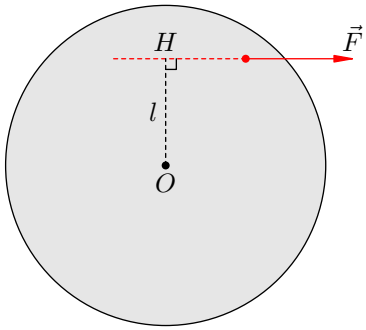


Рис. 1.33. Плечо силы

Момент силы относительно оси вращения — это произведение силы на плечо:

$$M = Fl.$$

Чтобы учесть также направление вращения, вызываемого действием силы, моменту силы приписывают знак. Именно, момент силы считается положительным, если сила стремится поворачивать тело против часовой стрелки, и отрицательным, если по часовой стрелке.

1.13.2 Условия равновесия

Если тело имеет неподвижную ось вращения и если алгебраическая сумма моментов всех сил относительно этой оси обращается в нуль, то тело будет находиться в равновесии. Это так называемое *правило моментов*. Оказывается, что в этом случае обращается в нуль алгебраическая сумма моментов всех сил относительно любой другой оси, параллельной оси вращения.

В общем случае, когда твёрдое тело может совершать как поступательное, так и вращательное движение, мы имеем два условия равновесия.

1. Равна нулю векторная сумма всех сил, приложенных к телу.

2. Равна нулю алгебраическая сумма моментов всех сил, приложенных к телу, относительно данной оси вращения или любой другой оси, параллельной данной.

Так, в примере на рис. 1.32 алгебраическая сумма моментов пары сил не обращается нуль (оба момента положительны). Поэтому стержень не находится в равновесии.

При решении задач удобно использовать сформулированные выше условия равновесия в следующем виде.

1′. Силы уравновешены вдоль любой оси.

2′. Суммарный момент сил, вращающих тело в одну сторону, равен суммарному моменту сил, вращающих тело в другую сторону.

Сейчас мы разберём одну достаточно содержательную задачу по статике и посмотрим, как работают наши условия равновесия.

Задача. Однородная лестница опирается на гладкую вертикальную стену, образуя с ней угол α. При каком максимальном значении α лестница будет покоиться? Коэффициент трения между лестницей и полом равен μ.

Решение. Пусть лестница опирается о пол и стену в точках A и B соответственно (рис. 1.34). Расставим силы, действующие на лестницу.

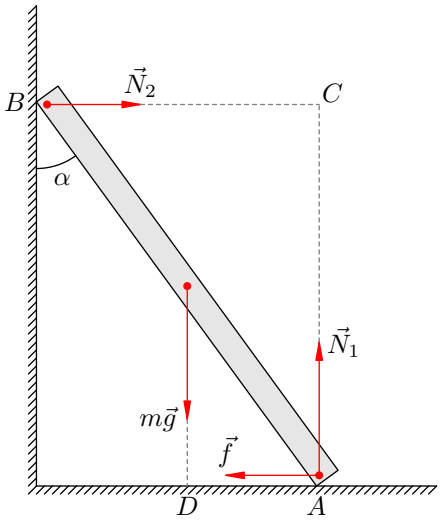


Рис. 1.34. К задаче

Поскольку лестница однородная, сила тяжести $m\vec{g}$ приложена в середине лестницы.

Сила упругости пола $\vec{N}_1$ и сила трения $\vec{f}$ приложены в точке A. На рис. 1.34 точка приложения этих сил немного смещена от точки A внутрь лестницы; тем самым мы однозначно указываем, что силы приложены именно к лестнице (а не к полу).

Точно так же сила упругости стены $\vec{N}_2$ приложена в точке B. Поскольку стена гладкая, сила трения между стеной и лестницей отсутствует.

Воспользуемся условием $1'$. Вдоль горизонтальной оси силы уравновешены:

$$f = N_2. \tag{1.62}$$

Вдоль вертикальной оси силы также уравновешены:

$$mg = N_1. \tag{1.63}$$

Теперь переходим к правилу моментов — условию $2'$. Какую ось вращения выбрать? Удобнее всего взять ось, проходящую через точку A (перпендикулярно плоскости рисунка). В таком случае моменты сразу двух сил $\vec{f}$ и $\vec{N}_1$ обратятся в нуль — ведь плечи этих сил относительно точки A равны нулю (поскольку линии действия сил проходят через эту точку). Ненулевые моменты относительно точки A имеют силы $m\vec{g}$ и $\vec{N}_2$, которые стремятся вращать лестницу в разные стороны; стало быть, моменты данных сил должны быть равны друг другу.

Плечо силы $\vec{N}_2$ — это длина перпендикуляра AC, опущенного из точки A на линию BC действия силы $\vec{N}_2$. Плечо силы $m\vec{g}$ — это длина перпендикуляра AD, опущенного из точки A на линию действия силы $m\vec{g}$. Согласно правилу моментов имеем:

$$N_2 \cdot AC = mg \cdot AD.$$

Пусть длина лестницы равна $2l$. Тогда $AC = 2l\cos\alpha$, $AD = l\sin\alpha$. Подставляем эти соотношения в равенство моментов:

$$N_2 \cdot 2l\cos\alpha = mg \cdot l\sin\alpha,$$

откуда

$$2N_2 = mg\,\mathrm{tg}\,\alpha. \tag{1.64}$$

С учётом равенства (1.62) имеем вместо (1.64):

$$2f = mg\,\mathrm{tg}\,\alpha. \tag{1.65}$$

Вспомним теперь, что в условии спрашивается *максимальное* значение α. При максимальном угле α лестница пока ещё стоит, но уже находится *на грани проскальзывания*. Это означает, что сила трения f достигла своего максимального значения, равного силе трения скольжения:

$$f = \mu N_1.$$

Теперь из (1.65) имеем:

$$2\mu N_1 = mg\,\mathrm{tg}\,\alpha,$$

а с учётом равенства (1.63):

$$2\mu mg = mg\,\mathrm{tg}\,\alpha.$$

Отсюда получаем искомую максимальную величину α:

$$\alpha = \mathrm{arctg}(2\mu).$$

1.14 Статика жидкостей и газов

В гидро- и аэростатике рассматриваются два вопроса: 1) равновесие жидкостей и газов под действием приложенных к ним сил; 2) равновесие твёрдых тел в жидкостях и газах.

Многие из обсуждаемых далее фактов относятся равным образом как к жидкостям, так и к газам. В таких случаях мы будем называть жидкость и газ *средой*.

При сжатии среды в ней возникают силы упругости, называемые *силами давления*. Силы давления действуют между соприкасающимися слоями среды, на погружённые в среду твёрдые тела, а также на дно и стенки сосуда.

Сила давления среды обладает двумя характерными свойствами.

1. Сила давления действует перпендикулярно поверхности выделенного элемента среды или твёрдого тела. Это объясняется текучестью среды: силы упругости не возникают в ней при относительном сдвиге слоёв, поэтому отсутствуют силы упругости, касательные к поверхности.

2. Сила давления равномерно распределена по той поверхности, на которую она действует.

Естественной величиной, возникающей в процессе изучения сил давления среды, является давление.

Пусть на поверхность площади S действует сила F, которая перпендикулярна поверхности и равномерно распределена по ней. *Давлением* называется величина

$$p = \frac{F}{S}\,.$$

Единицей измерения давления служит *паскаль* (Па). 1 Па — это давление, производимое силой 1 Н на поверхность площадью 1 м^2.

Полезно помнить приближённое значение нормального атмосферного давления: $p_0 = 10^5$ Па.

1.14.1 Гидростатическое давление

Гидростатическим называется давление неподвижной жидкости, вызванное силой тяжести. Найдём формулу для гидростатического давления столба жидкости.

Предположим, что в сосуд с площадью дна S налита жидкость до высоты h (рис. 1.35). Плотность жидкости равна ρ.

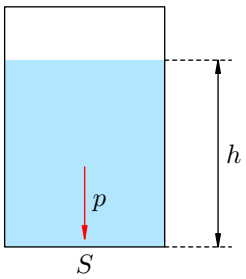


Рис. 1.35. Гидростатическое давление

Объём жидкости равен Sh, поэтому масса жидкости $m = \rho Sh$. Сила F давления жидкости на дно сосуда — это вес жидкости. Так как жидкость неподвижна, её вес равен силе тяжести:

$$F = mg = \rho Shg.$$

Разделив силу F на площадь S, получим давление жидкости:

$$p = \rho g h.$$

Это и есть формула гидростатического давления.

Так, на глубине 10 м вода оказывает давление $p = 1000 \cdot 10 \cdot 9{,}8 = 98000$ Па, примерно равное атмосферному. Можно сказать, что *атмосферное давление приблизительно равно 10 м водного столба.*

Для практики столь большая высота столба жидкости неудобна, и реальные жидкостные манометры — ртутные. Посмотрим, какую высоту должен иметь столб ртути ($\rho = 13600$ кг/м^3), чтобы создать аналогичное давление:

$$h = \frac{p}{\rho g} = \frac{10^5}{13600 \cdot 9{,}8} = 0{,}75 \text{ м} = 750 \text{ мм}.$$

Вот почему для измерения атмосферного давления широко используется *миллиметр ртутного столба* (мм рт. ст.).

1.14.2 Закон Паскаля

Если поставить гвоздь вертикально и ударить по нему молотком, то гвоздь передаст действие молотка по вертикали, но не вбок. Твёрдые тела из-за наличия кристаллической решётки передают производимое на них давление только в направлении действия силы.

Жидкости и газы (напомним, что мы называем их средами) ведут себя иначе. В средах справедлив закон Паскаля.

Закон Паскаля. Давление, оказываемое на жидкость или газ, передаётся в любую точку этой среды без изменения по всем направлениям.

(В частности, на площадку, помещённую внутри жидкости на фиксированной глубине, действует одна и та же сила давления, как эту площадку ни поворачивай.)

Например, ныряльщик на глубине h испытывает давление $p = p_0 + \rho g h$. Почему? Согласно закону Паскаля вода передаёт давление атмосферы p_0 без изменения на глубину h, где оно прибавляется к гидростатическому давлению водяного столба $\rho g h$.

Отличной иллюстрацией закона Паскаля служит опыт с шаром Паскаля. Это шар с множеством отверстий, соединённый с цилиндрическим сосудом (рис. 1.36)[7].

Рис. 1.36. Шар Паскаля

Если налить в сосуд воду и двинуть поршень, то вода брызнет из всех отверстий. Это как раз и означает, что вода передаёт внешнее давление по всем направлениям.

То же самое наблюдается и для газа: если сосуд наполнить дымом, то при движении поршня струйки дыма пойдут опять-таки из всех отверстий сразу. Стало быть, газ также передаёт давление по всем направлениям.

Вы ежедневно пользуетесь законом Паскаля, когда выдавливаете зубную пасту из тюбика. А именно, вы сжимаете тюбик в поперечном направлении, а паста двигается перпендикулярно вашему усилию — в продольном направлении. Почему? Ваше давление передаётся внутри тюбика по всем направлениям, в частности — в сторону отверстия тюбика. Туда-то паста и выходит.

[7] Изображение с сайта http://festival.1september.ru.

1.14.3 Гидравлический пресс

Гидравлический пресс — это устройство, дающее выигрыш в силе. Что значит «выигрыш в силе»? Имеется в виду, что, прикладывая сравнительно небольшую силу в одном месте данного устройства, оказывается возможным получить значительно большее усилие в другом его месте.

Гидравлический пресс изображён на рис. 1.37. Он состоит из двух сообщающихся сосудов, имеющих разную площадь поперечного сечения и закрытых поршнями. В сосудах между поршнями находится жидкость.

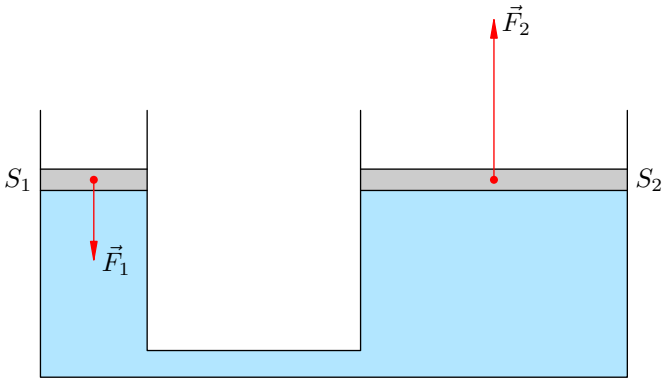


Рис. 1.37. Гидравлический пресс

Принцип действия гидравлического пресса очень прост и основан на законе Паскаля.

Пусть S_1 — площадь малого поршня, S_2 — площадь большого поршня. Надавим на малый поршень с силой F_1. Тогда под малым поршнем в жидкости возникнет давление:

$$p = \frac{F_1}{S_1}\,.$$

Согласно закону Паскаля это давление будет передано без изменения по всем направлениям в любую точку жидкости, в частности — непосредственно под большой поршень. Следовательно, на большой поршень со стороны жидкости будет действовать сила:

$$F_2 = pS_2 = F_1\frac{S_2}{S_1}\,.$$

Полученное соотношение можно переписать и так:

$$\frac{F_2}{F_1} = \frac{S_2}{S_1}\,.$$

Мы видим, что F_2 больше F_1 во столько раз, во сколько S_2 больше S_1. Например, если площадь большого поршня в 100 раз превышает площадь малого поршня, то усилие на большом поршне окажется в 100 раз больше усилия на малом поршне. Вот каким образом гидравлический пресс даёт выигрыш в силе.

1.14.4 Закон Архимеда

Почему плавают корабли? Почему поднимается вверх воздушный шар? Сейчас мы начнём разбираться с этими вопросами. И снова на помощь придёт закон Паскаля.

Мы знаем, что дерево в воде не тонет. Следовательно, сила тяжести уравновешивается какой-то другой силой, действующей на кусок дерева со стороны воды вертикально вверх. Эта сила называется *выталкивающей* или *архимедовой* силой. Она действует на всякое тело, погружённое в жидкость или газ.

Выясним причину возникновения архимедовой силы. Рассмотрим цилиндр площадью поперечного сечения S и высотой h, погружённый в жидкость плотности ρ. Основания цилиндра горизонтальны. Верхнее основание находится на глубине h_1, нижнее — на глубине $h_2 = h_1 + h$ (рис. 1.38).

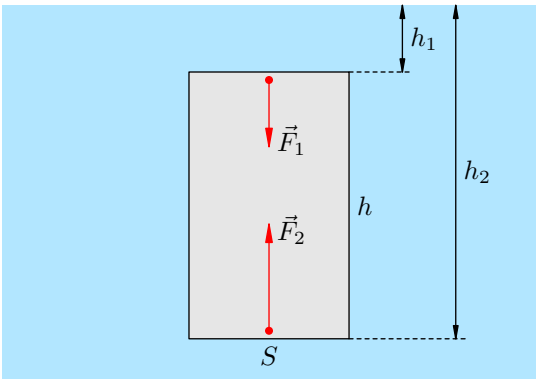


Рис. 1.38. $F_\text{A} = F_2 - F_1$

На боковую поверхность цилиндра действуют силы давления, которые приводят лишь к сжатию цилиндра. Эти силы можно не принимать во внимание.

На уровне верхнего основания цилиндра давление жидкости равно $p_1 = \rho g h_1$. На верхнее основание действует сила давления $F_1 = p_1 S = \rho g h_1 S$, направленная вертикально вниз.

На уровне нижнего основания цилиндра давление жидкости равно $p_2 = \rho g h_2$. На нижнее основание действует сила давления $F_2 = p_2 S = \rho g h_2 S$, направленная вертикально вверх (закон Паскаля!).

Так как $h_2 > h_1$, то $F_2 > F_1$, и поэтому возникает равнодействующая сил давления, направленная вверх. Это и есть архимедова сила F_A. Имеем:

$$F_\text{A} = F_2 - F_1 = \rho g h_2 S - \rho g h_1 S = \rho g S(h_2 - h_1) = \rho g S h.$$

Но произведение Sh равно объёму цилиндра V. Получаем окончательно:

$$F_\text{A} = \rho g V. \tag{1.66}$$

Это и есть формула для архимедовой силы. Возникает архимедова сила вследствие того, что давление жидкости на нижнее основание цилиндра больше, чем на верхнее.

Формулу (1.66) можно интерпретировать следующим образом. Произведение ρV — это масса жидкости m, объём которой равен V: $\rho V = m$. Но тогда $\rho g V = mg = P$, где P — вес жидкости, взятой в объёме V. Поэтому наряду с (1.66) имеем:

$$F_\text{A} = P. \tag{1.67}$$

Иными словами, архимедова сила, действующая на цилиндр, равна весу жидкости, объём которой совпадает с объёмом цилиндра.

Формулы (1.66) и (1.67) справедливы и в общем случае, когда погружённое в жидкость или газ тело объёма V имеет *любую* форму, а не только форму цилиндра (конечно, в случае газа ρ является плотностью этого газа). Поясним, почему так получается.

Выделим мысленно в среде некоторый объём V произвольной формы. Этот объём находится в равновесии: не тонет и не всплывает. Следовательно, сила тяжести, действующая на среду, находящуюся внутри выделенного нами объёма, уравновешена силами давления на поверхность нашего объёма со стороны остальной среды — ведь на нижние элементы поверхности приходится большее давление, чем на верхние.

Иными словами, равнодействующая сил гидростатического давления на поверхность выделенного объёма — архимедова сила — направлена вертикально вверх и равна весу среды в этом объёме.

Сила тяжести, действующая на наш объём, приложена к его центру тяжести. Значит, и архимедова сила должна быть приложена к центру тяжести выделенного объёма. В противном случае сила тяжести и архимедова сила образуют пару сил, которая вызовет вращение нашего объёма (а он находится в равновесии).

А теперь заменим выделенный объём среды твёрдым телом того же объёма V и той же самой формы. Ясно, что силы давления среды на поверхность тела не изменятся, так как неизменной осталась *конфигурация* среды, окружающей тело. Поэтому архимедова сила по-прежнему будет направлена вертикально вверх и равна весу среды, взятой в объёме V. Точкой приложения архимедовой силы будет центр тяжести тела.

Закон Архимеда. На погружённое в жидкость или газ тело действует выталкивающая сила, направленная вертикально вверх и равная весу среды, объём которой равен объёму тела.

Таким образом, архимедова сила всегда находится по формуле (1.66). Заметим, что в эту формулу не входят ни плотность тела, ни какие-либо его геометрические характеристики — при фиксированном объёме величина архимедовой силы не зависит от вещества и формы тела.

До сих пор мы рассматривали случай полного погружения тела. Чему равна архимедова сила при частичном погружении? На ту часть тела, которая находится над поверхностью жидкости, никакая выталкивающая сила не действует. Если эту часть мысленно срезать, то величина архимедовой силы не изменится. Но тогда мы получим целиком погружённое тело, объём которого равен объёму погружённой части исходного тела.

Значит, *на частично погружённое в жидкость тело действует выталкивающая сила, равная весу жидкости, объём которой равен объёму погружённой части тела.* Формула (1.66) справедлива и в этом случае, только объём всего тела V нужно заменить на объём погружённой части $V_{\text{погр}}$:

$$F_{\text{A}} = \rho g V_{\text{погр}}.$$

Архимед обнаружил, что целиком погружённое в воду тело вытесняет объём воды, равный собственному объёму. Тот же факт имеет место для других жидкостей и газов. Поэтому можно сказать, что *на всякое тело, погружённое в жидкость или газ, действует выталкивающая сила, равная весу вытесненной телом среды.*

1.14.5 Плавание тел

Рассмотрим тело плотности ρ и жидкость плотности ρ_0. Допустим, что тело полностью погрузили в жидкость и отпустили.

Сразу после отпускания на тело действуют лишь сила тяжести mg и архимедова сила F_{A}. Если объём тела равен V, то

$$mg = \rho g V, \quad F_{\text{A}} = \rho_0 g V.$$

Имеются три возможности дальнейшего движения тела.

1. Сила тяжести больше архимедовой силы: $mg > F_{\text{A}}$, или $\rho > \rho_0$. В этом случае тело тонет.

2. Сила тяжести равна архимедовой силе: $mg = F_{\text{A}}$, или $\rho = \rho_0$. В этом случае тело остаётся неподвижным в состоянии *безразличного равновесия.*

3. Сила тяжести меньше архимедовой силы: $mg < F_{\text{A}}$, или $\rho < \rho_0$. В этом случае тело всплывает, достигая поверхности жидкости. При дальнейшем всплытии начнёт уменьшаться объём погружённой части тела, а вместе с ним и архимедова сила. В какой-то

момент архимедова сила сравняется с силой тяжести (положение равновесия). Тело по инерции всплывёт дальше, остановится, снова начнёт погружаться... Возникнут затухающие колебания, после которых тело останется плавать в положении равновесия ($mg = F_{\text{A}}$), частично погрузившись в жидкость.

Таким образом, *условие плавания* тела можно записать в виде неравенства: $\rho \leqslant \rho_0$. Например, лёд ($\rho = 900$ кг/м^3) будет плавать в воде ($\rho_0 = 1000$ кг/м^3), но утонет в спирте ($\rho_0 = 800$ кг/м^3).

1.15 Импульс

Импульс тела — это векторная величина, равная произведению массы тела на его скорость:

$$\vec{p} = m\vec{v}.$$

Специальных единиц измерения импульса нет. Размерность импульса — это просто произведение размерности массы на размерность скорости:

$$[p] = [m] \cdot [v] = \frac{\text{кг} \cdot \text{м}}{\text{с}}.$$

Почему понятие импульса является интересным? Оказывается, с его помощью можно придать второму закону Ньютона несколько иную, также чрезвычайно полезную форму.

1.15.1 Второй закон Ньютона в импульсной форме

Пусть $\vec{F}$ — равнодействующая сил, приложенных к телу массы m. Начинаем с обычной записи второго закона Ньютона:

$$m\vec{a} = \vec{F}.$$

С учётом того, что ускорение тела $\vec{a}$ равно производной вектора скорости, второй закон Ньютона переписывается следующим образом:

$$m\frac{d\vec{v}}{dt} = \vec{F}.$$

Вносим константу m под знак производной:

$$\frac{d(m\vec{v})}{dt} = \vec{F}.$$

Как видим, в левой части получилась производная импульса:

$$\frac{d\vec{p}}{dt} = \vec{F}. \tag{1.68}$$

Соотношение (1.68) и есть новая форма записи второго закона Ньютона.

Второй закон Ньютона в импульсной форме. Производная импульса тела есть равнодействующая приложенных к телу сил.

Можно сказать и так: результирующая сила, действующая на тело, равна скорости изменения импульса тела.

Производную в формуле (1.68) можно заменить на отношение конечных приращений:

$$\frac{\Delta\vec{p}}{\Delta t} = \vec{F}. \tag{1.69}$$

В этом случае $\vec{F}$ есть *средняя* сила, действующая на тело в течение интервала времени Δt. Чем меньше величина Δt, тем ближе отношение $\Delta\vec{p}/\Delta t$ к производной $d\vec{p}/dt$, и тем ближе средняя сила $\vec{F}$ к своему мгновенному значению в данный момент времени.

В задачах, как правило, интервал времени Δt достаточно мал. Например, это может быть время соударения мяча со стенкой, и тогда $\vec{F}$ — средняя сила, действующая на мяч со стороны стенки во время удара.

Вектор $\Delta\vec{p}$ в левой части соотношения (1.69) называется *изменением импульса* за время Δt. Изменение импульса — это разность конечного и начального векторов импульса. А именно,

если $\vec{p}_0$ — импульс тела в некоторый начальный момент времени, $\vec{p}$ — импульс тела спустя промежуток времени Δt, то изменение импульса есть разность:

$$\Delta\vec{p} = \vec{p} - \vec{p}_0.$$

Подчеркнём ещё раз, что изменение импульса — это разность векторов (рис. 1.39). Напомним, что при построении разности векторов нужно совместить начала обоих векторов, соединить их концы и «уколоть» стрелкой тот вектор, из которого производится вычитание.

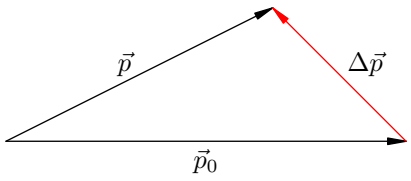


Рис. 1.39. Изменение импульса

Пусть, например, мяч летит перпендикулярно стенке (импульс перед ударом равен $\vec{p}_0$) и отскакивает назад без потери скорости (импульс после удара равен $\vec{p} = -\vec{p}_0$). Несмотря на то, что импульс по модулю не изменился ($p = p_0$), изменение импульса нулю не равно:

$$\Delta\vec{p} = \vec{p} - \vec{p}_0 = -\vec{p}_0 - \vec{p}_0 = -2\vec{p}_0.$$

Геометрически эта ситуация показана на рис. 1.40:

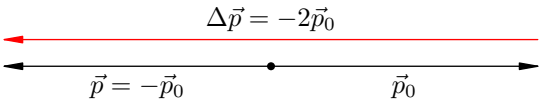


Рис. 1.40. Изменение импульса при отскоке назад

Модуль изменения импульса, как видим, равен удвоенному модулю начального импульса мяча: $\Delta p = 2p_0$.

Перепишем формулу (1.69) следующим образом:

$$\Delta\vec{p} = \vec{F}\Delta t, \tag{1.70}$$

или, расписывая изменение импульса, как и выше:

$$\vec{p} - \vec{p}_0 = \vec{F}\Delta t.$$

Величина $\vec{F}\Delta t$ называется *импульсом силы*. Специальной единицы измерения для импульса силы нет; размерность импульса силы равна просто произведению размерностей силы и времени:

$$[F\Delta t] = [F] \cdot [t] = \text{Н} \cdot \text{с}.$$

(Обратите внимание, что Н$\cdot$с оказывается ещё одной возможной единицей измерения импульса тела.)

Словесная формулировка равенства (1.70) такова: *изменение импульса тела равно импульсу действующей на тело силы за данный промежуток времени.* Это, разумеется, снова есть второй закон Ньютона в импульсной форме.

1.15.2 Пример вычисления силы

В качестве примера применения второго закона Ньютона в импульсной форме давайте рассмотрим следующую задачу.

Задача. Шарик массы $m = 100\,\text{г}$, летящий горизонтально со скоростью $v = 6$ м/с, ударяется о гладкую вертикальную стену и отскакивает от неё без потери скорости. Угол падения шарика (то есть угол между направлением движения шарика и перпендикуляром к стене) равен $\alpha = 60°$. Удар длится $\Delta t = 0{,}01\,\text{с}$. Найти среднюю силу, действующую на шарик во время удара.

Решение. Покажем прежде всего, что угол отражения равен углу падения, то есть шарик отскочит от стены под тем же углом α (рис. 1.41).

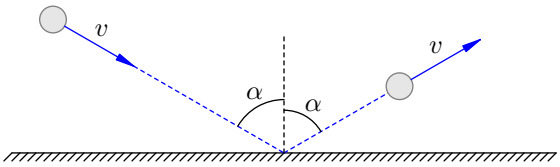


Рис. 1.41. К задаче (вид сверху)

Тут всё дело в том, что стена — гладкая. Это значит, что трения между шариком и стеной нет. Следовательно, со стороны стены на шарик действует единственная сила $\vec{N}$ — сила упругости, направленная перпендикулярно стене (рис. 1.42).

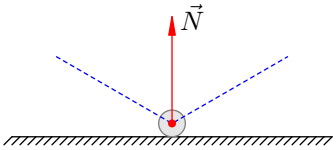


Рис. 1.42. К задаче

Согласно (1.70) имеем: $\Delta\vec{p} = \vec{N}\Delta t$. Отсюда следует, что вектор изменения импульса *сонаправлен* с вектором $\vec{N}$, то есть направлен перпендикулярно стене в сторону отскока шарика (рис. 1.43).

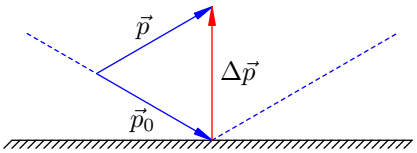


Рис. 1.43. К задаче

Векторы $\vec{p}_0$ и $\vec{p}$ равны по модулю (так как скорость шарика не изменилась). Поэтому треугольник, составленный из векторов $\vec{p}_0$, $\vec{p}$ и $\Delta\vec{p}$, является равнобедренным. Значит, угол между векторами $\vec{p}$ и $\Delta\vec{p}$ равен α, то есть угол отражения действительно равен углу падения.

Теперь заметим вдобавок, что в нашем равнобедренном треугольнике есть угол 60° (это угол падения); стало быть, данный треугольник — равносторонний. Отсюда:

$$\Delta p = p_0 = mv = 0{,}1 \cdot 6 = 0{,}6\ \text{Н} \cdot \text{с}.$$

И тогда искомая средняя сила, действующая на шарик:

$$N = \frac{\Delta p}{\Delta t} = \frac{0{,}6}{0{,}01} = 60\,\text{Н}.$$

1.15.3 Импульс системы тел

Начнём с простой ситуации системы двух тел. А именно, пусть имеются тело 1 и тело 2 с импульсами $\vec{p}_1$ и $\vec{p}_2$ соответственно. Импульс $\vec{p}$ системы данных тел — это векторная сумма импульсов каждого тела:

$$\vec{p} = \vec{p}_1 + \vec{p}_2.$$

Оказывается, для импульса системы тел имеется формула, аналогичная второму закону Ньютона в виде (1.68). Давайте выведем эту формулу.

Все остальные объекты, с которыми взаимодействуют рассматриваемые нами тела 1 и 2, мы будем называть *внешними телами*. Силы, с которыми внешние тела действуют на тела 1 и 2, называем *внешними силами*. Пусть $\vec{F}_1$ — результирующая внешняя сила, действующая на тело 1. Аналогично $\vec{F}_2$ — результирующая внешняя сила, действующая на тело 2 (рис. 1.44).

Рис. 1.44. Система двух тел

Кроме того, тела 1 и 2 могут взаимодействовать друг с другом. Пусть тело 2 действует на тело 1 с силой $\vec{T}$. Тогда тело 1 действует на тело 2 с силой $\vec{T}'$. По третьему закону Ньютона силы $\vec{T}$ и $\vec{T}'$ равны по модулю и противоположны по направлению: $\vec{T}' = -\vec{T}$. Силы $\vec{T}$ и $\vec{T}'$ — это *внутренние силы*, действующие в системе.

Запишем для каждого тела 1 и 2 второй закон Ньютона в форме (1.68):

$$\frac{d\vec{p}_1}{dt} = \vec{F}_1 + \vec{T}, \tag{1.71}$$

$$\frac{d\vec{p}_2}{dt} = \vec{F}_2 + \vec{T}'. \tag{1.72}$$

Сложим равенства (1.71) и (1.72):

$$\frac{d\vec{p}_1}{dt} + \frac{d\vec{p}_2}{dt} = \vec{F}_1 + \vec{F}_2 + \vec{T} + \vec{T}'.$$

В левой части полученного равенства стоит сумма производных, равная производной суммы векторов $\vec{p}_1$ и $\vec{p}_2$. В правой части имеем $\vec{T} + \vec{T}' = \vec{0}$ в силу третьего закона Ньютона:

$$\frac{d(\vec{p}_1 + \vec{p}_2)}{dt} = \vec{F}_1 + \vec{F}_2.$$

Но $\vec{p}_1 + \vec{p}_2 = \vec{p}$ — это импульс системы тел 1 и 2. Обозначим также $\vec{F}_1 + \vec{F}_2 = \vec{F}_{\text{внеш}}$ — это результирующая внешних сил, действующих на систему. Получаем:

$$\frac{d\vec{p}}{dt} = \vec{F}_{\text{внеш}}. \tag{1.73}$$

Таким образом, *скорость изменения импульса системы тел есть равнодействующая внешних сил, приложенных к системе.* Равенство (1.73), играющее роль второго закона Ньютона для системы тел, мы и хотели получить.

Формула (1.73) была выведена для случая двух тел. Теперь обобщим наши рассуждения на случай произвольного количества тел в системе.

Импульсом системы тел называется векторная сумма импульсов всех тел, входящих в систему. Если система состоит из N тел, то импульс этой системы равен:

$$\vec{p} = \vec{p}_1 + \vec{p}_2 + \ldots + \vec{p}_N.$$

Дальше всё делается совершенно так же, как и выше (только технически это выглядит несколько сложнее). Если для каждого тела записать равенства, аналогичные (1.71) и (1.72), а затем все эти равенства сложить, то в левой части мы снова получим производную импульса системы, а в правой части останется лишь сумма внешних сил (внутренние силы, попарно складываясь, дадут нуль ввиду третьего закона Ньютона). Поэтому равенство (1.73) останется справедливым и в общем случае.

1.15.4 Закон сохранения импульса

Система тел называется *замкнутой*, если действия внешних тел на тела данной системы или пренебрежимо малы, или компенсируют друг друга. Таким образом, в случае замкнутой системы тел существенно лишь взаимодействие этих тел друг с другом, но не с какими-либо другими телами.

Равнодействующая внешних сил, приложенных к замкнутой системе, равна нулю: $\vec{F}_{\text{внеш}} = \vec{0}$. В этом случае из (1.73) получаем:

$$\frac{d\vec{p}}{dt} = \vec{0}.$$

Но если производная вектора обращается в нуль (скорость изменения вектора равна нулю), то сам вектор не меняется со временем:

$$\vec{p} = \text{const}.$$

Закон сохранения импульса. Импульс замкнутой системы тел остаётся постоянным с течением времени при любых взаимодействиях тел внутри данной системы.

Простейшие задачи на закон сохранения импульса решаются по стандартной схеме, которую мы сейчас покажем.

Задача. Тело массы $m_1 = 800$ г движется со скоростью $v_1 = 3$ м/с по гладкой горизонтальной поверхности. Навстречу ему движется тело массы $m_2 = 200$ г со скоростью $v_2 = 13$ м/с. Происходит абсолютно неупругий удар (тела слипаются). Найти скорость тел после удара.

Решение. Ситуация изображена на рис. 1.45. Ось X направим в сторону движения первого тела.

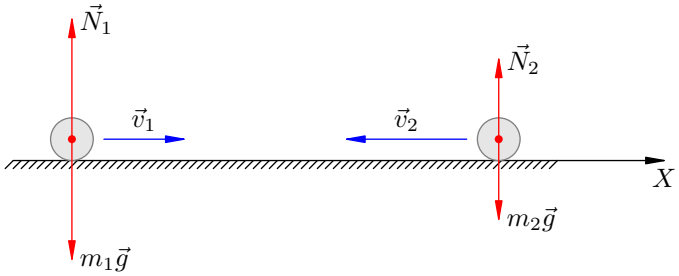


Рис. 1.45. К задаче

Поскольку поверхность гладкая, трения нет. Поскольку поверхность горизонтальная, а движение происходит вдоль неё, сила тяжести и реакция опоры уравновешивают друг друга:

$$m_1\vec{g} + \vec{N}_1 = \vec{0}, \quad m_2\vec{g} + \vec{N}_2 = \vec{0}.$$

Таким образом, векторная сумма сил, приложенных к системе данных тел, равна нулю. Это значит, что система тел замкнута. Стало быть, для неё выполняется закон сохранения импульса:

$$\vec{p}_{\text{до удара}} = \vec{p}_{\text{после удара}}. \tag{1.74}$$

Импульс системы до удара — это сумма импульсов тел:

$$\vec{p}_{\text{до удара}} = m_1\vec{v}_1 + m_2\vec{v}_2.$$

После неупругого удара получилось одно тело массы $m_1 + m_2$, которое движется с искомой скоростью $\vec{v}$:

$$\vec{p}_{\text{после удара}} = (m_1 + m_2)\vec{v}.$$

Из закона сохранения импульса (1.74) имеем:

$$m_1\vec{v}_1 + m_2\vec{v}_2 = (m_1 + m_2)\vec{v}.$$

Отсюда находим скорость тела, образовавшегося после удара:

$$\vec{v} = \frac{m_1\vec{v}_1 + m_2\vec{v}_2}{m_1 + m_2}\,.$$

Переходим к проекциям на ось X:

$$v_x = \frac{m_1 v_{1x} + m_2 v_{2x}}{m_1 + m_2}\,.$$

По условию имеем: $v_{1x} = 3$ м/с, $v_{2x} = -13$ м/с, так что

$$v_x = \frac{0{,}8 \cdot 3 - 0{,}2 \cdot 13}{0{,}8 + 0{,}2} = -0{,}2\,\frac{\text{м}}{\text{с}}\,.$$

Знак минус указывает на то, что слипшиеся тела двигаются в сторону, противоположную оси X. Искомая скорость: $v = 0{,}2$ м/с.

1.15.5 Закон сохранения проекции импульса

Часто в задачах встречается следующая ситуация. Система тел не является замкнутой (векторная сумма внешних сил, действующих на систему, не равна нулю), но существует такая ось X, что *сумма проекций внешних сил на ось X равна нулю* в любой момент времени. Тогда можно сказать, что вдоль данной оси наша система тел ведёт себя как замкнутая, и проекция импульса системы на ось X сохраняется.

Покажем это более строго. Спроектируем равенство (1.73) на ось X:

$$\frac{dp_x}{dt} = F_{\text{внеш}\,,\,x}.$$

Если проекция равнодействующей внешних сил обращается в нуль, $F_{\text{внеш}\,,\,x} = 0$, то

$$\frac{dp_x}{dt} = 0.$$

Следовательно, проекция p_x есть константа:

$$p_x = \text{const}.$$

Закон сохранения проекции импульса. Если проекция на ось X суммы внешних сил, действующих на систему, равна нулю, то проекция p_x импульса системы не меняется с течением времени.

Давайте посмотрим на примере конкретной задачи, как работает закон сохранения проекции импульса.

Задача. Мальчик массы M, стоящий на коньках на гладком льду, бросает камень массы m со скоростью v под углом α к горизонту. Найти скорость u, с которой мальчик откатывается назад после броска.

Решение. Ситуация схематически показана на рис. 1.46. Мальчик изображён прямогольником.

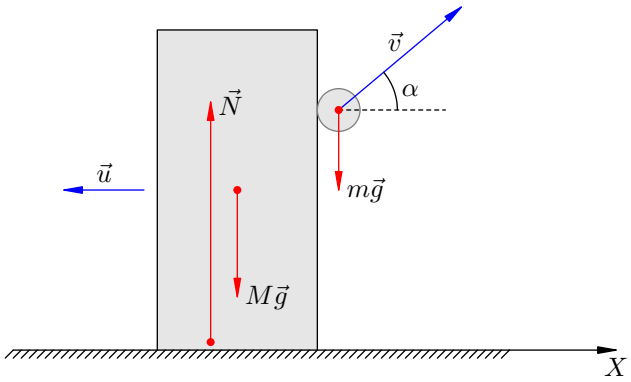


Рис. 1.46. К задаче

Импульс системы «мальчик + камень» не сохраняется. Это видно хотя бы из того, что после броска появляется вертикальная составляющая импульса системы (а именно, вертикальная составляющая импульса камня), которой до броска не было.

Стало быть, система, которую образуют мальчик и камень, не замкнута. Почему? Дело в том, что векторная сумма внешних сил $M\vec{g}+m\vec{g}+\vec{N}$ не равна нулю во время броска. Величина N больше, чем сумма $Mg + mg$, и за счёт этого превышения как раз и появляется вертикальная компонента импульса системы.

Однако внешние силы действуют только по вертикали (трения нет). Стало быть, сохраняется проекция импульса на горизонтальную ось X. До броска эта проекция была равна нулю. Направляя ось X в сторону броска (так что мальчик поехал в направлении отрицательной полуоси), получим:

$$-Mu + mv_0 \cos\alpha = 0,$$

откуда

$$u = \frac{mv_0 \cos\alpha}{M}\,.$$

1.16 Энергия

Мы приступаем к изучению энергии — фундаментального физического понятия. Но предварительно нужно разобраться с другой физической величиной — работой силы.

1.16.1 Работа

Пусть на тело действует постоянная сила $\vec{F}$ и тело, двигаясь прямолинейно по горизонтальной поверхности, совершило перемещение $\vec{s}$. Сила $\vec{F}$ не обязательно является непосредственной причиной перемещения (так, сила тяжести не является непосредственной причиной перемещения шкафа, который передвигают по комнате).

Предположим сначала, что векторы силы и перемещения сонаправлены (рис. 1.47; остальные силы, действующие на тело, не указаны).

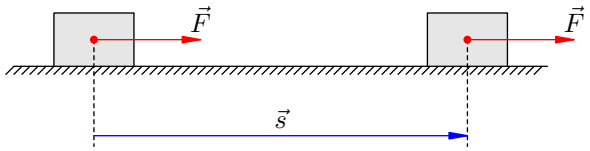


Рис. 1.47. $A = Fs$

В этом простейшем случае работа A определяется как произведение модуля силы на модуль перемещения:

$$A = Fs. \tag{1.75}$$

Единицей измерения работы служит *джоуль* (Дж): Дж $=$ Н $\cdot$ м. Таким образом, если под действием силы 1 Н тело перемещается на 1 м, то сила совершает работу 1 Дж.

Работа силы, перпендикулярной перемещению, по определению считается равной нулю. Так, в данном случае сила тяжести и сила реакции опоры не совершают работы.

Пусть теперь вектор силы образует с вектором перемещения острый угол α (рис. 1.48).

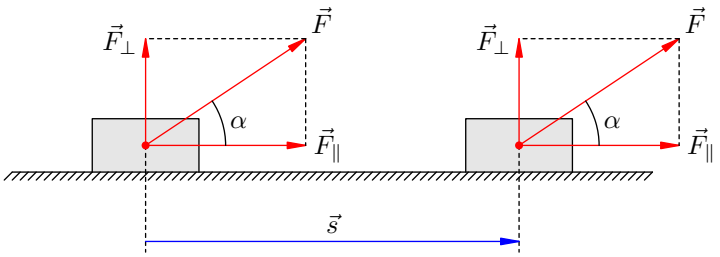


Рис. 1.48. $A = Fs\cos\alpha$

Разложим силу $\vec{F}$ на две составляющие: $\vec{F}_\parallel$ (параллельную перемещению) и $\vec{F}_\perp$ (перпендикулярную перемещению). Работу совершает только $\vec{F}_\parallel$. Поэтому для работы силы $\vec{F}$ получаем: $A = F_\parallel s = F\cos\alpha \cdot s$. Итак,

$$A = Fs\cos\alpha. \tag{1.76}$$

Если вектор силы образует с вектором перемещения тупой угол α, то работа по-прежнему определяется формулой (1.76). В этом случае работа оказывается отрицательной.

Например, работа силы трения скольжения, действующей на тело в рассмотренных ситуациях, будет отрицательной, так как сила трения направлена противоположно перемещению. В этом случае имеем: $\alpha = 180°$, $\cos\alpha = -1$, и для работы силы трения получаем:

$$A_{\text{тр}} = -F_{\text{тр}}s = -\mu mgs,$$

где m — масса тела, μ — коэффициент трения между телом и опорой.

Соотношение (1.76) означает, что работа является *скалярным произведением* векторов силы и перемещения:

$$A = \vec{F}\vec{s}.$$

Это позволяет вычислять работу через координаты данных векторов:

$$A = F_x s_x + F_y s_y + F_z s_z.$$

Пусть на тело действуют несколько сил $\vec{F}_1, \vec{F}_2, \ldots, \vec{F}_n$, и $\vec{F}$ — равнодействующая этих сил. Для работы силы $\vec{F}$ имеем:

$$A = \vec{F}\vec{s} = (\vec{F}_1 + \vec{F}_2 + \ldots + \vec{F}_n)\vec{s} = \vec{F}_1\vec{s} + \vec{F}_2\vec{s} + \ldots + \vec{F}_n\vec{s},$$

или

$$A = A_1 + A_2 + \ldots + A_n,$$

где $A_1, A_2, \ldots, A_n$ — работы сил $\vec{F}_1, \vec{F}_2, \ldots, \vec{F}_n$. Итак, *работа равнодействующей приложенных к телу сил равна сумме работ каждой силы в отдельности.*

1.16.2 Мощность

Часто имеет значение быстрота, с которой совершается работа. Скажем, на практике важно знать, какую работу сможет выполнить данное устройство за фиксированное время.

Мощность — это величина, характеризующая скорость совершения работы. Мощность N есть отношение работы A ко времени t, за которое эта работа совершена:

$$N = \frac{A}{t}\,.$$

Мощность измеряется в *ваттах* (Вт). 1 Вт = 1 Дж/с, то есть 1 Вт — это такая мощность, при которой работа в 1 Дж совершается за 1 с.

Предположим, что силы, действующие на тело, уравновешены, и тело движется равномерно и прямолинейно со скоростью $\vec{v}$. В этом случае существует полезная формула для мощности, развиваемой одной из действующих сил $\vec{F}$.

За время t тело совершит перемещение $\vec{s} = \vec{v}t$. Работа силы $\vec{F}$ будет равна:

$$A = \vec{F}\vec{s} = \vec{F}\vec{v}t.$$

Отсюда получаем мощность:

$$N = \vec{F}\vec{v},$$

или

$$N = Fv\cos\alpha,$$

где α — угол между векторами силы и скорости.

Наиболее часто эта формула используется в ситуации, когда $\vec{F}$ — «сила тяги» двигателя автомобиля (которая на самом деле есть сила трения ведущих колёс о дорогу). В этом случае $\alpha = 0$, и мы получаем просто:

$$N = Fv.$$

1.16.3 Механическая энергия

Энергия является мерой движения и взаимодействия любых объектов в природе. Имеются различные формы энергии: механическая, тепловая, электромагнитная, ядерная...

Опыт показывает, что энергия не появляется ниоткуда и не исчезает бесследно, она лишь переходит из одной формы в другую. Это самая общая формулировка *закона сохранения энергии*.

Каждый вид энергии представляет собой некоторое математическое выражение. Закон сохранения энергии означает, что в каждом явлении природы определённая сумма таких выражений остаётся постоянной с течением времени.

Измеряется энергия в джоулях, как и работа.

Механическая энергия является мерой движения и взаимодействия механических объектов (материальных точек, твёрдых тел).

Мерой движения тела является *кинетическая энергия*. Она зависит от скорости тела. Мерой взаимодействия тел является *потенциальная энергия*. Она зависит от взаимного расположения тел.

Механическая энергия системы тел равна сумме кинетической энергии тел и потенциальной энергии их взаимодействия друг с другом.

1.16.4 Кинетическая энергия

Кинетической энергией тела (принимаемого за материальную точку) называется величина

$$K = \frac{mv^2}{2},$$

где m — масса тела, v — его скорость.

Кинетической энергией системы из N тел называется сумма кинетических энергий каждого тела:

$$K = \frac{m_1 v_1^2}{2} + \frac{m_2 v_2^2}{2} + \ldots + \frac{m_N v_N^2}{2}.$$

Если тело движется под действием силы $\vec{F}$, то кинетическая энергия тела, вообще говоря, меняется со временем. Оказывается, именение кинетической энергии тела за некоторый промежуток времени равно работе силы $\vec{F}$. Покажем это для случая прямолинейного равноускоренного движения.

Пусть $\vec{v}_1$ — начальная скорость, $\vec{v}_2$ — конечная скорость тела. Выберем ось X вдоль траектории тела (и, соответственно, вдоль вектора силы $\vec{F}$). Для работы силы $\vec{F}$ получаем:

$$A = \vec{F}\vec{s} = F_x s_x = m a_x s_x = m a_x \frac{v_{2x}^2 - v_{1x}^2}{2a_x} = \frac{m v_{2x}^2 - m v_{1x}^2}{2}.$$

(мы воспользовались формулой для s_x, выведенной в разделе «Равноускоренное движение»).

Заметим теперь, что в данном случае проекция скорости отличается от модуля скорости разве что знаком; поэтому $v_{1x}^2 = v_1^2$ и $v_{2x}^2 = v_2^2$. В результате имеем:

$$A = \frac{m v_2^2}{2} - \frac{m v_1^2}{2} = K_2 - K_1 = \Delta K,$$

что и требовалось.

На самом деле соотношение $\Delta K = A$ справедливо и в самом общем случае криволинейного движения под действием переменной силы.

Теорема о кинетической энергии. Изменение кинетической энергии тела равно работе, совершённой приложенными к телу внешними силами за рассматриваемый промежуток времени.

Если работа внешних сил положительна, то кинетическая энергия увеличивается ($\Delta K > 0$, тело разгоняется).

Если работа внешних сил отрицательна, то кинетическая энергия уменьшается ($\Delta K < 0$, тело замедляет движение). Пример — торможение под действием силы трения, работа которой отрицательна.

Если же работа внешних сил равна нулю, то кинетическая энергия тела за это время не меняется. Нетривиальный пример — равномерное движение по окружности, совершаемое грузом на нити в горизонтальной плоскости. Сила тяжести, сила реакции опоры и сила натяжения нити всегда перпендикулярны скорости, и работа каждой из этих сил равна нулю в течение любого промежутка времени. Соответственно, кинетическая энергия груза (а значит, и его скорость) остаётся постоянной в процессе движения.

Задача. Автомобиль едет по горизонтальной дороге со скоростью v и начинает резко тормозить. Найти путь s, пройденный автомобилем до полной остановки, если коэффициент трения шин о дорогу равен μ.

Решение. Начальная кинетическая энергия автомобиля $K_1 = \dfrac{mv^2}{2}$, конечная кинетическая энергия $K_2 = 0$. Изменение кинетической энергии $\Delta K = K_2 - K_1 = -\dfrac{mv^2}{2}$.

На автомобиль действуют сила тяжести $m\vec{g}$, реакция опоры $\vec{N}$ и сила трения $\vec{f}$. Сила тяжести и реакция опоры, будучи перпендикулярны перемещению автомобиля, работы не совершают. Работа силы трения:

$$A = -fs = -\mu Ns = -\mu mgs.$$

Из теоремы о кинетической энергии теперь получаем:

$$\Delta K = A \Rightarrow -\frac{mv^2}{2} = -\mu mgs \Rightarrow s = \frac{v^2}{2\mu g}.$$

1.16.5 Потенциальная энергия тела вблизи поверхности Земли

Рассмотрим тело массы m, находящееся на некоторой высоте над поверхностью Земли. Высоту считаем много меньше земного радиуса. Изменением силы тяжести в процессе перемещения тела пренебрегаем.

Если тело находится на высоте h, то *потенциальная энергия* тела по определению равна:

$$W = mgh,$$

где g — ускорение свободного падения вблизи поверхности Земли.

Высоту не обязательно отсчитывать от поверхности Земли. Как мы увидим ниже (формулы (1.77), (1.78)), физическим смыслом обладает не сама по себе потенциальная энергия, но её *изменение*. А изменение потенциальной энергии не зависит от уровня отсчёта. Выбор нулевого уровня потенциальной энергии в конкретной задаче диктуется исключительно соображениями удобства.

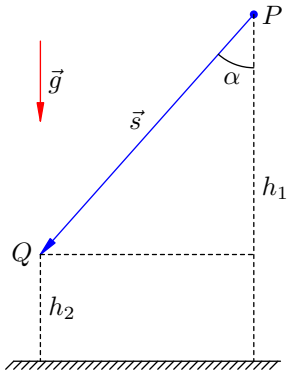


Рис. 1.49. $A = mg(h_1 - h_2)$

Найдём работу, совершаемую силой тяжести при перемещении тела. Предположим, что тело перемещается по прямой из точки P, находящейся на высоте h_1, в точку Q, находящуюся на высоте h_2 (рис. 1.49).

Угол между силой тяжести $m\vec{g}$ и перемещением тела $\vec{s}$ обозначим α. Для работы силы тяжести получим:

$$A = m\vec{g}\vec{s} = mgs\cos\alpha.$$

Но, как видно из рис. 1.49, $s\cos\alpha = h_1 - h_2$. Поэтому

$$A = mg(h_1 - h_2) = mgh_1 - mgh_2,$$

или

$$A = W_1 - W_2. \tag{1.77}$$

Учитывая, что $W_1 - W_2 = -(W_2 - W_1) = -\Delta W$, имеем также:

$$A = -\Delta W. \tag{1.78}$$

Можно доказать, что формулы (1.77) и (1.78) справедливы *для любой траектории*, по которой тело перемещается из точки P в точку Q, а не только для прямолинейного отрезка.

Работа силы тяжести не зависит от формы траектории, по которой перемещается тело, и равна разности значений потенциальной энергии в начальной и конечной точках траектории. Иными словами, работа силы тяжести всегда равна изменению потенциальной энергии с противоположным знаком.

В частности, работа силы тяжести по любому замкнутому пути равна нулю.

Сила называется *консервативной*, если при перемещении тела работа этой силы не зависит от формы траектории, а определяется только начальным и конечным положением тела. Сила тяжести, таким образом, является консервативной. Работа консервативной силы по любому замкнутому пути равна нулю. Только в случае консервативной силы возможно ввести такую величину, как потенциальная энергия.

1.16.6 Потенциальная энергия деформированной пружины

Рассмотрим пружину жёсткости k. Начальная деформация пружины равна x_1. Предположим, что пружина деформируется до некоторой конечной величины деформации x_2. Чему равна при этом работа силы упругости пружины?

В данном случае силу на перемещение не умножишь, так как сила упругости меняется в процессе деформации пружины. Для нахождения работы переменной силы требуется интегрирование. Мы не будем приводить здесь вывод, а сразу выпишем конечный результат.

Оказывается, сила упругости пружины также является консервативной. Её работа зависит лишь от величин x_1, x_2 и определяется формулой:

$$A = \frac{kx_1^2}{2} - \frac{kx_2^2}{2}.$$

Величина

$$W = \frac{kx^2}{2}$$

называется *потенциальной энергией деформированной пружины* (x — величина деформации). Следовательно,

$$A = W_1 - W_2 = -\Delta W,$$

что полностью аналогично формулам (1.77) и (1.78).

1.16.7 Закон сохранения механической энергии

Консервативные силы называются так потому, что сохраняют механическую энергию замкнутой системы тел.

Механическая энергия E тела равна сумме его кинетической и потенциальной энергий:

$$E = K + W.$$

Механическая энергия системы тел равна сумме их кинетических энергий и потенциальной энергии их взаимодействия друг с другом.

Предположим, что тело совершает движение под действием силы тяжести и/или силы упругости пружины. Будем считать, что трения нет. Пусть в начальном положении кинетическая и потенциальная энергии тела равны K_1 и W_1, в конечном положении — K_2 и W_2. Работу внешних сил при перемещении тела из начального положения в конечное обозначим A.

По теореме о кинетической энергии:

$$K_2 - K_1 = A.$$

Но работа консервативных сил равна разности потенциальных энергий:

$$A = W_1 - W_2.$$

Отсюда получаем:

$$K_2 - K_1 = W_1 - W_2,$$

или

$$K_1 + W_1 = K_2 + W_2.$$

Левая и правая части данного равенства представляют собой механическую энергию тела в начальном и конечном положении:

$$E_1 = E_2.$$

Следовательно, *при движении тела в поле силы тяжести и/или на пружине механическая энергия тела остаётся неизменной при отсутствии трения.*

Справедливо и более общее утверждение.

Закон сохранения механической энергии. Если в замкнутой системе действуют только консервативные силы, то механическая энергия системы сохраняется.

При этих условиях могут происходить лишь превращения энергии: из кинетической в потенциальную и наоборот. Общий запас механической энергии системы остаётся постоянным.

1.16.8 Закон изменения механической энергии

Если между телами замкнутой системы имеются силы сопротивления (сухое или вязкое трение), то механическая энергия системы будет уменьшаться. Так, автомобиль останавливается в результате торможения, колебания маятника постепенно затухают и т. д. Силы трения неконсервативны: работа силы трения очевидным образом зависит от пути, по которому перемещается тело между данными точками. В частности, работа силы трения по замкнутому пути не равна нулю.

Снова рассмотрим движение тела в поле силы тяжести и/или на пружине. Вдобавок на тело действует сила трения, которая за рассматриваемый промежуток времени совершает отрицательную работу $A_{\text{тр}}$. Работу консервативных сил (тяжести и упругости) по-прежнему обозначаем A.

Изменение кинетической энергии тела равно работе *всех* внешних сил:

$$K_2 - K_1 = A + A_{\text{тр}}.$$

Но $A = W_1 - W_2$, следовательно

$$K_2 - K_1 = W_1 - W_2 + A_{\text{тр}}.$$

Отсюда

$$K_2 + W_2 - (K_1 + W_1) = A_{\text{тр}},$$

или

$$E_2 - E_1 = A_{\text{тр}}.$$

В левой части стоит величина $\Delta E = E_2 - E_1$ — изменение механической энергии тела:

$$\Delta E = A_{\text{тр}}.$$

Итак, *при движении тела в поле силы тяжести и/или на пружине изменение механической энергии тела равно работе силы трения.* Так как работа силы трения отрицательна, изменение механической энергии также отрицательно: механическая энергия убывает.

Справедливо и более общее утверждение.

Закон изменения механической энергии. Изменение механической энергии замкнутой системы равно работе сил трения, действующих внутри системы.

Ясно, что закон сохранения механической энергии является частным случаем данного утверждения.

Конечно, убыль механической энергии не противоречит общефизическому закону сохранения энергии. В данном случае механическая энергия превращается в энергию теплового движения частиц вещества и их потенциальную энергию взаимодействия друг с другом, т. е. переходит во *внутреннюю энергию* тел системы.

1.17 Простые механизмы

Механизмом в физике называется приспособление для преобразования силы (её увеличения или уменьшения). Например, прикладывая небольшое усилие в одном месте механизма, можно получить значительно большее усилие в другом его месте.

Один вид механизма нам уже встретился: это гидравлический пресс. Здесь мы рассмотрим так называемые *простые механизмы* — рычаг и наклонную плоскость.

1.17.1 Рычаг

Рычаг — это твёрдое тело, которое может вращаться вокруг неподвижной оси. На рис. 1.50 изображён рычаг с осью вращения O. К концам рычага (точкам A и B) приложены силы $\vec{F}_1$ и $\vec{F}_2$. Плечи этих сил равны соответственно l_1 и l_2.

Условие равновесия рычага даётся правилом моментов: $F_1 l_1 = F_2 l_2$, откуда

$$\frac{F_1}{F_2} = \frac{l_2}{l_1}.$$

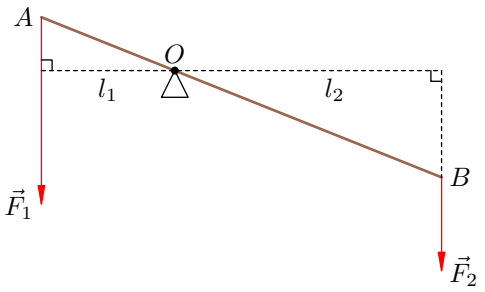


Рис. 1.50. Рычаг

Из этого соотношения следует, что рычаг даёт выигрыш в силе или в расстоянии (смотря по тому, с какой целью он используется) во столько раз, во сколько большее плечо длиннее меньшего.

Например, чтобы усилием 100 Н поднять груз весом 700 Н, нужно взять рычаг с отношением плеч 7 : 1 и положить груз на короткое плечо. Мы выиграем в силе в 7 раз, но во столько же раз проиграем в расстоянии: конец длинного плеча опишет в 7 раз бóльшую дугу, чем конец короткого плеча (то есть груз).

Примерами рычага, дающего выигрыш в силе, являются лопата, ножницы, плоскогубцы. Весло гребца — это рычаг, дающий выигрыш в расстоянии. А обычные рычажные весы являются равноплечим рычагом, не дающим выигрыша ни в расстоянии, ни в силе (в противном случае их можно использовать для обвешивания покупателей).

1.17.2 Неподвижный блок

Важной разновидностью рычага является *блок* — укреплённое в обойме колесо с жёлобом, по которому пропущена верёвка. В большинстве задач верёвка считается невесомой нерастяжимой нитью.

На рис. 1.51 изображён *неподвижный блок*, т. е. блок с неподвижной осью вращения (проходящей перпендикулярно плоскости рисунка через точку O).

На правом конце нити в точке D закреплён груз весом $\vec{P}$. Напомним, что вес тела — это сила, с которой тело давит на опору или растягивает подвес. В данном случае вес $\vec{P}$ приложен к точке D, в которой груз крепится к нити.

К левому концу нити в точке C приложена сила $\vec{F}$.

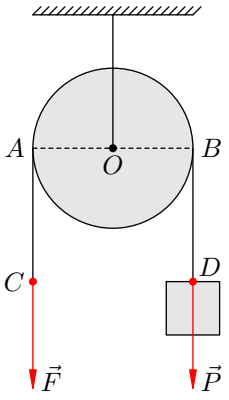


Рис. 1.51. Неподвижный блок

Плечо силы $\vec{F}$ равно $OA = r$, где r — радиус блока. Плечо веса $\vec{P}$ равно $OB = r$. Значит, неподвижный блок является равноплечим рычагом и потому не даёт выигрыша ни в силе, ни в расстоянии: во-первых,

имеем равенство $F = P$, а во-вторых, в процессе движении груза и нити перемещение точки C равно перемещению груза.

Зачем же тогда вообще нужен неподвижный блок? Он полезен тем, что позволяет *изменить направление усилия*. Обычно неподвижный блок используется как часть более сложных механизмов.

1.17.3 Подвижный блок

На рис. 1.52 изображён *подвижный блок*, ось которого перемещается вместе с грузом. Мы тянем за нить с силой $\vec{F}$, которая приложена в точке C и направлена вверх. Блок вращается и при этом также движется вверх, поднимая груз, подвешенный на нити OD.

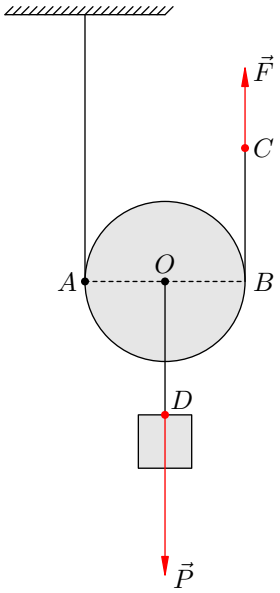


Рис. 1.52. Подвижный блок

В данный момент времени неподвижной точкой является точка A, и именно вокруг неё поворачивается блок (он бы «перекатывается» через точку A). Говорят ещё, что через точку A проходит *мгновенная ось вращения* блока (эта ось направлена перпендикулярно плоскости рисунка).

Вес груза $\vec{P}$ приложен в точке D крепления груза к нити. Плечо силы $\vec{P}$ равно $AO = r$.

А вот плечо силы $\vec{F}$, с которой мы тянем за нить, оказывается в два раза больше: оно равно $AB = 2r$. Соответственно, условием равновесия груза является равенство $F = P/2$ (что мы и видим на рис. 1.52: длина вектора $\vec{F}$ в два раза меньше длины вектора $\vec{P}$).

Следовательно, *подвижный блок даёт выигрыш в силе в два раза*. При этом, однако, мы в те же два раза проигрываем в расстоянии. Действительно, нетрудно сообразить, что для поднятия груза на один метр точку C придётся переместить вверх на два метра (то есть вытянуть два метра нити).

У блока на рис. 1.52 есть один недостаток: тянуть нить вверх (за точку C) — не самая лучшая идея. Согласитесь, что гораздо удобнее тянуть за нить вниз! Вот тут-то нас и выручает неподвижный блок.

На рис. 1.53 изображён подъёмный механизм, который представляет собой комбинацию подвижного блока с неподвижным. К подвижному блоку подвешен груз, а трос дополнительно перекинут через неподвижный блок, что даёт возможность тянуть за трос *вниз* для подъёма груза *вверх*. Внешнее усилие на тросе снова обозначено вектором $\vec{F}$.

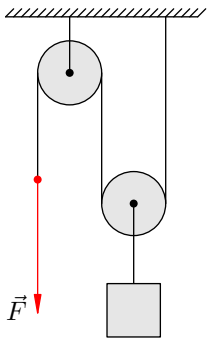


Рис. 1.53. Комбинация блоков

Принципиально данное устройство ничем не отличается от подвижного блока: с его помощью мы также получаем двукратный выигрыш в силе.

1.17.4 Наклонная плоскость

Как мы знаем, тяжёлую бочку проще вкатить по наклонным мосткам, чем поднимать вертикально. Мостки, таким образом, являются механизмом, который даёт выигрыш в силе.

В механике подобный механизм называется *наклонной плоскостью*. Наклонная плоскость — это ровная плоская поверхность, расположенная под некоторым углом α к горизонту. В таком

случае коротко говорят: «наклонная плоскость с углом α».

Найдём силу, которую надо приложить к грузу массы m, чтобы равномерно поднять его по *гладкой* наклонной плоскости с углом α. Эта сила $\vec{F}$, разумеется, направлена вдоль наклонной плоскости (рис. 1.54).

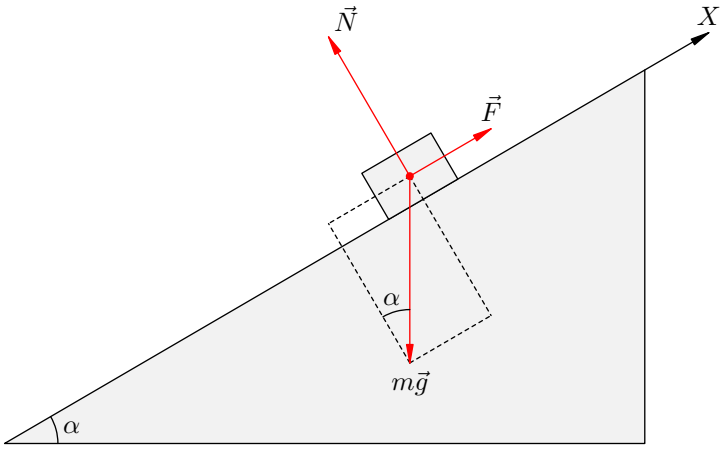


Рис. 1.54. Гладкая наклонная плоскость

Выберем ось X так, как показано на рисунке. Поскольку груз движется без ускорения, действующие на него силы уравновешены:

$$m\vec{g} + \vec{N} + \vec{F} = \vec{0}.$$

Проектируем на ось X:

$$-mg\sin\alpha + F = 0,$$

откуда

$$F = mg\sin\alpha.$$

Именно такую силу нужно приложить, что двигать груз вверх по наклонной плоскости.

Чтобы равномерно поднимать тот же груз по вертикали, к нему нужно приложить силу, равную mg. Видно, что $F < mg$, поскольку $\sin\alpha < 1$. Наклонная плоскость действительно даёт выигрыш в силе, и тем больший, чем меньше угол α.

Широко применяемыми разновидностями наклонной плоскости являются *клин* и *винт*.

1.17.5 Золотое правило механики

Простой механизм может дать выигрыш в силе или в расстоянии, но не может дать выигрыша в работе.

Например, рычаг с отношением плеч 2 : 1 даёт выигрыш в силе в два раза. Чтобы на меньшем плече поднять груз весом P, нужно к большему плечу приложить силу $P/2$. Но для поднятия груза на высоту h большее плечо придётся опустить на $2h$, и совершённая работа будет равна

$$A = \frac{P}{2}\cdot 2h = Ph,$$

т. е. той же величине, что и без использования рычага.

В случае наклонной плоскости мы выигрываем в силе, так как прикладываем к грузу силу $F = mg\sin\alpha$, меньшую силы тяжести. Однако, чтобы поднять груз на высоту h над начальным положением, нам нужно пройти путь $l = h/\sin\alpha$ вдоль наклонной плоскости. При этом мы совершаем работу

$$A = mg\sin\alpha\,\frac{h}{\sin\alpha} = mgh,$$

т. е. ту же самую, что и при вертикальном поднятии груза.

Данные факты служат проявлениями так называемого *золотого правила механики.*

Золотое правило механики. Ни один из простых механизмов не даёт выигрыша в работе. Во сколько раз выигрываем в силе, во столько же раз проигрываем в расстоянии, и наоборот.

Золотое правило механики есть не что иное, как простой вариант закона сохранения энергии.

1.17.6 КПД механизма

На практике приходится различать *полезную работу* $A_{\text{полезн}}$, которую нужно совершить при помощи механизма в идеальных условиях отсутствия каких-либо потерь, и *полную работу* $A_{\text{полн}}$, которая совершается для тех же целей в реальной ситуации.

Полная работа равна сумме:

- полезной работы;
- работы, совершённой против сил трения в различных частях механизма;
- работы, совершённой по перемещению составных элементов механизма.

Так, при подъёме груза рычагом приходится вдобавок совершать работу по преодолению силы трения в оси рычага и по перемещению самого рычага, имеющего некоторый вес.

Полная работа всегда больше полезной. Отношение полезной работы к полной называется *коэффициентом полезного действия (КПД) механизма*:

$$\eta = \frac{A_{\text{полезн}}}{A_{\text{полн}}}\,.$$

КПД принято выражать в процентах. КПД реальных механизмов всегда меньше 100%.

Вычислим КПД наклонной плоскости с углом α при наличии трения. Коэффициент трения между поверхностью наклонной плоскости и грузом равен μ.

Пусть груз массы m равномерно поднимается вдоль наклонной плоскости под действием силы $\vec{F}$ из точки P в точку Q на высоту h (рис. 1.55). В направлении, противоположном перемещению, на груз действует сила трения скольжения $\vec{f}$.

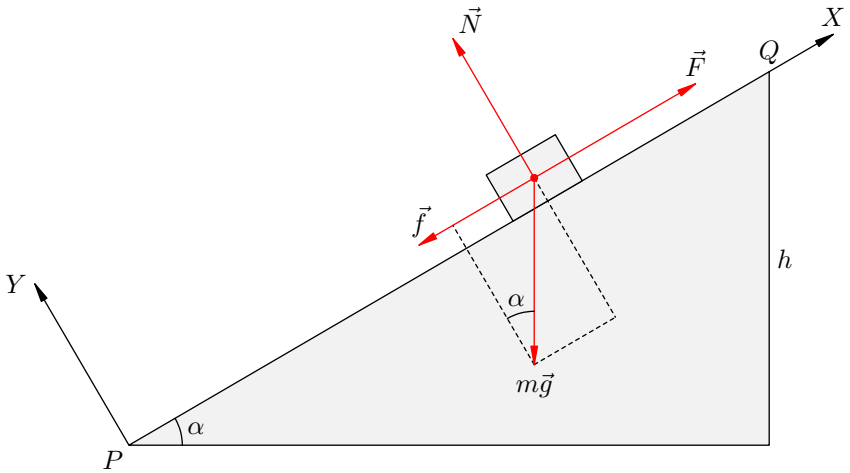


Рис. 1.55. Наклонная плоскость с трением

Ускорения нет, поэтому силы, действующие на груз, уравновешены:

$$m\vec{g} + \vec{N} + \vec{F} + \vec{f} = \vec{0}.$$

Проектируем на ось X:

$$-mg\sin\alpha + F - f = 0. \tag{1.79}$$

Проектируем на ось Y:

$$-mg\cos\alpha + N = 0. \tag{1.80}$$

Кроме того,

$$f = \mu N. \tag{1.81}$$

Из (1.80) имеем:

$$N = mg\cos\alpha.$$

Тогда из (1.81):

$$f = \mu mg\cos\alpha.$$

Подставляя это в (1.79), получаем:

$$F = mg\sin\alpha + f = mg\sin\alpha + \mu mg\cos\alpha = mg(\sin\alpha + \mu\cos\alpha).$$

Полная работа равна произведению силы F на путь, пройденный телом вдоль поверхности наклонной плоскости:

$$A_{\text{полн}} = F\cdot PQ = mg(\sin\alpha + \mu\cos\alpha)\,\frac{h}{\sin\alpha} = mgh(1 + \mu\operatorname{ctg}\alpha).$$

Полезная работа, очевидно, равна:

$$A_{\text{полезн}} = mgh.$$

Для искомого КПД получаем:

$$\eta = \frac{A_{\text{полезн}}}{A_{\text{полн}}} = \frac{mgh}{mgh(1 + \mu\operatorname{ctg}\alpha)} = \frac{1}{1 + \mu\operatorname{ctg}\alpha}\,.$$

1.18 Механические колебания

Колебания — это повторяющиеся во времени изменения состояния системы. Понятие колебаний охватывает очень широкий круг явлений.

Колебания механических систем, или *механические колебания* — это механическое движение тела или системы тел, которое обладает повторяемостью во времени и происходит в окрестности положения равновесия. *Положением равновесия* называется такое состояние системы, в котором она может оставаться сколь угодно долго, не испытывая внешних воздействий.

Например, если маятник отклонить и отпустить, то начнутся колебания. Положение равновесия — это положение маятника при отсутствии отклонения. В этом положении маятник, если его не трогать, может пребывать сколь угодно долго. При колебаниях маятник много раз проходит положение равновесия.

Сразу после того, как отклонённый маятник отпустили, он начал двигаться, прошёл положение равновесия, достиг противоположного крайнего положения, на мгновение остановился в нём, двинулся в обратном направлении, снова прошёл положение равновесия и вернулся назад. Совершилось одно *полное колебание*. Дальше этот процесс будет периодически повторяться.

Амплитуда колебаний тела — это величина его наибольшего отклонения от положения равновесия.

Период колебаний T — это время одного полного колебания. Можно сказать, что за период тело проходит путь в четыре амплитуды.

Частота колебаний ν — это величина, обратная периоду: $\nu = 1/T$. Частота измеряется в герцах (Гц) и показывает, сколько полных колебаний совершается за одну секунду.

1.18.1 Гармонические колебания

Будем считать, что положение колеблющегося тела определяется одной-единственной координатой x. Положению равновесия отвечает значение $x = 0$. Основная задача механики в данном случае состоит в нахождении функции $x(t)$, дающей координату тела в любой момент времени.

Для математического описания колебаний естественно использовать периодические функции. Таких функций много, но две из них — синус и косинус — являются самыми важными. У них много хороших свойств, и они тесно связаны с широким кругом физических явлений.

Поскольку функции синус и косинус получаются друг из друга сдвигом аргумента на $\pi/2$, можно ограничиться только одной из них. Мы для определённости будем использовать косинус.

Гармонические колебания — это колебания, при которых координата зависит от времени по *гармоническому закону*:

$$x = A\cos(\omega t + \alpha). \tag{1.82}$$

Выясним смысл входящих в эту формулу величин.

Положительная величина A является наибольшим по модулю значением координаты (так как максимальное значение модуля косинуса равно единице), т. е. наибольшим отклонением от положения равновесия. Поэтому A — амплитуда колебаний.

Аргумент косинуса $\omega t + \alpha$ называется *фазой* колебаний. Величина α, равная значению фазы при $t = 0$, называется *начальной фазой*. Начальная фаза отвечает начальной координате тела: $x_0 = A\cos\alpha$.

Величина ω называется *циклической частотой*. Найдём её связь с периодом колебаний T и частотой ν. Одному полному колебанию отвечает приращение фазы, равное 2π радиан: $\omega T = 2\pi$, откуда

$$\omega = \frac{2\pi}{T}, \tag{1.83}$$

$$\omega = 2\pi\nu. \tag{1.84}$$

Измеряется циклическая частота в рад/с (радиан в секунду).

В соответствии с выражениями (1.83) и (1.84) получаем ещё две формы записи гармонического закона (1.82):

$$x = A\cos\left(\frac{2\pi t}{T} + \alpha\right), \quad x = A\cos(2\pi\nu t + \alpha).$$

График функции (1.82), выражающей зависимость координаты от времени при гармонических колебаниях, приведён на рис. 1.56.

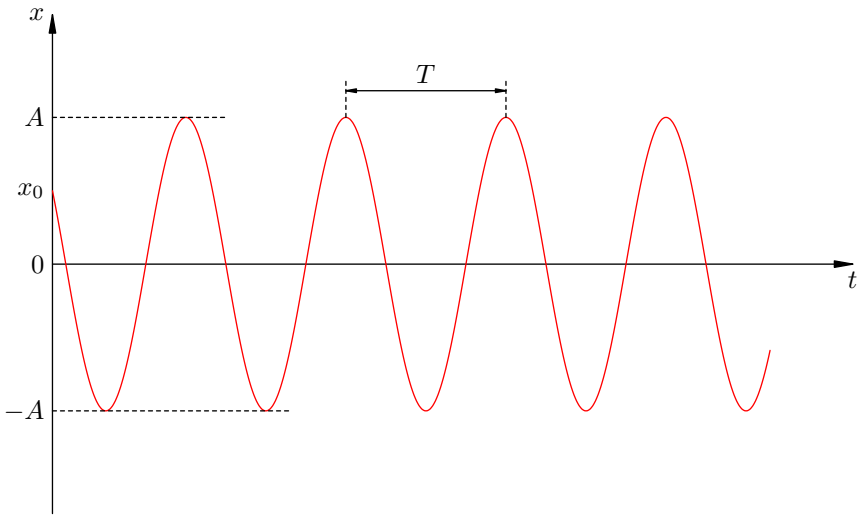


Рис. 1.56. График гармонических колебаний

Гармонический закон вида (1.82) носит самый общий характер. Он отвечает, например, ситуации, когда с маятником совершили одновременно два начальных действия: отклонили на величину x_0 и придали ему некоторую начальную скорость. Имеются два важных частных случая, когда одно из этих действий не совершалось.

Пусть маятник отклонили, но начальной скорости не сообщали (отпустили без начальной скорости). Ясно, что в этом случае $x_0 = A$, поэтому можно положить $\alpha = 0$. Мы получаем *закон косинуса*:

$$x = A\cos\omega t.$$

График гармонических колебаний в этом случае представлен на рис. 1.57.

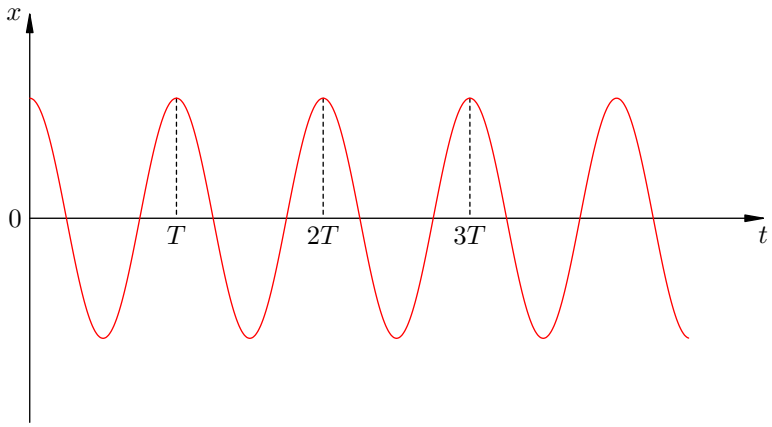


Рис. 1.57. Закон косинуса

Допустим теперь, что маятник не отклоняли, но ударом сообщили ему начальную скорость из положения равновесия. В этом случае $x_0 = 0$, так что можно положить $\alpha = -\pi/2$. Получаем *закон синуса*:

$$x = A\sin\omega t.$$

График колебаний представлен на рис. 1.58.

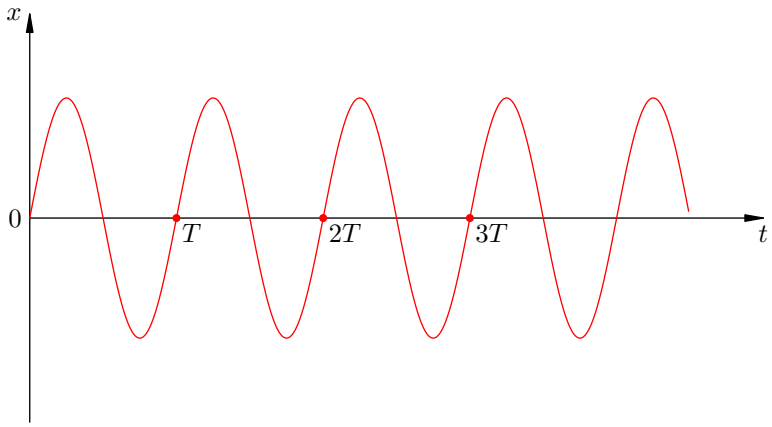


Рис. 1.58. Закон синуса

1.18.2 Уравнение гармонических колебаний

Вернёмся к общему гармоническому закону (1.82). Дифференцируем это равенство:

$$v_x = \dot{x} = -A\omega\sin(\omega t + \alpha). \tag{1.85}$$

Теперь дифференцируем полученное равенство (1.85):

$$a_x = \ddot{x} = -A\omega^2\cos(\omega t + \alpha). \tag{1.86}$$

Давайте сопоставим выражение (1.82) для координаты и выражение (1.86) для проекции ускорения. Мы видим, что проекция ускорения отличается от координаты лишь множителем $-\omega^2$:

$$a_x = -\omega^2 x. \tag{1.87}$$

Это соотношение называется *уравнением гармонических колебаний*. Его можно переписать и в таком виде:

$$\ddot{x} + \omega^2 x = 0. \tag{1.88}$$

С математической точки зрения уравнение (1.88) является *дифференциальным уравнением*. Решениями дифференциальных уравнений служат функции (а не числа, как в обычной алгебре). Так вот, можно доказать, что:

- решением уравнения (1.88) является всякая функция вида (1.82) с произвольными A и α;
- никакая другая функция решением данного уравнения не является.

Иными словами, соотношения (1.87), (1.88) описывают гармонические колебания с циклической частотой ω и только их. Две константы A и α определяются из начальных условий — по начальным значениям координаты и скорости.

1.18.3 Пружинный маятник

Пружинный маятник — это закреплённый на пружине груз, способный совершать колебания в горизонтальном или вертикальном направлении.

Найдём период малых горизонтальных колебаний пружинного маятника (рис. 1.59). Колебания будут *малыми*, если величина деформации пружины много меньше её размеров. При малых деформациях мы можем пользоваться законом Гука. Это приведёт к тому, что колебания окажутся гармоническими.

Трением пренебрегаем. Груз имеет массу m, жёсткость пружины равна k.

Координате $x = 0$ отвечает положение равновесия, в котором пружина не деформирована. Следовательно, величина деформации пружины равна модулю координаты груза.

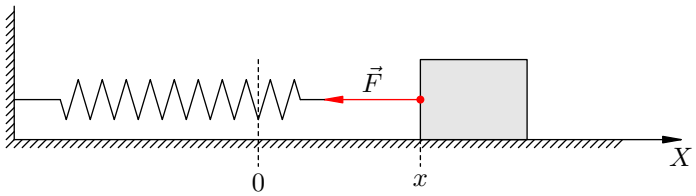


Рис. 1.59. Пружинный маятник

В горизонтальном направлении на груз действует только сила упругости $\vec{F}$ со стороны пружины. Второй закон Ньютона для груза в проекции на ось X имеет вид:

$$ma_x = F_x\,. \tag{1.89}$$

Если $x > 0$ (груз смещён вправо, как на рисунке), то сила упругости направлена в противоположную сторону, и $F_x < 0$. Наоборот, если $x < 0$, то $F_x > 0$. Знаки x и F_x всё время противоположны, поэтому закон Гука можно записать так:

$$F_x = -kx.$$

Тогда соотношение (1.89) принимает вид:

$$ma_x = -kx$$

или

$$a_x = -\frac{k}{m}\,x.$$

Мы получили уравнение гармонических колебаний вида (1.87), в котором

$$\omega^2 = \frac{k}{m}\,.$$

Циклическая частота колебаний пружинного маятника, таким образом, равна:

$$\omega = \sqrt{\frac{k}{m}}\,. \tag{1.90}$$

Отсюда и из соотношения $T = 2\pi/\omega$ находим период горизонтальных колебаний пружинного маятника:

$$T = 2\pi\sqrt{\frac{m}{k}}\,. \tag{1.91}$$

Если подвесить груз на пружине, то получится пружинный маятник, совершающий колебания в вертикальном направлении. Можно показать, что и в этом случае для периода колебаний справедлива формула (1.91).

1.18.4 Математический маятник

Математический маятник — это небольшое тело, подвешенное на невесомой нерастяжимой нити (рис. 1.60). Математический маятник может совершать колебания в вертикальной плоскости в поле силы тяжести.

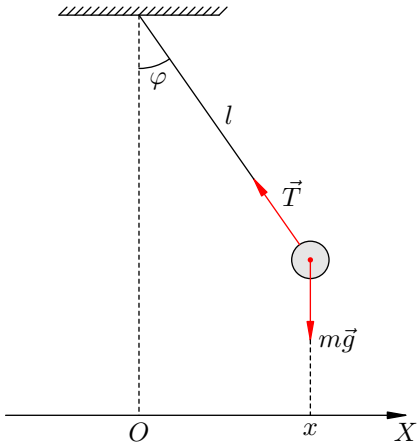


Рис. 1.60. Математический маятник

Найдём период малых колебаний математического маятника. Длина нити равна l. Сопротивлением воздуха пренебрегаем.

Запишем для маятника второй закон Ньютона:

$$m\vec{a} = m\vec{g} + \vec{T},$$

и спроектируем его на ось X:

$$ma_x = T_x\,.$$

Если маятник занимает положение как на рисунке (т. е. $x > 0$), то:

$$T_x = -T\sin\varphi = -T\,\frac{x}{l}\,.$$

Если же маятник находится по другую сторону от положения равновесия (т. е. $x < 0$), то:

$$T_x = T\sin\varphi = -T\,\frac{x}{l}\,.$$

Итак, при любом положении маятника имеем:

$$ma_x = -T\,\frac{x}{l}\,. \tag{1.92}$$

Когда маятник покоится в положении равновесия, выполнено равенство $T = mg$. При малых колебаниях, когда отклонения маятника от положения равновесия малы (по сравнению с длиной нити), выполнено приближённое равенство $T \approx mg$. Воспользуемся им в формуле (1.92):

$$ma_x = -mg\,\frac{x}{l}\,,$$

или

$$a_x = -\frac{g}{l}\,x.$$

Это — уравнение гармонических колебаний вида (1.87), в котором

$$\omega^2 = \frac{g}{l}\,.$$

Следовательно, циклическая частота колебаний математического маятника равна:

$$\omega = \sqrt{\frac{g}{l}}\,. \tag{1.93}$$

Отсюда период колебаний математического маятника:

$$T = 2\pi\sqrt{\frac{l}{g}}\,. \tag{1.94}$$

Обратите внимание, что в формулу (1.94) не входит масса груза. В отличие от пружинного маятника, период колебаний математического маятника не зависит от его массы.

1.18.5 Свободные и вынужденные колебания

Говорят, что система совершает *свободные колебания*, если она однократно выведена из положения равновесия и в дальнейшем предоставлена сама себе. Никаких периодических внешних воздействий система при этом не испытывает, и никаких внутренних источников энергии, поддерживающих колебания, в системе нет.

Рассмотренные выше колебания пружинного и математического маятников являются примерами свободных колебаний.

Частота, с которой совершаются свободные колебания, называется *собственной частотой* колебательной системы. Так, формулы (1.90) и (1.93) дают собственные (циклические) частоты колебаний пружинного и математического маятников.

В идеализированной ситуации при отсутствии трения свободные колебания являются *незатухающими*, т. е. имеют постоянную амплитуду и длятся неограниченно долго. В реальных колебательных системах всегда присутствует трение, поэтому свободные колебания постепенно затухают (рис. 1.61).

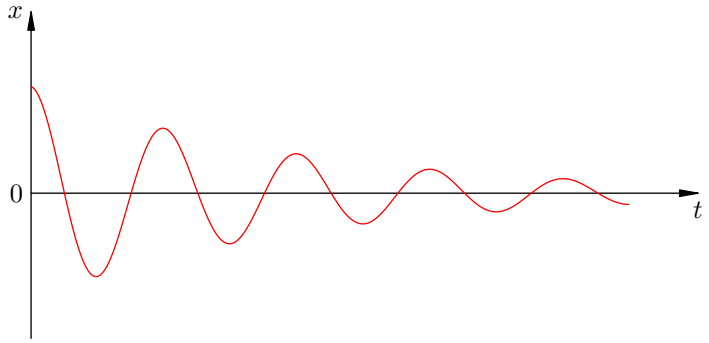


Рис. 1.61. Затухающие колебания

Вынужденные колебания — это колебания, совершаемые системой под воздействием внешней силы $F(t)$, периодически изменяющейся во времени (так называемой *вынуждающей силы*).

Предположим, что собственная частота колебаний системы равна ω_0, а вынуждающая сила зависит от времени по гармоническому закону:

$$F(t) = F_0 \cos \omega t\,.$$

В течение некоторого времени происходит установление вынужденных колебаний: система совершает сложное движение, которое является наложением выужденных и свободных колебаний. Свободные колебания постепенно затухают, и в установившемся режиме система совершает

вынужденные колебания, которые также оказываются гармоническими. *Частота установившихся вынужденных колебаний совпадает с частотой ω вынуждающей силы* (внешняя сила как бы навязывает системе свою частоту).

Амплитуда установившихся вынужденных колебаний зависит от частоты вынуждающей силы. График этой зависимости показан на рис. 1.62.

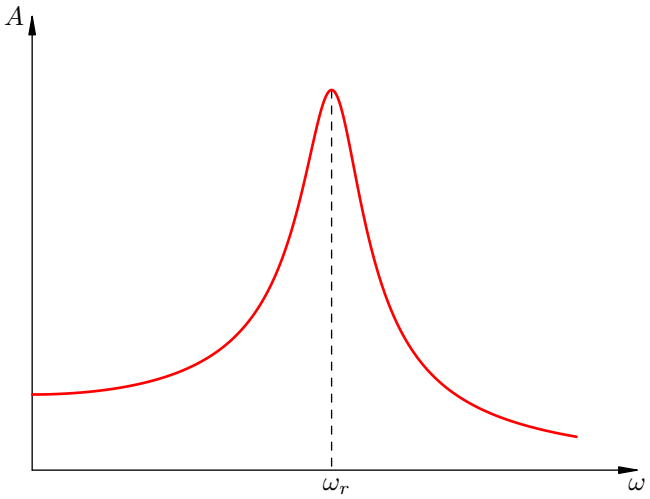


Рис. 1.62. Резонанс

Мы видим, что вблизи частоты $\omega = \omega_r$ наступает *резонанс* — явление возрастания амплитуды вынужденных колебаний. Резонансная частота приближённо равна собственной частоте колебаний системы: $\omega_r \approx \omega_0$, и это равенство выполняется тем точнее, чем меньше трение в системе. При отсутствии трения резонансная частота совпадает с собственной частотой колебаний, $\omega_r = \omega_0$, а амплитуда колебаний возрастает до бесконечности при $\omega \to \omega_0$.

1.19 Механические волны

Механические волны — это процесс распространения в пространстве колебаний частиц упругой среды (твёрдой, жидкой или газообразной).

Наличие у среды упругих свойств является необходимым условием распространения волн: деформация, возникающая в каком-либо месте, благодаря взаимодействию соседних частиц последовательно передаётся от одной точки среды к другой. Различным типам деформаций будут соответствовать разные типы волн.

1.19.1 Продольные и поперечные волны

Волна называется *продольной*, если частицы среды колеблются параллельно направлению распространения волны. Продольная волна состоит из чередующихся деформаций растяжения и сжатия. На рис. 1.63 показана продольная волна, представляющая собой колебания плоских слоёв среды; направление, вдоль которого колеблются слои, совпадает с направлением распространения волны (т. е. перпендикулярно слоям).

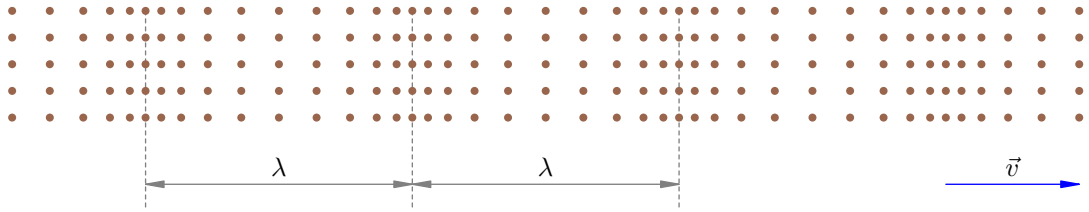


Рис. 1.63. Продольная волна

Волна называется *поперечной*, если частицы среды колеблются перпендикулярно направлению распространения волны. Поперечная волна вызывается деформациями сдвига одного слоя среды относительно другого. На рис. 1.64 каждый слой колеблется вдоль самого себя, а волна идёт перпендикулярно слоям.

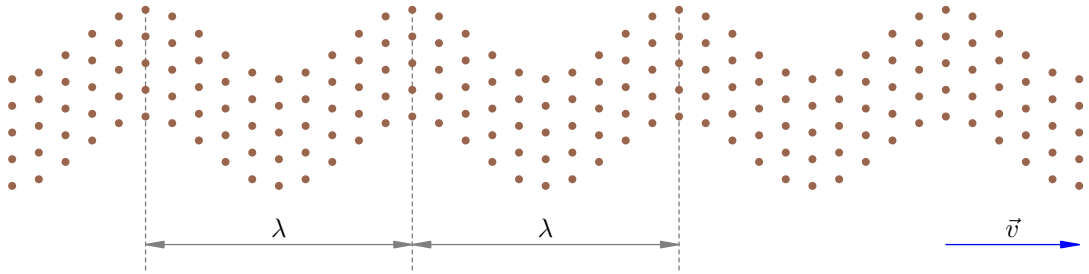


Рис. 1.64. Поперечная волна

Продольные волны могут распространяться в твёрдых телах, жидкостях и газах: во всех этих средах возникает упругая реакция на сжатие, в результате которой появятся бегущие друг за другом сжатия и разрежения среды.

Однако жидкости и газы, в отличие от твёрдых тел, не обладают упругостью по отношению к сдвигу слоёв. Поэтому поперечные волны могут распространяться в твёрдых телах, но не внутри жидкостей и газов[8].

[8]На поверхности жидкости могут существовать волны особого типа, похожие на поперечные — так называемые *поверхностные волны*. Они возникают под действием силы тяжести и силы поверхностного натяжения.

Важно отметить, что частицы среды при прохождении волны совершают колебания вблизи неизменных положений равновесия, т. е. в среднем остаются на своих местах. Волна, таким образом, осуществляет *перенос энергии, не сопровождающийся переносом вещества.*

Наиболее просты для изучения *гармонические волны.* Они вызываются внешним воздействием на среду, меняющимся по гармоническому закону. При распространении гармонической волны частицы среды совершают гармонические колебания с частотой, равной частоте внешнего воздействия. Гармоническими волнами мы в дальнейшем и ограничимся.

Рассмотрим процесс распространения волны более подробно.

Допустим, что некоторая частица среды (частица 1) начала совершать колебания с периодом T. Действуя на соседнюю частицу 2, она потянет её за собой. Частица 2, в свою очередь, потянет за собой частицу 3 и т. д. Так возникнет волна, в которой все частицы будут совершать колебания с периодом T.

Однако частицы имеют массу, т. е. обладают инертностью. На изменение их скорости требуется некоторое время. Следовательно, частица 2 в своём движении будет несколько отставать от частицы 1, частица 3 будет отставать от частицы 2 и т. д. Когда частица 1 спустя время T завершит первое колебание и начнёт второе, своё первое колебание начнёт частица $N+1$, находящаяся от частицы 1 на некотором расстоянии λ.

Итак, за время, равное периоду колебаний частиц, возмущение среды распространяется на расстояние λ. Это расстояние называется *длиной волны.* Колебания частицы $N+1$ будут идентичны колебаниям частицы 1, колебания следующей частицы $N+2$ будут идентичны колебаниям частицы 2 и т. д. Колебания как бы воспроизводят себя на расстоянии λ. Поэтому длину волны λ можно назвать *пространственным периодом* колебаний; наряду с временны́м периодом T она является важнейшей характеристикой волнового процесса.

В продольной волне длина волны равна расстоянию между соседними сжатиями или разрежениями (рис. 1.63). В поперечной — расстоянию между соседними горбами или впадинами (рис. 1.64). Вообще, длина волны равна расстоянию (вдоль направления распространения волны) между двумя ближайшими частицами среды, колеблющимися *одинаково* (т. е. с разностью фаз, равной 2π).

Скоростью распространения волны называется отношение длины волны к периоду колебаний частиц среды:

$$v = \frac{\lambda}{T}\,.$$

Частотой волны называется частота колебаний частиц:

$$\nu = \frac{1}{T}\,.$$

Отсюда получаем связь скорости волны, длины волны и частоты:

$$v = \lambda\nu. \tag{1.95}$$

1.19.2 Звук

Звуковыми волнами в широком смысле называются всякие волны, распространяющиеся в упругой среде. В узком смысле *звуком* называют звуковые волны в диапазоне частот от 16 Гц до 20 кГц, воспринимаемые человеческим ухом. Ниже этого диапазона лежит область *инфразвука*, выше — область *ультразвука.*

К основным характеристикам звука относятся *громкость* и *высота.*

Громкость звука определяется амплитудой колебаний давления в звуковой волне и измеряется в специальных единицах — *децибелах* (дБ). Так, громкость 0 дБ является порогом слышимости, 10 дБ — тиканье часов, 50 дБ — обычный разговор, 80 дБ — крик, 130 дБ — верхняя граница слышимости (так называемый *болевой порог*).

Тон — это звук, который издаёт тело, совершающее гармонические колебания (например, камертон или струна). Высота тона определяется частотой этих колебаний: чем выше частота, тем выше нам кажется звук. Так, натягивая струну, мы увеличиваем частоту её колебаний и, соответственно, высоту звука.

Скорость звука в разных средах различна: чем более упругой является среда, тем быстрее в ней распространяется звук. В жидкостях скорость звука больше, чем в газах, а в твёрдых телах — больше, чем в жидкостях.

Например, скорость звука в воздухе при 0 °C равна примерно 340 м/с (её удобно запомнить как «треть километра в секунду»)[9]. В воде звук распространяется со скоростью около 1500 м/с, а в стали — около 5000 м/с.

Заметим, что *частота* звука от данного источника во всех средах одна и та же: частицы среды совершают вынужденные колебания с частотой источника звука. Согласно формуле (1.95) заключаем тогда, что при переходе из одной среды в другую наряду со скоростью звука изменяется длина звуковой волны.

[9]Если хочешь найти расстояние до грозовых туч в километрах, посчитай, через сколько секунд после молнии придёт гром, и раздели полученное число на три.

Глава 2

Молекулярная физика и термодинамика

Молекулярная физика исходит из представлений о том, что все тела состоят из мельчайших частиц — молекул и атомов. Частицы движутся и взаимодействуют друг с другом, а физические свойства тел объясняются характером этого движения и взаимодействия.

Простейшей моделью молекулярной физики служит идеальный газ. В этой модели мы пренебрегаем размерами частиц, их внутренним строением и взаимодействием друг с другом. В результате удаётся получить простые законы, которые неплохо работают даже за пределами самой модели, приближённо описывая свойства реальных газов.

Термодинамика — это наука о взаимопревращениях тепла в работу и другие формы энергии. Термодинамика имеет дело с макроскопическими телами, состоящими из огромного числа частиц. При этом оказывается, что состояние макроскопического тела описывается сравнительно небольшим набором величин — например, давлением, объёмом и температурой.

Молекулярной физике и термодинамике посвящены следующие разделы пособия:

→ Основные положения МКТ
→ Газы, жидкости и твёрдые тела
→ Основные формулы молекулярной физики
→ Температура
→ Уравнение состояния идеального газа
→ Изопроцессы
→ Насыщенный пар
→ Внутренняя энергия
→ Количество теплоты
→ Фазовые переходы
→ Первый закон термодинамики
→ Тепловые машины
→ Второй закон термодинамики

2.1 Основные положения МКТ

Великому американскому физику Ричарду Фейнману, автору знаменитого курса «Фейнмановские лекции по физике», принадлежат замечательные слова:

> Если бы в результате какой-то мировой катастрофы все накопленные научные знания оказались бы уничтоженными и к грядущим поколениям живых существ перешла бы только одна фраза, то какое утверждение, составленное из наименьшего количества слов, принесло бы наибольшую информацию? Я считаю, что это — *атомная гипотеза* (можете называть ее не гипотезой, а фактом, но это ничего не меняет): *все тела состоят из атомов — маленьких телец, которые находятся в беспрерывном движении, притягиваются на небольшом расстоянии, но отталкиваются, если одно из них плотнее прижать к другому.* В одной этой фразе... содержится невероятное количество информации о мире, стоит лишь приложить к ней немного воображения и чуть соображения.

В этих словах заключена суть *молекулярно-кинетической теории* (МКТ) строения вещества. А именно, основными положениями МКТ являются следующие три утверждения.

1. Любое вещество состоит из мельчайших частиц — молекул и атомов. Они расположены в пространстве дискретно, то есть на некоторых расстояниях друг от друга.

2. Атомы или молекулы вещества находятся в состоянии беспорядочного движения[1], которое никогда не прекращается.

3. Атомы или молекулы вещества взаимодействуют друг с другом силами притяжения и отталкивания, которые зависят от расстояний между частицами.

Эти положения являются обобщением многочисленных наблюдений и экпериментальных фактов. Давайте рассмотрим подробнее эти положения и приведём их опытное обоснование.

2.1.1 Атомы и молекулы

Возьмём бумажный листок и начнём делить его на всё более и более мелкие части. На каждом ли шаге мы будем получать кусочки именно *бумаги*, или на каком-то этапе появится нечто новое?

Первое положение МКТ говорит нам о том, что вещество не является делимым до бесконечности. Рано или поздно мы дойдём до «последнего рубежа» — мельчайших частиц данного вещества. Эти частицы — атомы и молекулы. Их также можно разделить на части, но тогда исходное вещество прекратит своё существование.

Атом — это наименьшая частица данного *химического элемента*, сохраняющая все его химические свойства. Химических элементов не так много — все они сведены в таблицу Менделеева.

Молекула — это наименьшая частица данного вещества (не являющегося химическим элементом), сохраняющая все его химические свойства. Молекула состоит из двух или более атомов одного или нескольких химических элементов.

Например, H_2O — это молекула воды, состоящая из двух атомов водорода и одного атома кислорода. Разделив её на атомы, мы перестанем иметь дело в веществом под названием «вода». Далее, разделив атомы H и O на составные части, мы получим набор протонов, нейтронов и электронов и тем самым потеряем информацию о том, что поначалу это были водород и кислород.

[1] Это движение называется *тепловым движением*.

Размер атома или молекулы (состоящей из небольшого числа атомов) составляет порядка 10^{-8} см. Это настолько малая величина, что атом невозможно разглядеть ни в какой оптический микроскоп.

Атомы и молекулы называются для краткости просто *частицами* вещества. Чем именно является частица — атомом или молекулой — в каждом конкретном случае установить нетрудно. Если речь идёт о химическом элементе, то частицей будет атом; если же рассматривается сложное вещество, то его частица — это молекула, состоящая из нескольких атомов.

Далее, первое положение МКТ утверждает, что частицы вещества не заполняют пространство непрерывно. Частицы расположены *дискретно*, то есть как бы в отдельных точках. Между частицами имеются промежутки, величина которых может меняться в некоторых пределах.

В пользу первого положения МКТ свидетельствует явление *теплового расширения* тел. А именно, при нагревании увеличиваются расстояния между частицами вещества, и размеры тела возрастают. При охлаждении, наоборот, расстояния между частицами уменьшаются, в результате чего тело сжимается.

Ярким подтверждением первого положения МКТ служит также *диффузия* — взаимное проникновение соприкасающихся веществ друг в друга.

Например, на рис. 2.1 показан[2] процесс диффузии в жидкости. Частицы растворимого вещества помещены в стакан с водой и расположены вначале в верхней левой части стакана. С течением времени частицы перемещаются (как говорят, *диффундируют*) из области высокой концентрации в область низкой концентрации. В конце концов концентрация частиц становится везде одинаковой — частицы равномерно распределяются по всему объёму жидкости.

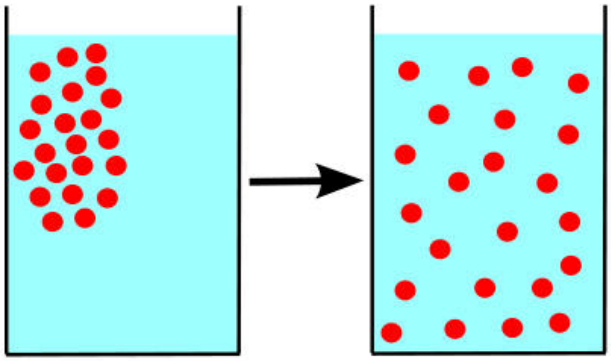


Рис. 2.1. Диффузия в жидкости

Как объяснить диффузию с точки зрения молекулярно-кинетической теории? Очень просто: частицы одного вещества проникают в промежутки между частицами другого вещества. Диффузия идёт тем быстрее, чем больше эти промежутки — поэтому легче всего смешиваются друг с другом газы (в которых расстояния между частицами много больше размеров самих частиц).

2.1.2 Тепловое движение атомов и молекул

Напомним ещё раз формулировку второго положения МКТ: *частицы вещества совершают беспорядочное движение (называемое также тепловым движением), которое никогда не прекращается.*

Опытным подтверждением второго положения МКТ служит опять-таки явление диффузии — ведь взаимное проникновение частиц возможно лишь при их беспрерывном движении!

[2]Изображение с сайта en.wikipedia.org.

Но наиболее ярким доказательством вечного хаотического движения частиц вещества является *броуновское движение*. Так называется непрерывное беспорядочное движение *броуновских частиц* — пылинок или крупинок (размерами 10^{-5} – 10^{-4} см), взвешенных в жидкости или газе.

Броуновское движение получило своё название в честь шотландского ботаника Роберта Броуна, увидевшего в микроскоп беспрерывную пляску взвешенных в воде частиц цветочной пыльцы. В доказательство того, что это движение совершается вечно, Броун нашёл кусок кварца с полостью, заполненной водой. Несмотря на то, что вода попала туда много миллионов лет назад, оказавшиеся там соринки продолжали своё движение, которое ничем не отличалось от того, что наблюдалось в других опытах.

Причина броуновского движения заключается в том, что взвешенная частица испытывает нескомпенсированные удары со стороны молекул жидкости (газа), причём в силу хаотичности движения молекул величина и направление результирующего воздействия абсолютно непредсказуемы. Поэтому броуновская частица описывает сложные зигзагообразные траектории (рис. 2.2)[3].

Рис. 2.2. Броуновское движение

Размеры броуновских частиц в 1000–10000 раз превышают размер атома. С одной стороны, броуновская частица достаточна мала и пока ещё «чувствует», что в разных направлениях по ней бьёт различное количество молекул; это различие в числе ударов приводит к заметным перемещениям броуновской частицы. С другой стороны, броуновские частицы достаточно велики для того, чтобы их можно было разглядеть в микроскоп.

Кстати говоря, броуновское движение может рассматриваться и как доказательство самого факта существования молекул, т. е. также может служить опытным обоснованием первого положения МКТ.

2.1.3 Взаимодействие частиц вещества

Третье положение МКТ говорит о взаимодействии частиц вещества: *атомы или молекулы взаимодействуют друг с другом силами притяжения и отталкивания, которые зависят от расстояний между частицами: при увеличении расстояний начинают преобладать силы притяжения, при уменьшении — силы отталкивания.*

О справедливости третьего положения МКТ свидетельствуют силы упругости, возникающие при деформациях тел. При растяжении тела увеличиваются расстояния между его частицами, и начинают преобладать силы притяжения частиц друг к другу. При сжатии тела расстояния между частицами уменьшаются, и в результате преобладают силы отталкивания. В обоих случаях упругая сила направлена в сторону, противоположную деформации.

[3]Изображение с сайта nv-magadan.narod.ru.

Другим подтверждением существования сил межмолекулярного взаимодействия служит наличие трёх агрегатных состояний вещества.

В газах молекулы удалены друг от друга на расстояния, значительно превышающие размеры самих молекул (в воздухе при нормальных условиях — примерно в 1000 раз). На таких расстояниях силы взаимодействия между молекулами практически отсутствуют, поэтому газы занимают весь предоставленный им объём и легко сжимаются.

В жидкостях промежутки между молекулами сравнимы с размерами молекул. Силы молекулярного притяжения весьма ощутимы и обеспечивают сохранение жидкостями объёма. Но для сохранения жидкостями ещё и формы эти силы недостаточно велики — жидкости, как и газы, принимают форму сосуда.

В твёрдых телах силы притяжения между частицами очень велики: твёрдые тела сохраняют не только объём, но и форму.

Переход вещества из одного агрегатного состояния в другое является результатом изменения величины сил взаимодействия между частицами вещества. Сами частицы остаются при этом неизменными.

2.2 Газы, жидкости и твёрдые тела

В предыдущем разделе, посвящённом основным положениям МКТ, мы вкратце описали молекулярное строение газов, жидкостей и твёрдых тел. Остановимся на этом вопросе подробнее.

2.2.1 Газы

Газы — самый простой объект для изучения в молекулярно-кинетической теории. Почему? Дело в том, что средние расстояния между частицами газов намного превышают размеры самих частиц. В промежутках между соударениями частицы газа проходят расстояния, на несколько порядков превышающие собственные размеры (рис. 2.3).

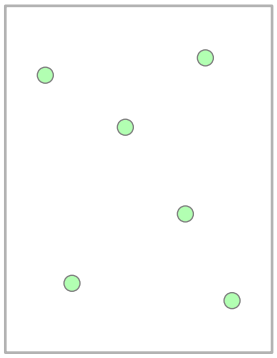


Рис. 2.3. Частицы газа

Например, в воздухе при нормальных условиях длина свободного пробега молекулы составляет примерно 10^{-5} см. Это на три порядка превышает средний размер молекулы (10^{-8} см).

При таких больших расстояниях между частицами силы межмолекулярного взаимодействия оказываются весьма незначительными. *Во многих ситуациях взаимодействием частиц газа на расстоянии можно пренебречь и учитывать лишь их соударения друг с другом.* Вот почему изучать газы гораздо проще, чем жидкости или твёрдые тела.

Не испытывая сильного притяжения со стороны других частиц, любая частица газа обладает полной свободой передвижения и может оказаться в любом месте сосуда. *Поэтому газы не имеют ни фиксированной формы, ни фиксированного объёма.* Мы можем поместить данную порцию газа в какой угодно сосуд, и газ неизменно займёт весь предоставленный ему объём. В зависимости от объёма сосуда будут изменяться лишь средние расстояния между частицами газа.

Способность газов менять свой объём находит широчайшее применение в технике. Функционирование тепловых двигателей и многих других устройств основано в конечном счёте на том, что газ при расширении перемещает поршень и совершает работу.

2.2.2 Твёрдые тела

Если сравнивать с газами, то твёрдые тела являются их полной противоположностью. Ни о какой свободе передвижения частицы твёрдых тел даже «не помышляют».

В твёрдых телах частицы расположены весьма близко друг к другу: расстояния между частицами порядка размера самих частиц. Силы взаимодействия между частицами твёрдого тела очень велики; расположение частиц в пространстве обладает периодической повторяемостью и образует так называемую *кристаллическую решётку.*

Например, на рис. 2.4 представлена пространственная модель кристаллической решётки кремния[4]. Шарики — это атомы кремния, а трубки между шариками изображают силы взаимодействия между атомами.

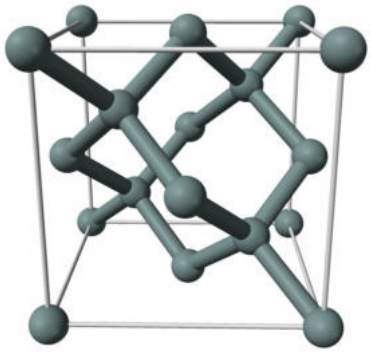


Рис. 2.4. Пространственная модель кристаллической решётки

Атомы кремния, как видите, расположены в пространстве периодическим образом, причём каждый атом сцеплен с четырьмя другими атомами. На рис. 2.5 мы видим плоское изображение той же самой кристаллической решётки (вдобавок показаны электроны, находящиеся на внешних оболочках атомов кремния).

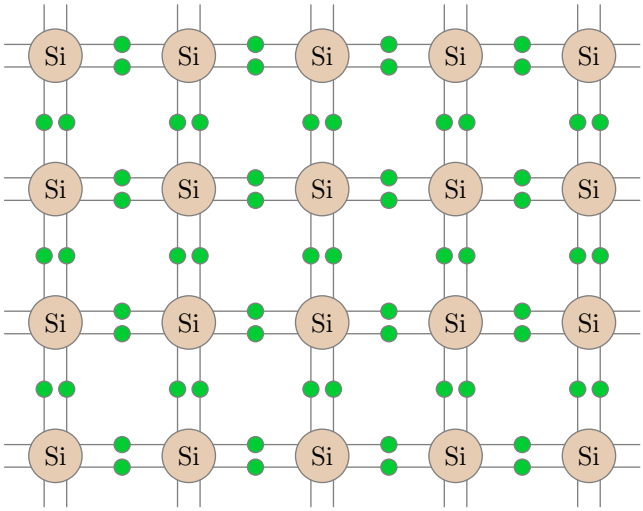


Рис. 2.5. Плоское изображение кристаллической решётки

Точки пространства, в которых находятся частицы твёрдого тела, называются *узлами* кристаллической решётки.

На самом деле частицы не покоятся в узлах кристаллической решётки, а совершают *тепловое движение* — колеблются относительно этих самых узлов. Таким образом, узел кристаллической решётки — это положение равновесия частицы, в небольшой окрестности которого частица постоянно находится. Чтобы покинуть область своего обитания (выскочить из узла), частице нужна очень большая энергия; произойти такое событие может лишь при чрезвычайно редком стечении обстоятельств. Имея столь жёсткую внутреннюю структуру, *твёрдое тело сохраняет неизменными свою форму и свой объём.*

[4]Автор картинки — Ben Mills.

2.2.3 Жидкости

По своим свойствам жидкости занимают промежуточное положение между газами и твёрдыми телами: *жидкости сохраняют объём, но не форму*. Мы можем говорить об одном литре воды, переливая её из сосуда в сосуд; но при этом вода принимает форму сосуда.

Как и в твёрдых телах, частицы жидкости упакованы весьма плотно и совершают колебания около некоторых положений равновесия. Попытка сжатия жидкости немедленно приводит к деформациям самих молекул и встречает мощное сопротивление: жидкости, в отличие от газов, практически не сжимаемы.

Однако, в отличие от твёрдых тел, частица жидкости не привязана навсегда к своему положению равновесия: спустя некоторое время она скачком меняет его на новое положение и колеблется в окружении новых частиц. Затем — новый скачок, новый период «осёдлой жизни», и так далее (рис. 2.6). Если частицы твёрдых тел можно сравнить с людьми, имеющими свой дом, то частицы жидкостей — это кочевники, постоянно меняющие места своих стоянок.

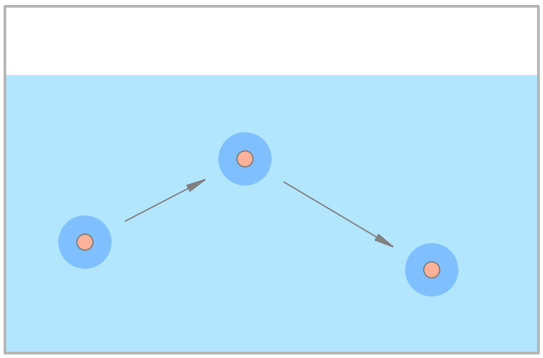


Рис. 2.6. «Осёдло-кочевая жизнь» молекулы жидкости

На рисунке мы видим молекулу жидкости, время от времени меняющую место своего расположения. «Ореол», окружающий молекулу, изображает область осёдлой жизни, внутри которой молекула совершает тепловые колебания. Теоретические расчёты показывают, что за время осёдлой жизни молекула совершает несколько десятков-сотен колебаний внутри области осёдлой жизни.

Силы притяжения между частицами жидкости достаточно велики для того, чтобы объём сохранялся фиксированным. Но ограниченность времени осёдлой жизни частиц придаёт жидкостям текучесть: жидкости не сохраняют форму.

Рассмотрим для примера воду в стакане. Скачки молекул из одного осёдлого положения в другое происходят равновероятно по всем направлениям. Но если стакан наклонить, то в определённых участках воды сила тяжести задаст преимущественное направление этих скачков, и форма воды в стакане поменяется. Таково в общих чертах объяснение текучести.

2.3 Основные формулы молекулярной физики

Введём основные величины молекулярной физики и соотношения между ними.

m — масса вещества, V — объём вещества, $\rho = \dfrac{m}{V}$ — плотность вещества (масса единицы объёма). Отсюда

$$m = \rho V.$$

N — число частиц вещества (атомов или молекул).

m_0 — масса частицы вещества. Тогда

$$m = m_0 N.$$

$n = \dfrac{N}{V}$ — концентрация вещества (число частиц в единице объёма), $[n] = \text{м}^{-3}$. Отсюда

$$N = nV.$$

Что получится, если m_0 умножить на n? Произведение массы частицы на число частиц в единице объёма даст массу единицы объёма, т. е. плотность. Формально:

$$m_0 n = m_0 \frac{N}{V} = \frac{m_0 N}{V} = \frac{m}{V} = \rho.$$

Итак,

$$\rho = m_0 n.$$

Массы и размеры частиц невообразимо малы по нашим обычным меркам. Например, масса атома водорода порядка 10^{-24} г, размер атома порядка 10^{-8} см. Из-за столь малых значений масс и размеров число частиц в макроскопическом теле огромно.

Оперировать столь грандиозными числами, как число частиц, неудобно. Поэтому для измерения *количества вещества* используют специальную единицу — *моль*.

Один моль — это количество вещества, в котором содержится столько же атомов или молекул, сколько атомов содержится в 12 граммах углерода. А в 12 граммах углерода содержится примерно $6{,}02 \cdot 10^{23}$ атомов. Стало быть, в одном моле вещества содержится $6{,}02 \cdot 10^{23}$ частиц. Это число называется *постоянной Авогадро*: $N_\text{A} = 6{,}02 \cdot 10^{23}\,\text{моль}^{-1}$.

Количество вещества обозначается ν. Это *число молей* данного вещества.

Что получится, если ν умножить на N_A? Число молей, умноженное на число частиц в моле, даст общее число частиц:

$$N = \nu N_\text{A}.$$

Масса одного моля вещества называется *молярной массой* этого вещества и обозначается μ ($[\mu] = $ кг/моль). Ясно, что

$$m = \mu\nu.$$

Как найти молярную массу химического элемента? Оказывается, для этого достаточно заглянуть в таблицу Менделеева! Нужно просто взять атомную массу A (число нуклонов) данного элемента — это будет его молярная масса, *выраженная в г/моль*. Например, для алюминия $A = 27$, поэтому молярная масса алюминия равна 27 г/моль или 0,027 кг/моль.

Почему так получается? Очень просто. Молярная масса углерода равна 12 г/моль по определению. В то же время ядро атома углерода содержит 12 нуклонов. Выходит, что каждый нуклон вносит в молярную массу 1 г/моль. Поэтому молярная масса химического элемента с атомной массой A оказывается равной A г/моль.

Молярная масса вещества, молекула которого состоит из нескольких атомов, получается простым суммированием молярных масс. Так, молярная масса углекислого газа CO_2 равна $12 + 16 \cdot 2 = 44$ г/моль $= 0{,}044$ кг/моль.

Будьте внимательны с молярными массами некоторых газов! Так, молярная масса газообразного водорода равна 2 г/моль, поскольку его молекула состоит из двух атомов (H_2). То же касается часто встречающихся в задачах азота и кислорода (N_2, O_2). Вместе с тем, наиболее частый персонаж задач — гелий (He) — является одноатомным газом и имеет молярную массу 4 г/моль, предписанную таблицей Менделеева.

Ещё раз предостережение: **при расчётах не забывайте переводить молярную массу в кг/моль**! Если ваш ответ отличается от правильного на три порядка, то вы наверняка сделали именно эту, очень распространённую ошибку :-)

Что получится, если m_0 умножить на N_A? Масса частицы, умноженная на число частиц в моле, даст массу моля, т. е. молярную массу:

$$\mu = m_0 N_\mathrm{A}.$$

2.4 Температура

Мы часто используем слово «температура» в повседневной речи. А что такое температура? В данном разделе мы объясним физический смысл этого понятия.

В молекулярной физике и термодинамике рассматриваются *макроскопические* тела, т. е. тела, состоящие из огромного числа частиц. Например, в стакане воды содержится порядка 10^{25} молекул. Такое грандиозное число с трудом поддаётся осмыслению.

2.4.1 Термодинамическая система

Термодинамической системой называется макроскопическое тело или система тел, которые могут взаимодействовать друг с другом и с окружающими телами. Стакан с водой — пример термодинамической системы.

Термодинамическая система состоит из столь большого числа частиц, что совершенно невозможно описывать её поведение путём рассмотрения движения каждой молекулы в отдельности. Однако именно грандиозность числа молекул делает ненужным такое описание.

Оказывается, что состояние термодинамической системы можно характеризовать небольшим числом *макроскопических параметров* — величин, относящимся к системе в целом, а не к отдельным атомам или молекулам. Такими макроскопическими параметрами являются давление, объём, температура, плотность, теплоёмкость, удельное сопротивление и др.

Состояние термодинамической системы, при котором все макроскопические параметры остаются неизменными с течением времени, называется *тепловым равновесием*. В состоянии теплового равновесия прекращаются все макроскопические процессы: диффузия, теплопередача, фазовые переходы, химические реакции и т. д.[5]

Термодинамическая система называется *изолированной*, если она не может обмениваться энергией с окружающими телами. Чай в термосе — типичный пример изолированной системы.

2.4.2 Тепловое равновесие

Фундаментальный постулат, вытекающий из многочисленных опытных данных, гласит: *каково бы ни было начальное состояние тел изолированной системы, со временем в ней устанавливается тепловое равновесие*. Таким образом, тепловое равновесие — это состояние, в которое любая система, изолированная от окружающей среды, самопроизвольно переходит через достаточно большой промежуток времени.

Температура как раз и является величиной, характеризующей состояние теплового равновесия термодинамической системы.

Температура — это макроскопический параметр, значения которого одинаковы для всех частей термодинамической системы, находящейся в состоянии теплового равновесия. Попросту говоря, температура — это то, что является одинаковым для любых двух тел, которые находятся в тепловом равновесии друг с другом. При тепловом контакте тел с одинаковыми температурами между ними не будет происходить обмен энергией (теплообмен).

В общем же случае при установлении между телами теплового контакта теплообмен начнётся. Говорят, что тело, которое отдаёт энергию, имеет *более высокую температуру*, а тело, которое получает энергию — *более низкую температуру*. Температура, таким образом, указывает *направление теплообмена* между телами. В процессе теплообмена температура первого тела начнёт уменьшаться, температура второго тела — увеличиваться; при выравнивании температур теплообмен прекратится — наступит тепловое равновесие.

[5] Следует отметить, что тепловое равновесие является *динамическим равновесием*. Так, при тепловом равновесии жидкости и её насыщенного пара весьма интенсивно идут взаимные превращения жидкости и пара. Но это — процессы молекулярного масштаба, они происходят с одинаковыми скоростями и компенсируют друг друга. На макроскопическом уровне количество жидкости и пара со временем не меняется.

Особенность температуры заключается в том, что она *не аддитивна*: температура тела не равна сумме температур его частей. Этим температура отличается от таких физических величин, как масса, длина или объём. И по этой причине температуру нельзя измерить путём сравнения с эталоном.

Измеряют температуру с помощью *термометра*.

Для создания термометра выбирают какое-либо вещество (*термометрическое вещество*), какую-либо характеристику этого вещества (*термометрическую величину*), и используют зависимость термометрической величины от температуры. При этом выбор термометрического вещества и термометрической величины может быть весьма произвольным.

Так, в бытовых *жидкостных* термометрах термометрическим веществом является ртуть (или спирт), а термометрической величиной — длина столбика жидкости. Здесь используется линейная зависимость объёма жидкости от температуры.

В *идеально-газовых* термометрах используется линейная зависимость давления разреженного газа (близкого по своим свойствам к идеальному) от температуры.

Действие *электрических* термометров (*термометров сопротивления*) основано на температурной зависимости сопротивления чистых металлов, сплавов и полупроводников.

В процессе измерения температуры термометр приводится в тепловой контакт[6] с телом, температура которого определяется. Показания термометра после наступления теплового равновесия — это и есть температура тела. При этом термометр показывает *свою* температуру!

2.4.3 Температурная шкала. Абсолютная температура

При установлении единицы температуры чаще всего поступают следующим образом. Берут две температуры (так называемые *реперные точки*) — температуру таяния льда и температуру кипения воды при нормальном атмосферном давлении. Первой температуре приписывают значение 0, второй — значение 100, а интервал между ними делят на 100 равных частей. Каждую из частей называют *градусом* (обозначают °C), а полученную таким образом температурную шкалу — *шкалой Цельсия*.

При измерениях по шкале Цельсия с помощью жидкостных термометров возникает одна трудность: разные жидкости при изменении температуры изменяют свой объём *по-разному*. Поэтому два термометра с различными жидкостями, приведённые в тепловой контакт с одним и тем же телом, могут показать разные температуры. От данного недостатка свободны идеально-газовые термометры — зависимость давления разреженного газа от температуры не зависит от вещества самого газа.

Кроме того, для температурной шкалы идеально-газового термометра существует естественное начало отсчёта (исчезает произвол выбора реперной точки!): *это та предельно низкая температура, при которой давление идеального газа постоянного объёма обращается в нуль*. Эта температура называется *абсолютным нулём температур*.

Температурная шкала, началом отсчёта которой является абсолютный нуль, а единицей температуры — градус Цельсия, называется *абсолютной температурной шкалой*.

Температура, измеряемая по абсолютной шкале, называется *абсолютной температурой* и обозначается буквой T. Единица абсолютной температуры называется *кельвином* (K).

Абсолютному нулю ($T = 0$) соответствует температура $t = -273{,}15\,^{\circ}\mathrm{C}$. Поэтому связь абсолютной температуры и температуры по шкале Цельсия даётся формулой:

$$T = t + 273{,}15.$$

В задачах достаточно использовать формулу $T = t + 273$.

[6] В области температур выше 1000 °C (раскалённые газы, расплавленные металлы) используются бесконтактные высокотемпературные термометры — *пирометры*. Их действие основано на измерении интенсивности теплового излучения в оптическом диапазоне.

2.5 Уравнение состояния идеального газа

Из трёх агрегатных состояний вещества наиболее простым для изучения является газообразное. В достаточно разреженных газах расстояния между молекулами намного больше размеров самих молекул (тогда как в жидкостях и твёрдых телах молекулы «упакованы» весьма плотно). Поэтому силы взаимодействия между молекулами таких газов очень малы.

Для описания разреженных газов в физике используется *модель идеального газа*. В рамках этой модели делаются следующие допущения.

1. Пренебрегаем размерами молекул. Иными словами, молекулы газа считаются материальными точками.

2. Пренебрегаем взаимодействием молекул на расстоянии.

3. Соударения молекул друг с другом и со стенками сосуда считаем абсолютно упругими.

Таким образом, *идеальный газ — это газ, частицы которого являются не взаимодействующими на расстоянии материальными точками и испытывают абсолютно упругие соударения друг с другом и со стенками сосуда.*

2.5.1 Средняя кинетическая энергия частиц газа

Оказывается, что ключевую роль в описании идеального газа играет средняя кинетическая энергия его частиц.

Частицы газа двигаются с разными скоростями. Пусть в газе содержится N частиц, скорости которых равны $v_1, v_2, \ldots, v_N$. Масса каждой частицы равна m_0. Кинетические энергии частиц:

$$E_1 = \frac{m_0 v_1^2}{2}\,, \quad E_2 = \frac{m_0 v_2^2}{2}\,, \quad \ldots, \quad E_N = \frac{m_0 v_N^2}{2}\,.$$

Средняя кинетическая энергия E частиц газа — это среднее арифметическое их кинетических энергий:

$$E = \frac{E_1 + E_2 + \ldots + E_N}{N} = \frac{1}{N}\left(\frac{m_0 v_1^2}{2} + \frac{m_0 v_2^2}{2} + \ldots + \frac{m_0 v_N^2}{2}\right) = \frac{m_0}{2}\,\frac{v_1^2 + v_2^2 + \ldots + v_N^2}{N}\,.$$

Последний множитель — это *средний квадрат скорости*, обозначаемый просто v^2:

$$v^2 = \frac{v_1^2 + v_2^2 + \ldots + v_N^2}{N}\,.$$

Тогда формула для средней кинетической энергии приобретает привычный вид:

$$E = \frac{m_0 v^2}{2}\,. \tag{2.1}$$

Корень из среднего квадрата скорости называется *средней квадратической скоростью*:

$$v = \sqrt{\frac{v_1^2 + v_2^2 + \ldots + v_N^2}{N}}\,.$$

2.5.2 Основное уравнение МКТ идеального газа

Связь между давлением газа и средней кинетической энергией его частиц называется *основным уравнением молекулярно-кинетической теории идеального газа*. Эта связь выводится из законов механики и имеет вид:

$$p = \frac{2}{3}nE,$$

где n — концентрация газа (число частиц в единице объёма). С учётом (2.1) имеем также:

$$p = \frac{1}{3}m_0 n v^2.$$

Что такое $m_0 n$? Произведение массы частицы на число частиц в единице объёма даёт массу единицы объёма, то есть плотность: $m_0 n = \rho$. Получаем третью разновидность основного уравнения:

$$p = \frac{1}{3}\rho v^2.$$

2.5.3 Энергия частиц и температура газа

Можно показать, что *при установлении теплового равновесия между двумя газами выравниваются средние кинетические энергии их частиц*. Но мы знаем, что при этом становятся равны и температуры газов. Следовательно, *температура газа — это мера средней кинетической энергии его частиц*.

Собственно, ничто не мешает попросту отождествить эти величины и сказать, что температура газа — это средняя кинетическая энергия его молекул. В продвинутых курсах теоретической физики так и поступают. Определённая таким образом температура измеряется в энергетических единицах — джоулях.

Но для практических задач удобнее иметь дело с привычными кельвинами. Связь средней кинетической энергии частиц и абсолютной температуры газа даётся формулой:

$$E = \frac{3}{2}kT, \tag{2.2}$$

где $k = 1{,}38 \cdot 10^{-23}$ Дж/К — *постоянная Больцмана*.

Из данной формулы можно получить выражение для средней квадратической скорости частиц. Подставим (2.1) в (2.2):

$$\frac{m_0 v^2}{2} = \frac{3}{2}kT,$$

откуда

$$v = \sqrt{\frac{3kT}{m_0}}\,.$$

В эту формулу входит масса частицы m_0, которую ещё надо вычислить. Но можно получить более удобный вариант формулы, домножив числитель и знаменатель подкоренного выражения на число Авогадро N_A:

$$v = \sqrt{\frac{3kN_\text{A}T}{m_0 N_\text{A}}}\,.$$

В знаменателе имеем: $m_0 N_\text{A} = \mu$ — молярная масса газа. В числителе стоит произведение двух констант, которое также является константой:

$$R = kN_\text{A} = 1{,}38 \cdot 10^{-23}\,\frac{\text{Дж}}{\text{К}} \cdot 6{,}02 \cdot 10^{23}\,\text{моль}^{-1} = 8{,}31\,\frac{\text{Дж}}{\text{моль} \cdot \text{К}}\,.$$

Константа R называется *универсальной газовой постоянной*.

Теперь формула для средней квадратической скорости приобретает вид:

$$v = \sqrt{\frac{3RT}{\mu}}\,.$$

Такое выражение гораздо более удобно для практических вычислений.

2.5.4 Уравнение Менделеева — Клапейрона

Берём формулу $p = \frac{2}{3}nE$ и подставляем в неё $E = \frac{3}{2}kT$. Получаем:

$$p = nkT.$$

Вспомним теперь, что $n = \frac{N}{V}$ и $N = \nu N_{\text{A}}$, где ν — число молей газа:

$$p = \frac{N}{V}kT = \frac{\nu N_{\text{A}}}{V}kT = \frac{\nu RT}{V}\,,$$

откуда

$$pV = \nu RT. \tag{2.3}$$

Соотношение (2.3) называется *уравнением Менделеева — Клапейрона*. Оно даёт взаимосвязь трёх важнейших макроскопических параметров, описывающих состояние идеального газа — давления, объёма и температуры. Поэтому уравнение Менделеева — Клапейрона называется ещё *уравнением состояния идеального газа*.

Учитывая, что $\nu = \frac{m}{\mu}$, где m — масса газа, получим другую форму уравнения Менделеева — Клапейрона:

$$pV = \frac{m}{\mu}RT. \tag{2.4}$$

Есть ещё один полезный вариант этого уравнения. Поделим обе части на V:

$$p = \frac{m}{V\mu}RT.$$

Но $\frac{m}{V} = \rho$ — плотность газа. Отсюда

$$p = \frac{\rho}{\mu}RT. \tag{2.5}$$

В задачах по физике активно используются все три формы записи (2.3)—(2.5).

2.6 Изопроцессы

На протяжении этого раздела мы будем придерживаться следующего предположения: *масса и химический состав газа остаются неизменными.* Иными словами, мы считаем, что:

- $m = \text{const}$, то есть нет утечки газа из сосуда или, наоборот, притока газа в сосуд;
- $\mu = \text{const}$, то есть частицы газа не испытывают каких-либо изменений (скажем, отсутствует диссоциация — распад молекул на атомы).

Эти два условия выполняются в очень многих физически интересных ситуациях (например, в простых моделях тепловых двигателей) и потому вполне заслуживают отдельного рассмотрения.

Если масса газа и его молярная масса фиксированы, то состояние газа определяется *тремя* макроскопическими параметрами: *давлением*, *объёмом* и *температурой*. Эти параметры связаны друг с другом уравнением состояния (уравнением Менделеева — Клапейрона).

2.6.1 Термодинамический процесс

Термодинамический процесс (или просто *процесс*) — это изменение состояния газа с течением времени. В ходе термодинамического процесса меняются значения макроскопических параметров — давления, объёма и температуры.

Особый интерес представляют *изопроцессы* — термодинамические процессы, в которых значение одного из макроскопических параметров остаётся неизменным. Поочерёдно фиксируя каждый из трёх параметров, мы получим три вида изопроцессов.

1. *Изотермический процесс* идёт при постоянной температуре газа: $T = \text{const}$.
2. *Изобарный процесс* идёт при постоянном давлении газа: $p = \text{const}$.
3. *Изохорный процесс* идёт при постоянном объёме газа: $V = \text{const}$.

Изопроцессы описываются очень простыми законами Бойля — Мариотта, Гей-Люссака и Шарля. Давайте перейдём к их изучению.

2.6.2 Изотермический процесс

При изотермическом процессе температура газа постоянна. В ходе процесса меняются только давление газа и его объём.

Установим связь между давлением p и объёмом V газа в изотермическом процессе. Пусть температура газа равна T. Рассмотрим два произвольных состояния газа: в одном из них значения макроскопических параметров равны p_1, V_1, T, а во втором — p_2, V_2, T. Эти значения связаны уравнением Менделеева — Клапейрона:

$$p_1 V_1 = \frac{m}{\mu} RT,$$
$$p_2 V_2 = \frac{m}{\mu} RT.$$

Как мы сказали с самого начала, масса газа m и его молярная масса μ предполагаются неизменными. Поэтому правые части выписанных уравнений равны. Следовательно, равны и левые части:

$$p_1 V_1 = p_2 V_2.$$

Поскольку два состояния газа были выбраны произвольно, мы можем заключить, что *в ходе изотермического процесса произведение давления газа на его объём остаётся постоянным*:

$$pV = \text{const}.$$

Данное утверждение называется *законом Бойля — Мариотта*.

Записав закон Бойля — Мариотта в виде

$$p = \frac{\text{const}}{V},$$

можно дать и такую формулировку: *в изотермическом процессе давление газа обратно пропорционально его объёму*. Если, например, при изотермическом расширении газа его объём увеличивается в три раза, то давление газа при этом в три раза уменьшается.

Как объяснить обратную зависимость давления от объёма с физической точки зрения? При постоянной температуре остаётся неизменной средняя кинетическая энергия молекул газа, то есть, попросту говоря, не меняется сила ударов молекул о стенки сосуда. При увеличении объёма концентрация молекул уменьшается, и соответственно уменьшается число ударов молекул в единицу времени на единицу площади стенки — давление газа падает. Наоборот, при уменьшении объёма концентрация молекул возрастает, их удары сыпятся чаще и давление газа увеличивается.

2.6.3 Графики изотермического процесса

Вообще, графики термодинамических процессов принято изображать в следующих системах координат:

- *pV-диаграмма*: ось абсцисс V, ось ординат p;
- *VT-диаграмма*: ось абсцисс T, ось ординат V;
- *pT-диаграмма*: ось абсцисс T, ось ординат p.

График изотермического процесса называется *изотермой*.

Изотерма на pV-диаграмме — это график обратно пропорциональной зависимости $p = \frac{\text{const}}{V}$. Такой график является гиперболой (вспомните алгебру — график функции $y = \frac{k}{x}$). Изотерма-гипербола изображена на рис. 2.7.

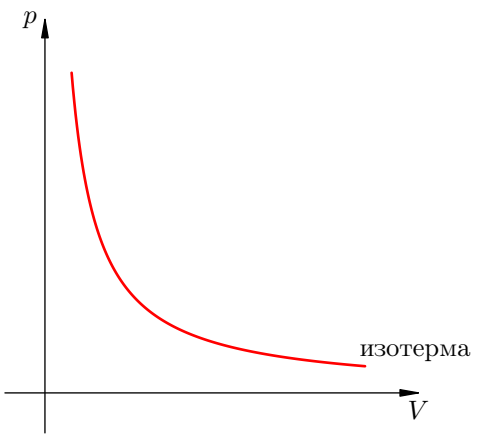


Рис. 2.7. Изотерма на pV-диаграмме

Каждая изотерма отвечает определённому фиксированному значению температуры. Оказывается, что *чем выше температура, тем выше лежит соответствующая изотерма на pV-диаграмме.*

В самом деле, рассмотрим два изотермических процесса, совершаемых одним и тем же газом (рис. 2.8). Первый процесс идёт при температуре T_1, второй — при температуре T_2.

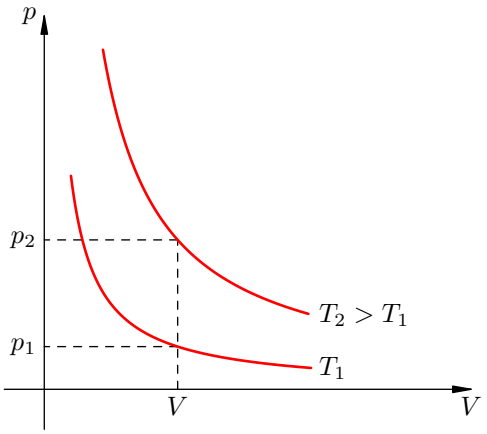


Рис. 2.8. Чем выше температура, тем выше изотерма

Фиксируем некоторое значение объёма V. На первой изотерме ему отвечает давление p_1, на второй — $p_2 > p_1$. Но при фиксированном объёме давление тем больше, чем выше температура (молекулы начинают сильнее бить по стенкам). Значит, $T_2 > T_1$.

В оставшихся двух системах координат изотерма выглядит очень просто: это прямая, перпендикулярная оси T (рис. 2.9):

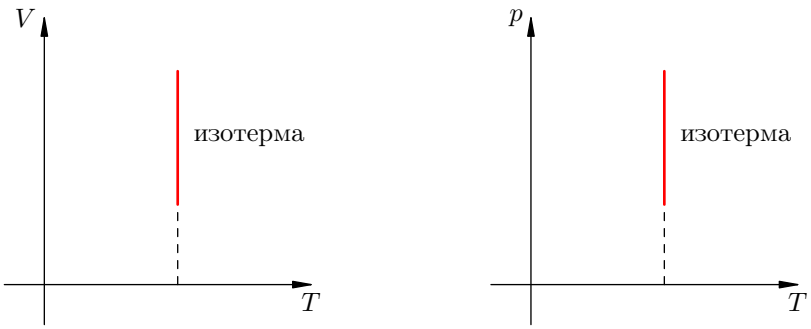


Рис. 2.9. Изотермы на VT- и pT-диаграммах

2.6.4 Изобарный процесс

Напомним ещё раз, что изобарный процесс — это процесс, проходящий при постоянном давлении. В ходе изобарного процесса меняются лишь объём газа и его температура.

Типичный пример изобарного процесса: газ находится под массивным поршнем, который может свободно перемещаться. Если масса поршня M и поперечное сечение поршня S, то давление газа всё время постоянно и равно

$$p = p_0 + \frac{Mg}{S},$$

где p_0 — атмосферное давление.

Пусть идеальный газ совершает изобарный процесс при давлении p. Снова рассмотрим два произвольных состояния газа; на этот раз значения макроскопических параметров будут равны p, V_1, T_1 и p, V_2, T_2.

Выпишем уравнения состояния:

$$pV_1 = \frac{m}{\mu} RT_1,$$
$$pV_2 = \frac{m}{\mu} RT_2.$$

Поделив их друг на друга, получим:

$$\frac{V_1}{V_2} = \frac{T_1}{T_2}.$$

В принципе, уже и этого могло бы быть достаточно, но мы пойдём немного дальше. Перепишем полученное соотношение так, чтобы в одной части фигурировали только параметры первого состояния, а в другой части — только параметры второго состояния (иными словами, «разнесём индексы» по разным частям):

$$\frac{V_1}{T_1} = \frac{V_2}{T_2}.$$

А отсюда теперь — ввиду произвольности выбора состояний! — получаем *закон Гей-Люссака*:

$$\frac{V}{T} = \text{const}.$$

Иными словами, *при постоянном давлении газа его объём прямо пропорционален температуре*:

$$V = \text{const} \cdot T.$$

Почему объём растёт с ростом температуры? При повышении температуры молекулы начинают бить сильнее и приподнимают поршень. При этом концентрация молекул падает, удары становятся реже, так что в итоге давление сохраняет прежнее значение.

2.6.5 Графики изобарного процесса

График изобарного процесса называется *изобарой*. На VT-диаграмме изобара $V = \text{const} \cdot T$ является прямой линией (рис. 2.10):

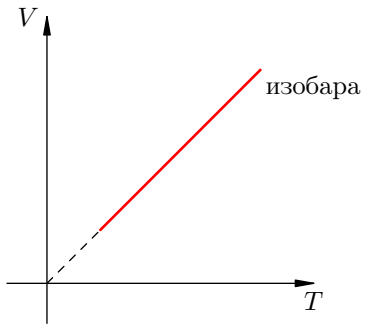


Рис. 2.10. Изобара на VT-диаграмме

Пунктирный участок графика означает, что в случае реального газа при достаточно низких температурах модель идеального газа (а вместе с ней и закон Гей-Люссака) перестаёт работать. В самом деле, при снижении температуры частицы газа двигаются всё медленнее, и силы

межмолекулярного взаимодействия оказывают всё более существенное влияние на их движение (аналогия: медленный мяч легче поймать, чем быстрый). Ну а при совсем уж низких температурах газы и вовсе превращаются в жидкости.

Разберёмся теперь, как меняется положение изобары при изменении давления. Оказывается, что *чем больше давление, тем ниже идёт изобара на VT-диаграмме.*

Чтобы убедиться в этом, рассмотрим две изобары с давлениями p_1 и p_2 (рис. 2.11):

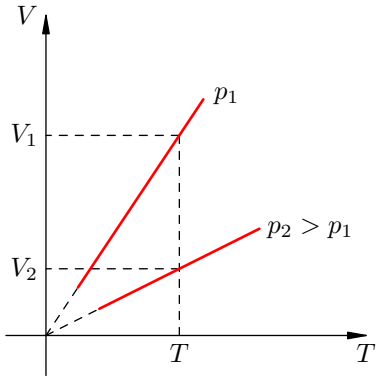


Рис. 2.11. Чем ниже изобара, тем больше давление

Зафиксируем некоторое значение температуры T. Мы видим, что $V_2 < V_1$. Но при фиксированной температуре объём тем меньше, чем больше давление (закон Бойля — Мариотта!). Стало быть, $p_2 > p_1$.

В оставшихся двух системах координат изобара является прямой линией, перпендикулярной оси p (рис. 2.12):

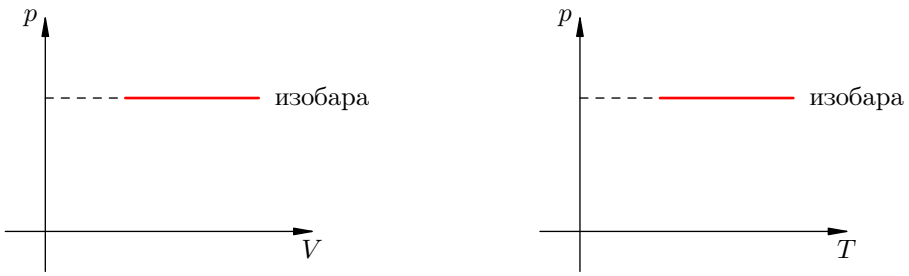


Рис. 2.12. Изобары на pV- и pT-диаграммах

2.6.6 Изохорный процесс

Изохорный процесс, напомним, — это процесс, проходящий при постоянном объёме. При изохорном процессе меняются только давление газа и его температура.

Изохорный процесс представить себе очень просто: это процесс, идущий в жёстком сосуде фиксированного объёма (или в цилиндре под поршнем, когда поршень закреплён).

Пусть идеальный газ совершает изохорный процесс в сосуде объёмом V. Опять-таки рассмотрим два произвольных состояния газа с параметрами p_1, V, T_1 и p_2, V, T_2. Имеем:

$$p_1 V = \frac{m}{\mu} R T_1,$$
$$p_2 V = \frac{m}{\mu} R T_2.$$

Делим эти уравнения друг на друга:

$$\frac{p_1}{p_2} = \frac{T_1}{T_2}\,.$$

Как и при выводе закона Гей-Люссака, «разносим» индексы в разные части:

$$\frac{p_1}{T_1} = \frac{p_2}{T_2}\,.$$

Ввиду произвольности выбора состояний мы приходим к *закону Шарля*:

$$\frac{p}{T} = \text{const}.$$

Иными словами, *при постоянном объёме газа его давление прямо пропорционально температуре*:

$$p = \text{const} \cdot T.$$

Увеличение давления газа фиксированного объёма при его нагревании — вещь совершенно очевидная с физической точки зрения. Вы сами легко это объясните.

2.6.7 Графики изохорного процесса

График изохорного процесса называется *изохорой*. На pT-диаграмме изохора $p = \text{const} \cdot T$ является прямой линией (рис. 2.13):

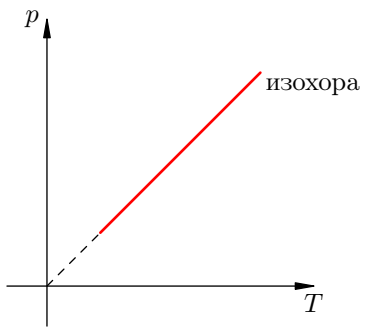


Рис. 2.13. Изохора на pT-диаграмме

Смысл пунктирного участка тот же: неадекватность модели идеального газа при низких температурах.

Далее, *чем больше объём, тем ниже идёт изохора на pT-диаграмме* (рис. 2.14):

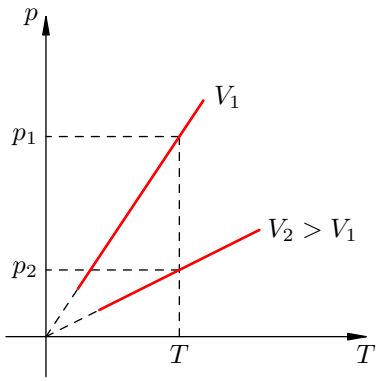


Рис. 2.14. Чем ниже изохора, тем больше объём

Доказательство аналогично предыдущему. Фиксируем температуру T и видим, что $p_2 < p_1$. Но при фиксированной температуре давление тем меньше, чем больше объём (снова закон Бойля — Мариотта). Стало быть, $V_2 > V_1$.

В оставшихся двух системах координат изохора является прямой линией, перпендикулярной оси V (рис. 2.15):

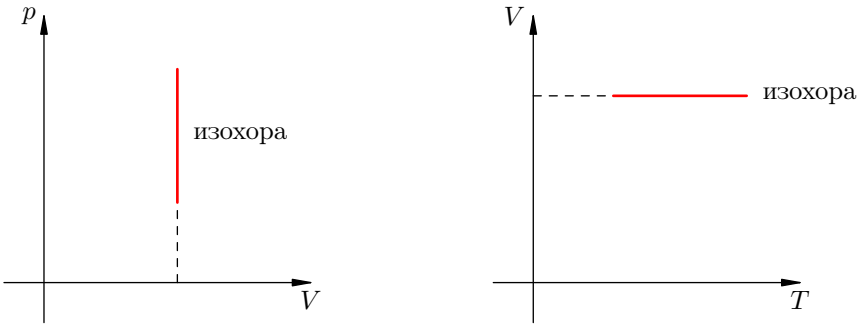


Рис. 2.15. Изохоры на pV- и VT-диаграммах

Законы Бойля — Мариотта, Гей-Люссака и Шарля называются также *газовыми законами*. Мы вывели газовые законы из уравнения Менделеева — Клапейрона. Но исторически всё было наоборот: газовые законы были установлены экспериментально, и намного раньше. Уравнение состояния появилось впоследствии как их обобщение.

2.7 Насыщенный пар

Если открытый стакан с водой оставить на долгое время, то в конце концов вода полностью улетучится. Точнее — испарится. Что такое испарение и почему оно происходит?

2.7.1 Испарение и конденсация

При данной температуре молекулы жидкости обладают разными скоростями. Скорости большинства молекул находятся вблизи некоторого среднего значения (характерного для этой температуры). Но попадаются молекулы, скорости которых значительно отличаются от средней как в меньшую, так и бóльшую сторону.

На рис. 2.16 изображён примерный график распределения молекул жидкости по скоростям. Голубым фоном показано то самое большинство молекул, скорости которых группируются около среднего значения. Красный «хвост» графика — это небольшое число «быстрых» молекул, скорости которых существенно превышают среднюю скорость основной массы молекул жидкости.

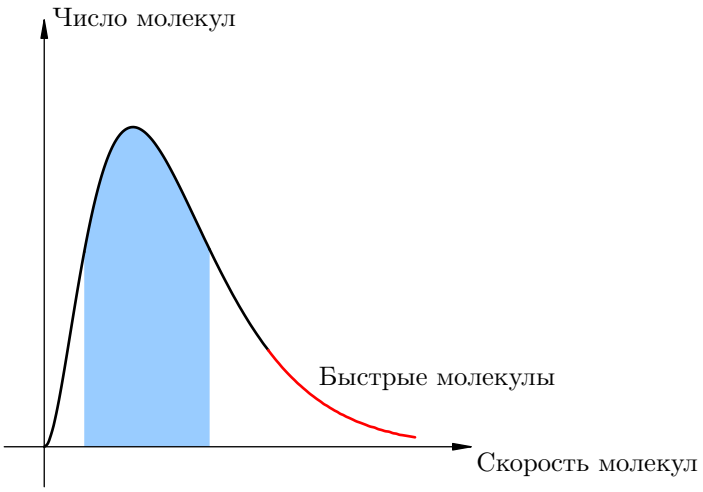


Рис. 2.16. Распределение молекул по скоростям

Когда такая весьма быстрая молекула окажется на свободной поверхности жидкости (т. е. на границе раздела жидкости и воздуха), кинетической энергии этой молекулы может хватить на то, чтобы преодолеть силы притяжения остальных молекул и вылететь из жидкости. Данный процесс и есть *испарение*, а молекулы, покинувшие жидкость, образуют *пар*.

Итак, *испарение — это процесс превращения жидкости в пар, происходящий на свободной поверхности жидкости*[7].

Может случиться, что через некоторое время молекула пара вернётся обратно в жидкость. *Процесс перехода молекул пара в жидкость называется конденсацией.* Конденсация пара — процесс, обратный испарению жидкости.

2.7.2 Динамическое равновесие

А что будет, если сосуд с жидкостью герметично закрыть? Плотность пара над поверхностью жидкости начнёт увеличиваться; частицы пара будут всё сильнее мешать другим молекулам жидкости вылетать наружу, и скорость испарения станет уменьшаться. Одновременно начнёт

[7]При особых условиях превращение жидкости в пар может происходить по всему объёму жидкости. Данный процесс вам хорошо известен — это *кипение*.

увеличиваться скорость конденсации, так как с возрастанием концентрации пара число молекул, возвращающихся в жидкость, будет становиться всё больше.

Наконец, в какой-то момент скорость конденсации окажется равна скорости испарения. Наступит *динамическое равновесие* между жидкостью и паром: за единицу времени из жидкости будет вылетать столько же молекул, сколько возвращается в неё из пара. Начиная с этого момента количество жидкости перестанет убывать, а количество пара — увеличиваться; пар достигнет «насыщения».

Насыщенный пар — это пар, который находится в состоянии динамического равновесия со своей жидкостью. Пар, не достигший состояния динамического равновесия с жидкостью, называется ненасыщенным.

Давление и плотность насыщенного пара обозначаются $p_н$ и $\rho_н$. Очевидно, $p_н$ и $\rho_н$ — это максимальные давление и плотность, которые может иметь пар при данной температуре. Иными словами, давление и плотность насыщенного пара всегда превышают давление и плотность ненасыщенного пара.

2.7.3 Свойства насыщенного пара

Оказывается, что состояние насыщенного пара (а ненасыщенного — тем более) можно приближённо описывать уравнением состояния идеального газа (уравнением Менделеева — Клапейрона). В частности, имеем приближённое соотношение между давлением насыщенного пара и его плотностью:

$$p_н = \frac{\rho_н}{\mu} RT. \tag{2.6}$$

Это весьма удивительный факт, подтверждаемый экспериментом. Ведь по своим свойствам насыщенный пар существенно отличается от идеального газа. Перечислим важнейшие из этих отличий.

1. *При неизменной температуре плотность насыщенного пара не зависит от его объёма.*

   Если, например, насыщенный пар изотермически сжимать, то его плотность в первый момент возрастёт, скорость конденсации превысит скорость испарения, и часть пара конденсируется в жидкость — до тех пор, пока вновь не наступит динамическое равновесие, в котором плотность пара вернётся к своему прежнему значению.

   Аналогично, при изотермическом расширении насыщенного пара его плотность в первый момент уменьшится (пар станет ненасыщенным), скорость испарения превысит скорость конденсации, и жидкость будет дополнительно испаряться до тех пор, пока опять не установится динамическое равновесие — т. е. пока пар снова не станет насыщенным с прежним значением плотности.

2. *Давление насыщенного пара не зависит от его объёма.*

   Это следует из того, что плотность насыщенного пара не зависит от объёма, а давление однозначно связано с плотностью уравнением (2.6).

   Как видим, *закон Бойля — Мариотта, справедливый для идеальных газов, для насыщенного пара не выполняется.* Это и не удивительно — ведь он получен из уравнения Менделеева — Клапейрона в предположении, что масса газа остаётся постоянной.

3. *При неизменном объёме плотность насыщенного пара растёт с повышением температуры и уменьшается с понижением температуры.*

   Действительно, при увеличении температуры возрастает скорость испарения жидкости. Динамическое равновесие в первый момент нарушается, и происходит дополнительное

испарение некоторой части жидкости. Пара будет прибавляться до тех пор, пока динамическое равновесие вновь не восстановится.

Точно так же при понижении температуры скорость испарения жидкости становится меньше, и часть пара конденсируется до тех пор, пока не восстановится динамическое равновесие — но уже с меньшим количеством пара.

Таким образом, при изохорном нагревании или охлаждении насыщенного пара его масса меняется, поэтому закон Шарля в данном случае не работает. Зависимость давления насыщенного пара от температуры уже не будет линейной функцией.

4. *Давление насыщенного пара растёт с температурой быстрее, чем по линейному закону.*

   В самом деле, с увеличением температуры возрастает плотность насыщенного пара, а согласно уравнению (2.6) давление пропорционально произведению плотности на температуру.

   Зависимость давления насыщенного пара от температуры является экспоненциальной (рис. 2.17). Она представлена участком 1–2 графика. Эту зависимость нельзя вывести из законов идеального газа.

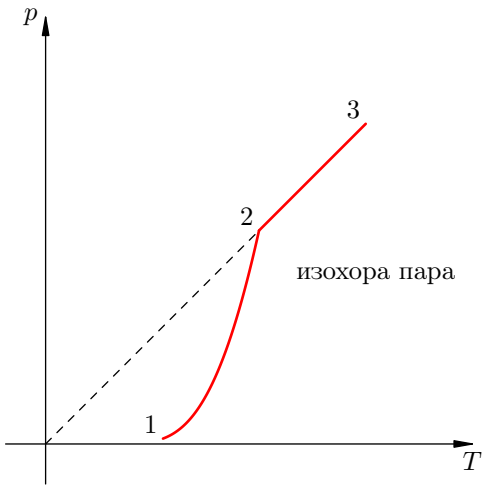


Рис. 2.17. Зависимость давления пара от температуры

В точке 2 вся жидкость испаряется; при дальнейшем повышении температуры пар становится ненасыщенным, и его давление растёт линейно по закону Шарля (участок 2–3).

Вспомним, что линейный рост давления идеального газа вызван увеличением интенсивности ударов молекул о стенки сосуда. В случае нагревания насыщенного пара молекулы начинают бить не только сильнее, но и чаще — ведь пара становится больше. Одновременным действием этих двух факторов и вызван экспоненциальный рост давления насыщенного пара.

2.7.4 Влажность воздуха

Воздух, содержащий водяной пар, называется *влажным*. Чем больше пара находится в воздухе, тем выше *влажность* воздуха.

Абсолютная влажность — это парциальное давление водяного пара, находящегося в воздухе (т. е. давление, которое водяной пар оказывал бы сам по себе, в отсутствие других газов). Иногда абсолютной влажностью называют также плотность водяного пара в воздухе.

Относительная влажность воздуха φ — это отношение парциального давления водяного пара в нём к давлению насыщенного водяного пара при той же температуре. Как правило, это отношение выражают в процентах:

$$\varphi = \frac{p}{p_{\text{н}}} \cdot 100\%.$$

Из уравнения Менделеева-Клапейрона (2.6) следует, что отношение давлений пара равно отношению плотностей. Так как само уравнение (2.6), напомним, описывает насыщенный пар лишь приближённо, мы имеем приближённое соотношение:

$$\varphi = \frac{\rho}{\rho_{\text{н}}} \cdot 100\%.$$

Одним из приборов, измеряющих влажность воздуха, является *психрометр*. Он включает в себя два термометра, резервуар одного из которых завёрнут в мокрую ткань. Чем ниже влажность, тем интенсивнее идёт испарение воды из ткани, тем сильнее охлаждается резервуар «мокрого» термометра, и тем больше разность его показаний и показаний сухого термометра. По этой разности с помощью специальной психрометрической таблицы определяют влажность воздуха.

2.8 Внутренняя энергия

Частицы любого тела — атомы или молекулы — совершают хаотическое непрекращающееся движение (так называемое *тепловое движение*). Поэтому каждая частица обладает некоторой кинетической энергией.

Кроме того, частицы вещества взаимодействуют друг с другом силами электрического притяжения и отталкивания, а также посредством ядерных сил. Стало быть, вся система частиц данного тела обладает ещё и потенциальной энергией.

Кинетическая энергия теплового движения частиц и потенциальная энергия их взаимодействия вместе образуют новый вид энергии, не сводящийся к механической энергии тела (т. е. кинетической энергии движения тела как целого и потенциальной энергии его взаимодействия с другими телами). Этот вид энергии называется внутренней энергией.

Внутренняя энергия тела — это суммарная кинетическая энергия теплового движения его частиц плюс потенциальная энергия их взаимодействия друг с другом.

Внутренняя энергия термодинамической системы — это сумма внутренних энергий тел, входящих в систему.

Таким образом, внутреннюю энергию тела образуют следующие слагаемые.

1. Кинетическая энергия непрерывного хаотического движения частиц тела.
2. Потенциальная энергия молекул (атомов), обусловленная силами межмолекулярного взаимодействия.
3. Энергия электронов в атомах.
4. Внутриядерная энергия.

В случае простейшей модели вещества — идеального газа — для внутренней энергии можно получить явную формулу.

2.8.1 Внутренняя энергия одноатомного идеального газа

Потенциальная энергия взаимодействия частиц идеального газа равна нулю (напомним, что в модели идеального газа мы пренебрегаем взаимодействием частиц на расстоянии). Поэтому внутренняя энергия одноатомного идеального газа сводится к суммарной кинетической энергии поступательного[8] движения его атомов. Эту энергию можно найти, умножив число атомов газа N на среднюю кинетическую энергию E одного атома:

$$U = NE = N \cdot \frac{3}{2}kT = \nu N_A \cdot \frac{3}{2}kT = \frac{3}{2}\nu RT,$$

или

$$U = \frac{3}{2}\frac{m}{\mu}RT.$$

Мы видим, что внутренняя энергия идеального газа (масса и химический состав которого неизменнны) является функцией только его температуры. У реального газа, жидкости или твёрдого тела внутренняя энергия будет зависеть ещё и от объёма — ведь при изменении объёма изменяется взаимное расположение частиц и, как следствие, потенциальная энергия их взаимодействия.

[8] У многоатомного газа приходится ещё учитывать вращение молекул и колебания атомов внутри молекул.

2.8.2 Функция состояния

Важнейшее свойство внутренней энергии заключается в том, что она является *функцией состояния* термодинамической системы. А именно, внутренняя энергия однозначно определяется набором макроскопических параметров, характеризующих систему, и не зависит от «предыстории» системы, т. е. от того, в каком состоянии система находилась прежде и каким конкретно образом она оказалась в данном состоянии.

Так, при переходе системы из одного состояния в другое изменение её внутренней энергии определяется лишь начальным и конечным состояниями системы и *не зависит* от пути перехода из начального состояния в конечное. Если система возвращается в исходное состояние, то изменение её внутренней энергии равно нулю.

Опыт показывает, что существует лишь два способа изменения внутренней энергии тела:

- совершение механической работы;
- теплопередача.

Попросту говоря, нагреть чайник можно только двумя принципиально разными способами: тереть его чем-нибудь или поставить на огонь :-) Рассмотрим эти способы подробнее.

2.8.3 Изменение внутренней энергии: совершение работы

Если работа совершается *над* телом, то внутренняя энергия тела возрастает.

Например, гвоздь после удара по нему молотком нагревается и немного деформируется. Но температура — это мера средней кинетической энергии частиц тела. Нагревание гвоздя свидетельствует об увеличении кинетической энергии его частиц: в самом деле, частицы разгоняются от удара молотком и от трения гвоздя о доску.

Деформация же есть не что иное, как смещение частиц друг относительно друга; гвоздь после удара испытывает деформацию сжатия, его частицы сближаются, между ними возрастают силы отталкивания, и это приводит к увеличению потенциальной энергии частиц гвоздя.

Итак, внутренняя энергия гвоздя увеличилась. Это явилось результатом совершения над ним работы — работу совершили молоток и сила трения о доску.

Если же работа совершается *самим* телом, то внутренняя энергия тела уменьшается.

Пусть, например, сжатый воздух в теплоизолированном сосуде под поршнем расширяется и поднимает некий груз, совершая тем самым работу[9]. В ходе такого процесса воздух будет охлаждаться — его молекулы, ударяя вдогонку по движущемуся поршню, отдают ему часть своей кинетической энергии. (Точно так же футболист, останавливая ногой быстро летящий мяч, делает ею движение *от* мяча и гасит его скорость.) Стало быть, внутренняя энергия воздуха уменьшается.

Воздух, таким образом, совершает работу за счёт своей внутренней энергии: поскольку сосуд теплоизолирован, нет притока энергии к воздуху от каких-либо внешних источников, и черпать энергию для совершения работы воздух может только из собственных запасов.

2.8.4 Изменение внутренней энергии: теплопередача

Теплопередача — это процесс перехода внутренней энергии от более горячего тела к более холодному, не связанный с совершением механической работы. Теплопередача может осуществляться либо при непосредственном контакте тел, либо через промежуточную среду (и даже через вакуум). Теплопередача называется ещё *теплообменом*.

[9]Процесс в теплоизолированном сосуде называется *адиабатным*. Мы изучим адиабатный процесс при рассмотрении первого закона термодинамики.

Различают три вида теплопередачи: теплопроводность, конвекция и тепловое излучение. Сейчас мы рассмотрим их более подробно.

2.8.5 Теплопроводность

Если железный стержень сунуть одним концом в огонь, то, как мы знаем, долго его в руке не продержишь. Попадая в область высокой температуры, атомы железа начинают колебаться интенсивнее (т. е. приобретают добавочную кинетическую энергию) и наносят более сильные удары по своим соседям.

Кинетическая энергия соседних атомов также возрастает, и теперь уже эти атомы сообщают дополнительную кинетическую энергию своим соседям. Так от участка к участку тепло постепенно распространяется по стержню — от помещённого в огонь конца до нашей руки. Это и есть теплопроводность (рис. 2.18)[10].

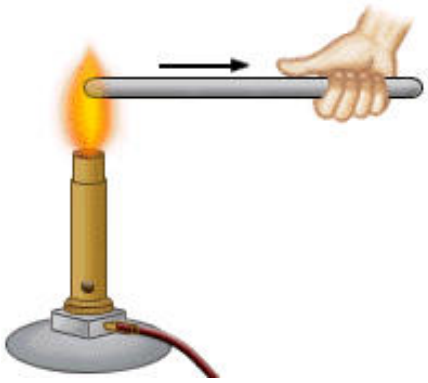


Рис. 2.18. Теплопроводность

Теплопроводность — это перенос внутренней энергии от более нагретых участков тела к менее нагретым за счёт теплового движения и взаимодействия частиц тела.

Теплопроводность разных веществ различна. Высокую теплопроводность имеют металлы: лучшими проводниками тепла являются серебро, медь и золото. Теплопроводность жидкостей гораздо меньше. Газы проводят тепло настолько плохо, что относятся уже к теплоизоляторам: молекулы газов из-за больших расстояний между ними слабо взаимодействуют друг с другом. Вот почему, например, в окнах делают двойные рамы: прослойка воздуха препятствует уходу тепла).

Плохими проводниками тепла являются поэтому пористые тела — такие, как кирпич, вата или мех. Они содержат в своих порах воздух. Недаром кирпичные дома считаются самыми тёплыми, а в мороз люди надевают меховые шубы и куртки с прослойкой пуха или синтепона.

Но если воздух так плохо проводит тепло, то почему тогда прогревается от батареи комната? Происходит это вследствие другого вида теплопередачи — конвекции.

2.8.6 Конвекция

Конвекция — это перенос внутренней энергии в жидкостях или газах в результате циркуляции потоков и перемешивания вещества.

Воздух вблизи батареи нагревается и расширяется. Действующая на *этот* воздух сила тяжести остаётся прежней, а выталкивающая сила со стороны окружающего воздуха увеличивается, так что нагретый воздух начинает всплывать к потолку. На его место приходит холодный

[10]Изображение с сайта educationalelectronicsusa.com.

воздух[11], с которым повторяется то же самое.

В результате устанавливается циркуляция воздуха, которая и служит примером конвекции — распространение тепла в комнате осуществляется воздушными потоками.

Совершенно аналогичный процесс можно наблюдать и в жидкости. Когда вы ставите на плиту чайник или кастрюлю с водой, нагревание воды происходит в первую очередь благодаря конвекции (вклад теплопроводности воды тут весьма незначителен).

Конвекционные потоки в воздухе и жидкости показаны[12] на рис. 2.19.

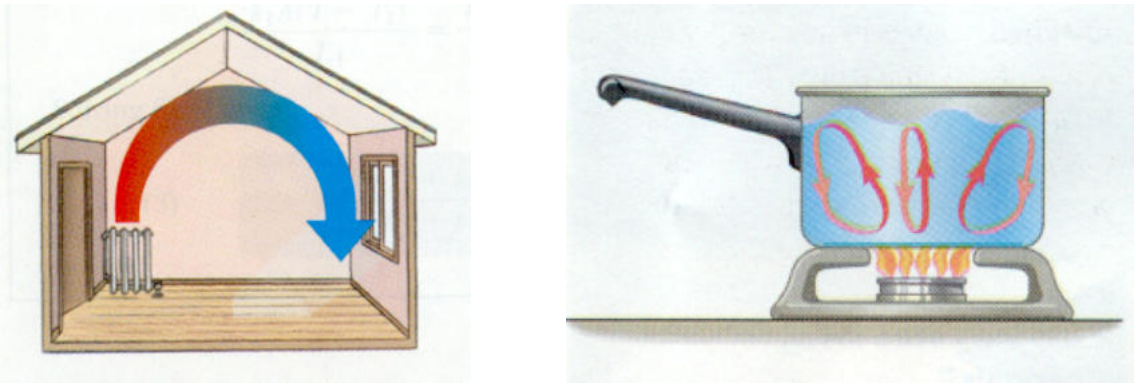


Рис. 2.19. Конвекция

В твёрдых телах конвекция отсутствует: силы взаимодействия частиц велики, частицы колеблются вблизи фиксированных пространственных точек (узлов кристаллической решётки), и никакие потоки вещества в таких условиях образоваться не могут.

Для циркуляции конвекционных потоков при отоплении комнаты необходимо, чтобы нагретому воздуху *было куда всплывать*. Если радиатор установить под потолком, то никакая циркуляция не возникнет — тёплый воздух так под потолком и останется. Именно поэтому нагревательные приборы помещают *внизу* комнаты. По той же причине чайник ставят *на* огонь, в результате чего нагретые слои воды, поднимаясь, уступают место более холодным.

Наоборот, кондиционер нужно располагать как можно выше: тогда охлаждённый воздух начнёт опускаться, и на его место будет приходить более тёплый. Циркуляция пойдёт в обратном направлении по сравнению с движением потоков при обогреве комнаты.

2.8.7 Тепловое излучение

Каким образом Земля получает энергию от Солнца? Теплопроводность и конвекция исключены: нас разделяет 150 миллионов километров безвоздушного пространства.

Здесь работает третий вид теплопередачи — *тепловое излучение*. Излучение может распространяться как в веществе, так и в вакууме. Как же оно возникает?

Оказывается, электрическое и магнитное поля тесно связаны друг с другом и обладают одним замечательным свойством. Если электрическое поле изменяется со временем, то оно порождает магнитное поле, которое, вообще говоря, также изменяется со временем[13]. В свою очередь переменное магнитное поле порождает переменное электрическое поле, которое опять порождает переменное магнитное поле, которое опять порождает переменное электрическое поле...

[11] Тот же процесс, но в куда более грандиозных масштабах, постоянно происходит в природе: именно так возникает ветер.

[12] Изображения с сайта physics.arizona.edu.

[13] Подробнее об этом будет рассказано в электродинамике, в теме про электромагнитную индукцию.

В результате развития этого процесса в пространстве распространяется *электромагнитная волна* — «зацепленные» друг за друга электрическое и магнитное поля. Как и звук, электромагнитные волны обладают скоростью распространения и частотой — в данном случае это частота, с которой колеблются в волне величины и направления полей. Видимый свет — частный случай электромагнитных волн.

Скорость распространения электромагнитных волн в вакууме огромна: 300000 км/с. Так, от Земли до Луны свет идёт чуть больше секунды.

Частотный диапазон электромагнитных волн очень широк. Подробнее о шкале электромагнитных волн мы поговорим в соответствующем листке. Здесь отметим лишь, что видимый свет — это крохотный диапазон данной шкалы. Ниже него лежат частоты инфракрасного излучения, выше — частоты ультрафиолетового излучения.

Вспомним теперь, что атомы, будучи в целом электрически нейтральными, содержат положительно заряженные протоны и отрицательно заряженные электроны. Эти заряженные частицы, совершая вместе с атомами хаотическое движение, создают переменные электрические поля и тем самым излучают электромагнитные волны. Эти волны и называются *тепловым излучением* — в напоминание о том, что их источником служит тепловое движение частиц вещества.

Источником теплового излучения является любое тело. При этом излучение уносит часть его внутренней энергии. Встретившись с атомами другого тела, излучение разгоняет их своим колеблющимся электрическим полем, и внутренняя энергия этого тела увеличивается. Именно так мы и греемся в солнечных лучах.

При обычных температурах частоты теплового излучения лежат в инфракрасном диапазоне, так что глаз его не воспринимает (мы не видим, как мы «светимся»). При нагревании тела его атомы начинают излучать волны более высоких частот. Железный гвоздь можно раскалить докрасна — довести до такой температуры, что его тепловое излучение выйдет в нижнюю (красную) часть видимого диапазона. А Солнце кажется нам жёлто-белым: температура на поверхности Солнца настолько высока (6000 °C), что в спектре его излучения присутствуют все частоты видимого света, да ещё ультрафиолет, благодаря которому мы загораем.

Давайте ещё раз взглянем на три вида теплопередачи (рис. 2.20)[14].

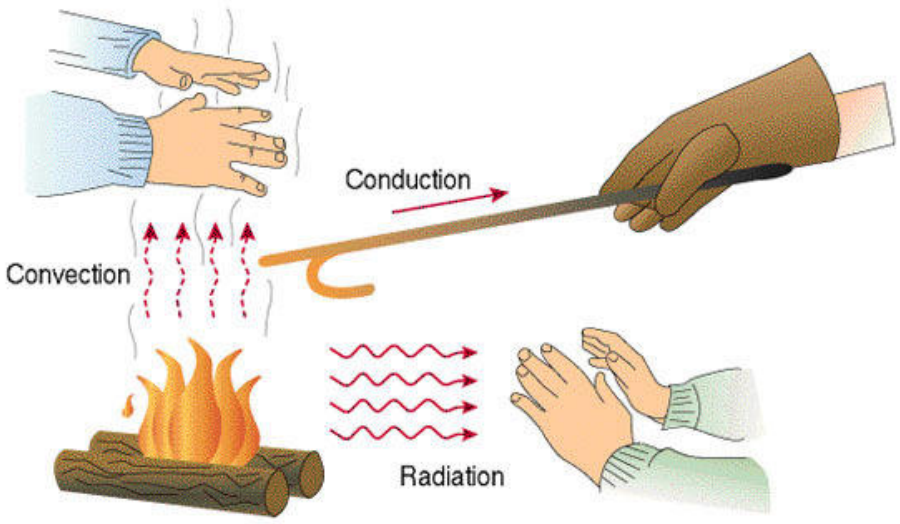


Рис. 2.20. Три вида теплопередачи: теплопроводность, конвекция, излучение

[14]Изображения с сайта beodom.com.

2.9 Количество теплоты

Как мы знаем, одним из способов изменения внутренней энергии является теплопередача (теплообмен). Предположим, что тело участвует в теплообмене с другими телами, и при этом не совершается механическая работа — ни самим телом, ни другими телами над этим телом.

Если в процессе теплообмена внутренняя энергия тела изменилась на величину ΔU, то говорят, что тело получило соответствующее количество теплоты: $Q = \Delta U$.

Если при этом величина ΔU отрицательна, т. е. тело отдавало энергию, то говорят также, что тело *отдавало тепло*. Например, вместо формально верной, но несколько нелепой фразы «тело получило -5 Дж тепла» мы скажем: «тело отдало 5 Дж тепла».

2.9.1 Удельная теплоёмкость вещества

Предположим, что в процессе теплообмена агрегатное состояние вещества тела не изменяется (не происходит плавление, кристаллизация, парообразование или конденсация). Начальную температуру тела обозначим t_1, конечную температуру — t_2.

Опыт показывает, что количество теплоты, полученное телом, прямо пропорционально массе тела m и разности конечной и начальной температур:

$$Q = cm(t_2 - t_1).$$

Коэффициент пропорциональности c называется *удельной теплоёмкостью вещества* тела. Удельная теплоёмкость не зависит от формы и размеров тела. Удельные теплоёмкости различных веществ можно найти в таблицах.

Введя обозначение $\Delta t = t_2 - t_1$, получим также:

$$Q = cm\Delta t.$$

Чтобы понять физический смысл удельной теплоёмкости, выразим её из последней формулы:

$$c = \frac{Q}{m\Delta t}\,.$$

Мы видим, что *удельная теплоёмкость численно равна количеству теплоты, которое необходимо для нагревания 1 кг данного вещества на* $1\,^\circ\mathrm{C}$ (или, что то же самое, на $1\,\mathrm{K}$). Измеряется удельная теплоёмкость в Дж/(кг $\cdot\,^\circ$C) или в Дж/(кг $\cdot$ K).

Чем больше удельная теплоёмкость вещества, тем большее количество теплоты требуется для нагревания тела данной массы на заданное количество градусов.

В задачах часто фигурируют вода и лёд. Их удельные теплоёмкости желательно помнить.

$$\text{Вода: } c = 4200\,\frac{\text{Дж}}{\text{кг}\cdot{}^\circ\text{C}}.$$
$$\text{Лёд: } c = 2100\,\frac{\text{Дж}}{\text{кг}\cdot{}^\circ\text{C}}.$$

Произведение удельной теплоёмкости вещества на массу тела называется *теплоёмкостью тела* и обозначается C:

$$C = cm.$$

Соответственно, для количества теплоты имеем:

$$Q = C(t_2 - t_1).$$

2.9.2 Уравнение теплового баланса

Рассмотрим два тела (обозначим их 1 и 2), которые образуют замкнутую систему. Это означает, что данные тела могут обмениваться энергией только друг с другом, но не с другими телами. Считаем также, что механическая работа не совершается — внутренняя энергия тел меняется только в процессе теплообмена.

Имеется фундаментальный закон природы, подтверждаемый всевозможными экспериментами — закон сохранения энергии. Он гласит, что *полная энергия замкнутой системы тел не меняется со временем.*

В данном случае закон сохранения энергии утверждает, что внутренняя энергия нашей системы будет оставаться одной и той же: $U_1 + U_2 = \text{const}$. Если изменение внутренней энергии первого тела равно ΔU_1, а изменение внутренней энергии второго тела равно ΔU_2, то суммарное изменение внутренней энергии будет равно нулю:

$$\Delta U_1 + \Delta U_2 = 0.$$

Но $\Delta U_1 = Q_1$ — количество теплоты, полученное первым телом в процессе теплообмена; аналогично $\Delta U_2 = Q_2$ — количество теплоты, полученное вторым телом в процессе теплообмена. Стало быть,

$$Q_1 + Q_2 = 0. \tag{2.7}$$

Попросту говоря, сколько дожулей тепла отдало одно тело, ровно столько же джоулей получило второе тело. Так как система замкнута, ни один джоуль наружу не вышел.

Соотношение (2.7) называется *уравнением теплового баланса.*

В общем случае, когда n тел образуют замкнутую систему и обмениваются энергией только с помощью теплопередачи, из закона сохранения энергии с помощью тех же рассуждений получаем общее уравнение теплового баланса:

$$Q_1 + Q_2 + \ldots + Q_n = 0. \tag{2.8}$$

В качестве простого примера применения уравнения теплового баланса рассмотрим следующую задачу.

Смешали $m_1 = 200$ *г воды при температуре* $t_1 = 100°\text{C}$ *и* $m_2 = 300$ *г воды при температуре* $t_2 = 20°\text{C}$. *Найти установившуюся температуру смеси.*

Обозначим искомую установившуюся температуру через θ. Запишем уравнение теплового баланса (2.7):

$$cm_1(\theta - t_1) + cm_2(\theta - t_2) = 0,$$

где c — удельная теплоёмкость воды. Раскрываем скобки и находим:

$$\theta = \frac{m_1 t_1 + m_2 t_2}{m_1 + m_2} = 52°\text{C}.$$

2.10 Фазовые переходы

Лёд, вода и водяной пар — примеры трёх *агрегатных состояний* вещества: твёрдого, жидкого и газообразного. В каком именно агрегатном состоянии находится данное вещество — зависит от его температуры и других внешних условий, в которых оно находится.

При изменении внешних условий (например, если внутренняя энергия тела увеличивается или уменьшается в результате нагревания или охлаждения) могут происходить *фазовые переходы* — изменения агрегатных состояний вещества тела. Нас будут интересовать следующие фазовые переходы:

- *плавление* (твёрдое тело → жидкость);
- *кристаллизация* (жидкость → твёрдое тело);
- *парообразование* (жидкость → пар);
- *конденсация* (пар → жидкость).

2.10.1 Плавление и кристаллизация

Большинство твёрдых тел являются *кристаллическими*, т. е. имеют *кристаллическую решётку* — строго определённое, периодически повторяющееся в пространстве расположение своих частиц.

Частицы (атомы или молекулы) кристаллического твёрдого тела совершают тепловые колебания вблизи фиксированных положений равновесия — *узлов* кристаллической решётки.

Например, узлы кристаллической решётки поваренной соли NaCl — это вершины кубических клеток «трёхмерной клетчатой бумаги» (см. рис. 2.21, на котором шарики большего размера обозначают атомы хлора)[15]; если дать испариться воде из раствора соли, то оставшаяся соль будет нагромождением маленьких кубиков.

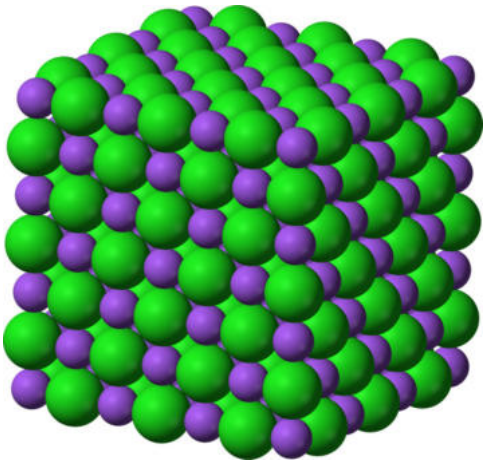


Рис. 2.21. Кристаллическая решётка NaCl

Плавлением называется превращение кристаллического твёрдого тела в жидкость. Расплавить можно любое тело — для этого нужно нагреть его до *температуры плавления*, которая зависит лишь от вещества тела, но не от его формы или размеров. Температуру плавления данного вещества можно определить из таблиц.

Наоборот, если охлаждать жидкость, то рано или поздно она перейдёт в твёрдое состояние. Превращение жидкости в кристаллическое твёрдое тело называется *кристаллизацией* или

[15]Изображение с сайта en.wikipedia.org.

отвердеванием. Таким образом, плавление и кристаллизация являются взаимно обратными процессами.

Температура, при которой жикость кристаллизуется, называется *температурой кристаллизации*. Оказывается, что температура кристаллизации равна температуре плавления: при данной температуре могут протекать оба процесса. Так, при 0 °C лёд плавится, а вода кристаллизуется; *что именно* происходит в каждом конкретном случае — зависит от внешних условий (например, подводится ли тепло к веществу или отводится от него).

Как происходят плавление и кристаллизация? Каков их механизм? Для уяснения сути этих процессов рассмотрим графики зависимости температуры тела от времени при его нагревании и охлаждении — так называемые графики плавления и кристаллизации.

2.10.2 График плавления

Начнём с *графика плавления* (рис. 2.22). Пусть в начальный момент времени (точка A на графике) тело является кристаллическим и имеет некоторую температуру t_A.

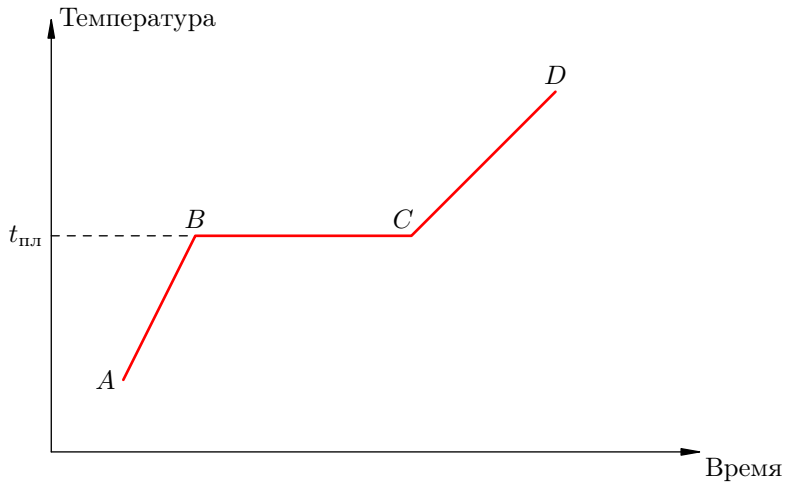


Рис. 2.22. График плавления

Затем к телу начинает подводиться тепло (скажем, тело поместили в плавильную печь), и температура тела повышается до величины $t_{пл}$ — температуры плавления данного вещества. Это участок AB графика.

На участке AB тело получает количество теплоты

$$Q = cm(t_{пл} - t_A),$$

где c — удельная теплоёмкость вещества твёрдого тела, m — масса тела.

При достижении температуры плавления (в точке B) ситуация качественно меняется. Несмотря на то, что тепло продолжает подводиться, температура тела остаётся неизменной! На участке BC происходит *плавление* тела — его постепенный переход из твёрдого состояния в жидкое.

Внутри участка BC мы имеем смесь твёрдого вещества и жидкости, и чем ближе к точке C, тем меньше остаётся твёрдого вещества и тем больше появляется жидкости. Наконец, в точке C от исходного твёрдого тела не осталось ничего: оно полностью превратилось в жидкость.

Участок CD соответствует дальнейшему нагреванию жидкости (или, как говорят, *расплава*). На этом участке жидкость поглощает количество теплоты

$$Q = c_{ж}m(t_D - t_{пл}),$$

где $c_{ж}$ — удельная теплоёмкость жидкости.

Но нас сейчас больше всего интересует BC — участок фазового перехода. Почему не меняется температура смеси на этом участке? Тепло-то подводится!

Вернёмся назад, к началу процесса нагревания. Повышение температуры твёрдого тела на участке AB есть результат возрастания интенсивности колебаний его частиц в узлах кристаллической решётки: подводимое тепло идёт на увеличение *кинетической* энергии частиц тела[16].

Кристаллическая решётка расшатывается всё сильнее и сильнее, и при температуре плавления размах колебаний достигает той предельной величины, при которой силы притяжения между частицами ещё способны обеспечивать их упорядоченное расположение друг относительно друга. Твёрдое тело начинает «трещать по швам», и дальнейшее нагревание разрушает кристаллическую решётку — так начинается плавление на участке BC.

С этого момента всё подводимое тепло идёт на совершение работы по разрыву связей, удерживающих частицы в узлах кристаллической решётки, т. е. на увеличение *потенциальной* энергии частиц. Кинетическая энергия частиц при этом остаётся прежней, так что температура тела не меняется. В точке C кристаллическая структура исчезает полностью, разрушать больше нечего, и подводимое тепло снова идёт на увеличение кинетической энергии частиц — на нагревание расплава.

2.10.3 Удельная теплота плавления

Итак, для превращения твёрдого тела в жидкость мало довести его до температуры плавления. Необходимо дополнительно (уже при температуре плавления) сообщить телу некоторое количество теплоты $Q_{пл}$ для полного разрушения кристаллической решётки (т. е. для прохождения участка BC).

Это количество теплоты идёт на увеличение потенциальной энергии взаимодействия частиц. Следовательно, внутренняя энергия расплава в точке C больше внутренней энергии твёрдого тела в точке B на величину $Q_{пл}$.

Опыт показывает, что величина $Q_{пл}$ прямо пропорциональна массе тела:

$$Q_{пл} = \lambda m.$$

Коэффициент пропорциональности λ не зависит от формы и размеров тела и является характеристикой вещества. Он называется *удельной теплотой плавления вещества*. Удельную теплоту плавления данного вещества можно найти в таблицах.

Удельная теплота плавления численно равна количеству теплоты, необходимому для превращения в жидкость одного килограмма данного кристаллического вещества, доведённого до температуры плавления.

Так, удельная теплота плавления льда равна 340 кДж/кг, свинца — 25 кДж/кг. Мы видим, что для разрушения кристаллической решётки льда требуется почти в 14 раз больше энергии! Лёд относится к веществам с большой удельной теплотой плавления и поэтому весной тает не сразу[17].

[16] На самом деле некоторая часть подводимого тепла расходуется на совершение работы по увеличению средних расстояний между частицами — как мы знаем, тела при нагревании расширяются. Однако эта часть столь мала, что её можно не принимать во внимание.

[17] Природа приняла свои меры: обладай лёд такой же удельной теплотой плавления, как и свинец, вся масса льда и снега таяла бы с первыми оттепелями, затопляя всё вокруг.

2.10.4 График кристаллизации

Теперь перейдём к рассмотрению *кристаллизации* — процесса, обратного плавлению. Начинаем с точки D предыдущего рисунка. Предположим, что в точке D нагревание расплава прекратилось (печку выключили и расплав выставили на воздух). Дальнейшее изменение температуры расплава представлено на рис. 2.23.

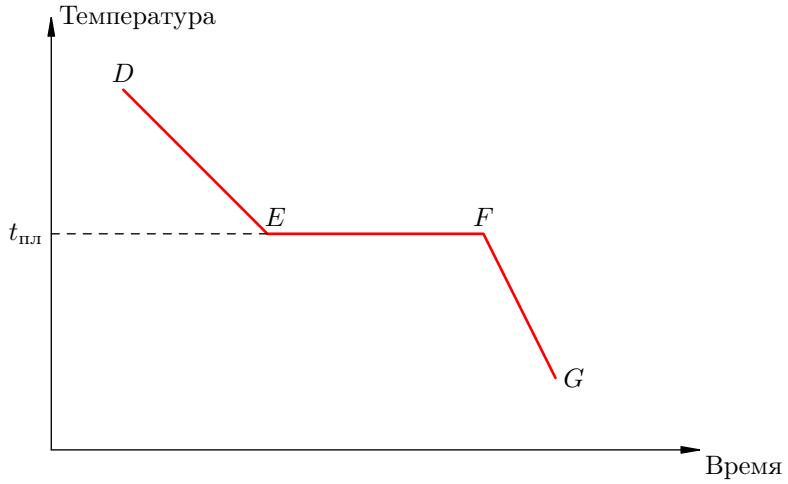


Рис. 2.23. График кристаллизации

Жидкость остывает (участок DE), пока её температура не достигнет температуры кристаллизации, которая совпадает с температурой плавления $t_{пл}$.

С этого момента температура расплава меняться перестаёт, хотя тепло по-прежнему уходит от него в окружающую среду. На участке EF происходит *кристаллизация* расплава — его постепенный переход в твёрдое состояние. Внутри участка EF мы снова имеем смесь твёрдой и жидкой фаз, и чем ближе к точке F, тем больше становится твёрдого вещества и тем меньше — жидкости. Наконец, в точке F жидкости не остаётся вовсе — она полностью кристаллизовалась.

Следующий участок FG соответствует дальнейшему остыванию твёрдого тела, возникшего в результате кристаллизации.

Нас опять-таки интересует участок фазового перехода EF: почему температура остаётся неизменной, несмотря на уход тепла?

Снова вернёмся в точку D. После прекращения подачи тепла температура расплава понижается, так как его частицы постепенно теряют кинетическую энергию в результате соударений с молекулами окружающей среды и излучения электромагнитных волн.

Когда температура расплава понизится до температуры кристаллизации (точка E), его частицы замедлятся настолько, что силы притяжения окажутся в состоянии «развернуть» их должным образом и придать им строго определённую взаимную ориентацию в пространстве. Так возникнут условия для зарождения кристаллической решётки, и она действительно начнёт формироваться благодаря дальнейшему уходу энергии из расплава в окружающее пространство.

Одновременно начнётся встречный процесс выделения энергии: когда частицы занимают свои места в узлах кристаллической решётки, их потенциальная энергия резко уменьшается, за счёт чего увеличивается их кинетическая энергия — кристаллизующаяся жидкость является источником тепла[18]. Выделяющееся в ходе кристаллизации тепло в точности компенсирует потерю тепла в окружающую среду, и потому температура на участке EF не меняется.

В точке F расплав исчезает, а вместе с завершением кристаллизации исчезает и этот внутренний «генератор» тепла. Вследствие продолжающегося рассеяния энергии во внешнюю среду

[18] Часто у проруби можно увидеть сидящих птиц. Они там греются!

понижение температуры возобновится, но только остывать уже будет образовавшееся твёрдое тело (участок FG).

Как показывает опыт, при кристаллизации на участке EF выделяется *ровно то же самое* количество теплоты $Q = \lambda m$, которое было поглощено при плавлении на участке BC.

2.10.5 Парообразование и конденсация

Парообразование — это переход жидкости в газообразное состояние (в *пар*). Существует два способа парообразования: испарение и кипение.

Испарением называется парообразование, которое происходит при любой температуре *со свободной поверхности* жидкости. Как вы помните из листка «Насыщенный пар», причиной испарения является вылет из жидкости наиболее быстрых молекул, которые способны преодолеть силы межмолекулярного притяжения. Эти молекулы и образуют пар над поверхностью жидкости.

Разные жидкости испаряются с разными скоростями: чем больше силы притяжения молекул друг к другу — тем меньшее число молекул в единицу времени окажутся в состоянии их преодолеть и вылететь наружу, и тем меньше скорость испарения. Быстро испаряются эфир, ацетон, спирт (их иногда называют летучими жидкостями), медленнее — вода, намного медленнее воды испаряются масло и ртуть.

Скорость испарения растёт с повышением температуры (в жару бельё высохнет скорее), поскольку увеличивается средняя кинетическая энергия молекул жидкости, и тем самым возрастает число быстрых молекул, способных покинуть её пределы.

Скорость испарения зависит от площади поверхности жидкости: чем больше площадь, тем большее число молекул получают доступ к поверхности, и испарение идёт быстрее (вот почему при развешивании белья его тщательно расправляют).

Одновременно с испарением наблюдается и обратный процесс: молекулы пара, совершая беспорядочное движение над поверхностью жидкости, частично возвращаются обратно в жидкость. Превращение пара в жидкость называется *конденсацией*.

Конденсация замедляет испарение жидкости. Так, в сухом воздухе бельё высохнет быстрее, чем во влажном. Быстрее оно высохнет и на ветру: пар сносится ветром, и испарение идёт более интенсивно.

В некоторых ситуациях скорость конденсации может оказаться равной скорости испарения. Тогда оба процесса компенсируют друг друга и наступает динамическое равновесие: из плотно закупоренной бутылки жидкость не улетучивается годами, а над поверхностью жидкости в этом случае находится *насыщенный пар*.

Конденсацию водяного пара в атмосфере мы постоянно наблюдаем в виде облаков, дождей и выпадающей по утрам росы; именно испарение и конденсация обеспечивают круговорот воды в природе, поддерживая жизнь на Земле.

Поскольку испарение — это уход из жидкости самых быстрых молекул, в процессе испарения средняя кинетическая энергия молекул жидкости уменьшается, т. е. жидкость *остывает*. Вам хорошо знакомо ощущение прохлады и порой даже зябкости (особенно при ветре), когда выходишь из воды: вода, испаряясь по всей поверхности тела, уносит тепло, ветер же ускоряет процесс испарения[19].

Ту же прохладу можно почувствовать, если провести по руке кусочком ваты, смоченным в летучем растворителе (скажем, в ацетоне или жидкости для снятия лака). В сорокаградусную жару благодаря усиленному испарению влаги через поры нашего тела мы сохраняем свою температуру на уровне нормальной; не будь этого терморегулирующего механизма, в такую жару

[19] Теперь понятно, зачем мы дуем на горячий чай. Кстати сказать, ещё лучше при этом *втягивать* воздух в себя, поскольку на поверхность чая тогда приходит сухой окружающий воздух, а не влажный воздух из наших лёгких ;-)

мы бы попросту погибли.

Наоборот, в процессе конденсации жидкость нагревается: молекулы пара при возвращении в жидкость разгоняются силами притяжения со стороны находящихся поблизости молекул жидкости, в результате чего средняя кинетическая энергия молекул жидкости увеличивается (сравните это явление с выделением энергии при кристаллизации расплава!).

2.10.6 Кипение

Процесс кипения воды вам хорошо знаком. В отличие от испарения, которое происходит только со свободной поверхности жидкости, *кипение* — это парообразование, происходящее *по всему объёму* жидкости.

Кипение оказывается возможным потому, что в жидкости всегда растворено какое-то количество воздуха, попавшего туда в результате диффузии. При нагревании жидкости этот воздух расширяется, пузырьки воздуха постепенно увеличиваются в размерах и становятся видимы невооружённым глазом (в кастрюле с водой они осаждают дно и стенки). Внутри воздушных пузырьков находится насыщенный пар, давление которого, как вы помните, быстро растёт с повышением температуры.

Чем крупнее становятся пузырьки, тем бóльшая действует на них архимедова сила, и определённого момента начинается отрыв и всплытие пузырьков. Поднимаясь вверх, пузырьки попадают в менее нагретые слои жидкости; пар в них конденсируется, и пузырьки сжимаются опять. Схлопывание пузырьков вызывает знакомый нам шум, предшествующий закипанию чайника. Наконец, с течением времени вся жидкость равномерно прогревается, пузырьки достигают поверхности и лопаются, выбрасывая наружу воздух и пар — шум сменяется бульканьем, жидкость кипит.

Пузырьки, таким образом, служат «проводниками» пара изнутри жидкости на её поверхность. При кипении наряду с обычным испарением идёт превращение жидкости в пар по всему объёму — испарение внутрь воздушных пузырьков с последующим выводом пара наружу. Вот почему кипящая жидкость улетучивается очень быстро: чайник, из которого вода испарялась бы много дней, выкипит за полчаса.

В отличие от испарения, происходящего при любой температуре, жидкость начинает кипеть только при достижении *температуры кипения* — именно той температуры, при которой пузырьки воздуха оказываются в состоянии всплыть и добраться до поверхности. *При температуре кипения давление насыщенного пара становится равно внешнему давлению на жидкость (в частности, атмосферному давлению).* Соответственно, чем больше внешнее давление, тем при более высокой температуре начнётся кипение.

При нормальном атмосферном давлении (1 атм или 10^5 Па) температура кипения воды равна 100 °C. Поэтому *давление насыщенного водяного пара при температуре* 100 °C *равно* 10^5 Па. Этот факт необходимо знать для решения задач — часто он считается известным по умолчанию.

На вершине Эльбруса атмосферное давление равно 0,5 атм, и вода там закипит при температуре 82 °C. А под давлением 15 атм вода начнёт кипеть только при 200 °C.

Температура кипения (при нормальном атмосферном давлении) является строго определённой для данной жидкости величиной[20]. Так, спирт кипит при 78 °C, эфир — при 35 °C, ртуть — при 357 °C. Обратите внимание: чем более летучей является жидкость, тем ниже её температура кипения. В таблице температур кипения мы видим также, что кислород кипит при −183 °C. Значит, при обычных температурах кислород — это газ!

[20]Температуры кипения, приводимые в таблицах учебников и справочников — это температуры кипения *химически чистых* жидкостей. Наличие в жидкости примесей может изменять температуру кипения. Скажем, водопроводная вода содержит растворённый хлор и некоторые соли, поэтому её температура кипения при нормальном атмосферном давлении может несколько отличаться от 100 °C.

Мы знаем, что если чайник снять с огня, то кипение тут же прекратится — процесс кипения требует непрерывного подвода тепла. Вместе с тем, температура воды в чайнике после закипания перестаёт меняться, всё время оставаясь равной 100 °C. Куда же при этом девается подводимое тепло?

Ситуация аналогична процессу плавления: тепло идёт на увеличение потенциальной энергии молекул. В данном случае — на совершение работы по удалению молекул на такие расстояния, что силы притяжения окажутся не способными удерживать молекулы неподалёку друг от друга, и жидкость будет переходить в газообразное состояние.

2.10.7 График кипения

Рассмотрим графическое представление процесса нагревания жидкости — так называемый *график кипения* (рис. 2.24).

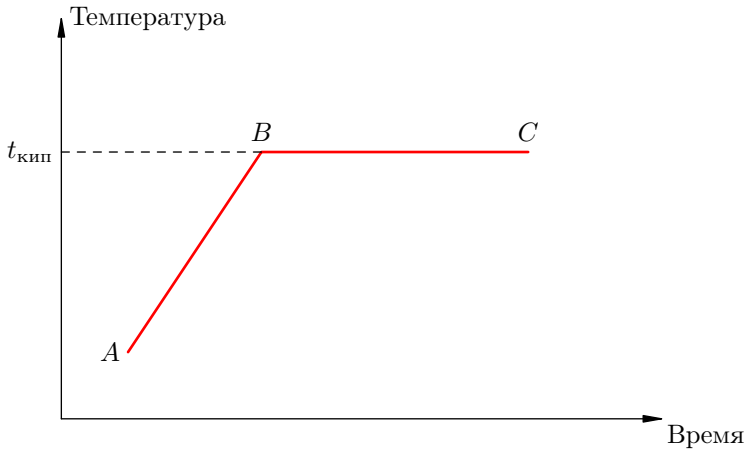


Рис. 2.24. График кипения

Участок AB предшествует началу кипения. На участке BC жидкость кипит, её масса уменьшается. В точке C жидкость выкипает полностью.

Чтобы целиком пройти участок BC, т. е. чтобы жидкость, уже доведённую до температуры кипения, полностью превратить в пар, к этой жидкости нужно подвести некоторое количество теплоты $Q_{пар}$. Опыт показывает, что данное количество теплоты прямо пропорционально массе жидкости:

$$Q_{пар} = Lm.$$

Коэффициент пропорциональности L называется *удельной теплотой парообразования* жидкости (при температуре кипения). Удельная теплота парообразования численно равна количеству теплоты, которое нужно подвести к 1 кг жидкости, взятой при температуре кипения, чтобы полностью превратить её в пар.

Так, при 100 °C удельная теплота парообразования воды равна 2300 кДж/кг. Интересно сравнить её с удельной теплотой плавления льда (340 кДж/кг) — удельная теплота парообразования почти в семь раз больше! Это и не удивительно: ведь для плавления льда нужно лишь разрушить упорядоченное расположение молекул воды в узлах кристаллической решётки; при этом расстояния между молекулами остаются примерно теми же (порядка размеров самих молекул). А вот для превращения воды в пар нужно совершить куда бóльшую работу по разрыву всех связей между молекулами и удалению молекул на значительные расстояния друг от друга (гораздо большие, чем размеры молекул).

2.10.8 График конденсации

Процесс конденсации пара и последующего остывания жидкости выглядит на графике симметрично процессу нагревания и кипения. Вот соответствующий *график конденсации* для случая стоградусного водяного пара, наиболее часто встречающегося в задачах (рис. 2.25).

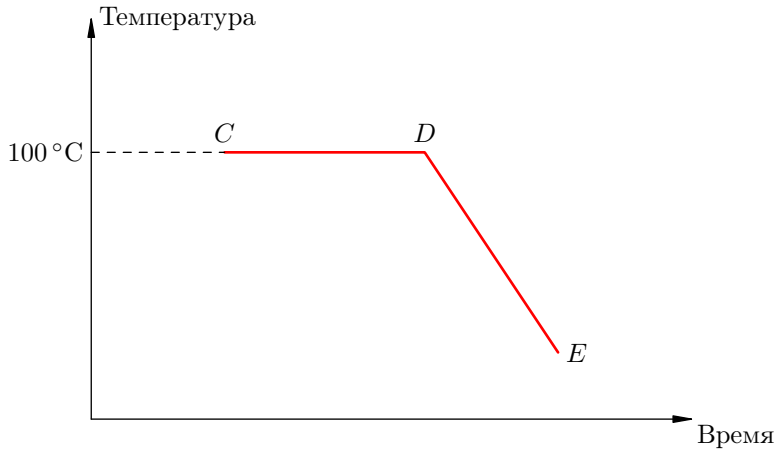


Рис. 2.25. График конденсации

В точке C имеем водяной пар при 100 °C. На участке CD идёт конденсация; внутри этого участка — смесь пара и воды при 100 °C. В точке D пара больше нет, имеется лишь вода при 100 °C. Участок DE — остывание этой воды.

Опыт показывает, что при конденсации пара массы m (т. е. при прохождении участка CD) выделяется *ровно то же самое* количество теплоты $Q = Lm$, которое было потрачено на превращение в пар жидкости массы m при данной температуре.

Давайте ради интереса сравним следующие количества теплоты:

- Q_1, которое выделяется при конденсации 1 г водяного пара;
- Q_2, которое выделяется при остывании получившейся стоградусной воды до температуры, скажем, 20 °C.

Имеем:

$$Q_1 = Lm = 2300000 \cdot 0{,}001 = 2300 \text{ Дж};$$
$$Q_2 = cm\Delta t = 4200 \cdot 0{,}001 \cdot 80 = 336 \text{ Дж}.$$

Эти числа наглядно показывают, что ожог паром гораздо страшнее ожога кипятком. При попадании на кожу кипятка выделяется «всего лишь» Q_2 (кипяток остывает). А вот при ожоге паром сначала выделится на порядок большее количество теплоты Q_1 (пар конденсируется), образуется стоградусная вода, после чего добавится та же величина Q_2 при остывании этой воды.

2.11 Первый закон термодинамики

Начнём с обуждения работы газа.

Газ, находящийся в сосуде под поршнем, действует на поршень с силой $F = pS$, где p — давление газа, S — площадь поршня. Если при этом поршень перемещается, то газ совершает *работу*.

При расширении газа эта работа будет положительной (сила давления газа и перемещение поршня направлены в одну сторону). При сжатии работа газа отрицательна (сила давления газа и перемещение поршня направлены в противоположные стороны).

2.11.1 Работа газа в изобарном процессе

Предположим, что газ расширяется при постоянном давлении p. Тогда сила F, с которой газ действует на поршень, также постоянна. Пусть поршень переместился на расстояние Δx (рис. 2.26).

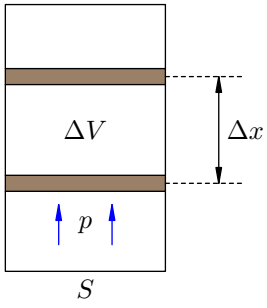


Рис. 2.26. $A = p\Delta V$

Работа газа равна:

$$A = F\Delta x = pS\Delta x.$$

Но $S\Delta x = \Delta V$ — изменение объёма газа. Поэтому для работы газа при изобарном расширении мы получаем формулу:

$$A = p\Delta V. \qquad (2.9)$$

Если V_1 и V_2 — начальный и конечный объём газа, то для работы газа имеем:

$$A = p(V_2 - V_1).$$

Изобразив данный процесс на pV-диаграмме, мы видим, что работа газа равна площади прямоугольника под графиком нашего процесса (рис. 2.27).

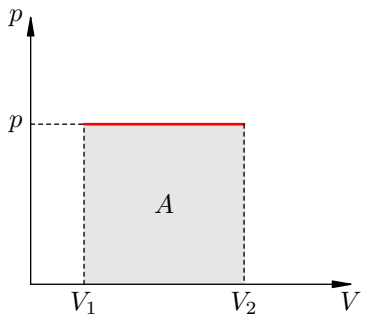


Рис. 2.27. Работа газа как площадь

Пусть теперь газ изобарно сжимается от объёма V_1 до объёма V_2. С помощью аналогичных рассуждений приходим к формуле:

$$A = -p(V_1 - V_2).$$

Но $-(V_1 - V_2) = V_2 - V_1 = \Delta V$, и снова получается формула (2.9).

Работа газа опять-таки будет равна площади под графиком процесса на pV-диаграмме, но теперь со знаком минус.

Итак, формула $A = p\Delta V$ выражает работу газа при постоянном давлении — как в процессе расширения газа, так и в процессе сжатия.

2.11.2 Работа газа в произвольном процессе

Геометрическая интерпретация работы газа (как площади под графиком процесса на pV-диаграмме) сохраняется и в общем случае неизобарного процесса.

Действительно, рассмотрим малое изменение dV объёма газа — настолько малое, что давление p будет оставаться приблизительно постоянным. Газ совершит малую работу $dA = pdV$. Тогда работа A газа во всём процессе найдётся суммированием этих малых работ:

$$A = \int_{V_1}^{V_2} p\,dV.$$

Но данный интеграл как раз и является площадью криволинейной трапеции (рис. 2.28):

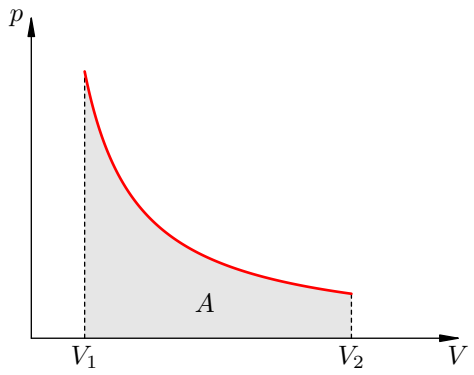


Рис. 2.28. Работа газа как площадь

2.11.3 Работа, совершаемая над газом

Наряду с работой A, которую совершает газ по передвижению поршня, рассматривают также работу A', которую поршень совершает над газом.

Если газ действует на поршень с силой $\vec{F}$, то по третьему закону Ньютона поршень действует на газ с силой $\vec{F}'$, равной силе $\vec{F}$ по модулю и противоположной по направлению: $\vec{F}' = -\vec{F}$ (рис. 2.29).

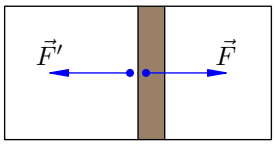


Рис. 2.29. Внешняя сила $\vec{F}'$, действующая на газ

Следовательно, работа поршня A' равна по модулю и противоположна по знаку работе газа:

$$A' = -A.$$

Так, в процессе расширения газ совершает положительную работу ($A > 0$); при этом работа, совершаемая над газом, отрицательна ($A' < 0$). Наоборот, при сжатии работа газа отрицательна ($A < 0$), а работа, совершаемая поршнем над газом, положительна ($A' > 0$).

Будьте внимательны: если в задаче просят найти работу, совершённую над газом, то имеется в виду работа A'.

2.11.4 Первый закон термодинамики

Как мы знаем, существует лишь два способа изменения внутренней энергии тела: теплопередача и совершение работы.

Опыт показывает, что эти способы независимы — в том смысле, что их результаты складываются. Если телу в процессе теплообмена передано количество теплоты Q, и если в то же время над телом совершена работа A', то изменение внутренней энергии тела будет равно:

$$\Delta U = Q + A'. \tag{2.10}$$

Нас больше всего интересует случай, когда тело является газом. Тогда $A' = -A$ (где A, как всегда, есть работа самого газа). Формула (2.10) принимает вид: $\Delta U = Q - A$, или

$$Q = \Delta U + A. \tag{2.11}$$

Соотношение (2.11) называется *первым законом термодинамики*. Смысл его прост: *количество теплоты, переданное газу, идёт на изменение внутренней энергии газа и на совершение газом работы*.

Напомним, что величина Q может быть и отрицательной: в таком случае тепло отводится от газа. Но первый закон термодинамики остаётся справедливым в любом случае. Он является одним из фундаментальных физических законов и находит подтверждение в многочисленных явлениях и экспериментах.

2.11.5 Применение первого закона термодинамики к изопроцессам

Напомним, что в изопроцессе остаётся неизменным значение некоторой величины, характеризующей состояние газа — температуры, объёма или давления. Для каждого вида изопроцессов запись первого закона термодинамики упрощается.

1. *Изотермический процесс*, $T = \text{const}$.

   Внутренняя энергия идеального газа зависит только от его температуры. Если температура газа не меняется, то не меняется и внутренняя энергия: $\Delta U = 0$. Тогда формула (2.11) даёт:

   $$Q = A.$$

   Всё подведённое к газу тепло идёт на совершение газом работы.

2. *Изохорный процесс*, $V = \text{const}$.

   Если объём газа остаётся постоянным, то поршень не перемещается, и потому работа газа равна нулю: $A = 0$. Тогда первый закон термодинамики даёт:

   $$Q = \Delta U.$$

   Всё тепло, переданное газу, идёт на изменение его внутренней энергии.

3. *Изобарный процесс*, $p = \text{const}$.

   Подведённое к газу тепло идёт как на изменение внутренней энергии, так и на совершение работы (для которой справедлива формула (2.9)). Имеем:

   $$Q = \Delta U + p\Delta V.$$

2.11.6 Адиабатный процесс

Процесс называется *адиабатным*, если он идёт без теплообмена с окружающими телами.

Адиабатный процесс совершается газом, находящимся в теплоизолированном сосуде. Такой сосуд препятствует всем видам теплопередачи: теплопроводности, конвекции, излучению. Пример теплоизолированного сосуда — термос.

Приблизительно адиабатным будет всякий процесс, протекающий достаточно быстро: в течение процесса теплообмен просто не успевает произойти.

При адиабатном процессе $Q = 0$. Из первого закона термодинамики получаем: $A + \Delta U = 0$, или $A = -\Delta U$.

В процессе адиабатного расширения газ совершает положительную работу, поэтому $\Delta U < 0$ (работа совершается за счёт убыли внутренней энергии). Следовательно, газ охлаждается. Если заставить газ совершить достаточно большую работу, охладить его можно весьма сильно. Именно на этом основаны методы сжижения газов.

Наоборот, в процессе адиабатного сжатия будет $A < 0$, поэтому $\Delta U > 0$: газ нагревается. Адиабатное нагревание воздуха используется в дизельных двигателях для воспламенения топлива.

Кривая, изображающая ход адиабатного процесса, называется *адиабатой*. Интересно сравнить ход адиабаты и изотермы на pV-диаграмме (рис. 2.30).

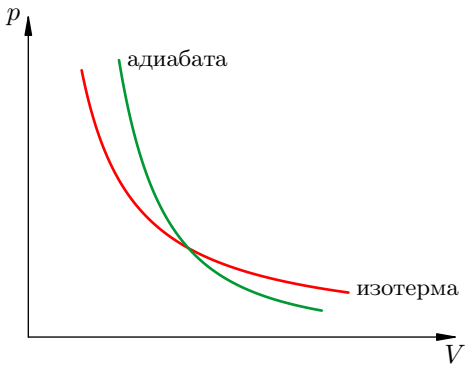


Рис. 2.30. Сравнительный ход изотермы и адиабаты

В обоих процессах давление убывает с увеличением объёма, но в адиабатном процессе убывание идёт быстрее. Почему?

При изотермическом расширении давление падает потому, что уменьшается концентрация частиц газа, в результате чего удары частиц по стенкам сосуда становятся реже. Однако интенсивность этих ударов остаётся прежней: ведь температура газа не меняется — значит, не меняется и средняя кинетическая энергия его частиц.

А при адиабатном расширении, наряду с уменьшением концентрации частиц, падает также и температура газа. Удары частиц становятся не только более редкими, но и более слабыми. Вот почему адиабата убывает быстрее изотермы.

2.12 Тепловые машины

Коротко говоря, *тепловые машины* преобразуют теплоту в работу или, наоборот, работу в теплоту.

Тепловые машины бывают двух видов — в зависимости от направления протекающих в них процессов.

1. *Тепловые двигатели* преобразуют теплоту, поступающую от внешнего источника, в механическую работу.

   Автомобильный двигатель внутреннего сгорания — это пример теплового двигателя. В нём происходит преобразование тепла, выделяющегося при сгорании топлива, в механическую энергию автомобиля.

2. *Холодильные машины* передают тепло от менее нагретого тела к более нагретому за счёт механической работы внешнего источника.

   Бытовой холодильник, который стоит у вас в квартире, служит примером холодильной машины. В нём тепло отводится от холодильной камеры и передаётся в окружающее пространство.

Рассмотрим эти виды тепловых машин более подробно.

2.12.1 Тепловые двигатели

Мы знаем, что совершение над телом работы есть один из способов изменения его внутренней энергии: совершённая работа как бы растворяется в теле, переходя в энергию беспорядочного движения и взаимодействия его частиц.

Тепловой двигатель — это устройство, которое, наоборот, извлекает полезную работу из «хаотической» внутренней энергии тела. Изобретение теплового двигателя радикально изменило облик человеческой цивилизации.

Принципиальную схему теплового двигателя можно изобразить следующим образом (рис. 2.31). Давайте разбираться, что означают элементы данной схемы.

Рабочее тело двигателя — это газ. Он расширяется, двигает поршень и совершает тем самым полезную механическую работу.

Но чтобы заставить газ расширяться, преодолевая внешние силы, нужно нагреть его до температуры, которая существенно выше температуры окружающей среды. Для этого газ приводится в контакт с *нагревателем* — сгорающим топливом.

В процессе сгорания топлива выделяется значительная энергия, часть которой идёт на нагревание газа. Газ получает от нагревателя количество теплоты Q_1. Именно за счёт этого тепла двигатель совершает полезную работу A.

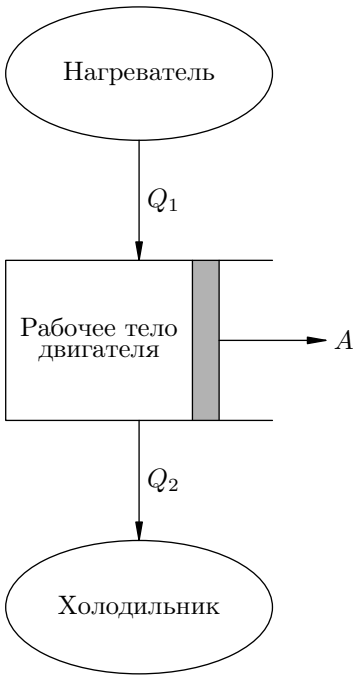


Рис. 2.31. Тепловой двигатель

Это всё понятно. Что такое холодильник и зачем он нужен?

При однократном расширении газа мы можем использовать поступающее тепло максимально эффективно и целиком превратить его в работу. Для этого надо расширять газ изотермически: первый закон термодинамики, как мы знаем, даёт нам в этом случае $A = Q_1$.

Но однократное расширение никому не нужно. Двигатель должен работать *циклически*, обеспечивая периодическую повторяемость движений поршня. Следовательно, по окончании расширения газ нужно сжимать, возвращая его в исходное состояние.

В процессе расширения газ совершает некоторую положительную работу A_1. В процессе сжатия над газом совершается положительная работа A_2 (а сам газ совершает отрицательную работу $-A_2$). В итоге полезная работа газа за цикл: $A = A_1 - A_2$.

Разумеется, должно быть $A > 0$, или $A_2 < A_1$ (иначе никакого смысла в двигателе нет). Сжимая газ, мы должны совершить меньшую работу, чем совершил газ при расширении.

Как этого достичь? Ответ: сжимать газ под меньшими давлениями, чем были в ходе расширения. Иными словами, на pV-диаграмме процесс сжатия должен идти *ниже* процесса расширения, т. е. цикл должен проходиться *по часовой стрелке* (рис. 2.32).

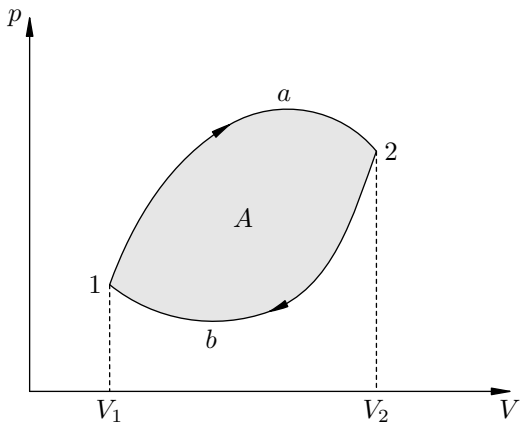


Рис. 2.32. Цикл теплового двигателя

Например, в цикле на рисунке работа газа при расширении равна площади криволинейной трапеции $V_1 1a 2 V_2$. Аналогично, работа газа при сжатии равна площади криволинейной трапеции $V_1 1b 2 V_2$ со знаком минус. В результате работа A газа за цикл оказывается положительной и равной площади цикла $1a2b1$.

Хорошо, но как заставить газ возвращаться в исходное состояние по более низкой кривой, то есть через состояния с меньшими давлениями? Вспомним, что при данном объёме давление газа тем меньше, чем ниже температура. Стало быть, при сжатии газ должен проходить состояния с меньшими температурами.

Вот именно для этого и нужен холодильник: чтобы *охлаждать* газ в процессе сжатия. Холодильником может служить атмосфера (для двигателей внутреннего сгорания) или охлаждающая проточная вода (для паровых турбин).

При охлаждении газ отдаёт холодильнику некоторое количество теплоты Q_2. Суммарное количество теплоты, полученное газом за цикл, оказывается равным $Q_1 - Q_2$. Согласно первому закону термодинамики:

$$Q_1 - Q_2 = A + \Delta U,$$

где ΔU — изменение внутренней энергии газа за цикл. Оно равно нулю: $\Delta U = 0$, так как газ вернулся в исходное состояние (а внутренняя энергия, как мы помним, является *функцией состояния*). В итоге работа газа за цикл получается равна:

$$A = Q_1 - Q_2. \tag{2.12}$$

Как видите, $A < Q_1$: не удаётся полностью превратить в работу поступающее от нагревателя тепло. Часть теплоты приходится отдавать холодильнику — для обеспечения цикличности процесса.

Показателем эффективности превращения энергии сгорающего топлива в механическую работу служит коэффициент полезного действия теплового двигателя.

КПД теплового двигателя — это отношение механической работы A к количеству теплоты Q_1, поступившему от нагревателя:

$$\eta = \frac{A}{Q_1}\,.$$

С учётом соотношения (2.12) имеем также

$$\eta = \frac{Q_1 - Q_2}{Q_1}\,. \tag{2.13}$$

КПД теплового двигателя, как видим, всегда меньше единицы. Например, КПД паровых турбин приблизительно 25%, а КПД двигателей внутреннего сгорания около 40%.

2.12.2 Холодильные машины

Житейский опыт и физические эксперименты говорят нам о том, что в процессе теплообмена теплота передаётся от более нагретого тела к менее нагретому, но не наоборот. Никогда не наблюдаются процессы, в которых за счёт теплообмена энергия *самопроизвольно* переходит от холодного тела к горячему, в результате чего холодное тело ещё больше остывало бы, а горячее тело — ещё больше нагревалось.

Ключевое слово здесь — «самопроизвольно». Если использовать внешний источник энергии, то осуществить процесс передачи тепла от холодного тела к горячему оказывается вполне возможным. Это и делают холодильные машины.

По сравнению с тепловым двигателем процессы в холодильной машине имеют противоположное направление (рис. 2.33).

Рабочее тело холодильной машины называют также *хладагентом*. Мы для простоты будем считать его газом, который поглощает теплоту при расширении и отдаёт при сжатии[21].

Холодильник в холодильной машине — это тело, от которого отводится теплота. Холодильник передаёт рабочему телу (газу) количество теплоты Q_2, в результате чего газ расширяется.

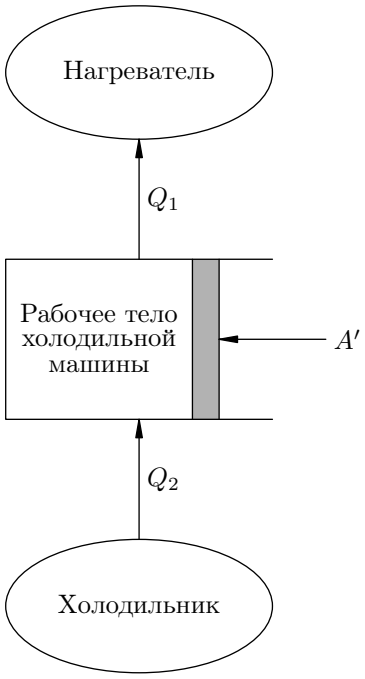


Рис. 2.33. Холодильная машина

В ходе сжатия газ отдаёт теплоту Q_1 более нагретому телу — *нагревателю*. Чтобы такая теплопередача осуществлялась, надо сжимать газ при более высоких температурах, чем были при расширении. Это возможно лишь за счёт работы A', совершаемой внешним источником (например, электродвигателем)[22]. Поэтому количество теплоты, передаваемое нагревателю, оказывается больше количества теплоты, забираемого от холодильника, как раз на величину A':

$$Q_1 = Q_2 + A'.$$

[21] В реальных холодильных установках хладагент — это летучий раствор с низкой температурой кипения, который забирает теплоту в процессе испарения и отдаёт при конденсации.

[22] В реальных холодильных агрегатах электродвигатель создаёт в *испарителе* низкое давление, в результате чего хладагент вскипает и забирает тепло; наоборот, в *конденсаторе* электродвигатель создаёт высокое давление, под которым хладагент конденсируется и отдаёт тепло.

Таким образом, на pV-диаграмме рабочий цикл холодильной машины идёт *против часовой стрелки*. Площадь цикла — это работа A', совершаемая внешним источником (рис. 2.34).

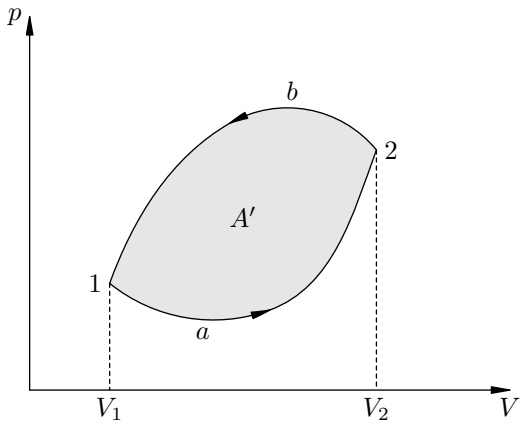


Рис. 2.34. Цикл холодильной машины

Основное назначение холодильной машины — охлаждение некоторого резервуара (например, морозильной камеры). В таком случае данный резервуар играет роль холодильника, а нагревателем служит окружающая среда — в неё рассеивается отводимое от резервуара тепло.

Показателем эффективности работы холодильной машины является *холодильный коэффициент*, равный отношению отведённого от холодильника тепла к работе внешнего источника:

$$\alpha = \frac{Q_2}{A'}\,.$$

Холодильный коэффициент может быть и больше единицы. В реальных холодильниках он принимает значения приблизительно от 1 до 3.

Имеется ещё одно интересное применение: холодильная машина может работать как *тепловой насос*. Тогда её назначение — нагревание некоторого резервуара (например, обогрев помещения) за счёт тепла, отводимого от окружающей среды. В данном случае этот резервуар будет нагревателем, а окружающая среда — холодильником.

Показателем эффективности работы теплового насоса служит *отопительный коэффициент*, равный отношению количества теплоты, переданного обогреваемому резервуару, к работе внешнего источника:

$$\beta = \frac{Q_1}{A'}\,.$$

Значения отопительного коэффициента реальных тепловых насосов находятся обычно в диапазоне от 3 до 5.

2.12.3 Тепловая машина Карно

Важными характеристиками тепловой машины являются наибольшее и наименьшее значения температуры рабочего тела в ходе цикла. Эти значения называются соответственно *температурой нагревателя* и *температурой холодильника*.

Мы видели, что КПД теплового двигателя строго меньше единицы. Возникает естественный вопрос: каков наибольший возможный КПД теплового двигателя с фиксированными значениями температуры нагревателя T_1 и температуры холодильника T_2?

Пусть, например, максимальная температура рабочего тела двигателя равна 1000 К, а минимальная — 300 К. Каков теоретический предел КПД такого двигателя?

Ответ на поставленный вопрос дал французский физик и инженер Сади Карно в 1824 году. Он придумал и исследовал замечательную тепловую машину с идеальным газом в качестве

рабочего тела. Эта машина работает по *циклу Карно*, состоящему из двух изотерм и двух адиабат.

Рассмотрим *прямой цикл* машины Карно, идущий по часовой стрелке (рис. 2.35). В этом случае машина функционирует как тепловой двигатель.

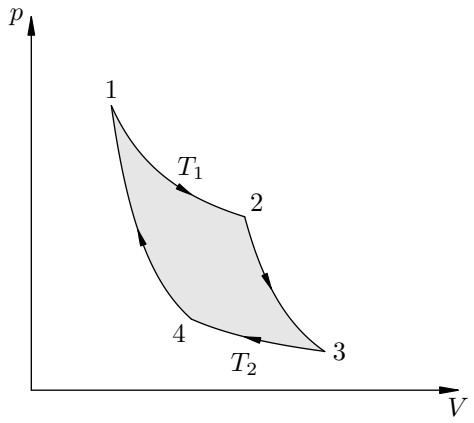


Рис. 2.35. Цикл Карно

Изотерма $1 \to 2$. На участке $1 \to 2$ газ приводится в тепловой контакт с нагревателем температуры T_1 и расширяется изотермически. От нагревателя поступает количество теплоты Q_1 и целиком превращается в работу на этом участке: $A_{12} = Q_1$.

Адиабата $2 \to 3$. В целях последующего сжатия нужно перевести газ в зону более низких температур. Для этого газ теплоизолируется, а затем расширяется адиабатно на участке $2 \to 3$. При расширении газ совершает положительную работу A_{23}, и за счёт этого уменьшается его внутренняя энергия: $\Delta U_{23} = -A_{23}$.

Изотерма $3 \to 4$. Теплоизоляция снимается, газ приводится в тепловой контакт с холодильником температуры T_2. Происходит изотермическое сжатие. Газ отдаёт холодильнику количество теплоты Q_2 и совершает отрицательную работу $A_{34} = -Q_2$.

Адиабата $4 \to 1$. Этот участок необходим для возврата газа в исходное состояние. В ходе адиабатного сжатия газ совершает отрицательную работу A_{41}, а изменение внутренней энергии положительно: $\Delta U_{41} = -A_{41}$. Газ нагревается до исходной температуры T_1.

Карно нашёл КПД этого цикла (вычисления, к сожалению, выходят за рамки школьной программы):

$$\eta = \frac{T_1 - T_2}{T_1}\,. \tag{2.14}$$

Кроме того, он доказал, что *КПД цикла Карно является максимально возможным для всех тепловых двигателей с температурой нагревателя* T_1 *и температурой холодильника* T_2.

Так, в приведённом выше примере ($T_1 = 1000$ K, $T_2 = 300$ K) имеем:

$$\eta_{\max} = \frac{1000 - 300}{1000} = 0{,}7\,(= 70\%).$$

В чём смысл использования именно изотерм и адиабат, а не каких-то других процессов? Оказывается, изотермические и адиабатные процессы делают машину Карно *обратимой*. Её можно запустить по *обратному циклу* (против часовой стрелки) между теми же нагревателем и холодильником, не привлекая другие устройства. В таком случае машина Карно будет функционировать как холодильная машина.

Возможность запуска машины Карно в обоих направлениях играет очень большую роль в термодинамике. Например, данный факт служит звеном доказательства максимальности КПД цикла Карно. Мы ещё вернёмся к этому в следующей статье, посвящённой второму закону термодинамики.

2.12.4 Тепловые двигатели и охрана окружающей среды

Тепловые двигатели наносят серьёзный ущерб окружающей среде. Их повсеместное использование приводит к целому ряду негативных эффектов.

- Рассеяние в атмосферу огромного количества тепловой энергии приводит к повышению температуры на планете. Потепление климата грозит обернуться таянием ледников и катастрофическими бедствиями.
- К потеплению климата ведёт также накопление в атмосфере углекислого газа, который замедляет уход теплового излучения Земли в космос (парниковый эффект).
- Из-за высокой концентрации продуктов сгорания топлива ухудшается экологическая ситуация.

Это — проблемы в масштабе всей цивилизации. Для борьбы с вредными последствиями работы тепловых двигателей следует повышать их КПД, снижать выбросы токсичных веществ, разрабатывать новые виды топлива и экономно расходовать энергию.

2.13 Второй закон термодинамики

Не все физические процессы, допускаемые нашим воображением, могут осуществляться в реальности.

Например, в течение нескольких столетий предпринимались попытки изобрести *вечный двигатель первого рода* — устройство, способное производить неограниченное количество механической работы само по себе, без привлечения внешних источников энергии.

Все подобные проекты, зачастую весьма хитроумные, неизменно терпели крах. В конечном счёте это привело к открытию фундаментального закона природы — закона сохранения энергии. Любой процесс, нарушающий закон сохранения энергии, оказывается невозможным; точнее — не обнаружено ни одного процесса, в котором не выполнялся бы закон сохранения энергии.

В термодинамике закон сохранения энергии принял форму *первого закона термодинамики*.

2.13.1 Необратимость процессов в природе

Существуют, однако, и другие ограничения на ход процессов, не связанные с законом сохранения энергии. Оказывается, *реальные процессы необратимы: они могут самопроизвольно идти лишь в одном определённом направлении*. Обратные им процессы, также не нарушающие закон сохранения энергии, сами по себе никогда не протекают.

Так, если привести в тепловой контакт горячее и холодное тело, то энергия в процессе теплообмена будет передаваться от горячего тела к холодному. Обратный процесс, в результате которого холодное тело ещё больше остывало бы, а горячее — ещё больше нагревалось, хотя и возможен с энергетической точки зрения, в действительности *сам по себе* не происходит. *Теплопередача от горячего тела к холодному является необратимым процессом*[23].

Подобных примеров можно привести сколько угодно. Ложка падает со стола и остаётся лежать на полу — кинетическая энергия её механического движения переходит во внутреннюю энергию пола и самой ложки. Энергетически возможен и обратный процесс: молекулы со стороны пола синхронно ударят по ложке так, что она запрыгнет обратно на стол (а участок пола при этом охладится, потеряв внутреннюю энергию). Но такое в природе никогда не наблюдалось. *Превращение механической энергии во внутреннюю является необратимым процессом*.

2.13.2 Постулаты Клаузиуса и Кельвина

Необратимость процессов теплопередачи от горячего тела к холодному и превращения механической энергии во внутреннюю была постулирована во *втором законе термодинамики*. Две эквивалентные формулировки второго закона термодинамики принадлежат Клаузиусу и Томсону (лорду Кельвину).

Постулат Клаузиуса. Невозможен процесс, единственным результатом которого является передача тепла от менее нагретого тела к более нагретому.

Процесс, при котором тепло передаётся от менее нагретого тела к более нагретому, происходит в холодильных машинах. Но эта передача тепла не является единственным результатом данного процесса. Происходят изменения и в других телах — например, работает внешний источник энергии. Постулат Клаузиуса утверждает невозможность самопроизвольной теплопередачи от холодного тела к горячему.

Постулат Кельвина. Невозможен циклический процесс, единственным результатом которого является получение работы за счёт охлаждения теплового резервуара.

[23] Подчеркнём ещё раз, что речь идёт о невозможности *самопроизвольного* протекания обратного процесса. В принципе передать энергию от холодного тела к горячему можно — но только за счёт работы внешнего источника. Именно этим и занимаются холодильные машины.

В тепловом двигателе работа A получается получается за счёт отвода тепла Q_1 от теплового резервуара — нагревателя. Но получение работы не является единственным результатом данного процесса. Мы уже отмечали, что для обеспечения цикличности работы теплового двигателя какое-то количество теплоты Q_2 должно быть отдано другому резервуару — холодильнику. Поэтому происходят изменения и в других телах.

Фактически постулат Кельвина запрещает существование теплового двигателя с КПД, равным единице.

Гипотетический тепловой двигатель, целиком превращающий в работу поступающее от нагревателя тепло, называется *вечным двигателем второго рода*. Он не противоречит закону сохранения энергии. Если бы такой двигатель был изобретён, люди получили бы возможность без потерь превращать в работу огромные запасы внутренней энергии морей и океанов. Но невозможность создать вечный двигатель второго рода как раз и утверждается постулатом Кельвина.

2.13.3 Эквивалентность постулатов Клаузиуса и Кельвина

Постулаты Клаузиуса и Кельвина логически следуют друг из друга. Показать это совсем не сложно.

Предположим сначала, что неверен постулат Клаузиуса. Тогда существует процесс X, единственным результатом которого является передача тепла Q от менее нагретого тела L к более нагретому телу M.

Возьмём тепловой двигатель, нагревателем которого является тело M, а холодильником — тело L. Двигатель подберём так, чтобы за цикл к холодильнику поступало как раз тепло Q. Совершим один цикл, в ходе которого двигатель производит работу A, после чего осуществим процесс X. Тепло Q вернётся к телу M.

Тогда окажется, что в результате циклического процесса наш двигатель произвёл работу A единственно за счёт тепла, отведённого от нагревателя M. Следовательно, неверен постулат Кельвина.

Наоборот, предположим, что неверен постулат Кельвина. Тогда существует циклический процесс Y, единственным результатом которого является получение работы A за счёт тепла, отводимого от некоторого резервуара L.

Возьмём тело M — более горячее, чем L. Запустим процесс Y. От тела L будет отведено тепло Q. Полученную работу $A = Q$ превратим во внутреннюю энергию тела M (например, с помощью трения).

В итоге мы осуществим процесс, единственным результатом которого является передача тепла Q от менее нагретого тела L к более нагретому телу M. Никаких других изменений не произошло (ведь процесс Y — циклический!). Стало быть, постулат Клаузиуса неверен.

2.13.4 Обратимые процессы

В разделе «Тепловые машины» мы отметили, что изотермические и адиабатные процессы делают машину Карно обратимой, но не объяснили — почему.

Кроме того, мы сказали, что все реальные процессы необратимы. Как это увязать с упомянутой обратимостью цикла Карно?

Прежде всего, нужно рассказать о важнейшей (наряду с идеальным газом) идеализации в термодинамике — равновесных обратимых процессах.

Равновесное состояние газа (и вообще любой термодинамической системы) — это состояние теплового равновесия[24]. В этом состоянии определены значения макроскопических параметров:

[24]Вспоминайте раздел «Температура»!

давления, объёма, температуры... Равновесное состояние может быть изображено точкой на диаграммах состояния (т. е. pV-, VT- и pT-диаграммах).

Равновесный процесс представляет собой последовательную цепочку близких равновесных состояний. Равновесный процесс может быть изображён непрерывной линией на диаграммах состояния.

Как вы теперь понимаете, все процессы, графики которых мы рассматривали в предыдущих листках, предполагались равновесными — ведь только в этом случае можно говорить о «графике процесса».

Процесс можно считать равновесным, если параметры газа меняются достаточно медленно. Настолько медленно, что в каждый момент времени будет успевать устанавливаться новое равновесное состояние. Такие медленные равновесные процессы называют ещё *квазистатическими*.

Вот пример квазистатического (равновесного) процесса. Пусть имеется вертикальный цилиндр с газом под поршнем. Положим на поршень песчинку. Потом другую, третью, сотую, тысячную. Получим чрезвычайно медленное сжатие газа, представляющее собой череду сменяющих друг друга равновесных состояний.

Теперь начнём снимать песчинки обратно, всё так же по одной. Получим квазистатическое равновесное расширение газа. При этом газ пройдёт в обратном направлении *те же самые* состояния, которые он проходил в предыдущем процессе равновесного сжатия. Действительно, раз уж при каждом положении поршня успевает установиться тепловое равновесие, то значения макроскопических параметров определяются только самим положением поршня, но не направлением его движения.

Мы видим, что равновесный процесс является *обратимым* — его можно провести в обратном направлении через ту же самую цепочку равновесных состояний. На диаграммах состояния обратимый процесс идёт по одной и той же линии как в прямом, так и в обратном направлении.

Если же вместо песчинок поставить на поршень увесистую гирьку, то процесс сжатия газа пойдёт весьма быстро. Давление газа непосредственно под поршнем будет больше, чем у дна сосуда, и мы уже не сможем охарактеризовать состояние газа в каждый момент времени каким-то одним значением давления. Состояния, проходимые газом, не будут состояниями теплового равновесия (макроскопические параметры не успевают принимать определённые значения для всего газа). Стало быть, процесс быстрого сжатия газа будет *неравновесным*.

Кроме того, такой процесс будет *необратимым*. Ведь если столь же быстро расширять газ, то давление непосредственно под поршнем теперь окажется меньше, чем у дна сосуда. Следовательно, при быстром расширении газ проходит через иную цепочку состояний, чем в процессе сжатия.

Процессы, идущие бесконечно медленно, являются обратимыми. Это идеализация. Реальные процессы идут с конечной скоростью и потому необратимы.

2.13.5 Обратимость машины Карно

Пусть некоторая тепловая машина осуществляет цикл между нагревателем с фиксированной температурой T_1 и холодильником с фиксированной температурой T_2. Как сделать этот цикл обратимым?

Разумеется, все процессы цикла должны быть квазистатическими — наша машина будет работать бесконечно медленно[25]. Но этого не достаточно. То, что машина обязана работать в прямом и обратном направлении (т. е. как тепловой двигатель и как холодильная машина) *между одними и теми же нагревателем и холодильником*, налагает дополнительные требования.

[25] Мощность такой машины равна нулю!

При работе в прямом направлении рабочее тело (газ) получает от нагревателя тепло Q_1 и отдаёт холодильнику тепло Q_2. Эти процессы теплопередачи должны быть обратимы: в самом деле, при работе в обратном направлении газ должен забрать у холодильника тепло Q_2 и отдать нагревателю тепло Q_1, проходя через те же самые состояния, что и в прямом направлении.

Можно ли сделать теплопередачу обратимой? Ведь если температура газа будет отличаться от температуры нагревателя (холодильника), передача тепла будет необратима в силу постулата Клаузиуса!

Единственный выход состоит в том, чтобы тепловое взаимодействие газа с нагревателем и холодильником происходило *изотермически*. В ходе теплообмена газ имеет ту же температуру, что и нагреватель (холодильник), и бесконечно медленно, обратимым образом обменивается с ними теплом. Вот откуда берутся изотермы в цикле Карно: только они и могут соответствовать тепловому контакту газа с нагревателем и холодильником.

Изменение температуры газа в цикле тоже должно происходить обратимо; значит, и тут нужно исключать необратимый теплообмен. Единственная возможность — теплоизолировать газ и использовать адиабатные процессы. Если проводить их квазистатически, они будут обратимыми.

Таким образом, цикл Карно — это единственный циклический процесс, который можно осуществлять обратимым образом между данными нагревателем и холодильником с фиксированными температурами.

Разумеется, машина Карно является идеализацией — уже потому, что использует бесконечно медленные процессы. Поэтому её часто называют *идеальной тепловой машиной.*

Глава 3

Электродинамика

Электродинамика имеет дело с широким кругом явлений, связанных с электромагнитным взаимодействием тел. Основным понятием электродинамики служит *электромагнитное поле*, посредством которого осуществляется это взаимодействие.

В общем случае электромагнитное поле представляет собой сложную комбинацию электрического и магнитного полей, меняющихся во времени и порождающих друг друга. Электромагнитное поле описывается уравнениями Максвелла; рассмотрение этих уравнений выходит далеко за рамки школьной программы.

Наиболее естественный путь изучения электродинамики — постепенное восхождение от простого к сложному. На данном пути мы проходим следующие этапы.

1. Электростатика. В электростатике рассматриваются *неподвижные* электрические заряды. Они создают *постоянное* электрическое поле (которое называется ещё электростатическим); магнитное поле в этой ситуации не возникает.

В электростатике вводятся важнейшие понятия напряжённости и потенциала электрического поля, а также напряжения (или разности потенциалов) между двумя точками. Электростатике посвящены следующие разделы:

→ Электрический заряд
→ Закон Кулона
→ Напряжённость электрического поля
→ Потенциал электрического поля
→ Проводники в электрическом поле
→ Диэлектрики в электрическом поле
→ Конденсатор. Энергия электрического поля

2. Постоянный ток. Теперь электрические заряды начинают синхронно двигаться с постоянной скоростью. Возникает постоянный электрический ток, который рассматривается в следующих разделах:

→ Постоянный электрический ток
→ Закон Ома
→ Соединения проводников
→ Работа и мощность тока
→ ЭДС. Закон Ома для полной цепи
→ Электрический ток в металлах
→ Электрический ток в электролитах
→ Электрический ток в газах
→ Полупроводники

3. Магнитное поле. Оказывается, что магнитное поле порождается только движущимися зарядами (в частности, электрическим током). Действует магнитное поле также лишь на дви-

жущиеся заряды (в частности, на проводники с током). Имеем два раздела, посвящённых магнитному полю:

→ Магнитное поле. Линии
→ Магнитное поле. Силы

4. Электромагнитная индукция и электромагнитные колебания. До сих пор электрическое и магнитное поля изучались независимо друг от друга. Оказывается, однако, что они тесно связаны. Фарадей открыл явление электромагнитной индукции: переменное магнитное поле порождает вихревое электрическое поле. Важнейший круг вытекающих отсюда вопросов изложен в разделах:

→ Электромагнитная индукция
→ Самоиндукция
→ Электромагнитные колебания
→ Переменный ток. 1
→ Переменный ток. 2
→ Мощность переменного тока
→ Электроэнергия

5. Электромагнитное поле. Выяснилось, что природа обладает замечательной симметрией: 1) переменное магнитное поле порождает электрическое поле; 2) переменное электрическое поле порождает магнитное поле. Взаимное превращение электрического и магнитного полей друг в друга говорит о том, что на самом деле имеется единый физический объект, называемый электромагнитным полем.

Электромагнитное поле подчиняется системе уравнений Максвелла. Чтобы понять эти уравнения, нужно не меньше года изучать математику в хорошем вузе. Поэтому в школьной программе электромагнитное поле рассматривается лишь в самых общих чертах:

→ Электромагнитное поле
→ Электромагнитные волны

3.1 Электрический заряд

Электромагнитные взаимодействия принадлежат к числу наиболее фундаментальных взаимодействий в природе. Силы упругости и трения, давление жидкости и газа и многое другое можно свести к электромагнитным силам между частицами вещества. Сами электромагнитные взаимодействия уже не сводятся к другим, более глубоким видам взаимодействий.

Столь же фундаментальным типом взаимодействия является тяготение — гравитационное притяжение любых двух тел. Однако между электромагнитными и гравитационными взаимодействиями имеется несколько важных отличий.

1. Участвовать в электромагнитных взаимодействиях могут не любые, а только *заряженные* тела (имеющие *электрический заряд*).

2. Гравитационное взаимодействие — это всегда притяжение одного тела к другому. Электромагнитные взаимодействия могут быть как притяжением, так и отталкиванием.

3. Электромагнитное взаимодействие гораздо интенсивнее гравитационного. Например, сила электрического отталкивания двух электронов в 10^{42} раз превышает силу их гравитационного притяжения друг к другу.

Каждое заряженное тело обладает некоторой величиной электрического заряда q. *Электрический заряд — это физическая величина, определяющая силу электромагнитного взаимодействия между объектами природы.* Единицей измерения заряда является *кулон* (Кл)[1].

3.1.1 Два вида заряда

Поскольку гравитационное взаимодействие всегда является притяжением, массы всех тел неотрицательны. Но для зарядов это не так. Два вида электромагнитного взаимодействия — притяжение и отталкивание — удобно описывать, вводя два вида электрических зарядов: *положительные* и *отрицательные*.

Заряды разных знаков притягиваются друг к другу, а заряды одного знака друг от друга отталкиваются. Это проиллюстрировано на рис. 3.1; подвешенным на нитях шарикам сообщены заряды того или иного знака.

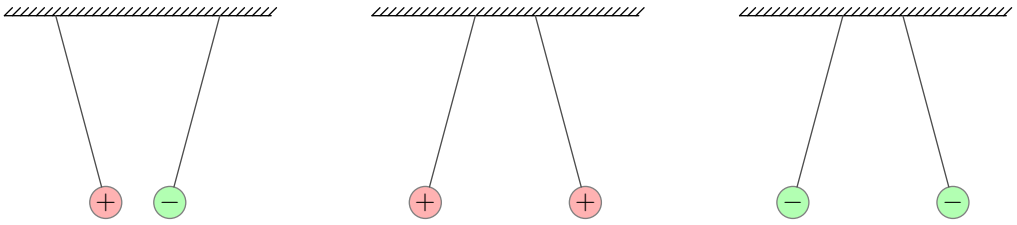


Рис. 3.1. Взаимодействие двух видов зарядов

Повсеместное проявление электромагнитных сил объясняется тем, что в атомах любого вещества присутствуют заряженные частицы: в состав ядра атома входят положительно заряженные протоны, а по орбитам вокруг ядра движутся отрицательно заряженные электроны. Заряды протона и электрона равны по модулю, а число протонов в ядре равно числу электронов на орбитах, и поэтому оказывается, что атом в целом электрически нейтрален. Вот почему в обычных условиях мы не замечаем электромагнитного воздействия со стороны окружающих

[1] Единица измерения заряда определяется через единицу измерения силы тока. 1 Кл — это заряд, проходящий через поперечное сечение проводника за 1 с при силе тока в 1 А.

тел: суммарный заряд каждого из них равен нулю, а заряженные частицы равномерно распределены по объёму тела. Но при нарушении электронейтральности (например, в результате *электризации*) тело немедленно начинает действовать на окружающие заряженные частицы.

Почему существует именно два вида электрических зарядов, а не какое-то другое их число, в данный момент не известно. Мы можем лишь утверждать, что принятие этого факта в качестве первичного даёт адекватное описание электромагнитных взаимодействий.

Заряд протона равен $1{,}6 \cdot 10^{-19}$ Кл. Заряд электрона противоположен ему по знаку и равен $-1{,}6 \cdot 10^{-19}$ Кл. Величина

$$e = 1{,}6 \cdot 10^{-19}\ \text{Кл}$$

называется *элементарным зарядом*. Это минимальный возможный заряд: свободные частицы с меньшей величиной заряда в экспериментах не обнаружены. Физика не может пока объяснить, почему в природе имеется наименьший заряд и почему его величина именно такова.

Заряд любого тела q всегда складывается из *целого* количества элементарных зарядов:

$$q = \pm Ne.$$

Если $q < 0$, то тело имеет избыточное количество N электронов (по сравнению с количеством протонов). Если же $q > 0$, то наоборот, у тела электронов недостаёт: протонов на N больше.

3.1.2 Электризация тел

Чтобы макроскопическое тело оказывало электрическое влияние на другие тела, его нужно электризовать. *Электризация* — это нарушение электрической нейтральности тела или его частей. В результате электризации тело становится способным к электромагнитным взаимодействиям.

Один из способов электризовать тело — сообщить ему электрический заряд, то есть добиться избытка в данном теле зарядов одного знака. Это несложно сделать с помощью трения.

Так, при натирании шёлком стеклянной палочки часть её отрицательных зарядов уходит на шёлк. В результате палочка заряжается положительно, а шёлк — отрицательно. А вот при натирании шерстью эбонитовой палочки часть отрицательных зарядов переходит с шерсти на палочку: палочка заряжается отрицательно, а шерсть — положительно.

Данный способ электризации тел называется *электризацией трением*. С электризацией трением вы сталкиваетесь всякий раз, когда снимаете свитер через голову ;-)

Другой тип электризации называется *электростатической индукцией*, или *электризацией через влияние*. В этом случае суммарный заряд тела остаётся равным нулю, но перераспределяется так, что в одних участках тела скапливаются положительные заряды, в других — отрицательные.

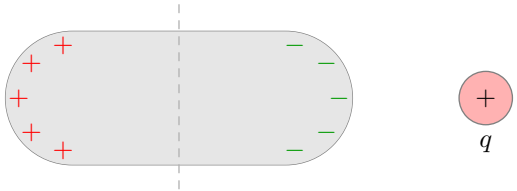


Рис. 3.2. Электростатическая индукция

Давайте посмотрим на рис. 3.2. На некотором расстоянии от металлического тела находится положительный заряд q. Он притягивает к себе отрицательные заряды металла (свободные электроны), которые скапливаются на ближайших к заряду участках поверхности тела. На дальних участках остаются нескомпенсированные положительные заряды.

Несмотря на то, что суммарный заряд металлического тела остался равным нулю, в теле произошло *пространственное разделение зарядов*. Если сейчас разделить тело вдоль пунктирной линии, то правая половина окажется заряженной отрицательно, а левая — положительно.

Наблюдать электризацию тела можно с помощью *электроскопа*. Простой электроскоп показан[2] на рис. 3.3.

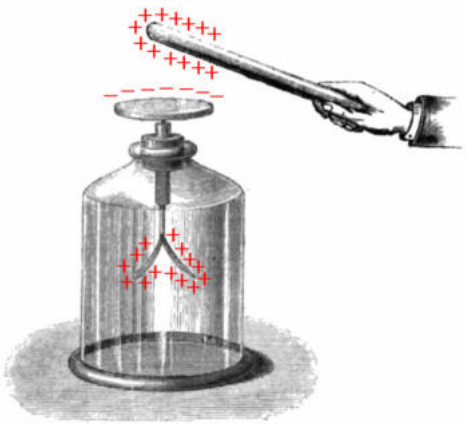


Рис. 3.3. Электроскоп

Что происходит в данном случае? Положительно заряженная палочка (например, предварительно натёртая) подносится к диску электроскопа и собирает на нём отрицательный заряд. Внизу, на подвижных *листочках* электроскопа, остаются нескомпенсированные положительные заряды; отталкиваясь друг от друга, листочки расходятся в разные стороны. Если убрать палочку, то заряды вернутся на место и листочки опадут обратно.

Явление электростатической индукции в грандиозных масштабах наблюдается во время грозы. На рис. 3.4 мы видим идущую над землёй грозовую тучу[3].

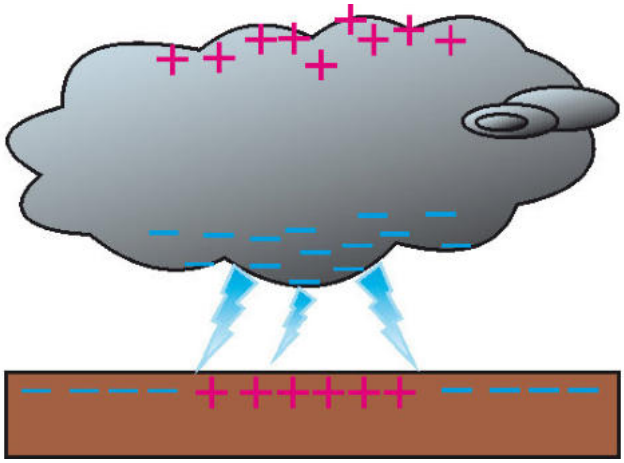


Рис. 3.4. Электризация земли грозовой тучей

Внутри тучи имеются льдинки разных размеров, которые перемешиваются восходящими потоками воздуха, сталкиваются друг с другом и электризуются. При этом оказывается, что в нижней части тучи скапливается отрицательный заряд, а в верхней — положительный.

Отрицательно заряженная нижняя часть тучи наводит под собой на поверхности земли заряды положительного знака. Возникает гигантский конденсатор с колоссальным напряжением

[2]Изображение с сайта en.wikipedia.org.
[3]Изображение с сайта elementy.ru.

между тучей и землёй. Если этого напряжения будет достаточно для пробоя воздушного промежутка, то произойдёт разряд — хорошо известная вам молния.

3.1.3 Закон сохранения заряда

Вернёмся к примеру электризации трением — натирании палочки тканью. В этом случае палочка и кусок ткани приобретают равные по модулю и противоположные по знаку заряды. Их суммарный заряд как был равен нулю до взаимодействия, так и остаётся равным нулю после взаимодействия.

Мы видим здесь *закон сохранения заряда*, который гласит: *в замкнутой системе тел алгебраическая сумма зарядов остаётся неизменной при любых процессах, происходящих с этими телами*:

$$q_1 + q_2 + \ldots + q_n = \text{const}.$$

Замкнутость системы тел означает, что эти тела могут обмениваться зарядами только между собой, но не с какими-либо другими объектами, внешними по отношению к данной системе.

При электризации палочки ничего удивительного в сохранении заряда нет: сколько заряженных частиц ушло с палочки — столько же пришло на кусок ткани (или наоборот). Удивительно то, что в более сложных процессах, сопровождающихся *взаимными превращениями* элементарных частиц и *изменением числа* заряженных частиц в системе, суммарный заряд всё равно сохраняется!

Например, на рис. 3.5 показан процесс $\gamma \to e^- + e^+$, при котором порция электромагнитного излучения γ (так называемый *фотон*) превращается в две заряженные частицы — электрон e^- и позитрон e^+. Такой процесс оказывается возможным при некоторых условиях — например, в электрическом поле атомного ядра.

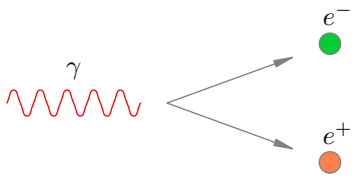


Рис. 3.5. Рождение пары электрон–позитрон

Заряд позитрона равен по модулю заряду электрона и противоположен ему по знаку. Закон сохранения заряда выполнен! Действительно, в начале процесса у нас был фотон, заряд которого равен нулю, а в конце мы получили две частицы с нулевым суммарным зарядом.

Закон сохранения заряда (наряду с существованием наименьшего элементарного заряда) является на сегодняшний день первичным научным фактом. Объяснить, почему природа ведёт себя именно так, а не иначе, физикам пока не удаётся. Мы можем лишь констатировать, что эти факты подтверждаются многочисленными физическими экспериментами.

3.2 Закон Кулона

Взаимодействие неподвижных (в данной инерциальной системе отсчёта) зарядов называется *электростатическим*. Оно наиболее просто для изучения.

Раздел электродинамики, в котором изучается взаимодействие неподвижных зарядов, называется *электростатикой*. Основной закон электростатики — это *закон Кулона*.

По внешнему виду закон Кулона удивительно похож на закон всемирного тяготения, который устанавливает характер гравитационного взаимодействия точечных масс. Закон Кулона является законом электростатического взаимодействия точечных зарядов.

Точечный заряд — это заряженное тело, размеры которого много меньше других размеров, характерных для данной задачи. В частности, размеры точечных зарядов пренебрежимо малы по сравнению с расстояниями между ними.

Точечный заряд — такая же идеализация, как материальная точка, точечная масса и т. д. В случае точечных зарядов мы можем однозначно говорить о расстоянии между ними, не задумываясь о том, между какими именно точками заряженных тел это расстояние измеряется.

Закон Кулона. Сила взаимодействия двух неподвижных точечных зарядов в вакууме прямо пропорциональна произведению абсолютных величин зарядов и обратно пропорциональна квадрату расстояния между ними.

Эта сила называется *кулоновской*. Вектор кулоновской силы всегда лежит на прямой, соединяющей заряды. Для кулоновской силы справедлив третий закон Ньютона: заряды действуют друг на друга с силами, равными по модулю и противоположными по направлению.

В качестве примера на рис. 3.6 показаны силы $\vec{F}_1$ и $\vec{F}_2$, с которыми взаимодействуют два отрицательных заряда.

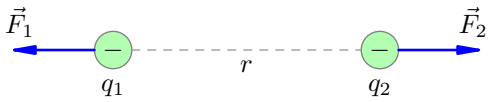


Рис. 3.6. Кулоновская сила

Если заряды, равные по модулю q_1 и q_2, находятся на расстоянии r друг от друга, то они взаимодействуют с силой

$$F = k\frac{q_1 q_2}{r^2}\,. \tag{3.1}$$

Коэффициент пропорциональности k в системе СИ равен:

$$k = 9\cdot 10^9 \frac{\text{Н}\cdot\text{м}^2}{\text{Кл}^2}\,.$$

Если сравнивать с законом всемирного тяготения, то роль точечных масс в законе Кулона играют точечные заряды, а вместо гравитационной постоянной G стоит коэффициент k. Математически формулы этих законов устроены одинаково. Важное физическое отличие заключается в том, что гравитационное взаимодействие всегда является притяжением, а взаимодействие зарядов может быть как притяжением, так и отталкиванием.

Так уж вышло, что наряду с константой k имеется ещё одна фундаментальная константа ε_0, связанная с k соотношением

$$k = \frac{1}{4\pi\varepsilon_0}\,.$$

Константа ε_0 называется *электрической постоянной*. Она равна:

$$\varepsilon_0 = \frac{1}{4\pi k} = 8{,}85\cdot 10^{-12}\frac{\text{Кл}^2}{\text{Н}\cdot\text{м}^2}\,.$$

Закон Кулона с электрической постоянной выглядит так:

$$F = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r^2}\,. \tag{3.2}$$

3.2.1 Принцип суперпозиции

Опыт показывает, что выполнен так называемый *принцип суперпозиции*. Он состоит из двух утверждений.

1. Кулоновская сила взаимодействия двух зарядов не зависит от присутствия других заряженных тел.

2. Предположим, что заряд q взаимодействует с системой зарядов q_1, q_2, …, q_n. Если каждый из зарядов системы действует на заряд q с силой $\vec{F}_1$, $\vec{F}_2$, …, $\vec{F}_n$ соответственно, то результирующая сила $\vec{F}$, приложенная к заряду q со стороны данной системы, равна векторной сумме отдельных сил:

$$\vec{F} = \vec{F}_1 + \vec{F}_2 + \ldots + \vec{F}_n.$$

Принцип суперпозиции проиллюстрирован на рис. 3.7. Здесь положительный заряд q взаимодействует с двумя зарядами: положительным зарядом q_1 и отрицательным зарядом q_2.

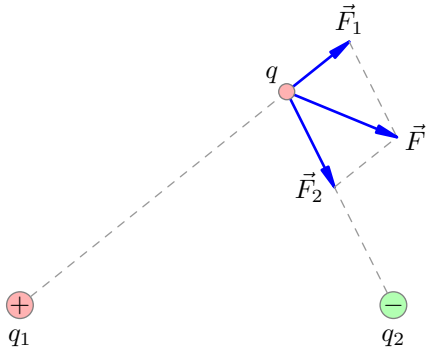


Рис. 3.7. Принцип суперпозиции

Принцип суперпозиции позволяет прийти к одному важному утверждению.

Вы помните, что закон всемирного тяготения справедлив на самом деле не только для точечных масс, но и для шаров со сферически-симметричным распределением массы (в частности, для шара и точечной массы); тогда r — расстояние между центрами шаров (от точечной массы до центра шара). Этот факт вытекает из математической формы закона всемирного тяготения и принципа суперпозиции.

Поскольку формула закона Кулона имеет ту же структуру, что и закон всемирного тяготения, и для кулоновской силы также выполнен принцип суперпозиции, мы можем сделать аналогичный вывод: *по закону Кулона будут взаимодействовать два заряженных шара (точечный заряд с шаром) при условии, что шары имеют сферически-симметричное распределение заряда; величина r в таком случае будет расстоянием между центрами шаров (от точечного заряда до шара).*

Значимость данного факта мы увидим совсем скоро; в частности, именно поэтому напряжённость поля заряженного шара окажется вне шара такой же, как и у точечного заряда.

Но в электростатике, в отличие от гравитации, с этим фактом надо быть осторожным. Например, при сближении положительно заряженных металлических шаров сферическая симметрия нарушится: положительные заряды, взаимно отталкиваясь, будут стремиться к наиболее

удалённым друг от друга участкам шаров (центры положительных зарядов будут находиться дальше друг от друга, чем центры шаров). Поэтому сила отталкивания шаров в данном случае будет *меньше* того значения, которое получится из закона Кулона при подстановке вместо r расстояния между центрами.

3.2.2 Закон Кулона в диэлектрике

Отличие электростатического взаимодействия от гравитационного состоит не только в наличии сил отталкивания. Сила взаимодействия зарядов зависит от среды, в которой заряды находятся (а сила всемирного тяготения от свойств среды не зависит).

Диэлектриками, или *изоляторами* называются вещества, которые не проводят электрический ток.

Оказывается, что диэлектрик уменьшает силу взаимодействия зарядов (по сравнению с вакуумом). Более того, на каком бы расстоянии друг от друга заряды ни находились, сила их взаимодействия в данном однородном диэлектрике всегда будет *в одно и то же число раз* меньше, чем на таком же расстоянии в вакууме. Это число обозначается ε и называется *диэлектрической проницаемостью* диэлектрика. Диэлектрическая проницаемость зависит только от вещества диэлектрика, но не от его формы или размеров. Она является безразмерной величиной и может быть найдена из таблиц.

Таким образом, в диэлектрике формулы (3.1) и (3.2) приобретают вид:

$$F = k\frac{q_1 q_2}{\varepsilon r^2}\,, \quad F = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{\varepsilon r^2}\,.$$

Диэлектрическая проницаемость вакуума, как видим, равна единице. Во всех остальных случаях диэлектрическая проницаемость больше единицы. Диэлектрическая проницаемость воздуха настолько близка к единице, что при расчёте сил взаимодействия зарядов в воздухе пользуются формулами (3.1) и (3.2) для вакуума.

3.3 Напряжённость электрического поля

Основной закон электростатики — закон Кулона — позволяет вычислить силу взаимодействия двух точечных зарядов. Данный закон, однако, ничего не говорит нам о том, *каким образом* осуществляется это взаимодействие. Как так получается, что один заряд может действовать на другой даже на весьма большом расстоянии?

3.3.1 Дальнодействие и близкодействие

Опыт показывает, что электрические заряды действуют друг на друга даже в отсутствие между ними вещества, то есть в вакууме. Поэтому долгое время в науке преобладала *теория дальнодействия*. Эта теория утверждала, что один заряд действует на другой непосредственно, без участия какого-то промежуточного агента. В частности, изменение взаимного расположения зарядов приводит к мгновенному изменению силы их взаимодействия.

Теория дальнодействия возникла под влиянием небесной механики, основанной на законах Ньютона и законе всемирного тяготения. Движение планет рассчитывалось с большой точностью, и вместе с тем не было объяснения того, *что такое* тяготение. Поэтому господствовала точка зрения, что гравитационные и электромагнитные силы являются врождённым, первичным свойством материи; эти силы нельзя объяснить на основе каких-то других понятий, и единственное, что науке доступно — это описывать их свойства. Основные положения теории дальнодействия были простыми и ясными, а сама теория обладала изяществом и математической строгостью. Этой теории придерживалось абсолютное большинство учёных первой половины XIX столетия.

Противоположной точкой зрения служила *теория близкодействия*. Согласно этой теории, для взаимодействия тел нужен промежуточный агент — физический объект, передающий взаимодействие от одной точки пространства к другой. В частности, скорость передачи взаимодействий конечна: при изменении положения одного из зарядов другой заряд «почувствует» это изменение не сразу, а спустя некоторый интервал времени.

Что же это за промежуточный объект, передающий взаимодействие, если заряды могут действовать друг на друга сквозь пустоту? Данный вопрос был одним из главных возражений сторонников дальнодействия, среди которых были крупнейшие физики и математики своего времени (Кулон, Ампер, Лаплас, Пуассон, Гаусс, Вебер, Кирхгоф).

3.3.2 Электрическое поле

Тем не менее, теория близкодействия одержала верх. Физическим объектом, передающим взаимодействие между зарядами даже сквозь пустоту, оказалось *электромагнитное поле*. Решающими здесь оказались идеи и труды двух великих учёных XIX столетия — Фарадея и Максвелла. Экспериментальным подтверждением теории близкодействия явилось открытие электромагнитных волн.

Неподвижные заряды не создают магнитного поля; поэтому, пока мы изучаем электростатику, мы будем говорить только об электрическом поле. Итак:

Электрический заряд создаёт вокруг себя электрическое поле, которое, в свою очередь, действует с некоторой силой на другие заряды.

Электрическое поле не нуждается в какой-то специальной среде, которая являлась бы его носителем. Оно может возникать как в веществе, так и в вакууме, и является, наряду с веществом, альтернативной формой существования материи.

По современным физическим представлениям электрическое поле является первичным физическим объектом: мы пока не можем сказать, каково его внутреннее устройство (точно так же мы не можем сказать, например, из чего состоит электрон). Мы можем лишь изучать свойства

электрического поля, устанавливать законы его поведения и использовать эти законы в своих целях.

Источниками электрического поля являются электрические заряды. Индикатором для обнаружения поля также является электрический заряд — так называемый *пробный заряд*. По действию на пробный заряд мы и можем судить о наличии электрического поля в данной области пространства. Кроме того, с помощью пробного заряда мы можем исследовать величину поля в различных пространственных точках. Разумеется, для этого пробный заряд должен быть точечным.

Опыт показывает, что сила, с которой поле действует на пробный заряд, прямо пропорциональна величине заряда. Поэтому отношение силы к заряду уже не зависит от величины заряда и является характеристикой поля.

Напряжённость электрического поля — это отношение вектора силы $\vec{F}$, с которой поле действует на пробный заряд q, к самому пробному заряду (с учётом его знака):

$$\vec{E} = \frac{\vec{F}}{q}\,. \tag{3.3}$$

Напряжённость поля, как видим, является векторной величиной. В каждой точке пространства электрическое поле характеризуется вектором напряжённости. Поле считается заданным, если нам известна зависимость вектора напряжённости от координат точки и, вообще говоря, от времени.

Как следует из определения, напряжённость измеряется в Н/Кл. Общепринятая единица напряжённости есть В/м. Мы скоро увидим, что это одно и то же.

Если напряжённость поля известна, то формула (3.3) позволяет найти силу, которая действует на точечный заряд со стороны электрического поля:

$$\vec{F} = q\vec{E}.$$

Сила и напряжённость, таким образом, являются коллинеарными векторами. Если заряд положительный, то сила направлена в ту же сторону, что и напряжённость. Если заряд отрицательный, то сила направлена противоположно напряжённости.

Одна из основных задач электростатики – нахождение напряжённости поля, создаваемого данной системой зарядов. Рассмотрим некотрорые примеры.

3.3.3 Напряжённость поля точечного заряда

Определение модуля и направления вектора напряжённости поля точечного заряда — это самая простая и легко решаемая задача.

Рассмотрим положительный точечный заряд q, находящийся в вакууме. Поместим на расстоянии r от него положительный пробный заряд q_0. Со стороны заряда q на пробный заряд действует сила отталкивания, поэтому *напряжённость поля положительного заряда q направлена* ***от него*** (рис. 3.8):

Рис. 3.8. Напряжённость поля положительного заряда

Величина силы отталкивания равна:

$$F = \frac{kqq_0}{r^2}\,.$$

Делим силу на пробный заряд q_0 и находим модуль напряжённости поля заряда q:

$$E = \frac{kq}{r^2}\,. \tag{3.4}$$

Пусть теперь заряд, создающий поле, будет отрицательным; модуль этого заряда также обозначаем q. Сила, действующая на положительный пробный заряд, станет силой притяжения. Поэтому *напряжённость поля отрицательного заряда направлена* ***к нему*** (рис. 3.9):

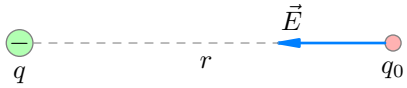


Рис. 3.9. Напряжённость поля отрицательного заряда

Модуль напряжённости поля снова находится по формуле (3.4).

Если заряд q находится в среде с диэлектрической проницаемостью ε, то сила его действия на пробный заряд уменьшается в ε раз:

$$F = \frac{kqq_0}{\varepsilon r^2}\,.$$

Следовательно, в ε раз уменьшается и напряжённость поля:

$$E = \frac{kq}{\varepsilon r^2}\,. \tag{3.5}$$

Модуль напряжённости поля точечного заряда q находится по формуле (3.4) в вакууме и по формуле (3.5) в диэлектрической среде. Вектор напряжённости в данной точке направлен вдоль прямой, соединяющей точку с зарядом: от заряда при $q > 0$ и к заряду при $q < 0$.

По мере удаления от заряда модуль напряжённости поля убывает пропорционально квадрату расстояния от точки наблюдения до заряда. На рис. 3.10 дано примерное графическое представление электрического поля точечного заряда в пространстве (показаны векторы напряжённости поля в различных точках).

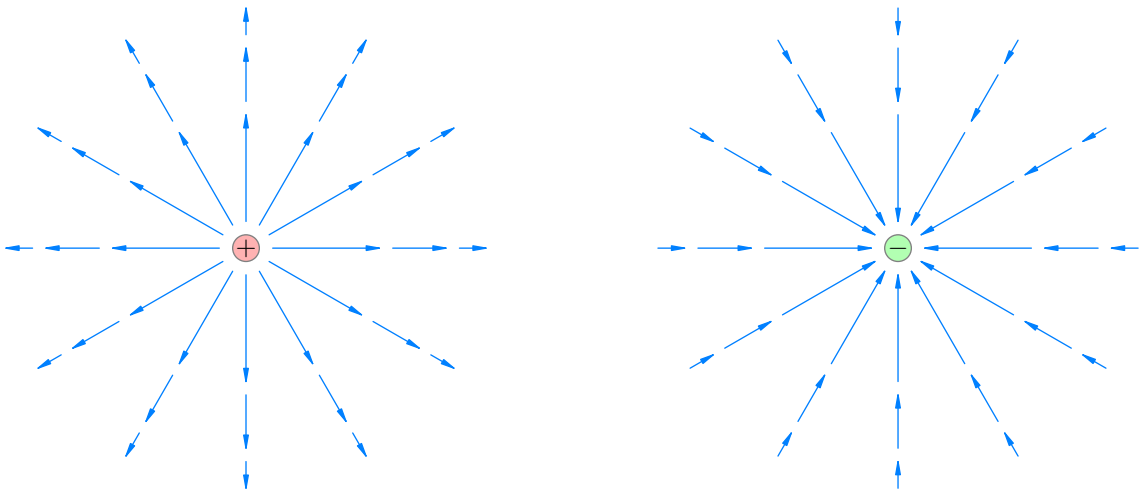


Рис. 3.10. Векторы напряжённости поля точечного заряда

Рябит в глазах, не правда ли? Ниже мы познакомимся с более удобным способом изображения поля — линиями напряжённости.

3.3.4 Принцип суперпозиции электрических полей

Начнём со случая двух зарядов. Пусть положительный заряд q_1 создаёт в точке M электрическое поле $\vec{E}_1$, и пусть отрицательный заряд q_2 создаёт в этой же точке поле $\vec{E}_2$. Какое поле создают в точке M оба заряда вместе?

Поместим в точку M пробный заряд q. Тогда со стороны заряда q_1 на него будет действовать сила $\vec{F}_1$, а со стороны заряда q_2 — сила $\vec{F}_2$. Согласно принципу суперпозиции, с которым вы познакомились в предыдущем разделе, на заряд q действует результирующая сила

$$\vec{F} = \vec{F}_1 + \vec{F}_2.$$

Поделим данное равенство на пробный заряд q:

$$\frac{\vec{F}}{q} = \frac{\vec{F}_1}{q} + \frac{\vec{F}_2}{q}.$$

С учётом определения (3.3) напряжённости поля получаем:

$$\vec{E} = \vec{E}_1 + \vec{E}_2.$$

Таким образом, напряжённость результирующего поля в точке M оказывается равна векторной сумме напряжённостей полей каждого из зарядов (рис. 3.11):

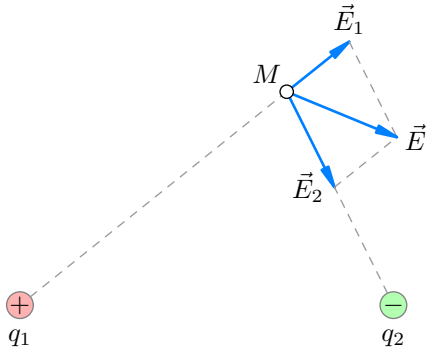


Рис. 3.11. Принцип суперпозиции полей

Напряжённости полей складываются векторно и в общем случае. В самом деле, поделив формулу $\vec{F} = \vec{F}_1 + \vec{F}_2 + \ldots + \vec{F}_n$ на пробный заряд q, приходим к общей формулировке принципа суперпозиции.

Принцип суперпозиции. Пусть заряды q_1, q_2, …, q_n по отдельности создают в данной точке поля $\vec{E}_1$, $\vec{E}_2$, …, $\vec{E}_n$. Тогда система этих зарядов создаёт в данной точке поле $\vec{E}$, равное векторной сумме напряжённостей полей отдельных зарядов:

$$\vec{E} = \vec{E}_1 + \vec{E}_2 + \ldots + \vec{E}_n.$$

С помощью принципа суперпозиции можно найти напряжённость поля любой системы зарядов — разбивая систему на малые заряды, которые можно считать точечными, с последующим суммированием напряжённостей малых зарядов. Обычно это приводит к достаточно сложным вычислениям.

3.3.5 Поле равномерно заряженной плоскости

Важным примером системы зарядов является заряженная плоскость. В качестве бесконечной плоскости мы можем рассматривать любую плоскую пластину, если расстояние от точки, в которой ищется поле, до пластины много меньше размеров самой пластины.

Заряженная плоскость характеризуется величиной поверхностной плотности заряда. Что это такое? Возьмём небольшой участок плоскости площадью S. Пусть заряд этого участка равен q. Тогда *поверхностная плотность заряда* определяется как отношение заряда к площади:

$$\sigma = \frac{q}{S}\,.$$

Иными словами, поверхностная плотность заряда — это заряд единицы площади.

Поверхностная плотность заряда может меняться от участка к участку. Но если на любом участке плоскости поверхностная плотность заряда одинакова ($\sigma = \text{const}$, т. е. заряд распределён равномерно), то плоскость называется *равномерно заряженной*.

Вектор напряжённости поля равномерно заряженной плоскости перпендикулярен плоскости; он направлен *от плоскости*, если плоскость заряжена положительно, и *к плоскости*, если плоскость заряжена отрицательно (рис. 3.12).

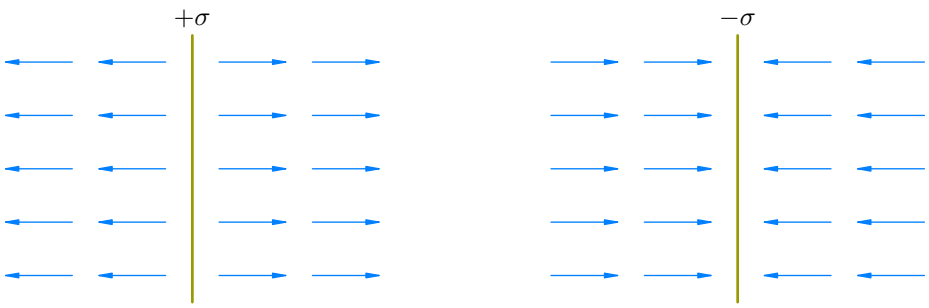


Рис. 3.12. Поле равномерно заряженной плоскости

Самое удивительное заключается в том, что величина напряжённости поля *не зависит* от расстояния до плоскости. Она оказывается равна:

$$E = \frac{\sigma}{2\varepsilon_0}\,. \tag{3.6}$$

Эта формула справедлива для вакуума (мы принимаем её без доказательства). В среде с диэлектрической проницаемостью ε поле, как обычно, уменьшается в ε раз:

$$E = \frac{\sigma}{2\varepsilon_0\varepsilon}\,. \tag{3.7}$$

Пример заряженной плоскости важен потому, что мы встречаемся здесь с понятием однородного поля. *Электрическое поле в данной области пространства называется однородным, если вектор напряжённости поля одинаков в каждой точке области.* Иными словами, напряжённость поля в каждой точке рассматриваемой области имеет одно и то же направление и неизменную величину.

Поле точечного заряда, например, не является однородным. В самом деле, напряжённость поля точечного заряда может меняться от точки к точке как по величине, так и по направлению (она обратно пропорциональна квадрату расстояния до заряда и направлена вдоль прямой, соединяющей заряд с точкой наблюдения).

А вот *заряженная плоскость создаёт однородное электрическое поле в каждом из полупространств, на которые она разбивает пространство.* Напряжённость этого поля вычисляется по формулам (3.6) или (3.7).

3.3.6 Линии напряжённости электрического поля

Давайте вернёмся к пространственной картине поля точечного заряда. Вместо векторов напряжённости в разных точках нарисуем более приятные глазу *линии напряжённости* (рис. 3.13):

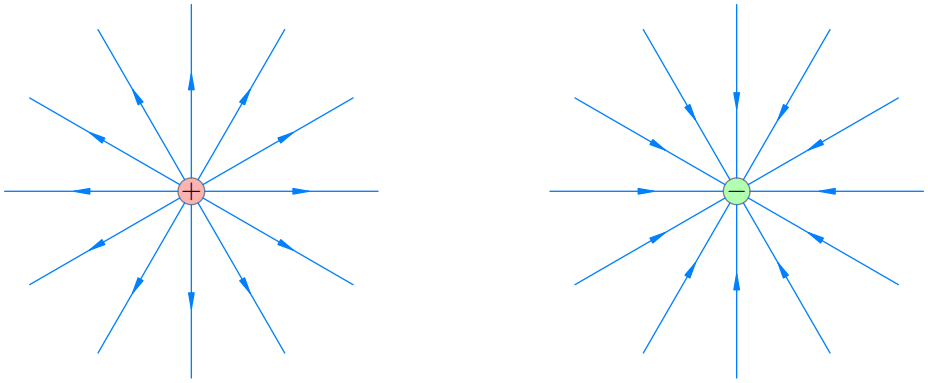


Рис. 3.13. Линии напряжённости поля точечного заряда

Линии напряжённости идут вдоль векторов напряжённости, указывают направление этих векторов и даже содержат информацию об их абсолютных величинах: чем гуще расположены линии напряжённости, тем больше величина напряжённости поля в данной области пространства.

Аналогичную картину линий напряжённости мы можем нарисовать и для заряженной плоскости (рис. 3.14). Как видим, *линии напряжённости однородного поля являются участками параллельных прямых.*

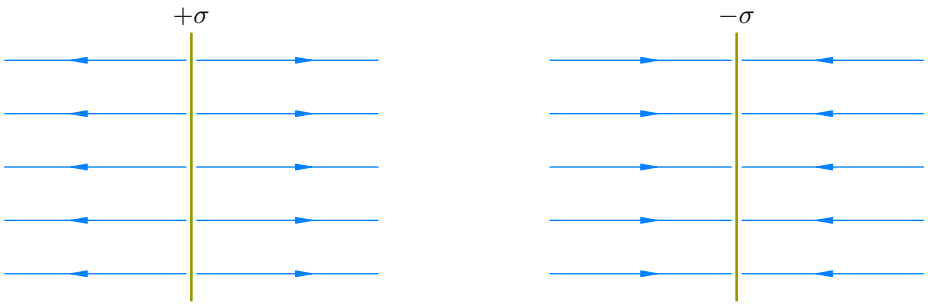


Рис. 3.14. Линии напряжённости поля заряженной плоскости

Линии напряжённости можно провести и в произвольном электрическом поле. Каким образом? *В каждой точке пространства вектор напряжённости поля направлен по касательной к линии напряжённости.* Линии напряжённости как бы «подстраиваются» под векторы напряжённости, «обтекая» их по касательной (рис. 3.15):

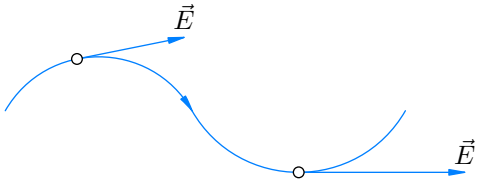


Рис. 3.15. Линия напряжённости

Линии напряжённости всегда начинаются на положительных зарядах и заканчиваются на отрицательных.

3.4 Потенциал электрического поля

Мы начнём с обсуждения потенциальной энергии, которую имеет заряд в электростатическом поле. Прежде всего необходимо вспомнить, при каких условиях можно вообще ввести понятие потенциальной энергии.

3.4.1 Консервативные силы

Сила называется *консервативной* (или *потенциальной*), если работа этой силы не зависит от формы траектории и определяется только начальным и конечным положением тела.

Пусть, например, тело под действием консервативной силы $\vec{F}$ переместилось из начальной точки 1 в конечную точку 2 (рис. 3.16). Тогда работа A силы $\vec{F}$ зависит только от положения самих точек 1 и 2, но не от траектории движения тела. Например, для траекторий $1 \to a \to 2$ и $1 \to b \to 2$ величина A будет одинаковой.

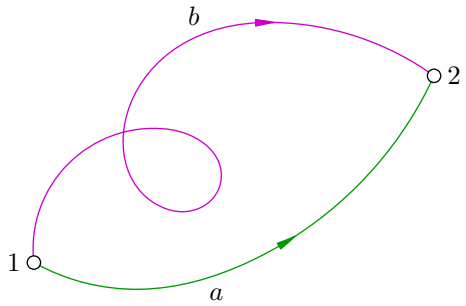


Рис. 3.16. К понятию консервативной силы

Отметим, что *работа консервативной силы по любому замкнутому пути равна нулю*. Действительно, давайте выйдем из точки 1 по траектории $1 \to a \to 2$ и вернёмся назад по траектории $2 \to b \to 1$. На первой траектории сила совершит работу A, а на второй траектории работа будет равна $-A$. В итоге суммарная работа окажется нулевой.

Так вот, понятие потенциальной энергии можно ввести только в случае консервативной силы. *Потенциальная энергия* W — это математическое выражение, зависящее от координат тела, такое, что работа силы равна изменению этого выражения со знаком минус:

$$A = -\Delta W. \tag{3.8}$$

Или, что то же самое:

$$A = -(W_2 - W_1) = W_1 - W_2.$$

Как видим, работа консервативной силы есть разность значений потенциальной энергии, вычисленных соответственно для начального и конечного положений тела.

Примеры консервативных сил вам хорошо известны. Например, сила тяжести является консервативной. Сила упругости пружины тоже консервативна. Именно поэтому мы можем говорить о потенциальной энергии тела, поднятого над землёй, или о потенциальной энергии деформированной пружины.

А вот сила трения не консервативна: работа силы трения зависит от формы траектории и не равна нулю на замкнутом пути. Поэтому не существует никакой «потенциальной энергии тела в поле силы трения».

3.4.2 Потенциальность электростатического поля

Оказывается, что сила, с которой электростатическое поле действует на заряженное тело, также является консервативной. Работа этой силы, совершаемая при перемещении заряда, называется *работой электростатического поля*. Имеем, таким образом, важнейший факт:

Работа электростатического поля не зависит от формы траектории, по которой перемещается заряд, и определяется лишь начальным и конечным положениями заряда. Работа поля по замкнутому пути равна нулю.

Этот факт называется также *потенциальностью* электростатического поля. Как и поле силы тяжести, электростатическое поле является *потенциальным*. Работа электростатического поля одинакова для всех путей, по которым заряд может двигаться из одной фиксированной точки пространства в другую.

Строгое математическое доказательство потенциальности электростатического поля выходит за рамки школьной программы. Однако «на физическом уровне строгости» мы можем убедиться в справедливости этого факта с помощью следующего простого рассуждения.

Нетрудно видеть, что если бы электростатическое поле не было потенциальным, то можно было бы построить вечный двигатель! В самом деле, тогда существовала бы замкнутая траектория, при перемещении заряда по которой поле совершало бы положительную работу (и при этом никаких изменений в окружающих телах не происходило бы). Крутим себе заряд по этой траектории, черпаем неограниченное количество энергии ниоткуда — и все энергетические проблемы человечества решены :-) Но такого, увы, не наблюдается — это вопиющим образом противоречит закону сохранения энергии.

Так как электростатическое поле потенциально, мы можем говорить о потенциальной энергии заряда в этом поле. Начнём с простого и важного случая.

3.4.3 Потенциальная энергия заряда в однородном поле

Потенциальная энергия тела, поднятого над землёй, равна mgh. Случай заряда в однородном поле оказывается очень похожим на эту механическую ситуацию.

Рассмотрим однородное электростатическое поле E, линии напряжённости которого направлены вдоль оси X (рис. 3.17). Пусть положительный заряд q перемещается вдоль силовой линии из точки 1 (с координатой x_1) в точку 2 (с координатой x_2).

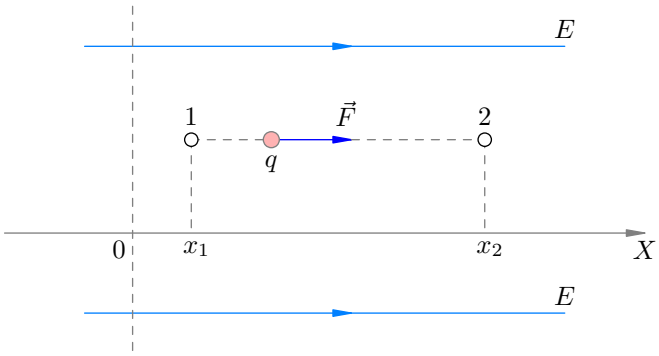


Рис. 3.17. Перемещение заряда в однородном поле

Поле действует на заряд с силой $\vec{F}$, которая направлена вдоль линий напряжённости. Работа этой силы, как легко видеть, будет равна:

$$A = F(x_2 - x_1) = qE(x_2 - x_1).$$

Что изменится, если точки 1 и 2 не лежат на одной линии напряжённости? Оказывается, ничего! Формула для работы поля останется той же самой. Убедимся в этом с помощью рис. 3.18.

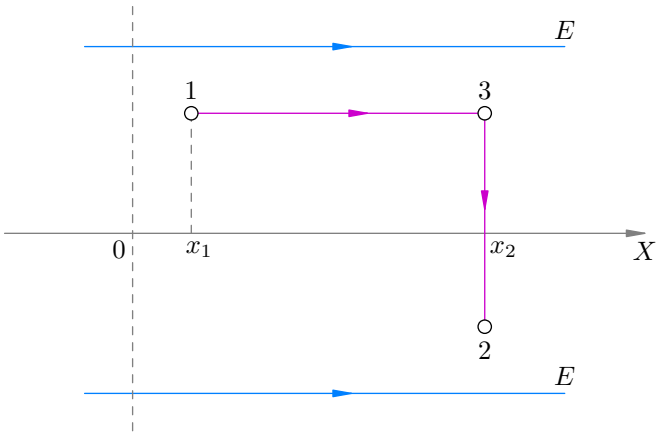


Рис. 3.18. Перемещение заряда в однородном поле

Двигаясь из точки 1 в точку 2, давайте выберем путь $1 \to 3 \to 2$, где точка 3 лежит на одной силовой линии с точкой 1. Тогда работа A_{32} на участке 32 равна нулю — ведь мы перемещаемся перпендикулярно силе. В результате получим:

$$A = A_{13} + A_{32} = A_{13} = qE(x_2 - x_1).$$

Мы видим, что работа поля зависит лишь от *абсцисс* начального и конечного положений заряда.

Запишем полученную формулу следующим образом:

$$A = qEx_2 - qEx_1 = -((-qEx_2) - (-qEx_1)) = -(W_2 - W_1) = -\Delta W.$$

Здесь $W_1 = -qEx_1$, $W_2 = -qEx_2$. Работа поля, в соответствии с формулой (3.8), оказывается равна изменению со знаком минус величины

$$W = -qEx. \tag{3.9}$$

Эта величина и есть *потенциальная энергия заряда в однородном электростатическом поле*. Через x обозначена абсцисса точки, в которой ищется потенциальная энергия. Нулевой уровень потенциальной энергии в данном случае соответствует началу координат $x = 0$ и на рисунках изображён пунктирной линией, перпендикулярной линиям напряжённости[4].

Напомним, что пока считается $q > 0$. Из формулы (3.9) следует, что при движении заряда вдоль силовой линии потенциальная энергия убывает с ростом x. Это естественно: ведь поле совершает положительную работу, разгоняя заряд, а кинетическая энергия заряда растёт за счёт убыли его потенциальной энергии.

Несложно показать, что формула (3.9) остаётся справедливой и для $q < 0$. В этом случае потенциальная энергия возрастает с ростом x. Это тоже понятно: ведь сила, с которой поле действует на заряд, теперь будет направлена влево, так что движение заряда вправо будет осуществляться против действия поля. Заряд тормозится полем, кинетическая энергия заряда уменьшается, а потенциальная энергия — увеличивается.

Итак, важный вывод: в формуле для потенциальной энергии через q обозначается *алгебраическая* величина заряда (с учётом знака), а не его модуль.

[4]На самом деле нулевой уровень потенциальной энергии можно выбирать где угодно. Иными словами, потенциальная энергия определена лишь с точностью до произвольной аддитивной постоянной C, т. е. $W = -qEx + C$. Ничего страшного в такой неопределённости нет: физическим смыслом обладает на потенциальная энергия сама по себе, а разность потенциальных энергий, равная работе поля. В этой разности константа C сократится.

3.4.4 Потенциальная энергия взаимодействия точечных зарядов

Пусть два точечных заряда q_1 и q_2 находятся в вакууме на расстоянии r друг от друга. Можно показать, что потенциальная энергия их взаимодействия даётся формулой:

$$W = \frac{kq_1q_2}{r}\,. \tag{3.10}$$

Мы принимаем формулу (3.10) без доказательства. Две особенности данной формулы следует обсудить.

Во-первых, где находится нулевой уровень потенциальной энергии? Ведь потенциальная энергия, как видно из формулы (3.10), в нуль обратиться не может. Но на самом деле нулевой уровень существует, и находится он *на бесконечности*. Иными словами, когда заряды расположены бесконечно далеко друг от друга, потенциальная энергия их взаимодействия полагается равной нулю (что логично — в этом случае заряды уже «не взаимодействуют»).

Во-вторых, q_1 и q_2 — это снова *алгебраические* величины зарядов, т. е. заряды с учётом их знака.

Например, потенциальная энергия взаимодействия двух одноимённых зарядов будет положительной. Почему? Если мы отпустим их, они начнут разгоняться и удаляться друг от друга. Их кинетическая энергия возрастает, стало быть потенциальная энергия — убывает. Но на бесконечности потенциальная энергия обращается в нуль, а раз она убывает к нулю, значит — она является положительной.

А вот потенциальная энергия взаимодействия разноимённых зарядов оказывается отрицательной. Действительно, давайте удалим их на очень большое расстояние друг от друга — так что потенциальная энергия равна нулю — и отпустим. Заряды начнут разгоняться, сближаясь, и потенциальная энергия снова убывает. Но если она была нулём, то куда ей убывать? Только в сторону отрицательных значений.

Формула (3.10) помогает также вычислить потенциальную энергию системы зарядов, если число зарядов больше двух. Для этого нужно просуммировать энергии каждой пары зарядов. Мы не будем выписывать общую формулу; лучше проиллюстрируем сказанное простым примером, изображённым на рис. 3.19.

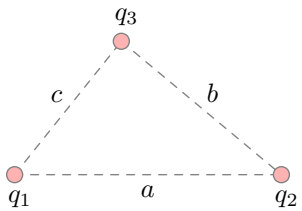


Рис. 3.19. Взаимодействие трёх зарядов

Если заряды q_1, q_2, q_3 находятся в вершинах треугольника со сторонами a, b, c, то потенциальная энергия их взаимодействия равна:

$$W = \frac{kq_1q_2}{a} + \frac{kq_2q_3}{b} + \frac{kq_1q_3}{c}\,.$$

3.4.5 Потенциал

Из формулы $W = -qEx$ мы видим, что потенциальная энергия заряда q в однородном поле прямо пропорциональна этому заряду.

То же самое мы видим из формулы $W = kq_1q_2/r$: потенциальная энергия заряда q_1, находящегося в поле точечного заряда q_2, прямо пропорциональна величине заряда q_1.

Оказывается, это общий факт: потенциальная энергия W заряда q в любом электростатическом поле прямо пропорциональна величине q:

$$W = q\varphi. \tag{3.11}$$

Величина φ уже не зависит от заряда, является характеристикой поля и называется *потенциалом*:

$$\varphi = \frac{W}{q}\,. \tag{3.12}$$

Так, потенциал однородного поля E в точке с абсциссой x равен:

$$\varphi = -Ex. \tag{3.13}$$

Напомним, что ось X совпадает с линией напряжённости поля. Мы видим, что с ростом координаты x потенциал убывает. Иными словами, *вектор напряжённости поля указывает направление убывания потенциала.*

Для потенциала поля точечного заряда q на расстоянии r от него имеем:

$$\varphi = \frac{kq}{r}\,. \tag{3.14}$$

Единицей измерения потенциала служит хорошо известный вам *вольт*. Из формулы (3.12) мы видим, что В = Дж/Кл.

Итак, теперь у нас есть две характеристики поля: силовая (напряжённость) и энергетическая (потенциал). У каждой из них имеются свои преимущества и недостатки. Какую именно характеристику удобнее использовать — зависит от конкретной задачи.

3.4.6 Разность потенциалов

Пусть заряд q перемещается в электростатическом поле из точки 1 в точку 2. Траектория заряда, напомним, роли не играет — работа поля A от этой траектории не зависит и равна разности потенциальных энергий заряда в начальной и конечной точках:

$$A = -\Delta W = -(W_2 - W_1) = W_1 - W_2.$$

С учётом формулы (3.11) имеем:

$$A = q\varphi_1 - q\varphi_2 = q(\varphi_1 - \varphi_2). \tag{3.15}$$

Здесь φ_1 — потенциал поля в точке 1, φ_2 — потенциал поля в точке 2. Величина $\varphi_1 - \varphi_2$, от которой зависит работа поля, так и называется: *разность потенциалов*. Обратите внимание, что *разность потенциалов есть потенциал начальной точки минус потенциал конечной точки*, а не наоборот!

Разность потенциалов называется также *напряжением* между точками 1 и 2 и обозначается через U:

$$U = \varphi_1 - \varphi_2. \tag{3.16}$$

Наряду с формулой (3.15) получаем тогда:

$$A = qU. \tag{3.17}$$

Записывая формулы (3.15) и (3.17) в виде:

$$U = \varphi_1 - \varphi_2 = \frac{A}{q}\,, \tag{3.18}$$

получаем полезное истолкование напряжения: *напряжение (или разность потенциалов) между данными точками — это работа поля по перемещению заряда из начальной точки в конечную, делённая на величину этого заряда.*

Как и потенциальная энергия, потенциал определён с точностью до прибавления произвольной постоянной C: в зависимости от выбора точки, в которой потенциал полагается равным нулю, эта постоянная примет то или иное значение. Но физическим смыслом обладает не потенциал сам по себе, а напряжение (разность потенциалов). При вычитании потенциалов константа C сократится, и напряжение будет уже однозначно определённой величиной, не зависящей от выбора начала отсчёта потенциала.

Выбор точки нулевого потенциала позволяет истолковать в терминах работы сам потенциал. Действительно, пусть 1 — данная точка, 2 — точка нулевого потенциала. Тогда в формуле (3.18) имеем: $\varphi_1 = \varphi$ (потенциал в данной точке), $\varphi_2 = 0$, $A = A_0$ — работа поля по перемещению заряда q из данной точки в точку с нулевым потенциалом. В результате:

$$\varphi = \frac{A_0}{q}\,. \tag{3.19}$$

Таким образом, *потенциал поля в данной точке — это работа поля по перемещению заряда из данной точки в точку нулевого потенциала, делённая на величину этого заряда.*

3.4.7 Принцип суперпозиции для потенциалов

Рассмотрим электрическое поле, создаваемое системой из n заряженных тел. Это поле можно рассматривать как наложение полей, создаваемых каждым телом в отдельности.

Принцип суперпозиции для потенциалов. Пусть φ — потенциал результирующего поля в данной точке, а φ_1, φ_2, …, φ_n — потенциалы полей каждого из тел. Тогда:

$$\varphi = \varphi_1 + \varphi_2 + \ldots + \varphi_n. \tag{3.20}$$

Иными словами, потенциал результирующего поля равен алгебраической сумме потенциалов полей, создаваемых каждым из тел в отдельности.

Принцип суперпозиции для потенциалов вытекает из формулы (3.19) и из того факта, что работа равнодействующей силы есть сумма работ её слагаемых.

3.4.8 Однородное поле: связь напряжения и напряжённости

Предположим, что положительный заряд q перемещается в однородном электростатическом поле по направлению силовой линии из точки 1 в точку 2 (рис. 3.20). Расстояние между точками равно d.

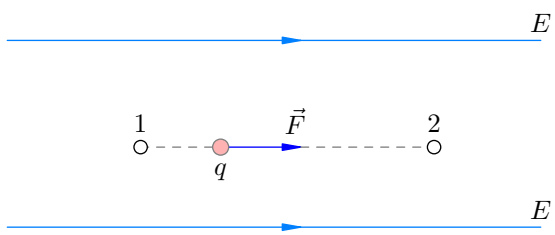


Рис. 3.20. К выводу формулы $U = Ed$

С одной стороны, работа поля равна произведению силы на путь:

$$A = Fd = qEd.$$

Работа получается положительной, так как сила и перемещение сонаправлены.

С другой стороны, работа поля есть произведение заряда на разность потенциалов между точками 1 и 2:

$$A = qU.$$

(Напряжение также положительно, так как $\varphi_1 > \varphi_2$ — ведь напряжённость направлена в сторону убывания потенциала.) Приравнивая правые части последних двух формул, получим: $qU = qEd$, откуда

$$U = Ed. \tag{3.21}$$

Эта простая формула позволяет находить напряжение между точками однородного поля E, находящимися на одной силовой линии; при этом напряжённость поля направлена от начальной точки к конечной.

Выразим из формулы (3.21) напряжённость:

$$E = \frac{U}{d}\,. \tag{3.22}$$

Эта формула пригодится нам впоследствии, при нахождении напряжённости поля в конденсаторе. А сейчас обратим внимание на одно следствие данной формулы: *единицей измерения напряжённости является* В/м. Эта единица используется чаще, чем первоначальная Н/Кл. Что ж, немало вещей пришлось узнать, чтобы понять равенство Н/Кл = В/м :-)

3.4.9 Эквипотенциальные поверхности

Как вы помните, введение силовой характеристики поля (напряжённости) дало возможность изображать поле графически — в виде картины линий напряжённости, или силовых линий.

Энергетическая характеристика поля (потенциал) также позволяет дать графическую картину поля — в виде семейства эквипотенциальных поверхностей.

Поверхность в пространстве называется *эквипотенциальной*, если во всех точках этой поверхности потенциал электрического поля принимает одно и то же значение. Коротко говоря, эквипотенциальные поверхности — это поверхности равного потенциала.

Например, из формулы $\varphi = -Ex$ мы видим, что эквипотенциальными поверхностями однородного поля являются всевозможные плоскости $x = \text{const}$. Они перпендикулярны линиям напряжённости. Так, на рис. 3.21 изображены пять плоскостей — эквипотенциальных поверхностей, отвечающих значениям потенциала φ_1, φ_2, φ_3, φ_4 и φ_5.

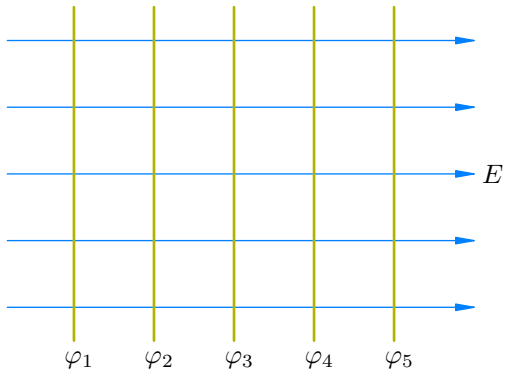


Рис. 3.21. Эквипотенциальные поверхности однородного поля

Теперь рассмотрим нашу вторую стандартную ситуацию: поле точечного заряда $q > 0$. Потенциал этого поля, как мы уже видели, равен:

$$\varphi = \frac{kq}{r}\,.$$

Эквипотенциальными поверхностями здесь будут всевозможные сферы $r = \text{const}$. Они также перпендикулярны линиям напряжённости. На рис. 3.22 показаны четыре такие сферы — эквипотенциальные поверхности, отвечающие значениям потенциала φ_1, φ_2, φ_3 и φ_4.

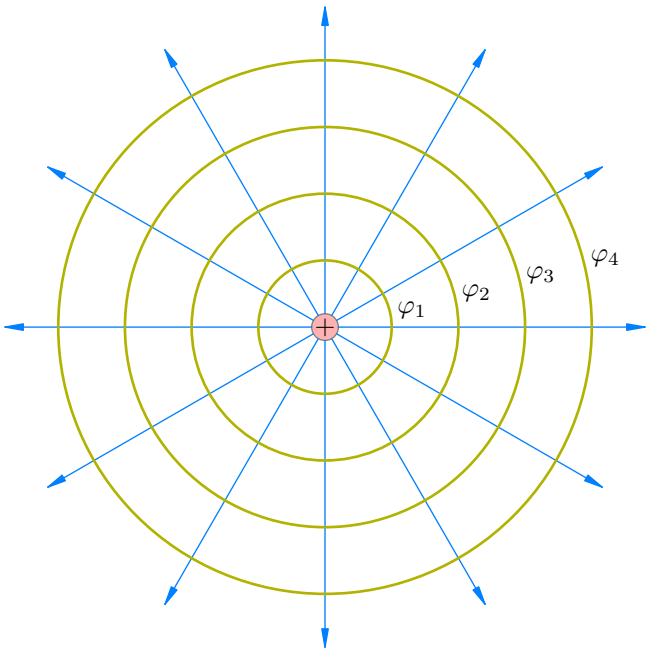


Рис. 3.22. Эквипотенциальные поверхности поля точечного заряда

Оказывается, *эквипотенциальные поверхности всегда перпендикулярны линиям напряжённости*. Нетрудно понять, почему это так. Предположим, что заряд перемещается по эквипотенциальной поверхности. Работа поля при этом равна нулю: $A = q(\varphi_1 - \varphi_2) = 0$, так как $\varphi_1 = \varphi_2$. Значит, угол между перемещением заряда и силой, с которой поле действует на заряд, всё время остаётся прямым. Иными словами, заряд перемещается перпендикулярно вектору напряжённости.

3.5 Проводники в электрическом поле

Если полюса батарейки замкнуть металлической проволокой, по ней пойдёт электрический ток. Заменим проволоку стеклянной палочкой — никакого тока не возникнет. Металл является проводником, а стекло — диэлектриком.

Проводники отличаются от диэлектриков наличием *свободных зарядов* — заряженных частиц, положение которых не связано с какой-то точкой внутри вещества. Свободные заряды приходят в движение под действием электрического поля и могут перемещаться по всему объёму проводника.

Проводники — это в первую очередь металлы. В металлах свободными зарядами являются свободные электроны. Откуда они там берутся? Это особенность металлической связи. Дело в том, что валентный электрон, находящийся на внешней электронной оболочке атома металла, весьма слабо связан с атомом. При взаимодействии атомов металла их валентные электроны покидают свои оболочки, «отправляясь в путешествие» по всему пространству металла[5].

Проводниками являются также электролиты. Так называются растворы и расплавы, свободные заряды в которых возникают в результате диссоциации молекул на положительные и отрицательные ионы. Бросим, например, в стакан воды щепотку поваренной соли. Молекулы NaCl распадутся на ионы Na^+ и Cl^-. Под действием электрического поля эти ионы начнут упорядоченное движение, и возникнет электрический ток.

Природная вода, даже пресная, является проводником из-за растворённых в ней солей[6] (но, конечно, не таким хорошим, как металлы). Человеческое тело в основном состоит из воды, в которой также растворены соли (хлориды натрия, калия, кальция, магния). Поэтому наше тело также служит проводником электрического тока.

Из-за наличия свободных зарядов, способных перемещаться по всему объёму, проводники обладают некоторыми характерными общими свойствами.

3.5.1 Поле внутри проводника

Первое общее свойство проводников в электростатическом поле состоит в том, что *напряжённость поля внутри проводника везде равна нулю.*

Докажем от противного, как в математике. Предположим, что в какой-то области проводника имеется электрическое поле. Тогда под действием этого поля свободные заряды проводника начнут направленное движение. Возникнет электрический ток — а это противоречит тому, что мы находимся в электростатике.

Конечно, такое рассуждение не оставляет ощущения удовлетворённости. Хотелось бы понять, *почему* обнуляется поле внутри проводника. Давайте попробуем.

Рассмотрим незаряженный проводящий шар, помещённый во внешнее электростатическое поле E. Для простоты считаем это поле однородным, но наши рассуждения останутся верными и в общем случае.

Под действием электрического поля E свободные электроны нашего шара скапливаются в левом его полушарии, которое заряжается отрицательно. Справа остаётся нескомпенсированный положительный заряд. Возникновение этих зарядов, как вы помните, называется электростатической индукцией: заряды на поверхности проводника индуцируются (т. е. наводятся) внешним электростатическим полем. Подчеркнём ещё раз, что происходит реальное разделение зарядов: если сейчас распилить шар по диаметру в вертикальной плоскости, то получатся два разноимённо заряженных полушария.

[5] В узлах кристаллической решётки остаются положительные ионы. Казалось бы, они должны разлететься под действием кулоновских сил. Но нет — промежутки между ионами заполнены «газом» свободных электронов, который играет роль клея, держащего всю кристаллическую решётку

[6] Поэтому нельзя купаться во время грозы!

Индуцированные заряды создают своё поле E_i, направление которого внутри шара оказывается противоположным внешнему полю (рис. 3.23).

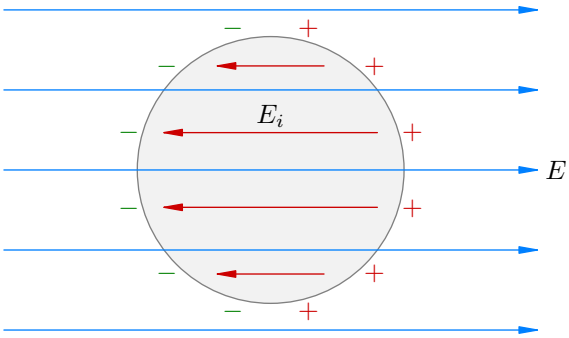


Рис. 3.23. $E_i = E$

Перестроение свободных зарядов шара продолжается до тех пор, пока поле E_i не компенсирует полностью внешнее поле E во всей области внутри шара. При наступлении этого момента (а наступает он почти мгновенно) результирующее поле внутри шара станет равным нулю, дальнейшее движение зарядов прекратится, и они окончательно займут свои фиксированные статические положения на поверхности шара.

А что будет в области снаружи шара? Поле E_i и тут наложится на внешнее поле E, искажая его тем сильнее, чем ближе к шару расположена точка наблюдения. На больших расстояниях от шара внешнее поле почти не изменится. В результате картина линий напряжённости будет иметь примерно следующий вид (рис. 3.24).

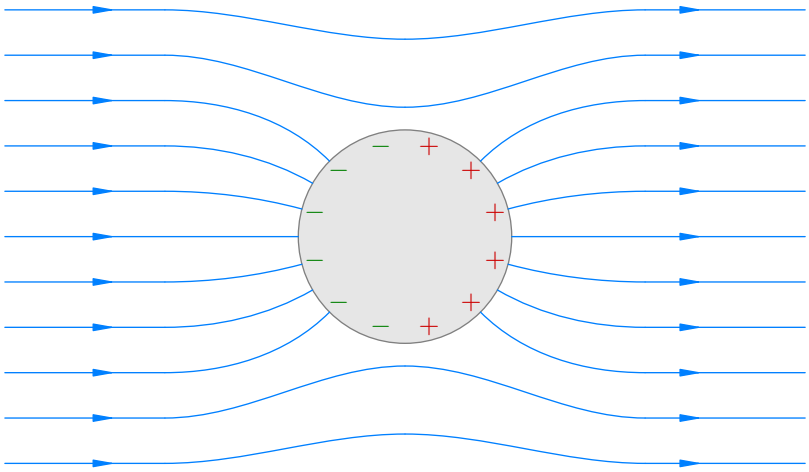


Рис. 3.24. Поле внутри проводника равно нулю

До сих пор наши рассуждения относились к случаю незаряженного проводника. Что изменится, если проводнику, помещённому в электростатическое поле, сообщить вдобавок некоторый заряд q?

Легко понять, что результирующее поле внутри проводника всё равно окажется равным нулю. В самом деле, заряд q начнёт перераспределяться по поверхности проводника таким образом, что поле E_i этого заряда внутри проводника будет направлено *против* внешнего электростатического поля E. Перераспределение будет продолжаться до тех пор, пока оба поля E и E_i не компенсируют друг друга во всей внутренней области проводника.

Таким образом, *поле внутри проводника равно нулю вне зависимости от того, заряжен проводник или нет.* Любой проводник, помещённый в электростатическое поле, как бы «выталкивает» внешнее поле из своей внутренней области.

3.5.2 Заряд внутри проводника

Следующий общее свойство проводников состоит в том, что *объёмная плотность зарядов внутри проводника везде равна нулю.* Сформулируем это более подробно.

Какую бы область внутри проводника мы ни взяли, её суммарный заряд окажется равен нулю. Нескомпенсированные заряды, если они имеются, располагаются целиком на поверхности проводника.

Строгое доказательство этого утверждения опирается на фудаментальную теорему Гаусса, которую в школе не проходят. А неформальное объяснение очень простое: если бы внутри проводника имелись нескомпенсированые заряды, то они создавали бы там электрическое поле. Но электрического поля внутри проводника нет — стало быть, нет и зарядов.

Отсюда следует ещё один замечательный факт: *если внутри проводника имеется полость, то поле в этой полости равно нулю.* В самом деле, создадим внутри проводника полость, изъяв часть вещества. Поле как было равно нулю до изъятия, так нулевым и останется — ведь заряд вынутого вещества равен нулю! Наши манипуляции не изменили ту статическую конфигурацию зарядов на поверхности проводника, которая создаёт нулевое поле во всех точках внутри проводника.

На явлении исчезновения поля в полости внутри проводника основана так называемая *электростатическая защита.* Если нужно уберечь от внешних электростатических полей какое-либо устройство, его помещают в металлический ящик (или окружают металлической сеткой), обнуляя напряжённость поля в пространстве вокруг устройства.

3.5.3 Поле вне проводника

Теперь рассмотрим область пространства, внешнюю по отношению к проводнику. Оказывается, *линии напряжённости электрического поля входят в проводник (или выходят из него) перпендикулярно поверхности проводника.*

Посмотрите ещё раз на рис. 3.24. Вы видите, что любая силовая линия, пересекающая шар, направлена точно под прямым углом к его поверхности.

Почему так получается? Давайте снова проведём доказательство от противного. Предположим, что в некоторой точке поверхности проводника силовая линия не перпендикулярна поверхности. Тогда в данной точке имеется составляющая вектора напряжённости, направленная *по касательной* к поверхности проводника — так называемая *касательная составляющая* вектора напряжённости. Под действием этой касательной составляющей возникнет электрический ток — а это противоречит тому, что мы находимся в электростатике.

Иными словами, заряды на поверхности проводника (при помещении проводника во внешнее поле или при сообщении проводнику заряда) перестраиваются до тех пор, пока линии напряжённости, уходящие в окружающее пространство, в каждой точке поверхности проводника не окажутся перпендикулярны этой поверхности (а внутри проводника не исчезнут вовсе).

3.5.4 Потенциал проводника

Раньше мы говорили о потенциале той или иной *точки* электростатического поля. Большой интерес представляют множества точек, потенциал которых одинаков. Один пример такого множества мы знаем — это эквипотенциальные поверхности. Другим замечательным примером служит проводник.

Все точки проводника имеют одинаковый потенциал. Иными словами, разность потенциалов между любыми двумя точками проводника равна нулю.

В самом деле, если бы между какой-либо парой точек проводника существовала ненулевая разность потенциалов, возник бы ток от одной точки к другой — ведь в этом случае электрическое поле совершало бы ненулевую работу по перемещению зарядов между данными точками. Но в электростатике никакого тока быть не может.

Потенциал какой-либо (и тогда любой) точки проводника называется *потенциалом проводника*.

Как видим, проводник представляет собой «эквипотенциальный объём». В частности, поверхность проводника является эквипотенциальной поверхностью. Это даёт дополнительное объяснение утверждения предыдущего пункта — мы же знаем, что линии напряжённости электростатического поля перпендикулярны эквипотенциальным поверхностям.

3.5.5 Напряжённость и потенциал поля проводящей сферы

Рассмотрим металлическую сферу радиуса R, которой сообщён заряд q. Нас интересуют напряжённость и потенциал электростатического поля, создаваемое сферой в каждой точке пространства.

Везде далее сферу можно заменить шаром — от этого ровным счётом ничего не изменится.

Начнём с напряжённости поля. Внутри сферы, как мы уже знаем, напряжённость поля равна нулю. Вне сферы напряжённость оказывается такой же, как если бы заряд q был точечным и находился в центре сферы. Итак:

$$E = \begin{cases} \dfrac{kq}{r^2}, & \text{если } r \geqslant R; \\ 0, & \text{если } r < R. \end{cases}$$

На рис. 3.25 показаны линии напряжённости поля положительно заряженной сферы и график зависимости модуля вектора напряжённости от расстояния до центра сферы.

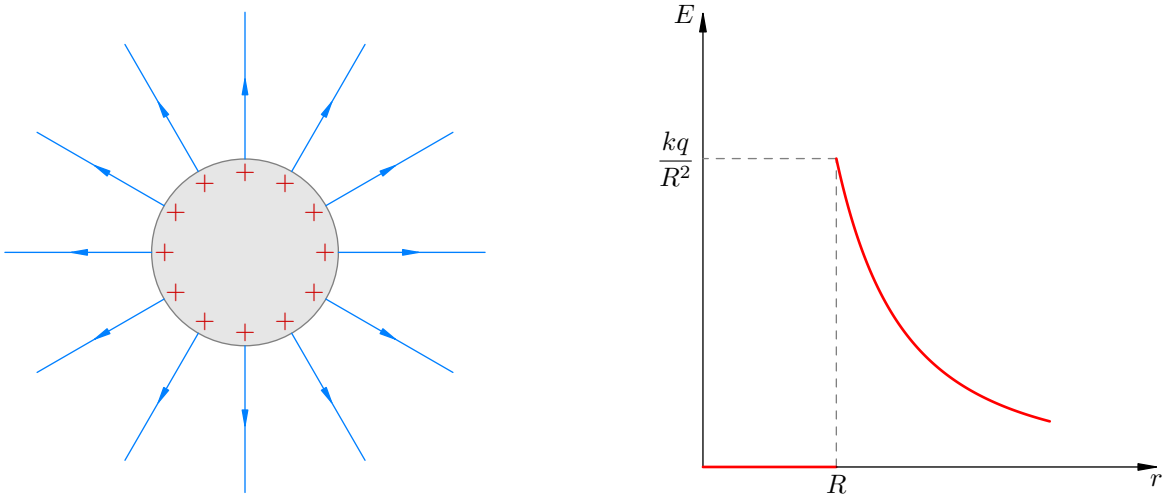


Рис. 3.25. Напряжённость поля заряженной сферы

Потенциал поля вне сферы равен потенциалу поля точечного заряда q, расположенного в центре сферы. Внутри сферы потенциал везде одинаков и совпадает с потенциалом точек поверхности сферы:

$$\varphi = \begin{cases} \dfrac{kq}{r}, & \text{если } r \geqslant R; \\ \dfrac{kq}{R}, & \text{если } r < R. \end{cases}$$

Вот как выглядит график зависимости потенциала положительно заряженной сферы от расстояния до её центра (рис. 3.26):

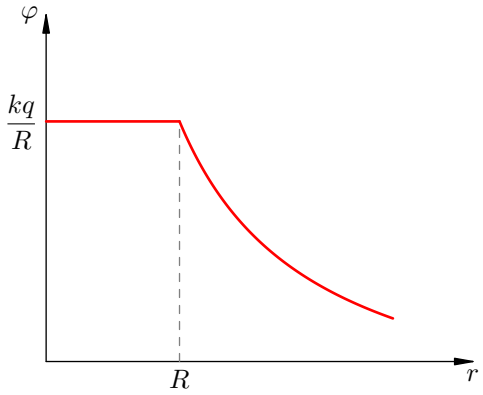


Рис. 3.26. Потенциал поля заряженной сферы

3.6 Диэлектрики в электрическом поле

В отличие от проводников, в *диэлектриках* нет свободных зарядов. Все заряды являются *связанными*: электроны принадлежат своим атомам, а ионы твёрдых диэлектриков колеблются вблизи узлов кристаллической решётки.

Соответственно, при помещении диэлектрика в электрическое поле не возникает направленного движения зарядов[7]. Поэтому для диэлектриков не проходят наши доказательства свойств проводников — ведь все эти рассуждения опирались на возможность появления тока. И действительно, ни одно из четырёх свойств проводников, сформулированных в предыдущей статье, не распростаняется на диэлектрики.

1. Напряжённость электрического поля внутри диэлектрика может быть не равна нулю.
2. Объёмная плотность заряда в диэлектрике может быть отличной от нуля.
3. Линии напряжённости могут быть не перпендикулярны поверхности диэлектрика.
4. Различные точки диэлектрика могут иметь разный потенциал. Стало быть, говорить о «потенциале диэлектрика» не приходится.

3.6.1 Диэлектрическая проницаемость

Но тем не менее, одно важнейшее общее свойство у диэлектриков имеется, и вам оно известно (вспомните формулу напряжённости поля точечного заряда в диэлектрике!). *Напряжённость поля уменьшается внутри диэлектрика в некоторое число ε раз по сравнению с вакуумом.* Величина ε даётся в таблицах и называется *диэлектрической проницаемостью* диэлектрика.

Давайте разберёмся, каковы причины ослабления поля в диэлектрике. Рассмотрим диэлектрик, помещённый во внешнее однородное (для простоты) поле E_0. Опыт показывает, что на противоположных поверхностях диэлектрика появляются заряды разных знаков.

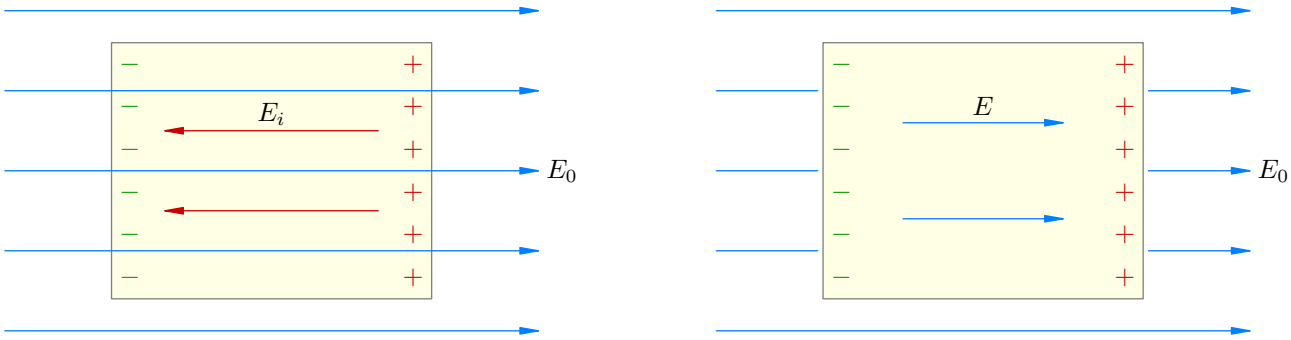


Рис. 3.27. Ослабление поля внутри диэлектрика

Эти индуцированные заряды расположены так, что создаваемое ими поле E_i внутри диэлектрика направлено *против* внешнего поля E_0 (рис. 3.27, слева). При этом $E_i < E_0$, так что внешнее поле ослабляется лишь частично (а не гасится полностью, как внутри проводника). Результирующее поле внутри диэлектрика равно:

$$E = E_0 - E_i.$$

[7] Впрочем, в достаточно сильном электрическом поле может случиться *пробой диэлектрика* (пример — молния во время грозы). Подобное явление мы изучим позже, при рассмотрении электрического тока в газах.

Мы видим, что $E < E_0$. Данный факт как раз и подчёркивается следующей формой записи:

$$E = \frac{E_0}{\varepsilon}\,.$$

Результирующее поле E направлено в ту же сторону, что и внешнее поле E_0 (рис. 3.27, справа; искажение поля снаружи диэлектрика считаем пренебрежимо малым).

Хорошо, но откуда в диэлектрике берутся поверхностные индуцированные заряды? Это — явление *поляризации*, свойственное всем диэлектрикам. Механизмы поляризации могут быть различными. Мы рассмотрим два типа поляризации: ориентационную и электронную.

3.6.2 Полярные диэлектрики

Молекулы *полярных* диэлектриков с точки зрения электрических свойств являются диполями.

Диполь — это система двух одинаковых по модулю и противоположных по знаку зарядов, находящихся на некотором расстоянии друг от друга (рис. 3.28).

Рис. 3.28. Диполь

Например, в молекуле поваренной соли NaCl одинокий внешний электрон натрия захватывается атомом хлора (которому как раз недостаёт одного электрона до полного комплекта из 8 электронов на внешней оболочке). Молекула становится диполем, состоящим из положительного иона Na^+ и отрицательного иона Cl^-.

Как диполи ведут себя молекулы воды. Это связано с геометрией их строения: молекула воды похожа на треугольник, в вершинах которого расположены два атома водорода и один атом кислорода. В результате центры положительных и отрицательных зарядов молекулы оказываются в разных местах, что и наделяет молекулу свойствами диполя.

К полярным диэлектрикам относятся также низкомолекулярные спирты и ряд других веществ.

При отсутствии внешнего электрического поля молекулы-диполи полярного диэлектрика, совершая хаотическое тепловое движение, ориентированы в самых разных направлениях. Электрические поля этих диполей полностью компенсируют друг друга, и результирующее поле равно нулю во всех областях диэлектрика.

Но если поместить такой диэлектрик во внешнее поле E_0, то оно «развернёт» диполи так, что они окажутся ориентированными вдоль линий напряжённости («минусы» диполей повернутся влево — к тем «плюсам», которые создают внешнее поле). Это показано на рис. 3.29.

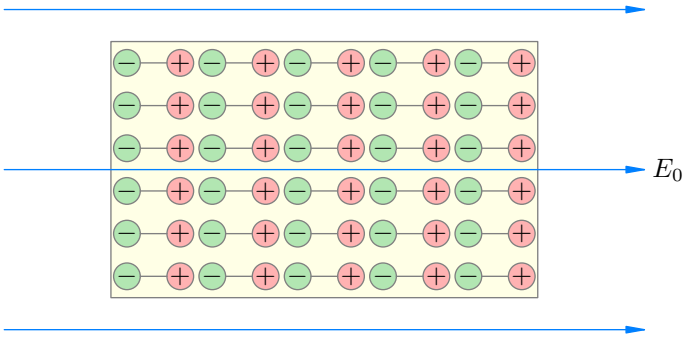


Рис. 3.29. Ориентационная поляризация

Разумеется, в реальности не будет столь идеально правильного порядка расположения диполей — ведь они по-прежнему совершают хаотическое тепловое движение. Но теперь у диполей появится *преимущественная* ориентация — вдоль линий напряжённости внешнего поля.

Итак, что же мы видим на рис. 3.29? Внутри диэлектрика заряды диполей по-прежнему компенсируют друг друга. Однако *на внешних поверхностях диэлектрика появляются нескомпенсированные заряды: справа — положительные, слева — отрицательные*. Именно эти заряды и показаны на рис. 3.27; благодаря им как раз и возникает встречное поле E_i, ослабляющее внешнее поле E_0.

Итак, механизм ориентационной поляризации ясен: поворот молекул-диполей и их ориентация вдоль линий напряжённости внешнего поля.

3.6.3 Неполярные диэлектрики

Далеко не все диэлектрики являются полярными. Диэлектрик называется *неполярным*, если его молекулы имеют симметричное распределение положительных и отрицательных зарядов[8] и потому не ведут себя как диполи. К неполярным диэлектрикам относятся, например, керосин, масло, воздух, инертные газы.

Тем не менее, поляризация наблюдается и у неполярных диэлектриков. Каков механизм поляризации в данном случае?

На рис. 3.30 слева изображена симметричная электронная орбита в атоме неполярного диэлектрика. При наложении внешнего поля E_0 эта орбита деформируется (рис. 3.30, справа): электрон смещается в сторону положительных зарядов, создающих внешнее поле.

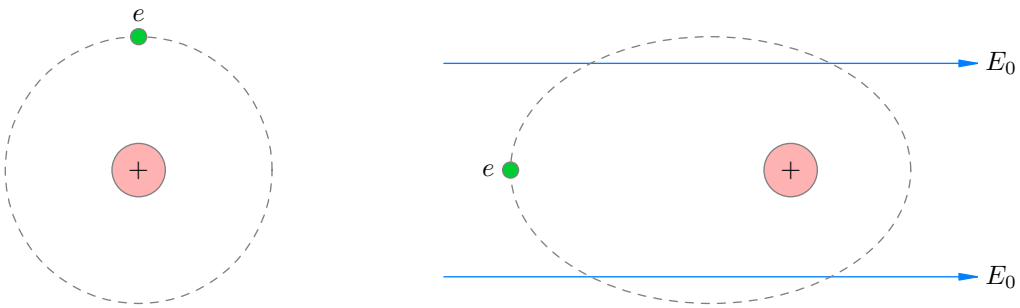


Рис. 3.30. Электронная поляризация

Мы видим, что во внешнем поле электрон будет проводить больше времени слева от ядра, нежели чем справа. Из-за этого центры положительных и отрицательных зарядов в атоме неполярного диэлектрика разойдутся в разные стороны. Соответственно, атомы или молекулы неполярного диэлектрика во внешнем поле также начнут вести себя подобно диполям, и мы приходим к такой же по сути картине, что и рис. 3.29 для случая ориентационной поляризации. Объяснение механизма ослабления поля внутри диэлектрика останется тем же самым.

Итак, электронная поляризация вызвана деформацией электронных оболочек атомов во внешнем электрическом поле. Разумеется, электронная поляризация присутствует и у полярных диэлектриков; но там она теряется на фоне куда более мощного эффекта разворота самих диполей.

[8] Разумеется, при отсутствии внешнего поля.

3.7 Конденсатор. Энергия электрического поля

Предыдущие два раздела были посвящены отдельному рассмотрению проводников и диэлектриков, помещённых в электрическое поле. Сейчас нам понадобится объединить эти знания. Дело в том, что большое практическое значение имеет совместное использование проводников и диэлектриков в специальных устройствах — *конденсаторах*.

Но прежде введём центральное понятие *электрической ёмкости*. Ёмкость является важной характеристикой изолированного проводника и систем проводников (к числу которых принадлежат конденсаторы).

3.7.1 Ёмкость уединённого проводника

Предположим, что заряженный проводник расположен настолько далеко от всех остальных тел, что взаимодействие зарядов проводника с окружающими телами можно не принимать во внимание. В таком случае проводник называется *уединённым*.

Потенциал всех точек нашего проводника, как мы знаем, имеет одно и то же значение φ, которое называется потенциалом проводника. Оказывается, что *потенциал уединённого проводника прямо пропорционален его заряду*. Коэффициент пропорциональности принято обозначать $1/C$, так что

$$\varphi = \frac{q}{C}\,.$$

Величина C называется *электрической ёмкостью* проводника и равна отношению заряда проводника к его потенциалу:

$$C = \frac{q}{\varphi}\,. \tag{3.23}$$

Например, потенциал уединённого шара в вакууме равен:

$$\varphi = \frac{kq}{R} = \frac{q}{4\pi\varepsilon_0 R}\,,$$

где q — заряд шара, R — его радиус. Отсюда ёмкость шара:

$$C = 4\pi\varepsilon_0 R. \tag{3.24}$$

Если шар окружён средой-диэлектриком с диэлектрической проницаемостью ε, то его потенциал уменьшается в ε раз:

$$\varphi = \frac{q}{4\pi\varepsilon_0\varepsilon R}\,.$$

Соответственно, ёмкость шара в ε раз увеличивается:

$$C = 4\pi\varepsilon_0\varepsilon R. \tag{3.25}$$

Увеличение ёмкости при наличии диэлектрика — важнейший факт. Мы ещё встретимся с ним при рассмотрении конденсаторов.

Из формул (3.24) и (3.25) мы видим, что ёмкость шара зависит только от его радиуса и диэлектрической проницаемости окружающей среды. То же самое будет и в общем случае: *ёмкость уединённого проводника не зависит от его заряда; она определяется лишь размерами и формой проводника, а также диэлектрической проницаемостью среды, окружающей проводник. От вещества проводника ёмкость также не зависит.*

В чём смысл понятия ёмкости? *Ёмкость показывает, какой заряд нужно сообщить проводнику, чтобы увеличить его потенциал на 1 В.* Чем больше ёмкость — тем, соответственно, больший заряд требуется поместить для этого на проводник.

Единицей измерения ёмкости служит *фарад* (Ф). Из определения ёмкости (3.23) видно, что Ф = Кл/В.

Давайте ради интереса вычислим ёмкость земного шара (он является проводником!). Радиус считаем приближённо равным 6400 км.

$$C = 4\pi\varepsilon_0 R \approx 4 \cdot 3{,}14 \cdot 8{,}85 \cdot 10^{-12} \cdot 6400 \cdot 10^3 \approx 712 \text{ мкФ}.$$

Как видите, 1 Ф — это очень большая ёмкость.

Единица измерения ёмкости полезна ещё и тем, что позволяет сильно сэкономить на обозначении размерности диэлектрической постоянной ε_0. В самом деле, выразим ε_0 из формулы (3.24):

$$\varepsilon_0 = \frac{C}{4\pi R}\,.$$

Следовательно, диэлектрическая постоянная может измеряться в Ф/м:

$$\varepsilon_0 = 8{,}85 \cdot 10^{-12}\ \frac{\text{Ф}}{\text{м}}\,.$$

Так легче запомнить, не правда ли?

3.7.2 Ёмкость плоского конденсатора

Ёмкость уединённого проводника на практике используется редко. В обычных ситуациях проводники не являются уединёнными. Заряженный проводник взаимодействует с окружающими телами и наводит на них заряды, а потенциал поля этих индуцированных зарядов (по принципу суперпозиции!) изменяет потенциал самого проводника. В таком случае уже нельзя утверждать, что потенциал проводника будет прямо пропорционален его заряду, и понятие ёмкости проводника самого по себе фактически утрачивает смысл.

Можно, однако, создать систему заряженных проводников, которая даже при накоплении на них значительного заряда почти не взаимодействует с окружающими телами. Тогда мы сможем снова говорить о ёмкости — но на сей раз о ёмкости этой системы проводников.

Наиболее простым и важным примером такой системы является *плоский конденсатор*. Он состоит из двух параллельных металлических пластин (называемых *обкладками*), разделённых слоем диэлектрика. При этом расстояние между пластинами много меньше их собственных размеров.

Обозначение конденсатора на электрической схеме показано на рис. 3.31.

Рис. 3.31. Конденсатор

Для начала мы рассмотрим *воздушный* конденсатор, у которого между обкладками находится воздух ($\varepsilon = 1$).

Пусть заряды обкладок равны $+q$ и $-q$. Именно так и бывает в реальных электрических схемах: заряды обкладок равны по модулю и противоположны по знаку. Величина q — заряд положительной обкладки — называется *зарядом конденсатора*.

Пусть S — площадь каждой обкладки. Найдём поле, создаваемое обкладками в окружающем пространстве.

Поскольку размеры обкладок велики по сравнению с расстоянием между ними, поле каждой обкладки вдали от её краёв можно считать однородным полем бесконечной заряженной плоскости:

$$E_+ = E_- = \frac{\sigma}{2\varepsilon_0}\,.$$

Здесь E_+ — напряжённость поля положительной обкладки, E_- — напряженность поля отрицательной обкладки, σ — поверхностная плотность зарядов на обкладке:

$$\sigma = \frac{q}{S}\,.$$

На рис. 3.32 (слева) изображены векторы напряжённости поля каждой обкладки в трёх областях: слева от конденсатора, внутри конденсатора и справа от конденсатора.

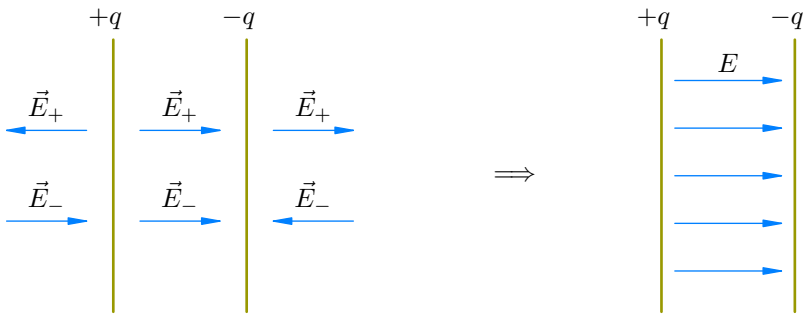


Рис. 3.32. Электрическое поле плоского конденсатора

Согласно принципу суперпозиции, для результирующего поля $\vec{E}$ имеем:

$$\vec{E} = \vec{E}_+ + \vec{E}_-\,.$$

Нетрудно видеть, что слева и справа от конденсатора поле обращается в нуль (поля обкладок погашают друг друга):

$$E = E_+ - E_- = 0.$$

Внутри конденсатора поле удваивается:

$$E = E_+ + E_- = \frac{\sigma}{\varepsilon_0}\,,$$

или

$$E = \frac{q}{\varepsilon_0 S}\,. \tag{3.26}$$

Результирующее поле обкладок плоского конденсатора изображено на рис. 3.32 справа. Итак:

Внутри плоского конденсатора создаётся однородное электрическое поле, напряжённость которого находится по формуле (3.26)*. Снаружи конденсатора поле равно нулю, так что конденсатор не взаимодействует с окружающими телами.*

Не будем забывать, однако, что данное утверждение выведено из предположения, будто обкладки являются бесконечными плоскостями. На самом деле их размеры конечны, и вблизи краёв обкладок возникают так называемые *краевые эффекты*: поле отличается от однородного и проникает в наружное пространство конденсатора. Но в большинстве ситуаций (и уж тем более в задачах ЕГЭ по физике) краевыми эффектами можно пренебречь и действовать так, словно утверждение, выделенное курсивом, является верным без всяких оговорок.

Пусть расстояние между обкладками конденсатора равно d. Поскольку поле внутри конденсатора является однородным, разность потенциалов U между обкладками равна произведению E на d (вспомните связь напряжения и напряжённости в однородном поле!):

$$U = Ed = \frac{qd}{\varepsilon_0 S}\,. \tag{3.27}$$

Разность потенциалов между обкладками конденсатора, как видим, прямо пропорциональна заряду конденсатора. Данное утверждение аналогично утверждению «потенциал уединённого

проводника прямо пропорционален заряду проводника», с которого и начался весь разговор о ёмкости. Продолжая эту аналогию, определяем *ёмкость конденсатора* как отношение заряда конденсатора к разности потенциалов между его обкладками:

$$C = \frac{q}{U}\,. \tag{3.28}$$

Ёмкость конденсатора показывает, какой заряд ему нужно сообщить, чтобы разность потенциалов между его обкладками увеличилась на 1 В. Формула (3.28), таким образом, является модификацией формулы (3.23) для случая системы двух проводников — конденсатора.

Из формул (3.28) и (3.27) легко находим *ёмкость плоского воздушного конденсатора*:

$$C = \frac{\varepsilon_0 S}{d}\,. \tag{3.29}$$

Она зависит только от геометрических характеристик конденсатора: площади обкладок и расстояния между ними.

Предположим теперь, что пространство между обкладками заполнено диэлектриком с диэлектрической проницаемостью ε. Как изменится ёмкость конденсатора?

Напряжённость поля внутри конденсатора уменьшится в ε раз. Соответственно, вместо формулы (3.26) теперь имеем:

$$E = \frac{q}{\varepsilon_0 \varepsilon S}\,. \tag{3.30}$$

Напряжение на конденсаторе получается равным:

$$U = Ed = \frac{qd}{\varepsilon_0 \varepsilon S}\,. \tag{3.31}$$

Отсюда *ёмкость плоского конденсатора с диэлектриком*:

$$C = \frac{\varepsilon_0 \varepsilon S}{d}\,. \tag{3.32}$$

Она зависит от геометрических характеристик конденсатора (площади обкладок и расстояния между ними) и от диэлектрической проницаемости диэлектрика, заполняющего конденсатор. Важное следствие формулы (3.32): *заполнение конденсатора диэлектриком увеличивает его ёмкость*.

3.7.3 Энергия заряженного конденсатора

Заряженный конденсатор обладает энергией. В этом можно убедиться на опыте. Если зарядить конденсатор и замкнуть его на лампочку, то (при условии, что ёмкость конденсатора достаточно велика) лампочка ненадолго загорится.

Следовательно, в заряженном конденсаторе запасена энергия, которая и выделяется при его разрядке. Нетрудно понять, что этой энергией является потенциальная энергия взаимодействия обкладок конденсатора — ведь обкладки, будучи заряжены разноимённо, притягиваются друг к другу.

Мы сейчас вычислим эту энергию, а затем увидим, что существует и более глубокое понимание происхождения энергии заряженного конденсатора.

Начнём с плоского воздушного конденсатора. Ответим на такой вопрос: какова сила притяжения его обкладок друг к другу? Величины используем те же: заряд конденсатора q, площадь обкладок S.

Возьмём на второй обкладке настолько маленькую площадку, что заряд q_0 этой площадки можно считать точечным. Данный заряд притягивается к первой обкладке с силой

$$F_0 = q_0 E_1,$$

где E_1 — напряжённость поля первой обкладки:

$$E_1 = \frac{\sigma}{2\varepsilon_0} = \frac{q}{2\varepsilon_0 S}\,.$$

Следовательно,

$$F_0 = \frac{q_0 q}{2\varepsilon_0 S}\,.$$

Направлена эта сила параллельно линиям поля (т. е. перпендикулярно пластинам).

Результирующая сила F притяжения второй обкладки к первой складывается из всех этих сил F_0, с которыми притягиваются к первой обкладке всевозможные маленькие заряды q_0 второй обкладки. При этом суммировании постоянный множитель $q/(2\varepsilon_0 S)$ вынесется за скобку, а в скобке просуммируются все q_0 и дадут q. В результате получим:

$$F = \frac{q^2}{2\varepsilon_0 S}\,. \tag{3.33}$$

Предположим теперь, что расстояние между обкладками изменилось от начальной величины d_1 до конечной величины d_2. Сила притяжения пластин совершает при этом работу:

$$A = F(d_1 - d_2).$$

Знак правильный: если пластины сближаются ($d_2 < d_1$), то сила совершает положительную работу, так как пластины притягиваются друг к другу. Наоборот, если удалять пластины ($d_2 > d_1$), то работа силы притяжения получается отрицательной, как и должно быть.

С учётом формул (3.33) и (3.29) имеем:

$$A = \frac{q^2}{2\varepsilon_0 S}(d_1 - d_2) = \frac{q^2 d_1}{2\varepsilon_0 S} - \frac{q^2 d_2}{2\varepsilon_0 S} = \frac{q^2}{2C_1} - \frac{q^2}{2C_2} = W_1 - W_2,$$

где $W_1 = q^2/(2C_1)$, $W_2 = q^2/(2C_2)$. Это можно переписать следующим образом:

$$A = -(W_2 - W_1) = -\Delta W,$$

где

$$W = \frac{q^2}{2C}\,. \tag{3.34}$$

Работа потенциальной силы F притяжения обкладок оказалась равна изменению со знаком минус величины W. Это как раз и означает, что W — потенциальная энергия взаимодействия обкладок, или *энергия заряженного конденсатора*.

Используя соотношение $q = CU$, из формулы (3.34) можно получить ещё две формулы для энергии конденсатора (убедитесь в этом самостоятельно!):

$$W = \frac{qU}{2}\,, \tag{3.35}$$

$$W = \frac{CU^2}{2}\,. \tag{3.36}$$

Особенно полезными являются формулы (3.34) и (3.36).

Допустим теперь, что конденсатор заполнен диэлектриком с диэлектрической проницаемостью ε. Сила притяжения обкладок уменьшится в ε раз, и вместо (3.33) получим:

$$F = \frac{q^2}{2\varepsilon_0 \varepsilon S}\,.$$

При вычислении работы силы F, как нетрудно видеть, величина ε войдёт в ёмкость C, и формулы (3.34)–(3.36) останутся неизменными. Ёмкость конденсатора в них теперь будет выражаться по формуле (3.32).

Итак, *формулы* (3.34)–(3.36) *универсальны: они справедливы как для воздушного конденсатора, так и для конденсатора с диэлектриком.*

3.7.4 Энергия электрического поля

Мы обещали, что после вычисления энергии конденсатора дадим более глубокое истолкование происхождения этой энергии. Что ж, приступим.

Рассмотрим воздушный конденсатор и преобразуем формулу (3.36) для его энергии:

$$W = \frac{CU^2}{2} = \frac{\varepsilon_0 S}{d}\frac{(Ed)^2}{2} = \frac{\varepsilon_0 E^2}{2} Sd.$$

Но $Sd = V$ — объём конденсатора. Получаем:

$$W = \frac{\varepsilon_0 E^2}{2} V. \tag{3.37}$$

Посмотрите внимательно на эту формулу. Она уже не содержит ничего, что являлось бы специфическим для конденсатора! Мы видим *энергию электрического поля* E, сосредоточенного в некотором объёме V.

Энергия конденсатора есть не что иное, как энергия заключённого внутри него электрического поля.

Итак, электрическое поле само по себе обладает энергией. Ничего удивительного для нас тут нет. Радиоволны, солнечный свет — это примеры распространения энергии, переносимой в пространстве электромагнитными волнами.

Величина $w = W/V$ — энергия единицы объёма поля — называется *объёмной плотностью энергии*. Из формулы (3.37) получим:

$$w = \frac{\varepsilon_0 E^2}{2}. \tag{3.38}$$

В этой формуле не осталось вообще никаких геометрических величин. Она даёт максимально чистую связь энергии электрического поля и его напряжённости.

Если конденсатор заполнен диэлектриком, то его ёмкость увеличивается в ε раз, и вместо формул (3.37) и (3.38) будем иметь:

$$W = \frac{\varepsilon_0 \varepsilon E^2}{2} V, \tag{3.39}$$

$$w = \frac{\varepsilon_0 \varepsilon E^2}{2}. \tag{3.40}$$

Как видим, энергия электрического поля зависит ещё и от диэлектрической проницаемости среды, в которой поле находится.

Замечательно, что полученные формулы для энергии и плотности энергии выходят далеко за пределы электростатики: они справедливы не только для электростатического поля, но и для электрических полей, меняющихся во времени.

3.8 Постоянный электрический ток

Электрический ток обеспечивает комфортом жизнь современного человека. Технологические достижения цивилизации — энергетика, транспорт, радио, телевидение, компьютеры, мобильная связь — основаны на использовании электрического тока.

Электрический ток — это направленное движение заряженных частиц, при котором происходит перенос заряда из одних областей пространства в другие.

Электрический ток может возникать в самых различных средах: твёрдых телах, жидкостях, газах. Порой и среды никакой не нужно — ток может существовать даже в вакууме! Мы поговорим об этом в своё время, а пока приведём лишь некоторые примеры.

- Замкнём полюса батарейки металлическим проводом. Свободные электроны провода начнут направленное движение от «минуса» батарейки к «плюсу».

  Это — пример тока в металлах.

- Бросим в стакан воды щепотку поваренной соли NaCl. Молекулы соли диссоциируют на ионы, так что в растворе появятся свободные заряды: положительные ионы Na^+ и отрицательные ионы Cl^-. Теперь засунем в воду два электрода, соединённые с полюсами батарейки. Ионы Na^+ начнут направленное движение к отрицательному электроду, а ионы Cl^- — к положительному.

  Это — пример прохождения тока через раствор электролита.

- Грозовые тучи создают столь мощные электрические поля, что оказывается возможным пробой воздушного промежутка длиной в несколько километров. В результате сквозь воздух проходит гигантский разряд — молния.

  Это — пример электрического тока в газе.

Во всех трёх рассмотренных примерах электрический ток обусловлен движением заряженных частиц внутри тела и называется *током проводимости*.

- Вот несколько иной пример. Будем перемещать в пространстве заряженное тело. Такая ситуация согласуется с определением тока! Направленное движение зарядов — есть, перенос заряда в пространстве — присутствует. Ток, созданный движением макроскопического заряженного тела, называется *конвекционным*.

Заметим, что не всякое движение заряженных частиц образует ток. Например, хаотическое тепловое движение зарядов проводника — не направленное (оно совершается в каких угодно направлениях), и потому током не является[9].

Не будет током и поступательное движение электрически нейтрального тела: хотя заряженные частицы в его атомах и совершают направленное движение, не происходит переноса заряда из одних участков пространства в другие.

3.8.1 Направление электрического тока

Направление движения заряженных частиц, образующих ток, зависит от знака их заряда. Положительно заряженные частицы будут двигаться от «плюса» к «минусу», а отрицательно

[9]При возникновении тока свободные заряды продолжают совершать тепловое движение! Просто в этом случае к хаотическим перемещениям заряженных частиц добавляется их упорядоченный дрейф в определённом направлении.

заряженные — наоборот, от «минуса» к «плюсу». В электролитах и газах, например, присутствуют как положительные, так и отрицательные свободные заряды, и ток создаётся их встречным движением в обоих направлениях. Какое же из этих направлений принять за направление электрического тока?

Направлением тока принято считать направление движения положительных зарядов. Попросту говоря, по соглашению *ток течёт от «плюса» к «минусу»* (рис. 3.33; положительная клемма источника тока изображена длинной чертой, отрицательная клемма — короткой).

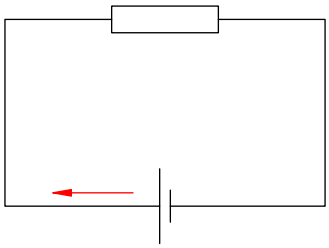


Рис. 3.33. Направление тока

Данное соглашение вступает в некоторое противоречие с наиболее распространённым случаем металлических проводников. В металле носителями заряда являются свободные электроны, и двигаются они от «минуса» к «плюсу». Но в соответствии с соглашением мы вынуждены считать, что направление тока в металлическом проводнике противоположно движению свободных электронов. Это, конечно, не очень удобно.

Тут, однако, ничего не поделаешь — придётся принять эту ситуацию как данность. Так уж исторически сложилось. Выбор направления тока был предложен Ампером[10] в первой половине XIX века, за 70 лет до открытия электрона. К этому выбору все привыкли, и когда в 1916 году выяснилось, что ток в металлах вызван движением свободных электронов, ничего менять уже не стали.

3.8.2 Действие электрического тока

Как мы можем определить, протекает электрический ток или нет? О возникновении электрического тока можно судить по следующим его проявлениям.

1. *Тепловое действие тока.* Электрический ток вызывает нагревание вещества, в котором он протекает. Именно так нагреваются спирали нагревательных приборов и ламп накаливания. Именно поэтому мы видим молнию. В основе действия тепловых амперметров лежит тепловое расширение проводника с током, приводящее к перемещению стрелки прибора.

2. *Магнитное действие тока.* Электрический ток создаёт магнитное поле: стрелка компаса, расположенная рядом с проводом, при включении тока поворачивается перпендикулярно проводу. Магнитное поле тока можно многократно усилить, если обмотать провод вокруг железного стержня — получится электромагнит. На этом принципе основано действие амперметров магнитоэлектрической системы: электромагнит поворачивается в поле постоянного магнита, в результате чего стрелка прибора перемещается по шкале.

3. *Химическое действие тока.* При прохождении тока через электролиты можно наблюдать изменение химического состава вещества. Так, в растворе $CuSO_4$ положительные ионы Cu^{2+} двигаются к отрицательному электроду, и этот электрод покрывается медью.

[10] Договорённость о направлении тока понадобилась Амперу для того, чтобы дать чёткое правило определения направления силы, действующей на проводник с током в магнитном поле. Сегодня эту силу мы называем *силой Ампера*, направление которой определяется по *правилу левой руки*.

3.8.3 Сила и плотность тока

Электрический ток называется *постоянным*, если за равные промежутки времени через поперечное сечение проводника проходит одинаковый заряд. Постоянный ток наиболее прост для изучения. С него мы и начинаем.

Количественной характеристикой электрического тока является *сила тока*. В случае постоянного тока абсолютная величина силы тока есть отношение абсолютной величины заряда q, прошедшего через поперечное сечение проводника за время t, к этому самому времени:

$$I = \frac{q}{t}\,. \tag{3.41}$$

Измеряется сила тока в *амперах* (А)[11]. При силе тока в 1 А через поперечное сечение проводника за 1 с проходит заряд в 1 Кл.

Подчеркнём, что формула (3.41) определяет абсолютную величину, или модуль силы тока. Сила тока может иметь ещё и знак! Этот знак не связан со знаком зарядов, образующих ток, и выбирается из иных соображений. А именно, в ряде ситуаций (например, если заранее не ясно, куда потечёт ток) удобно зафиксировать некоторое направление обхода цепи (скажем, против часовой стрелки) и считать силу тока положительной, если направление тока совпадает с направлением обхода, и отрицательной, если ток течёт против направления обхода[12].

В случае постоянного тока сила тока есть величина постоянная. Она показывает, какой заряд проходит через поперечное сечение проводника за 1 с.

Часто бывает удобно не связываться с площадью поперечного сечения и ввести величину *плотности тока*:

$$j = \frac{I}{S}\,, \tag{3.42}$$

где I — сила тока, S — площадь поперечного сечения проводника (разумеется, это сечение перпендикулярно направлению тока). С учётом формулы (3.41) имеем также:

$$j = \frac{q}{St}\,.$$

Плотность тока показывает, какой заряд проходит за единицу времени через единицу площади поперечного сечения проводника. Согласно формуле (3.42), плотность тока измеряется в А/м^2.

3.8.4 Скорость направленного движения зарядов

Когда мы включаем в комнате свет, нам кажется, что лампочка загорается мгновенно. Скорость распространения тока по проводам очень велика: она близка к 300000 км/с (скорости света в вакууме). Если бы лампочка находилась на Луне, она зажглась бы через секунду с небольшим.

Однако не следует думать, что с такой грандиозной скоростью двигаются свободные заряды, образующие ток. Оказывается, их скорость составляет всего-навсего доли миллиметра в секунду.

Почему же ток распространяется по проводам так быстро? Дело в том, что свободные заряды взаимодействуют друг с другом и, находясь под действием электрического поля источника тока, при замыкании цепи приходят в движение почти одновременно вдоль всего проводника. Скорость распространения тока есть скорость передачи электрического взаимодействия между

[11] Единица силы тока определяется через магнитное взаимодействие проводов с током. А именно, пусть имеются два параллельных провода, очень длинных и тонких, расположенных в вакууме на расстоянии 1 м друг от друга. По этим проводам течёт одинаковый ток. Мы говорим, что сила тока равна 1 А, если сила взаимодействия проводов равна $2 \cdot 10^{-7}$ Н на каждый метр провода.

[12] Сравните с тригонометрическим кругом: углы считаются положительными, если отсчитываются против часовой стрелки, и отрицательными, если по часовой стрелке.

свободными зарядами, и она близка к скорости света в вакууме. Скорость же, с которой сами заряды перемещаются внутри проводника, может быть на много порядков меньше.

Итак, подчеркнём ещё раз, что мы различаем две скорости.

1. *Скорость распространения тока.* Это — скорость передачи электрического сигнала по цепи. Близка к 300000 км/с.

2. *Скорость направленного движения свободных зарядов.* Это — средняя скорость перемещения зарядов, образующих ток. Называется ещё *скоростью дрейфа*.

Мы сейчас выведем формулу, выражающую силу тока I через скорость v направленного движения зарядов проводника.

Пусть проводник имеет площадь поперечного сечения S (рис. 3.34). Свободные заряды проводника будем считать положительными; величину свободного заряда обозначим e (в наиболее важном для практики случая металлического проводника это есть заряд электрона). Концентрация свободных зарядов (т. е. их число в единице объёма) равна n.

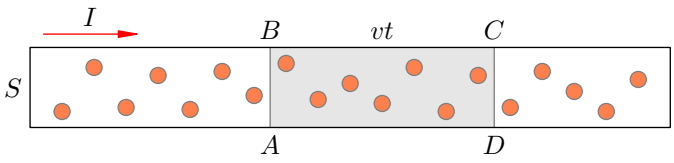


Рис. 3.34. К выводу формулы $I = envS$

Какой заряд q пройдёт через поперечное сечение AB нашего проводника за время t?

С одной стороны, разумеется,

$$q = It. \tag{3.43}$$

С другой стороны, сечение AB пересекут все те свободные заряды, которые спустя время t окажутся внутри цилиндра $ABCD$ с высотой vt. Их число равно:

$$N = nV_{ABCD} = nSvt.$$

Следовательно, их общий заряд будет равен:

$$q = eN = enSvt. \tag{3.44}$$

Приравнивая правые части формул (3.43) и (3.44) и сокращая на t, получим:

$$I = envS. \tag{3.45}$$

Соответственно, плотность тока оказывается равна:

$$j = env.$$

Давайте в качестве примера посчитаем, какова скорость движения свободных электронов в медном проводе при силе тока 1 А.

Заряд электрона известен: $e = 1{,}6 \cdot 10^{-19}$ Кл.

Чему равна концентрация свободных электронов? Она совпадает с концентрацией атомов меди, поскольку от каждого атома отщепляется по одному валентному электрону. Ну а концентрацию атомов мы находить умеем:

$$n = \frac{N}{V} = \frac{\nu N_A}{V} = \frac{mN_A}{\mu V} = \frac{\rho N_A}{\mu} = \frac{8900 \cdot 6{,}02 \cdot 10^{23}}{0{,}0635} \approx 8{,}5 \cdot 10^{28}\ \text{м}^{-3}.$$

Положим $S = 1$ мм^2. Из формулы (3.45) получим:

$$v = \frac{I}{enS} = \frac{1}{1{,}6 \cdot 10^{-19} \cdot 8{,}5 \cdot 10^{28} \cdot 10^{-6}} \approx 7{,}4 \cdot 10^{-5}\ \frac{\text{м}}{\text{с}}.$$

Это порядка одной десятой миллиметра в секунду.

3.8.5 Стационарное электрическое поле

Мы всё время говорим о направленном движении зарядов, но ещё не касались вопроса о том, *почему* свободные заряды совершают такое движение. Почему, собственно, возникает электрический ток?

Для упорядоченного перемещения зарядов внутри проводника необходима сила, действующая на заряды в определённом направлении. Откуда берётся эта сила? Со стороны электрического поля!

Чтобы в проводнике протекал постоянный ток, внутри проводника должно существовать стационарное[13] *электрическое поле. Иными словами, между концами проводника нужно поддерживать постоянную разность потенциалов.*

Стационарное электрическое поле должно создаваться зарядами проводников, входящих в электрическую цепь. Однако заряженные проводники сами по себе не смогут обеспечить протекание постоянного тока.

Рассмотрим, к примеру, два проводящих шара, заряженных разноимённо. Соединим их проводом. Между концами провода возникнет разность потенциалов, а внутри провода — электрическое поле. По проводу потечёт ток. Но по мере прохождения тока разность потенциалов между шарами будет уменьшаться, вслед за ней станет убывать и напряжённость поля в проводе. В конце концов потенциалы шаров станут равны друг другу, поле в проводе обратится в нуль, и ток исчезнет. Мы оказались в электростатике: шары плюс провод образуют единый проводник, в каждой точке которого потенциал принимает одно и то же значение; напряжённость поля внутри проводника равна нулю, никакого тока нет.

То, что электростатическое поле само по себе не годится на роль стационарного поля, создающего ток, ясно и из более общих соображений. Ведь электростатическое поле потенциально, его работа при перемещении заряда по замкнутому пути равна нулю. Следовательно, оно не может вызывать циркулирование зарядов по замкнутой электрической цепи — для этого требуется совершать ненулевую работу.

Кто же будет совершать эту ненулевую работу? Кто будет поддерживать в цепи разность потенциалов и обеспечивать стационарное электрическое поле, создающее ток в проводниках? Ответ — источник тока, важнейший элемент электрической цепи.

Чтобы в проводнике протекал постоянный ток, концы проводника должны быть присоединены к клеммам источника тока (батарейки, аккумулятора и т. д.).

Клеммы источника — это заряженные проводники. Если цепь замкнута, то заряды с клемм перемещаются по цепи — как в рассмотренном выше примере с шарами. Но теперь разность потенциалов между клеммами не уменьшается: источник тока непрерывно восполняет заряды на клеммах, поддерживая разность потенциалов между концами цепи на неизменном уровне.

В этом и состоит предназначение источника постоянного тока. Внутри него протекают процессы неэлектрического (чаще всего — химического) происхождения, которые обеспечивают непрерывное разделение зарядов. Эти заряды поставляются на клеммы источника в необходимом количестве.

Количественную характеристику неэлектрических процессов разделения зарядов внутри источника — так называемую ЭДС — мы изучим позже, в соответствующем листке.

А сейчас вернёмся к стационарному электрическому полю. Каким же образом оно возникает в проводниках цепи при наличии источника тока?

Заряженные клеммы источника создают на концах проводника электрическое поле. Свободные заряды проводника, находящиеся вблизи клемм, приходят в движение и действуют своим электрическим полем на соседние заряды. Со скоростью, близкой к скорости света, это взаимодействие передаётся вдоль всей цепи, и в цепи устанавливается постоянный электрический

[13] То есть — постоянное, не зависящее от времени.

ток. Стабилизируется и электрическое поле, создаваемое движущимися зарядами.

Стационарное электрическое поле — это поле свободных зарядов проводника, совершающих направленное движение.

Стационарное электрическое поле не меняется со временем потому, что при постоянном токе не меняется картина распределения зарядов в проводнике: на место заряда, покинувшего данный участок проводника, в следующий момент времени поступает точно такой же заряд. По этой причине стационарное поле во многом (но не во всём) аналогично полю электростатическому.

А именно, справедливы следующие два утверждения, которые понадобятся нам в дальнейшем (их доказательство даётся в вузовском курсе физики).

1. Как и электростатическое поле, стационарное электрическое поле *потенциально*. Это позволяет говорить о разности потенциалов (т. е. напряжении) на любом участке цепи[14].

   Потенциальность, напомним, означает, что работа стационарного поля по перемещению заряда не зависит от формы траектории. Именно поэтому при параллельном соединении проводников напряжение на каждом из них одинаково: оно равно разности потенциалов стационарного поля между теми двумя точками, к которым подключены проводники.

2. В отличие от электростатического поля, стационарное поле движущихся зарядов проникает внутрь проводника. Это объясняется тем, что свободные заряды, участвуя в направленном движении, не успевают должным образом перестраиваться и принимать «электростатические» конфигурации.

   Линии напряжённости стационарного поля внутри проводника параллельны его поверхности, как бы ни изгибался проводник. Поэтому, как и в однородном электростатическом поле, справедлива формула $U = El$, где U — напряжение на концах проводника, E — напряжённость стационарного поля в проводнике, l — длина проводника.

[14] Именно эту разность потенциалов мы измеряем вольтметром.

3.9 Закон Ома

Рассмотрим некоторый элемент электрической цепи постоянного тока. Это может быть что угодно — например, металлический проводник, раствор электролита, лампочка накаливания или газоразрядная трубка.

Будем менять напряжение U, поданное на наш элемент, и измерять силу тока I, протекающего через него. Получим функциональную зависимость $I = I(U)$. Эта зависимость называется *вольт-амперной характеристикой* элемента и является важнейшим показателем его электрических свойств.

Вольт-амперные характеристики различных элементов цепи могут выглядеть по-разному. Очень простой вид имеет вольт-амперная характеристика металлического проводника. Эту зависимость экспериментально установил Георг Ом.

3.9.1 Закон Ома для участка цепи

Оказалось, что сила тока в металлическом проводнике прямо пропорциональна напряжению на его концах: $I \sim U$. Коэффициент пропорциональности принято записывать в виде $1/R$:

$$I = \frac{U}{R}\,. \tag{3.46}$$

Величина R называется *сопротивлением* проводника. Измеряется сопротивление в *омах* (Ом). Как видим, Ом=В/А.

Дадим словесную формулировку закона Ома.

Закон Ома для участка цепи. Сила тока на участке цепи прямо пропорциональна напряжению на этом участке и обратно пропорциональна сопротивлению участка.

Закон Ома оказался справедливым не только для металлов, но и для растворов электролитов.

Сформулированный закон имеет место для так называемого *однородного* участка цепи — участка, не содержащего источников тока. Закон Ома для неоднородного участка (на котором имеется источник тока) мы обсудим позже.

Вольт-амперная характеристика (3.46) является линейной функцией. Её графиком служит прямая линия (рис. 3.35).

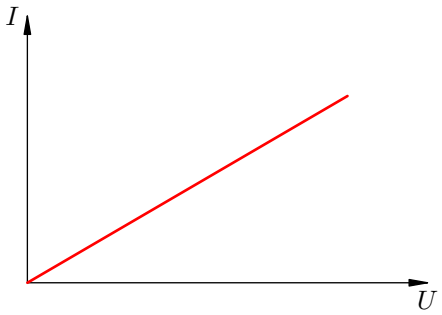


Рис. 3.35. Вольт-амперная характеристика металлического проводника

По этой причине металлические проводники (и электролиты) называются *линейными элементами*. А вот газоразрядная трубка, например, является нелинейным элементом — её вольт-амперная характеристика уже не будет линейной функцией. Но об этом мы поговорим позднее.

3.9.2 Электрическое сопротивление

А сейчас давайте подумаем вот о чём. Пусть к концам проводника приложено постоянное напряжение U. Тогда на свободные заряды проводника действует сила со стороны стационарного электрического поля. Раз есть сила — значит, эти заряды должны двигаться с ускорением; скорость их направленного движения будет увеличиваться, а вместе с ней будет возрастать и сила тока. Но закон Ома гласит, что сила тока будет постоянной. Как же так?

Дело в том, что сила со стороны стационарного поля — не единственная сила, действующая на свободные заряды проводника.

Например, свободные электроны металла, совершая направленное движение, сталкиваются с ионами кристаллической решётки. Возникает своего рода сила сопротивления, действующая со стороны проводника на свободные заряды. Эта сила уравновешивает электрическую силу, с которой на свободные заряды действует стационарное поле. В результате скорость направленного движения заряженных частиц не меняется по модулю[15]; вместе с ней остаётся постоянной и сила тока.

Так что величина R названа сопротивлением не случайно. Она и в самом деле показывает, в какой степени проводник «сопротивляется» прохождению тока.

3.9.3 Удельное сопротивление

Возьмём два проводника из одинакового материала с равными поперечными сечениями; пусть отличаются только их длины. Ясно, что сопротивление будет больше у того проводника, у которого больше длина. В самом деле, при большей длине проводника свободным зарядам труднее пройти сквозь него: каждый свободный электрон встретит на своём пути больше ионов кристаллической решётки. Аналогия такая: чем длиннее заполненная машинами улица, тем труднее будет через неё проехать.

Пусть теперь проводники отличаются только площадью поперечного сечения. Ясно, что чем больше площадь, тем меньше сопротивление проводника. Снова аналогия: чем шире шоссе, тем больше его пропускная способность, т. е. тем меньше его «сопротивление» движению машин.

Опыт подтверждает эти соображения и показывает, что сопротивление проводника прямо пропорционально его длине l и обратно пропорционально площади поперечного сечения S:

$$R = \rho \frac{l}{S}\,. \tag{3.47}$$

Коэффициент пропорциональности ρ уже не зависит от геометрии проводника; он является характеристикой вещества проводника и называется *удельным сопротивлением* данного вещества. Величины удельных сопротивлений различных веществ можно найти в соответствующей таблице.

В каких единицах измеряется удельное сопротивление? Давайте выразим его из формулы (3.47):

$$\rho = \frac{RS}{l}\,.$$

Получим:

$$[\rho] = \frac{\text{Ом} \cdot \text{м}^2}{\text{м}} = \text{Ом} \cdot \text{м}.$$

Однако такая «теоретическая» единица измерения не всегда удобна. Она вынуждает при расчётах переводить площадь поперечного сечения в квадратные метры, тогда как на практике

[15] Точнее говоря, свободные электроны всё же двигаются равноускоренно, но только в промежутках между соударениями с ионами кристаллической решётки. В среднем же оказывается, что электроны перемещаются с постоянной скоростью.

чаще всего речь идёт о квадратных миллиметрах (для проводов, например). На такой случай предусмотрена «практическая» единица:

$$\frac{\text{Ом} \cdot \text{мм}^2}{\text{м}}.$$

В таблице задачника Рымкевича удельное сопротивление даётся как в «теоретических» единицах, так и в «практических».

3.10 Соединения проводников

Есть два основных способа соединения проводников друг с другом — это *последовательное* и *параллельное* соединения. Различные комбинации последовательного и параллельного соединений приводят к *смешанному* соединению проводников.

Мы будем изучать свойства этих соединений, но сначала нам понадобится некоторая вводная информация.

3.10.1 Резисторы и подводящие провода

Проводник, обладающий сопротивлением R, мы называем *резистором* и изображаем следующим образом (рис. 3.36):

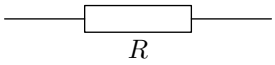


Рис. 3.36. Резистор

Напряжение на резисторе — это разность потенциалов стационарного электрического поля между концами резистора. Между какими именно концами? В общем-то, это неважно, но обычно удобно согласовывать разность потенциалов с направлением тока.

Ток в цепи течёт от «плюса» источника к «минусу». В этом направлении потенциал стационарного поля убывает. Напомним ещё раз, почему это так.

Пусть положительный заряд q перемещается по цепи из точки a в точку b, проходя через резистор R (рис. 3.37):

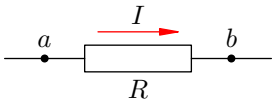


Рис. 3.37. $U = \varphi_a - \varphi_b$

Стационарное поле совершает при этом положительную работу $A = q(\varphi_a - \varphi_b)$. Так как $q > 0$ и $A > 0$, то и $\varphi_a - \varphi_b > 0$, т. е. $\varphi_a > \varphi_b$.

Поэтому напряжение на резисторе мы вычисляем как разность потенциалов в направлении тока: $U = \varphi_a - \varphi_b$.

Сопротивление подводящих проводов обычно пренебрежимо мало; на электрических схемах оно считается равным нулю. Из закона Ома следует тогда, что потенциал не меняется вдоль провода: ведь если $\varphi_a - \varphi_b = IR$ и $R = 0$, то $\varphi_a = \varphi_b$ (рис. 3.38):

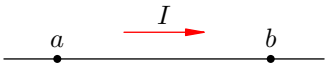


Рис. 3.38. $\varphi_a = \varphi_b$

Таким образом, при рассмотрении электрических цепей мы пользуемся идеализацией, которая сильно упрощает их изучение. А именно, мы считаем, что *потенциал стационарного поля изменяется лишь при переходе через отдельные элементы цепи, а вдоль каждого соединительного провода остаётся неизменным*. В реальных цепях потенциал монотонно убывает при движении от положительной клеммы источника к отрицательной.

3.10.2 Последовательное соединение

При *последовательном соединении* проводников конец каждого проводника соединяется с началом следующего за ним проводника.

Рассмотрим два резистора R_1 и R_2, соединённых последовательно и подключённых к источнику постоянного напряжения U (рис. 3.39). Напомним, что положительная клемма источника обозначается более длинной чертой, так что ток в данной схеме течёт по часовой стрелке.

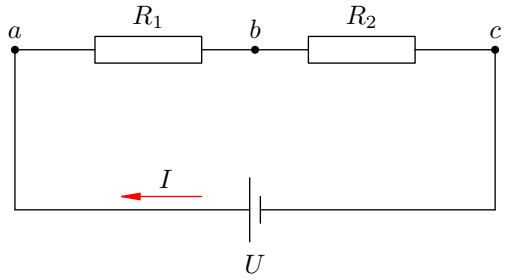


Рис. 3.39. Последовательное соединение

Сформулируем основные свойства последовательного соединения и проиллюстрируем их на этом простом примере.

1. *При последовательном соединении проводников сила тока в них одинакова.*

   В самом деле, через любое поперечное сечение любого проводника за одну секунду будет проходить один и тот же заряд. Ведь заряды нигде не накапливаются, из цепи наружу не уходят и не поступают в цепь извне.

2. *Напряжение на участке, состоящем из последовательно соединённых проводников, равно сумме напряжений на каждом проводнике.*

   Действительно, напряжение U_{ab} на участке ab — это работа поля по переносу единичного заряда из точки a в точку b; напряжение U_{bc} на участке bc — это работа поля по переносу единичного заряда из точки b в точку c. Складываясь, эти две работы дадут работу поля по переносу единичного заряда из точки a в точку c, то есть напряжение U на всём участке: $U = U_{ab} + U_{bc}$.

   Можно и более формально, без всяких словесных объяснений:

   $$U = U_{ac} = \varphi_a - \varphi_c = (\varphi_a - \varphi_b) + (\varphi_b - \varphi_c) = U_{ab} + U_{bc}.$$

3. *Сопротивление участка, состоящего из последовательно соединённых проводников, равно сумме сопротивлений каждого проводника.*

   Пусть R — сопротивление участка ac. По закону Ома имеем:

   $$R = \frac{U}{I} = \frac{U_{ab} + U_{bc}}{I} = \frac{U_{ab}}{I} + \frac{U_{bc}}{I} = R_1 + R_2,$$

   что и требовалось.

   Можно дать интуитивно понятное объяснение правила сложения сопротивлений на одном частном примере. Пусть последовательно соединены два проводника из одинакового вещества и с одинаковой площадью поперечного сечения S, но с разными длинами l_1 и l_2. Сопротивления проводников равны:

   $$R_1 = \rho\,\frac{l_1}{S}, \quad R_2 = \rho\,\frac{l_2}{S}\,.$$

Эти два проводника образуют единый проводник длиной $l_1 + l_2$ и сопротивлением

$$R = \rho \frac{l_1 + l_2}{S} = \rho \frac{l_1}{S} + \rho \frac{l_2}{S} = R_1 + R_2.$$

Но это, повторяем, лишь частный пример. Сопротивления будут складываться и в самом общем случае — если различны также вещества проводников и их поперечные сечения. Доказательство этого даётся с помощью закона Ома, как показано выше.

Наши доказательства свойств последовательного соединения, приведённые для двух проводников, переносятся без существенных изменений на случай произвольного числа проводников.

3.10.3 Параллельное соединение

При *параллельном соединении* проводников их начала подсоединяются к одной точке цепи, а концы — к другой точке.

Снова рассматриваем два резистора, на сей раз соединённые параллельно (рис. 3.40).

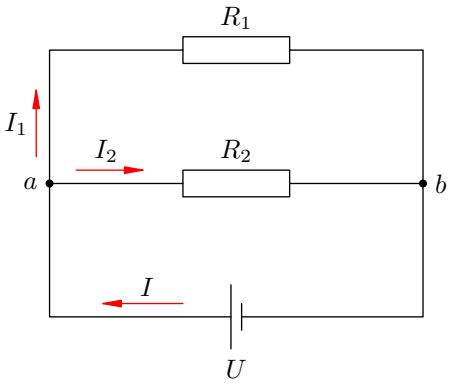


Рис. 3.40. Параллельное соединение

Резисторы подсоединены к двум точкам: a и b. Эти точки называются *узлами* или *точками разветвления* цепи. Параллельные участки называются также *ветвями*; участок от b к a (по направлению тока) называется *неразветвлённой частью* цепи.

Теперь сформулируем свойства параллельного соединения и докажем их для изображённого выше случая двух резисторов.

1. *Напряжение на каждой ветви одинаково и равно напряжению на неразветвлённой части цепи.*

   В самом деле, оба напряжения U_1 и U_2 на резисторах R_1 и R_2 равны разности потенциалов между точками подключения:

   $$U_1 = U_2 = \varphi_a - \varphi_b = U.$$

   Этот факт служит наиболее отчётливым проявлением потенциальности стационарного электрического поля движущихся зарядов.

2. *Сила тока в неразветвлённой части цепи равна сумме сил токов в каждой ветви.*

   Пусть, например, в точку a за время t из неразветвлённого участка поступает заряд q. За это же время t из точки a к резистору R_1 уходит заряд q_1, а к резистору R_2 — заряд q_2.

Ясно, что $q = q_1 + q_2$. В противном случае в точке a накапливался бы заряд, меняя потенциал данной точки, что невозможно (ведь ток постоянный, поле движущихся зарядов стационарно, и потенциал каждой точки цепи не меняется со временем). Тогда имеем:

$$I = \frac{q}{t} = \frac{q_1 + q_2}{t} = \frac{q_1}{t} + \frac{q_2}{t} = I_1 + I_2,$$

что и требовалось.

3. *Величина, обратная сопротивлению участка параллельного соединения, равна сумме величин, обратных сопротивлениям ветвей.*

   Пусть R — сопротивление разветвлённого участка ab. Напряжение на участке ab равно U; ток, текущий через этот участок, равен I. Поэтому:

$$\frac{U}{R} = I = I_1 + I_2 = \frac{U}{R_1} + \frac{U}{R_2}\,.$$

   Сокращая на U, получим:

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}\,, \tag{3.48}$$

   что и требовалось.

   Как и в случае последовательного соединения, можно дать объяснение данного правила на частном примере, не обращаясь к закону Ома.

   Пусть параллельно соединены проводники из одного вещества с одинаковыми длинами l, но разными поперечными сечениями S_1 и S_2. Тогда это соединение можно рассматривать как проводник той же длины l, но с площадью сечения $S = S_1 + S_2$. Имеем:

$$\frac{1}{R} = \frac{S}{\rho l} = \frac{S_1 + S_2}{\rho l} = \frac{S_1}{\rho l} + \frac{S_2}{\rho l} = \frac{1}{R_1} + \frac{1}{R_2}\,.$$

Приведённые доказательства свойств параллельного соединения без существенных изменений переносятся на случай любого числа проводников.

Из соотношения (3.48) можно найти R:

$$R = \frac{R_1 R_2}{R_1 + R_2}\,. \tag{3.49}$$

К сожалению, в общем случае n параллельно соединённых проводников компактного аналога формулы (3.49) не получается, и приходится довольствоваться соотношением

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \ldots + \frac{1}{R_n}\,. \tag{3.50}$$

Тем не менее, один полезный вывод из формулы (3.50) сделать можно. Именно, пусть сопротивления всех n резисторов одинаковы и равны R_1. Тогда:

$$\frac{1}{R} = \underbrace{\frac{1}{R_1} + \frac{1}{R_1} + \ldots + \frac{1}{R_1}}_{n \text{ слагаемых}} = \frac{n}{R_1}\,,$$

откуда

$$R = \frac{R_1}{n}\,.$$

Мы видим, что сопротивление участка из n параллельно соединённых одинаковых проводников в n раз меньше сопротивления одного проводника.

3.10.4 Смешанное соединение

Смешанное сединение проводников, как следует из названия, может являться совокупностью любых комбинаций последовательного и параллельного соединений, причём в состав этих соединений могут входить как отдельные резисторы, так и более сложные составные участки.

Расчёт смешанного соединения опирается на уже известные свойства последовательного и параллельного соединений. Ничего нового тут уже нет: нужно только аккуратно расчленить данную схему на более простые участки, соединённые последовательно или параллельно.

Рассмотрим пример смешанного соединения проводников (рис. 3.41).

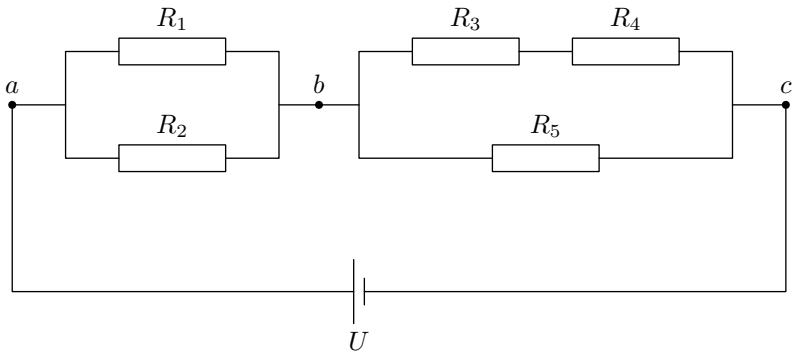


Рис. 3.41. Смешанное соединение

Пусть $U = 14$ В, $R_1 = 2$ Ом, $R_2 = 3$ Ом, $R_3 = 3$ Ом, $R_4 = 5$ Ом, $R_5 = 2$ Ом. Найдём силу тока в цепи и в каждом из резисторов.

Наша цепь состоит из двух последовательно соединённых участков ab и bc. Сопротивление участка ab:

$$R_{ab} = \frac{R_1 R_2}{R_1 + R_2} = \frac{2 \cdot 3}{2 + 3} = 1{,}2 \text{ Ом}.$$

Участок bc является параллельным соединением: два последовательно включённых резистора R_3 и R_4 подключены параллельно к резистору R_5. Тогда:

$$R_{bc} = \frac{(R_3 + R_4)R_5}{(R_3 + R_4) + R_5} = \frac{(3 + 5) \cdot 2}{(3 + 5) + 2} = 1{,}6 \text{ Ом}.$$

Сопротивление цепи:

$$R = R_{ab} + R_{bc} = 1{,}2 + 1{,}6 = 2{,}8 \text{ Ом}.$$

Теперь находим силу тока в цепи:

$$I = \frac{U}{R} = \frac{14}{2{,}8} = 5 \text{ А}.$$

Для нахождения тока в каждом резисторе вычислим напряжения на обоих участках:

$$U_{ab} = IR_{ab} = 5 \cdot 1{,}2 = 6 \text{ В};$$
$$U_{bc} = IR_{bc} = 5 \cdot 1{,}6 = 8 \text{ В}.$$

(Заметим попутно, что сумма этих напряжений равна 14 В, т. е. напряжению в цепи, как и должно быть при последовательном соединении.)

Оба резистора R_1 и R_2 находятся под напряжением U_{ab}, поэтому:

$$I_1 = \frac{U_{ab}}{R_1} = \frac{6}{2} = 3 \text{ А};$$
$$I_2 = \frac{U_{ab}}{R_2} = \frac{6}{3} = 2 \text{ А}.$$

(В сумме имеем 5 А, как и должно быть при параллельном соединении.)

Сила тока в резисторах R_3 и R_4 одинакова, так как они соединены последовательно:

$$I_3 = I_4 = \frac{U_{bc}}{R_3 + R_4} = \frac{8}{3+5} = 1 \text{ А}.$$

Стало быть, через резистор R_5 течёт ток $I_5 = I - I_3 = 5 - 1 = 4$ А.

3.11 Работа и мощность тока

Электрический ток снабжает нас энергией. Сейчас мы будем учиться эту энергию вычислять.

Откуда вообще берётся эта энергия? Она возникает за счёт работы электрического поля по передвижению свободных зарядов в проводнике. Поэтому нахождение работы поля — наша первая задача.

3.11.1 Работа тока

Рассмотрим участок цепи, по которому течёт ток I. Напряжение на участке обозначим U, сопротивление участка равно R (рис. 3.42).

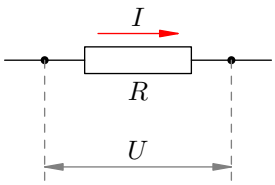


Рис. 3.42. Участок цепи

За время t по нашему участку проходит заряд $q = It$. Заряд перемещается стационарным электрическим полем, которое совершает при этом работу:

$$A = Uq = UIt. \tag{3.51}$$

За счёт работы (3.51) на рассматриваемом участке может выделяться тепловая энергия или совершаться механическая работа; могут также протекать химические реакции. Короче говоря, данная работа идёт на увеличение энергии нашего участка цепи.

Работа (3.51) называется *работой тока*. Термин крайне неудачный — ведь работу совершает не ток, а электрическое поле. Но с укоренившейся терминологией, увы, ничего не поделаешь.

Если участок цепи является *однородным*, т. е. не содержит источника тока, то для этого участка справедлив закон Ома: $U = IR$. Подставляя это в формулу (3.51), получим:

$$A = I^2Rt. \tag{3.52}$$

Теперь подставим в (3.51) вместо тока его выражение из закона Ома $I = U/R$:

$$A = \frac{U^2}{R}t. \tag{3.53}$$

Подчеркнём ещё раз: формула (3.51) получена из самых общих соображений, она является основной и годится для любого участка цепи. А вот формулы (3.52) и (3.53) получены из основной формулы с дополнительным привлечением закона Ома и потому годятся только для однородного участка.

3.11.2 Мощность тока

Как вы помните, *мощностью* называется отношение работы ко времени её совершения. В частности, *мощность тока* — это отношение работы тока ко времени, за которое эта работа совершена:

$$P = \frac{A}{t}\,.$$

Из формул (3.51)–(3.53) немедленно получаем соответствующие формулы для мощности тока:

$$P = UI; \tag{3.54}$$
$$P = I^2R; \tag{3.55}$$
$$P = \frac{U^2}{R}. \tag{3.56}$$

3.11.3 Закон Джоуля–Ленца

Предположим, что на рассматриваемом участке цепи не совершается механическая работа и не протекают химические реакции. Поскольку сила тока постоянна, работа поля не вызывает увеличение кинетической энергии свободных зарядов. Стало быть, работа поля A *целиком* превращается в тепло Q, которое выделяется на данном участке цепи и рассеивается в окружающее пространство: $A = Q$.

Таким образом, для количества теплоты, выделяющегося на данном участке цепи, мы получаем формулы:

$$Q = UIt; \tag{3.57}$$
$$Q = I^2Rt; \tag{3.58}$$
$$Q = \frac{U^2}{R}t. \tag{3.59}$$

Но часто бывает так, что не вся работа тока превращается в тепло. Например, за счёт работы тока может совершать механическую работу электродвигатель или заряжаться аккумулятор. Тепло, разумеется, будет выделяться и в этих случаях, но только на сей раз получится, что $Q < A$ (на величину механической работы, совершённой двигателем, или химической энергии, запасённой аккумулятором).

Оказывается, что в подобных случаях остаётся справедливой формула (3.58): $Q = I^2Rt$. Это — экспериментально установленный *закон Джоуля-Ленца*.

3.12 ЭДС. Закон Ома для полной цепи

До сих пор при изучении электрического тока мы рассматривали направленное движение свободных зарядов во *внешней цепи*, то есть в проводниках, подсоединённых к клеммам источника тока.

Как мы знаем, положительный заряд q:

- уходит во внешнюю цепь с положительной клеммы источника;
- перемещается во внешней цепи под действием стационарного электрического поля, создаваемого другими движущимися зарядами;
- приходит на отрицательную клемму источника, завершая свой путь во внешней цепи.

Теперь нашему положительному заряду q нужно замкнуть свою траекторию и вернуться на положительную клемму. Для этого ему требуется преодолеть заключительный отрезок пути — внутри источника тока от отрицательной клеммы к положительной. Но вдумайтесь: идти туда ему совсем не хочется! Отрицательная клемма притягивает его к себе, положительная клемма его от себя отталкивает, и в результате на наш заряд внутри источника действует электрическая сила $\vec{F}_{\text{эл}}$, направленная *против* движения заряда (т. е. против направления тока).

3.12.1 Сторонняя сила

Тем не менее, ток по цепи идёт; стало быть, имеется сила, «протаскивающая» заряд сквозь источник вопреки противодействию электрического поля клемм (рис. 3.43).

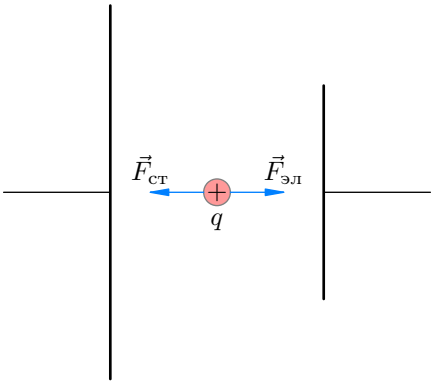


Рис. 3.43. Сторонняя сила

Эта сила называется *сторонней силой*; именно благодаря ей и функционирует источник тока. Сторонняя сила $\vec{F}_{\text{ст}}$ не имеет отношения к стационарному электрическому полю — у неё, как говорят, *неэлектрическое* происхождение; в батарейках, например, она возникает благодаря протеканию соответствующих химических реакций.

Обозначим через $A_{\text{ст}}$ работу сторонней силы по перемещению положительного заряда q внутри источника тока от отрицательной клеммы к положительной. Эта работа положительна, так как направление сторонней силы совпадает с направлением перемещения заряда. Работа сторонней силы $A_{\text{ст}}$ называется также *работой источника тока*.

Во внешней цепи сторонняя сила отсутствует, так что работа сторонней силы по перемещению заряда во внешней цепи равна нулю. Поэтому работа сторонней силы по перемещению заряда q вокруг всей цепи сводится к работе по перемещению этого заряда только лишь внутри источника тока. Таким образом, $A_{\text{ст}}$ — это также работа сторонней силы по перемещению заряда *по всей цепи*.

Мы видим, что сторонняя сила является непотенциальной — её работа при перемещении заряда по замкнутому пути не равна нулю. Именно эта непотенциальность и обеспечивает циркулирование электрического тока; потенциальное электрическое поле, как мы уже говорили ранее, не может поддерживать постоянный ток.

Опыт показывает, что работа $A_{\text{ст}}$ прямо пропорциональна перемещаемому заряду q. Поэтому отношение $A_{\text{ст}}/q$ уже не зависит от заряда и является количественной характеристикой источника тока. Это отношение обозначается $\mathcal{E}$:

$$\mathcal{E} = \frac{A_{\text{ст}}}{q}\,. \tag{3.60}$$

Данная величина называется *электродвижущей силой* (ЭДС) источника тока. Как видим, ЭДС измеряется в вольтах (В), поэтому название «электродвижущая сила» является крайне неудачным. Но оно давно укоренилось, так что приходится смириться.

Когда вы видите надпись на батарейке: «1,5 В», то знайте, что это именно ЭДС. Равна ли эта величина напряжению, которое создаёт батарейка во внешней цепи? Оказывается, нет! Сейчас мы поймём, почему.

3.12.2 Закон Ома для полной цепи

Любой источник тока обладает своим сопротивлением r, которое называется *внутренним сопротивлением* этого источника. Таким образом, источник тока имеет две важных характеристики: ЭДС и внутреннее сопротивление.

Пусть источник тока с ЭДС, равной $\mathcal{E}$, и внутренним сопротивлением r подключён к резистору R (который в данном случае называется *внешним резистором*, или *внешней нагрузкой*, или *полезной нагрузкой*). Всё это вместе называется *полной цепью* (рис. 3.44).

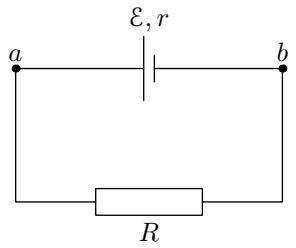


Рис. 3.44. Полная цепь

Наша задача — найти силу тока I в цепи и напряжение U на резисторе R.

За время t по цепи проходит заряд $q = It$. Согласно формуле (3.60) источник тока совершает при этом работу:

$$A_{\text{ст}} = \mathcal{E}q = \mathcal{E}It. \tag{3.61}$$

Так как сила тока постоянна, работа источника целиком превращается в теплоту, которая выделяется на сопротивлениях R и r. Данное количество теплоты определяется законом Джоуля–Ленца:

$$Q = I^2Rt + I^2rt = I^2(R+r)t. \tag{3.62}$$

Итак, $A_{\text{ст}} = Q$, и мы приравниваем правые части формул (3.61) и (3.62):

$$\mathcal{E}It = I^2(R+r)t.$$

После сокращения на It получаем:

$$\mathcal{E} = I(R+r).$$

Вот мы и нашли ток в цепи:

$$I = \frac{\mathcal{E}}{R+r}\,. \tag{3.63}$$

Формула (3.63) называется *законом Ома для полной цепи*.

Если соединить клеммы источника проводом пренебрежимо малого сопротивления ($R = 0$), то получится *короткое замыкание*. Через источник при этом потечёт максимальный ток — *ток короткого замыкания*:

$$I_{\text{кз}} = \frac{\mathcal{E}}{r}\,.$$

Из-за малости внутреннего сопротивления ток короткого замыкания может быть весьма большим. Например, пальчиковая батарейка разогревается при этом так, что обжигает руки.

Зная силу тока (формула (3.63)), мы можем найти напряжение на резисторе R с помощью закона Ома для участка цепи:

$$U = IR = \frac{\mathcal{E}R}{R+r}\,. \tag{3.64}$$

Это напряжение является разностью потенциалов между точками a и b (рис. 2). Потенциал точки a равен потенциалу положительной клеммы источника; потенциал точки b равен потенциалу отрицательной клеммы. Поэтому напряжение (3.64) называется также *напряжением на клеммах источника*.

Мы видим из формулы (3.64), что в реальной цепи будет $U < \mathcal{E}$ — ведь $\mathcal{E}$ умножается на дробь, меньшую единицы. Но есть два случая, когда $U = \mathcal{E}$.

1. *Идеальный источник тока*. Так называется источник с нулевым внутренним сопротивлением. При $r = 0$ формула (3.64) даёт $U = \mathcal{E}$.

2. *Разомкнутая цепь*. Рассмотрим источник тока сам по себе, вне электрической цепи. В этом случае можно считать, что внешнее сопротивление бесконечно велико: $R = \infty$. Тогда величина $R + r$ неотличима от R, и формула (3.64) снова даёт нам $U = \mathcal{E}$.

   Смысл этого результата прост: *если источник не подключён к цепи, то идеальный вольтметр*[16]*, подсоединённый к полюсам источника, покажет его ЭДС*.

3.12.3 КПД электрической цепи

Нетрудно понять, почему резистор R называется полезной нагрузкой. Представьте себе, что это лампочка. Теплота, выделяющаяся на лампочке, является *полезной*, так как благодаря этой теплоте лампочка выполняет своё предназначение — даёт свет.

Количество теплоты, выделяющееся на полезной нагрузке R за время t, обозначим $Q_{\text{полезн}}$. Если сила тока в цепи равна I, то

$$Q_{\text{полезн}} = I^2Rt.$$

Некоторое количество теплоты выделяется также на источнике тока:

$$Q_{\text{ист}} = I^2rt.$$

Полное количество теплоты, которое выделяется в цепи, равно:

$$Q_{\text{полн}} = Q_{\text{полезн}} + Q_{\text{ист}} = I^2Rt + I^2rt = I^2(R+r)t.$$

КПД электрической цепи — это отношение полезного тепла к полному:

$$\eta = \frac{Q_{\text{полезн}}}{Q_{\text{полн}}} = \frac{I^2Rt}{I^2(R+r)t} = \frac{R}{R+r}\,.$$

КПД цепи равен единице лишь в том случае, если источник тока идеальный ($r = 0$).

[16] Вольтметр называется *идеальным*, если его сопротивление бесконечно велико. Подключение идеального вольтметра не приводит к искажениям тока в цепи.

3.12.4 Закон Ома для неоднородного участка

Простой закон Ома $U = IR$ справедлив для так называемого однородного участка цепи — то есть участка, на котором нет источников тока. Сейчас мы получим более общие соотношения, из которых следует как закон Ома для однородного участка, так и полученный выше закон Ома для полной цепи.

Участок цепи называется *неоднородным*, если на нём имеется источник тока. Иными словами, неоднородный участок — это участок с ЭДС.

На рис. 3.45 показан неоднородный участок, содержащий резистор R и источник тока. ЭДС источника равна $\mathcal{E}$, его внутреннее сопротивление считаем равным нулю[17].

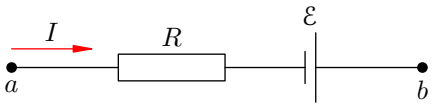


Рис. 3.45. ЭДС «помогает» току: $\varphi_a - \varphi_b + \mathcal{E} = IR$

Сила тока на участке равна I, ток течёт от точки a к точке b. Этот ток не обязательно вызван одним лишь источником $\mathcal{E}$. Рассматриваемый участок, как правило, входит в состав некоторой цепи (не изображённой на рисунке), а в этой цепи могут присутствовать и другие источники тока. Поэтому ток I является результатом совокупного действия *всех* источников, имеющихся в цепи.

Пусть потенциалы точек a и b равны соответственно φ_a и φ_b. Подчеркнём ещё раз, что речь идёт о потенциале стационарного электрического поля, порождённого действием всех источников цепи — не только источника, принадлежащего данному участку, но и, возможно, имеющихся вне этого участка.

Напряжение на нашем участке равно: $U = \varphi_a - \varphi_b$. За время t через участок проходит заряд $q = It$, при этом стационарное электрическое поле совершает работу:

$$A_{\text{поля}} = Uq = UIt.$$

Кроме того, положительную работу совершает источник тока (ведь заряд q прошёл сквозь него!):

$$A_{\text{ст}} = \mathcal{E}q = \mathcal{E}It.$$

Сила тока постоянна, поэтому суммарная работа по продвижению заряда q, совершаемая на участке стационарным элетрическим полем и сторонними силами источника, целиком превращается в тепло: $A_{\text{поля}} + A_{\text{ст}} = Q$. Подставляем сюда выражения для $A_{\text{поля}}$, $A_{\text{ст}}$ и закон Джоуля–Ленца:

$$UIt + \mathcal{E}It = I^2Rt.$$

Сокращая на It, получаем *закон Ома для неоднородного участка цепи*:

$$U + \mathcal{E} = IR, \tag{3.65}$$

или, что то же самое:

$$\varphi_a - \varphi_b + \mathcal{E} = IR. \tag{3.66}$$

Обратите внимание: перед $\mathcal{E}$ стоит знак «плюс». Причину этого мы уже указывали — источник тока в данном случае совершает *положительную* работу, «протаскивая» внутри себя заряд q от отрицательной клеммы к положительной. Попросту говоря, источник «помогает» току протекать от точки a к точке b.

Отметим два следствия выведенных формул (3.65) и (3.66).

[17] Если внутреннее сопротивление источника равно r, можно просто заменить резистор R на резистор $R + r$.

1. Если участок однородный, то $\mathcal{E} = 0$. Тогда из формулы (3.65) получаем $U = IR$ — закон Ома для однородного участка цепи.

2. Предположим, что источник тока обладает внутренним сопротивлением r. Это, как мы уже упоминали, равносильно замене R на $R + r$:

$$\varphi_a - \varphi_b + \mathcal{E} = I(R + r).$$

Теперь замкнём наш участок, соединив точки a и b. Получим рассмотренную выше полную цепь. При этом окажется, что $\varphi_a = \varphi_b$, и предыдущая формула превратится в закон Ома для полной цепи:

$$\mathcal{E} = I(R + r).$$

Таким образом, закон Ома для однородного участка и закон Ома для полной цепи оба вытекают из закона Ома для неоднородного участка.

Может быть и другой случай подключения, когда источник $\mathcal{E}$ «мешает» току идти по участку. Такая ситуация изображена на рис. 3.46. Здесь ток, идущий от a к b, направлен против действия сторонних сил источника.

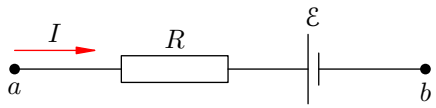


Рис. 3.46. ЭДС «мешает» току: $\varphi_a - \varphi_b - \mathcal{E} = IR$

Как такое возможно? Очень просто: другие источники, имеющиеся в цепи вне рассматриваемого участка, «пересиливают» источник на участке и вынуждают ток течь против $\mathcal{E}$. Именно так происходит, когда вы ставите телефон на зарядку: подключённый к розетке адаптер вызывает движение зарядов против действия сторонних сил аккумулятора телефона, и аккумулятор тем самым заряжается!

Что изменится теперь в выводе наших формул? Только одно — работа сторонних сил станет отрицательной:

$$A_{\text{ст}} = -\mathcal{E}q = -\mathcal{E}It.$$

Тогда закон Ома для неоднородного участка примет вид:

$$\varphi_a - \varphi_b - \mathcal{E} = IR, \tag{3.67}$$

или:

$$U - \mathcal{E} = IR,$$

где по-прежнему $U = \varphi_a - \varphi_b$ — напряжение на участке.

Давайте соберём вместе формулы (3.66) и (3.67) и запишем закон Ома для участка с ЭДС следующим образом:

$$\varphi_a - \varphi_b \pm \mathcal{E} = IR.$$

Ток при этом течёт от точки a к точке b. Если направление тока совпадает с направлением сторонних сил, то перед $\mathcal{E}$ ставится «плюс»; если же эти направления противоположны, то ставится «минус».

3.13 Электрический ток в металлах

В этом разделе мы приступаем к подробному изучению того, как осуществляется прохождение электрического тока в различных проводящих средах — твёрдых телах, жидкостях и газах.

Напомним, что необходимым условием возникновения тока является наличие в среде достаточно большого количества *свободных зарядов*, которые могут начать упорядоченное движение под действием электрического поля. Такие среды как раз и называются *проводниками* электрического тока.

Наиболее широко распространены металлические проводники. Поэтому начинаем мы с вопросов распространения электрического тока в металлах.

Мы много раз говорили о *свободных электронах*, которые являются носителями свободных зарядов в металлах. Вам хорошо известно, что электрический ток в металлическом проводнике образуется в результате направленного движения свободных электронов.

3.13.1 Свободные электроны

Металлы в твёрдом состоянии имеют кристаллическую структуру: расположение атомов в пространстве характеризуется периодической повторяемостью и образует геометрически правильный рисунок, называемый *кристаллической решёткой*.

Атомы металлов имеют небольшое число *валентных электронов*, расположенных на внешней электронной оболочке. Эти валентные электроны слабо связаны с ядром, и атом легко может их потерять.

Когда атомы металла занимают места в кристаллической решётке, валентные электроны покидают свои оболочки — они становятся *свободными* и отправляются «гулять» по всему кристаллу[18]. В узлах кристаллической решётки металла остаются *положительные ионы*, пространство между которыми заполнено «газом» свободных электронов (рис. 3.47).

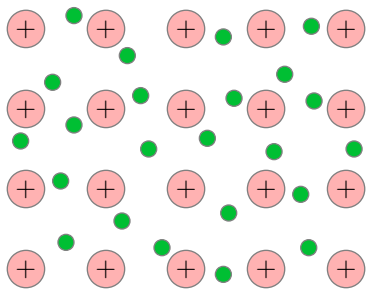


Рис. 3.47. Свободные электроны

Свободные электроны и впрямь ведут себя подобно частицам газа[19] — совершая тепловое движение, они хаотически снуют туда-сюда между ионами кристаллической решётки. Суммарный заряд свободных электронов равен по модулю и противоположен по знаку общему заряду положительных ионов, поэтому металлический проводник в целом оказывается электрически нейтральным.

Газ свободных электронов является «клеем», на котором держится вся кристаллическая структура проводника. Ведь положительные ионы отталкиваются друг от друга, так что кристаллическая решётка, распираемая изнутри мощными кулоновскими силами, могла бы разлететься в разные стороны. Однако в то же самое время ионы металла притягиваются к обволаки-

[18] А именно, свободные электроны перемещаются по внешним орбиталям соседних атомов. Эти орбитали перекрываются друг с другом вследствие близкого расположения атомов в кристаллической решётке, так что свободные электроны оказываются «общей собственностью» всего кристалла.

[19] Другой адекватный образ — *электронное море*, которое «омывает» кристаллическую решётку.

вающему их электронному газу и, как ни в чём не бывало, остаются на своих местах, совершая лишь тепловые колебания в узлах кристаллической решётки вблизи положений равновесия.

Что произойдёт, если металлический проводник включить в замкнутую цепь, содержащую источник тока? Свободные электроны продолжают совершать хаотическое тепловое движение, но теперь — под действием возникшего внешнего электрического поля — они вдобавок начнут перемещаться *упорядоченно*. Это направленное течение электронного газа, накладывающееся на тепловое движение электронов, и есть *электрический ток в металле*[20]. Скорость упорядоченного движения электронов в металлическом проводнике, как нам уже известно, составляет порядка 0,1 мм/с.

3.13.2 Опыт Рикке

Почему мы решили, что ток в металлах создаётся движением именно свободных электронов? Положительные ионы кристаллической решётки также испытывают на себе действие внешнего электрического поля. Может, они тоже перемещаются внутри металлического проводника и участвуют в создании тока?

Упорядоченное движение ионов означало бы постепенный перенос вещества вдоль направления электрического тока. Поэтому надо просто пропускать ток по проводнику на протяжении весьма длительного времени и посмотреть, что в итоге получится. Такого рода эксперимент и был поставлен Э. Рикке в 1901 году.

В электрическую цепь были включены три прижатых друг к другу цилиндра: два медных по краям и один алюминиевый между ними (рис. 3.48). По этой цепи пропускался электрический ток в течение года.

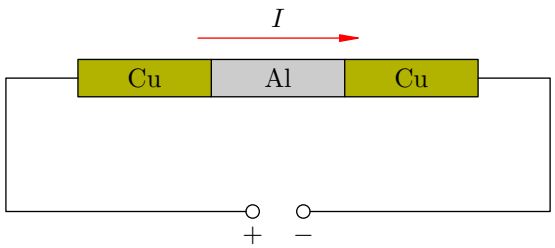


Рис. 3.48. Опыт Рикке

За год сквозь цилиндры прошёл заряд более трёх миллионов кулон. Предположим, что каждый атом металла теряет по одному валентному электрону, так что заряд иона равен элементарному заряду $e = 1{,}6 \cdot 10^{-19}$ Кл. Если ток создаётся движением положительных ионов, то нетрудно подсчитать (сделайте это сами!), что такая величина прошедшего по цепи заряда соответствует переносу вдоль цепи около 2 кг меди.

Однако после разъединения цилиндров было обнаружено лишь незначительное проникновение металлов друг в друга, обусловленное естественной диффузией их атомов (и не более того). *Электрический ток в металлах не сопровождается переносом вещества*, поэтому положительные ионы металла не принимают участия в создании тока.

3.13.3 Опыт Стюарта–Толмена

Прямое экспериментальное доказательство того, что электрический ток в металлах создаётся движением свободных электронов, было дано в опыте Т. Стюарта и Р. Толмена (1916 год).

[20] Поэтому свободные электроны называются также *электронами проводимости*.

Эксперименту Стюарта–Толмена предшествовали качественные наблюдения, сделанные четырьмя годами ранее русскими физиками Л. И. Мандельштамом и Н. Д. Папалекси. Они обратили внимание на так называемый *электроинерционный эффект*: если резко затормозить движущийся проводник, то в нём возникает кратковременный импульс тока. Эффект объясняется тем, что в течение небольшого времени после торможения проводника его свободные заряды продолжают двигаться по инерции.

Однако никаких количественных результатов Мандельштам и Папалекси не получили, и наблюдения их опубликованы не были. Честь назвать опыт своим именем принадлежит Стюарту и Толмену, которые не только наблюдали указанный электроинерционный эффект, но и произвели необходимые измерения и расчёты.

Установка Стюарта и Толмена показана на рис. 3.49. Катушка большим числом витков металлического провода приводилась в быстрое вращение вокруг своей оси. Концы обмотки с помощью скользящих контактов были подсоединены к специальному прибору — *баллистическому гальванометру*, который позволяет измерять проходящий через него заряд.

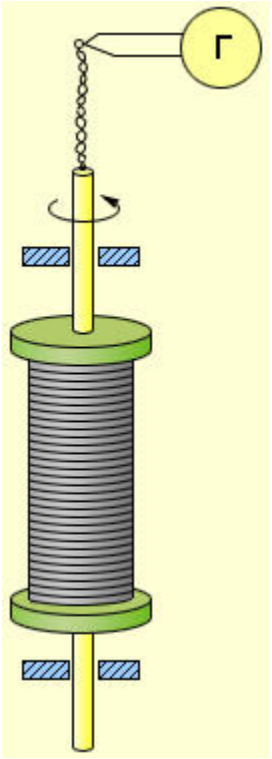


Рис. 3.49. Опыт Стюарта–Толмена

После резкого торможения катушки в цепи возникал импульс тока. Направление тока указывало на то, что он вызван движением *отрицательных* зарядов. Измеряя баллистическим гальванометром суммарный заряд, проходящий по цепи, Стюарт и Толмен вычислили отношение q/m заряда одной частицы к её массе. Оно оказалось равно отношению e/m для электрона, которое в то время уже было хорошо известно.

Так было окончательно выяснено, что *носителями свободных зарядов в металлах являются свободные электроны*. Как видите, этот давно и хорошо знакомый вам факт был установлен сравнительно поздно — учитывая, что металлические проводники к тому моменту уже более столетия активно использовались в самых разнообразных экспериментах по электромагнетизму[21].

3.13.4 Зависимость сопротивления от температуры

Опыт показывает, что при нагревании металлического проводника его сопротивление увеличивается. Как это объяснить?

Причина проста: с повышением температуры тепловые колебания ионов кристаллической решётки становятся более интенсивными, так что число соударений свободных электронов с ионами возрастает. Чем активнее тепловое движение решётки, тем труднее электронам пробираться сквозь промежутки между ионами[22]. Скорость упорядоченного движения электронов уменьшается, поэтому уменьшается и сила тока (при неизменном напряжении). Это и означает увеличение сопротивления.

[21] Сравните, например, с датой открытия закона Ома — 1826 год. Дело, однако, заключается в том, что сам электрон был открыт лишь в 1897 году.

[22] Представьте себе вращающуюся проходную дверь. В каком случае труднее проскочить через неё: когда она вращается медленно или быстро? :-)

Как опять-таки показывает опыт, зависимость сопротивления R металлического проводника от температуры t с хорошей точностью является линейной:

$$R = R_0(1 + \alpha t). \qquad (3.68)$$

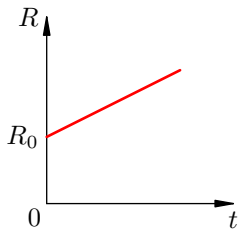


Рис. 3.50. $R = R(t)$

Здесь R_0 — сопротивление проводника при 0 °C. График зависимости (3.68) является прямой линией (рис. 3.50).

Множитель α называется *температурным коэффициентом сопротивления*. Его значения для различных металлов и сплавов можно найти в таблицах.

Длина проводника l и его площадь поперечного сечения S при изменении температуры меняются несущественно. Выразим R и R_0 через удельное сопротивление:

$$R = \rho\frac{l}{S}, \quad R_0 = \rho_0\frac{l}{S},$$

и подставим эти формулы в (3.68). Получим аналогичную зависимость удельного сопротивления от температуры:

$$\rho = \rho_0(1 + \alpha t).$$

Коэффициент α весьма мал (для меди, например, $\alpha = 0{,}0043$), так что температурной зависимостью сопротивления металла часто можно пренебречь. Однако в ряде случаев считаться с ней приходиться. Например, вольфрамовая спираль электрической лампочки раскаляется до такой степени, что её вольт-амперная характеристика оказывается существенно нелинейной.

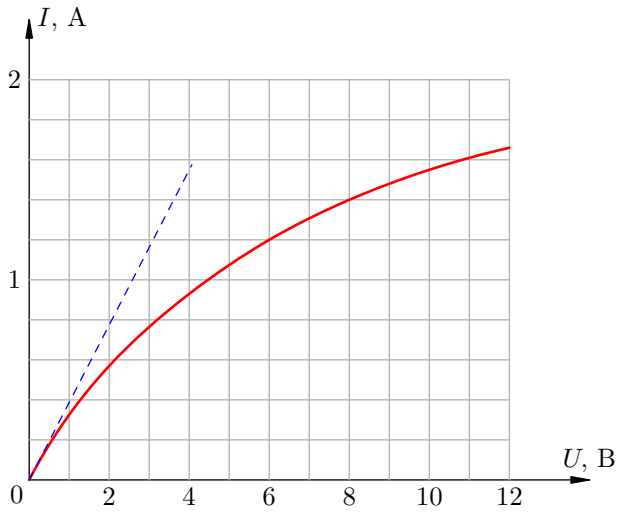


Рис. 3.51. Вольт-амперная характеристика лампочки

Так, на рис. 3.51 приведена вольт-амперная характеристика автомобильной лампочки. Если бы лампочка представляла собой идеальный резистор, её вольт-амперная характеристика была прямой линией в соответствии с законом Ома. Эта прямая изображена синим пунктиром.

Однако по мере роста напряжения, приложенного к лампочке, график отклоняется от этой прямой всё сильнее и сильнее. Почему? Дело в том, что с увеличением напряжения ток через лампочку возрастает и больше разогревает спираль; сопротивление спирали поэтому также увеличивается. Следовательно, сила тока хотя и продолжит возрастать, но будет иметь всё меньшее и меньшее значение по сравнению с тем, которое предписывается «пунктирной» линейной зависимостью тока от напряжения.

3.14 Электрический ток в электролитах

Электролитом мы называем раствор (или расплав) вещества, через который может идти электрический ток; при этом исходное вещество проводником тока *не является*.

Например, кристаллы поваренной соли NaCl не проводят ток. Дистиллированная вода — тоже диэлектрик. Однако при растворении соли в воде получается среда, через которую ток отлично проходит! Следовательно, солёная вода будет электролитом[23].

Электролитами оказываются растворы солей, кислот и оснований. Прохождение тока через эти растворы означает, что в них имеются свободные заряды. Откуда же они там берутся, если ни в воде, ни в исходном веществе свободных зарядов не было?

3.14.1 Электролитическая диссоциация

Механизм, обеспечивающий появление свободных зарядов в электролите, называется *электролитической диссоциацией*. Мы ограничимся рассмотрением электролитической диссоциации в растворах.

Вообще, диссоциация — это распад молекулы на составные части под влиянием тех или иных факторов. *В процессе электролитической диссоциации молекулы растворяемого вещества распадаются на положительные и отрицательные ионы в результате действия электрических сил со стороны молекул воды.*

Многие свойства воды объясняются тем, что её молекулы являются *полярными*, то есть в электрическом отношении ведут себя как диполи (напомним, что *диполь* — это система двух одинаковых по модулю и противоположных по знаку зарядов, расположенных на небольшом расстоянии друг от друга). Полярность молекул H_2O обусловлена их геометрическим устройством (рис. 3.52)[24].

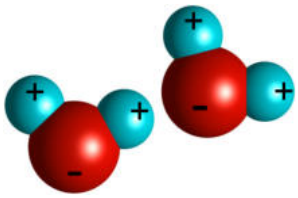


Рис. 3.52. Молекулы воды

Угол, образованный линиями центров атома кислорода и двух атомов водорода, составляет примерно 104,5°. Кроме того, электронные оболочки смещены в сторону кислорода. В результате центры положительных и отрицательных зарядов оказываются пространственно разделёнными: «минусы» преобладают в кислородной части молекулы воды, а «плюсы» — в водородной части.

Будучи диполями, молекулы воды создают вокруг себя электрическое поле и действуют электрическими силами как друг на друга, так и на молекулы примесей[25].

Почему же соли, кислоты и основания распадаются в воде на ионы? Всё дело в том, что молекулы этих веществ также являются полярными. Давайте вернёмся к нашему примеру с растворением поваренной соли NaCl.

[23] Вот почему запрещено купаться во время грозы — в водоёмах всегда растворено некоторое количество солей. При ударе молнии по воде пойдёт электрический ток.

[24] Изображение с сайта howyourbrainworks.net.

[25] Притягиваясь друг к другу противоположно заряженными частями, молекулы воды создают весьма прочные связи. Вот почему столь велики удельная теплоёмкость и удельная теплота парообразования воды — на разрыв этих связей требуется значительная энергия.

У атома натрия на внешнем электронном уровне находится один электрон. Он слабо связан с атомом и всегда готов покинуть место своего обитания. У атома хлора на внешнем электронном уровне семь электронов — одного как раз не хватает до полного комплекта. Атом хлора всегда готов захватить себе недостающий электрон.

Поэтому при образовании молекулы NaCl внешний электрон атома натрия уходит к атому хлора, и в результате молекула становится полярной — она состоит из положительного иона Na^+ и отрицательного иона Cl^-. Эта молекула схематически изображена на рис. 3.53 (атом хлора крупнее, чем атом натрия).

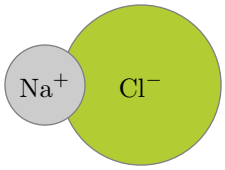


Рис. 3.53. Молекула NaCl

Иными словами, с электрической точки зрения молекула NaCl также оказывается диполем. Взаимодействие двух сортов диполей — молекул H_2O и NaCl — как раз и вызывает процесс растворения.

На рис. 3.54 мы видим, как протекает этот процесс[26]. Более крупные зелёные шарики изображают ионы хлора, более мелкие серые — ионы натрия.

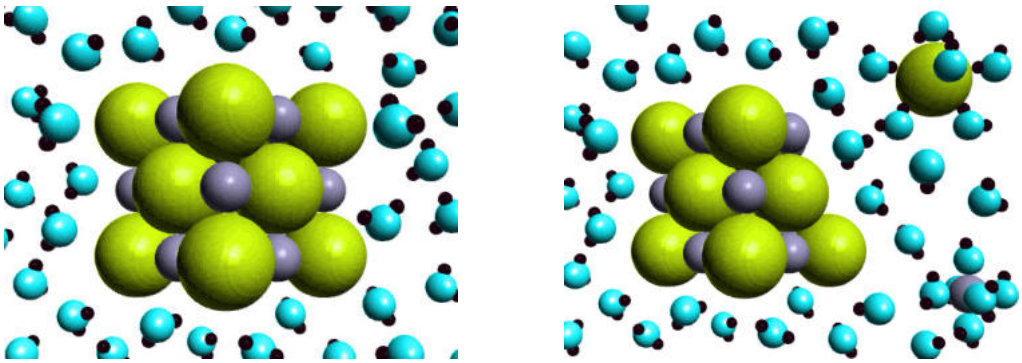


Рис. 3.54. Электролитическая диссоциация: растворение NaCl в воде

Левая часть рисунка показывает ситуацию до начала растворения. Крупица соли в виде небольшого кубического кристаллика[27] только что оказалась в воде.

Сразу же начинается «электрическая атака» со стороны молекул воды. Отрицательно заряженные (кислородные) части молекул H_2O обращаются к положительным ионам натрия, а положительные (водородные) части молекулы воды — к отрицательным ионам хлора. Молекулы NaCl начинают «растягиваться» разнонаправленными электрическими силами притяжения к молекулам-диполям воды, и связь между ионами натрия и хлора, скрепляющая молекулу соли, ослабевает.

В конце концов эта связь становится настолько слабой, что удары соседних частиц, совершающих тепловое движение, разрушают молекулу NaCl. Она распадается на положительный ион Na^+ и отрицательный ион Cl^-.

В правой части рис. 3.54 мы видим результат такого распада: вырванные из кристаллической решётки ион хлора и ион натрия отправляются «в свободное плавание», окружённые

[26]Изображение с сайта intro.chem.okstate.edu.

[27]Кристаллическая решётка поваренной соли имеет кубическую структуру. В вершинах куба в шахматном порядке, крест-накрест расположены атомы натрия и хлора.

прицепившимися к ним молекулами воды. Обратите внимание, что молекулы воды прилипли к отрицательному иону хлора своими положительными водородными частями, а к положительному иону натрия, наоборот, повернулись их отрицательные кислородные части.

Таким образом, при растворении соли NaCl в воде появляются свободные заряды: положительные ионы Na^+ и отрицательные ионы Cl^- (рис. 3.55). Это и является необходимым условием прохождения тока через раствор.

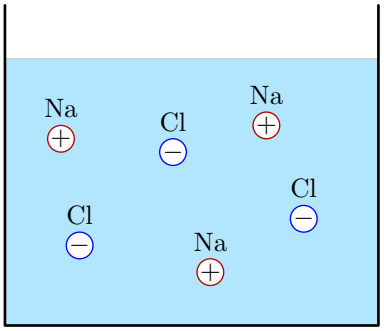


Рис. 3.55. Раствор NaCl в воде

Описанный выше процесс растворения совершенно аналогично протекает и в случае других примесей. Так, щёлочь KOH распадается в водном растворе на положительные ионы калия K^+ и отрицательные ионы гидроксильной группы OH^-. Молекула серной кислоты H_2SO_4 при распаде даёт два положительных иона H^+ и отрицательно заряженный ион кислотного остатка SO_4^{2-}. В растворе медного купороса $CuSO_4$ появляются положительные ионы меди Cu^{2+} и отрицательно заряженные ионы SO_4^{2-}.

Все ли молекулы растворяемого вещества распадутся на ионы? Это зависит от ряда условий. *Степенью диссоциации* называется отношение числа распавшихся молекул к общему начальному числу молекул. При полном растворении вещества степень диссоциации равна 1.

Наряду с диссоциацией имеет место и обратный процесс: *рекомбинация*. А именно, две частицы противоположного знака могут встретиться и снова образовать нейтральную молекулу (рекомбинировать). С течением времени в растворе устанавливается состояние динамического равновесия: среднее число диссоциаций в единицу времени равно среднему числу рекомбинаций, в результате чего концентрация раствора остаётся неизменной (вспомните аналогичную ситуацию с насыщенным паром: при динамическом равновесии пара и жидкости среднее число вылетевших из жидкости молекул равно среднему числу молекул, вернувшихся обратно из пара в жидкость, так что концентрация насыщенного пара неизменна).

Процессы диссоциации-рекомбинации записываются в виде следующих уравнений, отражающих состояние динамического равновесия:

$$NaCl \rightleftarrows Na^+ + Cl^-;$$
$$H_2SO_4 \rightleftarrows 2H^+ + SO_4^{2-};$$
$$CuSO_4 \rightleftarrows Cu^{2+} + SO_4^{2-}.$$

Изменение внешних условий может нарушить текущее динамическое равновесие и сместить его в ту или иную сторону. Например, при повышении температуры увеличивается скорость диссоциации, и концентрация положительных и отрицательных ионов в растворе возрастает.

3.14.2 Ионная проводимость

В металлах, как вы помните, имеется лишь один тип свободных зарядов — это свободные электроны. В электролитах ситуация иная: здесь возникают свободные заряды двух типов.

1. Положительные ионы, образовавшиеся из атомов металлов или водорода.

2. Отрицательные ионы — атомные или молекулярные кислотные остатки (например, Cl^- или SO_4^{2-}), а также гидроксильная группа OH^-.

Второе отличие от металлов заключается в том, что носители свободных зарядов в электролите могут иметь заряд, равный по модулю как элементарному заряду e, так и целому числу элементарных зарядов ze. Здесь z — валентность атома или группы атомов; например, при растворении медного купороса имеем $z = 2$.

Если внешнего электрического поля нет, то свободные заряды электролита совершают лишь хаотическое тепловое движение наряду с окружающими молекулами. Но при наложении внешнего поля положительные и отрицательные ионы начинают упорядоченное движение.

Поместим в сосуд с электролитом два электрода; один из электродов присоединим к положительной клемме источника тока, а другой — к отрицательной (рис. 3.56). Когда речь идёт о прохождении тока через электролиты, положительный электрод называют *анодом*, а отрицательный — *катодом*[28].

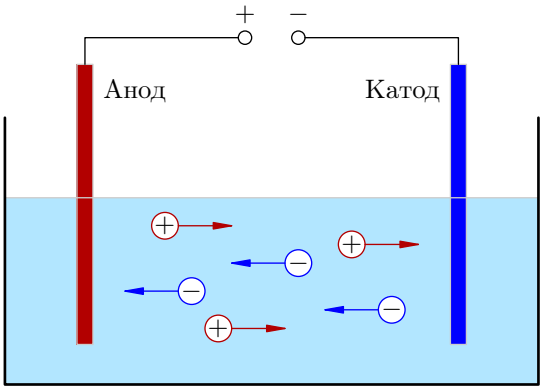


Рис. 3.56. Ионная проводимость электролита

В электрическом поле, возникшем между электродами, положительные ионы электролита устремляются к «минусу» катода, а отрицательные ионы — к «плюсу» анода. Таким образом, *электрический ток в электролите образуется в результате встречного движения ионов: положительных — к катоду, отрицательных — к аноду.* Поэтому проводимость электролитов называется *ионной* (в отличие от электронной проводимости металлов).

На положительном аноде имеется недостаток электронов. Отрицательные ионы, достигнув анода, отдают ему свои лишние электроны; эти электроны отправляются по цепи к «плюсу» источника.

Наоборот, на отрицательном катоде — избыток электронов. Положительные ионы, придя на катод, забирают у него электроны, и это количество ушедших электронов немедленно восполняется их доставкой на катод с «минуса» источника.

Таким образом, в той части цепи, которая состоит из источника тока и металлических проводников, возникает циркуляция электронов по маршруту «анод → источник → катод». Цепь замыкается электролитом, где электрический ток обеспечивается двусторонним движением ионов.

3.14.3 Электролиз

Положительные и отрицательные ионы, будучи носителями свободных зарядов, в то же время являются частицами вещества. Поэтому важнейшее отличие тока в электролитах от тока в

[28] Имеется народная мудрость для запоминания знаков анода и катода: *Андрей — парень положительный, Катька — девка отрицательная* ;-)

металлах состоит в том, что *электрический ток в электролите сопровождается переносом вещества*.

Явление переноса вещества при прохождении электрического тока через электролит называется *электролизом*. Законы электролиза были экспериментально изучены Фарадеем.

В процессе электролиза происходит разложение растворённого вещества на составные части и выделение этих частей на электродах. Так, в растворе медного купороса $CuSO_4$ положительные ионы меди Cu^{2+} идут на катод, в результате чего катод покрывается медью. Кислотный остаток SO_4^{2-} выделяется на аноде.

Естественным образом возникает вопрос о нахождении массы m вещества, выделяющегося на электроде за определённое время t. Эта масса, очевидно, совпадает с массой данного вещества, перенесённого током за время t через электролит.

Пусть m_1 — масса одного иона этого вещества, $q_1 = ze$ — заряд иона (z — валентность вещества). Предположим, что за время t через электролит прошёл заряд q. Число ионов, пришедших на электрод, тогда равно $N = q/q_1$. Масса выделившегося на электроде вещества равна суммарной массе пришедших ионов:

$$m = m_1 N = m_1 \frac{q}{q_1} = kq. \tag{3.69}$$

Величина $k = m_1/q_1$ является характеристикой вещества и называется его *электрохимическим эквивалентом*. Значения электрохимических эквивалентов различных веществ приводятся в таблицах.

При протекании через электролит постоянного тока I за время t проходит заряд $q = It$. Подставляя это в формулу (3.69), получим первую формулу Фарадея:

$$m = kIt. \tag{3.70}$$

Первый закон Фарадея. Масса выделяющегося на электроде вещества пропорциональна силе тока, протекающего через электролит, и времени прохождения тока.

Теперь преобразуем выражение для электрохимического эквивалента, введя молярную массу вещества:

$$k = \frac{m_1}{q_1} = \frac{\mu/N_\mathrm{A}}{ze} = \frac{\mu}{zeN_\mathrm{A}}\,.$$

Подставляя это выражение в (3.70), получим вторую формулу Фарадея:

$$m = \frac{\mu}{zeN_\mathrm{A}} It. \tag{3.71}$$

Второй закон Фарадея. Масса выделяющегося на электроде вещества прямо пропорциональна молярной массе этого вещества и обратно пропорциональна его валентности.

В формуле (3.71) мы видим произведение двух констант e и N_A. Оно также является константой и называется *постоянной Фарадея*:

$$F = eN_\mathrm{A} = 96485\ \frac{\text{Кл}}{\text{моль}}\,.$$

Формула (3.71) с постоянной Фарадея запишется так:

$$m = \frac{\mu}{zF} It.$$

3.15 Электрический ток в газах

При обычных условиях газы состоят из электрически нейтральных атомов или молекул; свободных зарядов в газах *почти нет*. Поэтому газы являются *диэлектриками* — электрический ток через них не проходит.

Мы сказали «почти нет», потому что на самом деле в газах и, в частности, в воздухе всегда присутствует некоторое количество свободных заряженных частиц. Они появляются в результате ионизирующего воздействия излучений радиоактивных веществ, входящих в состав земной коры, ультрафиолетового и рентгеновского излучения Солнца, а также космических лучей — потоков частиц высокой энергии, проникающих в атмосферу Земли из космического пространства. Впоследствии мы вернёмся к этому факту и обсудим его важность, а сейчас заметим лишь, что в обычных условиях проводимость газов, вызванная «естественным» количеством свободных зарядов, пренебрежимо мала, и её можно не принимать во внимание.

На изолирующих свойствах воздушного промежутка основано действие переключателей в электрических цепях (рис. 3.57).

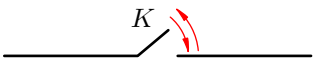


Рис. 3.57. Ключ K

Размыкание электрической цепи — это создание с помощью ключа K промежутка воздуха между её контактами. Пройти сквозь этот промежуток электрический ток не сможет. Например, небольшого воздушного зазора в выключателе оказывается достаточно, чтобы разомкнуть электрическую цепь в вашей комнате и погасить свет.

Можно, однако, создать такие условия, при которых электрический ток в газовом промежутке *появится*. Давайте рассмотрим следующий опыт.

Зарядим пластины воздушного конденсатора[29] и подсоединим их к чувствительному гальванометру (рис. 3.58, слева). При комнатной температуре и не слишком влажном воздухе гальванометр не покажет заметного тока: наш воздушный промежуток, как мы и говорили, не является проводником электричества.

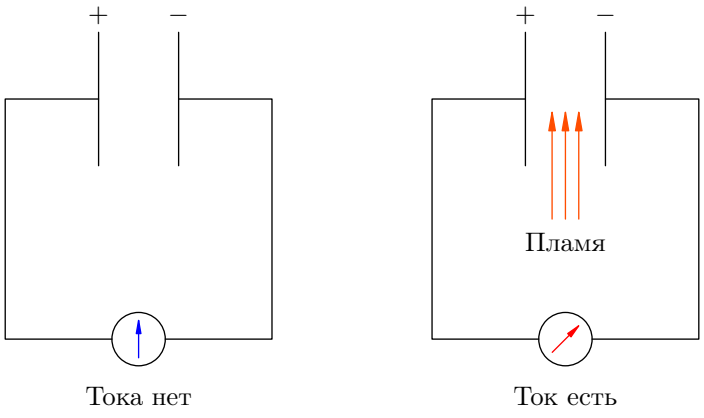


Рис. 3.58. Возникновение тока в воздухе

Теперь внесём в зазор между пластинами конденсатора пламя горелки или свечи (рис. 3.58, справа). Ток появляется! Почему?

[29] Разумеется, напряжение на конденсаторе предполагается не очень высоким, чтобы не наступил пробой воздуха. О том, что такое пробой газового промежутка, речь пойдёт ниже.

3.15.1 Свободные заряды в газе

Возникновение электрического тока между пластинами кондесатора означает, что в воздухе под воздействием пламени появились *свободные заряды*. Какие именно?

Опыт показывает, что электрический ток в газах является упорядоченным движением заряженных частиц *трёх видов*. Это *электроны*, *положительные ионы* и *отрицательные ионы*. Давайте посмотрим, каким образом эти заряды могут появляться в газе.

При увеличении температуры газа тепловые колебания его частиц — молекул или атомов — становятся всё интенсивнее. Удары частиц друг о друга достигают такой силы, что начинается *ионизация* — распад нейтральных частиц на электроны и положительные ионы (рис. 3.59).

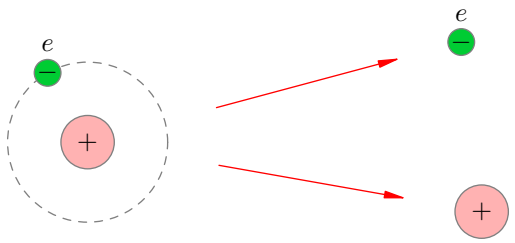


Рис. 3.59. Ионизация

Степенью ионизации называется отношение числа распавшихся частиц газа к общему исходному числу частиц. Например, если степень ионизации равна 40%, то это означает, что 40% исходных частиц газа распалось на положительные ионы и электроны.

Степень ионизации газа зависит от температуры и резко возрастает с её увеличением. У водорода, например, при температуре ниже 10000 °C степень ионизации не превосходит 10%, а при температуре выше 20000 °C степень ионизации близка к 100% (то есть водород почти полностью ионизирован[30]).

Помимо высокой температуры имеются и другие факторы, вызывающие ионизацию газа. Мы их уже вскользь упоминали: это радиоактивные излучения, ультрафиолетовые, рентгеновские и гамма-лучи, космические частицы. Всякий такой фактор, являющийся причиной ионизации газа, называется *ионизатором*.

Таким образом, ионизация происходит не сама по себе, а под воздействием ионизатора. Одновременно идёт и обратный процесс — *рекомбинация*, то есть воссоединение электрона и положительного иона в нейтральную частицу (рис. 3.60).

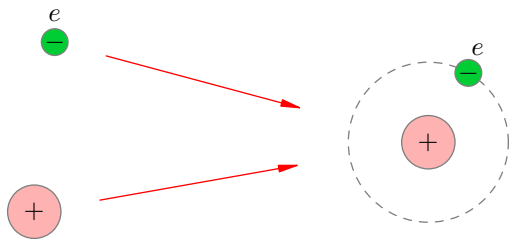


Рис. 3.60. Рекомбинация

Причина рекомбинации проста: это кулоновское притяжение электронов и положительных ионов. Устремляясь навстречу друг другу под действием электрических сил, положительный ион и электрон встречаются и получают возможность образовать нейтральный атом (или нейтральную молекулу).

[30] Частично или полностью ионизированный газ называется *плазмой*.

При неизменной интенсивности действия ионизатора устанавливается динамическое равновесие: среднее количество частиц, распадающихся в единицу времени, равно среднему количеству рекомбинирующих частиц (иными словами, скорость ионизации равна скорости рекомбинации). Если действие ионизатора усилить (например, повысить температуру), то динамическое равновесие сместится в сторону ионизации, и концентрация заряженных частиц в газе возрастёт. Наоборот, если выключить ионизатор, то рекомбинация начнёт преобладать, и свободные заряды постепенно исчезнут полностью.

Итак, положительные ионы и электроны появляются в газе в результате ионизации. Откуда же берётся третий сорт зарядов — отрицательные ионы? Очень просто: электрон может налететь на нейтральный атом и присоединиться к нему! Этот процесс показан на рис. 3.61.

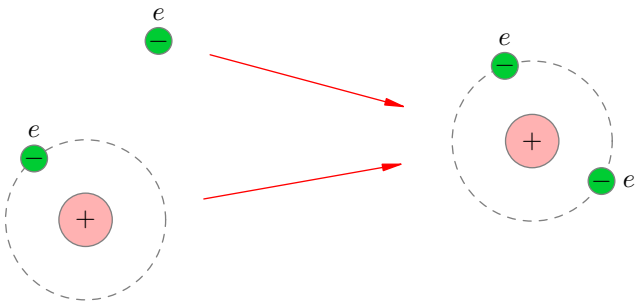


Рис. 3.61. Появление отрицательного иона

Образованные таким образом отрицательные ионы будут участвовать в создании тока наряду с положительными ионами и электронами.

3.15.2 Несамостоятельный разряд

Если внешнего электрического поля нет, то свободные заряды совершают хаотическое тепловое движение наряду с нейтральными частицами газа. Но при наложении электрического поля начинается упорядоченное движение заряженных частиц — *электрический ток в газе*.

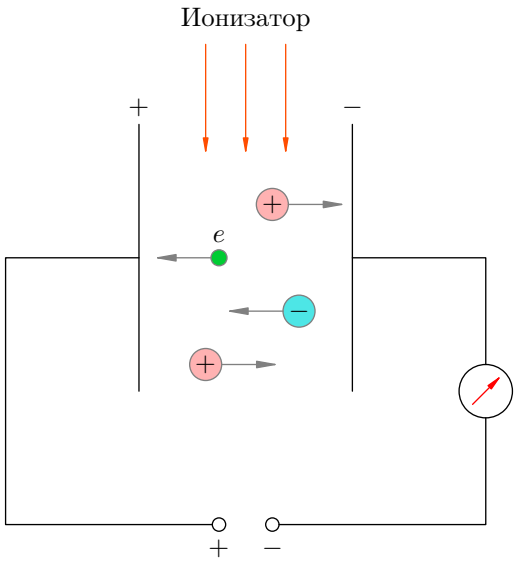


Рис. 3.62. Несамостоятельный разряд

На рис. 3.62 мы видим три сорта заряженных частиц, возникающих в газовом промежутке под действием ионизатора: положительные ионы, отрицательные ионы и электроны. *Электри-*

ческий ток в газе образуется в результате встречного движения заряженных частиц: положительных ионов — к отрицательному электроду (катоду), электронов и отрицательных ионов — к положительному электроду (аноду).

С движением заряженных частиц в газовом промежутке всё ясно. А почему возникает ток во внешней цепи?

Свободные электроны газа, попадая на анод, направляются по цепи к «плюсу» источника тока. Отрицательные ионы отдают аноду лишний электрон и, став нейтральными частицами, возвращаются в обратно газ; отданный же аноду электрон также устремляется к «плюсу» источника. Положительные ионы, приходя на катод, забирают оттуда электроны (и, превратившись в нейтральные частицы, диффундируют обратно в газ), а возникший дефицит электронов на катоде немедленно компенсируется их доставкой туда с «минуса» источника. В результате всех этих процессов возникает упорядоченное движение электронов во внешней цепи. Это и есть электрический ток, регистрируемый гальванометром.

Описанный процесс, изображённый на рис. 3.62, называется *несамостоятельным разрядом* в газе. Почему несамостоятельным? Потому для его поддержания необходимо постоянное действие ионизатора. Уберём ионизатор — и ток прекратится, поскольку исчезнет механизм, обеспечивающий появление свободных зарядов в газовом промежутке. Пространство между анодом и катодом снова станет изолятором.

3.15.3 Вольт-амперная характеристика газового разряда

Зависимость силы тока через газовый промежуток от напряжения между анодом и катодом (так называемая *вольт-амперная характеристика газового разряда*) показана на рис. 3.63.

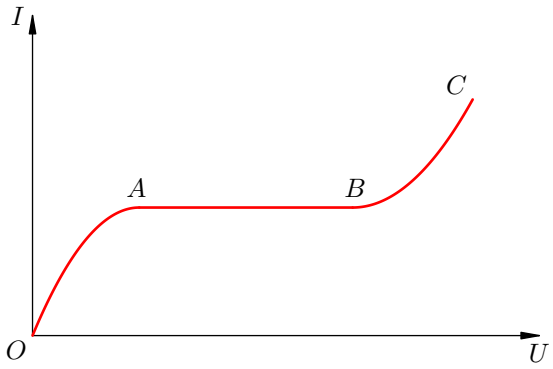


Рис. 3.63. Вольт-амперная характеристика газового разряда

При нулевом напряжении сила тока, естественно, равна нулю: заряженные частицы совершают лишь тепловое движение, упорядоченного их движения между электродами нет.

При небольшом напряжении сила тока также мала. Дело в том, что не всем заряженным частицам суждено добраться до электродов: часть положительных ионов и электронов в процессе своего движения находят друг друга и рекомбинируют.

С повышением напряжения свободные заряды развивают всё большую скорость, и тем меньше шансов у положительного иона и электрона встретиться и рекомбинировать. Поэтому всё большая часть заряженных частиц достигает электродов, и сила тока возрастает (участок OA).

При определённой величине напряжения (точка A) скорость движения зарядов становится настолько большой, что рекомбинация вообще не успевает происходить. С этого момента *все* заряженные частицы, образованные под действием ионизатора, достигают электродов, и *ток достигает насыщения* — а именно, сила тока перестаёт меняться с увеличением напряжения. Так будет происходить вплоть до некоторой точки B.

3.15.4 Самостоятельный разряд

После прохождения точки B сила тока при увеличении напряжения резко возрастает — начинается *самостоятельный разряд*. Сейчас мы разберёмся, что это такое.

Заряженные частицы газа движутся от столкновения к столкновению; в промежутках между столкновениями они разгоняются электрическим полем, увеличивая свою кинетическую энергию. И вот, когда напряжение становится достаточно большим (та самая точка B), электроны за время свободного пробега достигают таких энергий, что при соударении с нейтральными атомами ионизируют[31] их!

Начинается так называемая *ионизация электронным ударом*. Электроны, выбитые из ионизированных атомов, также разгоняются электрическим полем и налетают на новые атомы, ионизируя теперь уже их и порождая новые электроны. В результате возникающей электронной лавины число ионизированных атомов стремительно увеличивается, вследствие чего быстро возрастает и сила тока.

Количество свободных зарядов становится таким большим, что необходимость во внешнем ионизаторе отпадает. Его можно попросту убрать. Свободные заряженные частицы теперь порождаются в результате *внутренних* процессов, происходящих в газе — вот почему разряд называется самостоятельным.

Если газовый промежуток находится под высоким напряжением, то для самостоятельного разряда не нужен никакой ионизатор. Достаточно в газе оказаться лишь одному свободному электрону, и начнётся описанная выше электронная лавина. А хотя бы один свободный электрон всегда найдётся!

Вспомним ещё раз, что в газе даже при обычных условиях имеется некоторое «естественное» количество свободных зарядов, обусловленное ионизирующим радиоактивным излучением земной коры, высокочастотным излучением Солнца, космическими лучами. Мы видели, что при малых напряжениях проводимость газа, вызванная этими свободными зарядами, ничтожно мала; но теперь — при высоком напряжении — они-то и породят лавину новых частиц, дав начало самостоятельному разряду. Произойдёт, как говорят, *пробой* газового промежутка.

Напряжённость поля, необходимая для пробоя сухого воздуха, равна примерно $30\,\text{кВ/см}$. Иными словами, чтобы между электродами, разделёнными сантиметром воздуха, проскочила искра, на них нужно подать напряжение 30 киловольт. Вообразите же, какое напряжение необходимо для пробоя нескольких километров воздуха! А ведь именно такие пробои происходят во время грозы — это прекрасно известные вам молнии.

[31] С помощью законов сохранения импульса и энергии можно показать, что именно электроны (а не ионы), ускоряемые электрическим полем, обладают максимальной способностью ионизировать атомы.

3.16 Полупроводники

До сих пор, говоря о способности веществ проводить электрический ток, мы делили их на проводники и диэлектрики. Удельное сопротивление обычных проводников находится в интервале $10^{-8} - 10^{-6}$ Ом · м; удельное сопротивление диэлектриков превышает эти величины в среднем на 20 порядков: $10^{10} - 10^{16}$ Ом · м.

Но существуют также вещества, которые по своей электропроводности занимают промежуточное положение между проводниками и диэлектриками. Это *полупроводники*: их удельное сопротивление при комнатной температуре может принимать значения в очень широком диапазоне $10^{-3} - 10^{7}$ Ом · м. К полупроводникам относятся кремний, германий, селен, некоторые другие химические элементы и соединения[32]. Наиболее широко примененяются кремний (Si) и германий (Ge).

Главная особенность полупроводников заключается в том, что их электропроводность резко увеличивается с повышением температуры. Удельное сопротивление полупроводника убывает с ростом температуры примерно так, как показано на рис. 3.64.

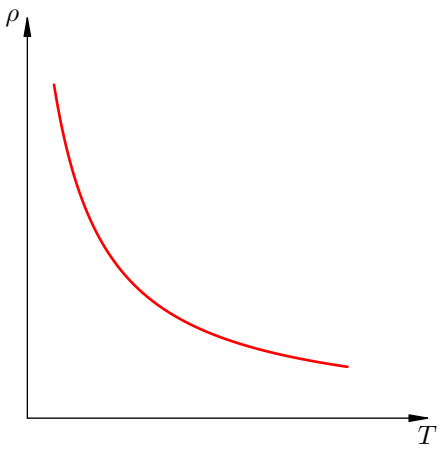


Рис. 3.64. Зависимость $\rho = \rho(T)$ для полупроводника

Иными словами, при низкой температуре полупроводники ведут себя как диэлектрики, а при высокой — как достаточно хорошие проводники. В этом состоит отличие полупроводников от металлов: удельное сопротивление металла, как вы помните, линейно возрастает с увеличением температуры.

Между полупроводниками и металлами имеются и другие отличия. Так, освещение полупроводника вызывает уменьшение его сопротивления (а на сопротивление металла свет почти не оказывает влияния). Кроме того, электропроводность полупроводников может очень сильно меняться при введении даже ничтожного количества примесей.

Опыт показывает, что, как и в случае металлов, при протекании тока через полупроводник не происходит переноса вещества. Стало быть, электрический ток в полупроводниках обусловлен движением электронов.

Уменьшение сопротивления полупроводника при его нагревании говорит о том, что повышение температуры приводит к увеличению количества свободных зарядов в полупроводнике. В металлах ничего такого не происходит; следовательно, полупроводники обладают иным механизмом электропроводности, чем металлы. И причина этого — различная природа химической связи между атомами металлов и полупроводников.

[32]Полупроводники чрезвычайно распространены в природе. Например, около 80% массы земной коры приходится на вещества, являющиеся полупроводниками.

3.16.1 Ковалентная связь

Металлическая связь, как вы помните, обеспечивается газом свободных электронов, который, подобно клею, удерживает положительные ионы в узлах кристаллической решётки. Полупроводники устроены иначе — их атомы скрепляет *ковалентная связь*. Давайте вспомним, что это такое.

Электроны, находящиеся на внешнем электронном уровне и называемые *валентными*, слабее связаны с атомом, чем остальные электроны, которые расположены ближе к ядру. В процессе образования ковалентной связи два атома вносят «в общее дело» по одному своему валентному электрону. Эти два электрона обобществляются, то есть теперь принадлежат уже обоим атомам, и потому называются *общей электронной парой* (рис. 3.65).

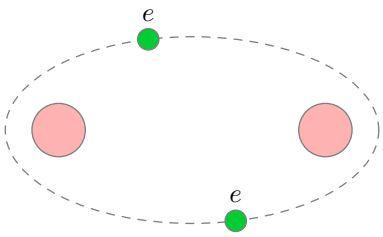


Рис. 3.65. Ковалентная связь

Обобществлённая пара электронов как раз и удерживает атомы друг около друга (с помощью сил электрического притяжения). *Ковалентная связь — это связь, существующая между атомами за счёт общих электронных пар.* По этой причине ковалентная связь называется также *парноэлектронной*.

3.16.2 Кристаллическая структура кремния

Теперь мы готовы подробнее изучить внутреннее устройство полупроводников. В качестве примера рассмотрим самый распространённый в природе полупроводник — кремний. Аналогичное строение имеет и второй по важности полупроводник — германий.

Пространственная структура кремния представлена[33] на рис. 3.66. Шариками изображены атомы кремния, а трубки, их соединяющие, — это каналы ковалентной связи между атомами.

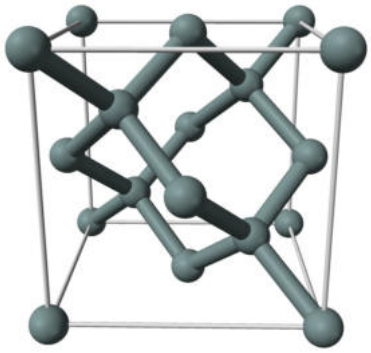


Рис. 3.66. Кристаллическая структура кремния

Обратите внимание, что каждый атом кремния скреплён с *четырьмя* соседними атомами. Почему так получается?

[33] Автор картинки — Ben Mills.

Дело в том, что кремний четырёхвалентен — на внешней электронной оболочке атома кремния расположены четыре валентных электрона. Каждый из этих четырёх электронов готов образовать общую электронную пару с валентным электроном другого атома. Так и происходит! В результате атом кремния окружается четырьмя пристыковавшимися к нему атомами, каждый из которых вносит по одному валентному электрону. Соответственно, вокруг каждого атома оказывается по восемь электронов (четыре своих и четыре чужих).

Более подробно мы видим это на плоской схеме кристаллической решётки кремния, представленной на рис. 3.67.

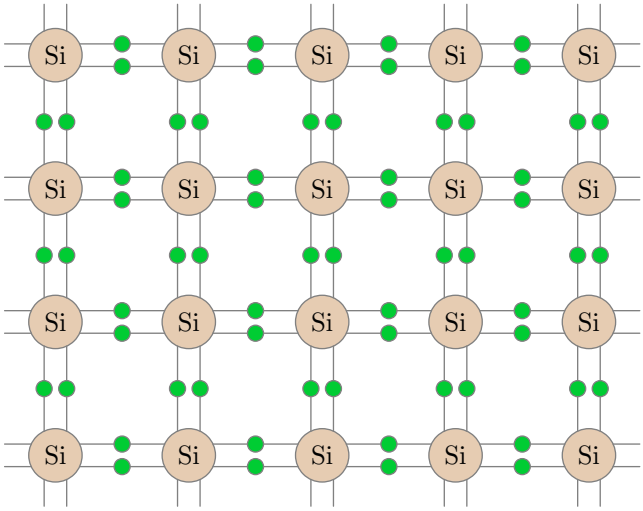


Рис. 3.67. Кристаллическая решётка кремния

Ковалентные связи изображены парами линий, соединяющих атомы; на этих линиях находятся общие электронные пары. Каждый валентный электрон, расположенный на такой линии, бóльшую часть времени проводит в пространстве между двумя соседними атомами.

Однако валентные электроны отнюдь не «привязаны намертво» к соответствующим парам атомов. Происходит перекрытие электронных оболочек *всех* соседних атомов, так что любой валентный электрон есть общее достояние всех атомов-соседей. От некоторого атома 1 такой электрон может перейти к соседнему с ним атому 2, затем — к соседнему с ним атому 3 и так далее. Валентные электроны могут перемещаться по всему пространству кристалла — они, как говорят, *принадлежат всему кристаллу* (а не какой-либо одной атомной паре).

Тем не менее, валентные электроны кремния не являются свободными (как это имеет место в металле). В полупроводнике связь валентных электронов с атомами гораздо прочнее, чем в металле; ковалентные связи кремния не разрываются при невысоких температурах. Энергии электронов оказывается недостаточно для того, чтобы под действием внешнего электрического поля начать упорядоченное движение от меньшего потенциала к большему. Поэтому при достаточно низких температурах полупроводники близки к диэлектрикам — они не проводят электрический ток.

3.16.3 Собственная проводимость

Если включить в электрическую цепь полупроводниковый элемент и начать его нагревать, то сила тока в цепи возрастает. Следовательно, сопротивление полупроводника *уменьшается* с ростом температуры. Почему это происходит?

При повышении температуры тепловые колебания атомов кремния становятся интенсивнее, и энергия валентных электронов возрастает. У некоторых электронов энергия достигает значений, достаточных для разрыва ковалентных связей. Такие электроны покидают свои атомы и

становятся *свободными* (или *электронами проводимости*) — точно так же, как в металле. Во внешнем электрическом поле свободные электроны начинают упорядоченное движение, образуя электрический ток.

Чем выше температура кремния, тем больше энергия электронов, и тем большее количество ковалентных связей не выдерживает и рвётся. Число свободных электронов в кристалле кремния возрастает, что и приводит к уменьшению его сопротивления.

Разрыв ковалентных связей и появление свободных электронов показан на рис. 3.68. На месте разорванной ковалентной связи образуется *дырка* — вакантное место для электрона. Дырка имеет *положительный* заряд, поскольку с уходом отрицательно заряженного электрона остаётся нескомпенсированный положительный заряд ядра атома кремния.

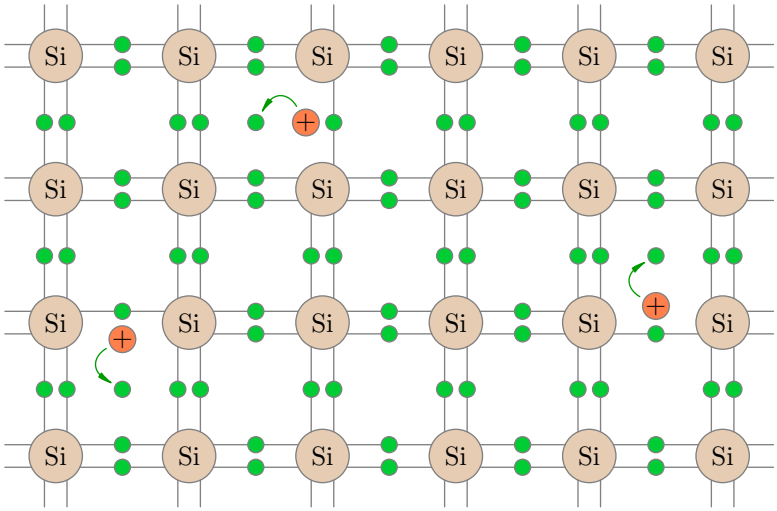


Рис. 3.68. Образование свободных электронов и дырок

Дырки не остаются на месте — они могут блуждать по кристаллу. Дело в том, что один из соседних валентных электронов, «путешествуя» между атомами, может перескочить на образовавшееся вакантное место, заполнив дырку; тогда дырка в этом месте исчезнет, но появится в том месте, откуда электрон пришёл.

При отсутствии внешнего электрического поля перемещение дырок носит случайный характер, ибо валентные электроны блуждают между атомами хаотически. Однако в электрическом поле начинается *направленное* движение дырок. Почему? Понять это несложно.

На рис. 3.69 изображён полупроводник, помещённый в электрическое поле $\vec{E}$. В левой части рисунка — начальное положение дырки.

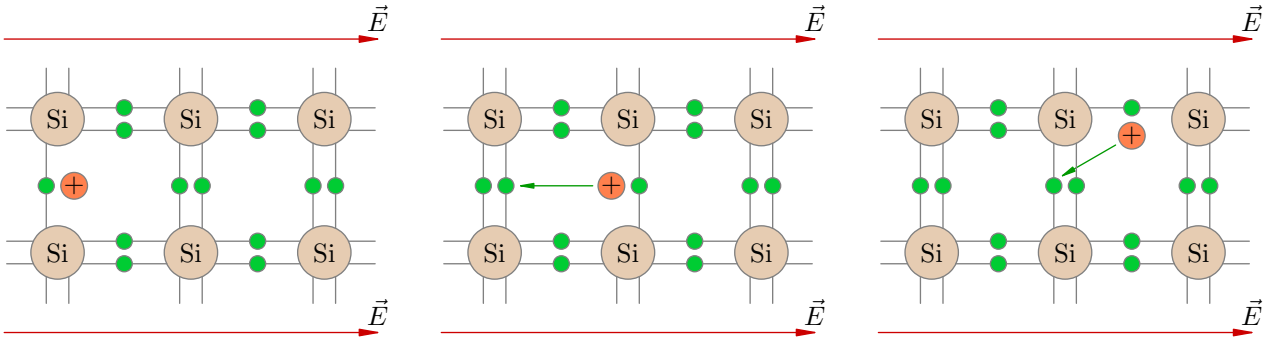


Рис. 3.69. Движение дырки в электрическом поле

Куда сместится дырка? Ясно, что наиболее вероятны перескоки «электрон → дырка» в направлении *против* линий поля (то есть к «плюсам», создающим поле). Один из таких пере-

скоков показан в средней части рисунка: электрон прыгнул влево, заполнив вакансию, а дырка, соответственно, сместилась вправо. Следующий возможный скачок электрона, вызванный электрическим полем, изображён в правой части рисунка; в результате этого скачка дырка заняла новое место, расположенное ещё правее.

Мы видим, что дырка в целом перемещается *по направлению* линий поля — то есть туда, куда и полагается двигаться положительным зарядам. Подчеркнём ещё раз, что направленное движение дырки вдоль поля вызвано перескоками валентных электронов от атома к атому, происходящими преимущественно в направлении против поля.

Таким образом, в кристалле кремния имеется два типа носителей заряда: свободные электроны и дырки. При наложении внешнего электрического поля появляется электрический ток, вызванный их упорядоченным встречным движением: свободные электроны перемещаются противоположно вектору напряжённости поля $\vec{E}$, а дырки — в направлении вектора $\vec{E}$.

Возникновение тока за счёт движения свободных электронов называется *электронной проводимостью*, или *проводимостью n-типа*. Процесс упорядоченного перемещения дырок называется *дырочной проводимостью*, или *проводимостью p-типа*[34]. Обе проводимости — электронная и дырочная — вместе называются *собственной проводимостью* полупроводника.

Каждый уход электрона с разорванной ковалентной связи порождает пару «свободный электрон–дырка». Поэтому концентрация свободных электронов в кристалле чистого кремния равна концентрации дырок. Соответственно, при нагревании кристалла увеличивается концентрация не только свободных электронов, но и дырок, что приводит к возрастанию собственной проводимости полупроводника за счёт увеличения как электронной, так и дырочной проводимости.

Наряду с образованием пар «свободный электрон–дырка» идёт и обратный процесс: *рекомбинация* свободных электронов и дырок. А именно, свободный электрон, встречаясь с дыркой, заполняет эту вакансию, восстанавливая разорванную ковалентную связь и превращаясь в валентный электрон. Таким образом, в полупроводнике устанавливается *динамическое равновесие*: среднее число разрывов ковалентных связей и образующихся электронно-дырочных пар в единицу времени равно среднему числу рекомбинирующих электронов и дырок. Это состояние динамического равновесия определяет равновесную концентрацию свободных электронов и дырок в полупроводнике при данных условиях.

Изменение внешних условий смещает состояние динамического равновесия в ту или иную сторону. Равновесное значение концентрации носителей заряда при этом, естественно, изменяется. Например, число свободных электронов и дырок возрастает при нагревании полупроводника или при его освещении.

При комнатной температуре концентрация свободных электронов и дырок в кремнии приблизительно равно $10^{10}\,\text{см}^{-3}$. Концентрация же атомов кремния — порядка $10^{22}\,\text{см}^{-3}$. Иными словами, на 10^{12} атомов кремния приходится лишь один свободный электрон! Это очень мало. В металлах, например, концентрация свободных электронов примерно равна концентрации атомов. Соответственно, *собственная проводимость кремния и других полупроводников при нормальных условиях мала по сравнению с проводимостью металлов.*

3.16.4 Примесная проводимость

Важнейшей особенностью полупроводников является то, что их удельное сопротивление может быть уменьшено на несколько порядков в результате введения даже весьма незначительного количества примесей. Помимо собственной проводимости у полупроводника возникает доминирующая *примесная проводимость*. Именно благодаря этому факту полупроводниковые приборы нашли столь широкое применение в науке и технике.

[34] От первых букв латинских слов *negativus* (отрицательный) и *positivus* (положительный)

Предположим, например, что в расплав кремния добавлено немного пятивалентного мышьяка (As). После кристаллизации расплава оказывается, что атомы мышьяка занимают места в некоторых узлах сформировавшейся кристаллической решётки кремния.

На внешнем электронном уровне атома мышьяка имеется пять электронов. Четыре из них образуют ковалентные связи с ближайшими соседями — атомами кремния (рис. 3.70). Какова судьба пятого электрона, не занятого в этих связях?

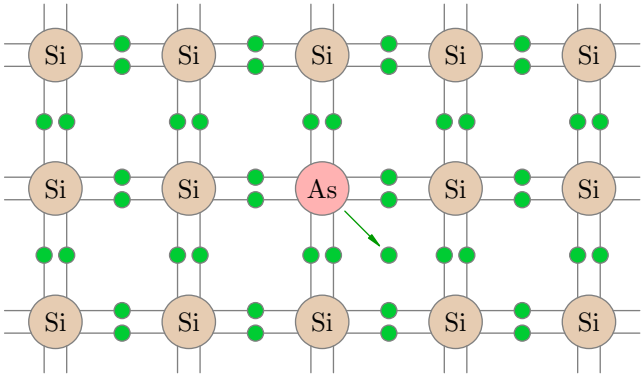


Рис. 3.70. Полупроводник n-типа

А пятый электрон становится свободным! Дело в том, что энергия связи этого «лишнего» электрона с атомом мышьяка, расположенным в кристалле кремния, гораздо меньше энергии связи валентных электронов с атомами кремния. Поэтому уже при комнатной температуре почти все атомы мышьяка в результате теплового движения остаются без пятого электрона, превращаясь в положительные ионы. А кристалл кремния, соответственно, наполняется свободными электронами, которые отщепились от атомов мышьяка.

Наполнение кристалла свободными электронами для нас не новость: мы видели это и выше, когда нагревался *чистый* кремний (без каких-либо примесей). Но сейчас ситуация принципиально иная: *появление свободного электрона, ушедшего из атома мышьяка, не сопровождается появлением подвижной дырки*. Почему? Причина та же — связь валентных электронов с атомами кремния гораздо прочнее, чем с атомом мышьяка на пятой вакансии, поэтому электроны соседних атомов кремния и не стремятся эту вакансию заполнить. Вакансия, таким образом, остаётся на месте, она как бы «приморожена» к атому мышьяка и не участвует в создании тока.

Таким образом, *внедрение атомов пятивалентного мышьяка в кристаллическую решётку кремния создаёт электронную проводимость, но не приводит к симметричному появлению дырочной проводимости*. Главная роль в создании тока теперь принадлежит свободным электронам, которые в данном случае называются *основными носителями* заряда.

Механизм собственной проводимости, разумеется, продолжает работать и при наличии примеси: ковалентные связи по-прежнему рвутся за счёт теплового движения, порождая свободные электроны и дырки. Но теперь дырок оказывается гораздо меньше, чем свободных электронов, которые в большом количестве предоставлены атомами мышьяка. Поэтому дырки в данном случае будут *неосновными носителями* заряда.

Примеси, атомы которых отдают свободные электроны без появления равного количества подвижных дырок, называются *донорными*. Например, пятивалентный мышьяк — донорная примесь. При наличии в полупроводнике донорной примеси основными носителями заряда являются свободные электроны, а неосновными — дырки; иными словами, концентрация свободных электронов намного превышает концентрацию дырок. Поэтому полупроводники с донорными примесями называются *электронными полупроводниками*, или *полупроводниками n-типа* (или просто *n-полупроводниками*).

А насколько, интересно, концентрация свободных электронов может превышать концентрацию дырок в n-полупроводнике? Давайте проведём простой расчёт.

Предположим, что примесь составляет 0,1 %, то есть на тысячу атомов кремния приходится один атом мышьяка. Концентрация атомов кремния, как мы помним, порядка $10^{22}\,\text{см}^{-3}$. Концентрация атомов мышьяка, соответственно, будет в тысячу раз меньше: $10^{19}\,\text{см}^{-3}$. Такой же окажется и концентрация свободных электронов, отданных примесью — ведь каждый атом мышьяка отдаёт по электрону. А теперь вспомним, что концентрация электронно-дырочных пар, появляющихся при разрывах ковалентных связей кремния, при комнатной температуре примерно равна $10^{10}\,\text{см}^{-3}$. Чувствуете разницу? Концентрация свободных электронов в данном случае больше концентрации дырок на 9 порядков, то есть в миллиард раз! Соответственно, в миллиард раз уменьшается удельное сопротивление кремниевого полупроводника при введении столь небольшого количества примеси.

Приведённый расчёт показывает, что в полупроводниках n-типа основную роль действительно играет электронная проводимость. На фоне столь колоссального превосходства численности свободных электронов вклад движения дырок в общую проводимость пренебрежимо мал.

Можно, наоборот, создать полупроводник с преобладанием дырочной проводимости. Так получится, если в кристалл кремния внедрить трёхвалентную примесь — например, индий (In). Результат такого внедрения показан на рис. 3.71.

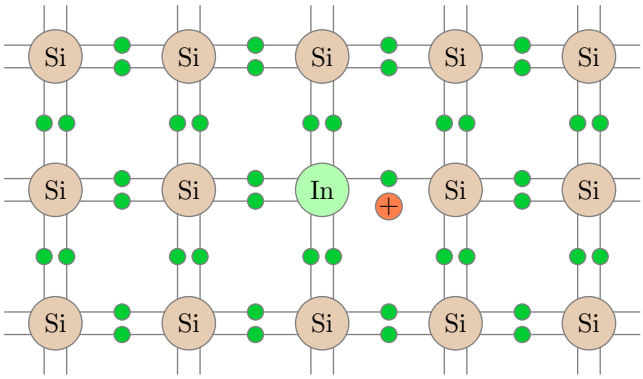


Рис. 3.71. Полупроводник p-типа

Что происходит в этом случае? На внешнем электронном уровне атома индия расположены три электрона, которые формируют ковалентные связи с тремя окружающими атомами кремния. Для четвёртого соседнего атома кремния у атома индия уже не хватает электрона, и в этом месте возникает дырка.

И дырка эта не простая, а особенная — с весьма большой энергией связи. Когда в неё попадёт электрон из соседнего атома кремния, он в ней «застрянет навеки», ибо притяжение электрона к атому индия весьма велико — больше, чем к атомам кремния. Атом индия превратится в отрицательный ион, а в том месте, откуда электрон пришёл, возникнет дырка — но теперь уже обыкновенная подвижная дырка в виде разорванной ковалентной связи в кристаллической решётке кремния. Эта дырка обычным образом начнёт блуждать по кристаллу за счёт «эстафетной» передачи валентных электронов от одного атома кремния к другому.

Итак, каждый примесный атом индия порождает дырку, но не приводит к симметричному появлению свободного электрона. Такие примеси, атомы которых захватывают «намертво» электроны и тем самым создают в кристалле подвижную дырку, называются *акцепторными*. Трёхвалентный индий — пример акцепторной примеси.

Если в кристалл чистого кремния ввести акцепторную примесь, то число дырок, порождённых примесью, будет намного больше числа свободных электронов, возникших за счёт разрыва ковалентных связей между атомами кремния. Полупроводник с акцепторной примесью — это *дырочный полупроводник*, или *полупроводник p-типа* (или просто *p-полупроводник*).

Дырки играют главную роль при создании тока в p-полупроводнике; дырки — *основные носители* заряда. Свободные электроны — *неосновные носители* заряда в p-полупроводнике.

Движение свободных электронов в данном случае не вносит существенного вклада: электрический ток обеспечивается в первую очередь дырочной проводимостью.

3.16.5 *p–n*-переход

Место контакта двух полупроводников с различными типами проводимости (электронной и дырочной) называется *электронно-дырочным переходом*, или *p–n-переходом*. В области *p–n*-перехода возникает интересное и очень важное явление — односторонняя проводимость.

На рис. 3.72 изображён контакт областей *p*- и *n*-типа; цветные кружочки — это дырки и свободные электроны, которые являются основными (или неосновными) носителями заряда в соответствующих областях.

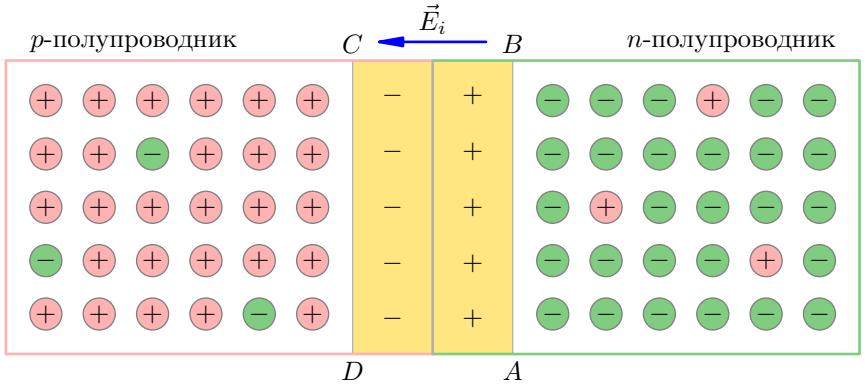


Рис. 3.72. Запирающий слой *p–n*-перехода

Совершая тепловое движение, носители заряда проникают через границу раздела областей. Свободные электроны переходят из *n*-области в *p*-область и рекомбинируют там с дырками; дырки же диффундируют из *p*-области в *n*-область и рекомбинируют там с электронами.

В результате этих процессов в электронном полупроводнике около границы контакта остаётся нескомпенсированный заряд положительных ионов донорной примеси, а в дырочном полупроводнике (также вблизи границы) возникает нескомпенсированный отрицательный заряд ионов акцепторной примеси. Эти нескомпенсированные объёмные заряды образуют так называемый *запирающий слой* $ABCD$, внутреннее электрическое поле $\vec{E}_i$ которого препятствует дальнейшей диффузии свободных электронов и дырок через границу контакта.

Подключим теперь к нашему полупроводниковому элементу источник тока, подав «плюс» источника на *n*-полупроводник, а «минус» — на *p*-полупроводник (рис. 3.73).

Мы видим, что внешнее электрическое поле $\vec{E}$ уводит основные носители заряда дальше от границы контакта. Ширина запирающего слоя увеличивается, его электрическое поле $\vec{E}_i$ возрастает. Сопротивление запирающего слоя велико, и основные носители не в состоянии преодолеть *p–n*-переход. Электрическое поле позволяет переходить границу лишь неосновным носителям, однако ввиду очень малой концентрации неосновных носителей создаваемый ими ток пренебрежимо мал.

Рассмотренная схема называется *включением p–n-перехода в обратном направлении*. Электрического тока основных носителей нет; имеется лишь ничтожно малый ток неосновных носителей. В данном случае *p–n*-переход оказывается закрытым.

Теперь поменяем полярность подключения и подадим «плюс» на *p*-полупроводник, а «минус» — на *n*-полупроводник (рис. 3.74). Эта схема называется *включением в прямом направлении*.

В этом случае внешнее электрическое поле $\vec{E}$ направлено против запирающего поля $\vec{E}_i$ и открывает путь основным носителям через *p–n*-переход. Запирающий слой становится тоньше,

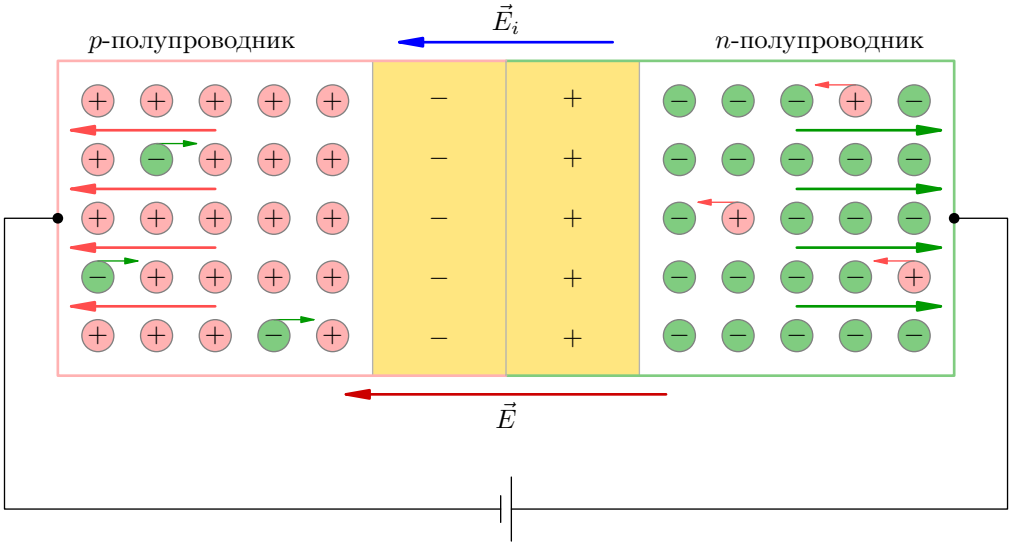


Рис. 3.73. Включение в обратном направлении: тока нет

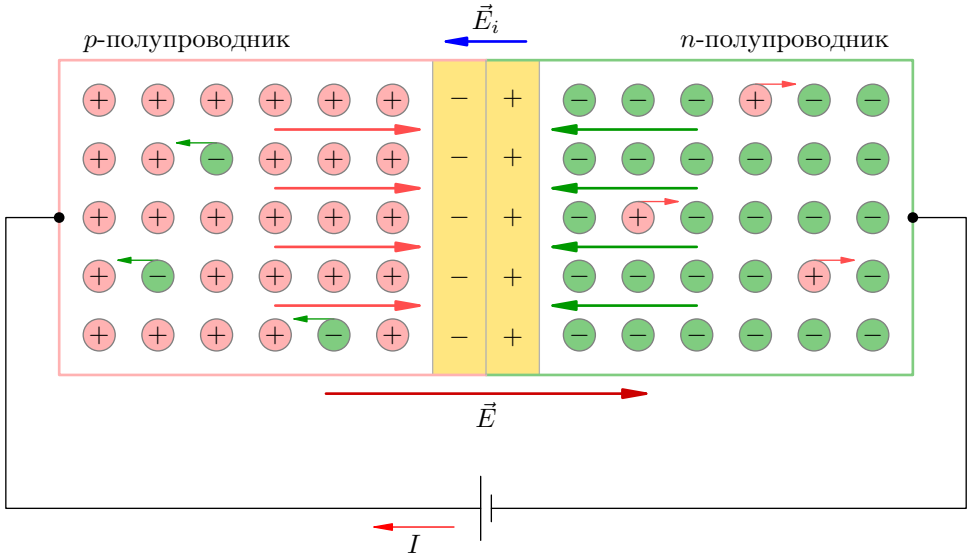


Рис. 3.74. Включение в прямом направлении: ток идёт

его сопротивление уменьшается. Происходит массовое перемещение свободных электронов из n-области в p-область, а дырки, в свою очередь, дружно устремляются из p-области в n-область. В цепи возникает ток I, вызванный движением основных носителей заряда[35].

Односторонняя проводимость p–n-перехода используется в *полупроводниковых диодах*. Диодом называется устройство, проводящие ток в лишь одном направлении; в противоположном направлении ток через диод не проходит (диод, как говорят, закрыт). Схематическое изображение диода показано на рис. 3.75.

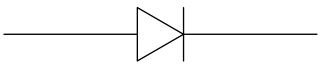


Рис. 3.75. Диод

В данном случае диод открыт в направлении слева направо: заряды как бы текут вдоль

[35]Теперь, правда, электрическое поле препятствует току неосновных носителей, но этот ничтожный фактор не оказывает заметного влияния на общую проводимость.

стрелки (видите её на рисунке?). В направлении справа налево заряды словно упираются в стенку — диод закрыт.

3.17 Магнитное поле. Линии

Магнитные свойства вещества известны людям давно. Магниты получили своё название от античного города Магнесия: в его окрестностях был распространён минерал (названный впоследствии магнитным железняком или магнетитом), куски которого притягивали железные предметы.

3.17.1 Взаимодействие магнитов

На двух сторонах каждого магнита расположены *северный полюс* и *южный полюс*. Два магнита притягиваются друг к другу разноимёнными полюсами и отталкиваются одноимёнными. Магниты могут действовать друг на друга даже сквозь вакуум! Всё это напоминает взаимодействие электрических зарядов, однако *взаимодействие магнитов не является электрическим*. Об этом свидетельствуют следующие опытные факты.

- Магнитная сила ослабевает при нагревании магнита. Сила же взаимодействия точечных зарядов не зависит от их температуры.
- Магнитная сила ослабевает, если трясти магнит. Ничего подобного с электрически заряженными телами не происходит.
- Положительные электрические заряды можно отделить от отрицательных (например, при электризации тел). А вот разделить полюса магнита не получается: если разрезать магнит на две части, то в месте разреза также возникают полюса, и магнит распадается на два магнита с разноимёнными полюсами на концах (ориентированных точно так же, как и полюса исходного магнита).

  Таким образом, магниты *всегда* двухполюсные, они существуют только в виде *диполей*. Изолированных магнитных полюсов (так называемых *магнитных монополей* — аналогов электрического заряда) в природе не существует (во всяком случае, экспериментально они пока не обнаружены). Это, пожалуй, самая впечатляющая асимметрия между электричеством и магнетизмом.
- Как и электрически заряженные тела, магниты действуют на электрические заряды. Однако магнит действует только на *движущийся* заряд; если заряд покоится относительно магнита, то действия магнитной силы на заряд не наблюдается. Напротив, наэлектризованное тело действует на любой заряд, вне зависимости от того, покоится он или движется.

По современным представлениям теории близкодействия, взаимодействие магнитов осуществляется посредством *магнитного поля*. А именно, магнит создаёт в окружающем пространстве магнитное поле, которое действует на другой магнит и вызывает видимое притяжение или отталкивание этих магнитов.

Примером магнита служит *магнитная стрелка* компаса. С помощью магнитной стрелки можно судить о наличии магнитного поля в данной области пространства, а также о направлении поля.

Наша планета Земля является гигантским магнитом. Неподалёку от северного географического полюса Земли расположен южный магнитный полюс. Поэтому северный конец стрелки компаса, поворачиваясь к южному магнитному полюсу Земли, указывает на географический север. Отсюда, собственно, и возникло название «северный полюс» магнита.

3.17.2 Линии магнитного поля

Электрическое поле, напомним, исследуется с помощью маленьких пробных зарядов, по действию на которые можно судить о величине и направлении поля. Аналогом пробного заряда в случае магнитного поля является маленькая магнитная стрелка.

Например, можно получить некоторое геометрическое представление о магнитном поле, если разместить в разных точках пространства очень маленькие стрелки компаса. Опыт показывает, что стрелки выстроятся вдоль определённых линий — так называемых *линий магнитного поля*. Дадим определение этого понятия в виде следующих трёх пунктов.

1. *Линии магнитного поля, или магнитные силовые линии — это направленные линии в пространстве, обладающие следующим свойством: маленькая стрелка компаса, помещённая в каждой точке такой линии, ориентируется по касательной к этой линии.*

2. *Направлением линии магнитного поля считается направление северных концов стрелок компаса, расположенных в точках данной линии.*

3. *Чем гуще идут линии, тем сильнее магнитное поле в данной области пространства.*

Роль стрелок компаса с успехом могут выполнять железные опилки: в магнитном поле маленькие опилки намагничиваются и ведут себя в точности как магнитные стрелки.

Так, насыпав железных опилок вокруг постоянного магнита, мы увидим примерно следующую картину линий магнитного поля (рис. 3.76).

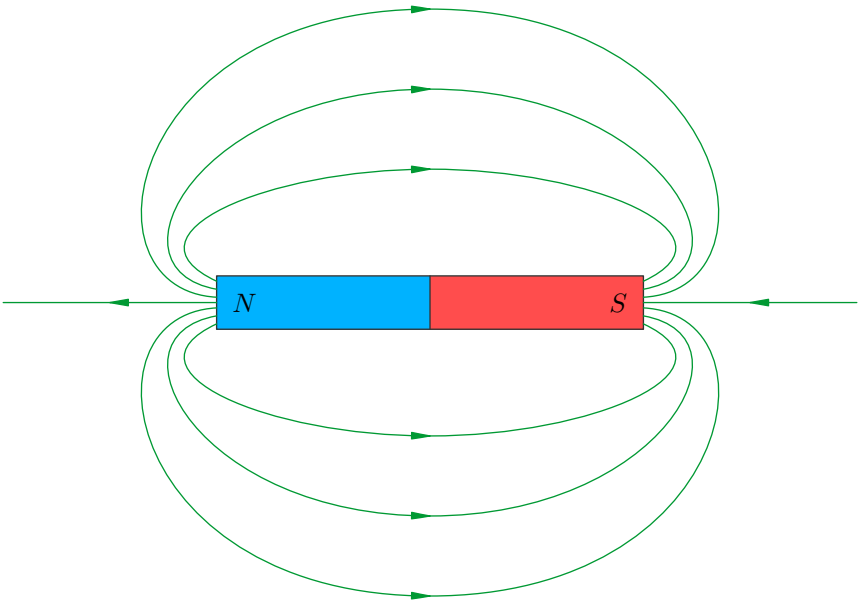


Рис. 3.76. Поле постоянного магнита

Северный полюс магнита обозначается синим цветом и буквой N; южный полюс — красным цветом и буквой S. Обратите внимание, что линии поля выходят из северного полюса магнита и входят в южный полюс: ведь именно к южному полюсу магнита будет направлен северный конец стрелки компаса.

3.17.3 Опыт Эрстеда

Несмотря на то, что электрические и магнитные явления были известны людям ещё с античности, никакой взаимосвязи между ними долгое время не наблюдалось. В течение нескольких

столетий исследования электричества и магнетизма шли параллельно и независимо друг от друга.

Тот замечательный факт, что электрические и магнитные явления на самом деле связаны друг с другом, был впервые обнаружен в 1820 году — в знаменитом опыте Эрстеда.

Схема опыта Эрстеда показана[36] на рис. 3.77. Над магнитной стрелкой NS (N и S — северный и южный полюсы стрелки) расположен металлический проводник, подключённый к источнику тока. Если замкнуть цепь, то стрелка поворачивается перпендикулярно проводнику!

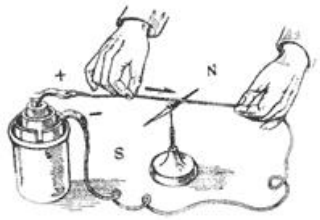


Рис. 3.77. Опыт Эрстеда

Этот простой опыт прямо указал на взаимосвязь электричества и магнетизма. Эксперименты последовавшие за опытом Эрстеда, твёрдо установили следующую закономерность: *магнитное поле порождается электрическими токами и действует на токи.*

Картина линий магнитного поля, порождённого проводником с током, зависит от формы проводника.

3.17.4 Магнитное поле прямого провода с током

Линии магнитного поля прямолинейного провода с током являются концентрическими окружностями. Центры этих окружностей лежат на проводе, а их плоскости перпендикулярны проводу (рис. 3.78).

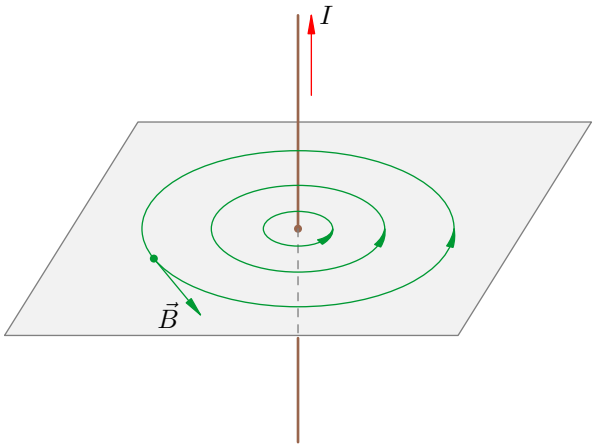


Рис. 3.78. Поле прямого провода с током

Для определения направления линий магнитного поля прямого тока существуют два альтернативных правила.

Правило часовой стрелки. Линии поля идут против часовой стрелки, если смотреть так, чтобы ток тёк на нас.

Правило винта.[37] Линии поля идут туда, куда надо вращать винт (с обычной правой резьбой), чтобы он двигался по резьбе в направлении тока.

Пользуйтесь тем правилом, которое вам больше по душе. Лучше привыкнуть к правилу часовой стрелки — вы сами впоследствии убедитесь, что оно более универсально и им проще пользоваться[38].

[36] Изображение с сайта rt.mipt.ru.

[37] Или *правило буравчика*, или *правило штопора* — это уж кому что ближе ;-)

[38] А потом с благодарностью вспомните его на первом курсе, когда будете изучать аналитическую геометрию.

На рис. 3.78 появилось и кое-что новое: это вектор $\vec{B}$, который называется *индукцией магнитного поля*, или *магнитной индукцией*. Вектор магнитной индукции является аналогом вектора напряжённости электрического поля: он служит *силовой характеристикой* магнитного поля, определяя силу, с которой магнитное поле действует на движущиеся заряды.

О силах в магнитном поле мы поговорим позже, а пока отметим лишь, что величина и направление магнитного поля определяется вектором магнитной индукции $\vec{B}$. В каждой точке пространства вектор $\vec{B}$ направлен туда же, куда и северный конец стрелки компаса, помещённой в данную точку, а именно по касательной к линии поля в направлении этой линии. Измеряется магнитная индукция в *теслах* (Тл).

Как и в случае электрического поля, для индукции магнитного поля справедлив *принцип суперпозиции*. Он заключается в том, что *индукции магнитных полей* $\vec{B}_1, \vec{B}_2, \ldots, \vec{B}_n$, *создаваемых в данной точке различными токами, складываются векторно и дают результирующий вектор магнитной индукции:* $\vec{B} = \vec{B}_1 + \vec{B}_2 + \ldots + \vec{B}_n$.

3.17.5 Магнитное поле витка с током

Рассмотрим круговой виток, по которому циркулирует постоянный ток I. Источник, создающий ток, мы на рисунке не показываем.

Картина линий поля нашего витка будет иметь приблизительно следующий вид (рис. 3.79).

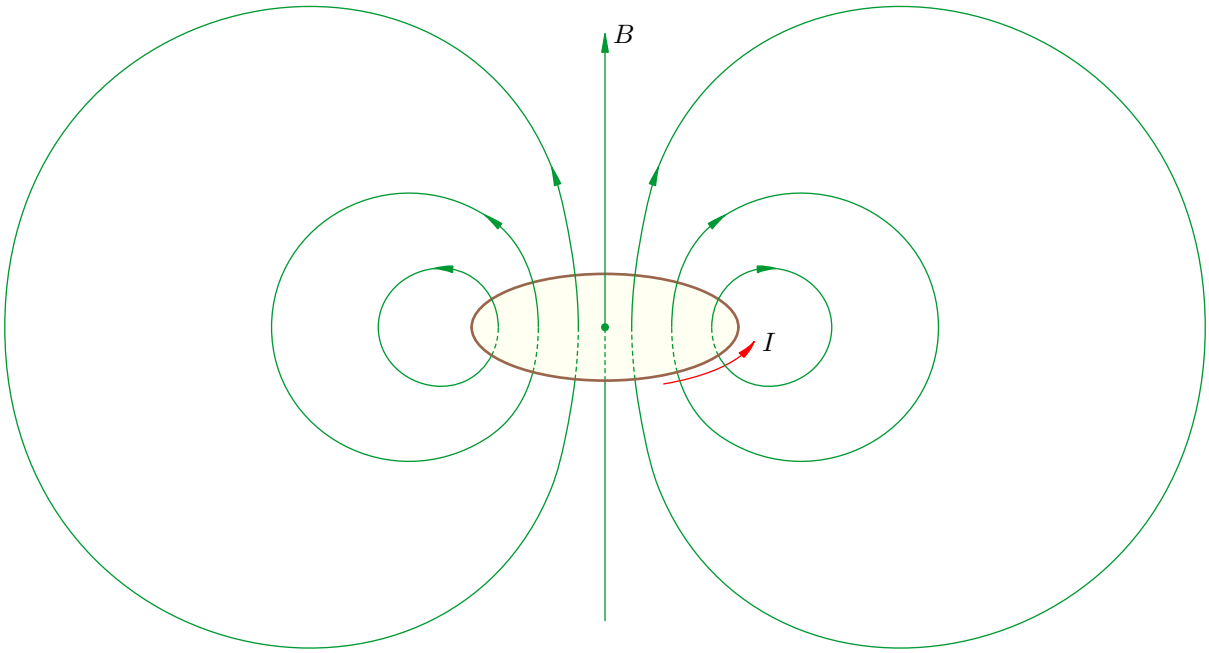


Рис. 3.79. Поле витка с током

Нам будет важно уметь определять, в какое полупространство (относительно плоскости витка) направлено магнитное поле. Снова имеем два альтернативных правила.

Правило часовой стрелки. Линии поля идут туда, глядя откуда ток кажется циркулирующим против часовой стрелки.

Правило винта. Линии поля идут туда, куда будет перемещаться винт (с обычной правой резьбой), если вращать его в направлении тока.

Как видите, ток и поле меняются ролями — по сравнению с формулировками этих правил для случая прямого тока.

3.17.6 Магнитное поле катушки с током

Катушка получится, если плотно, виток к витку, намотать провод в достаточно длинную спираль (рис. 3.80)[39]. В катушке может быть несколько десятков, сотен или даже тысяч витков. Катушка называется ещё *соленоидом*.

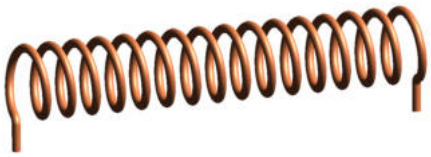


Рис. 3.80. Катушка (соленоид)

Магнитное поле одного витка, как мы знаем, выглядит не очень-то просто. Поля́ отдельных витков катушки накладываются друг на друга, и, казалось бы, в результате должна получиться совсем уж запутанная картина. Однако это не так: поле длинной катушки имеет неожиданно простую структуру (рис. 3.81).

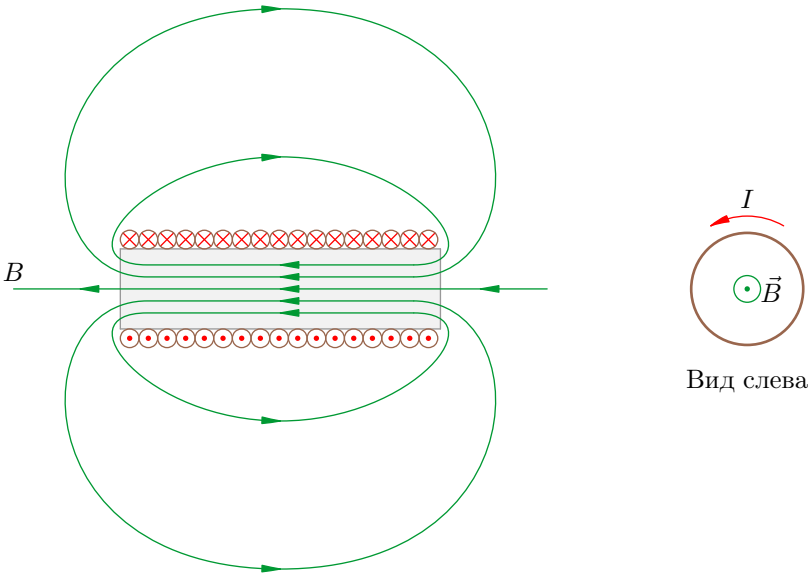


Рис. 3.81. Поле катушки с током

На этом рисунке ток в катушке идёт против часовой стрелки, если смотреть слева (так будет, если на рис. 3.80 правый конец катушки подключить к «плюсу» источника тока, а левый конец — к «минусу»). Мы видим, что магнитное поле катушки обладает двумя характерными свойствами.

1. Внутри катушки вдали от её краёв магнитное поле является *однородным*: в каждой точке вектор магнитной индукции одинаков по величине и направлению. Линии поля — параллельные прямые; они искривляются лишь вблизи краёв катушки, когда выходят наружу.

2. Вне катушки поле близко к нулю. Чем больше витков в катушке — тем слабее поле снаружи неё.

Заметим, что бесконечно длинная катушка вообще не выпускает поле наружу: вне катушки магнитное поле отсутствует. Внутри такой катушки поле всюду является однородным.

[39]Изображение с сайта en.wikipedia.org.

Ничего не напоминает? Катушка является «магнитным» аналогом конденсатора. Вы же помните, что конденсатор создаёт внутри себя однородное электрическое поле, линии которого искривляются лишь вблизи краёв пластин, а вне конденсатора поле близко к нулю; конденсатор с бесконечными обкладками вообще не выпускает поле наружу, а всюду внутри него поле однородно.

А теперь — главное наблюдение. Сопоставьте, пожалуйста, картину линий магнитного поля вне катушки (рис. 3.81) с линиями поля магнита на рис. 3.76. Одно и то же, не правда ли? И вот мы подходим к вопросу, который, вероятно, у вас уже давно возник: если магнитное поле порождается токами и действует на токи, то какова причина возникновения магнитного поля вблизи постоянного магнита? Ведь этот магнит вроде бы не является проводником с током!

3.17.7 Гипотеза Ампера. Элементарные токи

Поначалу думали, что взаимодействие магнитов объясняется особыми магнитными зарядами, сосредоточенными на полюсах. Но, в отличие от электричества, никто не мог изолировать магнитный заряд; ведь, как мы уже говорили, не удавалось получить по отдельности северный и южный полюс магнита — полюса всегда присутствуют в магните парами.

Сомнения насчёт магнитных зарядов усугубил опыт Эрстеда, когда выяснилось, что магнитное поле порождается электрическим током. Более того, оказалось, что для всякого магнита можно подобрать проводник с током соответствующей конфигурации, такой, что поле этого проводника совпадает с полем магнита.

Ампер выдвинул смелую гипотезу. *Нет никаких магнитных зарядов. Действие магнита объясняется замкнутыми электрическими токами внутри него.*

Что это за токи? Эти *элементарные токи* циркулируют внутри атомов и молекул; они связаны с движением электронов по атомным орбитам. Магнитное поле любого тела складывается из магнитных полей этих элементарных токов.

Элементарные токи могут быть беспорядочным образом расположены друг относительно друга. Тогда их поля взаимно погашаются, и тело не проявляет магнитных свойств.

Но если элементарные токи расположены согласованно, то их поля, складываясь, усиливают друг друга. Тело становится магнитом (рис. 3.82; магнитое поле будет направлено на нас; также на нас будет направлен и северный полюс магнита).

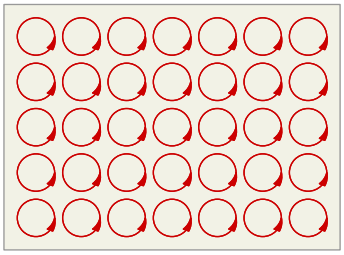


Рис. 3.82. Элементарные токи магнита

Гипотеза Ампера об элементарных токах прояснила свойства магнитов. Нагревание и тряска магнита разрушают порядок расположения его элементарных токов, и магнитные свойства ослабевают. Неразделимость полюсов магнита стала очевидной: в месте разреза магнита мы получаем те же элементарные токи на торцах. Способность тела намагничиваться в магнитном поле объясняется согласованным выстраиванием элементарных токов, «поворачивающихся» должным образом (о повороте кругового тока в магнитном поле читайте в следующем листке).

Гипотеза Ампера оказалась справедливой — это показало дальнейшее развитие физики. Представления об элементарных токах стали неотъемлемой частью теории атома, разработанной уже в XX веке — почти через сто лет после гениальной догадки Ампера.

3.18 Магнитное поле. Силы

В отличие от электрического поля, которое действует на любой заряд, магнитное поле действует только на *движущиеся* заряженные частицы. При этом оказывается, что сила зависит не только от величины, но и от направления скорости заряда.

3.18.1 Сила Лоренца

Сила, с которой магнитное поле действует на заряженную частицу, называется *силой Лоренца*. Опыт показывает, что вектор $\vec{F}$ силы Лоренца находится следующим образом.

1. Абсолютная величина силы Лоренца равна:

$$F = qvB\sin\alpha. \tag{3.72}$$

   Здесь q — абсолютная величина заряда, v — скорость заряда, B — индукция магнитного поля, α — угол между векторами $\vec{v}$ и $\vec{B}$.

2. Сила Лоренца перпендикулярна обоим векторам $\vec{v}$ и $\vec{B}$. Иными словами, вектор $\vec{F}$ перпендикулярен плоскости, в которой лежат векторы скорости заряда и индукции магнитного поля.

   Остаётся выяснить, в какое полупространство относительно данной плоскости направлена сила Лоренца.

3. Взаимное расположение векторов $\vec{v}$, $\vec{B}$ и $\vec{F}$ для *положительного* заряда q показано на рис. 3.83.

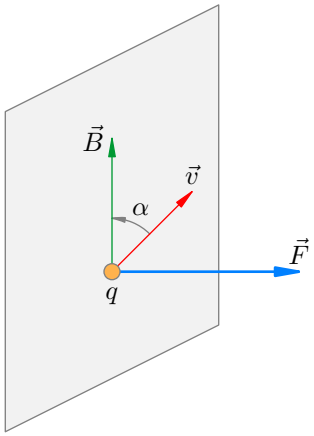


Рис. 3.83. Сила Лоренца

Направление силы Лоренца определяется в данном случае по одному из двух альтернативных правил.

Правило часовой стрелки. Сила Лоренца направлена туда, глядя откуда кратчайший поворот вектора скорости частицы $\vec{v}$ к вектору магнитной индукции $\vec{B}$ виден против часовой стрелки.

Правило левой руки. Располагаем левую руку так, чтобы четыре пальца указывали направление скорости частицы, а линии поля входили в ладонь. Тогда оттопыренный большой палец укажет направление силы Лоренца.

Для отрицательного заряда q направление силы Лоренца меняется на противоположное.

Всё вышеперечисленное является обобщением опытных фактов. Формула (3.72) позволяет связать размерность индукции магнитного поля с размерностями других физических величин:

$$B = \frac{F}{qv\sin\alpha} \;\Rightarrow\; \text{Тл} = \frac{\text{Н}\cdot\text{с}}{\text{Кл}\cdot\text{м}} = \frac{\text{В}\cdot\text{с}}{\text{м}^2}.$$

3.18.2 Сила Ампера

Если металлический проводник с током поместить в магнитное поле, то на этот проводник со стороны магнитного поля будет действовать сила, которая называется *силой Ампера*.

Происхождение силы Ампера легко понять. Ведь ток в металле является направленным движением электронов, а на каждый электрон действует сила Лоренца. Все эти силы Лоренца, действующие на свободные электроны, имеют одинаковое направление и одинаковую величину; они складываются друг с другом и дают результирующую силу Ампера.

Направление силы Ампера определяется по тем же двум правилам, сформулированным выше.

Правило часовой стрелки. Сила Ампера направлена туда, глядя откуда кратчайший поворот тока к полю виден против часовой стрелки.

Правило левой руки. Располагаем левую руку так, чтобы четыре пальца указывали направление тока, а линии поля входили в ладонь. Тогда оттопыренный большой палец укажет направление силы Ампера.

Взаимное расположение тока, поля и силы Ампера $\vec{F}$ указано на рис. 3.84.

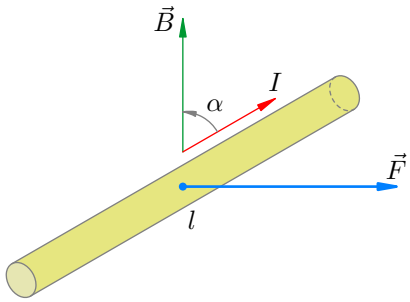


Рис. 3.84. Сила Ампера

На этом рисунке проводник имеет длину l, а угол между направлениями тока и поля равен α. Мы сейчас выведем выражение для абсолютной величины силы Ампера.

На каждый свободный электрон действует сила Лоренца:

$$F_1 = evB\sin\alpha,$$

где v — скорость направленного движения свободных электронов в проводнике.

Пусть N — число свободных электронов в данном проводнике, n — их концентрация (число в единице объёма). Тогда:

$$N = nV = nSl,$$

где V — объём проводника, S — площадь его поперечного сечения. Получаем:

$$F = NF_1 = nSl\cdot evB\sin\alpha = (envS)Bl\sin\alpha.$$

Мы не случайно выделили скобками четыре сомножителя. Ведь это есть не что иное, как сила тока: $I = envS$ (вспомните выражение силы тока через скорость направленного движения свободных зарядов!). В результате приходим к окончательной формуле для силы Ампера:

$$F = IBl\sin\alpha. \tag{3.73}$$

Хорошую возможность поупражняться в нахождении направлений магнитного поля и силы Ампера даёт взаимодействие параллельных токов. Оказывается, два параллельных провода отталкиваются, если направления токов в них противоположны, и притягиваются, если направления токов совпадают (рис. 3.85).

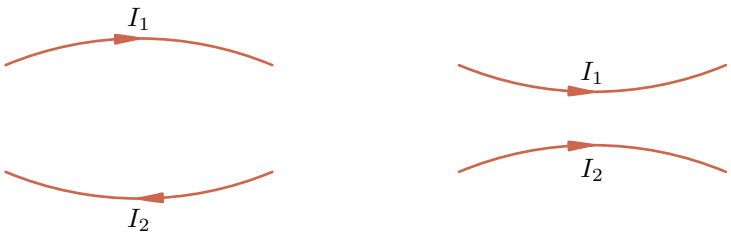


Рис. 3.85. Взаимодействие параллельных токов

Обязательно убедитесь в этом самостоятельно! Делаем так. Сначала берём произвольную точку на первом проводе и определяем направление магнитного поля, создаваемого в этой точке вторым проводом (правило вам известно — см. предыдущий листок). Ну а затем находим направление силы Ампера, действующей на первый провод со стороны магнитного поля второго провода.

3.18.3 Рамка с током в магнитном поле

В листках по термодинамике мы говорили о важности циклически работающих машин: они снабжают нас энергией. Понимание законов термодинамики позволило сконструировать тепловые двигатели, которые исправно служат нам и по сей день.

Понимание же законов электромагнетизма дало возможность создать циклическую машину другого типа — электродвигатель.

Мы рассмотрим один из элементов электродвигателя — рамку с током в магнитном поле. Разобравшись в её поведении, мы сможем уловить основную идею функционирования электродвигателя.

Пусть прямоугольная рамка 1234 может вращаться вокруг горизонтальной оси (рис. 3.86, слева). Рамка находится в вертикальном однородном магнитном поле $\vec{B}$. Ток течёт по рамке в направлении $1 \to 2 \to 3 \to 4 \to 1$; это направление показано соответствующими стрелками.

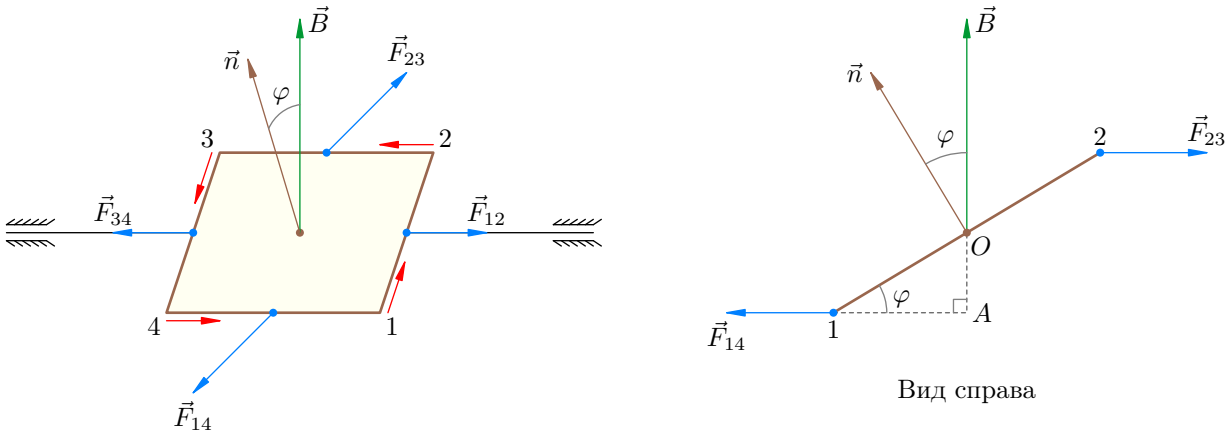


Рис. 3.86. Рамка с током в магнитном поле

Вектор $\vec{n}$ называется *вектором нормали*; он перпендикулярен плоскости рамки и направлен туда, глядя откуда ток кажется циркулирующим против часовой стрелки. (Иными словами,

вектор $\vec{n}$ сонаправлен с вектором индукции магнитного поля, которое создаётся током в рамке.) Поворот рамки измеряется углом φ между векторами $\vec{n}$ и $\vec{B}$.

Теперь определим направления сил Ампера, которые действуют на рамку со стороны магнитного поля. Эти силы расставлены на рисунке; вот вам ещё одно упражнение на правило часовой стрелки (левой руки) — обязательно проверьте правильность указанных направлений!

Силы $\vec{F}_{12}$ и $\vec{F}_{34}$, приложенные к сторонам 12 и 34, действуют вдоль оси вращения. Они лишь растягивают рамку и не вызывают её вращение.

Куда более интересны силы $\vec{F}_{23}$ и $\vec{F}_{14}$, приложеные соответственно к сторонам 23 и 14. Они лежат в горизонтальной плоскости и перпендикулярны оси вращения. Эти силы *вращают* рамку в направлении по часовой стрелке, если смотреть справа (рис. 3.86, правая часть). Вычислим момент этой пары сил относительно оси O вращения рамки.

Пусть длина стороны 14 равна a. Тогда

$$F_{14} = F_{23} = IBa.$$

Пусть длина стороны 12 равна b. Плечо d силы F_{14}, как видно из рис. 3.86 (справа) равно:

$$d = OA = \frac{b}{2}\sin\varphi.$$

Таким же будет плечо силы F_{23}. Отсюда получаем момент сил, вращающий рамку:

$$\mathcal{M} = F_{14}d + F_{23}d = IBa \cdot \frac{b}{2}\sin\varphi + IBa \cdot \frac{b}{2}\sin\varphi = IBab\sin\varphi.$$

Теперь заметим, что $ab = S$ — площадь рамки. Окончательно имеем:

$$\mathcal{M} = IBS\sin\varphi. \tag{3.74}$$

В этой формуле площадь служит единственной геометрической характеристикой рамки. Это наводит на мысль, что только площадь рамки и существенна в выражении для вращающего момента. И действительно, можно доказать[40], что формула (3.74) справедлива для рамки *любой формы* с площадью S.

Как видно из формулы (3.74), максимальный вращающий момент равен:

$$\mathcal{M}_{\max} = IBS.$$

Эта максимальная величина момента достигается при $\varphi = \pi/2$, то есть когда плоскость рамки параллельна магнитному полю.

Вращающий момент становится равным нулю при $\varphi = 0$ и $\varphi = \pi$. Оба этих положения по-своему интересны.

При $\varphi = \pi$ плоскость рамки перпендикулярна полю, а векторы $\vec{n}$ и $\vec{B}$ направлены в разные стороны. Данное положение является *положением неустойчивого равновесия*: стоит хоть немного шевельнуть рамку, как силы Ампера начнут её вращать в том же направлении, поворачивая вектор $\vec{n}$ к вектору $\vec{B}$ (убедитесь!).

При $\varphi = 0$ плоскость рамки также перпендикулярна полю, а векторы $\vec{n}$ и $\vec{B}$ сонаправлены. Это — *положение устойчивого равновесия*: при отклонении рамки возникает вращающий момент, стремящийся вернуть рамку назад (убедитесь!). Начнутся колебания рамки, постепенно затухающие из-за трения. В конце концов рамка остановится в положении $\varphi = 0$; в этом положении вектор индукции магнитного поля рамки сонаправлен с вектором $\vec{B}$ индукции внешнего

[40] Разбивая рамку на бесконечно узкие полоски, неотличимые от прямоугольников.

магнитного поля[41]. Полезное сопоставление: *рамка занимает такое положение, что её положительная нормаль ориентируется в том же направлении, что и северный конец стрелки компаса, помещённой в это магнитное поле.*

Таким образом, поведение рамки в магнитном поле становится ясным: если отклонить рамку от положения устойчивого равновесия и отпустить, то рамка будет совершать колебания. С точки зрения совершения механической работы это не очень хорошо: если намотать нить на ось вращения и подвесить к нити груз, то груз будет то подниматься, то опускаться.

Но вот если исхитриться и заставить ток менять направление в нужные моменты, то вместо колебаний рамки начнётся её непрерывное вращение и, соответственно, непрерывный подъём подвешенного груза. Тогда-то и получится полноценный электродвигатель; идея с переменой направления тока реализуется с помощью *коллектора* и *щёток*.

[41] Вот почему при намагничивании вещества элементарные токи ориентируются так, что их поля направлены в сторону внешнего магнитного поля.

3.19 Электромагнитная индукция

Опыт Эрстеда показал, что электрический ток создаёт в окружающем пространстве магнитное поле. Майкл Фарадей пришёл к мысли, что может существовать и обратный эффект: магнитное поле, в свою очередь, порождает электрический ток.

Иными словами, пусть в магнитном поле находится замкнутый проводник; не будет ли в этом проводнике возникать электрический ток под действием магнитного поля?

Через десять лет поисков и экспериментов Фарадею наконец удалось этот эффект обнаружить. В 1831 году он поставил следующие опыты.

1. На одну и ту же деревянную основу были намотаны две катушки; витки второй катушки были проложены между витками первой и изолированы. Выводы первой катушки подключались к источнику тока, выводы второй катушки — к гальванометру[42]. Таким образом, получались два контура: «источник тока — первая катушка» и «вторая катушка — гальванометр». Электрического контакта между контурами не было, только лишь магнитное поле первой катушки пронизывало вторую катушку.

   При замыкании цепи первой катушки гальванометр регистрировал короткий и слабый импульс тока во второй катушке.

   Когда по первой катушке протекал постоянный ток, никакого тока во второй катушке не возникало.

   При размыкании цепи первой катушки снова возникал короткий и слабый импульс тока во второй катушке, но на сей раз в обратном направлении по сравнению с током при замыкании цепи.

   Вывод.

   Меняющееся во времени магнитное поле первой катушки порождает (или, как говорят, *индуцирует*) электрический ток во второй катушке. Этот ток называется *индукционным током*.

   Если магнитное поле первой катушки увеличивается (в момент нарастания тока при замыкании цепи), то индукционный ток во второй катушке течёт в одном направлении.

   Если магнитное поле первой катушки уменьшается (в момент убывания тока при размыкании цепи), то индукционный ток во второй катушке течёт в другом направлении.

   Если магнитное поле первой катушки не меняется (постоянный ток через неё), то индукционного тока во второй катушке нет.

   Обнаруженное явление Фарадей назвал *электромагнитной индукцией* (т. е. «наведение электричества магнетизмом»).

2. Для подтверждения догадки о том, что индукционный ток порождается *переменным* магнитным полем, Фарадей перемещал катушки друг относительно друга. Цепь первой катушки всё время оставалась замкнутой, по ней протекал постоянный ток, но за счёт перемещения (сближения или удаления) вторая катушка оказывалась в переменном магнитном поле первой катушки.

   Гальванометр снова фиксировал ток во второй катушке. Индукционный ток имел одно направление при сближении катушек, и другое — при их удалении. *При этом сила индукционного тока была тем больше, чем быстрее перемещались катушки.*

[42] Гальванометр — чувствительный прибор для измерения малых токов.

3. Первая катушка была заменена постоянным магнитом. При внесении магнита внутрь второй катушки возникал индукционный ток. При выдвигании магнита снова появлялся ток, но в другом направлении. И опять-таки сила индукционного тока была тем больше, чем быстрее двигался магнит.

Эти и последующие опыты показали, что индукционный ток в проводящем контуре возникает во всех тех случаях, когда меняется «количество линий» магнитного поля, пронизывающих контур. Сила индукционного тока оказывается тем больше, чем быстрее меняется это количество линий. Направление тока будет одним при увеличении количества линий сквозь контур, и другим — при их уменьшении.

Замечательно, что для величины силы тока в данном контуре важна лишь скорость изменения количества линий. Что конкретно при этом происходит, роли не играет — меняется ли само поле, пронизывающее неподвижный контур, или же контур перемещается из области с одной густотой линий в область с другой густотой.

Такова суть закона электромагнитной индукции. Но, чтобы написать формулу и производить расчёты, нужно чётко формализовать расплывчатое понятие «количество линий поля сквозь контур».

3.19.1 Магнитный поток

Понятие магнитного потока как раз и является характеристикой количества линий магнитного поля, пронизывающих контур.

Для простоты мы ограничиваемся случаем однородного магнитного поля. Рассмотрим контур площади S, находящийся в магнитном поле с индукцией $\vec{B}$.

Пусть сначала магнитное поле перпендикулярно плоскости контура (рис. 3.87).

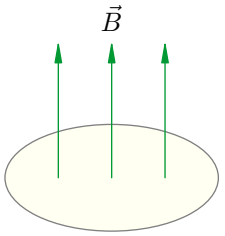


Рис. 3.87. $\Phi = BS$

В этом случае магнитный поток Φ определяется очень просто — как произведение индукции магнитного поля на площадь контура:

$$\Phi = BS. \tag{3.75}$$

Теперь рассмотрим общий случай, когда вектор $\vec{B}$ образует угол α с нормалью к плоскости контура (рис. 3.88).

Мы видим, что теперь сквозь контур «протекает» лишь перпендикулярная составляющая $\vec{B}_\perp$ вектора магнитной индукции $\vec{B}$ (а та составляющая, которая параллельна контуру, не «течёт» сквозь него). Поэтому, согласно формуле (3.75), имеем: $\Phi = B_\perp S$. Но $B_\perp = B\cos\alpha$, поэтому

$$\Phi = BS\cos\alpha. \tag{3.76}$$

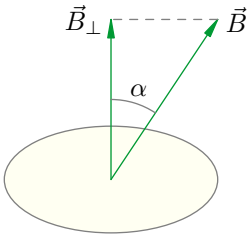


Рис. 3.88. $\Phi = BS\cos\alpha$

Это и есть общее определение магнитного потока в случае однородного магнитного поля. Обратим внимание на следующие две ситуации.

- Если вектор $\vec{B}$ перпендикулярен плоскости контура, то $\alpha = 0$. Тогда $\cos\alpha = 1$, и мы снова приходим к формуле (3.75).
- Если вектор $\vec{B}$ параллелен плоскости контура (то есть $\alpha = 90°$), то $\cos\alpha = 0$ и магнитный поток становится равным нулю (магнитное поле вообще не «течёт» сквозь контур).

А как определить магнитный поток, если поле не является однородным? Укажем лишь идею. Поверхность контура разбивается на очень большое число очень маленьких площадок, в пределах которых поле можно считать однородным. Для каждой площадки вычисляем свой маленький магнитный поток по формуле (3.76), а затем все эти магнитные потоки суммируем.

Единицей измерения магнитного потока является *вебер* (Вб). Как видим,

$$\text{Вб} = \text{Тл} \cdot \text{м}^2 = \text{В} \cdot \text{с}. \tag{3.77}$$

Почему же магнитный поток характеризует «количество линий» магнитного поля, пронизывающих контур? Очень просто. «Количество линий» определяется их густотой (а значит, величиной B — ведь чем больше индукция, тем гуще линии) и «эффективной» площадью, пронизываемой полем (а это есть не что иное, как $S\cos\alpha$). Но множители B и $S\cos\alpha$ как раз и образуют магнитный поток!

Теперь мы можем дать более чёткое определение явления электромагнитной индукции, открытого Фарадеем.

Электромагнитная индукция — это явление возникновения электрического тока в замкнутом проводящем контуре при изменении магнитного потока, пронизывающего контур.

3.19.2 ЭДС индукции

Каков механизм возникновения индукционного тока? Это мы обсудим позже. Пока ясно одно: при изменении магнитного потока, проходящего через контур, на свободные заряды в контуре действуют некоторые силы — *сторонние силы*, вызывающие движение зарядов.

Как мы знаем, работа сторонних сил по перемещению единичного положительного заряда вокруг контура называется электродвижущей силой (ЭДС): $\mathcal{E} = A_{\text{ст}}/q$. В нашем случае, когда меняется магнитный поток сквозь контур, соответствующая ЭДС называется *ЭДС индукции* и обозначается $\mathcal{E}_i$.

Итак, *ЭДС индукции $\mathcal{E}_i$ — это работа сторонних сил, возникающих при изменении магнитного потока через контур, по перемещению единичного положительного заряда вокруг контура.*

Природу сторонних сил, возникающих в данном случае в контуре, мы скоро выясним.

3.19.3 Закон электромагнитной индукции Фарадея

Сила индукционного тока в опытах Фарадея оказывалась тем больше, чем быстрее менялся магнитный поток через контур.

Если за малое время Δt изменение магнитного потока равно $\Delta\Phi$, то *скорость* изменения магнитного потока — это дробь $\Delta\Phi/\Delta t$ (или, что то же самое, производная $\dot{\Phi}$ магнитного потока по времени).

Опыты показали, что сила индукционного тока I прямо пропорциональна модулю скорости изменения магнитного потока:

$$I \sim \left|\frac{\Delta\Phi}{\Delta t}\right|.$$

Модуль поставлен для того, чтобы не связываться пока с отрицательными величинами (ведь при убывании магнитного потока будет $\Delta\Phi < 0$). Впоследствии мы это модуль снимем.

Из закона Ома для полной цепи мы в то же время имеем: $I \sim \mathcal{E}_i$. Поэтому ЭДС индукции прямо пропорциональна скорости изменения магнитного потока:

$$\mathcal{E}_i \sim \left|\frac{\Delta\Phi}{\Delta t}\right|. \tag{3.78}$$

ЭДС измеряется в вольтах. Но и скорость изменения магнитного потока также измеряется в вольтах! Действительно, из (3.77) мы видим, что Вб/с = В. Стало быть, единицы измерения обеих частей пропорциональности (3.78) совпадают, поэтому коэффициент пропорциональности — величина безразмерная. В системе СИ она полагается равной единице, и мы получаем:

$$\mathcal{E}_i = \left|\frac{\Delta\Phi}{\Delta t}\right| = |\dot{\Phi}|. \tag{3.79}$$

Это и есть *закон электромагнитной индукции* или *закон Фарадея*. Дадим его словесную формулировку.

Закон электромагнитной индукции Фарадея. При изменении магнитного потока, пронизывающего контур, в этом контуре возникает ЭДС индукции, равная модулю скорости изменения магнитного потока.

3.19.4 Правило Ленца

Магнитный поток, изменение которого приводит к появлению индукционного тока в контуре, мы будем называть *внешним магнитным потоком*. А само магнитное поле, которое создаёт этот магнитный поток, мы будем называть *внешним магнитным полем*.

Зачем нам эти термины? Дело в том, что индукционный ток, возникающий в контуре, создаёт своё *собственное* магнитное поле, которое по принципу суперпозиции складывается с внешним магнитным полем. Соответственно, наряду с внешним магнитным потоком через контур будет проходить *собственный* магнитный поток, создаваемый магнитным полем индукционного тока.

Оказывается, эти два магнитных потока — собственный и внешний — связаны между собой строго определённым образом.

Правило Ленца. Индукционный ток всегда имеет такое направление, что собственный магнитный поток препятствует изменению внешнего магнитного потока.

Правило Ленца позволяет находить направление индукционного тока в любой ситуации. Рассмотрим некоторые примеры применения правила Ленца.

Предположим, что контур пронизывается магнитным полем, которое возрастает со временем (рис. 3.89). Например, мы приближаем снизу к контуру магнит, северный полюс которого направлен в данном случае вверх, к контуру.

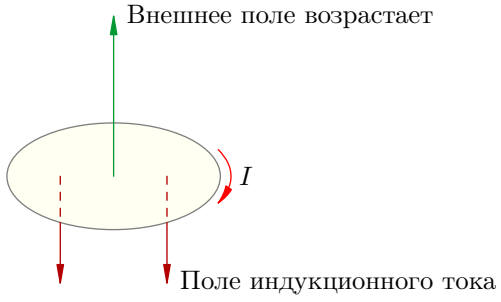


Рис. 3.89. Магнитный поток возрастает

Магнитный поток через контур увеличивается. Индукционный ток будет иметь такое направление, чтобы создаваемый им магнитный поток препятствовал увеличению внешнего магнитного потока. Для этого магнитное поле, создаваемое индукционным током, должно быть направлено *против* внешнего магнитного поля.

Индукционный ток течёт против часовой стрелки, если смотреть со стороны создаваемого им магнитного поля. В данном случае ток будет направлен по часовой стрелке, если смотреть сверху, со стороны внешнего магнитного поля, как и показано на (рис. 3.89).

Теперь предположим, что магнитное поле, пронизывающее контур, уменьшается со временем (рис. 3.90). Например, мы удаляем магнит вниз от контура, а северный полюс магнита направлен на контур.

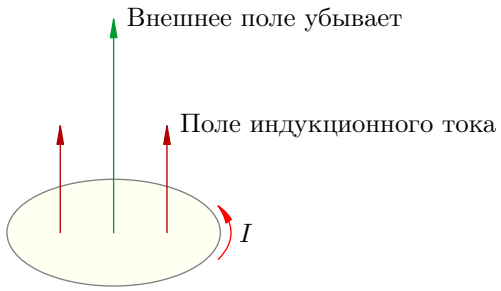


Рис. 3.90. Магнитный поток убывает

Магнитный поток через контур уменьшается. Индукционный ток будет иметь такое направление, чтобы его собственный магнитный поток поддерживал внешний магнитный поток, препятствуя его убыванию. Для этого магнитное поле индукционного тока должно быть направлено *в ту же сторону*, что и внешнее магнитное поле.

В этом случае индукционный ток потечёт против часовой стрелки, если смотреть сверху, со стороны обоих магнитных полей.

3.19.5 Взаимодействие магнита с контуром

Итак, приближение или удаление магнита приводит к появлению в контуре индукционного тока, направление которого определяется правилом Ленца. Но ведь магнитное поле действует на ток! Появится сила Ампера, действующая на контур со стороны поля магнита. Куда будет направлена эта сила?

Если вы хотите хорошо разобраться в правиле Ленца и в определении направления силы Ампера, попробуйте ответить на данный вопрос самостоятельно. Это не очень простое упражнение. Рассмотрите четыре возможных случая.

1. Магнит приближаем к контуру, северный полюс направлен на контур.
2. Магнит удаляем от контура, северный полюс направлен на контур.
3. Магнит приближаем к контуру, южный полюс направлен на контур.
4. Магнит удаляем от контура, южный полюс направлен на контур.

Не забывайте, что поле магнита не однородно: линии поля расходятся от северного полюса и сходятся к южному. Это очень существенно для определения результирующей силы Ампера. Результат получается следующий.

Если приближать магнит, то контур отталкивается от магнита. Если удалять магнит, то контур притягивается к магниту. Таким образом, если контур подвешен на нити, то он всегда будет отклоняться в сторону движения магнита, словно следуя за ним. Расположение полюсов магнита при этом роли не играет.

Результат этот можно объяснить и из совершенно общих соображений — при помощи закона сохранения энергии.

Допустим, мы приближаем магнит к контуру. В контуре появляется индукционный ток. Но для создания тока надо совершить работу! Кто её совершает? В конечном счёте — мы, перемещая магнит. Мы совершаем положительную механическую работу, которая преобразуется в положительную работу возникающих в контуре сторонних сил, создающих индукционный ток.

Итак, наша работа по перемещению магнита должна быть *положительна*. Это значит, что мы, приближая магнит, должны *преодолевать* силу взаимодействия магнита с контуром, которая, стало быть, является силой *отталкивания*.

Теперь удаляем магнит. Повторите, пожалуйста, эти рассуждения и убедитесь, что между магнитом и контуром должна возникнуть сила притяжения.

3.19.6 Закон Фарадея + Правило Ленца = Снятие модуля

Выше мы обещали снять модуль в законе Фарадея (3.79). Правило Ленца позволяет это сделать. Но сначала нам нужно будет договориться о знаке ЭДС индукции — ведь без модуля, стоящего в правой части (3.79), величина ЭДС может получаться как положительной, так и отрицательной.

Прежде всего, фиксируется одно из двух возможных направлений обхода контура. Это направление объявляется *положительным*. Противоположное направление обхода контура называется, соответственно, *отрицательным*. Какое именно направление обхода мы берём в качестве положительного, роли не играет — важно лишь сделать этот выбор.

Магнитный поток через контур считается положительным ($\Phi > 0$), если магнитное поле, пронизывающее контур, направлено туда, глядя откуда обход контура в положительном направлении совершается против часовой стрелки. Если же с конца вектора магнитной индукции положительное направление обхода видится по часовой стрелке, то магнитный поток считается отрицательным ($\Phi < 0$).

ЭДС индукции считается положительной ($\mathcal{E}_i > 0$), если индукционный ток течёт в положительном направлении. В этом случае направление сторонних сил, возникающих в контуре при изменении магнитного потока через него, совпадает с положительным направлением обхода контура.

Наоборот, ЭДС индукции считается отрицательной ($\mathcal{E}_i < 0$), если индукционный ток течёт в отрицательном направлении. Сторонние силы в данном случае также будут действовать вдоль отрицательного направления обхода контура.

Итак, пусть контур находится в магнитном поле $\vec{B}$. Фиксируем направление положительного обхода контура. Предположим, что магнитное поле направлено туда, глядя откуда положительный обход совершается против часовой стрелки. Тогда магнитный поток положителен: $\Phi > 0$.

Предположим, далее, что магнитный поток увеличивается ($\Delta\Phi/\Delta t > 0$). Согласно правилу Ленца индукционный ток потечёт в отрицательном направлении (рис. 3.91).

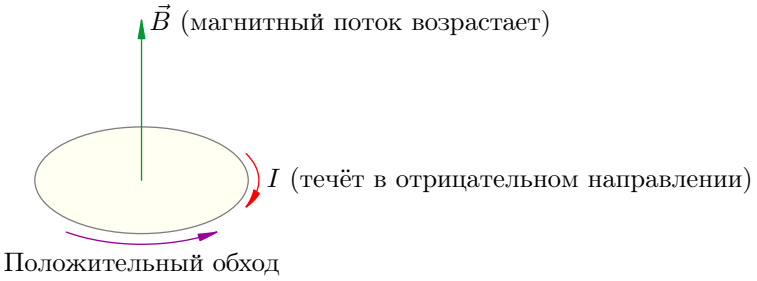


Рис. 3.91. Магнитный поток возрастает $\Rightarrow \mathcal{E}_i < 0$

Стало быть, в данном случае имеем $\mathcal{E}_i < 0$. Знак ЭДС индукции оказался противоположен знаку скорости изменения магнитного потока. Проверим это в другой ситуации.

А именно, предположим теперь, что магнитный поток убывает ($\Delta\Phi/\Delta t < 0$). По правилу Ленца индукционный ток потечёт в положительном направлении. Стало быть, $\mathcal{E}_i > 0$ (рис. 3.92).

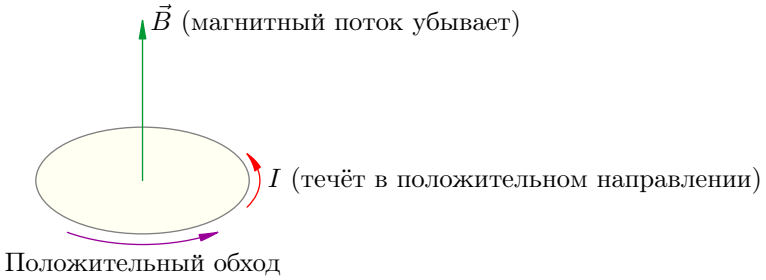


Рис. 3.92. Магнитный поток возрастает $\Rightarrow \mathcal{E}_i > 0$

Таков в действительности общий факт: *при нашей договорённости о знаках правило Ленца всегда приводит к тому, что знак ЭДС индукции противоположен знаку скорости изменения магнитного потока*:

$$\mathcal{E}_i = -\frac{\Delta\Phi}{\Delta t} = -\dot{\Phi}. \tag{3.80}$$

Тем самым ликвидирован знак модуля в законе электромагнитной индукции Фарадея.

3.19.7 Вихревое электрическое поле

Рассмотрим неподвижный контур, находящийся в переменном магнитном поле. Каков же механизм возникновения индукционного тока в контуре? А именно, какие силы вызывают движение свободных зарядов, какова природа этих сторонних сил?

Пытаясь ответить на эти вопросы, великий английский физик Максвелл открыл фундаментальное свойство природы: *меняющееся во времени магнитное поле порождает поле электрическое*. Именно это электрическое поле и действует на свободные заряды, вызывая индукционный ток.

Линии возникающего электрического поля оказываются замкнутыми, в связи с чем оно было названо *вихревым электрическим полем*. Линии вихревого электрического поля идут вокруг линий магнитного поля и направлены следующим образом.

Пусть магнитное поле увеличивается. Если в нём находится проводящий контур, то индукционный ток потечёт в соответствии с правилом Ленца — по часовой стрелке, если смотреть с конца вектора $\vec{B}$. Значит, туда же направлена и сила, действующая со стороны вихревого электрического поля на положительные свободные заряды контура; значит, именно туда направлен вектор напряжённости вихревого электрического поля.

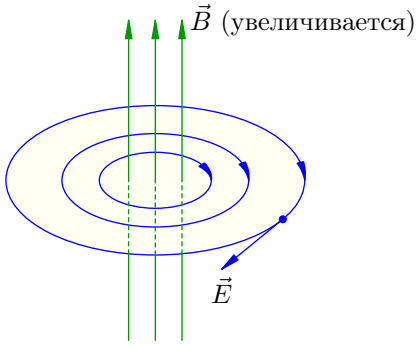


Рис. 3.93. Вихревое электрическое поле при увеличении магнитного поля

Итак, линии напряжённости вихревого электрического поля направлены в данном случае по часовой стрелке (смотрим с конца вектора $\vec{B}$, (рис. 3.93).

Наоборот, если магнитное поле убывает, то линии напряжённости вихревого электрического поля направлены против часовой стрелки (рис. 3.94).

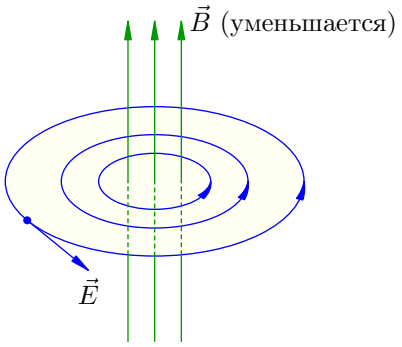


Рис. 3.94. Вихревое электрическое поле при уменьшении магнитного поля

Теперь мы можем глубже понять явление электромагнитной индукции. Суть его состоит именно в том, что переменное магнитное поле порождает вихревое электрическое поле. Данный эффект не зависит от того, присутствует ли в магнитном поле замкнутый проводящий контур или нет; с помощью контура мы лишь обнаруживаем это явление, наблюдая индукционный ток.

Вихревое электрическое поле по некоторым свойствам отличается от уже известных нам электрических полей: электростатического поля и стационарного поля зарядов, образующих постоянный ток.

1. Линии вихревого поля замкнуты, тогда как линии электростатического и стационарного полей начинаются на положительных зарядах и оканчиваются на отрицательных.

2. Вихревое поле непотенциально: его работа перемещению заряда по замкнутому контуру не равна нулю. Иначе вихревое поле не могло бы создавать электрический ток! В то же время, как мы знаем, электростатическое и стационарное поля являются потенциальными.

Итак, *ЭДС индукции в неподвижном контуре — это работа вихревого электрического поля по перемещению единичного положительного заряда вокруг контура.*

Пусть, например, контур является кольцом радиуса r и пронизывается однородным переменным магнитным полем. Тогда напряжённость E вихревого электрического поля одинакова во всех точках кольца. Работа A силы F, с которой вихревое поле действует на заряд q, равна:

$$A = F \cdot 2\pi r = qE \cdot 2\pi r.$$

Следовательно, для ЭДС индукции получаем:

$$\mathcal{E}_i = \frac{A}{q} = 2\pi r E.$$

3.19.8 ЭДС индукции в движущемся проводнике

Если проводник перемещается в постоянном магнитном поле, то в нём также появляется ЭДС индукции. Однако причиной теперь служит не вихревое электрическое поле (оно вообще не появляется — ведь магнитное поле постоянно), а действие силы Лоренца на свободные заряды проводника.

Рассмотрим ситуацию, которая часто встречается в задачах. В горизонтальной плоскости расположены параллельные рельсы, расстояние между которыми равно l. Рельсы находятся в вертикальном однородном магнитном поле $\vec{B}$. По рельсам движется тонкий проводящий стержень PN со скоростью $\vec{v}$; он всё время остаётся перпендикулярным рельсам (рис. 3.95).

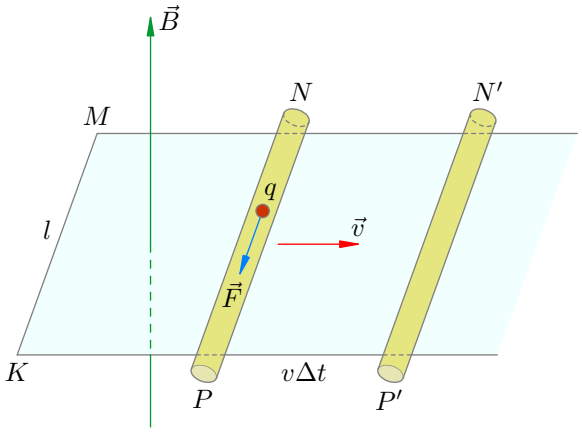


Рис. 3.95. Движение проводника в магнитном поле

Возьмём внутри стержня положительный свободный заряд q. Вследствие движения этого заряда вместе со стержнем со скоростью $\vec{v}$ на заряд будет действовать сила Лоренца:

$$F = qvB.$$

Направлена эта сила вдоль оси стержня, как показано на рисунке (убедитесь в этом сами — не забывайте правило часовой стрелки или левой руки!).

Сила Лоренца $\vec{F}$ играет в данном случае роль сторонней силы: она приводит в движение свободные заряды стержня. При перемещении заряда q от точки N к точке P наша сторонняя сила совершит работу:

$$A = Fl = qvBl.$$

(Длину стержня мы также считаем равной l.) Стало быть, ЭДС индукции в стержне окажется равной:

$$\mathcal{E}_i = \frac{A}{q} = vBl. \tag{3.81}$$

Таким образом, стержень PN аналогичен источнику тока с положительной клеммой P и отрицательной клеммой N. Внутри стержня за счёт действия сторонней силы Лоренца происходит разделение зарядов: положительные заряды двигаются к точке P, отрицательные — к точке N.

Допустим сначала, что рельсы не проводят ток. Тогда движение зарядов в стержне постепенно прекратится. Ведь по мере накопления положительных зарядов на торце P и отрицательных зарядов на торце N будет возрастать кулоновская сила, с которой положительный свободный заряд q отталкивается от P и притягивается к N — и в какой-то момент эта кулоновская сила уравновесит силу Лоренца. Между концами стержня установится разность потенциалов, равная ЭДС индукции (3.81).

Теперь предположим, что рельсы и перемычка KM являются проводящими. Тогда в цепи возникнет индукционный ток; он пойдёт в направлении $P \to K \to M \to N$ (от «плюса источника» P к «минусу» N). Предположим, что сопротивление стержня равно r (это аналог внутреннего сопротивления источника тока), а сопротивление участка $PKMN$ равно R (сопротивление внешней цепи). Тогда сила индукционного тока найдётся по закону Ома для полной цепи:

$$I = \frac{\mathcal{E}_i}{R+r} = \frac{vBl}{R+r}.$$

Замечательно, что выражение (3.81) для ЭДС индукции можно получить также с помощью закона Фарадея. Сделаем это.

За время Δt наш стержень PN проходит путь $v\Delta t$ и занимает положение $P'N'$ (рис. 3.95). Площадь контура возрастает на величину площади прямоугольника $PP'N'N$:

$$\Delta S = S_{PP'N'N} = lv\Delta t.$$

Магнитный поток через контур увеличивается. Приращение магнитного потока равно:

$$\Delta\Phi = B\Delta S = Blv\Delta t.$$

Скорость изменения магнитного потока положительна и равна ЭДС индукции:

$$\mathcal{E}_i = \frac{\Delta\Phi}{\Delta t} = Blv.$$

Мы получили тот же самый результат, что и в (3.81).

Направление индукционного тока, заметим, подчиняется правилу Ленца. Действительно, раз ток течёт в направлении $P \to K \to M \to N$, то его магнитное поле направлено противоположно внешнему полю $\vec{B}$ и, стало быть, препятствует возрастанию магнитного потока через контур.

На этом примере мы видим, что в ситуациях, когда проводник движется в магнитном поле, можно действовать двояко: либо с привлечением силы Лоренца как сторонней силы, либо с помощью закона Фарадея. Результаты будут получаться одинаковые.

3.20 Самоиндукция

Самоиндукция является частным случаем электромагнитной индукции. Оказывается, что электрический ток в контуре, меняющийся со временем, определённым образом воздействует сам на себя.

Ситуация 1. Предположим, что сила тока в контуре возрастает. Пусть ток течёт против часовой стрелки; тогда магнитное поле этого тока направлено вверх и увеличивается (рис. 3.96).

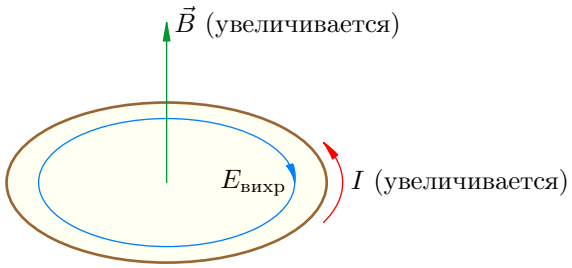


Рис. 3.96. Вихревое поле препятствует увеличению тока

Таким образом, наш контур оказывается в переменном магнитном поле своего собственного тока. Магнитное поле в данном случае возрастает (вместе с током) и потому порождает вихревое электрическое поле, линии которого направлены по часовой стрелке в соответствии с правилом Ленца.

Как видим, вихревое электрическое поле направлено против тока, препятствуя его возрастанию; оно как бы «тормозит» ток. Поэтому при замыкании любой цепи ток устанавливается не мгновенно — требуется некоторое время, чтобы преодолеть тормозящее действие возникающего вихревого электрического поля.

Ситуация 2. Предположим теперь, что сила тока в контуре уменьшается. Магнитное поле тока также убывает и порождает вихревое электрическое поле, направленное против часовой стрелки (рис. 3.97).

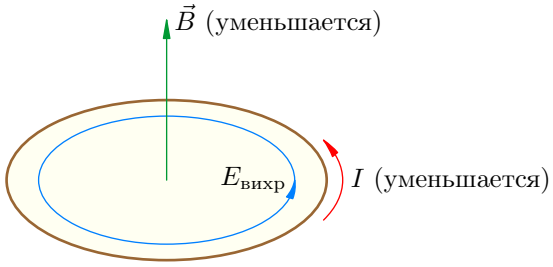


Рис. 3.97. Вихревое поле поддерживает убывающий ток

Теперь вихревое электрическое поле направлено в ту же сторону, что и ток; оно поддерживает ток, препятствуя его убыванию.

Как мы знаем, работа вихревого электрического поля по перемещению единичного положительного заряда вокруг контура — это ЭДС индукции. Поэтому мы можем дать такое определение.

Явление самоиндукции состоит в том, что при изменении силы тока в контуре возникает ЭДС индукции в этом же самом контуре.

При возрастании силы тока (в ситуации 1) вихревое электрическое поле совершает отрицательную работу, тормозя свободные заряды. Стало быть, ЭДС индукции в этом случае отрицательна.

При убывании силы тока (в ситуации 2) вихревое электрическое поле совершает положительную работу, «подталкивая» свободные заряды и препятствуя убыванию тока. ЭДС индукции в этом случае также положительна[43].

3.20.1 Индуктивность

Мы знаем, что магнитный поток, пронизывающий контур, пропорционален индукции магнитного поля: $\Phi \sim B$. Кроме того, опыт показывает, что величина индукции магнитного поля контура с током пропорциональна силе тока: $B \sim I$. Стало быть, магнитный поток через поверхность контура, создаваемый магнитным полем тока в этом самом контуре, пропорционален силе тока: $\Phi \sim I$.

Коэффициент пропорциональности обозначается L и называется *индуктивностью* контура:

$$\Phi = LI. \tag{3.82}$$

Индуктивность зависит от геометрических свойств контура (формы и размеров), а также от магнитных свойств среды, в которую помещён контур[44]. Единицей измерения индуктивности служит *генри* (Гн).

Допустим, что форма контура, его размеры и магнитные свойства среды остаются постоянными (например, наш контур — это катушка, в которую не вводится сердечник); изменение магнитного потока через контур вызвано только изменением силы тока. Тогда $\Delta\Phi = L\Delta I$, и закон Фарадея $\mathcal{E}_i = -\Delta\Phi/\Delta t$ приобретает вид:

$$\mathcal{E}_i = -L\frac{\Delta I}{\Delta t} = -L\dot{I}. \tag{3.83}$$

Благодаря знаку «минус» в (3.83) ЭДС индукции оказывается отрицательной при возрастании тока и положительной при убывании тока, что мы и видели выше.

Рассмотрим два опыта, демонстрирующих явление самоиндукции при замыкании и размыкании цепи.

В первом опыте к батарейке подключены параллельно две лампочки, причём вторая — последовательно с катушкой достаточно большой индуктивности L (рис. 3.98). Ключ вначале разомкнут.

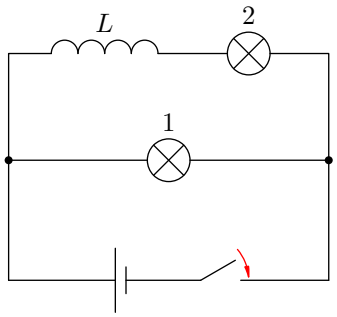


Рис. 3.98. Самоиндукция при замыкании цепи

При замыкании ключа лампочка 1 загорается сразу, а лампочка 2 — постепенно. Дело в том, что в катушке возникает ЭДС индукции, препятствующая возрастанию тока. Поэтому максимальное значение тока во второй лампочке устанавливается лишь спустя некоторое заметное время после вспыхивания первой лампочки.

[43] Нетрудно убедиться в том, что знак ЭДС индукции, определённый таким образом, согласуется с правилом выбора знака для ЭДС индукции, сформулированным в предыдущем разделе.

[44] Улавливаете аналогию? Ёмкость конденсатора зависит от его геометрических характеристик, а также от диэлектрической проницаемости среды между обкладками конденсатора.

Это время запаздывания тем больше, чем больше индуктивность катушки. Объяснение простое: ведь тогда больше будет напряжённость вихревого электрического поля, возникающего в катушке, и потому батарейке придётся совершить бóльшую работу по преодолению вихревого поля, тормозящего заряженные частицы.

Во втором опыте к батарейке подключены параллельно катушка и лампочка рис. 3.99). Сопротивление катушки много меньше сопротивления лампочки.

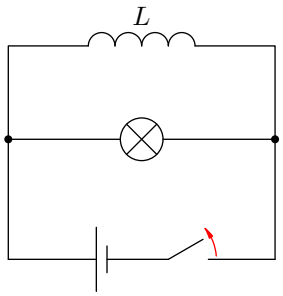


Рис. 3.99. Самоиндукция при размыкании цепи

Ключ вначале замкнут. Лампочка не горит — напряжение на ней близко к нулю из-за малости сопротивления катушки. Почти весь ток, идущий в неразветвлённой цепи, проходит через катушку.

При размыкании ключа лампочка ярко вспыхивает! Почему? Ток через катушку начинает резко убывать, и возникает значительная ЭДС индукции, поддерживающая убывающий ток (ведь ЭДС индукции, как видно из (3.83), пропорциональна скорости изменения тока).

Иными словами, при размыкании ключа в катушке появляется весьма большое вихревое электрическое поле, разгоняющее свободные заряды. Под действием этого вихревого поля через лампочку пробегает импульс тока, и мы видим яркую вспышку. При достаточно большой индуктивности катушки ЭДС индукции может стать существенно больше ЭДС батарейки, и лампочка вовсе перегорит.

Лампочку-то, может, и не жалко, но в промышленности и энергетике данный эффект является серьёзной проблемой. Так как при размыкании цепи ток начинает уменьшаться очень быстро, возникающая в цепи ЭДС индукции может значительно превышать номинальные напряжения и достигать опасно больших величин. Поэтому в агрегатах, потребляющих большой ток, предусмотрены специальные аппаратные меры предосторожности (например, масляные выключатели на электростанциях), препятствующие моментальному размыканию цепи.

3.20.2 Электромеханическая аналогия

Нетрудно заметить определённую аналогию между индуктивностью L в электродинамике и массой m в механике.

1. Чтобы разогнать тело до заданной скорости, требуется некоторое время — мгновенно изменить скорость тела не получается. При неизменной силе, приложенной к телу, это время тем больше, чем больше масса m тела.

   Чтобы ток в катушке достиг своего максимального значения, требуется некоторое время; мгновенно ток не устанавливается. Время установления тока тем больше, чем больше индуктивность L катушки.

2. Если тело налетает на неподвижную стену, то скорость тела уменьшается очень быстро. Стена принимает на себя удар, и его разрушительное действие тем сильнее, чем больше масса тела.

При размыкании цепи с катушкой ток уменьшается очень быстро. Цепь принимает на себя «удар» в виде вихревого электрического поля, порождаемого убывающим магнитным полем тока, и этот «удар» тем сильнее, чем больше индуктивность катушки. ЭДС индукции может достичь столь больших величин, что пробой воздушного промежутка выведет из строя оборудование.

На самом деле эти *электромеханические аналогии* простираются довольно далеко; они касаются не только индуктивности и массы, но и других величин, и оказываются весьма полезными на практике. Мы ещё поговорим об этом в листке про электромагнитные колебания.

3.20.3 Энергия магнитного поля

Вспомним второй опыт с лампочкой, которая не горит при замкнутом ключе и ярко вспыхивает при размыкании цепи. Мы непосредственно наблюдаем, что после размыкания ключа в лампочке выделяется энергия. Но откуда эта энергия берётся?

Берётся она, ясное дело, из катушки — больше неоткуда. Но что за энергия была запасена в катушке и как вычислить эту энергию? Чтобы понять это, продолжим нашу электромеханическую аналогию между индуктивностью и массой.

Чтобы разогнать тело массы m из состояния покоя до скорости v, внешняя сила должна совершить работу A. Тело приобретает кинетическую энергию, которая равна затраченной работе: $K = A = mv^2/2$.

Чтобы после замыкания цепи ток в катушке индуктивности L достиг величины I, источник тока должен совершить работу по преодолению вихревого электрического поля, направленного против тока. *Работа источника идёт на создание тока и превращается в энергию магнитного поля созданного тока.* Эта энергия запасается в катушке; именно эта энергия и выделяется потом в лампочке после размыкания ключа (во втором опыте).

Индуктивность L служит аналогом массы m; сила тока I является очевидным аналогом скорости v. Поэтому естественно предположить, что для энергии магнитного поля катушки может иметь место формула, аналогичная выражению для кинетической энергии:

$$W = \frac{LI^2}{2} \tag{3.84}$$

(тем более, что правая часть данной формулы имеет размерность энергии — проверьте!).

Формула (3.84) действительно оказывается справедливой. Уметь её выводить пока не обязательно, но если вы знаете, что такое интеграл, то вам не составит труда понять следующие рассуждения.

Пусть в данный момент сила тока через катушку равна I. Возьмём малый промежуток времени dt. В течение этого промежутка приращение силы тока равно dI; величина dt считается настолько малой, что dI много меньше, чем I.

За время dt по цепи проходит заряд $dq = Idt$. Вихревое электрическое поле совершает при этом отрицательную работу:

$$dA_{\text{вихр}} = \mathcal{E}_i dq = \mathcal{E}_i I dt = -L\frac{dI}{dt} I dt = -LIdI.$$

Источник тока совершает такую же по модулю положительную работу dA (сопротивлением катушки, напомним, мы пренебрегаем, так что вся работа источника совершается против вихревого поля):

$$dA = -dA_{\text{вихр}} = LIdI.$$

Интегрируя это от нуля до I, найдем работу источника A, которая затрачивается на создание тока I:

$$A = \int_0^I LI dI = \frac{LI^2}{2}\,.$$

Эта работа превращается в энергию W магнитного поля созданного тока, и мы приходим к формуле (3.84).

3.21 Электромагнитные колебания

Электромагнитные колебания — это периодические изменения заряда, силы тока и напряжения, происходящие в электрической цепи; кроме того, это периодические изменения напряжённости электрического поля и индукции магнитного поля, возникающие и распространяющиеся в окружающем пространстве.

Знакомство с электромагнитными колебаниями мы начнём с рассмотрения процессов, происходящих в колебательном контуре.

3.21.1 Колебательный контур

Колебательный контур — это замкнутый контур, образованный последовательно соединёнными конденсатором и катушкой. Колебательный контур является простейшей системой, в которой могут происходить электромагнитные колебания.

Зарядим конденсатор, подключим к нему катушку и замкнём цепь. Начнутся *свободные электромагнитные колебания* — периодические изменения заряда на конденсаторе и тока в катушке. Свободными, напомним, эти колебания называются потому, что они совершаются без какого-либо внешнего воздействия — только за счёт энергии, запасённой в контуре.

Период колебаний в контуре обозначим, как всегда, через T. Сопротивление катушки будем считать равным нулю.

Рассмотрим подробно все важные стадии процесса колебаний. Для большей наглядности будем проводить аналогию с колебаниями горизонтального пружинного маятника.

Начальный момент: $t = 0$. Заряд конденсатора равен q_0, ток через катушку отсутствует (рис. 3.100). Конденсатор сейчас начнёт разряжаться.

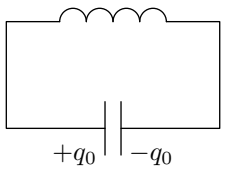


Рис. 3.100. $t = 0$

Несмотря на то, что сопротивление катушки равно нулю, ток не возрастёт мгновенно. Как только ток начнёт увеличиваться, в катушке возникнет ЭДС самоиндукции, препятствующая возрастанию тока.

Аналогия. Маятник оттянут вправо на величину x_0 и в начальный момент отпущен. Начальная скорость маятника равна нулю.

Первая четверть периода: $0 < t < T/4$. Конденсатор разряжается, его заряд в данный момент равен q. Ток I через катушку нарастает (рис. 3.101).

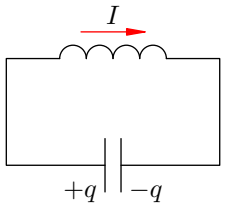


Рис. 3.101. $0 < t < T/4$

Увеличение тока происходит постепенно: вихревое электрическое поле катушки препятствует нарастанию тока и направлено против тока.

Аналогия. Маятник движется влево к положению равновесия; скорость v маятника постепенно увеличивается. Деформация пружины x (она же — координата маятника) уменьшается.

Конец первой четверти: $t = T/4$. Конденсатор полностью разрядился. Сила тока достигла максимального значения I_0 (рис. 3.102). Сейчас начнётся перезарядка конденсатора.

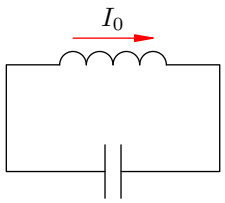


Рис. 3.102. $t = T/4$

Напряжение на катушке равно нулю, но ток не исчезнет мгновенно. Как только ток начнёт уменьшаться, в катушке возникнет ЭДС самоиндукции, препятствующая убыванию тока.

Аналогия. Маятник проходит положение равновесия. Его скорость достигает максимального значения v_0. Деформация пружины равна нулю.

Вторая четверть: $T/4 < t < T/2$. Конденсатор перезаряжается — на его обкладках появляется заряд противоположного знака по сравнению с тем, что был вначале (рис. 3.103).

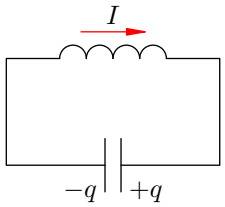


Рис. 3.103. $T/4 < t < T/2$

Сила тока убывает постепенно: вихревое электрическое поле катушки, поддерживая убывающий ток, сонаправлено с током.

Аналогия. Маятник продолжает двигаться влево — от положения равновесия к правой крайней точке. Скорость его постепенно убывает, деформация пружины увеличивается.

Конец второй четверти: $t = T/2$. Конденсатор полностью перезарядился, его заряд опять равен q_0 (но полярность другая). Сила тока равна нулю (рис. 3.104). Сейчас начнётся обратная перезарядка конденсатора.

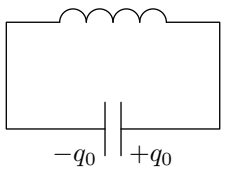


Рис. 3.104. $t = T/2$

Аналогия. Маятник достиг крайней правой точки. Скорость маятника равна нулю. Деформация пружины максимальна и равна x_0.

Третья четверть: $T/2 < t < 3T/4$. Началась вторая половина периода колебаний; процессы пошли в обратном направлении. Конденсатор разряжается (рис. 3.105).

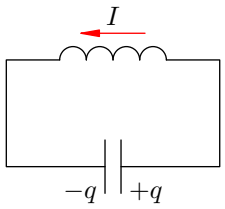


Рис. 3.105. $T/2 < t < 3T/4$

Аналогия. Маятник двигается обратно: от правой крайней точки к положению равновесия.

Конец третьей четверти: $t = 3T/4$. Конденсатор полностью разрядился. Ток максимален и снова равен I_0, но на сей раз имеет другое направление (рис. 3.106).

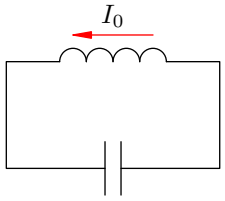


Рис. 3.106. $t = 3T/4$

Аналогия. Маятник снова проходит положение равновесия с максимальной скоростью v_0, но на сей раз в обратном направлении.

Четвёртая четверть: $3T/4 < t < T$. Ток убывает, конденсатор заряжается (рис. 3.107).

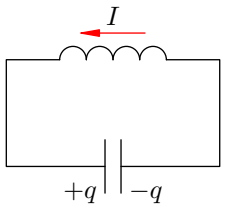


Рис. 3.107. $3T/4 < t < T$

Аналогия. Маятник продолжает двигаться вправо — от положения равновесия к крайней левой точке.

Конец четвёртой четверти и всего периода: $t = T$. Обратная перезарядка конденсатора завершена, ток равен нулю (рис. 3.108).

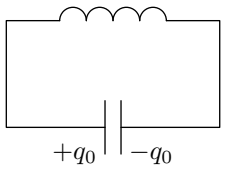


Рис. 3.108. $t = T$

Данный момент идентичен моменту $t = 0$, а данный рисунок — рисунку 3.100. Совершилось одно полное колебание. Сейчас начнётся следующее колебание, в течение которого процессы будут происходить точно так же, как описано выше.

Аналогия. Маятник вернулся в исходное положение.

Рассмотренные электромагнитные колебания являются *незатухающими* — они будут продолжаться бесконечно долго. Ведь мы предположили, что сопротивление катушки равно нулю! Точно так же будут незатухающими колебания пружинного маятника при отсутствии трения.

В реальности катушка обладает некоторым сопротивлением. Поэтому колебания в реальном колебательном контуре будут затухающими. Так, спустя одно полное колебание заряд на конденсаторе окажется меньше исходного значения. Со временем колебания и вовсе исчезнут: вся энергия, запасённая изначально в контуре, выделится в виде тепла на сопротивлении катушки и соединительных проводов.

Точно так же будут затухающими колебания реального пружинного маятника: вся энергия маятника постепенно превратится в тепло из-за неизбежного наличия трения.

3.21.2 Энергетические превращения в колебательном контуре

Продолжаем рассматривать незатухающие колебания в контуре, считая сопротивление катушки нулевым. Конденсатор имеет ёмкость C, индуктивность катушки равна L.

Поскольку тепловых потерь нет, энергия из контура не уходит: она постоянно перераспределяется между конденсатором и катушкой.

Возьмём момент времени, когда заряд конденсатора максимален и равен q_0, а ток отсутствует. Энергия магнитного поля катушки в этот момент равна нулю. Вся энергия W контура сосредоточена в конденсаторе:

$$W = \frac{q_0^2}{2C}\,.$$

Теперь, наоборот, рассмотрим момент, когда ток максимален и равен I_0, а конденсатор разряжен. Энергия конденсатора равна нулю. Вся энергия контура запасена в катушке:

$$W = \frac{LI_0^2}{2}\,.$$

В произвольный момент времени, когда заряд конденсатора равен q и через катушку течёт ток I, энергия контура равна:

$$W = \frac{q^2}{2C} + \frac{LI^2}{2}\,.$$

Таким образом,

$$\frac{q^2}{2C} + \frac{LI^2}{2} = \frac{q_0^2}{2C} = \frac{LI_0^2}{2}\,. \tag{3.85}$$

Соотношение (3.85) применяется при решении многих задач.

3.21.3 Электромеханические аналогии

В предыдущем листке про самоиндукцию мы отметили аналогию между индуктивностью и массой. Теперь мы можем установить ещё несколько соответствий между электродинамическими и механическими величинами.

Для пружинного маятника мы имеем соотношение, аналогичное (3.85):

$$\frac{kx^2}{2} + \frac{mv^2}{2} = \frac{kx_0^2}{2} = \frac{mv_0^2}{2}\,. \tag{3.86}$$

Здесь, как вы уже поняли, k — жёсткость пружины, m — масса маятника, x и v — текущие значения координаты и скорости маятника, x_0 и v_0 — их наибольшие значения.

Сопоставляя друг с другом равенства (3.85) и (3.86), мы видим следующие соответствия:

$$q \longleftrightarrow x; \tag{3.87}$$

$$I \longleftrightarrow v; \tag{3.88}$$

$$L \longleftrightarrow m; \tag{3.89}$$

$$1/C \longleftrightarrow k. \tag{3.90}$$

Опираясь на эти электромеханические аналогии, мы можем предвидеть формулу для периода электромагнитных колебаний в колебательном контуре.

В самом деле, период колебаний пружинного маятника, как мы знаем, равен:

$$T = 2\pi\sqrt{\frac{m}{k}}\,.$$

В соответствии с аналогиями (3.89) и (3.90) заменяем здесь массу m на индуктивность L, а жёсткость k на обратную ёмкость $1/C$. Получим:

$$T = 2\pi\sqrt{LC}\,. \tag{3.91}$$

Электромеханические аналогии не подводят: формула (3.91) даёт верное выражение для периода колебаний в колебательном контуре. Она называется *формулой Томсона*. Мы вскоре приведём её более строгий вывод.

3.21.4 Гармонический закон колебаний в контуре

Напомним, что колебания называются *гармоническими*, если колеблющаяся величина меняется со временем по закону синуса или косинуса.

Колебания заряда на конденсаторе и силы тока в контуре оказываются гармоническими. Мы сейчас это докажем. Но прежде нам надо установить правила выбора знака для заряда конденсатора и для силы тока — ведь при колебаниях эти величины будут принимать как положительные, так и отрицательные значения.

Сначала мы выбираем *положительное направление обхода* контура. Выбор роли не играет; пусть это будет направление *против часовой стрелки* (рис. 3.109).

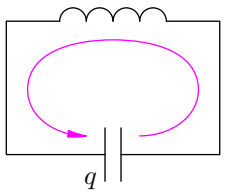


Рис. 3.109. Положительное направление обхода

Сила тока считается положительной ($I > 0$), если ток течёт в положительном направлении. В противном случае сила тока будет отрицательной ($I < 0$).

Заряд конденсатора q — это заряд той его пластины, *на которую* течёт положительный ток (т. е. той пластины, на которую указывает стрелка направления обхода). В данном случае q — заряд *левой* пластины конденсатора.

При таком выборе знаков тока и заряда справедливо соотношение: $\dot{q} = I$ (при ином выборе знаков могло случиться $\dot{q} = -I$). Действительно, знаки обеих частей совпадают: если $I > 0$, то заряд q левой пластины возрастает, и потому $\dot{q} > 0$.

Величины $q = q(t)$ и $I = I(t)$ меняются со временем, но энергия контура остаётся неизменной:

$$\frac{q^2}{2C} + \frac{LI^2}{2} = W = \text{const}. \tag{3.92}$$

Стало быть, производная энергии по времени обращается в нуль: $\dot{W} = 0$. Берём производную по времени от обеих частей соотношения (3.92); не забываем, что слева дифференцируются сложные функции[45]:

$$\frac{2q\dot{q}}{2C} + \frac{L \cdot 2I\dot{I}}{2} = \dot{W} = 0.$$

Подставляя сюда $\dot{q} = I$ и $\dot{I} = \ddot{q}$, получим:

$$\frac{qI}{C} + LI\ddot{q} = 0,$$
$$I\left(\frac{q}{C} + L\ddot{q}\right) = 0.$$

Но сила тока не является функцией, тождественно равной нулю; поэтому

$$\frac{q}{C} + L\ddot{q} = 0.$$

Перепишем это в виде:

$$\ddot{q} + \frac{1}{LC}\,q = 0. \tag{3.93}$$

Мы получили дифференциальное уравнение гармонических колебаний вида $\ddot{q} + \omega_0^2 q = 0$, где $\omega_0^2 = 1/LC$. Это доказывает, что заряд конденсатора колеблется по гармоническому закону (т. е. по закону синуса или косинуса). Циклическая частота этих колебаний равна:

$$\omega_0 = \frac{1}{\sqrt{LC}}\,. \tag{3.94}$$

Эта величина называется ещё *собственной частотой* контура; именно с этой частотой в контуре совершаются свободные (или, как ещё говорят, *собственные* колебания). Период колебаний равен:

$$T = \frac{2\pi}{\omega_0} = 2\pi\sqrt{LC}\,.$$

Мы снова пришли к формуле Томсона.

Гармоническая зависимость заряда от времени в общем случае имеет вид:

$$q = q_0 \cos(\omega_0 t + \alpha). \tag{3.95}$$

Циклическая частота ω_0 находится по формуле (3.94); амплитуда q_0 и начальная фаза α определяются из начальных условий.

Мы рассмотрим ситуацию, подробно изученную в начале этого листка. Пусть при $t = 0$ заряд конденсатора максимален и равен q_0 (как на рис. 3.100); ток в контуре отсутствует. Тогда начальная фаза $\alpha = 0$, так что заряд меняется по закону косинуса с амплитудой q_0:

$$q = q_0 \cos\omega_0 t = q_0 \cos\left(\frac{t}{\sqrt{LC}}\right). \tag{3.96}$$

[45] Если $y = y(x)$ — функция от x, то по правилу дифференцирования сложной функции производная от квадрата нашей функции будет равна: $(y^2)' = 2yy'$.

Найдём закон изменения силы тока. Для этого дифференцируем по времени соотношение (3.96), опять-таки не забывая о правиле нахождения производной сложной функции:

$$I = \dot{q} = -q_0\omega_0 \sin \omega_0 t.$$

Мы видим, что и сила тока меняется по гармоническому закону, на сей раз — по закону синуса:

$$I = -I_0 \sin \omega_0 t = -I_0 \sin\left(\frac{t}{\sqrt{LC}}\right). \tag{3.97}$$

Амплитуда силы тока равна:

$$I_0 = q_0\omega_0 = \frac{q_0}{\sqrt{LC}}.$$

Наличие «минуса» в законе изменения тока (3.97) понять не сложно. Возьмём, к примеру, интервал времени $0 < t < T/4$ (рис. 2).

Ток течёт в отрицательном направлении: $I < 0$. Поскольку $\omega_0 = 2\pi/T$, фаза колебаний находится в первой четверти: $0 < \omega_0 t < \pi/2$. Синус в первой четверти положителен; стало быть, синус в (3.97) будет положительным на рассматриваемом интервале времени. Поэтому для обеспечения отрицательности тока действительно необходим знак «минус» в формуле (3.97).

А теперь посмотрите на рис. 3.107. Ток течёт в положительном направлении. Как же работает наш «минус» в этом случае? Разберитесь-ка, в чём тут дело!

Изобразим графики колебаний заряда и тока, т. е. графики функций (3.96) и (3.97). Для наглядности представим эти графики в одних координатных осях (рис. 3.110).

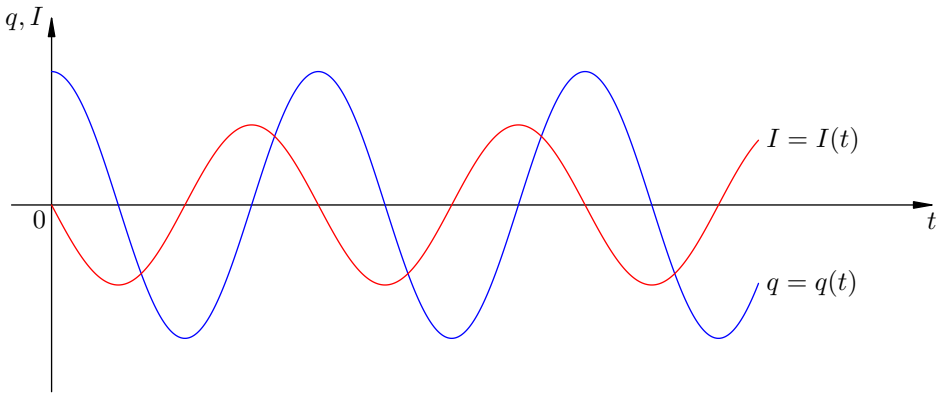


Рис. 3.110. Графики колебаний заряда и тока

Обратите внимание: нули заряда приходятся на максимумы или минимумы тока; и наоборот, нули тока соответствуют максимумам или минимумам заряда.

Используя формулу приведения

$$\cos\left(\varphi + \frac{\pi}{2}\right) = -\sin\varphi,$$

запишем закон изменения тока (3.97) в виде:

$$I = -I_0 \sin \omega_0 t = I_0 \cos\left(\omega_0 t + \frac{\pi}{2}\right).$$

Сопоставляя это выражение с законом изменения заряда $q = q_0 \cos \omega_0 t$, мы видим, что фаза тока, равная $\omega_0 t + \pi/2$, больше фазы заряда $\omega_0 t$ на величину $\pi/2$. В таком случае говорят, что ток *опережает по фазе* заряд на $\pi/2$; или *сдвиг фаз* между током и зарядом равен $\pi/2$; или *разность фаз* между током и зарядом равна $\pi/2$.

Опережение током заряда по фазе на $\pi/2$ графически проявляется в том, что график тока сдвинут *влево* на $\pi/2$ относительно графика заряда. Сила тока достигает, например, своего максимума на четверть периода раньше, чем достигает максимума заряд (а четверть периода как раз и соответствует разности фаз $\pi/2$).

3.21.5 Вынужденные электромагнитные колебания

Как вы помните, *вынужденные колебания* возникают в системе под действием периодической вынуждающей силы. Частота вынужденных колебаний совпадает с частотой вынуждающей силы.

Вынужденные электромагнитные колебания будут совершаться в контуре, поключённом к источнику синусоидального напряжения (рис. 3.111).

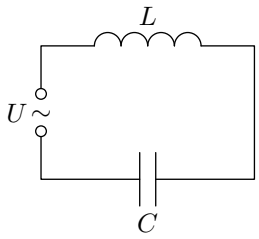


Рис. 3.111. Вынужденные колебания

Если напряжение источника меняется по закону:

$$U = U_0 \sin \omega t,$$

то в контуре происходят колебания заряда и тока с циклической частотой ω (и с периодом, соответственно, $T = 2\pi/\omega$). Источник переменного напряжения как бы «навязывает» контуру свою частоту колебаний, заставляя забыть о собственной частоте $\omega_0 = 1/\sqrt{LC}$.

Амплитуда вынужденных колебаний заряда и тока зависит от частоты ω: амплитуда тем больше, чем ближе ω к собственной частоте контура ω_0. При $\omega = \omega_0$ наступает *резонанс* — резкое возрастание амплитуды колебаний. Мы поговорим о резонансе более подробно в следующем листке, посвящённом переменному току.

3.22 Переменный ток. 1

Переменный ток — это вынужденные электромагнитные колебания, вызываемые в электрической цепи источником переменного (чаще всего синусоидального) напряжения.

Переменный ток присутствует всюду. Он течёт по проводам наших квартир, в промышленных электросетях, в высоковольтных линиях электропередач. И если вам нужен постоянный ток, чтобы зарядить аккумулятор телефона или ноутбука, вы используете специальный адаптер, выпрямляющий переменный ток из розетки.

Почему переменный ток распространён так широко? Оказывается, он прост в получении и идеально приспособлен для передачи электроэнергии на большие расстояния. Подробнее об этих вопросах мы поговорим позже — в разделе, посвящённом производству, передаче и потреблению электрической энергии.

А сейчас мы рассмотрим простейшие цепи переменного тока. Будем подключать к источнику синусоидального напряжения *поочерёдно*:

- резистор сопротивлением R;
- конденсатор ёмкости C;
- катушку индуктивности L.

Изучив поведение резистора, конденсатора и катушки в цепи переменного тока, мы затем подключим их одновременно и исследуем прохождение переменного тока через колебательный контур, обладающий сопротивлением.

Напряжение на клеммах источника меняется по закону:

$$U = U_0 \sin \omega t. \tag{3.98}$$

Как видим, напряжение может быть положительным и отрицательным. Каков смысл знака напряжения?

Всегда подразумевается, что выбрано положительное направление обхода контура. Напряжение считается положительным, если электрическое поле зарядов, образующих ток, имеет положительное направление. В противном случае напряжение считается отрицательным.

Начальная фаза напряжения не играет никакой роли, поскольку мы рассматриваем процессы, установившиеся во времени. При желании вместо синуса в выражении (3.98) можно было бы взять косинус — принципиально от этого ничего не изменится.

Текущее значение напряжения $U(t)$ в момент времени t называется *мгновенным значением напряжения*.

3.22.1 Условие квазистационарности

В случае переменного тока возникает один тонкий момент. Предположим, что цепь состоит из нескольких последовательно соединённых элементов.

Если напряжение источника меняется по синусоидальному закону, то сила тока не успевает мгновенно принимать одно и то же значение во всей цепи — на передачу взаимодействий между заряженными частицами вдоль цепи требуется некоторое время.

Между тем, как и в случае постоянного тока, нам хотелось бы считать силу тока одинаковой во всех элементах цепи. К счастью, во многих практически важных случаях мы действительно имеем на это право.

Возьмём, к примеру, переменное напряжение частоты $\nu = 50$ Гц (это промышленный стандарт России и многих других стран). Период колебаний напряжения: $T = 1/\nu = 0{,}02$ с.

Взаимодействие между зарядами передаётся со скоростью света: $c = 3 \cdot 10^8$ м/с. За время, равное периоду колебаний, это взаимодействие распространится на расстояние:

$$cT = 6 \cdot 10^6 \text{ м } = 6000 \text{ км}.$$

Поэтому в тех случаях, когда длина цепи на несколько порядков меньше данного расстояния, мы можем пренебречь временем распространения взаимодействия и считать, что сила тока мгновенно принимает одно и то же значение во всей цепи.

Теперь рассмотрим общий случай, когда напряжение колеблется с циклической частотой ω. Период колебаний равен $T = 2\pi/\omega$, и за это время взаимодействие между зарядами передаётся на расстояние cT. Пусть l — длина цепи. Мы можем пренебречь временем распространения взаимодействия, если l много меньше cT:

$$l \ll cT. \tag{3.99}$$

Неравенство (3.99) называется *условием квазистационарности*. При выполнении этого условия можно считать, что сила тока в цепи мгновенно принимает одно и то же значение во всей цепи. Такой ток называется *квазистационарным*.

В дальнейшем мы подразумеваем, что переменный ток меняется достаточно медленно и его можно считать квазистационарным. Поэтому сила тока I во всех последовательно включённых элементах цепи будет принимать одинаковое значение — своё в каждый момент времени. Оно называется *мгновенным значением силы тока*.

3.22.2 Резистор в цепи переменного тока

Простейшая цепь переменного тока получится, если к источнику переменного напряжения $U = U_0 \sin \omega t$ подключить обычный резистор[46] R, называемый также *активным сопротивлением* (рис. 3.112)

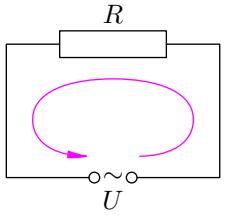


Рис. 3.112. Резистор в цепи переменного тока

Положительное направление обхода цепи выбираем против часовой стрелки, как показано на рисунке. Напомним, что сила тока считается положительной, если ток течёт в положительном направлении; в противном случае сила тока отрицательна.

Оказывается, мгновенные значения силы тока и напряжения связаны формулой, аналогичной закону Ома для постоянного тока:

$$I = \frac{U}{R} = \frac{U_0}{R} \sin \omega t.$$

Таким образом, сила тока в резисторе также меняется по закону синуса:

$$I = I_0 \sin \omega t.$$

[46] Мы полагаем, разумеется, что индуктивность этого резистора пренебрежимо мала, так что эффект самоиндукции можно не принимать во внимание.

Амплитуда тока I_0 равна отношению амплитуды напряжения U_0 к сопротивлению R:

$$I_0 = \frac{U_0}{R}\,.$$

Мы видим, что сила тока через резистор и напряжение на нём меняются «синхронно», точнее говоря — синфазно (рис. 3.113).

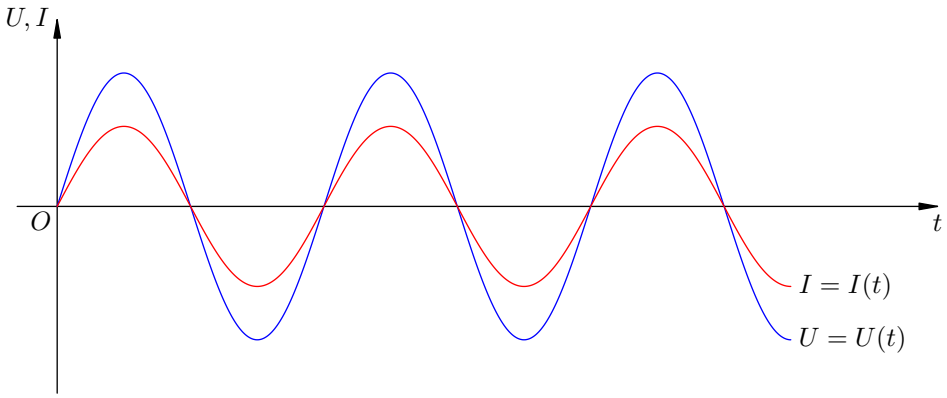


Рис. 3.113. Ток через резистор совпадает по фазе с напряжением

Фаза тока равна фазе напряжения, то есть сдвиг фаз между током и напряжением равен нулю.

3.22.3 Конденсатор в цепи переменного тока

Постоянный ток через конденсатор не течёт — для постоянного тока конденсатор является разрывом цепи. Однако переменному току конденсатор не помеха! Протекание переменного тока через конденсатор обеспечивается периодическим изменением заряда на его пластинах.

Рассмотрим конденсатор ёмкости C, подключённый к источнику синусоидального напряжения (рис. 3.114). Активное сопротивление проводов, как всегда, считаем равным нулю. Положительное направление обхода цепи снова выбираем против часовой стрелки.

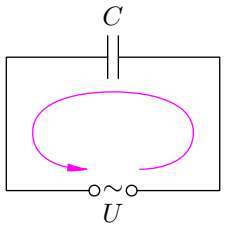


Рис. 3.114. Конденсатор в цепи переменного тока

Как и ранее, обозначим через q заряд той пластины конденсатора, на которую течёт положительный ток — в данном случае это будет правая пластина. Тогда знак величины q совпадает со знаком напряжения U. Кроме того, как мы помним из предыдущего листка, при таком согласовании знака заряда и направления тока будет выполнено равенство $\dot{q} = I$.

Напряжение на конденсаторе равно напряжению источника:

$$\frac{q}{C} = U = U_0 \sin \omega t.$$

Отсюда

$$q = CU_0 \sin \omega t.$$

Дифференцируя это равенство по времени, находим силу тока через конденсатор:

$$I = \dot{q} = CU_0\omega \cos \omega t. \tag{3.100}$$

Графики тока и напряжения представлены на рис. 3.115. Мы видим, что сила тока каждый раз достигает максимума на четверть периода раньше, чем напряжение. Это означает, что фаза силы тока на $\pi/2$ больше фазы напряжения (ток опережает по фазе напряжение на $\pi/2$).

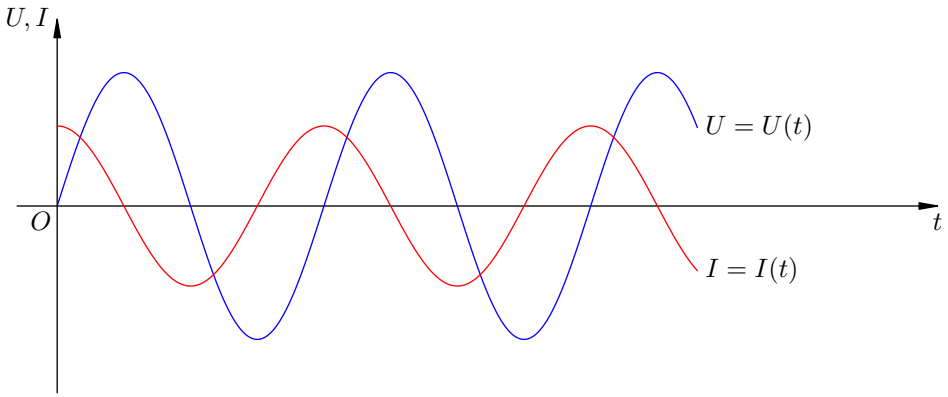


Рис. 3.115. Ток через конденсатор опережает по фазе напряжение на $\pi/2$

Найти сдвиг фаз между током и напряжением можно также с помощью формулы приведения:

$$\cos \varphi = \sin \left(\varphi + \frac{\pi}{2} \right).$$

Используя её, получим из (3.100):

$$I = CU_0\omega \sin \left(\omega t + \frac{\pi}{2} \right).$$

И теперь мы чётко видим, что фаза тока больше фазы напряжения на $\pi/2$.

Для амплитуды силы тока имеем:

$$I_0 = CU_0\omega = \frac{U_0}{1/(\omega C)}.$$

Таким образом, амплитуда силы тока связана с амплитудой напряжения соотношением, аналогичным закону Ома:

$$I_0 = \frac{U_0}{X_C},$$

где

$$X_C = \frac{1}{\omega C}.$$

Величина X_C называется *ёмкостным сопротивлением* конденсатора. Чем больше ёмкостное сопротивление конденсатора, тем меньше амплитуда тока, протекающего через него, и наоборот.

Ёмкостное сопротивление обратно пропорционально циклической частоте колебаний напряжения (тока) и ёмкости конденсатора. Попробуем понять физическую причину такой зависимости.

1. Чем больше частота колебаний (при фиксированной ёмкости C), тем за меньшее время по цепи проходит заряд CU_0; тем больше амплитуда силы тока и тем меньше ёмкостное сопротивление. При $\omega \to \infty$ ёмкостное сопротивление стремится к нулю: $X_C \to 0$.

Это означает, что для тока высокой частоты конденсатор фактически является коротким замыканием цепи.

Наоборот, при уменьшении частоты ёмкостное сопротивление увеличивается, и при $\omega \to 0$ имеем $X_C \to \infty$. Это не удивительно: случай $\omega = 0$ отвечает постоянному току, а конденсатор для постоянного тока представляет собой бесконечное сопротивление (разрыв цепи).

2. Чем больше ёмкость конденсатора (при фиксированной частоте), тем больший заряд CU_0 проходит по цепи за то же время (за ту же четверть периода); тем больше амплитуда силы тока и тем меньше ёмкостное сопротивление.

Подчеркнём, что, в отличие от ситуации с резистором, *мгновенные* значения тока и напряжения в одни и те же моменты времени уже не будут удовлетворять соотношению, аналогичному закону Ома. Причина заключается в сдвиге фаз: напряжение меняется по закону синуса, а сила тока — по закону косинуса; эти функции не пропорциональны друг другу. Законом Ома связаны лишь *амплитудные* значения тока и напряжения.

3.22.4 Катушка в цепи переменного тока

Теперь подключим к нашему источнику переменного напряжения катушку индуктивности L (рис. 3.116). Активное сопротивление катушки считается равным нулю.

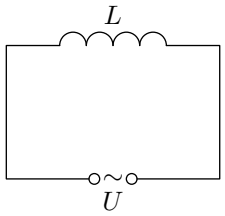


Рис. 3.116. Катушка в цепи переменного тока

Казалось бы, при нулевом активном (или, как ещё говорят, *омическом*) сопротивлении через катушку должен потечь бесконечный ток. Однако катушка оказывает переменному току сопротивление иного рода. Магнитное поле тока, меняющееся во времени, порождает в катушке вихревое электрическое поле $\vec{E}_{\text{вихр}}$, которое, оказывается, в точности уравновешивает кулоновское поле $\vec{E}$ движущихся зарядов:

$$\vec{E} + \vec{E}_{\text{вихр}} = \vec{0}. \tag{3.101}$$

Работа кулоновского поля $\vec{E}$ по перемещению единичного положительного заряда по внешней цепи в положительном направлении — это как раз напряжение U. Аналогичная работа вихревого поля — это ЭДС индукции $\mathcal{E}_i$. Поэтому из (3.101) получаем:

$$U + \mathcal{E}_i = 0. \tag{3.102}$$

Равенство (3.102) можно объяснить и с энергетической точки зрения. Допустим, что оно не выполняется. Тогда при перемещении заряда по цепи совершается ненулевая работа, которая должна превращаться в тепло. Но тепловая мощность I^2R равна нулю при нулевом омическом сопротивлении цепи. Возникшее противоречие показывает, что равенство (3.102) обязано выполняться.

Вспоминая закон Фарадея $\mathcal{E}_i = -L\dot{I}$, переписываем соотношение (3.102):

$$U - L\dot{I} = 0,$$

откуда

$$\dot{I} = \frac{U}{L} = \frac{U_0}{L}\sin\omega t. \tag{3.103}$$

Остаётся выяснить, какую функцию, меняющуюся по гармоническому закону, надо продифференцировать, чтобы получить правую часть выражения (3.103). Сообразить это нетрудно (продифференцируйте и проверьте!):

$$I = -\frac{U_0}{\omega L}\cos\omega t. \tag{3.104}$$

Мы получили выражение для силы тока через катушку. Графики тока и напряжения представлены на рис. 3.117.

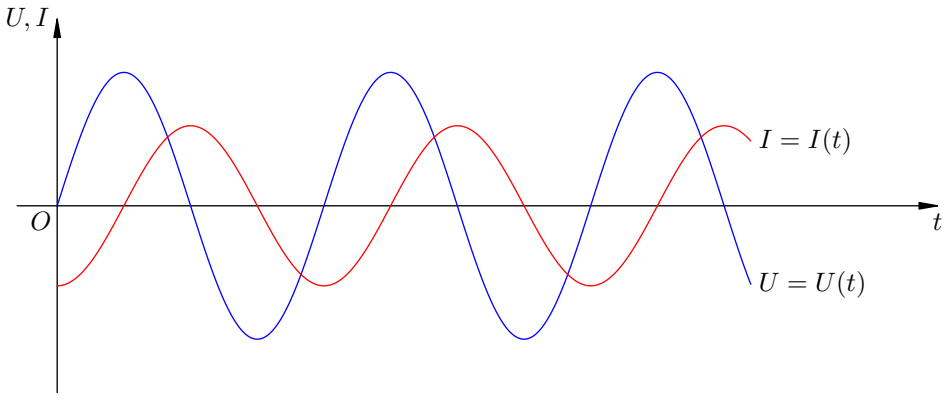


Рис. 3.117. Ток через катушку отстаёт по фазе от напряжения на $\pi/2$

Как видим, сила тока достигает каждого своего максимума на четверть периода позже, чем напряжение. Это означает, что сила тока отстаёт по фазе от напряжения на $\pi/2$.

Определить сдвиг фаз можно и с помощью формулы приведения:

$$\sin\left(\varphi - \frac{\pi}{2}\right) = -\cos\varphi.$$

Получаем:

$$I = \frac{U_0}{\omega L}\sin\left(\omega t - \frac{\pi}{2}\right).$$

Непосредственно видим, что фаза силы тока меньше фазы напряжения на $\pi/2$.

Амплитуда силы тока через катушку равна:

$$I_0 = \frac{U_0}{\omega L}.$$

Это можно записать в виде, аналогичном закону Ома:

$$I_0 = \frac{U_0}{X_L},$$

где

$$X_L = \omega L.$$

Величина X_L называется *индуктивным сопротивлением* катушки. Это и есть то самое сопротивление, которое наша катушка оказывает переменному току (при нулевом омическом сопротивлении).

Индуктивное сопротивление катушки пропорционально её индуктивности и частоте колебаний. Обсудим физический смысл этой зависимости.

1. Чем больше индуктивность катушки, тем большая в ней возникает ЭДС индукции, противодействующая нарастанию тока; тем меньшего амплитудного значения достигнет сила тока. Это и означает, что X_L будет больше.

2. Чем больше частота, тем быстрее меняется ток, тем больше скорость изменения магнитного поля в катушке, и тем большая возникает в ней ЭДС индукции, препятствующая возрастанию тока. При $\omega \to \infty$ имеем $X_L \to \infty$, т. е. высокочастотный ток практически не проходит через катушку.

   Наоборот, при $\omega = 0$ имеем $X_L = 0$. Для постоянного тока катушка является коротким замыканием цепи.

И снова мы видим, что закону Ома подчиняются лишь амплитудные, но не мгновенные значения тока и напряжения. Причина та же — наличие сдвига фаз.

Итак, мы разобрались с прохождением переменного тока через резистор, конденсатор и катушку по отдельности. Теперь нам предстоит соединить их вместе — в колебательный контур, подключённый к источнику переменного напряжения.

3.23 Переменный ток. 2

Давайте начнём с одного математического приёма, чтобы не отвлекаться потом на его объяснение. Это — тригонометрический метод введения вспомогательного угла. Он наверняка вам известен, но всё же повторить его не помешает.

3.23.1 Метод вспомогательного угла

Речь идёт о преобразовании выражения $a\sin\varphi + b\cos\varphi$. Вынесем за скобки «амплитудный множитель» $\sqrt{a^2+b^2}$:

$$a\sin\varphi + b\cos\varphi = \sqrt{a^2+b^2}\left(\frac{a}{\sqrt{a^2+b^2}}\sin\varphi + \frac{b}{\sqrt{a^2+b^2}}\cos\varphi\right).$$

Зачем нужно такое вынесение за скобки? Оказывается, в скобках при синусе и косинусе образовались замечательные множители! Сумма квадратов этих множителей равна единице:

$$\left(\frac{a}{\sqrt{a^2+b^2}}\right)^2 + \left(\frac{b}{\sqrt{a^2+b^2}}\right)^2 = 1.$$

Значит, эти множители являются соответственно косинусом и синусом некоторого угла α:

$$\frac{a}{\sqrt{a^2+b^2}} = \cos\alpha, \quad \frac{b}{\sqrt{a^2+b^2}} = \sin\alpha. \tag{3.105}$$

В результате получаем:

$$a\sin\varphi + b\cos\varphi = \sqrt{a^2+b^2}\,(\cos\alpha\sin\varphi + \sin\alpha\cos\varphi).$$

Остаётся заметить, что в скобках стоит синус суммы, так что мы приходим к окончательному выражению:

$$a\sin\varphi + b\cos\varphi = \sqrt{a^2+b^2}\sin(\varphi+\alpha). \tag{3.106}$$

При этом для «начальной фазы» α имеем из (3.105) простую формулу:

$$\operatorname{tg}\alpha = \frac{b}{a}. \tag{3.107}$$

3.23.2 Колебательный контур с резистором

Теперь мы готовы рассмотреть вынужденные колебания, происходящие в колебательном контуре с активным сопротивлением. К источнику переменного напряжения U последовательно подключены: резистор сопротивлением R, катушка индуктивности L и конденсатор ёмкости C (рис. 3.118; такой контур называется ещё *RLC-контуром*).

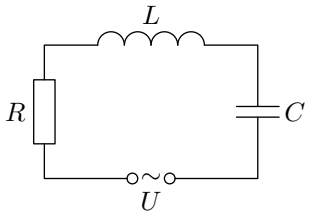


Рис. 3.118. Колебательный контур с резистором

Так как элементы соединены последовательно, сила тока в них одинакова в любой момент времени (вспомните условие квазистационарности!). Поэтому нам будет удобно начать не с

напряжения источника, как раньше, а с силы тока, и считать, что ток в цепи колеблется по закону синуса: $I = I_0 \sin \omega t$.

А теперь вспоминаем материал предыдущего раздела.

1. Пусть U_R — мгновенное значение напряжения на резисторе. Оно связано с силой тока обычным законом Ома:
$$U_R = IR = I_0 R \sin \omega t. \tag{3.108}$$

2. Напряжение на конденсаторе U_C отстаёт по фазе от тока на $\pi/2$; это значит, что фаза напряжения U_C равна $\omega t - \pi/2$. Амплитуда напряжения U_C равна:
$$U_{C0} = I_0 X_C = \frac{I_0}{\omega C}\,.$$
Таким образом,
$$U_C = U_{C0} \sin\left(\omega t - \frac{\pi}{2}\right) = -\frac{I_0}{\omega C} \cos \omega t. \tag{3.109}$$

3. Напряжение на катушке U_L, наоборот, опережает по фазе силу тока на $\pi/2$. Амплитуда:
$$U_{L0} = I_0 X_L = I_0 \omega L.$$
В результате получаем:
$$U_L = U_{L0} \sin\left(\omega t + \frac{\pi}{2}\right) = I_0 \omega L \cos \omega t. \tag{3.110}$$

Напряжение источника равно сумме напряжений на резисторе, катушке и конденсаторе:
$$U = U_R + U_L + U_C.$$
Подставляя сюда выражения (3.108)–(3.110), получим:
$$U = I_0 R \sin \omega t + I_0 \omega L \cos \omega t - \frac{I_0}{\omega C} \cos \omega t = I_0 \left(R \sin \omega t + \left(\omega L - \frac{1}{\omega C} \right) \cos \omega t \right). \tag{3.111}$$

Вот теперь нам и понадобится метод вспомогательного угла. Выражение во внешних скобках имеет для этого подходящий вид: $a \sin \omega t + b \cos \omega t$. Пользуясь выражениями (3.106) и (3.107), получим:
$$U = I_0 \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C} \right)^2} \sin(\omega t + \alpha), \tag{3.112}$$
где
$$\operatorname{tg} \alpha = \frac{\omega L - \frac{1}{\omega C}}{R}\,. \tag{3.113}$$

Угол α является сдвигом фаз между напряжением источника и силой тока в цепи: фаза напряжения больше фазы тока на величину α. Амплитуда напряжения:
$$U_0 = I_0 \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C} \right)^2}. \tag{3.114}$$

Получив все эти результаты, мы их несколько переиначим и приведём в соответствие с тем, что было в предыдущем разделе.

Начнём с напряжения источника. Предположим, как и ранее, что оно меняется по закону синуса:

$$U = U_0 \sin \omega t.$$

Как мы сейчас выяснили, фаза тока меньше фазы напряжения на величину α:

$$I = I_0 \sin(\omega t - \alpha).$$

При этом амплитуда силы тока находится из формулы (3.114):

$$I_0 = \frac{U_0}{\sqrt{R^2 + \left(\omega L - \dfrac{1}{\omega C}\right)^2}}. \tag{3.115}$$

Выражение (3.115) имеет вид закона Ома:

$$I_0 = \frac{U_0}{X},$$

где

$$X = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}. \tag{3.116}$$

Величина X — это *полное сопротивление* цепи. Такое сопротивление оказывает наш колебательный контур переменному току.

Закон Ома в данном случае выполнен лишь для *амплитудных* значений тока и напряжения. *Мгновенные* значения $I(t)$ и $U(t)$ уже не будут пропорциональны друг другу — ведь между ними имеется сдвиг фаз, равный α.

3.23.3 Резонанс в колебательном контуре

Как видно из выражения (3.115), амплитуда силы тока в контуре зависит от частоты колебаний. Построим график этой зависимости — так называемую *резонансную кривую* (рис. 3.119).

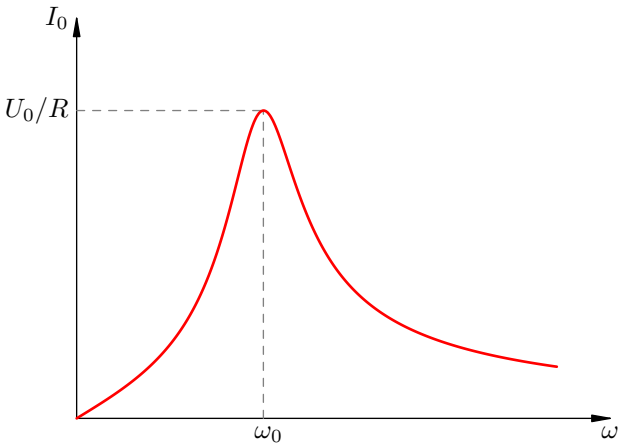


Рис. 3.119. Резонансная кривая

При $\omega \to 0$ имеем $I_0 \to 0$. Математическая причина стремления силы тока к нулю — неограниченное возрастание ёмкостного сопротивления $1/(\omega C)$, в результате чего полное сопротивление X также стремится к бесконечности. Физическая причина очевидна: ток малой частоты — это почти постоянный ток, а для постоянного тока конденсатор является разрывом цепи.

При $\omega \to \infty$ опять-таки имеем $I_0 \to 0$: график асимптотически приближается к оси ω. Теперь это происходит за счёт неограниченного роста индуктивного сопротивления ωL. Физическая причина также ясна: при быстром изменении тока в катушке возникает большая ЭДС самоиндукции, препятствующая его увеличению.

При некоторой частоте ω_0 амплитуда силы тока достигает максимума: наступает *резонанс*. Из (3.115) нетрудно видеть, что величина I_0 принимает максимальное значение

$$I_{0\max} = \frac{U_0}{R}\,, \tag{3.117}$$

и происходит это при выполнении равенства

$$\omega L - \frac{1}{\omega C} = 0.$$

Отсюда находим ω_0:

$$\omega_0 = \frac{1}{\sqrt{LC}}\,.$$

Это хорошо знакомая нам частота собственных колебаний в контуре с нулевым активным сопротивлением. Она же, как видим, является *резонансной частотой* нашего контура.

Из (3.117) мы видим, что резонансное значение амплитуды тока $I_{0\max}$ тем больше, чем меньше активное сопротивление R. На рис. 3.120 представлены три резонансные кривые. Верхняя кривая отвечает достаточно малому сопротивлению R, средняя кривая — большему сопротивлению, нижняя кривая — ещё большему сопротивлению.

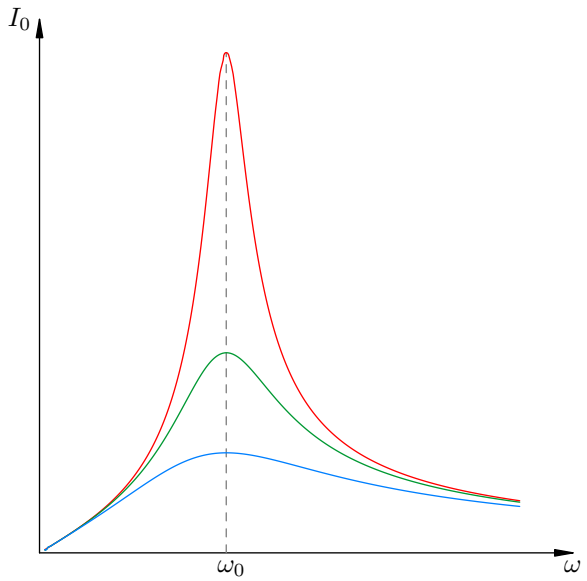


Рис. 3.120. Резонансные кривые при различных R

Таким образом, резонансный пик тем острее, чем меньше активное сопротивление контура. При весьма большом активном сопротивлении (как это видно из нижней резонансной кривой) понятие резонанса фактически утрачивает смысл.

При резонансе в контуре происходят любопытные вещи.

1. Амплитуды напряжений на конденсаторе и катушке равны друг другу. Действительно:

$$U_{C0} = I_0 \frac{1}{\omega_0 C} = \frac{U_0}{R}\sqrt{\frac{L}{C}}\,; \quad U_{L0} = I_0 \omega_0 L = \frac{U_0}{R}\sqrt{\frac{L}{C}}\,.$$

При малых значениях R эти амплитуды могут значительно превосходить амплитуду U_0 напряжения источника! Это, кстати, является наглядной демонстрацией одного важного факта:

Хотя сумма мгновенных значений напряжения на элементах контура равна мгновенному значению напряжению источника, сумма амплитуд напряжений на отдельных элементах может и не быть равной амплитуде напряжения источника.

2. Равен нулю сдвиг фаз между током в контуре и напряжением источника: $\alpha = 0$. Математически мы это видим из соотношения (3.113): при $\omega = \omega_0$ получается $\operatorname{tg}\alpha = 0$.

   Физическую причину синфазности тока и напряжения понять также не сложно. Дело в том, что напряжения U_C и U_L на конденсаторе и катушке колеблются в противофазе (т. е. разность фаз между ними равна π), а их амплитуды при резонансе равны. Стало быть, они отличаются только знаком: $U_L = -U_C$, и в сумме дают нуль. Получается, что $U = U_R + U_L + U_C = U_R$ (словно бы в цепи имелся один только резистор), а колебания напряжения и тока на резисторе происходят синфазно.

Резонанс играет важнейшую роль в радиосвязи. Когда осуществляется приём радиосигнала, радиоволны различных частот возбуждают в контуре колебания. Но амплитуды колебаний будут малы для сигналов тех радиостанций, частоты которых отличаются от собственной частоты контура. Контур выделяет лишь ту радиоволну, частота которой равна его собственной частоте; именно эти колебания будут иметь значительную амплитуду.

Поэтому, когда мы настраиваем приёмник на какую-то радиостанцию, мы меняем собственную частоту контура (как правило, путём изменения ёмкости конденсатора), пока не наступит резонанс с искомой радиоволной.

3.24 Мощность переменного тока

Переменный ток несёт энергию. Поэтому крайне важным является вопрос о мощности в цепи переменного тока.

Пусть U и I — мгновенные значение напряжения и силы тока на данном участке цепи. Возьмём малый интервал времени dt — настолько малый, что напряжение и ток не успеют за это время сколько-нибудь измениться; иными словами, величины U и I можно считать постоянными в течение интервала dt.

Пусть за время dt через наш участок прошёл заряд $dq = Idt$ (в соответствии с правилом выбора знака для силы тока заряд dq считается положительным, если он переносится в положительном направлении, и отрицательным в противном случае). Электрическое поле движущихся зарядов совершило при этом работу

$$dA = Udq = UIdt.$$

Мощность тока P — это отношение работы электрического поля ко времени, за которое эта работа совершена:

$$P = \frac{dA}{dt} = UI. \tag{3.118}$$

Точно такую же формулу мы получили в своё время для постоянного тока. Но в данном случае мощность зависит от времени, совершая колебания вместе током и напряжением; поэтому величина (3.118) называется ещё *мгновенной мощностью*.

Из-за наличия сдвига фаз сила тока и напряжение на участке не обязаны совпадать по знаку (например, может случиться так, что напряжение положительно, а сила тока отрицательна, или наоборот). Соответственно, мощность может быть как положительной, так и отрицательной. Рассмотрим чуть подробнее оба этих случая.

1. *Мощность положительна:* $P > 0$. Напряжение и сила тока имеют одинаковые знаки. Это означает, что направление тока совпадает с направлением электрического поля зарядов, образующих ток. В таком случае *энергия участка возрастает: она поступает на данный участок из внешней цепи* (например, конденсатор заряжается).

2. *Мощность отрицательна:* $P < 0$. Напряжение и сила тока имеют разные знаки. Стало быть, ток течёт против поля движущихся зарядов, образующих этот самый ток.

   Как такое может случиться? Очень просто: электрическое поле, возникающее на участке, как бы «перевешивает» поле движущихся зарядов и «продавливает» ток против этого поля. В таком случае *энергия участка убывает: участок отдаёт энергию во внешнюю цепь* (например, конденсатор разряжается).

Если вы не вполне поняли, о чём только что шла речь, не переживайте — дальше будут конкретные примеры, на которых вы всё и увидите.

3.24.1 Мощность тока через резистор

Пусть переменный ток $I = I_0 \sin \omega t$ протекает через резистор сопротивлением R. Напряжение на резисторе, как нам известно, колеблется в фазе с током:

$$U = IR = I_0 R \sin \omega t = U_0 \sin \omega t.$$

Поэтому для мгновенной мощности получаем:

$$P = UI = U_0 I_0 \sin^2 \omega t,$$

или

$$P = P_0 \sin^2 \omega t. \tag{3.119}$$

График зависимости мощности (3.119) от времени представлен на рис. 3.121. Мы видим, что мощность всё время неотрицательна — резистор забирает энергию из цепи, но не возвращает её обратно в цепь.

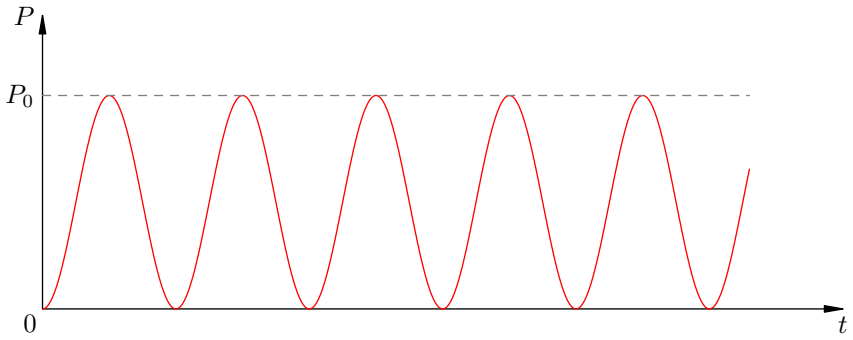


Рис. 3.121. Мощность переменного тока через резистор

Максимальное значение P_0 нашей мощности связано с амплитудами тока и напряжения привычными формулами:

$$P_0 = U_0 I_0 = I_0^2 R = \frac{U_0^2}{R}\,.$$

На практике, однако, интерес представляет не максимальная, а *средняя* мощность тока. Это и понятно. Возьмите, например, обычную лампочку, которая горит у вас дома. По ней течёт ток частотой 50 Гц, т. е. за секунду совершается 50 колебаний силы тока и напряжения. Ясно, что за достаточно продолжительное время на лампочке выделяется некоторая средняя мощность, значение которой находится где-то между 0 и P_0. Где же именно?

Посмотрите ещё раз внимательно на рис. 1. Не возникает ли у вас интуитивное ощущение, что средняя мощность соответствует «середине» нашей синусоиды и принимает поэтому значение $P_0/2$?

Это ощущение совершенно верное! Так оно и есть. Разумеется, можно дать математически строгое определение среднего значения функции (в виде некоторого интеграла) и подтвердить нашу догадку прямым вычислением, но нам это не нужно. Достаточно интуитивного понимания простого и важного факта:

среднее значение квадрата синуса (или косинуса) за период равно $1/2$.

Этот факт иллюстрируется рисунком 3.122.

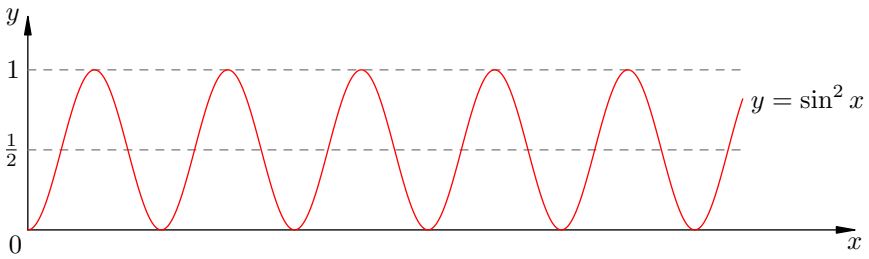


Рис. 3.122. Среднее значение квадрата синуса равно $1/2$

Итак, для среднего значения $\overline{P}$ мощности тока на резисторе имеем:

$$\overline{P} = \frac{P_0}{2} = \frac{U_0 I_0}{2} = \frac{I_0^2 R}{2} = \frac{U_0^2}{2R}\,. \tag{3.120}$$

В связи с этими формулами вводятся так называемые *действующие* (или *эффективные*) значения напряжения и силы тока[47]:

$$\overline{U} = \frac{U_0}{\sqrt{2}}, \quad \overline{I} = \frac{I_0}{\sqrt{2}}. \tag{3.121}$$

Формулы (3.120), записанные через действующие значения, полностью аналогичны соответствующим формулам для постоянного тока:

$$\overline{P} = \overline{U}\,\overline{I} = \overline{I}^2 R = \frac{\overline{U}^2}{R}.$$

Поэтому если вы возьмёте лампочку, подключите её сначала к источнику постоянного напряжения U, а затем к источнику переменного напряжения с таким же действующим значением U, то в обоих случаях лампочка будет гореть одинаково ярко.

Действующие значения (3.121) чрезвычайно важны для практики. Оказывается, *вольтметры и амперметры переменного тока показывают именно действующие значения* (так уж они устроены). Знайте также, что пресловутые 220 вольт из розетки — это *действующее* значение напряжения бытовой электросети.

3.24.2 Мощность тока через конденсатор

Пусть на конденсатор подано переменное напряжение $U = U_0 \sin \omega t$. Как мы знаем, ток через конденсатор опережает по фазе напряжение на $\pi/2$:

$$I = I_0 \sin\left(\omega t + \frac{\pi}{2}\right) = I_0 \cos \omega t.$$

Для мгновенной мощности получаем:

$$P = UI = U_0 I_0 \sin \omega t \cos \omega t = \frac{1}{2} U_0 I_0 \sin 2\omega t,$$

или

$$P = P_0 \sin 2\omega t. \tag{3.122}$$

Здесь введено обозначение $P_0 = U_0 I_0/2$. График зависимости (3.122) мгновенной мощности от времени представлен на рис. 3.123.

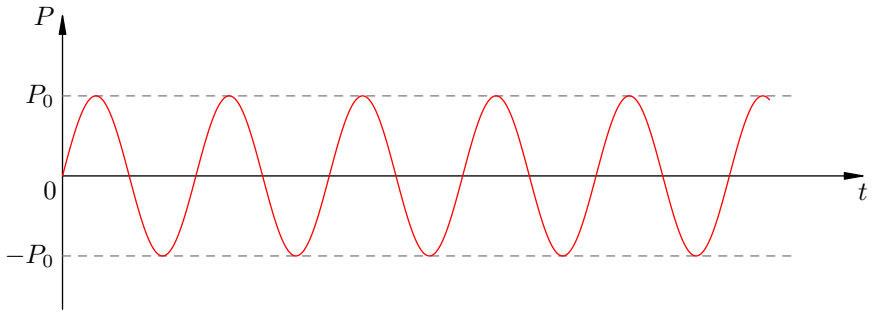


Рис. 3.123. Мощность переменного тока через конденсатор

Чему равно среднее значение мощности? Оно соответствует «середине» синусоиды и в данном случае равно нулю! Мы видим это сейчас как математический факт. Но интересно было

[47] На самом деле это есть не что иное, как *средние квадратические* значения напряжения и тока. Такое вам уже встречалось: средняя квадратическая скорость молекул идеального газа.

бы с физической точки зрения понять, почему мощность тока через конденсатор оказывается нулевой.

Для этого давайте нарисуем графики напряжения и силы тока в конденсаторе на протяжении одного периода колебаний (рис. 3.124).

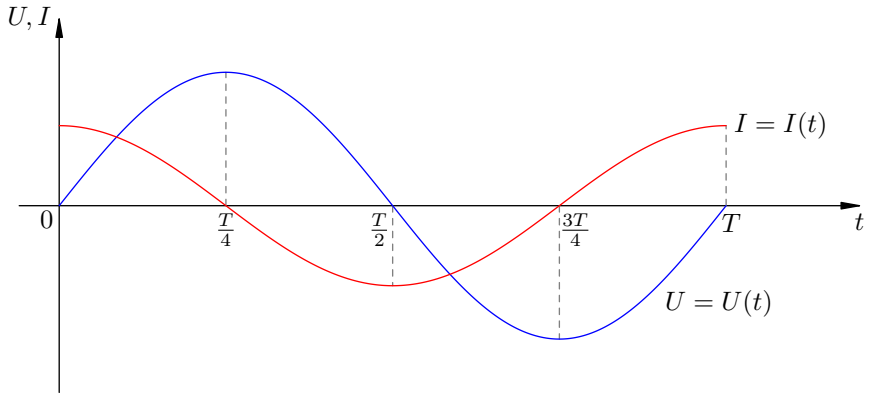


Рис. 3.124. Напряжение на конденсаторе и сила тока через него

Рассмотрим последовательно все четыре четверти периода.

1. *Первая четверть*, $0 < t < T/4$. Напряжение положительно и возрастает. Ток положителен (течёт в положительном направлении), конденсатор заряжается. По мере увеличения заряда на конденсаторе сила тока убывает.

   Мгновенная мощность положительна: конденсатор накапливает энергию, поступающую из внешней цепи. Эта энергия возникает за счёт работы внешнего электрического поля, продвигающего заряды на конденсатор.

2. *Вторая четверть*, $T/4 < t < T/2$. Напряжение продолжает оставаться положительным, но идёт на убыль. Ток меняет направление и становится отрицательным: конденсатор разряжается против направления внешнего электрического поля. В конце второй четверти конденсатор полностью разряжен.

   Мгновенная мощность отрицательна: конденсатор отдаёт энергию. Эта энергия возвращается в цепь: она идёт на совершение работы против электрического поля внешней цепи (конденсатор как бы «продавливает» заряды в направлении, противоположном тому, в котором внешнее поле «хочет» их двигать).

3. *Третья четверть*, $T/2 < t < 3T/4$. Внешнее электрическое поле меняет направление: напряжение отрицательно и возрастает по модулю. Сила тока отрицательна: идёт зарядка конденсатора в отрицательном направлении.

   Ситуация полностью аналогична первой четверти, только знаки напряжения и тока — противоположные. Мощность положительна: конденсатор вновь накапливает энергию.

4. *Четвёртая четверть*, $3T/4 < t < T$. Напряжение отрицательно и убывает по модулю. Конденсатор разряжается против внешнего поля: сила тока положительна.

   Мощность отрицательна: конденсатор возвращает энергию в цепь. Ситуация аналогична второй четверти — опять-таки с заменой заменой знаков тока и напряжения на противоположные.

Мы видим, что энергия, забранная конденсатором из внешней цепи в ходе первой четверти периода колебаний, полностью возвращается в цепь в ходе второй четверти. Затем этот процесс повторяется вновь и вновь. Вот почему средняя мощность, потребляемая конденсатором, оказывается равной нулю.

3.24.3 Мощность тока через катушку

Пусть на катушку подано переменное напряжение $U = U_0 \sin \omega t$. Ток через катушку отстаёт по фазе от напряжения на $\pi/2$:

$$I = I_0 \sin\left(\omega t - \frac{\pi}{2}\right) = -I_0 \cos \omega t.$$

Для мгновенной мощности получаем:

$$P = UI = -U_0 I_0 \sin \omega t \cos \omega t = -\frac{1}{2} U_0 I_0 \sin 2\omega t = -P_0 \sin 2\omega t.$$

Снова средняя мощность оказывается равной нулю. Причины этого, в общем-то, те же, что и в случае с конденсатором. Рассмотрим графики напряжения и силы тока через катушку за период (рис. 3.125).

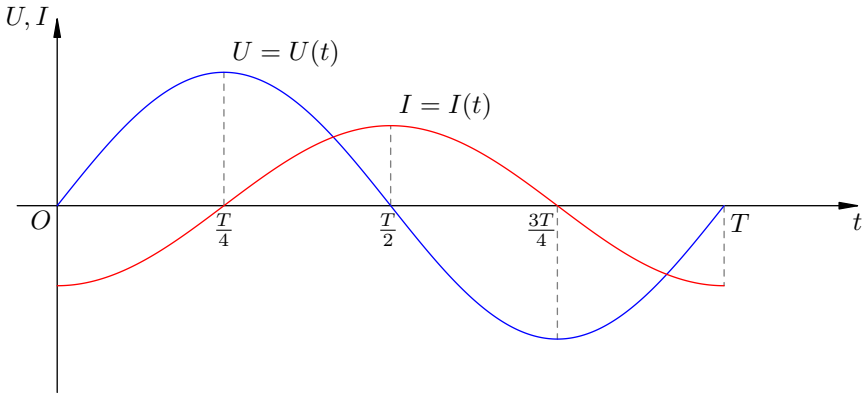


Рис. 3.125. Напряжение на катушке и сила тока через неё

Мы видим, что в течение второй и четвёртой четвертей периода энергия поступает в катушку из внешней цепи. В самом деле, напряжение и сила тока имеют одинаковые знаки, сила тока возрастает по модулю; для создания тока внешнее электрическое поле совершает работу против вихревого электрического поля, и эта работа идёт на увеличение энергии магнитного поля катушки.

В первой и третьей четвертях периода напряжение и сила тока имеют разные знаки: катушка возвращает энергию в цепь. Вихревое электрическое поле, поддерживающее убывающий ток, двигает заряды против внешнего электрического поля и совершает тем самым положительную работу. А за счёт чего совершается эта работа? За счёт энергии, накопленной ранее в катушке.

Таким образом, энергия, запасаемая в катушке за одну четверть периода, полностью возвращается в цепь в ходе следующей четверти. Поэтому средняя мощность, потребляемая катушкой, оказывается равной нулю.

3.24.4 Мощность тока на произвольном участке

Теперь рассмотрим самый общий случай. Пусть имеется произвольный участок цепи — он может содержать резисторы, конденсаторы, катушки... На этот участок подано переменное напряжение $U = U_0 \sin \omega t$.

Как мы знаем из предыдущего раздела «Переменный ток. 2», между напряжением и силой тока на данном участке имеется некоторый сдвиг фаз α. Мы записывали это так:

$$I = I_0 \sin(\omega t - \alpha).$$

Тогда для мгновенной мощности имеем:

$$P = U_0 I_0 \sin \omega t \sin(\omega t - \alpha). \tag{3.123}$$

Теперь нам хотелось бы определить, чему равна средняя мощность. Для этого мы преобразуем выражение (3.123), используя формулу:

$$\sin x \sin y = \frac{1}{2}(\cos(x - y) - \cos(x + y)).$$

В результате получим:

$$P = \frac{1}{2} U_0 I_0 (\cos \alpha - \cos(2\omega t - \alpha)). \tag{3.124}$$

Но среднее значение величины $\cos(2\omega t - \alpha)$ равно нулю! Поэтому средняя мощность оказывается равной:

$$\overline{P} = \frac{1}{2} U_0 I_0 \cos \alpha. \tag{3.125}$$

Данную формулу можно записать с помощью действующих значений (3.121) напряжения и силы тока:

$$\overline{P} = \overline{U}\,\overline{I} \cos \alpha.$$

Формула (3.125) охватывает все три рассмотренные выше ситуации. В случае резистора имеем $\alpha = 0$, и мы приходим к формуле (3.120). Для конденсатора и катушки $\alpha = \pi/2$, и средняя мощность равна нулю.

Кроме того, формула (3.125) даёт представление о весьма общей проблеме, связанной с передачей электроэнергии. Чрезвычайно важно, чтобы $\cos \alpha$ у потребителя был как можно ближе к единице. Иначе потребитель начнёт возвращать значительную часть энергии назад в сеть (что ему совсем невыгодно), и к тому же возвращаемая энергия будет безвозвратно расходоваться на нагревание проводов и других элементов цепи.

С этой проблемой приходится сталкиваться разработчикам электрических схем, содержащих электродвигатели. Обмотки электродвигателей обладают большими индуктивностями, и возникает ситуация, близкая к «чистой» катушке. Чтобы избежать бесполезного циркулирования энергии по сети, в цепь включают дополнительные элементы, сдвигающие фазу — например, так называемые компенсирующие конденсаторы.

3.25 Электроэнергия

Электрическая энергия играет в нашей жизни исключительную роль. Если в доме нет света, мы оказываемся практически беспомощны. Функционирование предприятий, средств транспорта, коммуникаций и прочих достижений цивилизации основано на использовании электроэнергии.

Электроэнергия обладает замечательными свойствами, которые и обеспечивают возможность её повсеместного применения.

- *Простота производства.* В мире функционирует огромное множество разнообразных генераторов электроэнергии.
- *Передача на большие расстояния.* Электроэнергия транспортируется по высоковольтным линиям электропередачи без существенных потерь.
- *Преобразование в другие виды энергии.* Электроэнергия легко преобразуется в механическую энергию (электродвигатели), внутреннюю энергию (нагревательные приборы), энергию света (осветительные приборы) и т. д.
- *Распределение между потребителями.* Специальные устройства позволяют распределять электроэнергию между потребителями с самыми разными «запросами» — промышленными предприятиями, городскими электросетями, жилыми домами и т. д.

Рассмотрим подробнее вопросы, связанные с производством, передачей и потреблением электрической энергии.

3.25.1 Производство электроэнергии

Среди генераторов электроэнергии наиболее распространены *электромеханические генераторы переменного тока*. Они преобразуют механическую энергию вращения ротора в энергию индукционного переменного тока, возникающего благодаря явлению электромагнитной индукции.

На рис. 3.126 проиллюстрирована основная идея генератора переменного тока: проводящая рамка (называемая *якорем*) вращается в магнитном поле.

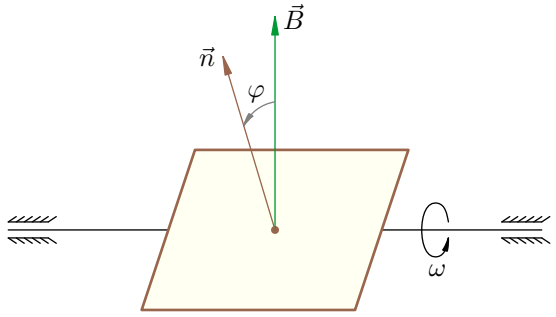


Рис. 3.126. Схема генератора переменного тока

Магнитный поток сквозь рамку меняется со временем и порождает ЭДС индукции, которая приводит к возникновению индукционного тока в рамке. С помощью специальных приспособлений (колец и щёток) переменный ток передаётся из рамки во внешнюю цепь.

Если рамка вращается в однородном магнитном поле B с постоянной угловой скоростью ω, то возникающий переменный ток будет синусоидальным. Давайте убедимся в этом.

Выберем направление вектора нормали $\vec{n}$ к плоскости рамки. Вектор $\vec{n}$, таким образом, вращается вместе с рамкой. Направление обхода рамки считается положительным, если с конца вектора $\vec{n}$ этот обход видится против часовой стрелки.

Напомним, что ток считается положительным, если он течёт в положительном направлении (и отрицательным в противном случае). ЭДС индукции считается положительной, если она создаёт ток в положительном направлении (и отрицательной в противном случае).

Предположим, что в начальный момент времени векторы $\vec{n}$ и $\vec{B}$ сонаправлены. За время t рамка повернётся на угол $\varphi = \omega t$. Магнитный поток через рамку в момент времени t равен:

$$\Phi = BS\cos\varphi = BS\cos\omega t, \tag{3.126}$$

где S — площадь рамки. Дифференцируя по времени, находим ЭДС индукции:

$$e = -\dot{\Phi} = BS\omega\sin\omega t. \tag{3.127}$$

Если сопротивление рамки равно R, то в ней возникает ток:

$$i = \frac{e}{R} = \frac{BS\omega}{R}\sin\omega t. \tag{3.128}$$

Как видим, ток действительно меняется по гармоническому закону, то есть является синусоидальным.

В реальных генераторах переменного тока рамка содержит не один виток, как в нашей схеме, а большое число N витков. Это позволяет увеличить в N раз ЭДС индукции в рамке. Почему?

Объяснить это несложно. В самом деле, магнитный поток через каждый виток площади S по-прежнему определяется выражением (3.126), так что ЭДС индукции в одном витке согласно формуле (3.127) равна: $e_1 = BS\omega\sin\omega t$. Все эти ЭДС индукции, возникающие в каждом витке, складываются друг с другом, и суммарная ЭДС в рамке окажется равной:

$$e = Ne_1 = NBS\omega\sin\omega t.$$

Сила тока в рамке:

$$i = \frac{NBS\omega}{R}\sin\omega t,$$

где R есть по-прежнему сопротивление рамки.

Кроме того, рамку снабжают железным (или стальным) сердечником. Железо многократно усиливает магнитное поле внутри себя, и поэтому наличие сердечника позволяет увеличить магнитный поток сквозь рамку в сотни и даже тысячи раз. В результате, как следует из формул (3.127) и (3.128), ЭДС индукции и ток в рамке увеличатся во столько же раз.

3.25.2 Передача электроэнергии

Электроэнергия производится в основном на тепловых электростанциях (ТЭС), гидроэлектростанциях (ГЭС) и атомных электростанциях (АЭС).

Роторы генераторов ТЭС вращаются за счёт энергии сгорающего топлива (чаще всего этим топливом является уголь). Экономически целесообразным является строительство ТЭС вблизи крупных угольных месторождений.

Роторы генераторов ГЭС приводятся во вращение энергией падающей воды. Поэтому ГЭС строятся на реках.

В любом случае возникает проблема передачи выработанной электроэнергии потребителям, находящимся за много километров от электростанций.

Электроэнергия транспортируется по проводам. Потери энергии на нагревание проводов должны быть сведены к минимуму. Оказывается, для этого нужно *высокое напряжение* в линии электропередачи. Покажем это.

Рассмотрим двухпроводную линию электропередачи, связывающую источник переменного напряжения u с потребителем П (рис. 3.127).

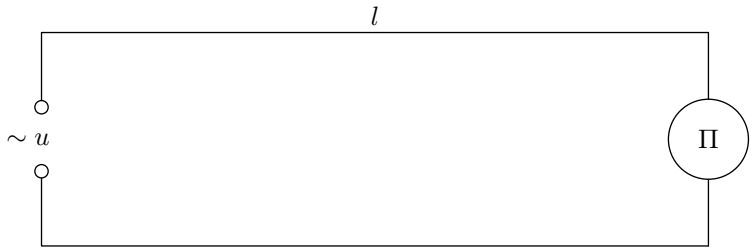


Рис. 3.127. Передача электроэнергии по двухпроводной линии

Длина линии равна l, так что общая длина проводов составит $2l$. Если ρ — удельное сопротивление материала провода, S — площадь поперечного сечения провода, то сопротивление линии будет равно:

$$R = \frac{2\rho l}{S}\,. \tag{3.129}$$

Потребителю должна быть передана мощность с заданным действующим значением P. Обозначим через U и I действующие значения напряжения в линии и силы тока. Если α — сдвиг фаз между током и напряжением, то, как мы знаем из предыдущего листка, $P = UI\cos\alpha$. Отсюда

$$I = \frac{P}{U\cos\alpha}\,. \tag{3.130}$$

Часть мощности ΔP теряется на нагревание проводов:

$$\Delta P = I^2 R.$$

Подставляя сюда выражения (3.129) и (3.130), получим:

$$\Delta P = \frac{2\rho l P^2}{SU^2\cos^2\alpha}\,. \tag{3.131}$$

Мы видим из формулы (3.131), что потеря мощности обратно пропорциональна квадрату напряжения в линии. Следовательно, для уменьшения потерь надо повышать напряжение при передаче. Вот почему линии электропередач являются *высоковольтными*. Например, Волжская ГЭС передаёт в Москву электроэнергию при напряжении 500 киловольт.

3.25.3 Трансформатор

Генераторы электростанций имеют ЭДС порядка 10–20 кВ. Как мы только что видели, для передачи электроэнергии на большие расстояния нужно повышать напряжение до нескольких сотен киловольт.

С другой стороны, напряжение бытовой электросети составляет 220 В. Поэтому при доставке энергии обычному потребителю требуется понижение напряжения до сотен вольт.

Замечательно, что повышение и понижение напряжения в случае синусоидального переменного тока не представляет никаких сложностей. Для этого используются специальные устройства — *трансформаторы*.

Простейший трансформатор состоит из двух обмоток, навитых на один и тот же замкнутый стальной сердечник.

Первичная обмотка содержит N_1 витков; на неё подаётся входное напряжение u_1. Это напряжение как раз и требуется преобразовать — повысить или понизить.

Вторичная обмотка содержит N_2 витков. К ней подсоединяется *нагрузка*, условно обозначаемая резистором R. Это — потребитель, для работы которого нужно преобразованное напряжение u_2.

Схема такого трансформатора изображена на рис. 3.128.

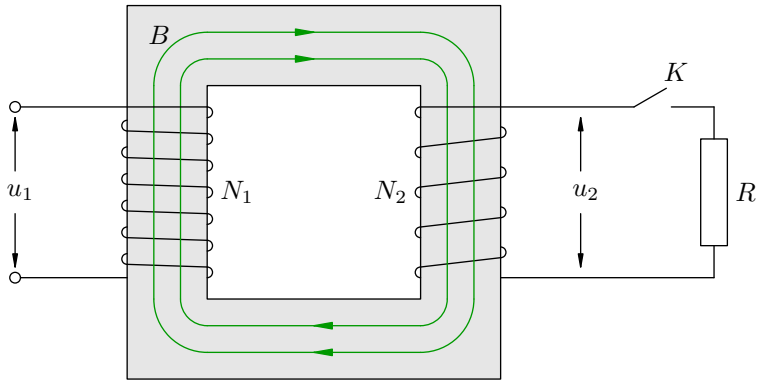


Рис. 3.128. Трансформатор

Режим холостого хода

Наиболее прост для рассмотрения *холостой ход* трансформатора, когда нагрузка отключена (ключ K разомкнут).

Пусть напряжение на первичной обмотке меняется по закону косинуса с амплитудой U_{01}:

$$u_1 = U_{01}\cos\omega t.$$

Активное сопротивление R_1 первичной обмотки считаем очень малым по сравнению с её индуктивным сопротивлением. В таком случае, как мы знаем, сила тока i_1 в первичной обмотке отстаёт по фазе от напряжения на $\pi/2$:

$$i_1 = I_{01}\cos\left(\omega t - \frac{\pi}{2}\right) = I_{01}\sin\omega t.$$

При этом трансформатор не потребляет энергию из сети, к которой он подключён.

Магнитный поток Ф, пронизывающий витки первичной обмотки, пропорционален току i_1 и поэтому также меняется по закону синуса:

$$\Phi = \Phi_0\sin\omega t.$$

В каждом витке первичной обмотки возникает ЭДС индукции:

$$e = -\dot{\Phi} = -\omega\Phi_0\cos\omega t. \tag{3.132}$$

Следовательно, полная ЭДС индукции в первичной обмотке равна:

$$e_1 = N_1 e = -N_1\omega\Phi_0\cos\omega t. \tag{3.133}$$

Стальной сердечник практически не выпускает магнитное поле наружу — линии магнитного поля почти целиком идут внутри сердечника. Магнитный поток в любом сечении сердечника

одинаков; в частности, каждый виток вторичной обмотки пронизывает тот же самый магнитный поток Ф. Поэтому в одном витке вторичной обмотки возникает та же ЭДС индукции e, даваемая выражением (3.132), а полная ЭДС индукции во вторичной обмотке равна:

$$e_2 = N_2 e = -N_2 \omega \Phi_0 \cos \omega t. \tag{3.134}$$

Как видим, обе ЭДС индукции в первичной и вторичной обмотках меняются синфазно. Мгновенные значения ЭДС индукции относятся друг к другу как числа витков в обмотках:

$$\frac{e_1}{e_2} = \frac{N_1}{N_2}. \tag{3.135}$$

Ввиду малости активного сопротивления первичной обмотки мы можем считать, что выполнено приближённое равенство:

$$u_1 + e_1 \approx 0 \tag{3.136}$$

(вспомните рассуждение из листка «Переменный ток. 1», раздел «Катушка в цепи переменного тока»). Так как цепь вторичной обмотки разомкнута и ток в ней отсутствует, имеем точное равенство:

$$u_2 + e_2 = 0.$$

Итак, $u_1 \approx -e_1$, $u_2 = -e_2$. Следовательно, мгновенные значения напряжений в первичной и вторичной обмотках также меняются почти синфазно. С учётом равенства (3.135) получаем:

$$\frac{u_1}{u_2} = \frac{N_1}{N_2}. \tag{3.137}$$

Величина $k = N_1/N_2$ называется *коэффициентом трансформации*. Отношение мгновенных значений напряжений в (3.137) можно заменить отношением действующих значений U_1 и U_2:

$$\frac{U_1}{U_2} = \frac{N_1}{N_2} = k.$$

Если $k > 1$, то трансформатор является *понижающим*. В этом случае вторичная обмотка содержит меньше витков, чем первичная; потребитель получает меньшее напряжение, чем то, что поступает на вход трансформатора. На рис. 3 изображён как раз понижающий трансформатор.

Если же $k < 1$, то трансформатор будет *повышающим*. Вторичная обмотка содержит больше витков, чем первичная, и потребитель получает напряжение более высокое, чем на входе трансформатора.

Режим нагрузки

Теперь рассмотрим вкратце работу *нагруженного* трансформатора. В этом случае ключ K на рис. 3.128 *замкнут*, и трансформатор выполняет свою прямую задачу — передаёт энергию потребителю, подключённому ко вторичной обмотке.

Согласно закону сохранения энергии, передача энергии потребителю возможна только за счёт увеличения потребления энергии из внешней сети. Так оно в действительности и происходит. Давайте попробуем понять, какие физические процессы приводят к этому.

Главное заключается в том, что ввиду малого омического сопротивления первичной обмотки сохраняется приближённое равенство (3.136), т. е.

$$e_1 \approx -u_1.$$

Напряжение u_1 задаётся внешней сетью, поэтому амплитуда ЭДС индукции e_1 остаётся прежней — равной амплитуде внешнего напряжения.

Но, с другой стороны, из выражения (3.133) мы знаем, что амплитуда величины e_1 равна $N_1\omega\Phi_0$. Стало быть, при подключении нагрузки остаётся неизменной амплитуда Φ_0 магнитного потока Φ, пронизывающего витки первичной и вторичной обмоток.

При холостом ходе магнитный поток Φ порождался магнитным полем тока i_1 первичной обмотки (во вторичной обмотке тока не было). Теперь в создании магнитного потока участвуют два магнитных поля: поле B_1 тока i_1 первичной обмотки (оно создаёт поток Φ_1) и поле B_2 тока i_2 вторичной обмотки (оно создаёт поток Φ_2). Таким образом,

$$\Phi = \Phi_1 + \Phi_2.$$

В отличие от тока i_1, который «навязывается» первичной обмотке внешней сетью, ток i_2 — индукционный, и его направление определяется правилом Ленца: магнитное поле B_2 стремится уменьшить изменение суммарного магнитного потока Φ. Но амплитуда Φ_0 этого потока, как мы уже говорили, остаётся той же, что и при холостом ходе. Как же так?

Очень просто — чтобы обеспечить неизменность величины Φ_0, приходится увеличиваться магнитному потоку Φ_1. Возрастает амплитуда тока i_1 первичной обмотки! Вот почему увеличивается потребление энергии из сети по сравнению с режимом холостого хода.

Первичная обмотка потребляет из сети мощность

$$P_1 = U_1 I_1 \cos\alpha_1$$

(как и выше, в данной формуле фигурируют действующие значения мощности, напряжения и силы тока).

Нагрузка получает от вторичной обмотки мощность

$$P_2 = U_2 I_2 \cos\alpha_2.$$

Эта мощность является *полезной* с точки зрения потребителя. Отношение полезной мощности, получаемой нагрузкой, к мощности, потребляемой из сети — это *КПД трансформатора*:

$$\eta = \frac{P_2}{P_1} = \frac{U_2 I_2 \cos\alpha_2}{U_1 I_1 \cos\alpha_1}.$$

Разумеется, $P_2 < P_1$ — часть мощности теряется в трансформаторе. Потери мощности состоят из двух частей.

1. Так называемые «потери в меди», обозначаемые $\Delta P_{\text{мед}}$. Это мощность, расходуемая на нагревание первичной и вторичной обмоток:

   $$\Delta P_{\text{мед}} = I_1^2 R_1 + I_2^2 R_2.$$

   Сколь бы малыми не были активные сопротивления R_1 и R_2 этих обмоток, они не равны нулю, и при больших токах с ними приходится считаться.

2. Так называемые «потери в стали», обозначаемые $\Delta P_{\text{ст}}$. Сюда относятся:

   - Мощность, расходуемая на перемагничивание сердечника, т. е. на изменение ориентации элементарных токов под действием внешнего магнитного поля.
   - Мощность, расходуемая на нагревание сердечника индукционными вихревыми токами (которые называются ещё *токами Фуко*). Эти токи возникают в сердечнике под действием вихревого электрического поля, порождаемого переменным магнитным полем. Для уменьшения токов Фуко сердечники собираются из листов специальной трансформаторной стали, но полностью ликвидировать эти токи, конечно же, не удаётся.

Оказывается, потери в стали не зависят от нагрузки — они определяются только амплитудой магнитного потока, которая, как мы знаем, при любой нагрузке остаётся неизменной.

Таким образом, имеем:

$$P_1 = P_2 + \Delta P_{\text{мед}} + \Delta P_{\text{ст}},$$

и для КПД трансформатора получаем следующее выражение:

$$\eta = \frac{P_2}{P_2 + \Delta P_{\text{мед}} + \Delta P_{\text{ст}}}\,. \tag{3.138}$$

Если полезная мощность P_2 мала (*недогрузка* трансформатора), то и КПД мал. Действительно, числитель в (3.138) маленький, а знаменатель — не меньше постоянной величины потерь в стали $\Delta P_{\text{ст}}$.

Если полезная мощность P_2 чрезмерно велика (*перегрузка* трансформатора), то КПД опять-таки мал. Дело в том, что в этом случае велики токи I_1 и I_2 в обмотках трансформатора, и, следовательно, большой величины достигают потери в меди $\Delta P_{\text{мед}}$.

Для трансформатора существует оптимальная (так называемая *номинальная*) нагрузка, на которую он рассчитан. При номинальной нагрузке оказывается, что КПД трансформатора близок к единице, т. е.

$$P_2 \approx P_1,$$

или, с учётом выражений для мощностей:

$$U_2 I_2 \cos\alpha_2 \approx U_1 I_1 \cos\alpha_1.$$

Кроме того, сдвиги фаз приближённо равны нулю, так что

$$U_2 I_2 \approx U_1 I_1.$$

Следовательно, при нагрузках, близких к номинальной, имеем:

$$\frac{I_2}{I_1} \approx \frac{U_1}{U_2} \approx k,$$

где k — введённый выше коэффициент трансформации. Например, у понижающего трансформатора $k > 1$, и при номинальной нагрузке ток в его вторичной обмотке в k раз больше тока первичной обмотки.

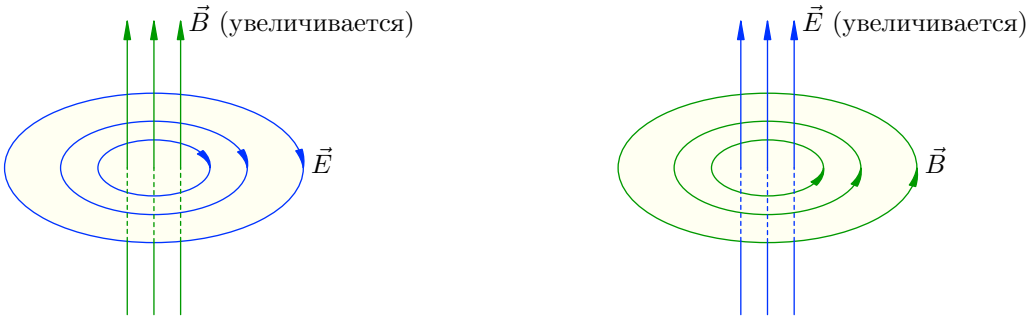
3.26 Электромагнитное поле

Вспомним, каким образом Максвелл объяснил явление электромагнитной индукции. **Переменное магнитное поле порождает вихревое электрическое поле**. Если в переменном магнитном поле находится замкнутый проводник, то вихревое электрическое поле приводит в движение заряженные частицы этого проводника — так возникает индукционный ток, наблюдаемый в эксперименте.

Линии вихревого электрического поля охватывают линии магнитного поля. Если смотреть с конца вектора $\vec{B}$, то линии вихревого электрического поля идут по часовой стрелке при возрастании магнитного поля и против часовой стрелки при убывании магнитного поля. Такое направление вихревого электрического поля, напомним, задаёт направление индукционного тока в соответствии с правилом Ленца.

3.26.1 Гипотеза Максвелла

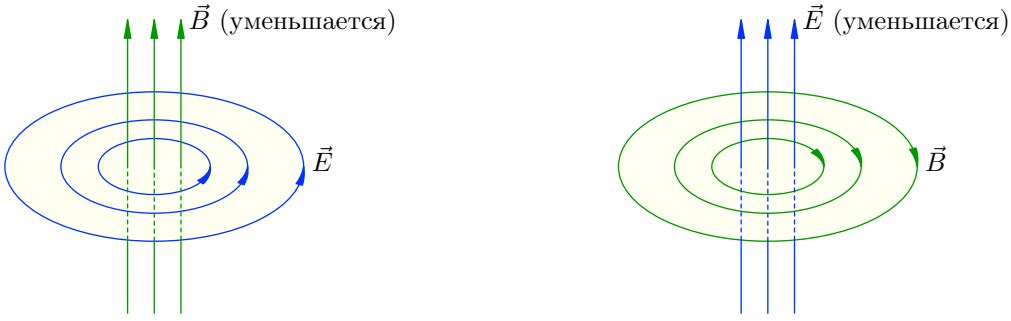
Описанным выше способом Максвелл объяснил, почему в экспериментах Фарадея появлялся индукционный ток. Но затем Максвелл пошёл ещё дальше и уже без какой-либо опоры на экспериментальные данные высказал симметричную гипотезу: **переменное электрическое поле порождает магнитное поле** (рис. 3.129, 3.130).



Объяснение электромагнитной индукции — Симметричная гипотеза Максвелла

Рис. 3.129. Симметричная гипотеза Максвелла (возрастание поля)

Линии этого магнитного поля охватывают линии переменного электрического поля и идут *в другую сторону* по сравнению с линиями вихревого электрического поля. Так, *при возрастании электрического поля линии порождаемого магнитного поля направлены против часовой стрелки, если смотреть с конца вектора* $\vec{E}$ (рис. 3.129, справа).



Объяснение электромагнитной индукции — Симметричная гипотеза Максвелла

Рис. 3.130. Симметричная гипотеза Максвелла (убывание поля)

Наоборот, *при убывании электрического поля линии порождаемого магнитного поля идут по часовой стрелке* (рис. 3.130, справа).

У электрического поля может быть два источника: электрические заряды и переменное магнитное поле. В первом случае линии электрического поля начинаются на положительных зарядах и оканчиваются на отрицательных. Во втором случае электрическое поле является вихревым — его линии оказываются замкнутыми.

У магнитного поля также может быть два источника: электрический ток и переменное электрическое поле. При этом линии магнитного поля замкнуты в обоих случаях (оно всегда вихревое).

Максвелл предположил, что оба источника магнитного поля равноправны в следующем смысле. Рассмотрим, например, процесс зарядки конденсатора (рис. 3.131):

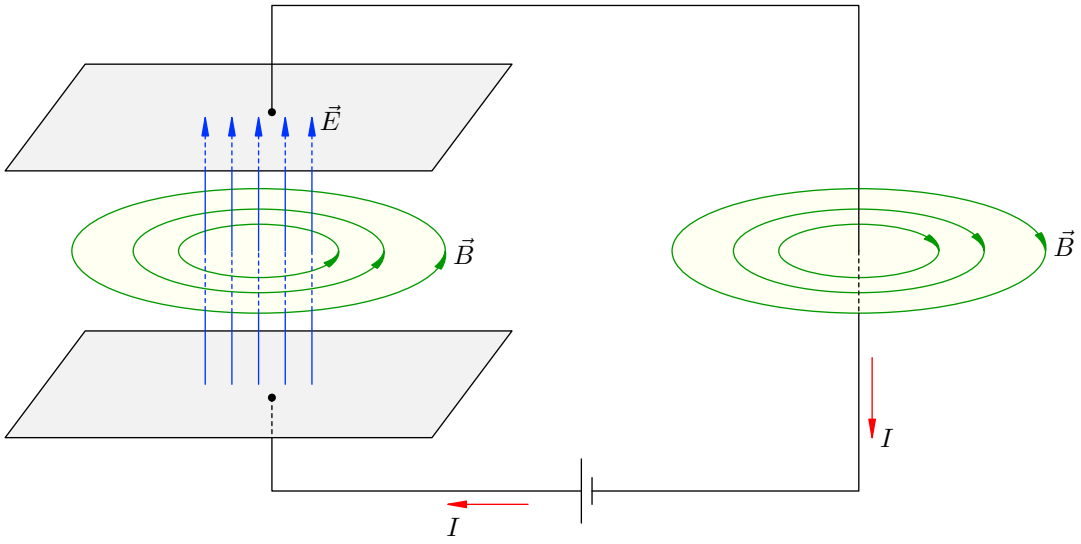


Рис. 3.131. Магнитное поле внутри конденсатора совпадает с магнитным полем тока

В данный момент по проводам, соединяющим обкладки конденсатора, течёт ток I. Заряд конденсатора увеличивается, и, соответственно, возрастает электрическое поле между обкладками. Переменное электрическое поле $\vec{E}$ порождает внутри конденсатора магнитное поле $\vec{B}$. Так вот, согласно гипотезе Максвелла *магнитное поле $\vec{B}$ внутри конденсатора оказывается точно таким же, как и магнитное поле тока I — как если бы ток I протекал в пространстве между обкладками конденсатора.*

Подчеркнём ещё раз, что симметричная гипотеза Максвелла была поначалу чисто умозрительной. На тот момент не наблюдалось каких-либо неясных физических явлений, для объяснения которых потребовалась бы такая гипотеза. Лишь впоследствии (и уже после смерти Максвелла) она получила блестящее экспериментальное подтверждение. Об этом — чуть ниже.

3.26.2 Понятие электромагнитного поля

Прежде всего, симметричная гипотеза указала на то, что электрическое и магнитное поля тесно взаимосвязаны. Они не являются обособленными физическими объектами и всегда существуют рядом друг с другом. *Если в какой-то системе отсчёта электрическое (магнитное) поле отсутствует, то в другой системе отсчёта, движущейся относительно первой, оно непременно появится.*

Допустим, например, что в движущемся автомобиле покоится электрический заряд. В системе отсчёта, связанной с автомобилем, этот заряд не создаёт магнитного поля. Но относительно земли заряд *движется*, а любой движущийся заряд является источником магнитного поля. Поэтому наблюдатель, стоящий на земле, зафиксирует магнитное поле, создаваемое зарядом в автомобиле.

Пусть также на земле лежит магнит. Наблюдатель, стоящий на земле, регистрирует постоянное магнитное поле, создаваемое этим магнитом; коль скоро это поле не меняется со временем, никакого электрического поля в земной системе отсчёта не возникает. Но относительно автомобиля магнит *движется* — приближается к автомобилю или удаляется от него. В системе отсчёта автомобиля магнитное поле меняется со временем — нарастает или убывает; наблюдатель в автомобиле фиксирует вихревое электрическое поле, порождаемое переменнным магнитным полем нашего магнита.

Но все инерциальные системы отсчёта абсолютно равноправны, среди них нет какой-то одной привилегированной. Законы природы выглядят одинаково в любой инерциальной системе отсчёта, и никакой физический эксперимент не может отличить одну инерциальную систему отсчёта от другой[48]. Поэтому естественно считать, что *электрическое поле и магнитное поле служат двумя различными проявлениями одного физического объекта — электромагнитного поля.*

Таким образом, в произвольной, наудачу выбранной системе отсчёта будут присутствовать обе компоненты электромагнитного поля — поле электрическое и поле магнитное. Но может случиться и так, что в некоторой системе отсчёта, специально приспособленной для данной задачи, одна из этих компонент обратится в нуль. Мы видели это в наших примерах с автомобилем.

Электромагнитное поле можно наблюдать и исследовать по его действию на заряженные частицы. Силовой характеристикой электромагнитного поля является пара векторов $\vec{E}$ и $\vec{B}$ — напряжённость электрического поля и индукция магнитного поля. Сила, с которой электромагнитное поле действует на заряд q, движущийся со скоростью $\vec{v}$, равна:

$$\vec{F} = \vec{F}_{\text{эл}} + \vec{F}_{\text{магн}}.$$

Силы в правой части нам хорошо известны. Сила $\vec{F}_{\text{эл}} = q\vec{E}$ действует со стороны электрического поля. Она не зависит от скорости заряда.

Сила $\vec{F}_{\text{магн}}$ действует со стороны магнитного поля. Её направление определяется по правилу часовой стрелки или левой руки, а модуль — по формуле $F_{\text{магн}} = qvB\sin\alpha$, где α — угол между векторами $\vec{v}$ и $\vec{B}$.

3.26.3 Об уравнениях Максвелла

Теория электромагнитного поля была создана Максвеллом. Он предложил свою знаменитую систему дифференциальных уравнений (*уравнений Максвелла*), которые позволяют найти векторы $\vec{E}$ и $\vec{B}$ в любой точке заданной области пространства по известным источникам — зарядам и токам[49]. Уравнения Максвелла легли в основу электродинамики и позволили объяснить все известные на тот момент явления электричества и магнетизма. Но мало того — уравнения Максвелла дали возможность предсказывать новые явления!

Так, среди решений уравнений Максвелла обнаружились поля с неизвестными ранее свойствами — электромагнитные волны. А именно, уравнения Максвелла допускали решения в виде электромагнитного поля, которое может распространяться в пространстве, захватывая с течением времени все новые и новые области. Скорость этого распространения конечна и зависит от среды, заполняющей пространство. Но электромагнитные волны не нуждаются ни в какой среде — они могут распространяться даже сквозь пустоту. Скорость распространения электромагнитных волн в вакууме совпадает со скоростью света $c = 3\cdot10^8$ м/с (сам свет также является электромагнитной волной).

[48] Это — *принцип относительности Эйнштейна*, лежащий в основе теории относительности.

[49] Для однозначного нахождения полей необходимо знать ещё *начальные условия* — значения полей в начальный момент времени, а также *граничные условия* — некоторые условия для полей на границе рассматриваемой области.

Сейчас электромагнитные волны хорошо изучены и классифицированы по нескольким диапазонам (подробному рассмотрению электромагнитных волн посвящён следующий раздел). Но во времена Максвелла никто даже не подозревал об их существовании. Это был один из удивительных случаев в физике, когда фундаментальное открытие делалось «на кончике пера» — новое явление открывалось чисто теоретически, опережая эксперимент!

Экспериментальное доказательство пришло позже: электромагнитные волны были впервые обнаружены в опытах Герца через восемь лет после смерти Максвелла. Эти опыты подтвердили справедливость симметричной гипотезы и основанной на ней теории электромагнитного поля, построенной Максвеллом.

3.27 Электромагнитные волны

Важнейший результат электродинамики, вытекающий из уравнений Максвелла[50], состоит в том, что *электромагнитные взаимодействия передаются из одной точки пространства в другую не мгновенно, а с конечной скоростью.* В вакууме скорость распространения электромагнитных взаимодействий совпадает со скоростью света $c = 3 \cdot 10^8$ м/с.

Рассмотрим, например, два покоящихся заряда, находящихся на некотором расстоянии друг от друга. Сила их взаимодействия определяется законом Кулона. Шевельнём один из зарядов; согласно закону Кулона сила взаимодействия изменится мгновенно — второй заряд сразу «почувствует» изменение положения первого заряда. Так утверждала теория дальнодействия[51].

Однако в действительности дело обстоит иначе. При шевелении заряда электрическое поле вблизи него меняется и порождает магнитное поле. Это магнитное поле также является переменным и, в свою очередь, порождает переменное электрическое поле, которое опять порождает переменное магнитное поле и т. д. Таким образом, *в окружающем пространстве начинает распространяться процесс колебаний напряжённости электрического поля и индукции магнитного поля — электромагнитная волна.* Спустя некоторое время эта электромагнитная волна достигнет второго заряда; лишь тогда — а не мгновенно! — он и «почувствует», что положение первого заряда изменилось.

Существование электромагнитных волн было теоретически предсказано Максвеллом и получило блестящее подтверждение в опыте Герца.

3.27.1 Открытый колебательный контур

Электромагнитные волны должны быть достаточно интенсивными для того, чтобы их можно было наблюдать в эксперименте.

Нетрудно понять, что электромагнитные волны будут тем интенсивнее, чем быстрее меняется положение зарядов, излучающих эти волны. Действительно, в таком случае электрическое поле вблизи зарядов меняется с большей скоростью и порождает большее магнитное поле; оно, в свою очередь, меняется столь же быстро и порождает большее электрическое поле, и т. д.

В частности, *интенсивные электромагнитные волны порождаются высокочастотными электромагнитными колебаниями.*

Электромагнитные колебания создаются в хорошо знакомом нам колебательном контуре. Частота колебаний заряда и тока в контуре равна:

$$\nu = \frac{1}{2\pi\sqrt{LC}}\,. \tag{3.139}$$

С этой же частотой колеблются векторы $\vec{E}$ и $\vec{B}$ в заданной точке пространства. Поэтому величина ν, вычисляемая по формуле (3.139), будет также *частотой электромагнитной волны.*

Чтобы увеличить частоту колебаний в контуре, нужно уменьшать ёмкость конденсатора и индуктивность катушки.

Но эксперименты показали, что дело не ограничивается одной лишь высокой частотой колебаний. Для образования интенсивных электромагнитных волн существенным оказывается ещё один фактор: *переменное электромагнитное поле, являющееся источником электромагнитных волн, должно занимать достаточно большую область пространства.*

Между тем, в обычном колебательном контуре, состоящем из конденсатора и катушки, переменное электрическое поле почти целиком сосредоточено в малой области внутри конденсатора,

[50] Мы уже не первый раз говорим об уравнениях Максвелла, а самих уравнений при этом не выписываем. Ничего не поделаешь — эти уравнения пока слишком сложны для вас. Вы познакомитесь с ними курсе на втором, когда будут освоены необходимые темы из высшей математики.

[51] Теории дальнодействия и близкодействия обсуждались в разделе «Напряжённость электрического поля».

а переменное магнитное поле — в малой области внутри катушки. Поэтому даже при достаточно высокой частоте колебаний такой колебательный контур оказался непригоден для излучения электромагнитных волн.

Как добиться увеличения области, занимаемой высокочастотным электромагнитным полем? Герц нашёл красивое и гениально простое решение — *открытый колебательный контур*.

Возьмём обычный колебательный контур (рис. 3.132, слева). Начнём уменьшать число витков катушки — от этого её индуктивность будет уменьшаться. Одновременно уменьшаем площадь пластин конденсатора и раздвигаем их — это приводит к уменьшению ёмкости конденсатора и к увеличению пространственной области, занимаемой электрическим полем. Эта промежуточная ситуация изображена на рис. 3.132 в середине.

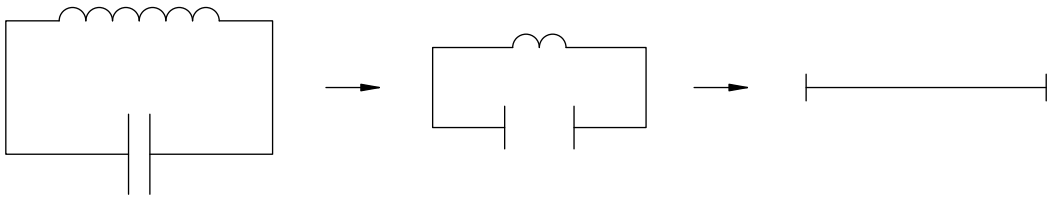


Рис. 3.132. Превращение обычного колебательного контура в открытый

К чему мы придём, продолжая этот процесс? Катушка ликвидируется вовсе, превращаясь в кусок проводника. Пластины конденсатора раздвигаются максимально далеко и оказываются на концах этого проводника (рис. 3.132, справа). Остаётся уменьшить до предела размеры пластин — и получится самый обычный прямолинейный стержень! Это и есть открытый колебательный контур (рис. 3.133).

Рис. 3.133. Открытый колебательный контур

Таким образом, идея Герца превратить обычный колебательный контур в открытый позволила сразу «убить двух зайцев»:

1. ёмкость и индуктивность стержня очень малы, поэтому в нём возбуждаются колебания весьма высокой частоты;

2. переменное электромагнитное поле занимает довольно большую область пространства вокруг стержня.

Поэтому такой стержень может служить источником достаточно интенсивных электромагнитных волн, которые будут доступны для экспериментального наблюдения.

Но как возбудить в стержне электромагнитные колебания? Герц разрезал стержень посередине, раздвинул половинки на небольшое расстояние (создав так называемый разрядный промежуток) и подключил их к источнику высокого напряжения (рис. 3.134).

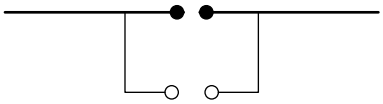


Рис. 3.134. Излучающий вибратор Герца

Получился *излучающий вибратор Герца*; концы провода в разрядном промежутке снабжались небольшими шариками. Когда напряжение между шариками превышало напряжение

пробоя, в разрядном промежутке проскакивала искра. Во время существования искры цепь замыкалась, и в стержне возникали электромагнитные колебания — вибратор излучал электромагнитные волны.

Герц регистрировал эти волны с помощью *приёмного вибратора* — проводника с шариками на концах разрядного промежутка (рис. 3.135). Приёмный вибратор находился поодаль, на некотором расстоянии от излучающего вибратора.

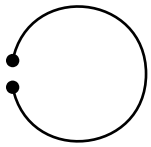


Рис. 3.135. Приёмный вибратор Герца

Переменное электрическое поле электромагнитной волны возбуждало в приёмном вибраторе переменный ток. Если частота этого тока совпадала с собственной частотой приёмного вибратора, то возникал резонанс, и в разрядном промежутке проскакивала искра!

Наличие этой искры, появляющейся на концах совершенно изолированного проводника, явилось ярким свидетельством существования электромагнитных волн.

3.27.2 Свойства электромагнитных волн

Для излучения электромагнитных волн заряд не обязательно должен совершать колебательное движение; главное — чтобы у заряда было ускорение. *Любой заряд, движущийся с ускорением, является источником электромагнитных волн.* При этом излучение будет тем интенсивнее, чем больше модуль ускорения заряда.

Так, при равномерном движении по окружности (скажем, в магнитном поле) заряд имеет центростремительное ускорение и, стало быть, излучает электромагнитные волны. Быстрые электроны в газоразрядных трубках, налетая на стенки, тормозятся с очень большим по модулю ускорением; поэтому вблизи стенок регистрируется рентгеновское излучение высокой энергии (так называемое *тормозное излучение*).

В качестве примера рассмотрим излучение заряда, совершающего гармонические колебания с частотой ν вдоль оси Y вокруг начала координат. Во все стороны от него бегут электромагнитные волны — в частности, вдоль оси X. На рис. 3.136 показана структура излучаемой электромагнитной волны на большом расстоянии от заряда в фиксированный момент времени.

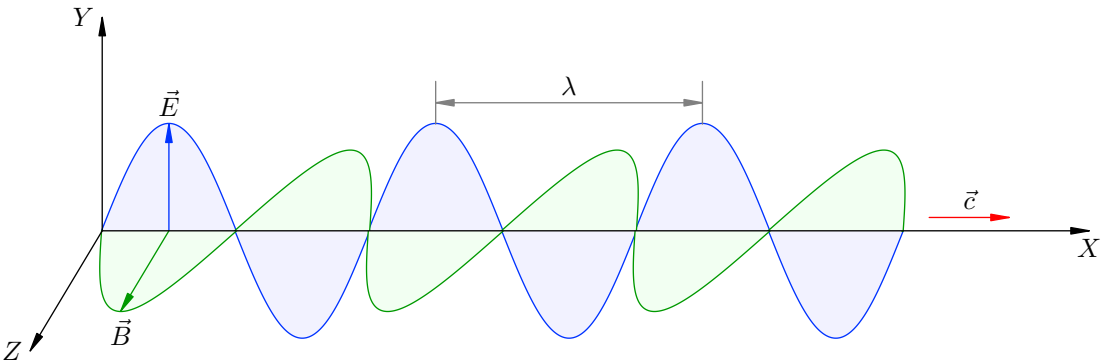


Рис. 3.136. Синусоидальная электромагнитная волна

Скорость волны $\vec{c}$ направлена вдоль оси X. Векторы $\vec{E}$ и $\vec{B}$ в каждой точке оси X совершают синусоидальные колебания вдоль осей Y и Z соответственно, меняясь при этом синфазно.

Кратчайший поворот вектора $\vec{E}$ к вектору $\vec{B}$ всегда совершается против часовой стрелки, если глядеть с конца вектора $\vec{c}$.

В любой фиксированный момент времени распределение вдоль оси X значений модуля векторов $\vec{E}$ и $\vec{B}$ имеет вид двух синфазных синусоид, расположенных перпендикулярно друг другу в плоскостях XY и XZ соответственно. *Длина волны* λ — это расстояние между двумя ближайшими точками оси X, в которых колебания значений поля происходят в одинаковой фазе (в частности — между двумя ближайшими максимумами поля, как на рис. 3.136).

Таким образом, электромагнитные волны оказались **поперечными** — колебания векторов напряжённости электрического поля и индукции магнитного поля происходят в плоскости, перпендикулярной направлению распространения волны.

Частота, с которой меняются значения E и B в данной точке пространства, называется *частотой электромагнитной волны*; она совпадает с частотой ν колебаний излучающего заряда. Длина электромагнитной волны λ, её частота ν и скорость распространения c связаны стандартным для всех волн соотношением:

$$c = \lambda\nu. \tag{3.140}$$

Эксперименты показали, что электромагнитным волнам присущи те же основные свойства, что и другим видам волновых процессов.

1. *Отражение волн.* Электромагнитные волны отражаются от металлического листа — это было обнаружено ещё Герцем. Угол отражения при этом равен углу падения.

2. *Поглощение волн.* Электромагнитные волны частично поглощаются при прохождении сквозь диэлектрик.

3. *Преломление волн.* Электромагнитные волны меняют направление распространения при переходе из воздуха в диэлектрик (и вообще на границе двух различных диэлектриков).

4. *Интерференция волн.* Герц наблюдал интерференцию двух волн: первая приходила к приёмному вибратору непосредственно от излучающего вибратора, вторая — после предварительного отражения от металлического листа.

   Меняя положение приёмного вибратора и фиксируя положения интерференционных максимумов, Герц измерил длину волны λ. Частота ν собственных колебаний в приёмном вибраторе была Герцу известна. По формуле (3.140) Герц вычислил скорость распространения электромагнитных волн и получил приближённо $c \approx 3 \cdot 10^8$ м/с. Именно такой результат предсказывала теория, построенная Максвеллом!

5. *Дифракция волн.* Электромагнитные волны огибают препятствия, размеры которых соизмеримы с длиной волны. Например, радиоволны, длина волны которых составляет несколько десятков или сотен метров, огибают дома или горы, находящиеся на пути их распространения.

3.27.3 Плотность потока излучения

Электромагнитные волны переносят энергию из одних участков пространства в другие. Перенос энергии осуществляется вдоль *лучей* — воображаемых линий, указывающих направление распространения волны[52].

Важнейшей энергетической характеристикой электромагнитных волн служит плотность потока излучения.

[52]Мы не даём строгого определения понятия луча и надеемся на ваше интуитивное понимание, которого пока будет вполне достаточно.

Представим себе площадку площадью S, расположенную перпендикулярно лучам. Допустим, что за время t волна переносит через эту площадку энергию W. Тогда *плотность потока излучения* I определяется формулой:

$$I = \frac{W}{St}\,. \tag{3.141}$$

Иначе говоря, плотность потока излучения — это энергия, переносимая через единичную площадку (перпендикулярную лучам) в единицу времени; или, что то же самое — это мощность излучения, переносимая через единичную площадку. Единицей измерения плотности потока излучения служит Вт/м^2.

Плотность потока излучения связана простым соотношением с плотностью энергии электромагнитного поля.

Фиксируем площадку S, перпендикулярную лучам, и небольшой промежуток времени t. Сквозь площадку пройдёт энергия:

$$W = ISt. \tag{3.142}$$

Эта энергия будет сосредоточена внутри цилиндра с площадью основания S и высотой ct (рис. 3.137), где c — скорость электромагнитной волны.

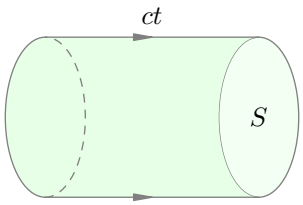


Рис. 3.137. К выводу формулы $I = wc$

Объём данного цилиндра равен: $V = Sct$. Поэтому если w — плотность энергии электромагнитного поля, то для энергии W получим также:

$$W = wV = wSct. \tag{3.143}$$

Приравнивая правые части формул (3.142) и (3.143) и сокращая на St, получим соотношение:

$$I = wc. \tag{3.144}$$

Плотность потока излучения характеризует, в частности, степень воздействия электромагнитного излучения на его приёмники; когда говорят об *интенсивности* электромагнитных волн, имеют в виду именно плотность потока излучения.

Интересным является вопрос о том, как интенсивность излучения зависит от его частоты.

Пусть электромагнитная волна излучается зарядом, совершающим гармонические колебания вдоль оси X по закону $x = x_0 \sin \omega t$. Циклическая частота ω колебаний заряда будет в то же время циклической частотой излучаемой электромагнитной волны.

Для скорости и ускорения заряда имеем: $v = \dot{x} = x_0\omega \cos \omega t$ и $a = \dot{v} = -x_0\omega^2 \sin \omega t$. Как видим, $a \sim \omega^2$. Напряжённость электрического поля и индукция магнитного поля в электромагнитной волне пропорциональны ускорению заряда: $E \sim a$ и $B \sim a$. Стало быть, $E \sim \omega^2$ и $B \sim \omega^2$.

Плотность энергии электромагнитного поля есть сумма плотности энергии электрического поля и плотности энергии магнитного поля: $w = w_{\text{эл}} + w_{\text{магн}}$. Плотность энергии электрического поля, как мы знаем, пропорциональна квадрату напряжённости поля: $w_{\text{эл}} \sim E^2$. Аналогично можно показать, что $w_{\text{магн}} \sim B^2$. Следовательно, $w_{\text{эл}} \sim \omega^4$ и $w_{\text{магн}} \sim \omega^4$, так что $w \sim \omega^4$.

Согласно формуле (3.144) плотность потока излучения пропорциональна плотности энергии: $I \sim w$. Поэтому $I \sim \omega^4$. Мы получили важный результат: *интенсивность электромагнитного излучения пропорциональна четвёртой степени его частоты.*

Другой важный результат заключается в том, что *интенсивность излучения убывает с увеличением расстояния до источника.* Это понятно: ведь источник излучает в разных направлениях, и по мере удаления от источника излучённая энергия распределяется по всё большей и большей площади.

Количественную зависимость плотности потока излучения от расстояния до источника легко получить для так называемого точечного источника излучения.

Точечный источник излучения — это источник, размерами которого в условиях данной ситуации можно пренебречь. Кроме того, считается, что точечный источник одинаково излучает во всех направлениях.

Конечно, точечный источник является идеализацией, но в некоторых задачах эта идеализация отлично работает. Например, при исследовании излучения звёзд их вполне можно считать точечными источниками — ведь расстояния до звёзд настолько громадны, что их собственные размеры можно не принимать во внимание.

На расстоянии r от источника излучённая энергия равномерно распределяется по поверхности сферы радиуса r. Площадь сферы, напомним, $S = 4\pi r^2$. Если мощность излучения нашего источника равна P, то за время t через поверхность сферы проходит энергия $W = Pt$. С помощью формулы (3.141) получаем тогда:

$$I = \frac{Pt}{4\pi r^2 t} = \frac{P}{4\pi r^2}\,.$$

Таким образом, *интенсивность излучения точечного источника обратно пропорциональна расстоянию до него.*

3.27.4 Виды электромагнитных излучений

Спектр электромагнитных волн необычайно широк: длина волны может измеряться тысячами километров, а может быть меньше пикометра. Тем не менее, весь этот спектр можно разделить на несколько характерных диапазонов длин волн; внутри каждого диапазона электромагнитные волны обладают более-менее схожими свойствами и способами излучения.

Мы рассмотрим эти диапазоны в порядке убывания длины волны. Диапазоны плавно переходят друг в друга, чёткой границы между ними нет. Поэтому граничные значения длин волн порой весьма условны.

1. **Радиоволны** ($\lambda > 1$ мм).

   Источниками радиоволн служат колебания зарядов в проводах, антеннах, колебательных контурах. Радиоволны излучаются также во время гроз.

   - *Сверхдлинные волны* ($\lambda > 10$ км). Хорошо распространяются в воде, поэтому используются для связи с подводными лодками.
   - *Длинные волны* (1 км $< \lambda <$ 10 км). Используются в радиосвязи, радиовещании, радионавигации.
   - *Средние волны* (100 м $< \lambda <$ 1 км). Радиовещание. Радиосвязь на расстоянии не более 1500 км.
   - *Короткие волны* (10 м $< \lambda <$ 100 м). Радиовещание. Хорошо отражаются от ионосферы; в результате многократных отражений от ионосферы и от поверхности Земли могут распространяться вокруг земного шара. Поэтому на коротких волнах можно ловить радиостанции других стран.

- *Метровые волны* (1 м $< \lambda <$ 10 м). Местное радивещание в УКВ-диапазоне. Например, длина волны радиостанции «Эхо Москвы» составляет 4 м. Используются также в телевидении (федеральные каналы); так, длина волны телеканала «Россия 1» равна примерно 5 м.
- *Дециметровые волны* (10 см $< \lambda <$ 1 м). Телевидение (дециметровые каналы). Например, длина волны телеканала «Animal Planet» приблизительно равна 42 см.

  Это также диапазон мобильной связи; так, стандарт GSM 1800 использует радиоволны с частотой примерно 1800 МГц, т. е. с длиной волны около 17 см.

  Есть ещё одно хорошо известное вам применение дециметровых волн — это микроволновые печи. Стандартная частота микроволновой печи равна 2450 МГц (это частота, на которой происходит резонансное поглощение электромагнитного излучения молекулами воды). Она отвечает длине волны примерно 12 см.

  Наконец, в технологиях беспроводной связи Wi-Fi и Bluetooth используется такая же длина волны — 12 см (частота 2400 МГц).
- *Сантиметровые волны* (1 см $< \lambda <$ 10 см). Это — область радиолокации и спутниковых телеканалов. Например, канал НТВ+ ведёт своё телевещание на длинах волн около 2 см.
- *Миллиметровые волны* (1 мм $< \lambda <$ 1 см). Радиолокация, космические линии связи. Здесь мы подходим к длинноволновой границе инфракрасного излучения.

2. **Инфракрасное излучение** (780 нм $< \lambda <$ 1 мм).

   Испускается молекулами и атомами нагретых тел. Инфракрасное излучение называется ещё *тепловым* — когда оно попадает на наше тело, мы чувствуем тепло. Человеческим глазом инфракрасное излучение не воспринимается[53].

   Мощнейшим источником инфракрасного излучения служит Солнце. Лампы накаливания излучают наибольшее количество энергии (до 80%) в как раз в инфракрасной области спектра.

   Инфракрасное излучение имеет широкую область применения: инфракрасные обогреватели, пульты дистанционного управления, приборы ночного видения, сушка лакокрасочных покрытий и многое другое.

   При повышении температуры тела длина волны инфракрасного излучения уменьшается, смещаясь в сторону видимого света. Засунув гвоздь в пламя горелки, мы можем наблюдать это воочию: в какой-то момент гвоздь «раскаляется докрасна», начиная излучать в видимом диапазоне.

3. **Видимый свет** (380 нм $< \lambda <$ 780 нм).

   Излучение в этом промежутке длин волн воспринимается человеческим глазом.

   Диапазон видимого света можно разделить на семь интервалов — так называемые *спектральные цвета*.

   - Красный: 625 нм — 780 нм;
   - Оранжевый: 590 нм — 625 нм;
   - Жёлтый: 565 нм — 590 нм;
   - Зелёный: 500 нм — 565 нм;
   - Голубой: 485 нм — 500 нм;

[53] Некоторые змеи видят в инфракрасном диапазоне.

- Синий: 440 нм — 485 нм;
- Фиолетовый: 380 нм — 440 нм.

Глаз имеет максимальную чувствительность к свету в зелёной части спектра. Вот почему школьные доски согласно ГОСТу должны быть зелёными: глядя на них, глаз испытывает меньшее напряжение.

4. **Ультрафиолетовое излучение** (10 нм $< \lambda <$ 380 нм).

   Главным источником ультрафиолетового излучения является Солнце. Именно ультрафиолетовое излучение приводит к появлению загара. Человеческим глазом оно уже не воспринимается[54].

   В небольших дозах ультрафиолетовое излучение полезно для человека: оно повышает иммунитет, улучшает обмен веществ, имеет целый ряд других целебных воздействий и потому применяется в физиотерапии.

   Ультрафиолетовое излучение обладает бактерицидными свойствами. Например, в больницах для дезинфекции операционных в них включаются специальные ультрафиолетовые лампы.

   Очень опасным является воздействие УФ излучения на сетчатку глаза — при больших дозах ультрафиолета можно получить ожог сетчатки. Поэтому для защиты глаз (высоко в горах, например) нужно надевать очки, стёкла которых поглощают ультрафиолет.

5. **Рентгеновское излучение** (5 пм $< \lambda <$ 10 нм).

   Возникает в результате торможения быстрых электронов у анода и стенок газоразрядных трубок (*тормозное излучение*), а также при некоторых переходах электронов внутри атомов с одного уровня на другой (*характеристическое излучение*).

   Рентгеновское излучение легко проникает сквозь мягкие ткани человеческого тела, но поглощается кальцием, входящим в состав костей. Это даёт возможность хорошо известные вам рентгеновские снимки.

   В аэропортах вы наверняка видели действие *рентгенотелевизионных интроскопов* — эти приборы просвечивают рентгеновскими лучами ручную кладь и багаж.

   Длина волны рентгеновского излучения сравнима с размерами атомов и межатомных расстояний в кристаллах; поэтому кристаллы являются естественными дифракционными решётками для рентгеновских лучей. Наблюдая дифракционные картины, получаемые при прохождении рентгеновских лучей сквозь различные кристаллы, можно изучать порядок расположения атомов в кристаллических решётках и сложных молекулах.

   Так, именно с помощью *рентгеноструктурного анализа* было определено устройство ряда сложных органических молекул — например, ДНК и гемоглобина.

   В больших дозах рентгеновское излучение опасно для человека — оно может вызывать раковые заболевания и лучевую болезнь.

6. **Гамма-излучение** ($\lambda <$ 5 пм).

   Это излучение наиболее высокой энергии. Его проникающая способность намного выше, чем у рентгеновских лучей.

   Гамма-излучение возникает при переходах атомных ядер из одного состояния в другое, а также при некоторых ядерных реакциях.

[54] Некоторые насекомые и птицы способны видеть в ультрафиолете. Например, пчёлы с помощью своего ультрафиолетового зрения находят нектар на цветах.

Источниками гамма-лучей могут быть заряженные частицы, движущиеся со скоростями, близкими к скорости света — в случае, если траектории таких частиц искривлены магнитным полем (так называемое *синхротронное излучение*).

В больших дозах гамма-излучение очень опасно для человека: оно вызывает лучевую болезнь и онкологические заболевания. Но в малых дозах оно может подавлять рост раковых опухолей и потому применяется в *лучевой терапии*.

Бактерицидное действие гамма-излучения используется в сельском хозяйстве (гамма-стерилизация сельхозпродукции перед длительным хранением), в пищевой промышленности (консервирование продуктов), а также в медицине (стерилизация материалов).

Глава 4

Оптика

Оптика есть наука о распространении света. Говоря о свете, мы всегда подразумеваем *видимый свет*. Напомним, что на огромной шкале электромагнитных волн диапазон видимого света занимает весьма узкий промежуток: длины волн видимого диапазона принимают значения примерно от 380 нм (фиолетовый участок спектра) до 780 нм (красный участок).

С точки зрения электродинамики Максвелла распространение света ничем не отличается от распространения других электромагнитных излучений — радиоволн, инфракрасного, ультрафиолетового, рентгеновского и гамма-излучения. В этом смысле оптика оказывается просто частью электродинамики.

Но ввиду той колоссальной роли, которую свет играет в жизни человека, оптические явления начали изучаться давным-давно. Все основные законы оптики были установлены задолго до создания электродинамики и открытия электромагнитных волн. И потому с тех давних пор оптика оформилась в самостоятельный раздел физики — со своими специфическими задачами, методами, экспериментами и приборами.

Волновая оптика (или, как ещё говорят, *физическая оптика*) рассматривает свет как электромагнитные волны. Явления интерференции и дифракции света служат опытным подтверждением его волновой природы.

Геометрическая оптика устанавливает закономерности распространения световых лучей. Геометрическая оптика является предельным случаем волновой оптики при $\lambda \to 0$; иными словами, геометрическая оптика работает тем лучше, чем меньше длина световой волны по сравнению с характерными размерами препятствий.

С геометрической оптики мы и начинаем изложение. Ей посвящены следующие разделы:

→ Световые лучи
→ Отражение света
→ Преломление света
→ Линзы. Ход лучей
→ Тонкие линзы. Ход лучей
→ Тонкие линзы. Построение изображений
→ Глаз человека
→ Оптические приборы

Волновая оптика объясняет законы геометрической оптики и устанавливает границы их применимости.

Так, в приближении геометрической оптики справедлив закон прямолинейного распространения света, но удовлетворительно работает он лишь вдали от краёв препятствий. Вблизи края наблюдается *дифракция света* — изменение направления световых лучей и их проникновение в область геометрической тени. Объяснение закона прямолинейного распространения света даётся в рамках теории дифракции и выходит за рамки школьной программы.

Имеются отступления и от закона независимости световых лучей. А именно, при некоторых условиях освещённость, создаваемая несколькими световыми пучками, не равна сумме освещённостей, вносимых каждым пучком в отдельности, и может принимать как большее, так и меньше значения. В этом состоит явление *интерференции света*. При интерференции вместо равномерной освещённости экрана наблюдается интерференционная картина — чередование светлых и тёмных участков пространства.

Дисперсия света, то есть различная преломляемость лучей разного цвета на одной и той же границе, была исследована ещё Ньютоном. Но появление теории дисперсии, выясняющей характер зависимости показателя преломления среды от частоты электромагнитных волн, оказалось возможным лишь после создания электродинамики Максвелла.

Волновая оптика рассматривается в следующих разделах:

→ Принцип Гюйгенса
→ Интерференция волн
→ Интерференция света
→ Дифракция света
→ Дисперсия света

Волновая оптика — обширная и непростая наука; сколько-нибудь полное изложение волновой оптики возможно лишь в рамках вузовского курса физики[1]. В данном пособии затрагиваются лишь те вопросы, которые необходимы для адекватного восприятия школьной программы и успешного решения задач ЕГЭ по физике и различных олимпиад.

[1] Достаточно сказать, что в знаменитом физтеховском пятитомнике Д. В. Сивухина «Общий курс физики» четвёртый том «Оптика» является самым толстым из всех томов.

4.1 Световые лучи

Главным природным источником света служит Солнце, и люди ставили много опытов с солнечными лучами. Отсюда в оптику вошло понятие *светового луча*. Впоследствии оно получило строгое определение.

Световой луч — это геометрическая линия, которая в каждой своей точке перпендикулярна волновому фронту, проходящему через эту точку. Направление светового луча совпадает с направлением распространения света.

Если данное определение осталось для вас не совсем понятным — ничего страшного: на первых порах вы можете представлять себе просто узкие пучки света наподобие солнечных лучей. Этого вполне хватит, чтобы уяснить все основные вещи и научиться решать задачи. Ну а время строгого определения придёт несколько позже — когда начнётся волновая оптика.

4.1.1 Законы геометрической оптики

Геометрическая оптика изучает распространение световых лучей. Это исторически первый и наиболее простой раздел оптики. В основе геометрической оптики лежат четыре основных закона.

1. Закон независимости световых лучей.
2. Закон прямолинейного распространения света.
3. Закон отражения света.
4. Закон преломления света.

Данные законы были установлены в результате наблюдений за световыми лучами и послужили обобщениями многочисленных опытных фактов. Они являются утверждениями, сформулированными на языке геометрии. Волновая природа света в них не затрагивается.

Законы геометрической оптики первоначально являлись *постулатами*. Они лишь констатировали: таким вот образом ведёт себя природа. Однако впоследствии оказалось, что законы геометрической оптики могут быть *выведены* из более фундаментальных законов волновой оптики.

Геометрическая оптика отлично работает, когда длина световой волны λ много меньше размеров объектов, присутствующих в данной физической ситуации. Можно сказать, что геометрическая оптика есть предельный случай волновой оптики при $\lambda \to 0$. Неудивительно поэтому, что сначала были открыты законы именно геометрической оптики: ведь размеры предметов, встречающихся нам в повседневной жизни, намного превышают длины волн видимого света.

Первый закон геометрической оптики совсем простой. Он говорит о том, что вклад каждого светового луча в суммарное освещение не зависит от наличия других лучей.

Закон независимости световых лучей. Если световые лучи пересекаются, то они не оказывают никакого влияния друг на друга. Каждый луч освещает пространство так, как если бы других лучей вообще не было.

Закон прямолинейного распространения света также очень прост, и мы его сейчас обсудим. Законам отражения и преломления будут посвящены следующие разделы.

Закон прямолинейного распространения света. В прозрачной однородной среде световые лучи являются прямыми линиями.

Что такое «прозрачная однородная среда»? Среда называется *прозрачной*, если в ней может распространяться свет. Среда называется *однородной*, если её свойства не меняются от точки

к точке. Равномерно прогретый воздух, чистая вода, стекло без примесей — всё это примеры прозрачных и оптически однородных сред.

Таким образом, закон прямолинейного распространения света означает, что в прозрачной однородной среде понятие светового луча совпадает с понятием луча в геометрии.

Данный закон не требует каких-либо дополнительных пояснений — он хорошо вам известен. Вам неоднократно доводилось видеть прямолинейные солнечные лучи, пронизывающие облака, или тонкий прямой луч, пробивающийся в запылённой комнате через щель в окне. Находясь под водой, можно наблюдать прямые солнечные лучи, идущие сквозь воду.

При нарушении однородности среды нарушается и закон прямолинейного распространения света. Например, на границе раздела двух прозрачных сред световой луч может разделиться на два луча: отражённый и преломлённый. Если оптические свойства среды меняются от точки к точке, то ход световых лучей искривляется. В этом состоит причина миражей: слой воздуха вблизи раскалённой земной поверхности нагрет больше, чем вышележащие слои; он имеет иные оптические свойства, и его действие оказывается подобным зеркалу. Обо всём этом мы поговорим позднее.

4.1.2 Геометрическая тень

Вам хорошо известно, что различные предметы отбрасывают тень. На рис. 4.1 изображён точечный источник света S и непрозрачный предмет — красный треугольник. На экране мы видим тень этого предмета в виде серого треугольника.

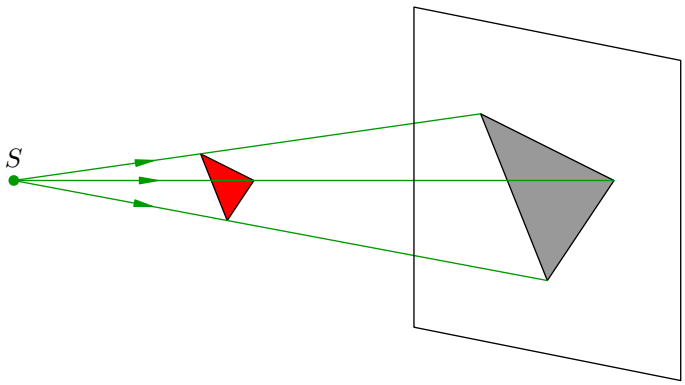


Рис. 4.1. Геометрическая тень

Откуда берётся тень? Дело в том, что если на пути световых лучей оказывается непрозрачный предмет, то происходит следующее.

- Луч, идущий мимо предмета, продолжает распространяться в прежнем направлении — как если бы данного предмета вообще не было.
- Луч, попадающий на предмет, не проникает внутрь предмета. Дальнейший ход такого луча в прежнем направлении пресекается.

Так возникает *геометрическая тень*, края которой чётко очерчены[2]. Поскольку свет распространяется прямолинейно, форма геометрической тени оказывается подобной контуру предмета. Так, на рис. 4.1 серый треугольник подобен красному.

[2]Граница реальной тени имеет более сложный вид: вмешивается *дифракция* света на краях предмета. Дифракция — это отклонение света от первоначального направления; данное явление обусловлено волновой природой света и не описывается в рамках геометрической оптики.

4.2 Отражение света

Когда световой луч падает на границу раздела двух сред, происходит *отражение света*: луч изменяет направление своего хода и возвращается в исходную среду.

На рис. 4.2 изображены *падающий луч* AO, *отражённый луч* OB, а также перпендикуляр OC, проведённый к отражающей поверхности KL в точке падения O.

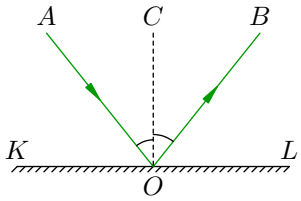


Рис. 4.2. Закон отражения

Угол AOC называется *углом падения*. Обратите внимание и запомните: угол падения отсчитывается от перпендикуляра к отражающей поверхности, а не от самой поверхности! Точно так же *угол отражения* — это угол BOC, образованный отражённым лучом и перпендикуляром к поверхности.

4.2.1 Закон отражения

Сейчас мы сформулируем один из самых древних законов физики. Он был известен грекам ещё в античности!

Закон отражения.

1) Падающий луч, отражённый луч и перпендикуляр к отражающей поверхности, проведённый в точке падения, лежат в одной плоскости.

2) Угол отражения равен углу падения.

Таким образом, $\angle AOC = \angle BOC$, что и показано на рис. 4.2.

Закон отражения имеет одно простое, но очень важное геометрическое следствие. Давайте посмотрим на рис. 4.3. Пусть из точки A исходит световой луч. Построим точку A', симметричную точке A относительно отражающей поверхности KL.

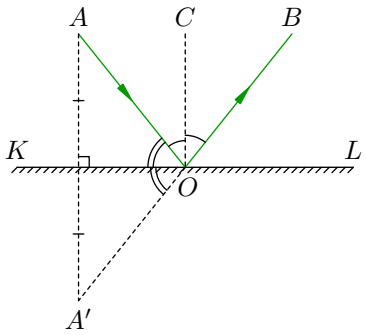


Рис. 4.3. Отражённый луч выходит из точки A'

Из симметрии точек A и A' ясно, что $\angle AOK = \angle A'OK$. Кроме того, $\angle AOK + \angle AOC = 90°$. Поэтому $\angle A'OB = 2(\angle AOK + \angle AOC) = 180°$, и, следовательно, точки A', O и B лежат на одной прямой! *Отражённый луч OB как бы выходит из точки A', симметричной точке A относительно отражающей поверхности.* Данный факт нам чрезвычайно пригодится в самом скором времени.

Закон отражения описывает ход отдельных световых лучей — узких пучков света. Но во многих случаях пучок является достаточно широким, то есть состоит из множества параллельных лучей. Картина отражения широкого пучка света будет зависеть от свойств отражающей поверхности.

Если поверхность является неровной, то после отражения параллельность лучей нарушится. В качестве примера на рис. 4.4 показано отражение от волнообразной поверхности. Отражённые лучи, как видим, идут в самых разных направлениях.

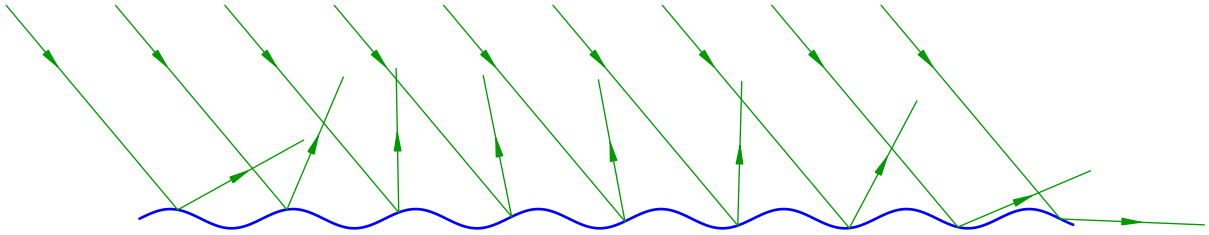


Рис. 4.4. Отражение от волнообразной поверхности

Но что значит «неровная» поверхность? Какие поверхности являются «ровными»? Ответ таков: поверхность считается неровной, если размеры её неровностей не меньше длины световых волн. Так, на рис. 4.4 характерный размер неровностей на несколько порядков превышает величину длин волн видимого света.

Поверхность с микроскопическими неровностями, соизмеримыми с длинами волн видимого света, называется *матовой*. В результате отражения параллельного пучка от матовой поверхности получается *рассеянный свет* — лучи такого света идут во всевозможных направлениях[3]. Само отражение от матовой поверхности называется поэтому *рассеянным* или *диффузным*[4].

Если же размер неровностей поверхности меньше длины световой волны, то такая поверхность называется *зеркальной*. При отражении от зеркальной поверхности параллельность пучка сохраняется: отражённые лучи также идут параллельно (рис. 4.5).

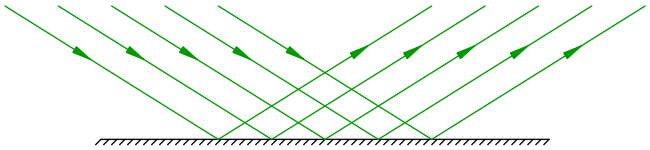


Рис. 4.5. Отражение от зеркальной поверхности

Приблизительно зеркальной является гладкая поверхность воды, стекла или отполированного металла. Отражение от зеркальной поверхности называется соответственно *зеркальным*. Нас будет интересовать простой, но важный частный случай зеркального отражения — отражение в плоском зеркале.

4.2.2 Плоское зеркало

Плоское зеркало — это часть плоскости, зеркально отражающая свет. Плоское зеркало — привычная вещь; таких зеркал несколько в вашем доме. Но теперь мы сможем разобраться, почему, смотрясь в зеркало, вы видите в нём отражение себя и находящихся рядом с вами предметов.

Точечный источник света S на рис. 4.6 испускает лучи в разных направлениях; давайте возьмём два близких луча, падающих на плоское зеркало. Мы уже знаем, что отражённые лучи

[3]Именно поэтому мы видим окружающие предметы: они отражают рассеянный свет, который мы и наблюдаем с любого ракурса.

[4]Латинское слово *diffusio* как раз и означает *распространение*, *растекание*, *рассеивание*.

пойдут так, будто они исходят из точки S', симмметричной точке S относительно плоскости зеркала.

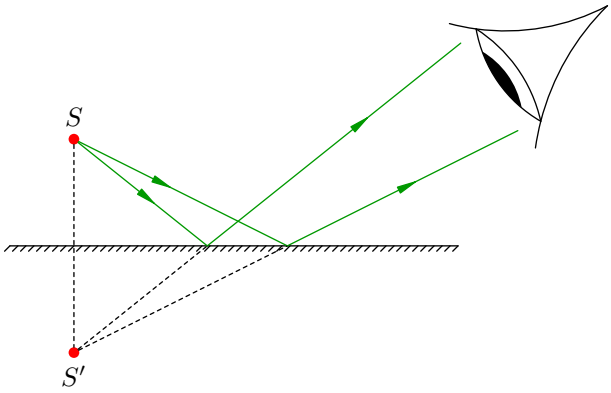


Рис. 4.6. Изображение источника света в плоском зеркале

Самое интересное начинается, когда расходящиеся отражённые лучи попадают к нам в глаз. Особенность нашего сознания состоит в том, что мозг достраивает расходящийся пучок, продолжая его за зеркало до пересечения в точке S'. Нам *кажется*, что отражённые лучи исходят из точки S' — мы видим там светящуюся точку!

Эта точка служит *изображением* источника света S. Конечно, в реальности ничего за зеркалом не светится, никакая энергия там не сосредоточена — это иллюзия, обман зрения, порождение нашего сознания. Поэтому точка S' называется *мнимым изображением* источника S. В точке S' пересекаются не сами световые лучи, а их мысленные продолжения «в зазеркалье».

Ясно, что изображение S' будет существовать независимо от размеров зеркала и от того, находится ли источник непосредственно над зеркалом или нет (рис. 4.7). Важно только, чтобы отражённые от зеркала лучи попадали в глаз — а уж глаз сам сформирует изображение источника.

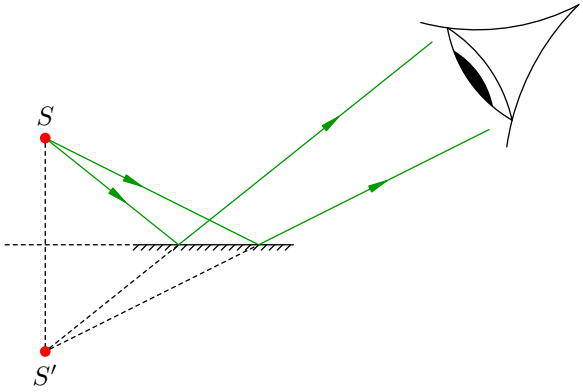


Рис. 4.7. Источник не над зеркалом: изображение есть всё равно

От расположения источника и размеров зеркала зависит *область видения* — пространственная область, из которой видно изображение источника. Область видения задаётся краями K и L зеркала KL. Построение области видения изображения S' ясно из рис. 4.8; искомая область видения выделена серым фоном.

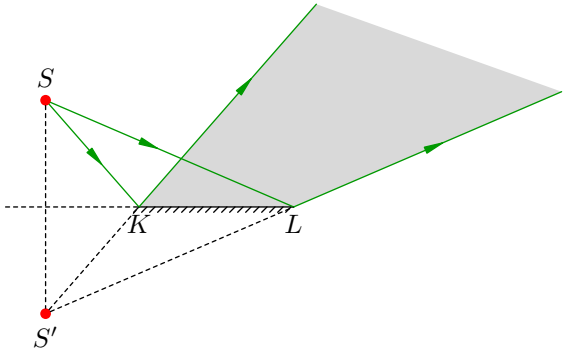


Рис. 4.8. Область видения изображения источника S

Как построить изображение произвольного предмета в плоском зеркале? Для этого достаточно найти изображение каждой точки этого предмета. Но мы знаем, что изображение точки симметрично самой точке относительно зеркала. Следовательно, *изображение предмета в плоском зеркале симметрично предмету относительно плоскости зеркала* (рис. 4.9).

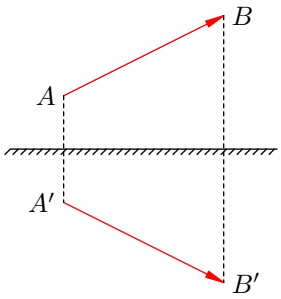


Рис. 4.9. Изображение предмета AB в плоском зеркале

Расположение предмета относительно зеркала и размеры самого зеркала не влияют на изображение (рис. 4.10).

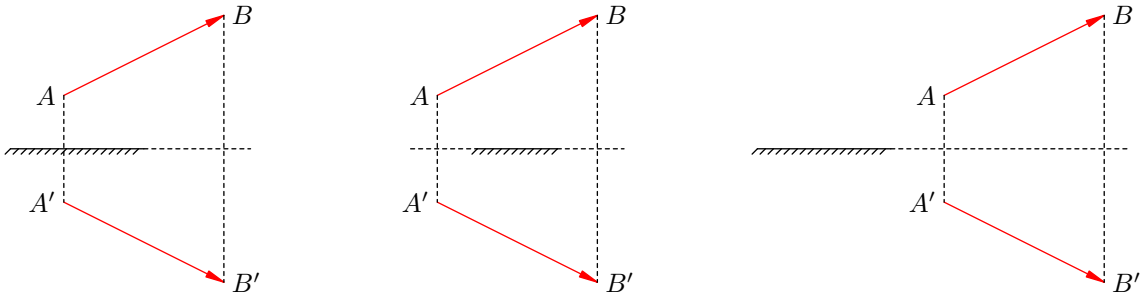


Рис. 4.10. Изображение не зависит от взаимного расположения предмета и зеркала

4.3 Преломление света

На границе раздела двух прозрачных сред наряду с отражением света наблюдается его *преломление* — свет, переходя в другую среду, меняет направление своего распространения.

Преломление светового луча происходит при его *наклонном* падении на поверхность раздела (правда, не всегда — читайте дальше про полное внутреннее отражение). Если же луч падает перпендикулярно поверхности, то преломления не будет — во второй среде луч сохранит своё направление и также пойдёт перпендикулярно поверхности.

4.3.1 Закон преломления (частный случай)

Мы начнём с частного случая, когда одна из сред является воздухом. Именно такая ситуация присутствует в подавляющем большинстве задач. Мы обсудим соответствующий частный случай закона преломления, а уж затем дадим самую общую его формулировку.

Предположим, что луч света, идущий в воздухе, наклонно падает на поверхность стекла, воды или какой-либо другой прозрачной среды. При переходе в среду луч преломляется, и его дальнейший ход показан на рис. 4.11.

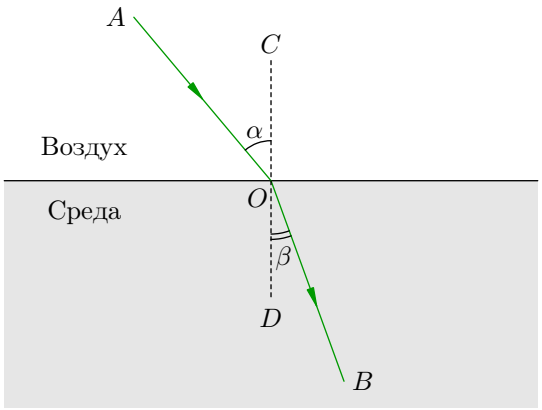


Рис. 4.11. Преломление луча на границе «воздух–среда»

В точке падения O проведён перпендикуляр (или, как ещё говорят, *нормаль*) CD к поверхности среды. Луч AO, как и раньше, называется *падающим лучом*, а угол α между падающим лучом и нормалью — *углом падения*. Луч OB — это *преломлённый луч*; угол β между преломлённым лучом и нормалью к поверхности называется *углом преломления*.

Всякая прозрачная среда характеризуется величиной n, которая называется *показателем преломления* этой среды. Показатели преломления различных сред можно найти в таблицах. Например, для стекла $n = 1{,}6$, а для воды $n = 1{,}33$. Вообще, у любой среды $n > 1$; показатель преломления равен единице только в вакууме. У воздуха $n = 1{,}0003$, поэтому для воздуха с достаточной точностью можно полагать в задачах $n = 1$ (в оптике воздух не сильно отличается от вакуума).

Закон преломления (переход «воздух–среда»).

1) Падающий луч, преломлённый луч и нормаль к поверхности, проведённая в точке падения, лежат в одной плоскости.

2) Отношение синуса угла падения к синусу угла преломления равно показателю преломления среды:

$$\frac{\sin\alpha}{\sin\beta} = n. \tag{4.1}$$

Поскольку $n > 1$, из соотношения (4.1) следует, что $\sin\alpha > \sin\beta$, то есть $\alpha > \beta$ — угол преломления меньше угла падения. Запоминаем: *переходя из воздуха в среду, луч после преломления идёт ближе к нормали.*

Показатель преломления непосредственно связан со скоростью v распространения света в данной среде. Эта скорость всегда меньше скорости света в вакууме: $v < c$. И вот оказывается, что

$$n = \frac{c}{v}\,. \tag{4.2}$$

Почему так получается, мы с вами поймём при изучении волновой оптики. А пока скомбинируем формулы (4.1) и (4.2):

$$\frac{\sin\alpha}{\sin\beta} = \frac{c}{v}\,. \tag{4.3}$$

Так как показатель преломления воздуха очень близок единице, мы можем считать, что скорость света в воздухе примерно равна скорости света в вакууме c. Приняв это во внимание и глядя на формулу (4.3), делаем вывод: *отношение синуса угла падения к синусу угла преломления равно отношению скорости света в воздухе к скорости света в среде.*

4.3.2 Обратимость световых лучей

Теперь рассмотрим обратный ход луча: его преломление при переходе из среды в воздух. Здесь нам окажет помощь следующий полезный принцип.

Принцип обратимости световых лучей. Траектория луча не зависит от того, в прямом или обратном направлении распространяется луч. Двигаясь в обратном направлении, луч пойдёт в точности по тому же пути, что и в прямом направлении.

Согласно принципу обратимости, при переходе из среды в воздух луч пойдёт по той же самой траектории, что и при соответствующем переходе из воздуха в среду (рис. 4.12) Единственное отличие рис. 4.12 от рис. 4.11 состоит в том, что направление луча поменялось на противоположное.

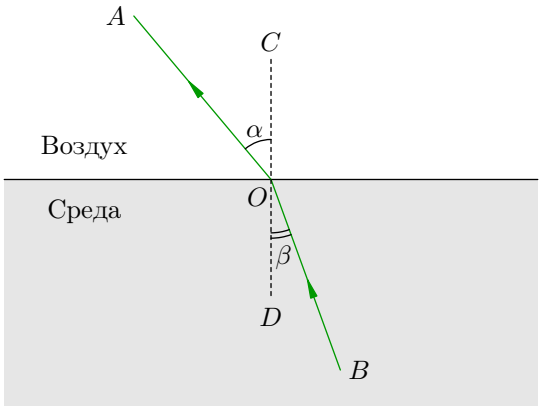


Рис. 4.12. Преломление луча на границе «среда–воздух»

Раз геометрическая картинка не изменилась, той же самой останется и формула (4.1): отношение синуса угла α к синусу угла β по-прежнему равно показателю преломления среды. Правда, теперь углы поменялись ролями: угол β стал углом падения, а угол α — углом преломления.

В любом случае, как бы ни шёл луч — из воздуха в среду или из среды в воздух — работает следующее простое правило. *Берём два угла — угол падения и угол преломления; отношение синуса большего угла к синусу меньшего угла равно показателю преломления среды.*

Теперь мы целиком подготовлены для того, чтобы обсудить закон преломления в самом общем случае.

4.3.3 Закон преломления (общий случай)

Пусть свет переходит из среды 1 с показателем преломления n_1 в среду 2 с показателем преломления n_2. Среда с бóльшим показателем преломления называется *оптически более плотной*; соответственно, среда с меньшим показателем преломления называется *оптически менее плотной*.

Переходя из оптически менее плотной среды в оптически более плотную, световой луч после преломления идёт ближе к нормали (рис. 4.13). В этом случае угол падения больше угла преломления: $\alpha > \beta$.

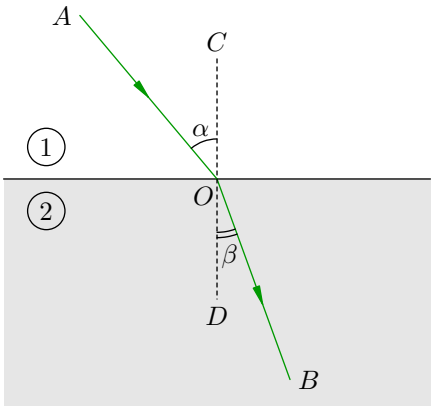


Рис. 4.13. $n_1 < n_2 \Rightarrow \alpha > \beta$

Наоборот, переходя из оптически более плотной среды в оптически менее плотную, луч отклоняется дальше от нормали (рис. 4.14). Здесь угол падения меньше угла преломления: $\alpha < \beta$.

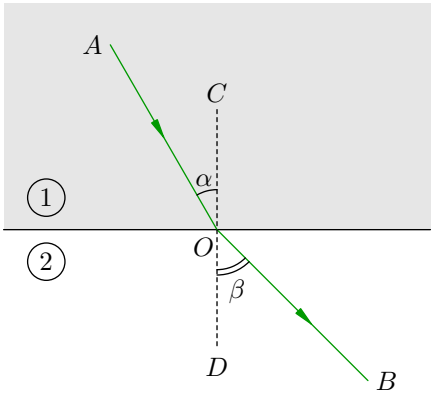


Рис. 4.14. $n_1 > n_2 \Rightarrow \alpha < \beta$

Оказывается, оба этих случая охватываются одной формулой — общим законом преломления, справедливым для любых двух прозрачных сред.

Закон преломления.

1) Падающий луч, преломлённый луч и нормаль к поверхности раздела сред, проведённая в точке падения, лежат в одной плоскости.

2) Отношение синуса угла падения к синусу угла преломления равно отношению показателя преломления второй среды к показателю преломления первой среды:

$$\frac{\sin\alpha}{\sin\beta} = \frac{n_2}{n_1}\,. \tag{4.4}$$

Нетрудно видеть, что сформулированный ранее закон преломления для перехода «воздух–среда» является частным случаем данного закона. В самом деле, полагая в формуле (4.4) $n_1 = 1$ и $n_2 = n$, мы придём к формуле (4.1).

Вспомним теперь, что показатель преломления — это отношение скорости света в вакууме к скорости света в данной среде: $n_1 = c/v_1$, $n_2 = c/v_2$. Подставляя это в (4.4), получим:

$$\frac{\sin\alpha}{\sin\beta} = \frac{v_1}{v_2}\,. \tag{4.5}$$

Формула (4.5) естественным образом обобщает формулу (4.3). *Отношение синуса угла падения к синусу угла преломления равно отношению скорости света в первой среде к скорости света во второй среде.*

4.3.4 Полное внутреннее отражение

При переходе световых лучей из оптически более плотной среды в оптически менее плотную наблюдается интересное явление — *полное внутреннее отражение*. Давайте разберёмся, что это такое.

Будем считать для определённости, что свет идёт из воды в воздух. Предположим, что в глубине водоёма находится точечный источник света S, испускающий лучи во все стороны. Мы рассмотрим некоторые из этих лучей (рис. 4.15).

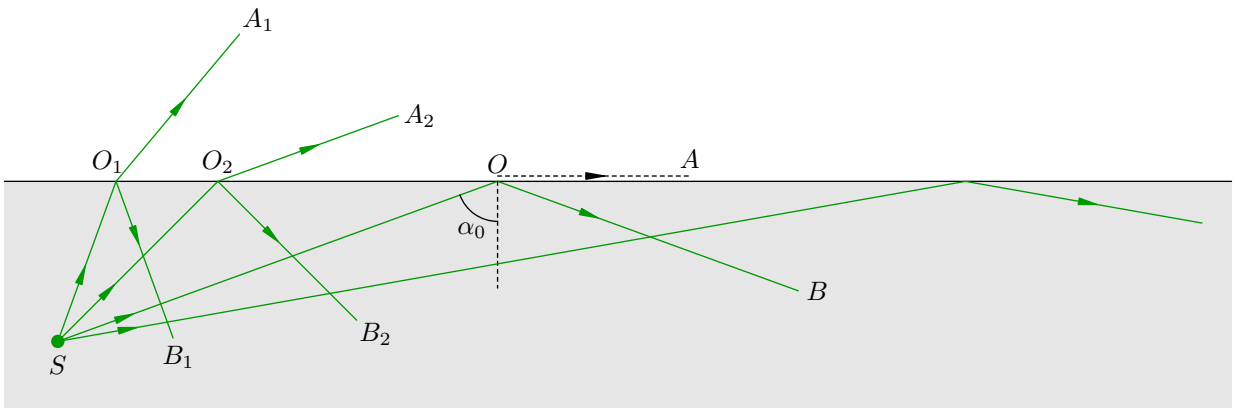


Рис. 4.15. Полное внутреннее отражение

Луч SO_1 падает на поверхность воды под наименьшим углом. Этот луч частично преломляется (луч O_1A_1) и частично отражается назад в воду (луч O_1B_1). Таким образом, часть энергии падающего луча передаётся преломлённому лучу, а оставшаяся часть энергии — отражённому лучу.

Угол падения луча SO_2 больше. Этот луч также разделяется на два луча — преломлённый и отражённый. Но энергия исходного луча распределяется между ними по-другому: преломлённый луч O_2A_2 будет тусклее, чем луч O_1A_1 (то есть получит меньшую долю энергии), а отражённый луч O_2B_2 — соответственно ярче, чем луч O_1B_1 (он получит бóльшую долю энергии).

По мере увеличения угла падения прослеживается та же закономерность: всё бóльшая доля энергии падающего луча достаётся отражённому лучу, и всё меньшая — преломлённому лучу. Преломлённый луч становится всё тусклее и тусклее, и в какой-то момент исчезает совсем!

Это исчезновение происходит при достижении угла падения α_0, которому отвечает угол преломления 90°. В данной ситуации преломлённый луч OA должен был бы пойти параллельно поверхности воды, да идти уже нечему — вся энергия падающего луча SO целиком досталась отражённому лучу OB.

При дальнейшем увеличении угла падения преломлённый луч и подавно будет отсутствовать.

Описанное явление и есть полное внутреннее отражение. Вода не выпускает наружу лучи с углами падения, равными или превышающими некоторое значение α_0 — все такие лучи целиком отражаются назад в воду. Угол α_0 называется *предельным углом полного отражения*.

Величину α_0 легко найти из закона преломления. Имеем:

$$\frac{\sin\alpha_0}{\sin 90^\circ} = \frac{1}{n}\,.$$

Но $\sin 90^\circ = 1$, поэтому

$$\sin\alpha_0 = \frac{1}{n}\,,$$

откуда

$$\alpha_0 = \arcsin\frac{1}{n}\,.$$

Так, для воды предельный угол полного отражения равен:

$$\alpha_0 = \arcsin\frac{1}{1{,}33} \approx 48{,}8^\circ.$$

Явление полного внутреннего отражения вы легко можете наблюдать дома. Налейте воду в стакан, поднимите его и смотрите на поверхность воды чуть снизу сквозь стенку стакана. Вы увидите серебристый блеск поверхности — вследствие полного внутреннего отражения она ведёт себя подобно зеркалу.

Важнейшим техническим применением полного внутреннего отражения является *волоконная оптика*. Световые лучи, запущенные внутрь оптоволоконного кабеля (*световода*) почти параллельно его оси, падают на поверхность под большими углами и целиком, без потери энергии отражаются назад внутрь кабеля. Многократно отражаясь, лучи идут всё дальше и дальше, перенося энергию на значительное расстояние. Волоконно-оптическая связь применяется, например, в сетях кабельного телевидения и высокоскоростного доступа в Интернет.

4.4 Линзы. Ход лучей

Преломление света широко используется в различных оптических приборах: фотоаппаратах, биноклях, телескопах, микроскопах... Непременной и самой существенной деталью таких приборов является линза.

Линза — это оптически прозрачное однородное тело, ограниченное с двух сторон двумя сферическими (или одной сферической и одной плоской) поверхностями.

Линзы обычно изготавливаются из стекла или специальных прозрачных пластмасс. Говоря о материале линзы, мы будем называть его стеклом — особой роли это не играет.

4.4.1 Двояковыпуклая линза

Рассмотрим сначала линзу, ограниченную с обеих сторон двумя выпуклыми сферическими поверхностями (рис. 4.16). Такая линза называется *двояковыпуклой*. Наша задача сейчас — понять ход лучей в этой линзе.

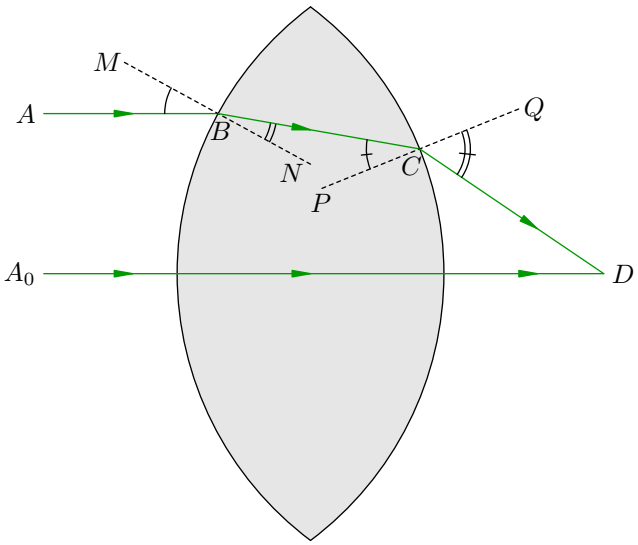


Рис. 4.16. Преломление в двояковыпуклой линзе

Проще всего обстоит дело с лучом, идущим вдоль *главной оптической оси* — оси симметрии линзы. На рис. 4.16 этот луч выходит из точки A_0. Главная оптическая ось перпендикулярна обеим сферическим поверхностям, поэтому данный луч идёт сквозь линзу, не преломляясь.

Теперь возьмём луч AB, идущий параллельно главной оптической оси. В точке B падения луча на линзу проведена нормаль MN к поверхности линзы; поскольку луч переходит из воздуха в оптически более плотное стекло, угол преломления CBN меньше угла падения ABM. Следовательно, преломлённый луч BC приближается к главной оптической оси.

В точке C выхода луча из линзы также проведена нормаль PQ. Луч переходит в оптически менее плотный воздух, поэтому угол преломления QCD больше угла падения PCB; луч преломляется опять-таки в сторону главной оптической оси и пересекает её в точке D.

Таким образом, всякий луч, параллельный главной оптической оси, после преломления в линзе приближается к главной оптической оси и пересекает её. На рис. 4.17 изображена картина преломления достаточно *широкого* светового пучка, параллельного главной оптической оси.

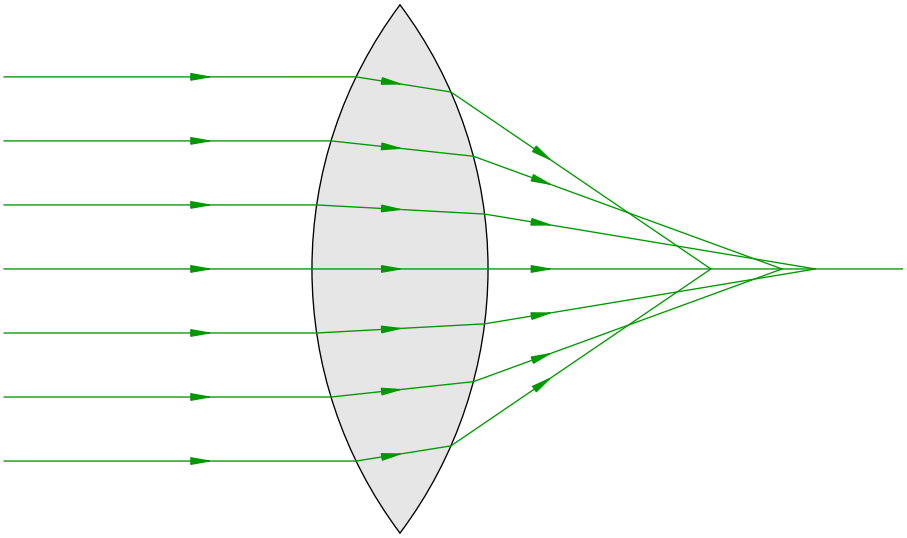


Рис. 4.17. Сферическая аберрация в двояковыпуклой линзе

Как видим, широкий пучок света *не фокусируется* линзой: чем дальше от главной оптической оси расположен падающий луч, тем ближе к линзе он пересекает главную оптическую ось после преломления. Это явление называется *сферической аберрацией* и относится к недостаткам линз — ведь хотелось бы всё же, чтобы линза сводила параллельный пучок лучей в одну точку[5].

Весьма приемлемой фокусировки можно добиться, если использовать *узкий* световой пучок, идущий вблизи главной оптической оси. Тогда сферическая аберрация почти незаметна — посмотрите на рис. 4.18.

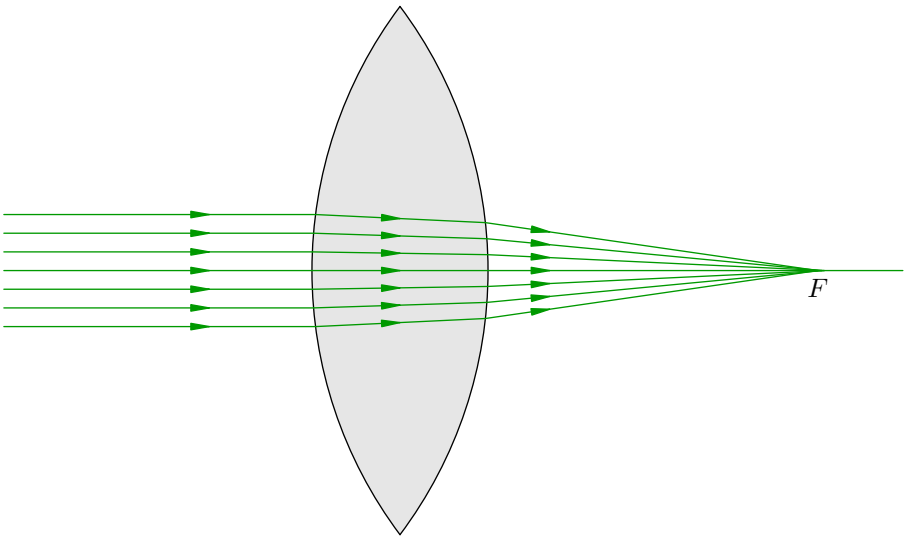


Рис. 4.18. Фокусировка узкого пучка собирающей линзой

Хорошо видно, что узкий пучок, параллельный главной оптической оси, после прохождения линзы собирается приблизительно в одной точке F. По этой причине наша линза носит название *собирающей*.

[5]Точная фокусировка широкого пучка действительно возможна, но для этого поверхность линзы должна иметь не сферическую, а более сложную форму. Шлифовать такие линзы — дело трудоёмкое и нецелесообразное. Проще уж изготавливать сферические линзы и бороться с появляющейся сферической аберрацией.

Кстати, аберрация называется сферической как раз потому, что возникает в результате замены оптимально фокусирующей сложной несферической линзы на простую сферическую.

Точка F называется *фокусом* линзы. Вообще, линза имеет два фокуса, находящиеся на главной оптической оси справа и слева от линзы. Расстояния от фокусов до линзы не обязательно равны друг другу, но мы всегда будем иметь дело с ситуациями, когда фокусы расположены симметрично относительно линзы.

4.4.2 Двояковогнутая линза

Теперь мы рассмотрим совсем другую линзу, ограниченную двумя *вогнутыми* сферическими поверхностями (рис. 4.19). Такая линза называется *двояковогнутой*. Так же, как и выше, мы проследим ход двух лучей, руководствуясь законом преломления.

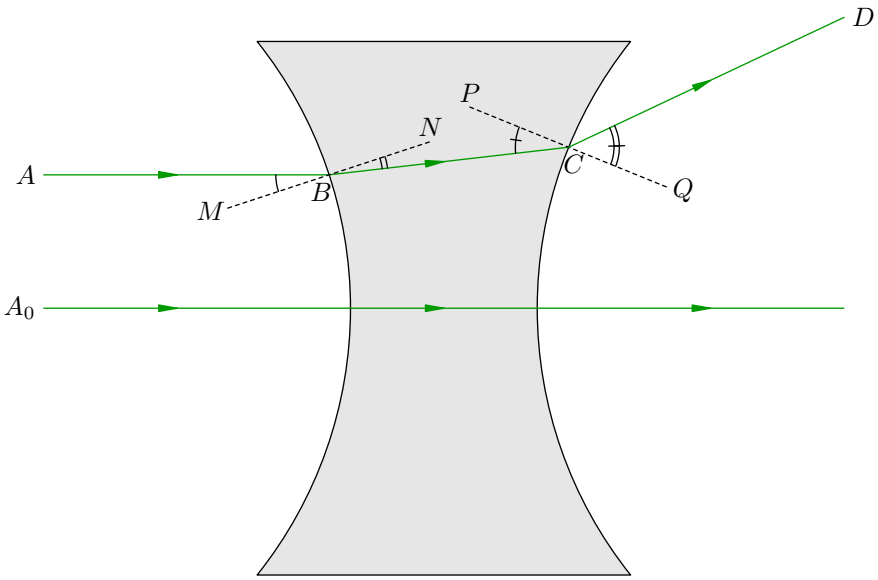


Рис. 4.19. Преломление в двояковогнутой линзе

Луч, выходящий из точки A_0 и идущий вдоль главной оптической оси, не преломляется — ведь главная оптическая ось, будучи осью симметрии линзы, перпендикулярна обеим сферическим поверхностям.

Луч AB, параллельный главной оптической оси, после первого преломления начинает удаляться от неё (так как при переходе из воздуха в стекло $\angle CBN < \angle ABM$), а после второго преломления удаляется от главной оптической оси ещё сильнее (так как при переходе из стекла в воздух $\angle QCD > \angle PCB$). Двояковогнутая линза преобразует параллельный пучок света в расходящийся пучок (рис. 4.20) и называется поэтому *рассеивающей*.

Здесь также наблюдается сферическая аберрация: продолжения расходящихся лучей не пересекаются в одной точке. Мы видим, что чем дальше от главной оптической оси расположен падающий луч, тем ближе к линзе пересекает главную оптическую ось продолжение преломлённого луча.

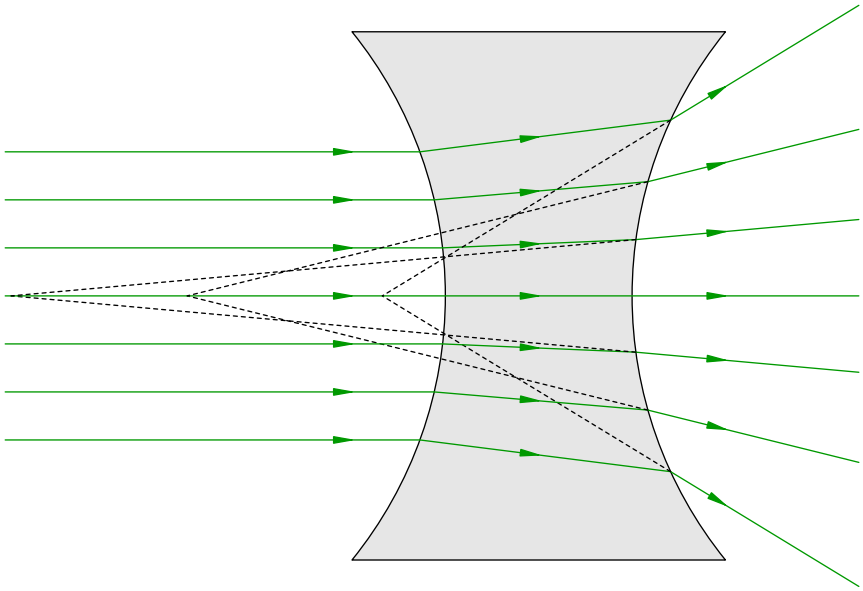


Рис. 4.20. Сферическая аберрация в двояковогнутой линзе

Как и в случае двояковыпуклой линзы, сферическая аберрация будет практически незаметна для узкого приосевого пучка (рис. 4.21). Продолжения лучей, расходящихся от линзы, пересекаются приблизительно в одной точке — в *фокусе* линзы F.

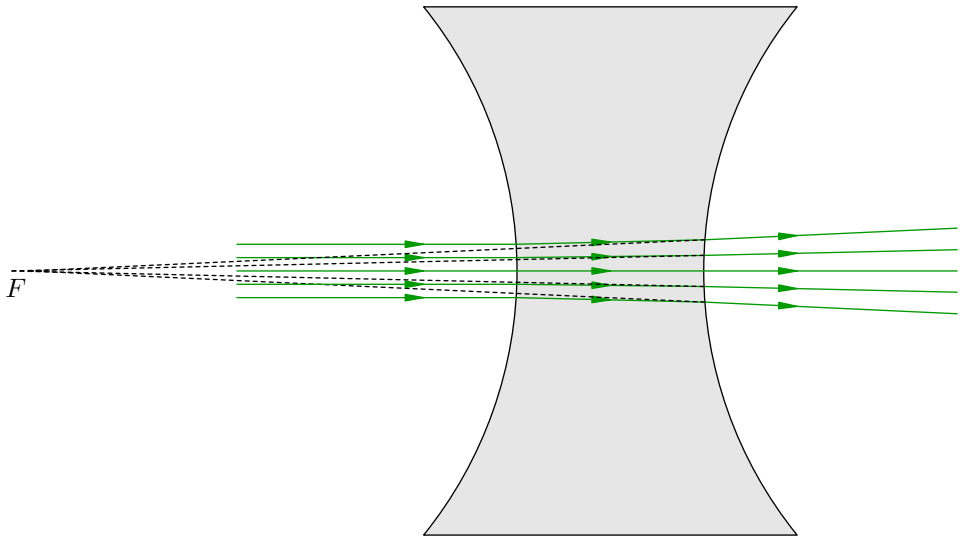


Рис. 4.21. Преломление узкого пучка в рассеивающей линзе

Если такой расходящийся пучок попадёт в наш глаз, то мы увидим за линзой светящуюся точку! Почему? Вспомните, как возникает изображение в плоском зеркале: наш мозг обладает способностью продолжать расходящиеся лучи до их пересечения и создавать в месте пересечения иллюзию светящегося объекта (так называемое мнимое изображение). Вот именно такое мнимое изображение, расположенное в фокусе линзы, мы и увидим в данном случае.

4.4.3 Виды собирающих и рассеивающих линз

Мы рассмотрели две линзы: двояковыпуклую линзу, которая является собирающей, и двояковогнутую линзу, которая является рассеивающей. Существуют и другие примеры собирающих и рассеивающих линз.

Полный набор собирающих линз представлен на рис. 4.22.

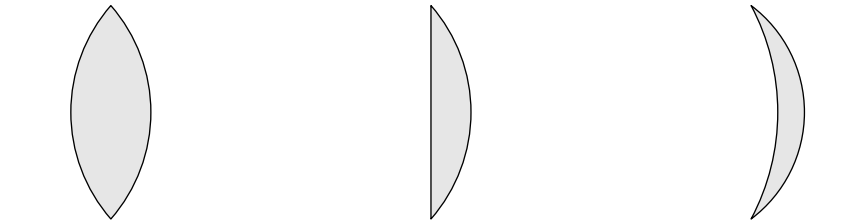


Рис. 4.22. Собирающие линзы

Помимо известной нам двояковыпуклой линзы, здесь изображены: *плосковыпуклая* линза, у которой одна из поверхностей плоская, и *вогнуто-выпуклая* линза, сочетающая вогнутую и выпуклую граничные поверхности. Обратите внимание, что у вогнуто-выпуклой линзы выпуклая поверхность в большей степени искривлена (радиус её кривизны меньше); поэтому собирающее действие выпуклой преломляющей поверхности перевешивает рассеивающее действие вогнутой поверхности, и линза в целом оказывается собирающей.

Все возможные рассеивающие линзы изображены на рис. 4.23.

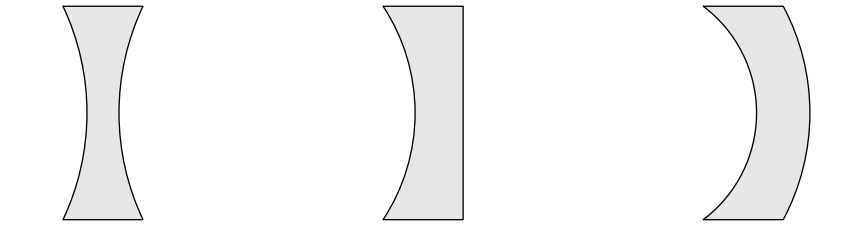


Рис. 4.23. Рассеивающие линзы

Наряду с двояковогнутой линзой мы видим *плосковогнутую* (одна из поверхностей которой плоская) и *выпукло-вогнутую* линзу. Вогнутая поверхность выпукло-вогнутой линзы искривлена в большей степени, так что рассеивающее действие вогнутой границы преобладает над собирающим действием выпуклой границы, и в целом линза оказывается рассеивающей.

Попробуйте самостоятельно построить ход лучей в тех видах линз, которые мы не рассмотрели, и убедиться, что они действительно являются собирающими или рассеивающими. Это отличное упражнение, и в нём нет ничего сложного — ровно те же самые построения, которые мы проделали выше!

4.5 Тонкие линзы. Ход лучей

Взгляните ещё раз на рисунки линз из предыдущего листка: эти линзы обладают заметной толщиной и существенной кривизной своих сферических границ. Мы намеренно рисовали такие линзы — чтобы основные закономерности хода световых лучей проявились как можно более чётко.

4.5.1 Понятие тонкой линзы

Теперь, когда эти закономерности достаточно ясны, мы рассмотрим очень полезную идеализацию, которая называется *тонкой линзой*. В качестве примера на рис. 4.24 приведена двояковыпуклая линза; точки O_1 и O_2 являются центрами её сферических поверхностей[6], R_1 и R_2 — радиусы кривизны этих поверхностей.

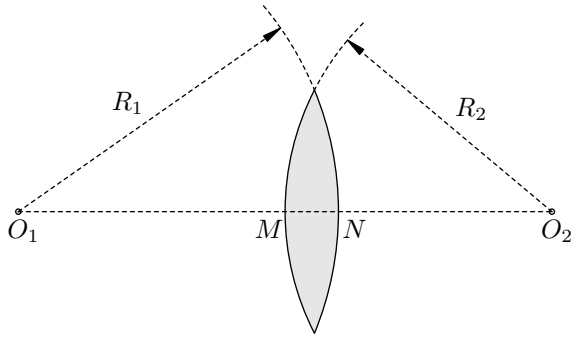


Рис. 4.24. К определению тонкой линзы

Так вот, линза считается тонкой, если её толщина MN очень мала. Нужно, правда, уточнить: мала по сравнению с чем?

Во-первых, предполагается, что $MN \ll R_1$ и $MN \ll R_2$. Тогда поверхности линзы хоть и будут выпуклыми, но могут восприниматься как «почти плоские». Этот факт нам очень скоро пригодится.

Во-вторых, $MN \ll a$, где a — характерное расстояние от линзы до интересующего нас предмета. Собственно, лишь в таком случае мы и сможем корректно говорить о «расстоянии от предмета до линзы», не уточняя, до какой именно точки линзы берётся это самое расстояние.

Мы дали определение тонкой линзы, имея в виду двояковыпуклую линзу на рис. 4.24. Это определение без каких-либо изменений переносится на все остальные виды линз. Итак: *линза является тонкой, если толщина линзы много меньше радиусов кривизны её сферических границ и расстояния от линзы до предмета.*

Условное обозначение тонкой собирающей линзы показано на рис. 4.25.

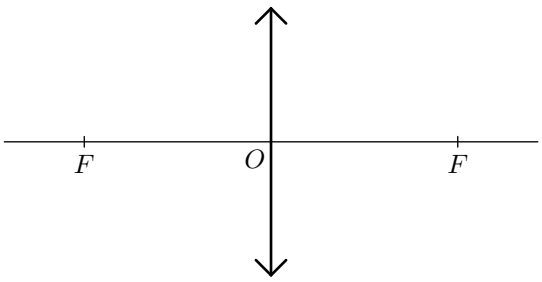


Рис. 4.25. Обозначение тонкой собирающей линзы

[6]Напомним, что прямая O_1O_2 называется *главной оптической осью* линзы.

Условное обозначение тонкой рассеивающей линзы показано на рис. 4.26.

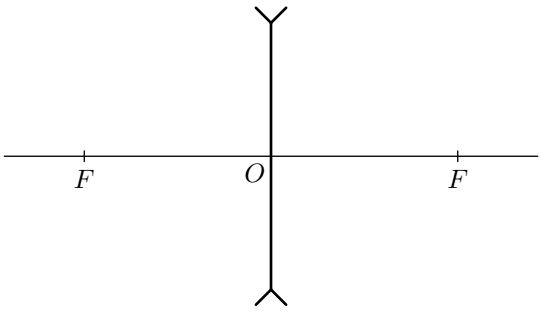


Рис. 4.26. Обозначение тонкой рассеивающей линзы

В каждом случае прямая FF — это главная оптическая ось линзы, а сами точки F — её фокусы. Оба фокуса тонкой линзы расположены симметрично относительно линзы.

4.5.2 Оптический центр и фокальная плоскость

Точки M и N, обозначенные на рис. 4.24, у тонкой линзы фактически сливаются в одну точку. Это точка O на рис. 4.25 и 4.26, называемая *оптическим центром* линзы. Оптический центр находится на пересечении линзы с её главной оптической осью.

Расстояние OF от оптического центра до фокуса называется *фокусным расстоянием* линзы. Мы будем обозначать фокусное расстояние буквой f. Величина D, обратная фокусному расстоянию, есть *оптическая сила* линзы:

$$D = \frac{1}{f}\,.$$

Оптическая сила измеряется в *диоптриях* (дптр). Так, если фокусное расстояние линзы равно 25 см, то её оптическая сила:

$$D = \frac{1}{0{,}25} = 4 \text{ дптр}.$$

Продолжаем знакомиться с новыми понятиями. Всякая прямая, проходящая через оптический центр линзы и отличная от главной оптической оси, называется *побочной оптической осью*. На рис. 4.27 изображена побочная оптическая ось — прямая OP.

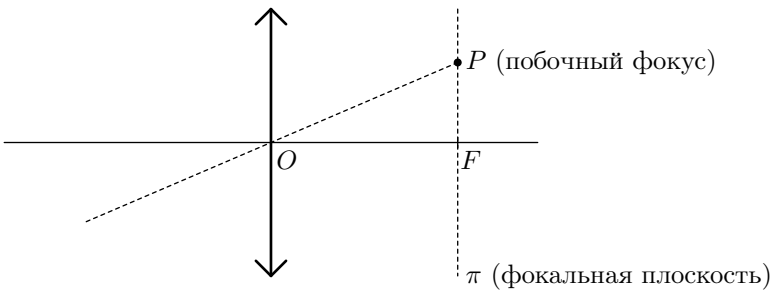


Рис. 4.27. Побочная оптическая ось, фокальная плоскость и побочный фокус

Плоскость π, проходящая через фокус перпендикулярно главной оптической оси, называется *фокальной плоскостью*. Фокальная плоскость, таким образом, параллельна плоскости линзы. Имея два фокуса, линза соответственно имеет и две фокальных плоскости, расположенных симметрично относительно линзы.

Точка P, в которой побочная оптическая ось пересекает фокальную плоскость, называется *побочным фокусом*. Собственно, каждая точка фокальной плоскости (кроме F) есть побочный фокус — мы ведь всегда сможем провести побочную оптическую ось, соединив данную точку с оптическим центром линзы. А сама точка F — фокус линзы — в связи с этим называется ещё *главным фокусом*.

То, что на рис. 4.27 изображена собирающая линза, никакой роли не играет. Понятия побочной оптической оси, фокальной плоскости и побочного фокуса совершенно аналогично определяются и для рассеивающей линзы — с заменой на рис. 4.27 собирающей линзы на рассеивающую.

Теперь мы переходим к рассмотрению хода лучей в тонких линзах. Мы будем предполагать, что лучи являются *параксиальными*, то есть образуют достаточно малые углы с главной оптической осью. Если параксиальные лучи исходят из одной точки, то после прохождения линзы преломлённые лучи или их продолжения также пересекаются в одной точке. Поэтому изображения предметов, даваемые линзой, в параксиальных лучах получаются весьма чёткими.

4.5.3 Ход луча через оптический центр

Как мы знаем из предыдущего раздела, луч, идущий вдоль главной оптической оси, не преломляется. В случае тонкой линзы оказывается, что луч, идущий вдоль побочной оптической оси, также не преломляется!

Объяснить это можно следующим образом. Вблизи оптического центра O обе поверхности линзы неотличимы от параллельных плоскостей, и луч в данном случае идёт как будто через плоскопараллельную стеклянную пластинку (рис. 4.28).

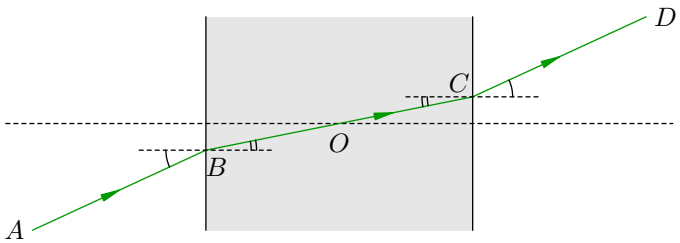


Рис. 4.28. Ход луча через оптический центр линзы

Угол преломления луча AB равен углу падения преломлённого луча BC на вторую поверхность. Поэтому второй преломлённый луч CD выходит из плоскопараллельной пластинки параллельно падающему лучу AB. Плоскопараллельная пластинка лишь смещает луч, не изменяя его направления, и это смещение тем меньше, чем меньше толщина пластинки.

Но для тонкой линзы мы можем считать, что эта толщина равна нулю. Тогда точки B, O и C фактически сольются в одну точку, и луч CD окажется просто продолжением луча AB. Вот поэтому и получается, что луч, идущий вдоль побочной оптической оси, не преломляется тонкой линзой (рис. 4.29).

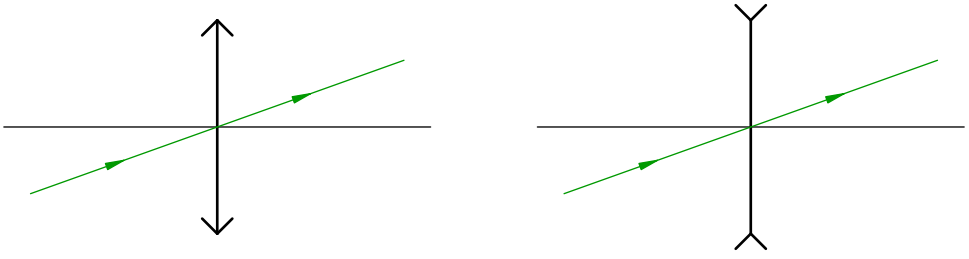


Рис. 4.29. Луч, идущий через оптический центр тонкой линзы, не преломляется

Это единственное общее свойство собирающих и рассеивающих линз. В остальном ход лучей в них оказывается различным, и дальше нам придётся рассматривать собирающую и рассеивающую линзу по отдельности.

4.5.4 Ход лучей в собирающей линзе

Как мы помним, собирающая линза называется так потому, что световой пучок, параллельный главной оптической оси, после прохождения линзы собирается в её главном фокусе (рис. 4.30).

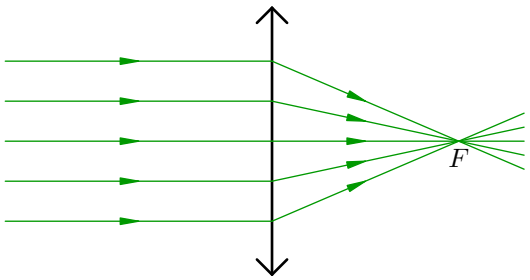


Рис. 4.30. Параллельный пучок собирается в главном фокусе

Пользуясь обратимостью световых лучей, приходим к следующему выводу: если в главном фокусе собирающей линзы находится точечный источник света, то на выходе из линзы получится световой пучок, параллельный главной оптической оси (рис. 4.31).

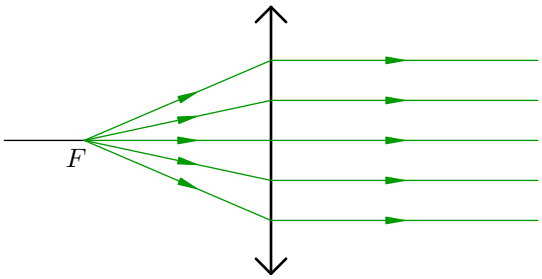


Рис. 4.31. Преломление пучка, идущего из главного фокуса

Оказывается, что пучок параллельных лучей, падающих на собирающую линзу *наклонно*, тоже соберётся в фокусе — но в побочном. Этот побочный фокус P отвечает тому лучу, который проходит через оптический центр линзы и не преломляется (рис. 4.32).

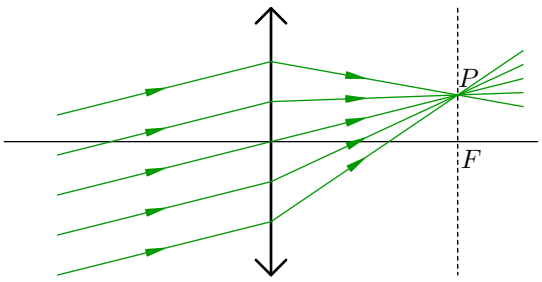


Рис. 4.32. Параллельный пучок собирается в побочном фокусе

Теперь мы можем сформулировать **правила хода лучей в собирающей линзе**. Эти правила вытекают из рисунков 4.29–4.32.

1. Луч, идущий через оптический центр линзы, не преломляется.

2. Луч, идущий параллельно главной оптической оси линзы, после преломления пойдёт через главный фокус (рис. 4.33).

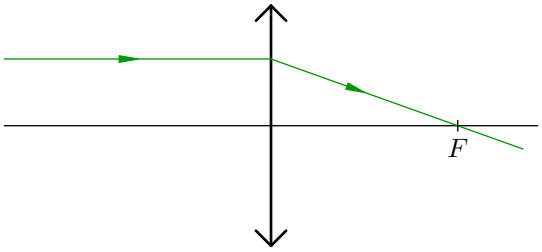


Рис. 4.33. К правилу 2

3. Если луч падает на линзу наклонно, то для построения его дальнейшего хода мы проводим побочную оптическую ось, параллельную этому лучу, и находим соответствующий побочный фокус. Вот через этот побочный фокус и пойдёт преломлённый луч (рис. 4.34).

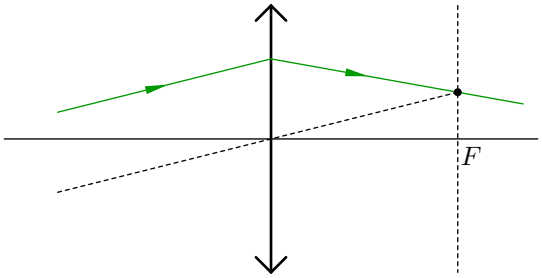


Рис. 4.34. К правилу 3

В частности, если падающий луч проходит через фокус линзы, то после преломления он пойдёт параллельно главной оптической оси.

4.5.5 Ход лучей в рассеивающей линзе

Переходим к рассеивающей линзе. Она преобразует пучок света, параллельный главной оптической оси, в расходящийся пучок, как бы выходящий из главного фокуса (рис. 4.35).

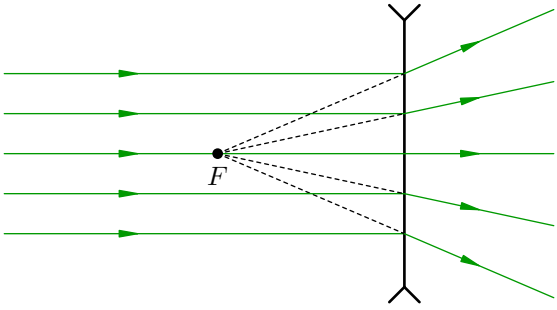


Рис. 4.35. Рассеяние параллельного пучка

Наблюдая этот расходящийся пучок, мы увидим светящуюся точку, расположенную в фокусе F позади линзы.

Если параллельный пучок падает на линзу наклонно, то после преломления он также станет расходящимся. Продолжения лучей расходящегося пучка соберутся в побочном фокусе P, отвечающем тому лучу, который проходит через через оптический центр линзы и не преломляется (рис. 4.36).

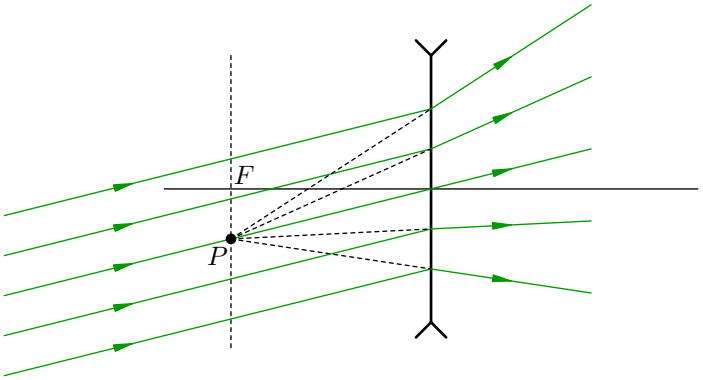


Рис. 4.36. Рассеяние наклонного параллельного пучка

Этот расходящийся пучок создаст у нас иллюзию светящейся точки, расположенной в побочном фокусе P за линзой.

Теперь мы готовы сформулировать **правила хода лучей в рассеивающей линзе**. Эти правила следуют из рисунков 4.29, 4.35 и 4.36.

1. Луч, идущий через оптический центр линзы, не преломляется.

2. Луч, идущий параллельно главной оптической оси линзы, после преломления начнёт удаляться от главной оптической оси; при этом продолжение преломлённого луча пройдёт через главный фокус (рис. 4.37).

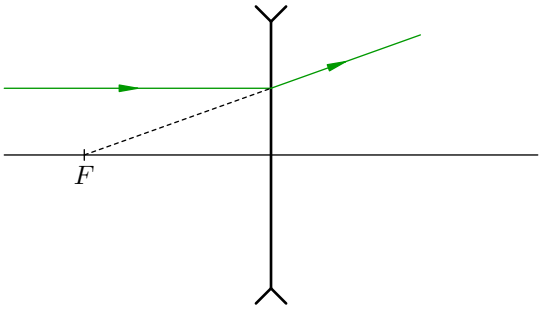


Рис. 4.37. К правилу 2

3. Если луч падает на линзу наклонно, то мы проводим побочную оптическую ось, параллельную этому лучу, и находим соответствующий побочный фокус. Преломлённый луч пойдёт так, словно он исходит из этого побочного фокуса (рис. 4.38).

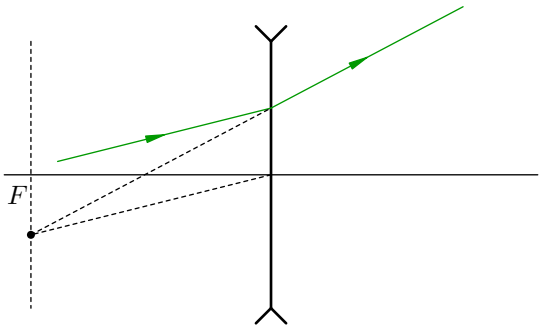


Рис. 4.38. К правилу 3

Пользуясь правилами хода лучей 1–3 для собирающей и рассеивающей линзы, мы теперь научимся самому главному — строить изображения предметов, даваемые линзами.

4.6 Тонкие линзы. Построение изображений

Правила хода лучей в тонких линзах, сформулированные в предыдущем разделе, приводят нас к важнейшему утверждению.

Теорема об изображении. Если перед линзой находится светящаяся точка S, то после преломления в линзе все лучи[7] (или их продолжения) пересекаются в одной точке S'.

Точка S' называется *изображением* точки S.

Если в точке S' пересекаются сами преломлённые лучи, то изображение называется *действительным*. Оно может быть получено на экране, так как в точке S' концентрируется энергия световых лучей.

Если же в точке S' пересекаются не сами преломлённые лучи, а их продолжения (так бывает, когда преломлённые лучи расходятся после линзы), то изображение называется *мнимым*. Его нельзя получить на экране, поскольку в точке S' не сосредоточено никакой энергии. Мнимое изображение, напомним, возникает благодаря особенности нашего мозга — достраивать расходящиеся лучи до их мнимого пересечения и видеть в этом пересечении светящуюся точку. Мнимое изображение существует лишь в нашем сознании.

Теорема об изображении служит основой построения изображений в тонких линзах. Мы докажем эту теорему как для собирающей, так и для рассеивающей линзы.

4.6.1 Собирающая линза: действительное изображение точки

Сперва рассмотрим собирающую линзу. Пусть a — расстояние от точки S до линзы, f — фокусное расстояние линзы. Имеются два принципиально разных случая: $a > f$ и $a < f$ (а также промежуточный случай $a = f$). Мы разберём эти случаи поочерёдно; в каждом из них мы обсудим свойства изображений точечного источника и протяжённого объекта.

Первый случай: $a > f$. Точечный источник света S расположен дальше от линзы, чем левая фокальная плоскость (рис. 4.39).

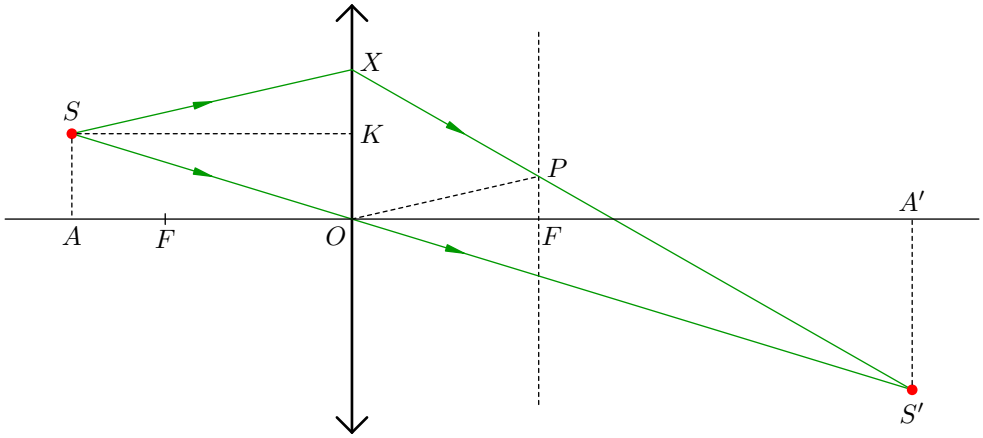


Рис. 4.39. Случай $a > f$: действительное изображение точки S

Луч SO, идущий через оптический центр, не преломляется. Мы возьмём *произвольный* луч SX, построим точку S', в которой преломлённый луч пересекается с лучом SO, а затем покажем, что положение точки S' *не зависит* от выбора луча SX (иными словами, точка S'

[7] Напомним ещё раз, что это касается не вообще всех лучей, а только *параксиальных*, то есть образующих малые углы с главной оптической осью. В предыдущем разделе мы договорились, что рассматриваем только параксиальные лучи. Лишь для них работают наши правила хода лучей сквозь тонкие линзы.

является *одной и той же* для всевозможных лучей SX). Тем самым окажется, что *все* лучи, исходящие из точки S, после преломления в линзе пересекаются в точке S', и теорема об изображении будет доказана для рассматриваемого случая $a > f$.

Точку S' мы найдём, построив дальнейший ход луча SX. Делать это мы умеем: параллельно лучу SX проводим побочную оптическую ось OP до пересечения с фокальной плоскостью в побочном фокусе P, после чего проводим преломлённый луч XP — до пересечения с лучом SO в точке S'.

Теперь будем искать расстояние b от точки S' до линзы. Мы покажем, что это расстояние выражается только через a и f, т. е. определяется лишь положением источника и свойствами линзы, и не зависит тем самым от конкретного луча SX.

Опустим перпендикуляры SA и $S'A'$ на главную оптическую ось. Проведём также SK параллельно главной оптической оси, т. е. перпендикулярно линзе. Получим три пары подобных треугольников:

$$\Delta SAO \sim \Delta S'A'O, \tag{4.6}$$
$$\Delta SXS' \sim \Delta OPS', \tag{4.7}$$
$$\Delta SXK \sim \Delta OPF. \tag{4.8}$$

В результате имеем следующую цепочку равенств (номер формулы над знаком равенства указывает, из какой пары подобных треугольников данное равенство получено).

$$\frac{AO}{OA'} \overset{(4.6)}{=} \frac{SO}{OS'} = \frac{SS' - OS'}{OS'} = \frac{SS'}{OS'} - 1 \overset{(4.7)}{=} \frac{SX}{OP} - 1 \overset{(4.8)}{=} \frac{SK}{OF} - 1. \tag{4.9}$$

Но $AO = SK = a$, $OA' = b$, $OF = f$, так что соотношение (4.9) переписывается в виде:

$$\frac{a}{b} = \frac{a}{f} - 1. \tag{4.10}$$

Отсюда находим искомое расстояние от точки S' до линзы:

$$b = \frac{af}{a - f}\,. \tag{4.11}$$

Как видим, оно и в самом деле не зависит от выбора луча SX. Следовательно, любой луч SX после преломления в линзе пройдёт через построенную нами точку S', и эта точка будет действительным изображением источника S.

Теорема об изображении в данном случае доказана.

Практическая важность теоремы об изображении состоит вот в чём. Коль скоро все лучи источника S пересекаются после линзы в одной точке — его изображении S' — то для построения изображения достаточно взять *два наиболее удобных* луча. Какие именно?

Если источник S не лежит на главной оптической оси, то в качестве удобных лучей годятся следующие:

- луч, идущий через оптический центр линзы — он не преломляется;
- луч, параллельный главной оптической оси — после преломления он идёт через фокус.

Построение изображения с помощью этих лучей показано на рис. 4.40.

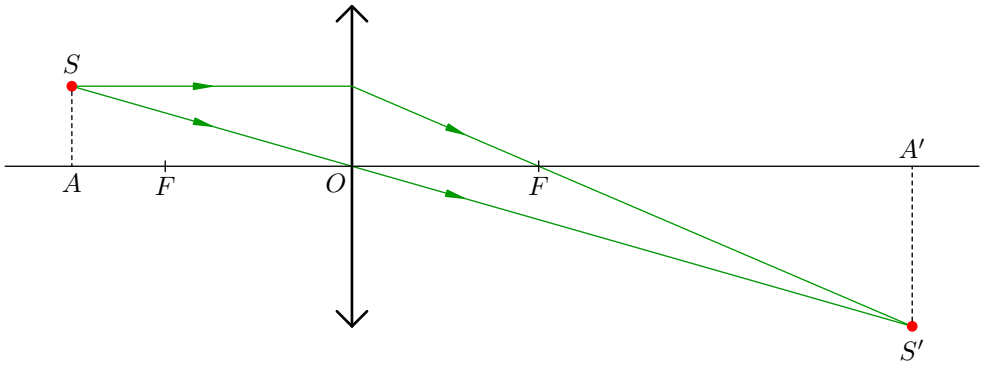


Рис. 4.40. Построение изображения точки S, не лежащей на главной оптической оси

Если же точка S лежит на главной оптической оси, то удобный луч остаётся лишь один — идущий вдоль главной оптической оси. В качестве второго луча приходится брать «неудобный» (рис. 4.41).

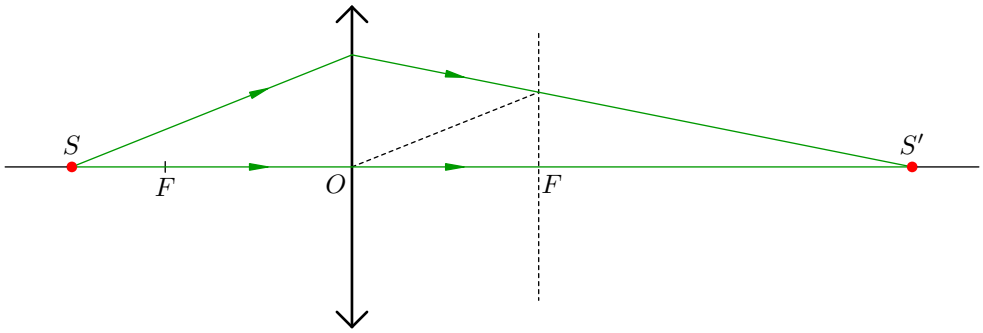


Рис. 4.41. Построение изображения точки S, лежащей на главной оптической оси

Посмотрим ещё раз на выражение (4.10). Его можно записать в несколько ином виде, более симпатичном и запоминающемся. Перенесём сначала единицу влево:

$$1 + \frac{a}{b} = \frac{a}{f}\,.$$

Теперь разделим обе части этого равенства на a:

$$\frac{1}{a} + \frac{1}{b} = \frac{1}{f}\,. \tag{4.12}$$

Соотношение (4.12) называется *формулой тонкой линзы* (или просто *формулой линзы*). Пока что формула линзы получена для случая собирающей линзы и для $a > f$. В дальнейшем мы выведем модификации этой формулы для остальных случаев.

Теперь вернёмся к соотношению (4.11). Его важность не исчерпывается тем, что оно доказывает теорему об изображении. Мы видим также, что b не зависит от расстояния SA (рис. 4.39, 4.40) между источником S и главной оптической осью!

Это означает, что какую бы точку M отрезка SA мы ни взяли, её изображение будет находиться на одном и том же расстоянии b от линзы. Оно будет лежать на отрезке $S'A'$ — а именно, на пересечении отрезка $S'A'$ с лучом MO, который пойдёт сквозь линзу без преломления. В частности, изображением точки A будет точка A'.

Тем самым мы установили важный факт: *изображением отрезка SA служит отрезок $S'A'$*. Отныне исходный отрезок, изображение которого нас интересует, мы называем *предметом* и обозначаем на рисунках красной стрелочкой. Направление стрелки нам понадобится для того, чтобы следить — прямым или перевёрнутым получается изображение.

4.6.2 Собирающая линза: действительное изображение предмета

Перейдём к рассмотрению изображений предметов. Напомним, что пока мы находимся в рамках случая $a > f$. Здесь можно выделить три характерных ситуации.

1. $f < a < 2f$. Изображение предмета является действительным, перевёрнутым, увеличенным (рис. 4.42; двойной фокус обозначен $2F$). Из формулы линзы следует, что в этом случае будет $b > 2f$ (почему?).

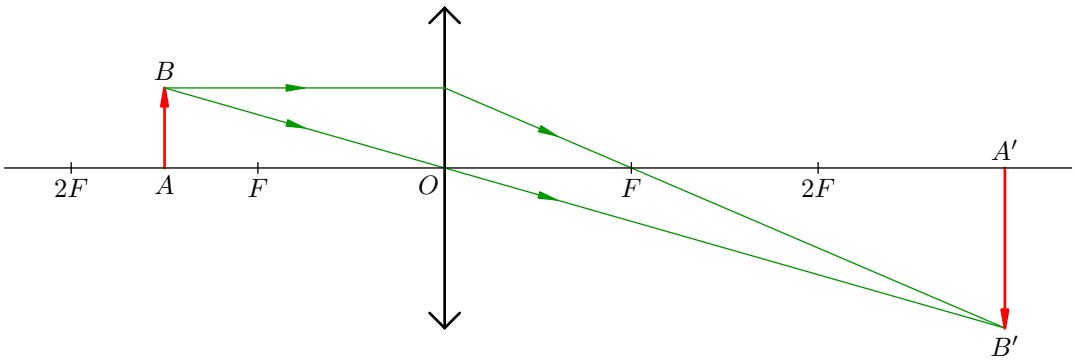


Рис. 4.42. $f < a < 2f$: изображение действительное, перевёрнутое, увеличенное

Такая ситуация реализуется, например, в диапроекторах и киноаппаратах — эти оптические приборы дают на экране увеличенное изображение того, что находится на плёнке. Если вам доводилось показывать слайды, то вы знаете, что слайд нужно вставлять в проектор перевёрнутым — чтобы изображение на экране выглядело правильно, а не получилось вверх ногами.

Отношение размера изображения к размеру предмета называется *линейным увеличением* линзы и обозначается Γ (это заглавная греческая «гамма»):

$$\Gamma = \frac{A'B'}{AB}\,.$$

Из подобия треугольников ABO и $A'B'O$ получим:

$$\Gamma = \frac{A'O}{AO} = \frac{b}{a}\,. \tag{4.13}$$

Формула (4.13) применяется во многих задачах, где фигурирует линейное увеличение линзы.

2. $a = 2f$. Из формулы (4.11) находим, что и $b = 2f$. Линейное увеличение линзы согласно (4.13) равно единице, т. е. размер изображения равен размеру предмета (рис. 4.43).

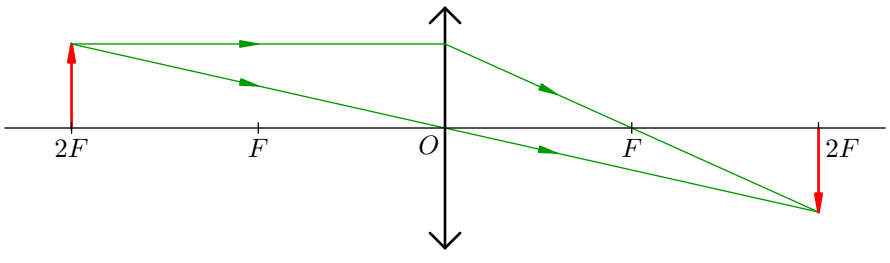


Рис. 4.43. $a = 2f$: размер изображения равен размеру предмета

3. $a > 2f$. В этом случае из формулы линзы следует, что $b < 2f$ (почему?). Линейное увеличение линзы будет меньше единицы — изображение действительное, перевёрнутое, уменьшенное (рис. 4.44).

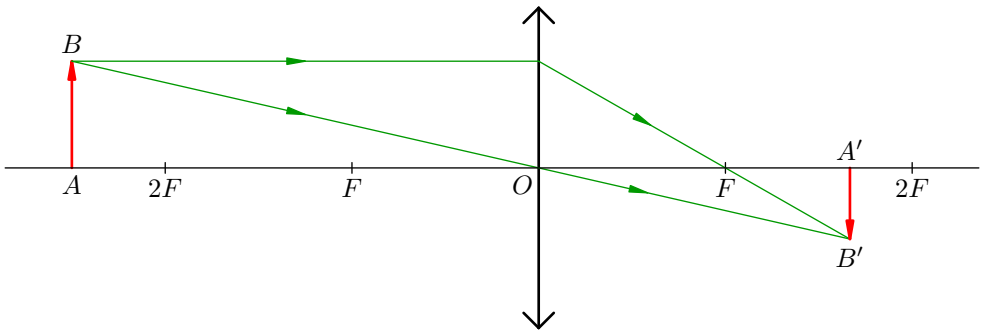


Рис. 4.44. $a > 2f$: изображение действительное, перевёрнутое, уменьшенное

Данная ситуация является обычной для многих оптических приборов: фотоаппаратов, биноклей, телескопов — словом, тех, в которых получают изображения удалённых объектов. По мере удаления предмета от линзы его изображение уменьшается в размерах и приближается к фокальной плоскости.

Рассмотрение первого случая $a > f$ нами полностью закончено. Переходим ко второму случаю. Он уже не будет столь объёмным.

4.6.3 Собирающая линза: мнимое изображение точки

Второй случай: $a < f$. Точечный источник света S расположен между линзой и фокальной плоскостью (рис. 4.45).

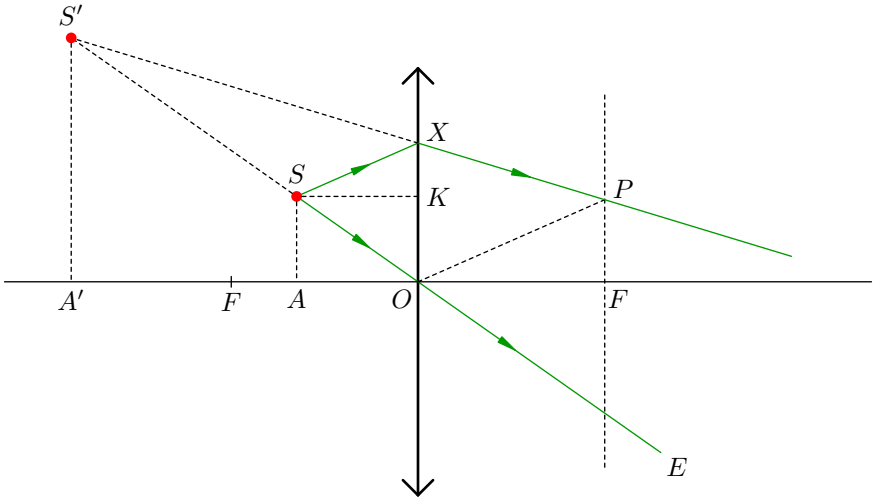


Рис. 4.45. Случай $a < f$: мнимое изображение точки

Наряду с лучом SO, идущим без преломления, мы снова рассматриваем произвольный луч SX. Однако теперь на выходе из линзы получаются два расходящихся луча OE и XP. Наш глаз продолжит эти лучи до пересечения в точке S'.

Теорема об изображении утверждает, что точка S' будет одной и той же для всех лучей SX, исходящих из точки S. Мы опять докажем это с помощью трёх пар подобных треугольников:

$$\Delta SAO \sim \Delta S'A'O, \quad \Delta SXS' \sim \Delta OPS', \quad \Delta SXK \sim \Delta OPF.$$

Снова обозначая через b расстояние от S' до линзы, имеем соответствующую цепочку равенств (вы уже без труда в ней разберётесь):

$$\frac{a}{b}=\frac{AO}{A'O}=\frac{SO}{S'O}=\frac{S'O-S'S}{S'O}=1-\frac{S'S}{S'O}=1-\frac{SX}{OP}=1-\frac{SK}{OF}=1-\frac{a}{f}\,. \tag{4.14}$$

Отсюда

$$b=\frac{fa}{f-a}\,. \tag{4.15}$$

Величина b не зависит от луча SX, что и доказывает теорему об изображении для нашего случая $a<f$. Итак, S' — мнимое изображение источника S.

Если точка S не лежит на главной оптической оси, то для построения изображения S' удобнее всего брать луч, идущий через оптический центр, и луч, параллельный главной оптической оси (рис. 4.46).

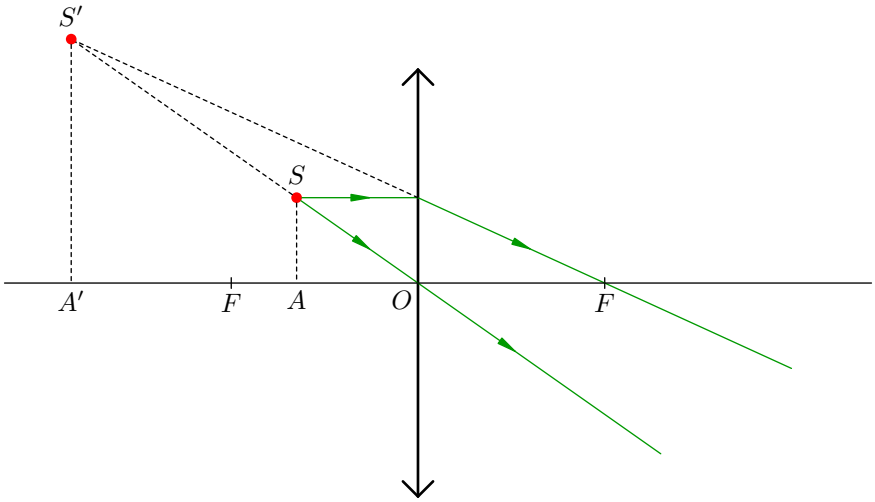


Рис. 4.46. Построение изображения точки S, не лежащей на главной оптической оси

Ну а если точка S лежит на главной оптической оси, то деваться некуда — придётся довольствоваться лучом, падающим на линзу наклонно (рис. 4.47).

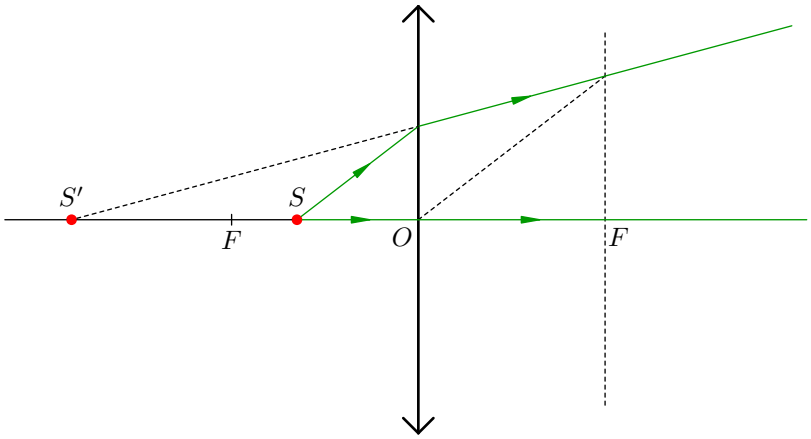


Рис. 4.47. Построение изображения точки S, лежащей на главной оптической оси

Соотношение (4.14) приводит нас к варианту формулы линзы для рассматриваемого случая $a<f$. Сначала переписываем это соотношение в виде:

$$1-\frac{a}{b}=\frac{a}{f}\,,$$

а затем делим обе части полученного равенства на a:

$$\frac{1}{a} - \frac{1}{b} = \frac{1}{f}\,. \tag{4.16}$$

Сравнивая (4.12) и (4.16), мы видим небольшую разницу: перед слагаемым $1/b$ стоит знак плюс, если изображение действительное, и знак минус, если изображение мнимое.

Величина b, вычисляемая по формуле (4.15), не зависит также от расстояния SA между точкой S и главной оптической осью. Как и выше (вспомните рассуждение с точкой M), это означает, что изображением отрезка SA на рис. 4.47 будет отрезок $S'A'$.

4.6.4 Собирающая линза: мнимое изображение предмета

Учитывая это, мы легко строим изображение предмета, находящегося между линзой и фокальной плоскостью (рис. 4.48). Оно получается мнимым, прямым и увеличенным.

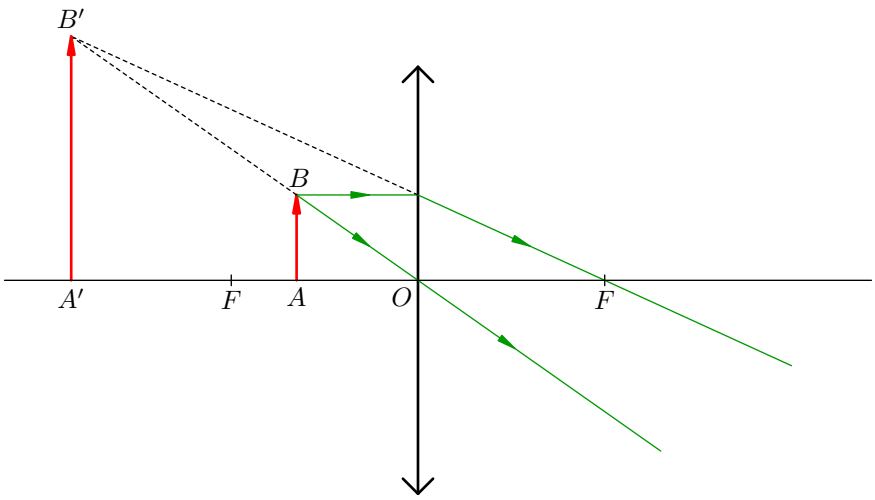


Рис. 4.48. $a < f$: изображение мнимое, прямое, увеличенное

Такое изображение вы наблюдаете, когда разглядываете мелкий предмет в увеличительное стекло — лупу.

Случай $a < f$ полностью разобран. Как видите, он качественно отличается от нашего первого случая $a > f$. Это не удивительно — ведь между ними лежит промежуточный «катастрофический» случай $a = f$.

4.6.5 Собирающая линза: предмет в фокальной плоскости

Промежуточный случай: $a = f$. Источник света S расположен в фокальной плоскости линзы (рис. 4.49).

Как мы помним из предыдущего раздела, лучи параллельного пучка после преломления в собирающей линзе пересекутся в фокальной плоскости — а именно, в главном фокусе, если пучок падает перпендикулярно линзе, и в побочном фокусе при наклонном падении пучка. Воспользовавшись обратимостью хода лучей, мы заключаем, что

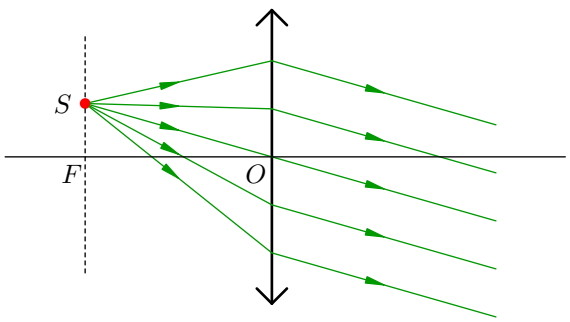


Рис. 4.49. $a = f$: изображение отсутствует

все лучи источника S, расположенного в фокальной плоскости, после выхода из линзы пойдут параллельно друг другу.

Где же изображение точки S? *Изображения нет.* Впрочем, никто не запрещает нам считать, что параллельные лучи пересекаются в бесконечно удалённой точке. Тогда теорема об изображении сохраняет свою силу и в данном случае — изображение S' находится *на бесконечности.*

Соответственно, если предмет целиком расположен в фокальной плоскости, изображение этого предмета будет находиться на бесконечности (или, что то же самое, будет отсутствовать).

Итак, мы полностью рассмотрели построение изображений в собирающей линзе.

4.6.6 Рассеивающая линза: мнимое изображение точки

К счастью, здесь нет такого разнообразия ситуаций, как для собирающей линзы. Характер изображения не зависит от того, на каком расстоянии предмет находится от рассеивающей линзы, так что случай тут будет один-единственный.

Снова берём луч SO и произвольный луч SX (рис. 4.50). На выходе из линзы имеем два расходящихся луча OE и XY, которые наш глаз достраивает до пересечения в точке S'.

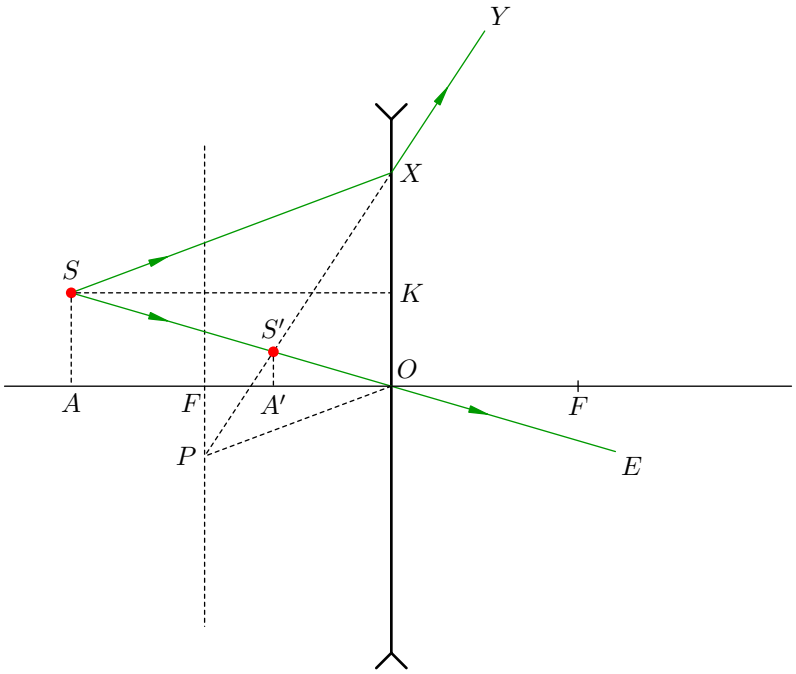


Рис. 4.50. Мнимое изображение точки S в рассеивающей линзе

Нам снова предстоит доказать теорему об изображении — о том, что точка S' будет одной и той же для всех лучей SX. Действуем с помощью всё тех же трёх пар подобных треугольников:

$$\Delta SAO \sim \Delta S'A'O, \quad \Delta SXS' \sim \Delta OPS', \quad \Delta SXK \sim \Delta OPF.$$

Имеем:

$$\frac{a}{b} = \frac{AO}{A'O} = \frac{SO}{S'O} = \frac{SS' + S'O}{S'O} = \frac{SS'}{S'O} + 1 = \frac{SX}{OP} + 1 = \frac{SK}{OF} + 1 = \frac{a}{f} + 1. \tag{4.17}$$

Отсюда

$$b = \frac{af}{a+f}\,. \tag{4.18}$$

Величина b не зависит от луча SX, поэтому продолжения всех преломлённых лучей XY пересекутся в точке S' — мнимом изображении точки S. Теорема об изображении тем самым полностью доказана.

Вспомним, что для собирающей линзы мы получили аналогичные формулы (4.11) и (4.15). В случае $a = f$ их знаменатель обращался в нуль (изображение уходило на бесконечность), и поэтому данный случай разграничивал принципиально разные ситуации $a > f$ и $a < f$.

А вот у формулы (4.18) знаменатель не обращается в нуль ни при каком a. Стало быть, для рассеивающей линзы не существует качественно разных ситуаций расположения источника — случай тут, как мы и сказали выше, имеется только один.

Если точка S не лежит на главной оптической оси, то для построения её изображения удобны два луча: один идёт через оптический центр, другой — параллельно главной оптической оси (рис. 4.51).

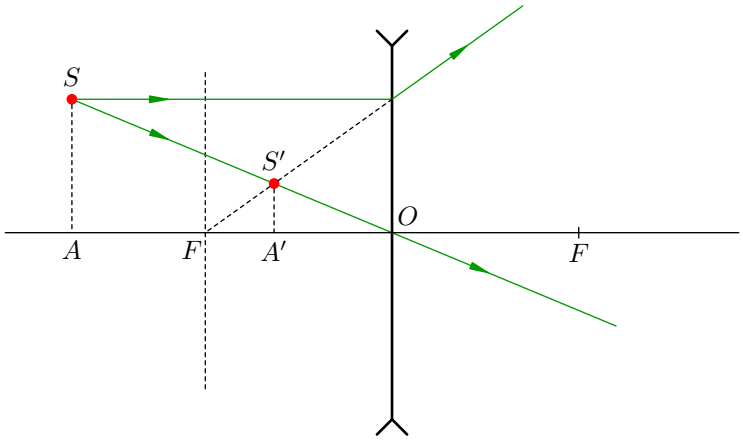


Рис. 4.51. Построение изображения точки S, не лежащей на главной оптической оси

Если же точка S лежит на главной оптической оси, то второй луч приходится брать произвольным (рис. 4.52).

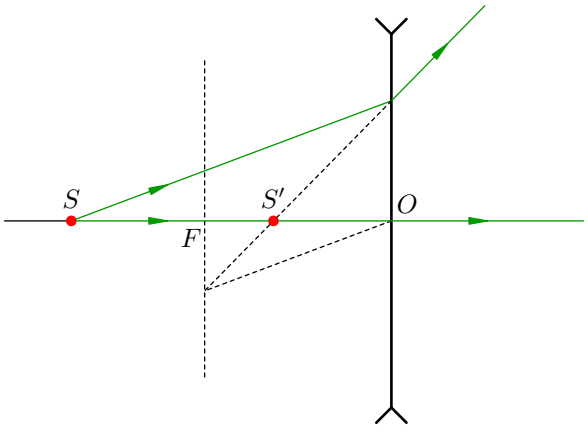


Рис. 4.52. Построение изображения точки S, лежащей на главной оптической оси

Соотношение (4.18) даёт нам ещё один вариант формулы линзы. Сначала перепишем:

$$1 - \frac{a}{b} = -\frac{a}{f}\,,$$

а потом разделим обе части полученного равенства на a:

$$\frac{1}{a} - \frac{1}{b} = -\frac{1}{f}\,. \tag{4.19}$$

Так выглядит формула линзы для рассеивающей линзы.

Три формулы линзы (4.12), (4.16) и (4.19) можно записать единообразно:

$$\frac{1}{a}+\frac{1}{b}=\frac{1}{f}\,,$$

если соблюдать следующую договорённость о знаках:

- для мнимого изображения величина b считается отрицательной;
- для рассеивающей линзы величина f считается отрицательной.

Это очень удобно и охватывает все рассмотренные случаи.

4.6.7 Рассеивающая линза: мнимое изображение предмета

Величина b, вычисляемая по формуле (4.18), опять-таки не зависит от расстояния SA между точкой S и главной оптической осью. Это снова даёт нам возможность построить изображение предмета AB, которое на сей раз получается мнимым, прямым и уменьшенным (рис. 4.53).

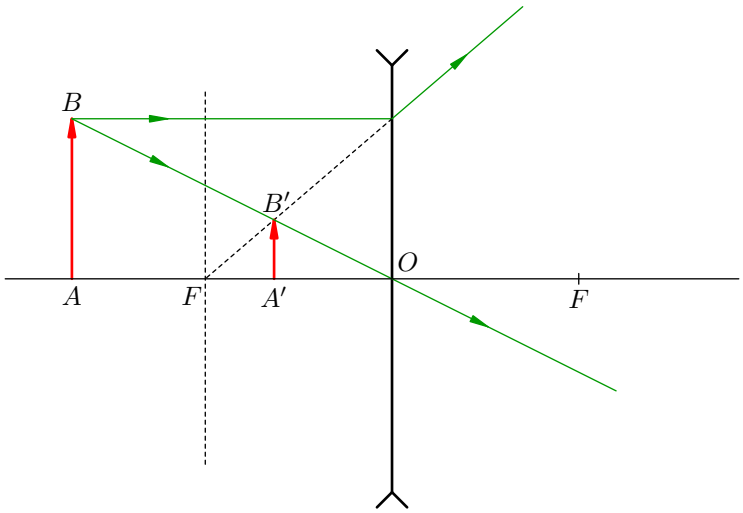


Рис. 4.53. Изображение мнимое, прямое, уменьшенное

4.7 Глаз человека

Глаз — удивительно сложная и совершенная оптическая система, созданная природой. Сейчас мы в общих чертах узнаем, как функционирует человеческий глаз. Впоследствии это позволит нам лучше понять принципы работы оптических приборов; да, кроме того, это интересно и важно само по себе.

4.7.1 Строение глаза

Мы ограничимся рассмотрением лишь самых основных элементов глаза. Они показаны[8] на рис. 4.54 (правый глаз, вид сверху).

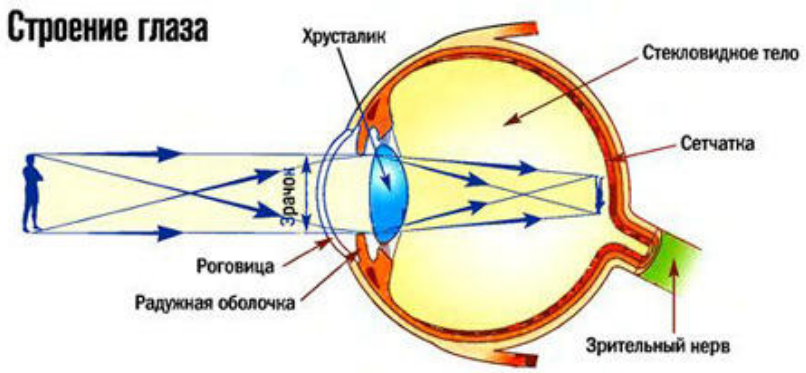


Рис. 4.54. Строение глаза

Лучи, идущие от предмета (в данном случае предметом является фигура человека), попадают на *роговицу* — переднюю прозрачную часть защитной оболочки глаза. Преломляясь в роговице и проходя сквозь *зрачок* (отверстие в *радужной оболочке* глаза), лучи испытывают вторичное преломление в *хрусталике*. Хрусталик является собирающей линзой с переменным фокусным расстоянием; он может менять свою кривизну (и тем самым фокусное расстояние) под действием специальной глазной мышцы.

Преломляющая система роговицы и хрусталика формирует на *сетчатке* изображение предмета. Сетчатка состоит из светочувствительных палочек и колбочек — нервных окончаний *зрительного нерва*. Падающий свет вызывает раздражение этих нервных окончаний, и зрительный нерв передаёт соответствующие сигналы в мозг. Так в нашем сознании формируются образы предметов — мы *видим* окружающий мир.

Ещё раз взгляните на рис. 4.54 и обратите внимание, что изображение разглядываемого предмета на сетчатке — действительное, перевёрнутое и уменьшенное. Так получается потому, что предметы, рассматриваемые глазом без напряжения, расположены за двойным фокусом системы роговица-хрусталик (помните случай $a > 2f$ для собирающей линзы?).

То, что изображение является действительным, понятно: на сетчатке должны пересекаться сами лучи (а не их продолжения), концентрируя световую энергию и вызывая раздражения палочек и колбочек.

Насчёт того, что изображение является уменьшенным, тоже вопросов не возникает. А каким же ему ещё быть? Диаметр глаза равен примерно 25 мм, а поле нашего зрения попадают предметы куда большего размера. Естественно, глаз отображает их на сетчатке в уменьшенном виде.

Но вот как быть с тем, что изображение на сетчатке является перевёрнутым? Почему же тогда мы видим мир не вверх ногами? Здесь подключается корректирующее действие нашего

[8] Изображение заимствовано с сайта «Детская энциклопедия What This».

мозга. Оказывается, кора головного мозга, обрабатывая изображение на сетчатке, переворачивает картинку обратно! Это установленный факт, проверенный экспериментами.

Как мы уже сказали, хрусталик — это собирающая линза с переменным фокусным расстоянием. Но зачем хрусталику менять своё фокусное расстояние?

4.7.2 Аккомодация

Представьте себе, что вы смотрите на приближающегося к вам человека. Вы всё время чётко его видите. Каким образом глазу удаётся это обеспечивать?

Чтобы лучше понять суть вопроса, давайте вспомним формулу линзы:

$$\frac{1}{a}+\frac{1}{b}=\frac{1}{f}\,.$$

В данном случае a — это расстояние от глаза до предмета, b — расстояние от хрусталика до сетчатки, f — фокусное расстояние оптической системы глаза. Величина b является неизменной, поскольку это геометрическая характеристика глаза. Следовательно, чтобы формула линзы оставалась справедливой, вместе с расстоянием a до разглядываемого предмета должно меняться и фокусное расстояние f.

Например, если предмет приближается к глазу, то a уменьшается, поэтому и f должно уменьшаться. Для этого глазная мышца деформирует хрусталик, делая его более выпуклым и уменьшая тем самым фокусное расстояние до нужной величины. При удалении предмета, наоборот, кривизна хрусталика уменьшается, а фокусное расстояние возрастает.

Описанный механизм самонастройки глаза называется аккомодацией. Итак, *аккомодация — это способность глаза отчётливо видеть предметы на различных расстояниях*. В процессе аккомодации кривизна хрусталика меняется так, что изображение предмета всегда оказывается на сетчатке.

Аккомодация глаза совершается бессознательно и очень быстро. Эластичный хрусталик может легко менять свою кривизну в определённых пределах. Этим естественным пределам деформации хрусталика отвечает *область аккомодации* — диапазон расстояний, на которых глаз способен чётко видеть предметы. Область аккомодации характеризуется своими границами — дальней и ближней точками аккомодации.

Дальняя точка аккомодации[9] — это точка нахождения предмета, изображение которого на сетчатке получается при расслабленной глазной мышце, т. е. когда хрусталик не деформирован.

Ближняя точка аккомодации[10] — это точка нахождения предмета, изображение которого на сетчатке получается при наибольшем напряжении глазной мышцы, т. е. при максимально возможной деформации хрусталика.

Дальняя точка аккомодации нормального глаза находится на бесконечности: в ненапряжённом состоянии глаз фокусирует параллельные лучи на сетчатке (рис. 4.55, слева). Иными словами, *фокусное расстояние оптической системы нормального глаза при недеформированном хрусталике равно расстоянию от хрусталика до сетчатки*.

Ближняя точка аккомодации нормального глаза расположена на некотором расстоянии $d_{\min}$ от него (рис. 4.55, справа; хрусталик максимально деформирован). Это расстояние с возрастом увеличивается. Так, у десятилетнего ребёнка $d_{\min}\approx 7$ см; в возрасте 30 лет $d_{\min}\approx 15$ см; к 45 годам ближняя точка аккомодации находится уже на расстоянии 20–25 см от глаза.

[9] Другое название — *дальняя точка ясного видения*.
[10] Другое название — *ближняя точка ясного видения*.

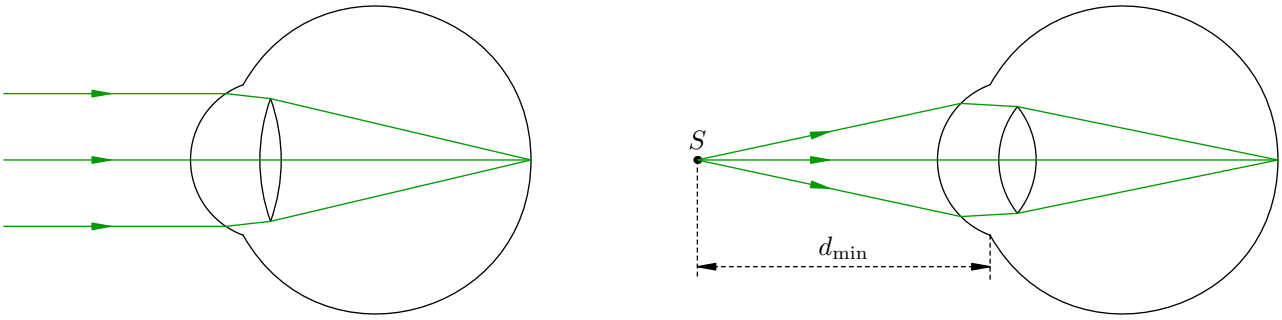


Рис. 4.55. Дальняя и ближняя точки аккомодации нормального глаза

Теперь мы переходим к простому, но очень важному понятию угла зрения. Оно является ключевым для понимания принципов работы различных оптических приборов.

4.7.3 Угол зрения

Когда мы хотим получше рассмотреть предмет, мы приближаем его к глазам. Чем ближе предмет, тем больше его деталей оказываются различимыми. Почему так получается?

Давайте посмотрим на рис. 4.56. Пусть стрелка AB — рассматриваемый предмет, O — оптический центр глаза. Проведём лучи AO и BO (которые не преломляются) и получим на сетчатке изображение нашего предмета — красную изогнутую стрелочку.

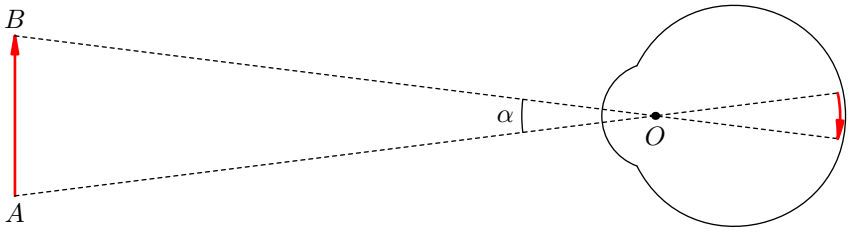


Рис. 4.56. Предмет далеко, угол зрения мал

Угол $\alpha = \angle AOB$ называется *углом зрения*. Если предмет расположен далеко от глаза, то угол зрения мал, и размер изображения на сетчатке также оказывается малым.

Но если предмет расположить ближе, то угол зрения увеличивается (рис. 4.57). Соответственно увеличивается и размер изображения на сетчатке. Сравните рис. 4.56 и рис. 4.57 — во втором случае изогнутая стрелочка оказывается явно длиннее!

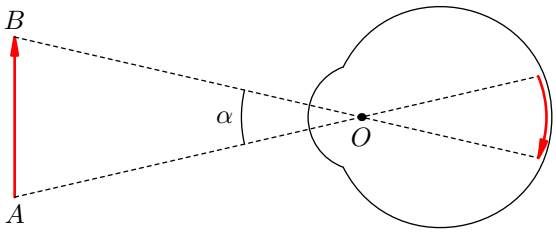


Рис. 4.57. Предмет близко, угол зрения велик

Размер изображения на сетчатке — вот что важно для подробного разглядывания предмета. Сетчатка, напомним, состоит из нервных окончаний зрительного нерва. Поэтому чем крупнее изображение на сетчатке, тем больше нервных окончаний раздражается идущими от предмета световыми лучами, тем больший поток информации о предмете направляется по зрительному

нерву в мозг — и, следовательно, тем больше подробностей мы различаем, тем лучше мы видим предмет!

Ну а размер изображения на сетчатке, как мы уже убедились из рисунков 4.56 и 4.57, напрямую зависит от угла зрения: чем больше угол зрения, тем крупнее изображение. Поэтому вывод: *увеличивая угол зрения, мы различаем больше подробностей рассматриваемого объекта.*

Вот почему мы одинаково плохо видим как мелкие объекты, пусть и находящиеся рядом, так и крупные объекты, но расположенные далеко. В обоих случаях угол зрения мал, и на сетчатке раздражается небольшое число нервных окончаний. Известно, кстати, что если угол зрения меньше одной угловой минуты (1/60 градуса), то раздражается лишь одно нервное окончание. В этом случае мы воспринимаем объект просто как точку, лишённую деталей.

4.7.4 Расстояние наилучшего зрения

Итак, приближая предмет, мы увеличиваем угол зрения и различаем больше деталей. Казалось бы, оптимального качества видения мы достигнем, если расположим предмет максимально близко к глазу — в ближней точке аккомодации (в среднем это 10–15 см от глаза).

Однако мы так не поступаем. Например, читая книгу, мы держим её на расстоянии примерно 25 см. Почему же мы останавливаемся на этом расстоянии, хотя ещё имеется ресурс дальнейшего увеличения угла зрения?

Дело в том, что при достаточно близком расположении предмета хрусталик чрезмерно деформируется. Конечно, глаз ещё способен чётко видеть предмет, но при этом быстро утомляется, и мы испытываем неприятное напряжение.

Величина $d_0 = 25$ см называется *расстоянием наилучшего зрения* для нормального глаза. При таком расстоянии достигается компромисс: угол зрения уже достаточно велик, и в то же время глаз не утомляется ввиду не слишком большой деформации хрусталика. Поэтому с расстояния наилучшего зрения мы можем полноценно созерцать предмет в течении весьма долгого времени.

4.7.5 Близорукость

Напомним, что фокусное расстояние нормального глаза в расслабленном состоянии равно расстоянию от оптического центра до сетчатки. Нормальный глаз фокусирует параллельные лучи на сетчатке и поэтому может чётко видеть удалённые предметы, не испытывая напряжения.

Близорукость — это дефект зрения, при котором фокусное расстояние расслабленного глаза меньше расстояния от оптического центра до сетчатки. Близорукий глаз фокусирует параллельные лучи *перед* сетчаткой, и от этого изображения удалённых объектов оказываются размытыми (рис. 4.58; хрусталик не изображаем).

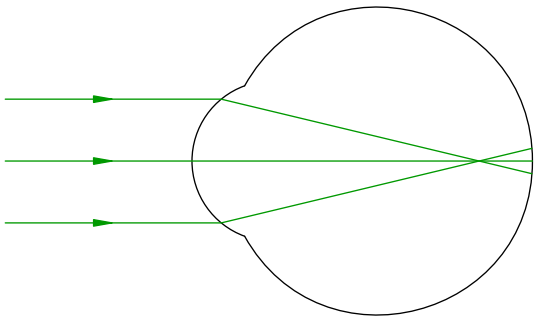


Рис. 4.58. Близорукость

Потеря чёткости изображения наступает, когда предмет находится дальше определённого расстояния. Это расстояние соответствует дальней точке аккомодации близорукого глаза. Та-

ким образом, если у человека с нормальным зрением дальняя точка аккомодации находится на бесконечности, то *у близорукого человека дальняя точка аккомодации расположена на конечном расстоянии перед ним.*

Соответственно, ближняя точка аккомодации у близорукого глаза находится ближе, чем у нормального. *Расстояние наилучшего зрения для близорукого человека меньше 25 см.*

Близорукость корректируется с помощью очков с рассеивающими линзами. Проходя через рассеивающую линзу, параллельный пучок света становится расходящимся, в результате чего изображение бесконечно удалённой точки отодвигается на сетчатку (рис. 4.59). Если при этом мысленно продолжить расходящиеся лучи, попадающие в глаз, то они соберутся в дальней точке аккомодации A.

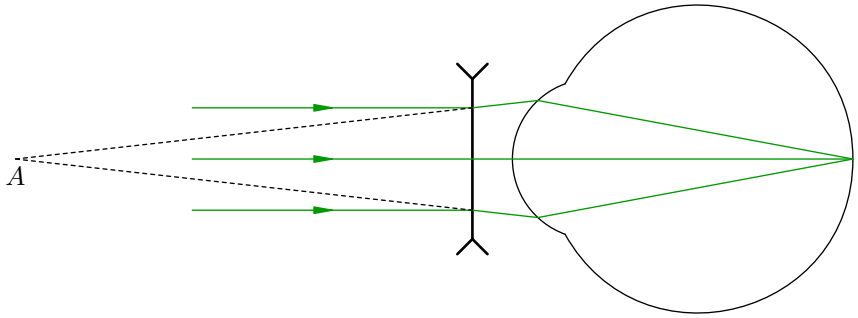


Рис. 4.59. Коррекция близорукости с помощью очков

Таким образом, *близорукий глаз, вооружённый подходящими очками, воспринимает параллельный пучок света как исходящий из дальней точки аккомодации.* Вот почему близорукий человек в очках может отчётливо рассматривать удалённые предметы без напряжения в глазах. Из рис. 4.59 мы видим также, что *фокусное расстояние подходящей линзы равно расстоянию от глаза до дальней точки аккомодации.*

4.7.6 Дальнозоркость

Дальнозоркость — это дефект зрения, при котором фокусное расстояние расслабленного глаза больше расстояния от оптического центра до сетчатки.

Дальнозоркий глаз фокусирует параллельные лучи *за* сетчаткой, отчего изображения удалённых объектов оказываются размытыми (рис. 4.60).

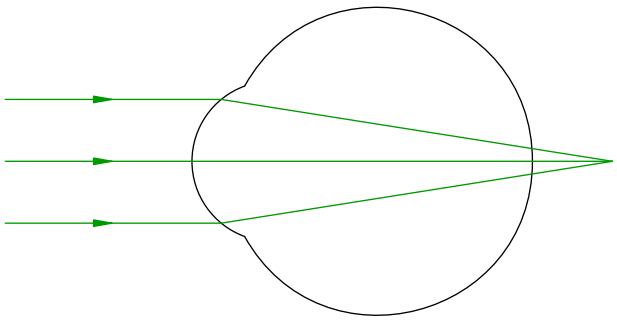


Рис. 4.60. Дальнозоркость

На сетчатке же фокусируется *сходящийся* пучок лучей. Поэтому дальняя точка аккомодации дальнозоркого глаза оказывается *мнимой*: в ней пересекаются мысленные продолжения лучей сходящегося пучка, попадающего на глаз (мы увидим это ниже на рис. 4.61).

Ближняя точка аккомодации у дальнозоркого глаза расположена дальше, чем у нормального. *Расстояние наилучшего зрения для дальнозоркого человека больше 25 см.*

Дальнозоркость корректируется с помощью очков с собирающими линзами. После прохождения собирающей линзы параллельный пучок света становится сходящимся и затем фокусируется на сетчатке (рис. 4.61).

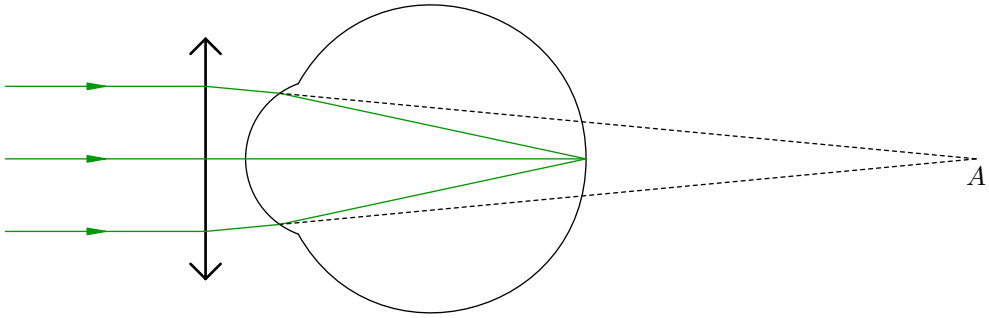


Рис. 4.61. Коррекция дальнозоркости с помощью очков

Параллельные лучи после преломления в линзе идут так, что продолжения преломлённых лучей пересекаются в дальней точке аккомодации A. Поэтому дальнозоркий человек, вооружённый подходящими очками, будет отчётливо и без напряжения рассматривать удалённые предметы. Мы также видим из рис. 4.61, что *фокусное расстояние подходящей линзы равно расстоянию от глаза до мнимой дальней точки аккомодации.*

4.8 Оптические приборы

Как мы знаем из предыдущего раздела, для более подробного разглядывания объекта нужно увеличить угол зрения. Тогда изображение объекта на сетчатке будет крупнее, и это приведёт к раздражению большего числа нервных окончаний зрительного нерва; в мозг направится большее количество визуальной информации, и мы сможем увидеть новые детали рассматриваемого объекта.

Почему угол зрения бывает малым? На то есть две причины: 1) объект сам по себе имеет малый размер; 2) объект, хотя и достаточно велик по размерам, но расположен далеко.

Оптические приборы — это приспособления для увеличения угла зрения. Для рассматривания *малых* объектов используются *лупа* и *микроскоп*. Для рассматривания *далёких* объектов применяются *зрительные трубы* (а также бинокли, телескопы и т. д.).

4.8.1 Невооружённый глаз

Начинаем с рассматривания мелких объектов невооружённым глазом. Здесь и далее глаз считается нормальным. Напомним, что нормальный глаз в ненапряжённом состоянии фокусирует на сетчатке параллельный пучок света, а расстояние наилучшего зрения для нормального глаза равно $d_0 = 25$ см.

Пусть небольшой предмет размером h находится на расстоянии наилучшего зрения d_0 от глаза (рис. 4.62). На сетчатке возникает перевёрнутое изображение предмета, но, как вы помните, это изображение затем вторично переворачивается в коре головного мозга, и в результате мы видим предмет нормально — не вверх ногами.

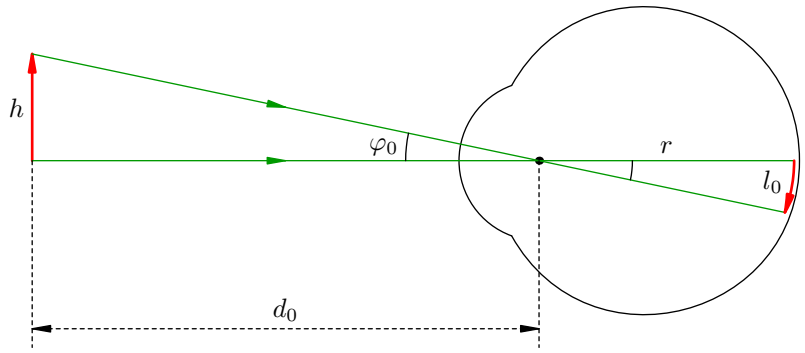


Рис. 4.62. Рассматривание мелкого предмета невооружённым глазом

Ввиду малости предмета угол зрения φ_0 также является малым. Напомним, что малый угол (в радианах) почти не отличается от своего тангенса: $\varphi_0 \approx \operatorname{tg}\varphi_0$. Поэтому:

$$\varphi_0 = \frac{h}{d_0}\,. \tag{4.20}$$

Если r — расстояние от оптического центра глаза до сетчатки, то размер изображения на сетчатке будет равен:

$$l_0 = r\varphi_0. \tag{4.21}$$

Из (4.20) и (4.21) имеем также:

$$l_0 = \frac{rh}{d_0}\,. \tag{4.22}$$

Как известно, диаметр глаза составляет около 2,5 см, так что $r/d_0 \approx 0{,}1$. Поэтому из (4.22) следует, что при рассматривании мелкого предмета невооружённым глазом изображение предмета на сетчатке примерно в 10 раз меньше самого предмета.

4.8.2 Лупа

Укрупнить изображение объекта на сетчатке можно с помощью лупы (увеличительного стекла). *Лупа* — это просто собирающая линза (или система линз); фокусное расстояние лупы обычно находится в диапазоне от 5 до 125 мм.

Предмет, разглядываемый через лупу, помещается в её фокальной плоскости (рис. 4.63). В таком случае лучи, исходящие из каждой точки предмета, после прохождения лупы становятся параллельными, и глаз фокусирует их на сетчатке, не испытывая напряжения.

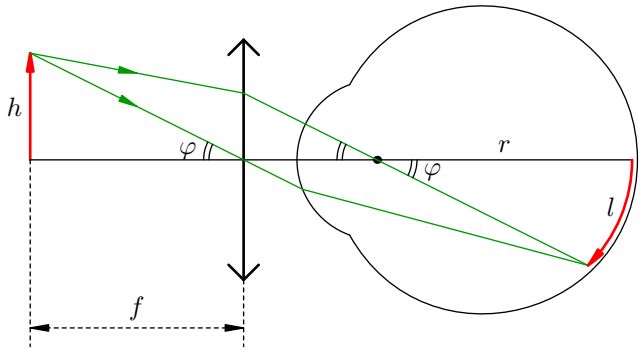


Рис. 4.63. Рассматривание предмета через лупу

Теперь, как видим, угол зрения равен φ. Он также мал и приблизительно равен своему тангенсу:

$$\varphi = \frac{h}{f}\,. \tag{4.23}$$

Размер l изображения на сетчатке теперь равен:

$$l = r\varphi, \tag{4.24}$$

или, с учётом (4.23):

$$l = \frac{rh}{f}\,. \tag{4.25}$$

Как и на рис. 4.62, красная стрелочка на сетчатке также направлена вниз. Это означает, что (с учётом вторичного переворачивания изображения нашим сознанием) в лупу мы видим неперевёрнутое изображение предмета.

Увеличение лупы — это отношение размера изображения при использовании лупы к размеру изображения при рассматривании предмета невооружённым глазом:

$$\Gamma = \frac{l}{l_0}\,. \tag{4.26}$$

Подставляя сюда выражения (4.25) и (4.22), получим:

$$\Gamma = \frac{d_0}{f}\,. \tag{4.27}$$

Например, если фокусное расстояние лупы равно 5 см, то её увеличение $\Gamma = 25/5 = 5$. При рассматривании через такую лупу объект кажется в пять раз больше, чем при рассматривании его невооружённым глазом.

Подставим также в формулу (4.26) соотношения (4.24) и (4.21):

$$\Gamma = \frac{r\varphi}{r\varphi_0} = \frac{\varphi}{\varphi_0}\,.$$

Таким образом, увеличение лупы есть *угловое увеличение*: оно равно отношению угла зрения при рассматривании объекта через лупу к углу зрения при рассматривании этого объекта невооружённым глазом.

Отметим, что увеличение лупы есть величина субъективная — ведь величина d_0 в формуле (4.27) есть расстояние наилучшего зрения для *нормального* глаза. В случае близорукого или дальнозоркого глаза расстояние наилучшего зрения будет соответственно меньше или больше.

Из формулы (4.27) следует, что увеличение лупы тем больше, чем меньше её фокусное расстояние. Уменьшение фокусного расстояния собирающей линзы достигается за счёт увеличения кривизны преломляющих поверхностей: линзу надо делать более выпуклой и тем самым уменьшать её размеры. Когда увеличение Γ достигает 40–50, размер лупы становится равным нескольким миллиметрам. При ещё меньших размерах лупы пользоваться ей станет невозможно, поэтому $\Gamma = 50$ считается верхней границей увеличения лупы.

4.8.3 Микроскоп

Во многих случаях (например, в биологии, медицине и т. д.) нужно наблюдать мелкие объекты с увеличением в несколько сотен. Лупой тут не обойдёшься, и люди прибегают к помощи *микроскопа*.

Микроскоп содержит две собирающие линзы (или две системы таких линз) — *объектив* и *окуляр*. Запомнить это просто: объектив обращён к объекту, а окуляр — к глазу (к оку).

Идея микроскопа проста. Рассматриваемый объект находится между фокусом и двойным фокусом объектива, так что объектив даёт увеличенное (действительное перевёрнутое) изображение объекта. Это изображение располагается в фокальной плоскости окуляра и затем рассматривается в окуляр как в лупу. В результате удаётся достичь итогового увеличения, гораздо большего 50.

Ход лучей в микроскопе показан на рис. 4.64.

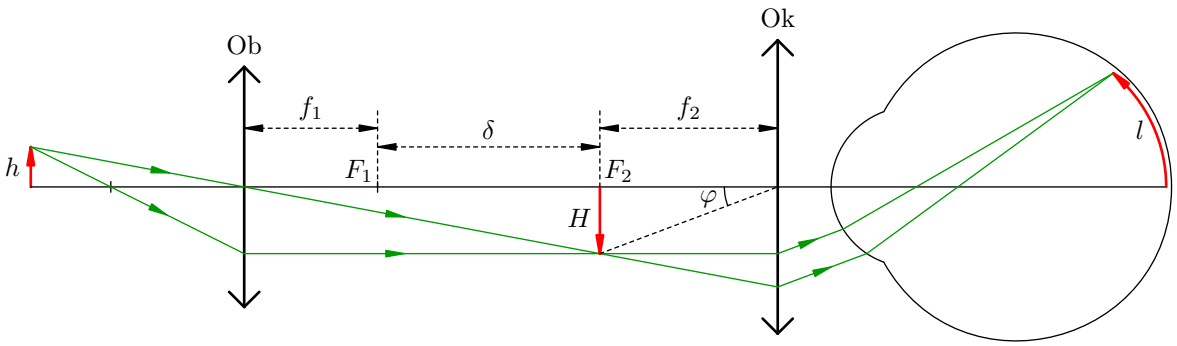


Рис. 4.64. Ход лучей в микроскопе

Обозначения на рисунке понятны: f_1 — фокусное расстояние объектива Ob; f_2 — фокусное расстояние окуляра Ok; h — размер объекта; H — размер изображения объекта, даваемого объективом. Расстояние $\delta = F_1F_2$ между фокальными плоскостями объектива и окуляра называется *оптической длиной тубуса* микроскопа.

Обратите внимание, что красная стрелочка на сетчатке направлена вверх. Мозг вторично перевернёт её, и в результате объект при рассмотрении в микроскоп будет казаться *перевёрнутым*. Чтобы этого не происходило, в микроскопе используются промежуточные линзы, дополнительно переворачивающие изображение.

Увеличение микроскопа определяется точно так же, как и для лупы: $\Gamma = l/l_0 = \varphi/\varphi_0$. Здесь, как и выше, l и φ — размер изображения на сетчатке и угол зрения при рассматривании объекта в микроскоп, l_0 и φ_0 — те же величины при рассматривании объекта невооружённым глазом.

Имеем по-прежнему $\varphi_0 = h/d_0$, а угол φ, как видно из рис. 4.64, равен:

$$\varphi = \frac{H}{f_2}\,.$$

Деля φ на φ_0, получим для увеличения микроскопа:

$$\Gamma = \frac{H d_0}{h f_2}\,. \tag{4.28}$$

Это, разумеется, не окончательная формула: в ней присутствуют h и H (величины, относящиеся к объекту), а хотелось бы видеть характеристики микроскопа. Ненужное нам отношение H/h мы устраним с помощью формулы линзы.

Для начала ещё раз посмотрим на рис. 4.64 и используем подобие прямоугольных треугольников с красными катетами H и h:

$$\frac{H}{h} = \frac{b}{a}\,.$$

Здесь $b = f_1 + \delta$ — расстояние от изображения H до объектива, a — расстояние от объекта h до объектива. Теперь привлекаем формулу линзы для объектива:

$$\frac{1}{a} + \frac{1}{b} = \frac{1}{f_1}\,,$$

из которой получаем:

$$\frac{b}{a} = \frac{b - f_1}{f_1}\,.$$

Итак,

$$\frac{H}{h} = \frac{b - f_1}{f_1} = \frac{\delta}{f_1}\,,$$

и это выражение мы подставляем в (4.28):

$$\Gamma = \frac{\delta d_0}{f_1 f_2}\,. \tag{4.29}$$

Вот это и есть окончательное выражение для увеличения, даваемого микроскопом. Например, если фокусное расстояние объектива равно $f_1 = 1$ см, фокусное расстояние окуляра $f_2 = 2$ см, а оптическая длина тубуса $\delta = 20$ см, то согласно формуле (4.29) увеличение микроскопа равно:

$$\Gamma = \frac{20\,\text{см} \cdot 25\,\text{см}}{1\,\text{см} \cdot 2\,\text{см}} = 250.$$

Сравните это с увеличением одного только объектива, которое вычисляется по формуле (4.27):

$$\Gamma_1 = \frac{d_0}{f_1} = \frac{25\,\text{см}}{1\,\text{см}} = 25.$$

Увеличение микроскопа в 10 раз больше!

Теперь мы переходим к объектам, которые достаточно крупны, но находятся слишком далеко от нас. Чтобы рассматривать их получше, применяются *зрительные трубы* — подзорные трубы, бинокли, телескопы и т. д.

Объективом зрительной трубы служит собирающая линза (или система линз) с достаточно большим фокусным расстоянием. А вот окуляром может быть как собирающая, так и рассеивающая линза. Соответственно имеются два вида зрительных труб:

- *труба Кеплера* — если окуляр является собирающей линзой;
- *труба Галилея* — если окуляр является рассеивающей линзой.

Рассмотрим подробнее, как работают эти зрительные трубы.

4.8.4 Труба Кеплера

Принцип действия трубы Кеплера очень прост: объектив даёт изображение удалённого объекта в своей фокальной плоскости, а затем это изображение рассматривается в окуляр как в лупу. Таким образом, задняя фокальная плоскость объектива совпадает с передней фокальной плоскостью окуляра.

Ход лучей в трубе Кеплера изображён на рис. 4.65.

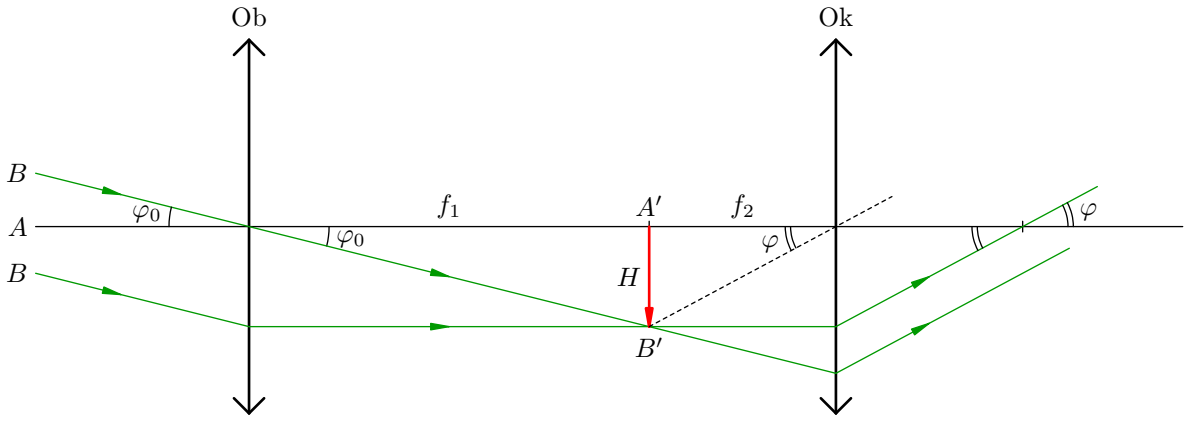


Рис. 4.65.

Объектом служит далеко расположенная стрелка AB, направленная вертикально вверх; она не показана на рисунке. Луч из точки A идёт вдоль главной оптической оси объектива и окуляра. Из точки B идут два луча, которые ввиду удалённости объекта можно считать параллельными.

В результате изображение $A'B'$ нашего объекта, даваемое объективом, расположено в фокальной плоскости объектива и является действительным, перевёрнутым и уменьшенным. Размер изображения обозначим H.

Невооружённым глазом объект виден под углом φ_0. Согласно рис. 4.65:

$$\varphi_0 = \frac{H}{f_1}\,, \tag{4.30}$$

где f_1 — фокусное расстояние объектива.

Изображение объекта мы видим в окуляр под углом φ, который равен:

$$\varphi = \frac{H}{f_2}\,, \tag{4.31}$$

где f_2 — фокусное расстояние окуляра.

Увеличение зрительной трубы — это отношение угла зрения при наблюдении в трубу к углу зрения при наблюдении невооружённым глазом:

$$\Gamma = \frac{\varphi}{\varphi_0}\,.$$

Согласно формулам (4.31) и (4.30) получаем:

$$\Gamma = \frac{f_1}{f_2}\,. \tag{4.32}$$

Например, если фокусное расстояние объектива равно 1 м, а фокусное расстояние окуляра равно 2 см, то увеличение зрительной трубы окажется равным: $\Gamma = 100/2 = 50$.

Ход лучей в трубе Кеплера принципиально тот же, что и в микроскопе. Изображением объекта на сетчатке также будет стрелочка, направленная вверх, и поэтому в трубе Кеплера мы увидим объект перевёрнутым. Во избежании этого в пространстве между объективом и окуляром ставят специальные оборачивающие системы линз или призм, которые ещё раз переворачивают изображение.

4.8.5 Труба Галилея

Галилей изобрёл свой телескоп в 1609 году, и его астрономические открытия потрясли современников. Он обнаружил спутники Юпитера и фазы Венеры, разглядел лунный рельеф (горы, впадины, долины) и пятна на Солнце, а сплошной с виду Млечный Путь оказался скоплением звёзд.

Окуляром трубы Галилея служит рассеивающая линза; задняя фокальная плоскость объектива совпадает с задней фокальной плоскостью окуляра (рис. 4.66).

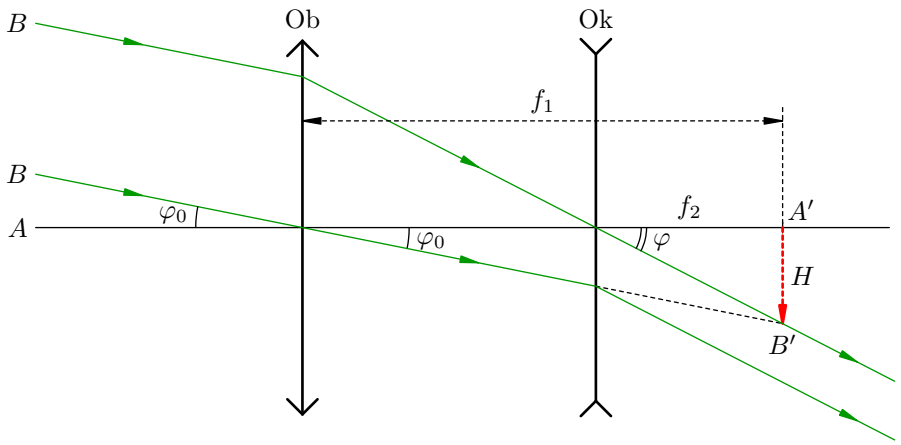


Рис. 4.66.

Если бы окуляра не было, то изображение $A'B'$ удалённой стрелки AB находилось бы в фокальной плоскости объектива. На рисунке это изображение показано пунктиром — ведь в реальности его там нет!

А нет его там потому, что лучи от точки B, которые после прохождения объектива стали сходящимися к точке B', не доходят до B' и попадают на окуляр. После окуляра они вновь становятся параллельными и поэтому воспринимаются глазом без напряжения. Но теперь мы видим изображение объекта под углом φ, который больше угла зрения φ_0 при рассматривании объекта невооружённым глазом.

Из рис. 4.66 имеем:

$$\varphi = \frac{H}{f_2}, \quad \varphi_0 = \frac{H}{f_1},$$

и для увеличения трубы Галилея мы получаем ту же формулу (4.32), что и для трубы Кеплера:

$$\Gamma = \frac{\varphi}{\varphi_0} = \frac{f_1}{f_2}.$$

Заметьте, что при том же увеличении труба Галилея меньше размером, чем труба Кеплера. Поэтому одно из основных применений трубы Галилея — театральные бинокли.

В отличие от микроскопа и трубы Кеплера, в трубе Галилея мы видим объекты неперевёрнутыми. Почему?

4.9 Принцип Гюйгенса

Базовым утверждением волновой оптики является *принцип Гюйгенса*. Законы отражения и преломления света получаются из него в качестве следствий.

Принцип Гюйгенса в его исходной формулировке не смог объяснить дифракцию и, в частности, закон прямолинейного распространения света. Впоследствии он был дополнен важной идеей Френеля об интерференции вторичных волн. Принцип Гюйгенса–Френеля — это мощный инструмент волновой оптики; мы рассмотрим его в разделе «Дифракция света».

Данный раздел посвящён «чистому» принципу Гюйгенса (т. е. без дополнения Френеля).

4.9.1 Волновые поверхности и лучи

Представьте себе маленькую лампочку, которая даёт частые периодические вспышки. Каждая вспышка порождает расходящуюся световую волну в виде расширяющейся сферы (с центром в лампочке). Остановим время — и увидим остановившиеся в пространстве световые сферы, образованные вспышками в различные предшествующие моменты времени.

Эти сферы — так называемые волновые поверхности. Заметьте, что лучи, идущие от лампочки, перпендикулярны волновым поверхностям.

Чтобы дать строгое определение волновой поверхности, давайте вспомним сначала, что такое фаза колебаний. Пусть величина x совершает гармонические колебания по закону:

$$x = A\cos(\omega t + \varphi_0).$$

Так вот, *фаза* — это величина $\varphi = \omega t + \varphi_0$, которая является аргументом косинуса. Фаза, как видим, линейно возрастает со временем. Значение фазы при $t = 0$ равно φ_0 и называется *начальной фазой*.

Вспомним также, что волна представляет собой распространение колебаний в пространстве. В случае механических волн это будут колебания частиц упругой среды, в случае электромагнитных волн — колебания векторов напряжённости электрического поля и индукции магнитного поля.

Вне зависимости от того, какие волны рассматриваются, мы можем сказать, что в каждой точке пространства, захваченной волновым процессом, происходят колебания некоторой величины; такой величиной является набор координат колеблющейся частицы в случае механической волны или набор координат векторов, описывающих электрическое и магнитное поля в электромагнитной волне.

Фазы колебаний в двух различных точках пространства, вообще говоря, имеют разное значение. Интерес представляют множества точек, в которых фаза одна и та же. Оказывается, совокупность точек, в которых фаза колебаний в данный момент времени имеет фиксированное значение, образует двумерную поверхность в пространстве.

Определение. Волновая поверхность — это множество всех точек пространства, в которых фаза колебаний в данный момент времени имеет одно и то же значение.

Коротко говоря, волновая поверхность есть поверхность постоянной фазы. Каждому значению фазы отвечает своя волновая поверхность. Набору различных значений фазы соответствует семейство волновых поверхностей.

С течением времени фаза в каждой точке меняется, и волновая поверхность, отвечающая фиксированному значению фазы, перемещается в пространстве. Следовательно, распространение волн можно рассматривать как движение волновых поверхностей! Тем самым в нашем распоряжении оказываются удобные геометрические образы для описания физических волновых процессов.

Например, если точечный источник света находится в прозрачной однородной среде, то волновые поверхности являются концентрическими сферами с общим центром в источнике. Распространение света выглядит как расширение этих сфер. Мы это уже видели выше в ситуации с лампочкой.

Через каждую точку пространства в данный момент времени может проходить только одна волновая поверхность. В самом деле, если предположить, что через точку A проходят две волновых поверхности, отвечающие различным значениям фазы φ_1 и φ_2, то немедленно получим противоречие: фаза колебаний в точке A окажется одновременно равна этим двум различным числам.

Коль скоро через точку A проходит единственная волновая поверхность, то однозначно определено и направление перпендикуляра к волновой поверхности в данной точке.

Определение. Луч — это линия в пространстве, которая в каждой своей точке перпендикулярна волновой поверхности, проходящей через эту точку.

Иными словами, луч есть общий перпендикуляр к семейству волновых поверхностей. Направление луча — это направление распространения волны. Вдоль лучей осуществляется перенос энергии волны от одних точек пространства к другим.

По мере распространения волны происходит перемещение границы, которая разделяет область пространства, захваченную волновым процессом, и невозмущённую пока ещё область. Эта граница называется *волновым фронтом*. Таким образом, волновой фронт — это множество всех точек пространства, которых достиг колебательный процесс в данный момент времени. Волновой фронт есть частный случай волновой поверхности; это, если можно так выразиться, «самая первая» волновая поверхность.

К наиболее простым видам геометрических поверхностей относятся сфера и плоскость. Соответственно, имеем два важных случая волновых процессов с волновыми поверхностями такой формы — это сферические и плоские волны.

4.9.2 Сферическая волна

Волна называется *сферической*, если её волновые поверхности — сферы (рис. 4.67).

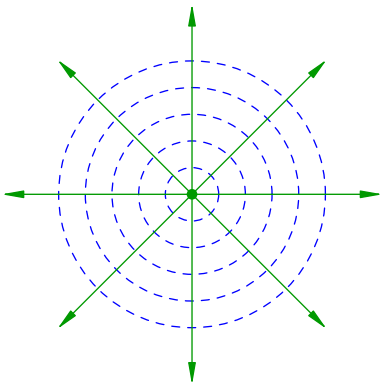


Рис. 4.67. Сферическая волна

Волновые поверхности показаны синим пунктиром, а зелёные радиальные стрелки — это лучи, перпендикулярные волновым поверхностям.

Понятие сферической волны оказывается чрезвычайно полезным. В самом деле, возьмём прозрачную однородную среду, физические свойства которой одинаковы вдоль всех направлений. Точечный источник света, помещённый в такую среду, излучает сферические волны (это понятно — ведь свет пойдёт в каждом направлении с одинаковой скоростью, так что любая волновая поверхность будет сферой). Ну а протяжённый источник света можно рассматривать

как совокупность точечных источников, и наложение сферических волн этих источников даст общую световую волну, идущую от протяжённого источника.

Кроме того, обсуждаемые ниже вторичные волны (центральное понятие принципа Гюйгенса) являются именно сферическими.

4.9.3 Плоская волна

Волна называется *плоской*, если её волновые поверхности — плоскости (рис. 4.68).

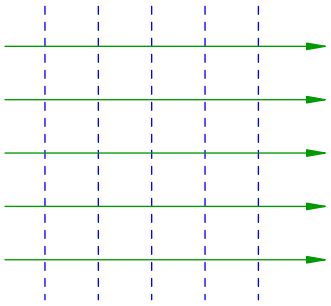


Рис. 4.68. Плоская волна

Синим пунктиром показаны параллельные плоскости, являющиеся волновыми поверхностями. Лучи — зелёные стрелки — снова оказываются прямыми линиями.

Плоская волна — одна из важнейших идеализаций волновой теории; математически она описывается наиболее просто. Этой идеализацией можно пользоваться, например, когда мы находимся на достаточно большом расстоянии от источника. Тогда в окрестности точки наблюдения можно пренебречь искривлением сферической волновой поверхности и считать волну приблизительно плоской.

В дальнейшем, выводя законы отражения и преломления из принципа Гюйгенса, мы будем использовать именно плоские волны. Но сначала разберёмся с самим принципом Гюйгенса.

4.9.4 Вторичные волны

Мы говорили выше, что распространение волн удобно представлять себе как движение волновых поверхностей. Но согласно каким правилам перемещаются волновые поверхности? Иными словами — как, зная положение волновой поверхности в данный момент времени, определить её положение в следующий момент?

Ответ на этот вопрос даёт принцип Гюйгенса — ключевое утверждение волновой теории света. Принцип Гюйгенса имеет весьма общую формулировку и равным образом справедлив как для механических, так и для электромагнитных волн.

Вначале, чтобы лучше понять идею Гюйгенса, давайте рассмотрим такой пример. Бросим в воду горсть камней. От каждого камня пойдёт круговая волна с центром в точке падения камня. Эти круговые волны, накладываясь друг на друга, создадут общую волновую картину на поверхности воды. Важно то, что все круговые волны и порождённая ими волновая картина будут существовать и после того, как камни опустятся на дно. Стало быть, непосредственной причиной исходных круговых волн служат не сами камни, а *локальные возмущения* поверхности воды в тех местах, куда камни упали. Именно локальные возмущения сами по себе являются источниками расходящихся круговых волн и формирующейся волновой картины, и уже не столь важно, что конкретно послужило причиной каждого из этих возмущений — камень ли, поплавок или какой-то иной объект. Для описания последующего волнового процесса существенно только то, что в начальный момент времени в определённых точках поверхности воды возникли круговые волны.

Так вот, основная идея Гюйгенса состояла в том, что локальные возмущения могут порождаться не только посторонними объектами типа камня или поплавка, но также и распространяющейся в пространстве волной!

Принцип Гюйгенса. Каждая точка пространства, вовлечённая в волновой процесс, сама становится источником сферических волн.

Эти сферические волны, распространяющиеся во все стороны от каждой точки волнового возмущения, называются *вторичными волнами*. Последующая эволюция волнового процесса состоит в наложении вторичных волн, испущенных всеми точками, до которых волновой процесс уже успел добраться.

И вот тут возникает самый главный вопрос: а что такое «наложение вторичных волн»? Что представляет собой с физической точки зрения этот процесс и как он описывается математически?

Чёткий ответ был дан Френелем в 1815 году: вторичные волны интерферируют друг с другом, и наблюдаемый волновой процесс есть результат интерференции вторичных волн. Френель разработал математический способ нахождения суммарного волнового поля (метод зон Френеля), а модифицированный Френелем принцип Гюйгенса с тех пор называется принципом Гюйгенса–Френеля.

Но сам Гюйгенс сформулировал свой принцип в 1678 году, когда об интерференции волн ещё не было ничего известно. Гюйгенс предложил лишь геометрический рецепт построения волновой поверхности в момент времени $t+\Delta t$ по известному её положению в текущий момент времени t (рис. 4.69).

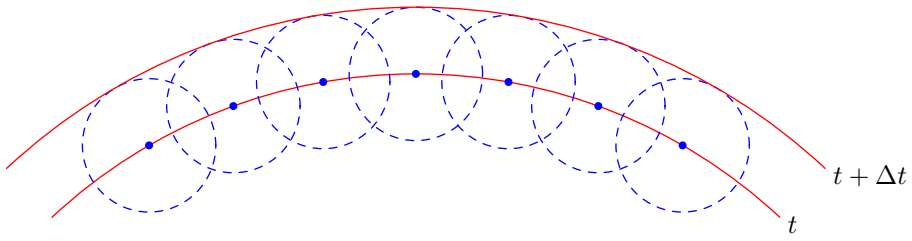


Рис. 4.69. Принцип Гюйгенса: движение волновых поверхностей

Именно, каждую точку исходной волновой поверхности мы рассматриваем как источник вторичных волн. За время Δt вторичные волны пройдут расстояние $c\Delta t$, где c — скорость волны. Из каждой точки старой волновой поверхности строим сферы радиуса $c\Delta t$; новая волновая поверхность будет *касательной* ко всем этим сферам[11].

Этот геометрический рецепт мы и называем «чистым» принципом Гюйгенса.

Конечно, для построения волновой поверхности мы не обязаны брать вторичные волны, испущенные точками, лежащими непременно на одной из предыдущих волновых поверхностей. Искомая волновая поверхность будет огибающей семейства вторичных волн, излучённых точками вообще всякой поверхности, вовлечённой в колебательный процесс. Выбор этой поверхности в каждой конкретной ситуации диктуется соображениями удобства.

С помощью своего принципа Гюйгенс пытался объяснить прямолинейное распространение света, но сделать этого ему не удалось. Как оказалось впоследствии, вопрос о прямолинейном распространении света решается только в рамках теории дифракции, которую «чистый» принцип Гюйгенса также не объясняет. Для истолкования дифракционных явлений нужен более мощный принцип Гюйгенса–Френеля.

Но и «чистый» принцип Гюйгенса весьма силён: с его помощью могут быть получены законы отражения и преломления света.

[11] Говорят ещё, что волновая поверхность в любой момент времени служит *огибающей* семейства вторичных волн.

4.9.5 Вывод закона отражения

Предположим, что на поверхность KL раздела двух сред падает плоская волна (рис. 4.70). Фиксируем две точки A и B этой поверхности.

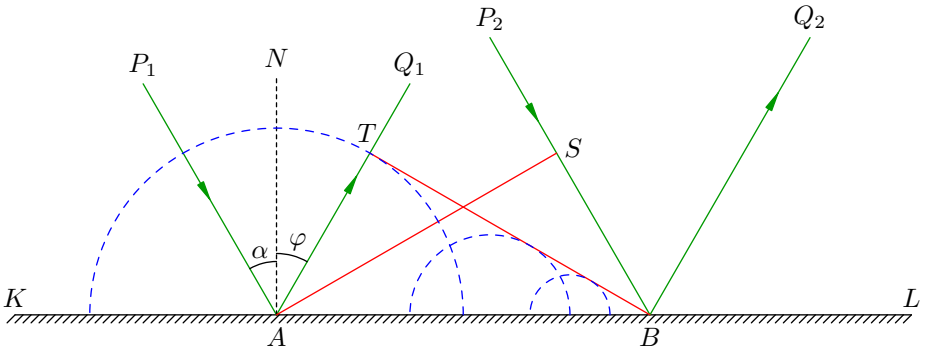


Рис. 4.70. Отражение волны

В эти точки приходят два падающих луча P_1A и P_2B; плоскость AS, перпендикулярная этим лучам, есть волновая поверхность падающей волны.

В точке A проведена нормаль AN к отражающей поверхности. Угол $\alpha = \angle P_1AN$ есть, как вы помните, угол падения.

Из точек A и B выходят отражённые лучи AQ_1 и BQ_2. Перпендикулярная этим лучам плоскость BT есть волновая поверхность отражённой волны. Угол отражения $\angle NAQ_1$ обозначим пока φ; мы хотим доказать, что $\varphi = \alpha$.

Все точки отрезка AB служат источниками вторичных волн. Раньше всего волновая поверхность AS приходит в точку A. Затем, по мере движения падающей волны, в колебательный процесс вовлекаются другие точки данного отрезка, и в самую последнюю очередь — точка B.

Соответственно, раньше всего начинается излучение вторичных волн в точке A; сферическая волна с центром в A имеет на рис. 4.70 наибольший радиус. По мере приближения к точке B радиусы сферических вторичных волн, испущенных промежуточными точками, уменьшаются до нуля — ведь вторичная волна будет излучена тем позже, чем ближе её источник находится к точке B.

Волновая поверхность BT отражённой волны есть плоскость, касательная ко всем этим сферам. На нашем планиметрическом чертеже BT есть отрезок касательной, проведённой из точки B к самой большой окружности с центром в A и радиусом AT.

Теперь заметим, что радиус AT — это расстояние, пройденное вторичной волной с центром в A за то время, пока волновая поверхность AS двигается к точке B. Скажем это чуть по-другому: время движения вторичной волны от точки A до точки T равно времени движения падающей волны от точки S до точки B. Но скорости движения падающей и вторичной волн совпадают — ведь дело происходит в одной и той же среде! Поэтому, раз совпадают скорости и времена, то равны и расстояния: $AT{=}BS$.

Получается, что прямоугольные треугольники ABT и ABS равны по гипотенузе и катету. Стало быть, равны и соответствующие острые углы: $\angle ABT = \angle BAS$. Остаётся заметить, что $\angle ABT = \varphi$ (так как оба они равны $90^\circ - \angle BAT$) и $\angle BAS = \alpha$ (оба они равны $90^\circ - \angle NAS$). Таким образом, $\varphi = \alpha$ — угол отражения равен углу падения, что и требовалось.

Кроме того, из построения на рис. 4.70 нетрудно видеть, что выполнено и второе утверждение закона преломления: падающий луч P_1A, отражённый луч AQ_1 и нормаль AN к отражающей поверхности лежат в одной плоскости.

4.9.6 Вывод закона преломления

Теперь покажем, как из принципа Гюйгенса следует закон преломления. Будем для определённости считать, что плоская электромагнитная волна распространяется в воздухе и падает на границу KL с некоторой прозрачной средой (рис. 4.71). Как обычно, угол падения α есть угол между падающим лучом и нормалью к поверхности, угол преломления β — это угол между преломлённым лучом и нормалью.

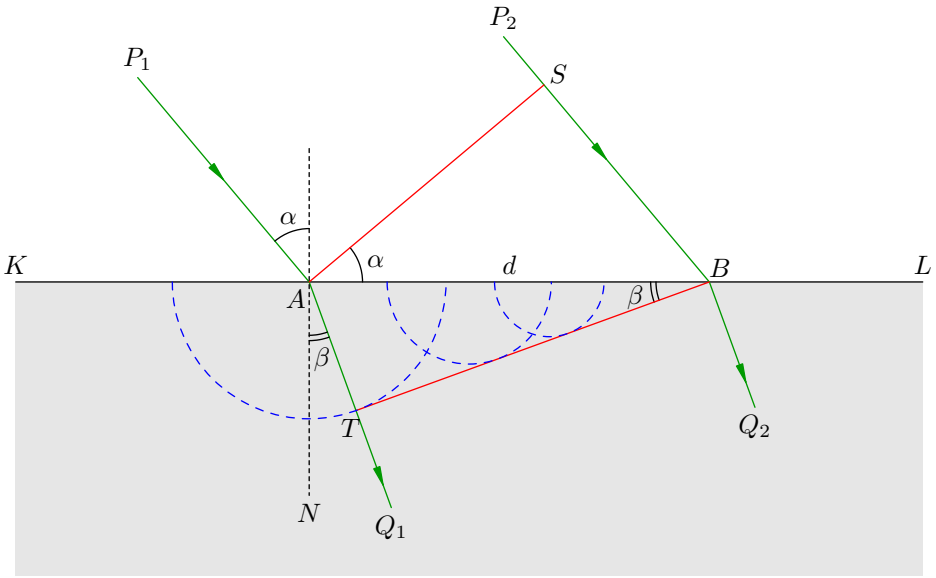


Рис. 4.71. Преломление волны

Точка A является первой точкой отрезка AB, которой достигает волновая поверхность AS падающей волны; в точке A излучение вторичных волн начинается раньше всего. Пусть t — время, которое с этого момента требуется падающей волне, чтобы достичь точки B, то есть пройти отрезок SB.

Скорость света в воздухе обозначим c, скорость света в среде пусть будет v. Пока падающая волна проходит расстояние $SB = ct$ и достигает точки B, вторичная волна из точки A распространится на расстояние $AT = vt$.

Поскольку $v < c$, то $AT < SB$. Вследствие этого волновая поверхность BT *не параллельна* волновой поверхности AS — происходит преломление света! В рамках геометрической оптики не давалось никакого объяснения того, почему вообще наблюдается явление преломления. Причина преломления кроется в волновой природе света и становится понятной с точки зрения принципа Гюйгенса: всё дело в том, что скорость вторичных волн в среде меньше скорости света в воздухе, и это приводит к повороту волновой поверхности BT относительно исходного положения AS.

Из прямоугольных треугольников ABS и ABT легко видеть, что $SB = d\sin\alpha$ и $BT = d\sin\beta$ (для краткости обозначено $d = AB$). Имеем, таким образом:

$$d\sin\alpha = ct, \quad d\sin\beta = vt.$$

Поделив эти уравнения друг на друга, получим:

$$\frac{\sin\alpha}{\sin\beta} = \frac{c}{v}.$$

Отношение синуса угла падения к синусу угла преломления оказалось равно постоянной величине c/v, не зависящей от угла падения. Эта величина называется *показателем преломления*

среды:

$$n = \frac{c}{v}\,.$$

Получился хорошо известный нам закон преломления:

$$\frac{\sin\alpha}{\sin\beta} = n.$$

Обратите внимание: физический смысл показателя преломления (как отношения скоростей света в вакууме и в среде) прояснился опять-таки благодаря принципу Гюйгенса.

Из рис. 4.71 очевидно и второе утверждение закона преломления: падающий луч P_1A, преломлённый луч AQ_1 и нормаль AN к границе раздела лежат в одной плоскости.

4.10 Интерференция волн

Изучая принцип Гюйгенса, мы говорили о том, что общая картина волнового процесса создаётся наложением вторичных волн и сказали пару слов про *интерференцию*. Сейчас мы займёмся этим явлением подробнее. Итак: в чём состоит физический смысл наложения волн? Что вообще происходит, когда в пространстве одновременно распространяются несколько волн?

В данном разделе оптический характер явлений временно отодвигается на второй план. Природа рассматриваемых волновых процессов сейчас значения не имеет — это могут быть механические волны в упругой среде или электромагнитные волны (в частности, свет) в прозрачной среде или в вакууме. Полученные нами результаты будут одинаково справедливы для всех видов волн.

Нас в первую очередь интересует взаимодействие *двух* волн — это самая простая и часто встречающаяся в оптике ситуация[12].

4.10.1 Сложение колебаний

Опыт показывает, что волны складываются друг с другом в следующем смысле.

Принцип суперпозиции. Если две волны накладываются друг на друга в определённой области пространства, то они порождают новый волновой процесс. При этом значение колеблющейся величины в любой точке данной области равно сумме соответствующих колеблющихся величин в каждой из волн по отдельности.

Например, при наложении двух механических волн перемещение частицы упругой среды равно сумме перемещений, создаваемых в отдельности каждой волной. При наложении двух электромагнитных волн напряжённость электрического поля в данной точке равна сумме напряжённостей в каждой волне (и то же самое для индукции магнитного поля).

Разумеется, принцип суперпозиции справедлив не только для двух, но и вообще для любого количества накладывающихся волн. Результирующее колебание в данной точке всегда равно сумме колебаний, создаваемых каждой волной по отдельности.

Мы ограничимся рассмотрением наложения двух волн *одинаковой амплитуды и частоты*. Этот случай наиболее часто встречается в физике и, в частности, в оптике.

Оказывается, на амплитуду результирующего колебания сильно влияет *разность фаз* складывающихся колебаний. В зависимости от разности фаз в данной точке пространства две волны могут как усиливать друг друга, так и полностью гасить!

Предположим, например, что в некоторой точке фазы колебаний в накладывающихся волнах совпадают (рис. 4.72).

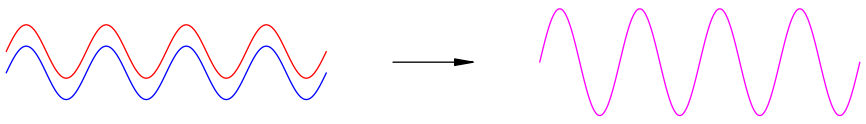


Рис. 4.72. Волны в фазе: усиление колебаний

Мы видим, что максимумы красной волны приходятся в точности на максимумы синей волны, минимумы красной волны — на минимумы синей (левая часть рис. 4.72). Красная и синяя волны *складываются в фазе* и усиливают друг друга, порождая колебания удвоенной амплитуды (волна цвета маджента[13] справа на рис. 4.72).

[12] Сложение нескольких волн — так называемая многолучевая интерференция — у нас возникнет лишь при рассмотрении дифракционной решётки, и то в своём простейшем виде.

[13] *Маджента* — цвет, получающийся в результате смешения красного и синего цветов. К рассматриваемой теме данная ремарка отношения не имеет — это просто для расширения кругозора ;-)

Теперь сдвинем синюю синусоиду относительно красной на половину длины волны. Тогда максимумы синей волны будут совпадать с минимумами красной и наоборот — минимумы синей волны совпадут с максимумами красной (рис. 4.73, слева).

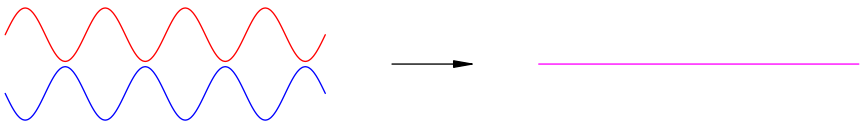


Рис. 4.73. Волны в противофазе: гашение колебаний

Колебания, создаваемые этими волнами, будут происходить, как говорят, *в противофазе* — разность фаз колебаний станет равна π. Результирующее колебание окажется равным нулю, т. е. красная и синяя волны попросту уничтожат друг друга (рис. 4.73, справа).

4.10.2 Интенсивность волны

Энергетической характеристикой волны является её *интенсивность* (называемая также плотностью потока излучения):

$$I = \frac{W}{St}\,. \tag{4.33}$$

Здесь W — энергия, которая переносится за время t через поверхность площади S, расположенную перпендикулярно лучам. Таким образом, интенсивность I — это энергия, переносимая волной вдоль лучей через единицу площади в единицу времени.

Пусть w — объёмная плотность энергии волны, то есть отношение переносимой волной энергии W к объёму области, через которую эта энергия переносится:

$$w = \frac{W}{V}\,.$$

Оказывается, объёмная плотность энергии пропорциональна квадрату амплитуды A колебаний в волне: $w \sim A^2$. Подробно мы вдаваться в этот вопрос не будем, но полезной ассоциацией тут служит известная вам формула $w = \varepsilon_0 \varepsilon E^2/2$ для объёмной плотности энергии электрического поля.

Далее, имеется простое соотношение между интенсивностью и объёмной плотностью энергии:

$$I = wc,$$

где c — скорость волны (эта формула вам встречалась, когда вы проходили электромагнитные волны).

Отсюда следует, что *интенсивность пропорциональна квадрату амплитуды колебаний*: $I \sim A^2$. Можно сказать и по-другому: амплитуда колебаний пропорциональна квадратному корню из интенсивности: $A \sim \sqrt{I}$.

В сферической волне энергия равномерно распределяется по поверхности сферы, радиус r которой увеличивается в процессе распространения волны. Площадь поверхности сферы есть $S = 4\pi r^2$, поэтому из формулы (4.33) для интенсивности сферической волны получим:

$$I = \frac{W}{4\pi r^2 t}\,.$$

Как видим, *интенсивность сферической волны обратно пропорциональна квадрату расстояния до источника.*

Поскольку амплитуда колебаний A, как мы выяснили, пропорциональна корню из интенсивности, мы приходим к следующему выводу: *амплитуда колебаний в сферической волне обратно пропорциональна расстоянию до источника.*

4.10.3 Когерентные источники

Пусть имеются два точечных источника, создающие волны в окружающем пространстве. Мы полагаем, что эти источники согласованы друг с другом в следующем смысле.

Когерентность. Два источника называются когерентными, если они имеют одинаковую частоту и постоянную, не зависящую от времени разность фаз. Волны, возбуждаемые такими источниками, также называются когерентными.

Итак, рассматриваем два когерентных источника S_1 и S_2. Для простоты считаем, что *источники излучают волны одинаковой амплитуды, а разность фаз между источниками равна нулю*. Словом, эти источники являются «точными копиями» друг друга.

В оптике такие идентичные источники могут быть получены разными способами. Например, оба источника S_1 и S_2 служат изображениями одного и того же источника S в какой-либо оптической системе. Именно так работает интерференционная схема с зеркалами Френеля, рассматриваемая в следующем разделе.

Наложение волн, излучённых источниками S_1 и S_2, наблюдается в некоторой точке P. Вообще говоря, амплитуды этих волн в точке P не будут равны друг другу — ведь, как мы только что выяснили, амплитуда сферической волны обратно пропорциональна расстоянию до источника, и при разных расстояниях S_1P и S_2P амплитуды пришедших волн окажутся различными. Но во многих случаях точка P расположена достаточно далеко от источников — *на расстоянии гораздо большем, чем расстояние между самими источниками*. В такой ситуации различие в расстояниях S_1P и S_2P не приводит к существенному отличию в амплитудах приходящих волн. Следовательно, мы можем считать, что амплитуды волн в точке P также совпадают.

4.10.4 Условие максимума и минимума

Тем не менее, величина $\delta = |S_1P - S_2P|$, называемая *разностью хода*, имеет важнейшее значение. От неё самым решительным образом зависит то, какой результат сложения приходящих волн мы увидим в точке P.

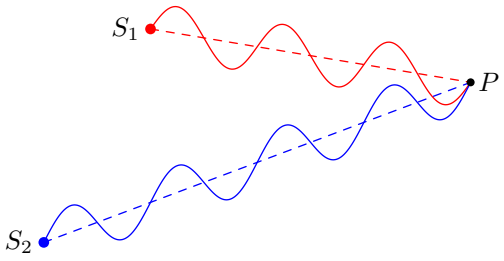


Рис. 4.74. Усиление колебаний в точке P

В ситуации на рис. 4.74 разность хода равна длине волны ($\delta = \lambda$). Действительно, на отрезке S_1P укладываются три полных волны, а на отрезке S_2P — четыре (это, конечно, лишь иллюстрация; в оптике, например, длины таких отрезков могут составлять тысячи и миллионы длин волн).

Легко видеть, что волны в точке P складываются в фазе — наблюдается, как говорят, *интерференционный максимум*. Если красная и синяя волны имеют амплитуду A_1, то результирующая волна в точке P будет иметь удвоенную амплитуду $A = 2A_1$. Интенсивность результирующей волны, пропорциональная квадрату амплитуды, будет соответственно в четыре раза превышать интенсивность волн-слагаемых: $I = 4I_1$.

Ясно, что аналогичная ситуация возникнет при разности хода, равной не только длине волны, но и любому целому числу длин волн.

Условие максимума. При наложении когерентных волн колебания в данной точке будут иметь максимальную амплитуду, если разность хода равна целому числу длин волн:

$$\delta = n\lambda \quad (n = 0, 1, 2, \ldots). \tag{4.34}$$

Теперь посмотрим на рис. 4.75. На отрезке S_1P укладываются две с половиной волны, а на отрезке S_2P — три волны. Разность хода составляет половину длины волны ($\delta = \lambda/2$).

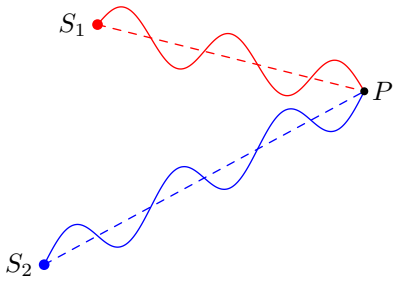


Рис. 4.75. Гашение колебаний в точке P

Теперь нетрудно видеть, что волны в точке P складываются в противофазе и гасят друг друга — наблюдается *интерференционный минимум*. Амплитуда и интенсивность результирующей волны в точке минимума равны нулю.

То же самое будет, если разность хода окажется равна половине длины волны плюс любое целое число длин волн.

Условие минимума. Когерентные волны, складываясь, гасят друг друга, если разность хода равна полуцелому числу длин волн:

$$\delta = n\lambda + \frac{\lambda}{2} \quad (n = 0, 1, 2, \ldots). \tag{4.35}$$

Равенство (4.35) можно переписать следующим образом:

$$\delta = (2n + 1)\frac{\lambda}{2}\,.$$

Поэтому условие минимума формулируют ещё так: *разность хода должна быть равна нечётному числу длин полуволн.*

4.10.5 Интерференционная картина

А что, если разность хода принимает какое-то иное значение, не равное целому или полуцелому числу длин волн?

Тогда волны, приходящие в данную точку, создают в ней колебания с некоторой промежуточной амплитудой, расположенной между нулём и удвоенным значением $2A_1$ амплитуды одной волны. Эта промежуточная амплитуда может принимать все значения от 0 до $2A_1$ по мере того, как разность хода меняется от полуцелого до целого числа длин волн. Соответственно, интенсивность результирующей волны принимает значения от 0 до $4I_1$.

Таким образом, в той области пространства, где происходит наложение волн когерентных источников S_1 и S_2, наблюдается устойчивая *интерференционная картина* — фиксированное, не зависящее от времени распределение амплитуд колебаний. А именно, в каждой точке P данной области амплитуда колебаний принимает своё значение, определяемое разностью хода $\delta = |S_1P - S_2P|$ приходящих сюда волн, и это значение амплитуды не меняется со временем.

Такая стационарность интерференционной картины обеспечивается когерентностью источников. Если, например, разность фаз источников будет постоянно меняться, то никакой устойчивой интерференционной картины уже не возникнет.

Теперь, наконец, мы можем сказать, что такое интерференция.

Интерференция — это взаимодействие волн, в результате которого возникает устойчивая интерференционная картина, то есть не зависящее от времени распределение амплитуд результирующих колебаний в точках области, где волны накладываются друг на друга.

Если волны, перекрываясь, образуют устойчивую интерференционную картину, то говорят попросту, что волны *интерферируют*. Как мы выяснили выше, интерферировать могут только когерентные волны. Когда, например, разговаривают два человека, то мы не замечаем вокруг них чередований максимумов и минимумов громкости; интерференции нет, поскольку в данном случае источники некогерентны.

На первый взгляд может показаться, явление интерференции противоречит закону сохранения энергии — например, куда девается энергия, когда волны полностью гасят друг друга? Но никакого нарушения закона сохранения энергии, конечно же, нет: энергия просто *перераспределяется* между различными участками интерференционной картины. Наибольшее количество энергии концентрируется в интерференционных максимумах, а в точки интерференционных минимумов энергия не поступает совсем.

На рис. 4.76 показана интерференционная картина, созданная наложением волн двух точечных источников S_1 и S_2. Картина построена в предположении, что область наблюдения интерференции находится достаточно далеко от источников. Пунктиром отмечена ось симметрии интерференционной картины.

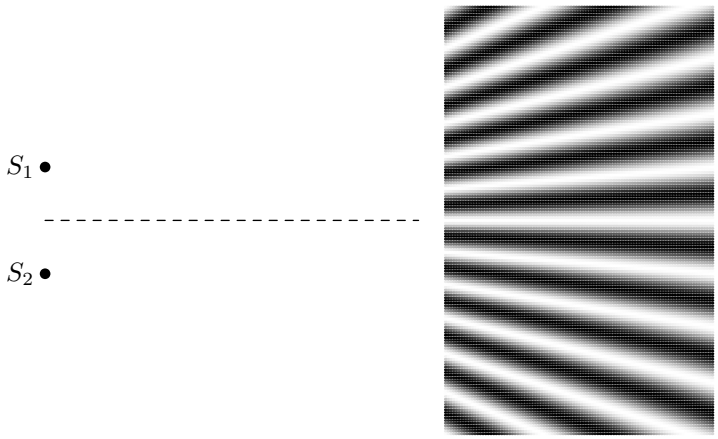


Рис. 4.76. Интерференция волн двух точечных источников

Цвета точек интерференционной картины на этом рисунке меняются от чёрного до белого через промежуточные оттенки серого. Чёрный цвет — интерференционные минимумы, белый цвет — интерференционные максимумы; серый цвет — промежуточное значение амплитуды, и чем больше амплитуда в данной точке, тем светлее сама точка.

Обратите внимание на прямую белую полосу, которая идёт вдоль оси симметрии картины. Здесь расположены так называемые *центральные максимумы*. Действительно, любая точка данной оси равноудалена от источников (разность хода равна нулю), так что в этой точке будет наблюдаться является интерференционный максимум.

Остальные белые полосы и все чёрные полосы слегка искривлены; можно показать, что они являются ветвями гипербол. Однако в области, расположенной на большом расстоянии от источников, кривизна белых и чёрных полос мало заметна, и выглядят эти полосы почти прямыми.

4.10.6 Схема Юнга

Интерференционный опыт, изображённый на рис. 4.76, вместе с соответствующим методом расчёта интерференционной картины называется *схемой Юнга*. Эта схема лежит в основе знаменитого опыта Юнга (речь о котором пойдёт в разделе «Дифракция света»). Многие эксперименты по интерференции света так или иначе сводятся к схеме Юнга.

В оптике интерференционную картину обычно наблюдают на экране. Давайте ещё раз посмотрим на рис. 4.76 и представим себе экран, поставленный перпендикулярно пунктирной оси. На этом экране мы увидим чередование светлых и тёмных *интерференционных полос*.

На рис. 4.77 синусоида показывает распределение освещённости вдоль экрана. В точке O, расположенной на оси симметрии, находится центральный максимум. Первый максимум в верхней части экрана, соседний с центральным, находится в точке A. Выше идут второй, третий (и так далее) максимумы.

Рис. 4.77. Интерференционная картина на экране

Расстояние $\Delta x = OA$, равное расстоянию между любыми двумя соседними максимумами или минимумами, называется *шириной интерференционной полосы*. Сейчас мы займёмся нахождением этой величины.

Пусть источники находятся на расстоянии a друг от друга, а экран расположен на расстоянии L от источников (рис. 4.78). Экран заменён осью OX; начало отсчёта O, как и выше, отвечает центральному максимуму.

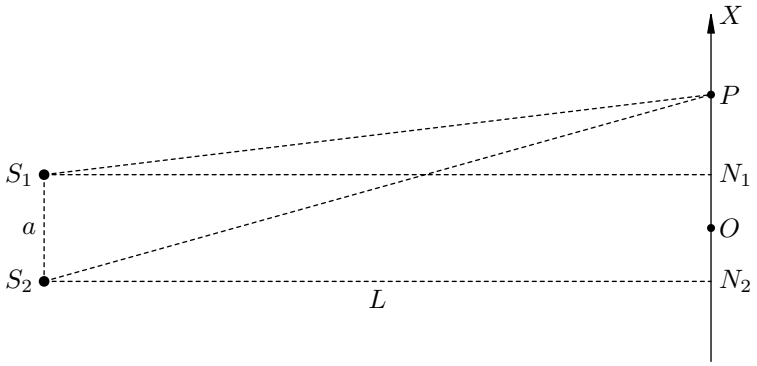


Рис. 4.78. Вычисление координат максимумов

Точки N_1 и N_2 служат проекциями точек S_1 и S_2 на ось OX и расположены симметрично относительно точки O. Имеем: $ON_1 = ON_2 = a/2$.

Точка наблюдения P может находиться на оси OX (на экране) где угодно. Координату точки P мы обозначим x. Нас интересует, при каких значениях x в точке P будет наблюдаться интерференционный максимум.

Волна, излучённая источником S_1, проходит расстояние:

$$S_1P = \sqrt{S_1N_1^2 + PN_1^2} = \sqrt{L^2 + \left(x - \frac{a}{2}\right)^2} = L\sqrt{1 + \left(\frac{2x-a}{2L}\right)^2}. \tag{4.36}$$

Теперь вспомним, что расстояние между источниками много меньше расстояния от источников до экрана: $a \ll L$. Кроме того, в подобных интерференционных опытах координата x точки наблюдения также гораздо меньше L. Это означает, что второе слагаемое под корнем в выражении (4.36) много меньше единицы:

$$\left(\frac{2x-a}{2L}\right)^2 \ll 1.$$

Раз так, можно использовать приближённую формулу:

$$\sqrt{1+\alpha} = 1 + \frac{\alpha}{2}, \quad \text{если } \alpha \ll 1. \tag{4.37}$$

Применяя её к выражению (4.36), получим:

$$S_1P = L\left(1 + \frac{1}{2}\left(\frac{2x-a}{2L}\right)^2\right) = L + \frac{(2x-a)^2}{8L}. \tag{4.38}$$

Точно так же вычисляем расстояние, которое проходит волна от источника S_2 до точки наблюдения:

$$S_2P = \sqrt{S_2N_2^2 + PN_2^2} = \sqrt{L^2 + \left(x + \frac{a}{2}\right)^2} = L\sqrt{1 + \left(\frac{2x+a}{2L}\right)^2}. \tag{4.39}$$

Применяя к выражению (4.39) приближённую формулу (4.37), получаем:

$$S_2P = L\left(1 + \frac{1}{2}\left(\frac{2x+a}{2L}\right)^2\right) = L + \frac{(2x+a)^2}{8L}. \tag{4.40}$$

Вычитая выражения (4.40) и (4.38), находим разность хода:

$$\delta = S_2P - S_1P = \frac{(2x+a)^2 - (2x-a)^2}{8L} = \frac{ax}{L}. \tag{4.41}$$

Пусть λ — длина волны, излучаемой источниками. Согласно условию (4.34), в точке P будет наблюдаться интерференционный максимум, если разность хода равна целому числу длин волн:

$$\delta = \frac{ax}{L} = n\lambda \quad (n = 0, 1, 2, \ldots).$$

Отсюда получаем координаты максимумов в верхней части экрана (в нижней части максимумы идут симметрично):

$$x_n = \frac{n\lambda L}{a} \quad (n = 0, 1, 2, \ldots).$$

При $n = 0$ получаем, разумеется, $x_0 = 0$ (центральный максимум). Первый максимум рядом с центральным соответствует значению $n = 1$ и имеет координату $x_1 = \lambda L/a$. Такой же будет и искомая ширина интерференционной полосы в схеме Юнга:

$$\Delta x = x_{n+1} - x_n = \frac{(n+1)\lambda L}{a} - \frac{n\lambda L}{a} = \frac{\lambda L}{a}. \tag{4.42}$$

4.11 Интерференция света

Всё, о чём мы говорили в предыдущем разделе, справедливо для интерференции любых видов волн — в частности, световых. Но имеются две существенные особенности, отличающие интерференцию света от интерференции, скажем, механических волн.

1. Период колебаний электромагнитного поля в световой волне является столь малым, что наблюдать и измерять мы можем лишь усреднённое значение интенсивности света.

2. Два независимых источника света всегда будут некогерентными, даже если они излучают свет одинаковой длины волны[14].

Любой эксперимент, связанный с интерференцией света, обязан учитывать эти особенности. Поэтому вначале мы изучим их более подробно, и лишь затем перейдём к рассмотрению интерференционных явлений в оптике.

4.11.1 Усреднение интенсивности

Что мы видим, воспринимая свет? Эксперименты показывают, что рецепторы человеческого глаза регистрируют не напряжённость электрического поля E световой волны, а *интенсивность* света I, которая пропорциональна квадрату напряжённости: $I \sim E^2$.

Кроме того, нашему глазу присуща некоторая *инерционность*. Именно, если что-либо колеблется или мелькает чаще десяти раз в секунду (т. е. $\nu > 10\,\text{Гц}$), то глаз не успевает следить за этими изменениями и они сливаются в один сплошной сигнал. Например, плавность движения изображений на телеэкране обеспечивается частотой 24 кадра в секунду (и выше).

Электромагнитные волны в диапазоне видимого света обладают частотами около $10^{15}\,\text{Гц}$. Разумеется, глаз не в состоянии отслеживать изменения интенсивности, происходящие с такой гигантской частотой. Мы воспринимаем *среднюю* интенсивность света, получаемую усреднением огромного количества мгновенных значений интенсивности за промежуток времени порядка $0{,}1\,\text{с}$ (равный времени разрешения нашего глаза).

4.11.2 Некогерентность независимых источников

Давайте включим электрическую лампочку, а затем — такую же вторую рядом с ней. Как вы хорошо знаете из житейского опыта, освещённость окружающего пространства равномерно возрастёт, и никакой интерференционной картины вокруг лампочек не возникнет. Почему же?

Здесь мы сталкиваемся с весьма общим для оптики фактом: два независимых источника света (даже с одинаковой длиной волны) оказываются некогерентными. И причина этого заключена во внутреннем устройстве световых волн.

Откуда вообще берётся свет? Видимый свет излучается *атомами* различных тел. Механизм излучения света относится квантовой физике, но для понимания оптических интерференционных явлений знать хотя бы в общих чертах, как атомы излучают свет, надо обязательно. Поэтому обсудим вкратце этот вопрос.

Обычным состоянием атома, в котором он может пребывать неограниченно долго, является *основное*, или *невозбуждённое* состояние. Когда атом находится в основном состоянии, электроны, окружающие ядро атома, максимально заполняют ближайшие к ядру орбиты. Потенциальная энергия взаимодействия электронов с ядром принимает своё минимальное значение, и говорят, соответственно, что *в основном состоянии атом обладает наименьшей энергией.*

[14] Исключением являются *лазеры.*

Но атом обладает способностью поглощать энергию. Под действием внешних факторов — например, в результате соударений с другими атомами или поглощения света — атом может перейти в *возбуждённое* состояние. Это значит, что какой-либо электрон покидает «насиженное место» на своей основной орбите и переходит на орбиту, расположенную дальше от ядра. Потенциальная энергия взаимодействия электрона с ядром при этом возрастает; соответственно, *в возбуждённом состоянии энергия атома больше, чем в основном.* Величина ΔW, на которую энергия возбуждённого состояния превышает энергию основного состояния, в точности равна энергии, поглощённой атомом.

Опыт показывает, однако, что в возбуждённом состоянии атом долго не пробудет и в итоге вернётся в основное состояние. В процессе этого перехода энергия атома уменьшится и снова примет своё минимально возможное значение. Куда же при этом денется энергия ΔW?

Атом вернёт поглощённую энергию ΔW в виде излучения. *В результате перехода из возбуждённого состояния в основное атом излучает электромагнитную волну конечной длительности — так называемый волновой цуг* (рис. 4.79).

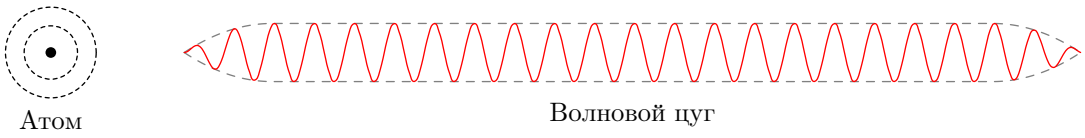


Рис. 4.79. Излучение атомом волнового цуга

Длительность цуга τ порядка 10^{-8} с; соответственно, длина цуга $c\tau \approx 3$ м. Частота цуга может находиться в видимом диапазоне, и тогда цуг будет регистрироваться человеческим глазом.

Выходит, что свет — это поток цугов, излучённых атомами. Почему, например, горит электрическая лампочка? Когда через лампочку идёт ток, атомы спирали совершают интенсивное тепловое движение и сталкиваются друг с другом, то и дело переходя в возбуждённое состояние; затем, возвращаясь в основное состояние, они испускают цуги в видимом диапазоне. Благодаря этому беспрерывному процессу мы и наблюдаем свет, постоянно идущий от лампочки.

Естественно спросить: если атом находится в возбуждённом состоянии, то когда именно надо ждать его возвращения в основное состояние? Оказывается, точный момент этого перехода предвидеть невозможно. *Переход возбуждённого атома в основное состояние происходит в случайный, непредсказуемый момент времени.* Это означает, что моменты испускания цугов различными атомами никак не согласованы между собой! В результате *цуги, образующие свет, даже если и обладают одной частотой, имеют совершенно произвольные начальные фазы.*

Вот теперь мы можем объяснить, почему не образуется устойчивая интерференционная картина от двух независимых источников, излучающих свет с одинаковой длиной волны.

Пусть на экране встречаются два цуга: один излучён атомом первого источника, а другой излучён атомом второго источника. Частоты цугов одинаковы, поэтому они проинтерферируют друг с другом и дадут на экране некоторую интерференционную картину. Положение интерференционных максимумов и минимумов будет определяться разностью фаз между цугами.

Эта интерференционная картина, однако, проживёт недолго — порядка 10^{-8} с, пока длятся цуги. Затем её сменят вторая, третья (и так далее) интерференционные картины, образованные парами других цугов. Максимумы и минимумы интерференции будут занимать всё новые и новые положения, поскольку каждая пара цугов приходит с непредсказуемой разностью фаз.

Таким образом, за секунду сменят друг друга порядка ста миллионов интерференционных картин, причём положения интерференционных полос беспорядочно меняются от картины к картине. Глаз не способен уследить за таким частым мельтешением полос, и потому мы видим лишь равномерно освещённый экран.

4.11.3 Зеркала Френеля

Но если независимые источники света некогерентны, то можно ли вообще наблюдать интерференцию света? Оказывается, можно!

Замечательная идея состоит в следующем. Раз уж не получается использовать два различных источника света, давайте возьмём *один и тот же* источник и разделим свет от него на два пучка. Оба пучка сведём на экране, причём так, чтобы они попадали на экран *различными* путями.

Тогда каждый цуг, излучённый источником, окажется расщеплённым на два цуга, которые проинтерферируют друг с другом. Однако теперь всякая новая пара цугов будет иметь один и тот же неизменный сдвиг фаз, определяемый разностью хода между двумя пучками.

Что же тогда получится? Снова будут сменять друг друга сто миллионов интерференционных картин в секудну, но на сей раз полосы каждой картины занимают одни и те же, строго фиксированные положения на экране. Снова глаз не сможет различить этого мелькания, но теперь вместо равномерной освещённости экрана мы увидим устойчивую интерференционную картину!

Давайте посмотрим, как работает эта идея, на примере одного из первых классических интерференционных опытов — зеркал Френеля.

На рис. 4.80 изображена схема этого эксперимента. Два плоских зеркала OA и OB образуют почти развёрнутый угол и создают два близко расположенных изображения S_1 и S_2 точечного источника света S. Вдали расположен экран; ширма закрывает экран от прямых лучей источника. На экран, таким образом, попадают лишь лучи, отражённые от зеркал.

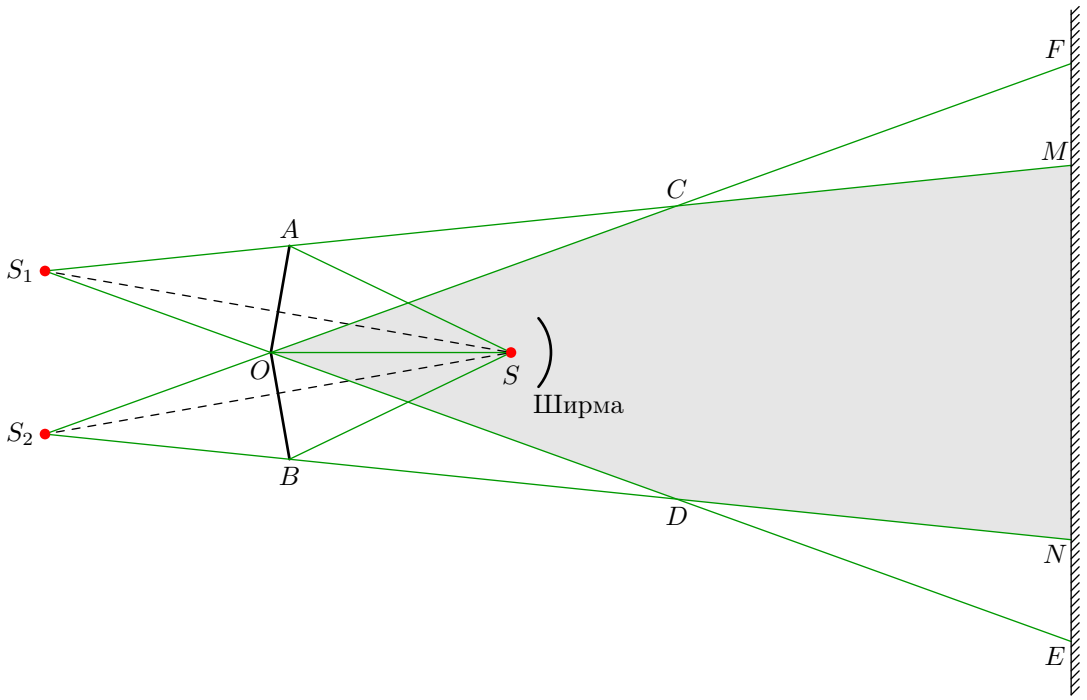


Рис. 4.80. Интерференционный опыт с зеркалами Френеля

Световые лучи, как всегда, изображены зелёным цветом. Направления лучей мы уже не указываем, чтобы не загромождать рисунок. К тому же, у вас позади геометрическая оптика, так что вы легко поймёте ход лучей и без указания их направления :-)

Лучи, отражённые зеркалом OA, образуют пучок $MAOE$, который как бы исходит из мнимого изображения S_1 источника S. Аналогично, лучи, отражённые зеркалом OB, образуют пучок $FOBN$, как бы исходящий из мнимого изображения S_2.

Эти пучки оказываются когерентными, поскольку когерентны мнимые источники S_1 и S_2. Действительно, эти источники суть изображения *одного и того же* источника S, поэтому их частоты совпадают и сдвиг фаз между ними равен нулю. Следовательно, в области $MCODN$, где перекрываются пучки, можно наблюдать устойчивую интерференционную картину. Фактически же, повторяем, в каждой точке данной области в каждый момент времени накладывается сам на себя один и тот же цуг — с одним и тем же, фиксированным для данной точки сдвигом фаз, определяемым разностью хода от источников S_1 и S_2.

Теперь мы видим, что данная ситуация совершенно аналогична задаче об интерференции волн двух когерентных точечных источников, вычисления в которой проводятся по схеме Юнга. В частности, ширина интерференционных полос, наблюдаемых в опыте с зеркалами Френеля, равна $\lambda L/a$, где $a = S_1S_2$ и L — расстояние от прямой S_1S_2 до экрана. Величины a и L несложно найти геометрически.

4.11.4 Интерференция в тонких плёнках

Глядя на переливающийся различными цветами мыльный пузырь, на радужные отблески масляных или бензиновых пятен на поверхности воды, мы, оказывается, наблюдаем не что иное, как интерференцию света!

Давайте посмотрим на рис. 4.81. На поверхность тонкой прозрачной плёнки падает световой луч AO.

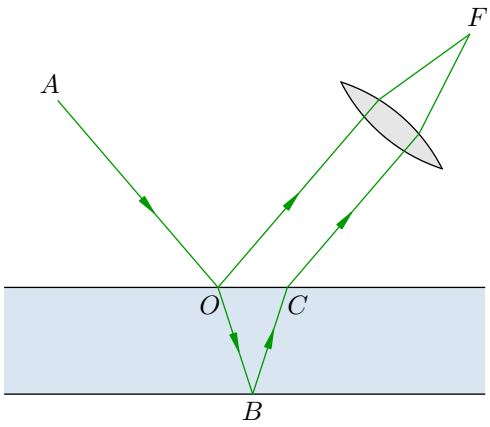


Рис. 4.81. Интерференция в тонкой плёнке

Падающий луч расщепляется на два луча: отражённый луч OF и преломлённый луч OB. После вторичного отражения и преломления из плёнки выходит второй луч CF, параллельный отражённому лучу.

Оба луча фокусируются собирающей линзой в точке F. Это может быть самая обычная линза (при наблюдении интерференционной картины на экране) или оптическая система нормального глаза (при непосредственном разглядывании).

Обе волны OF и CF, будучи частями одной и той же волны AO, являются когерентными. Действительно, в точке F сходятся две копии одного цуга с некоторым фиксированным сдвигом фаз между собой; этот сдвиг фаз обеспечивается разностью хода между волнами OF и $OBCF$. Поэтому волны OF и CF интерферируют друг с другом, давая картину чередующихся максимумов и минимумов в окружающем пространстве.

Предположим, что плёнка освещается белым светом. Как вы знаете, белый свет является смесью волн с различными частотами; эти частоты отвечают цветам от красного до фиолетового. Пусть, например, разность хода между волнами OF и $OBCF$ равна целому числу длин волн красного света. Тогда красная составляющая белого света усилит сама себя, и отражённый плёнкой свет нам будет казаться красным.

При небольшом изменении угла падения (или толщины плёнки) изменится и разность хода. Поэтому, если поверхность плёнки является неровной (или если мы посмотрим чуть с другого направления), то новая разность хода может стать равна целому числу длин волн, например, зелёного света. Теперь произойдёт усиление зелёной составляющей белого света, и отражённый от плёнки свет мы увидим зелёным.

Всё это мы наблюдаем, рассматривая мыльный пузырь. Перемещение его поверхности приводит к постоянному изменению разности хода для данного ракурса. Происходит усиление то одного цвета, то другого, и в результате пузырь переливается цветами радуги.

4.11.5 Кольца Ньютона

Возьмём плоско-выпуклую линзу с достаточно большим радиусом сферической поверхности и положим её выпуклостью вниз на стеклянную пластину. Если глядеть сверху, то сквозь линзу можно увидеть интерференционную картину в виде концентрических колец.

Это так называемые *кольца Ньютона*; они изучались Ньютоном при освещении как белым, так и монохроматическим светом. Кольца Ньютона в красном свете показаны[15] на рис. 4.82.

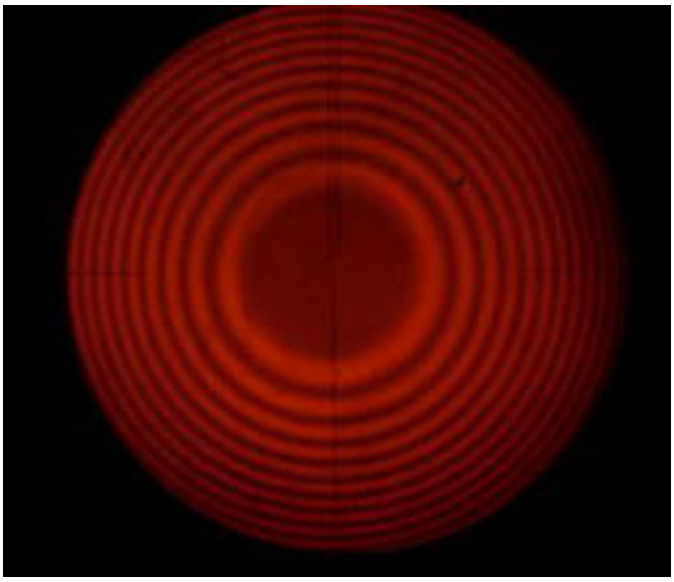


Рис. 4.82. Кольца Ньютона в красном свете

Происхождение колец Ньютона вполне аналогично интерференции в тонких плёнках. Взгляните на рис. 4.83.

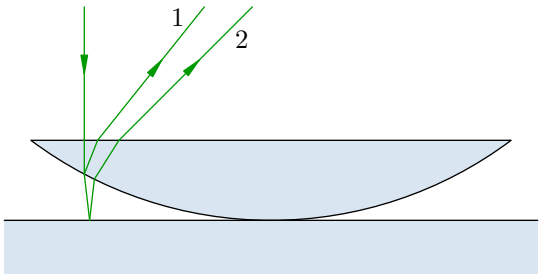


Рис. 4.83. Происхождение колец Ньютона

Падающий луч расщепляется на два луча 1 и 2, отражённых соответственно от сферической поверхности линзы и от пластины; между этими лучами возникает разность хода, и они

[15]Изображение с сайта femto.com.ua.

интерферируют между собой. Все три луча, изображённые на рисунке, в реальности почти сливаются друг с другом из-за малой кривизны поверхности линзы.

Вычислим радиусы светлых колец Ньютона. Пусть точка падения луча на сферическую поверхность находится на расстоянии y от пластины (рис. 4.84).

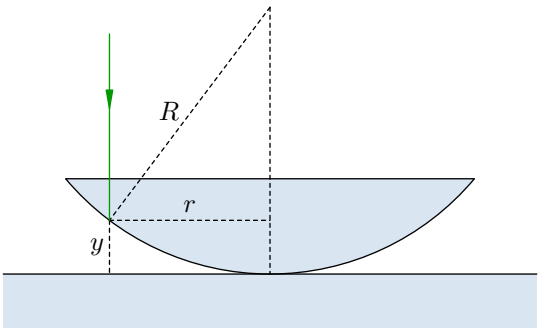


Рис. 4.84. К расчёту радиусов колец

Пусть R — радиус кривизны сферической поверхности линзы, r — расстояние от точки падения до оси симметрии линзы. Имеем:

$$r^2 = R^2 - (R-y)^2 = 2Ry - y^2.$$

Поскольку воздушная прослойка очень тонка ($y \ll R$), величиной y^2 можно пренебречь по сравнению с $2Ry$:

$$r^2 = 2Ry.$$

Отсюда

$$y = \frac{r^2}{2R}\,.$$

Как видно из рис. 4.83, путь второго луча превышает путь первого луча примерно на $2y$. Однако разность хода будет больше, чем $2y$, поскольку вмешивается один важный эффект.

Взгляните ещё раз на рис. 4.82. Почему в центре картины наблюдается тёмное пятно? Ведь разность хода лучей там равна нулю, и, казалось бы, наоборот — в центре должен наблюдаться максимум.

Причину этого таинственного минимума разгадал гениальный английский учёный Томас Юнг. Оказывается, от более плотной среды свет отражается со сдвигом на полволны!

На рис. 4.85 слева показано отражение на границе воздух-стекло. Обратите внимание: фаза отражённой волны отличается на π от фазы падающей волны. Это общий факт: *при отражении от оптически более плотной среды (то есть от среды с бóльшим показателем преломления) происходит изменение фазы колебаний на π, что равносильно сдвигу отражённой волны относительно падающей на половину длины волны.*

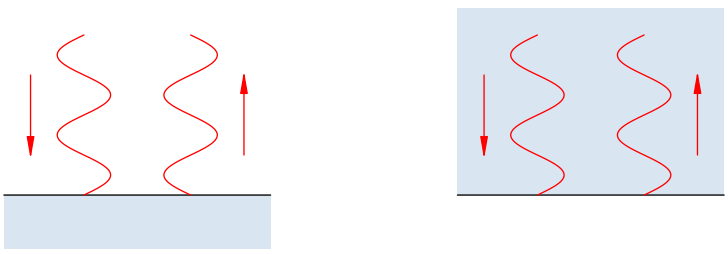


Рис. 4.85. Отражение со сдвигом на полволны и без него

Справа на рис. 4.85 показано отражение на границе стекло-воздух. Изменения фазы нет! И это общий факт: *при отражении от оптически менее плотной среды фазы отражённой и падающей волн совпадают.*

Вот в этом-то и заключается причина того, что центральное пятно — тёмное. Один луч отражается от поверхности стеклянной пластины и получает сдвиг на полволны. А другой луч отражается от поверхности линзы без изменения фазы. Интерферируя, эти лучи гасят друг друга.

Возвращаясь теперь к рис. 4.83 и 4.84, мы видим, что луч 2 не только проходит дополнительный путь $2y$, но и сдвигается на полволны при отражении на границе воздух–пластина. Луч 1 не испытывает такого сдвига, поскольку отражается на границе линза–воздух. Поэтому разность хода δ между лучами 1 и 2 оказывается больше, чем $2y$, на половину длины волны:

$$\delta = 2y + \frac{\lambda}{2} = \frac{r^2}{R} + \frac{\lambda}{2}\,.$$

Светлые кольца будут в местах интерференционных максимумов, когда разность хода равна целому числу длин волн. Имеем:

$$\frac{r^2}{R} + \frac{\lambda}{2} = n\lambda \quad (n = 1, 2, \ldots).$$

Отсюда получаем радиусы светлых колец:

$$r_n = \sqrt{\left(n - \frac{1}{2}\right)\lambda R} \quad (n = 1, 2, \ldots).$$

Как видим, радиус растёт с увеличением номера кольца. Кроме того, радиус кольца с заданным порядковым номером возрастает при переходе от фиолетового цвета к красному (поскольку увеличивается длина волны).

Радиусы тёмных колец вычисляются аналогично — надо только разность хода δ приравнять к полуцелому числу длин волн:

$$\delta = \frac{r^2}{R} + \frac{\lambda}{2} = n\lambda + \frac{\lambda}{2}\,.$$

Отсюда получаем радиус n-го тёмного кольца:

$$r_n = \sqrt{n\lambda R} \quad (n = 1, 2, \ldots).$$

Радиусы тёмных колец, как видим, увеличиваются пропорционально квадратному корню из номера кольца. Этот факт был отмечен ещё Ньютоном. Но Ньютон, трактовавший свет как поток частиц, не смог дать удовлетворительного объяснения наблюдаемой картине колец. Полное объяснение было дано Юнгом[16] на основе представлений о волновой природе света и интерференции когерентных световых пучков.

4.11.6 Просветление оптики

Пожалуй, самым широким на сегодняшний день применением интерференции света служит *просветление оптики*. Расскажем вкратце, что это такое.

[16]Имя Томаса Юнга нам уже встречалось и встретится дальше. Это одна из самых удивительных фигур в истории науки. Юнг не только был выдающимся физиком, но и обладал многочисленными талантами в совершенно разных областях: медицине, филологии, живописи, музыке и даже искусстве пляски на канате. Так, Юнгу принадлежит решающая идея, положившая начало расшифровке египетских иероглифов.

Свет, падающий на линзу, частично отражается назад; доля отражённого света обычно составляет несколько процентов. Объективы современной оптической техники представляют собой системы линз (числом до нескольких десятков). В результате отражений на поверхности каждой линзы происходит значительное ослабление света: в сумме на отражениях может теряться до 90% световой энергии. Освещённость изображений предметов, даваемых такой оптической системой, будет чрезвычайно низкой.

Как уменьшить потери на отражение? Для этого на поверхность линзы наносят *интерференционное покрытие* в виде тонкой плёнки (рис. 4.86).

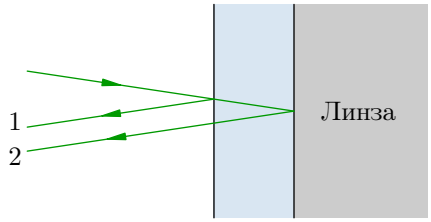


Рис. 4.86. Просветление оптики

Толщина покрытия подбирается так, чтобы отражённые волны 1 и 2 были сдвинуты на полволны и, интерферируя, погасили друг друга. Тогда не будет потерь на отражение, и вся световая энергия пройдёт через линзу. Изображение получится более ярким — оптика «просветляется».

Толщина интерференционного покрытия, гасящего отражённую волну, зависит от длины волны, и поэтому добиться полного гашения отражённых волн во всём видимом диапазоне не получается. Обычно стараются, чтобы сквозь объектив прошла без потерь средняя, жёлто-зелёная часть видимого спектра (в которой лежит максимум интенсивности солнечного излучения). В таком случае в отражённых лучах доминируют крайние части спектра — красная и фиолетовая; их смесью, например, является хорошо известный вам сиреневый отблеск объектива фотоаппарата.

4.12 Дифракция света

Если на пути волны возникает препятствие, то происходит *дифракция* — отклонение волны от прямолинейного распространения. Это отклонение не сводится к отражению или преломлению, а также искривлению хода лучей вследствие изменения показателя преломления среды. Дифракция состоит в том, что волна *огибает* край препятствия и заходит в область геометрической тени.

Пусть, например, плоская волна падает на экран с достаточно узкой щелью (рис. 4.87). На выходе из щели мы увидим расходящуюся волну! Прямолинейный ход волны нарушается, и волновой процесс достигает тех точек, где должна была бы находиться тень от экрана.

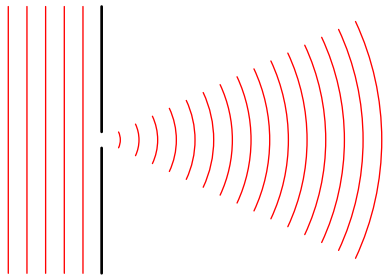


Рис. 4.87. Дифракция на щели

Расходимость волны в результате дифракции на щели будет тем больше, чем больше отношение λ/a, где λ — длина волны, a — ширина щели. Поэтому дифракционные явления выражены тем отчётливей, чем мельче препятствие. Наиболее существенна дифракция в тех случаях, когда размер препятствия меньше или порядка длины волны.

Дифракция, как и интерференция, свойственна всем видам волн — механическим и электромагнитным. На опыте наблюдается *дифракция света*. Именно способность света к интерференции и дифракции послужила в своё время доказательством волновой природы света и дала начало волновой оптике.

Правда, наблюдение дифракционных явлений в оптике затруднено ввиду малости длины световой волны. Так, при $\lambda = 500\,\text{нм}$ и размере щели $a = 0{,}5\,\text{мм}$ имеем: $\lambda/a = 0{,}001$. Такова примерно величина угла (в радианах), характеризующая расхождение пучка после дифракции. Поэтому, чтобы зафиксировать отклонение световых лучей от прямолинейного распространения, экран должен находиться достаточно далеко (на расстоянии нескольких метров, а то и нескольких десятков метров).

Что же мы увидим на экране? Вот пример: на рис. 4.88 изображена дифракционная картина[17], полученная в результате прохождения лазерного луча сквозь отверстие диаметром 0,2 мм.

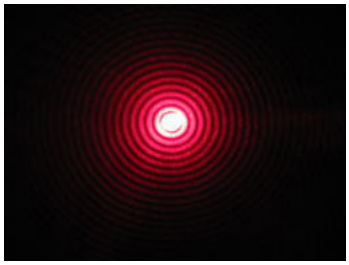


Рис. 4.88. Дифракция лазерного луча на отверстии

Напоминает интерференцию, не правда ли? Это она и есть; чередующиеся светлые и тёмные кольца являются интерференционными максимумами и минимумами. Какие же волны тут

[17]Изображение с сайта en.wikipedia.org.

интерферируют?

«Чистый» принцип Гюйгенса ничего не говорит об интерференции и уж тем более не объясняет её. Соответственно, на основании принципа Гюйгенса нельзя понять и дифракцию. Необходимое усиление этого принципа было сделано Френелем.

4.12.1 Принцип Гюйгенса–Френеля

Напомним формулировку принципа Гюйгенса: каждая точка, вовлечённая в волновой процесс, является источником вторичных сферических волн; эти волны распространяются от данной точки, как из центра, во все стороны и накладываются друг на друга.

Что значит «накладываются»? Гюйгенс понимал это лишь как геометрический способ построения новой волновой поверхности в качестве огибающей семейства сфер, расширяющихся от каждой точки исходной волновой поверхности. Вторичные волны Гюйгенса — это математические сферы, а не реальные волны; их суммарное действие проявляется только на огибающей, т. е. на новом положении волновой поверхности.

В таком виде принцип Гюйгенса не давал ответа на естественный вопрос, почему в процессе распространения волны не возникает волна, идущая в обратном направлении. Не объяснёнными, как мы уже сказали выше, оставались и дифракционные явления.

Модификация принципа Гюйгенса состоялась лишь спустя 137 лет. Огюстен Френель заменил вспомогательные геометрические сферы Гюйгенса на реальные волны и предположил, что эти волны *интерферируют* друг с другом.

Принцип Гюйгенса–Френеля. Каждая точка волновой поверхности служит источником вторичных сферических волн. Все эти вторичные волны являются когерентными ввиду общности их происхождения от первичного источника (и, стало быть, могут интерферировать друг с другом); волновой процесс в окружающем пространстве есть результат интерференции вторичных волн.

Идея Френеля наполнила принцип Гюйгенса физическим смыслом. Вторичные волны, интерферируя, усиливают друг друга на огибающей своих волновых поверхностей в направлении «вперёд», обеспечивая дальнейшее распространение волны. А в направлении «назад» происходит их интерференция с исходной волной, наблюдается взаимное гашение, и обратная волна не возникает.

В частности, свет распространяется там, где вторичные волны взаимно усиливаются. А в местах ослабления вторичных волн мы будем видеть тёмные участки пространства.

Принцип Гюйгенса–Френеля выражает важную физическую идею: волна, удалившись от своего источника, в дальнейшем «живёт своей жизнью» и уже никак от этого источника не зависит. Захватывая новые участки пространства, волна распространяется всё дальше и дальше вследствие интерференции вторичных волн, возбуждённых в различных точках пространства по мере прохождения волны.

Итак, какое же объяснение даёт принцип Гюйгенса–Френеля явлению дифракции? Каким образом, например, происходит дифракция на отверстии?

Экранное отверстие вырезает из бесконечной плоской волновой поверхности маленький светящийся диск. Последующее световое поле получается в результате интерференции волн вторичных источников, расположенных уже не на всей плоскости, а лишь на этом диске. Естественно, новые волновые поверхности теперь не будут плоскими; ход лучей искривляется, и волна начинает распространяться в разных направлениях, не совпадающих с первоначальным. Волна огибает края отверстия и проникает в область геометрической тени.

Вторичные волны, испущенные различными точками вырезанного светлого диска, интерферируют друг с другом. Результат интерференции определяется разностью фаз вторичных волн и зависит от угла отклонения лучей. В результате возникает чередование интерференционных

максимумов и минимумов — что мы и видели на рис. 4.88.

Френель не только дополнил принцип Гюйгенса важной идеей когерентности и интерференции вторичных волн, но и придумал свой знаменитый метод решения дифракционных задач, основанный на построении так называемых *зон Френеля*. Изучение зон Френеля не входит в школьную программу — о них вы узнаете уже в вузовском курсе физики. Здесь мы отметим лишь, что закон прямолинейного распространения света (а именно, образование теней непрозрачных предметов, совпадающих по форме с самими предметами) получает своё качественное объяснение именно в рамках метода зон Френеля.

4.12.2 Опыт Юнга

Рассмотрим самый первый из классических опытов по наблюдению интерференции света. Его придумал Томас Юнг, и в нём существенно используется явление дифракции.

Всякий эксперимент с интерференцией света содержит некоторый способ получения двух когерентных световых волн. В опыте с зеркалами Френеля, как вы помните, когерентными источниками являлись два изображения одного и того же источника, полученные в обоих зеркалах.

Простейшая идея, которая возникла прежде всего, состояла в следующем. Давайте проколем в куске картона два отверстия и подставим под солнечные лучи. Эти отверстия будут когерентными вторичными источниками света, поскольку первичный источник один — Солнце. Следовательно, на экране в области перекрытия пучков, расходящихся от отверстий, мы должны увидеть интерференционную картину.

Такой опыт был поставлен задолго до Юнга итальянским учёным Франческо Гримальди (который открыл дифракцию света). Интерференции, однако, не наблюдалось. Почему же? Вопрос это не очень простой, и причина заключается в том, что Солнце — не точечный, а протяжённый источник света (угловой размер Солнца равен 30 угловым минутам). Солнечный диск состоит из множества точечных источников, каждый из которых даёт на экране свою интерференционную картину. Накладываясь, эти отдельные картины «смазывают» друг друга, и в результате на экране получается равномерная освещённость области перекрытия пучков.

Но если Солнце является чрезмерно «большим», то нужно искусственно создать *точечный* первичный источник. С этой целью в опыте Юнга использовано маленькое предварительное отверстие (рис. 4.89).

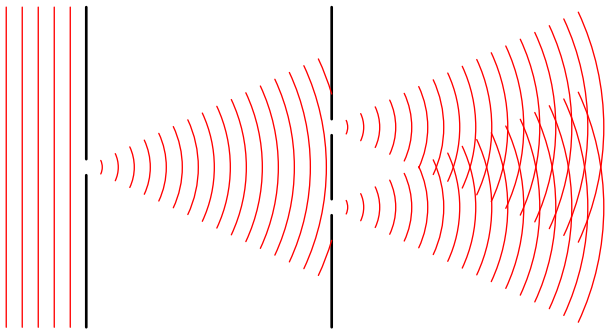


Рис. 4.89. Схема опыта Юнга

Плоская волна падает на первое отверстие, и за отверстием возникает световой конус, расширяющийся вследствие дифракции. Он достигает следующих двух отверстий, которые становятся источниками двух когерентных световых конусов. Вот теперь — благодаря точечности первичного источника — в области перекрытия конусов будет наблюдаться интерференционная картина!

Опыт Юнга вошёл в число наиболее знаменитых экспериментов в истории физики. Юнг вывел формулу (4.42) и измерил в своём опыте ширину Δx интерференционных полос. Оставалось выразить оттуда λ:

$$\lambda = \frac{a\Delta x}{L}\,.$$

С помощью этой формулы Юнг *впервые вычислил длины волн видимого света.*

4.12.3 Дифракционная решётка

Дифракционная решётка — это оптический прибор, позволяющий получать разложение света на спектральные составляющие и измерять длины волн. Дифракционные решётки бывают прозрачными и отражательными.

Мы рассмотрим прозрачную дифракционную решётку. Она состоит из большого числа щелей ширины a, разделённых промежутками ширины b (рис. 4.90). Свет проходит только сквозь щели; промежутки свет не пропускают. Величина $d = a + b$ называется *периодом* решётки.

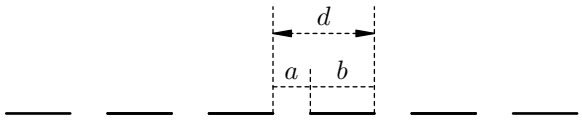


Рис. 4.90. Дифракционная решётка

Дифракционная решётка изготавливается с помощью так называемой делительной машины, которая наносит штрихи на поверхность стекла или прозрачной плёнки. При этом штрихи оказываются непрозрачными промежутками, а нетронутые места служат щелями. Если, например, дифракционная решётка содержит 100 штрихов на миллиметр, то период такой решётки будет равен: $d = 0{,}01$ мм $= 10$ мкм.

Сперва мы посмотрим, как проходит сквозь решётку монохроматический свет, т. е. свет со строго определённой длиной волны. Отличным примером монохроматического света служит луч лазерной указки (длина волны около 0,65 мкм).

На рис. 4.91 мы видим такой луч, падающий на одну из дифракционных решёток стандартного набора[18]. Щели решётки расположены вертикально, и на экране за решёткой наблюдаются периодически расположенные вертикальные полосы.

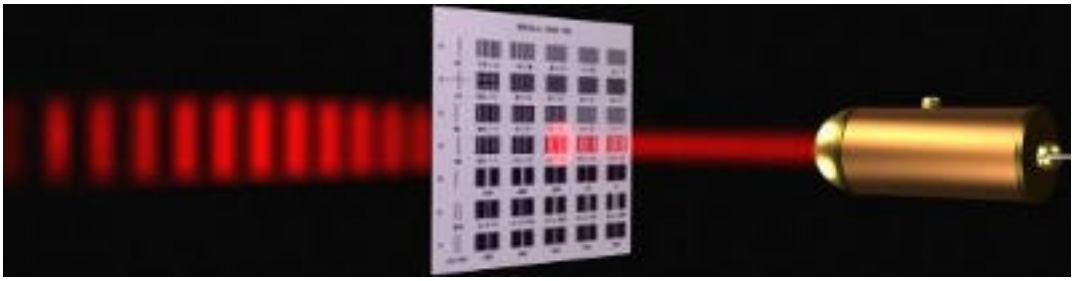


Рис. 4.91. Дифракция лазерного луча на решётке

Как вы уже поняли, это интерференционная картина. Дифракционная решётка расщепляет падающую волну на множество когерентных пучков, которые распространяются по всем направлениям и интерферируют друг с другом. Поэтому на экране мы видим чередование максимумов и минимумов интерференции — светлых и тёмных полос.

[18]Изображение с сайта physics.nad.ru.

Теория дифракционной решётки весьма сложна и во всей своей полноте оказывается далеко за рамками школьной программы. Вам следует знать лишь самые элементарные вещи, связанные с одной-единственной формулой; эта формула описывает положения максимумов освещённости экрана за дифракционной решёткой.

Итак, пусть на дифракционную решётку с перидом d падает плоская монохроматическая волна (рис. 4.92). Длина волны равна λ.

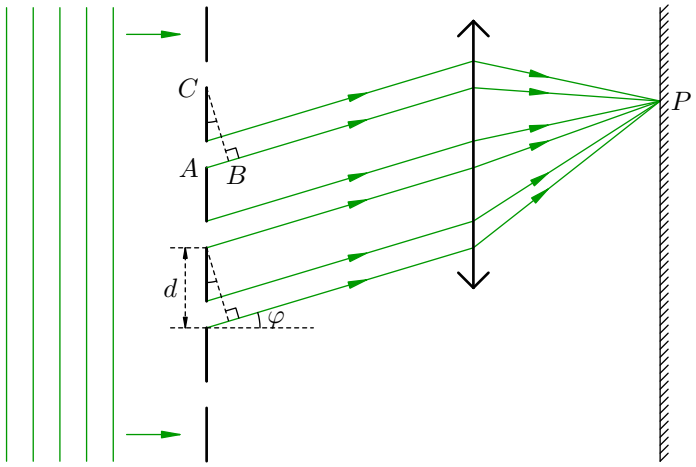


Рис. 4.92. Дифракция на решётке

Для большей чёткости интерференционной картины можно поставить линзу между решёткой и экраном, а экран поместить в фокальной плоскости линзы. Тогда вторичные волны, идущие параллельно от различных щелей, соберутся в одной точке P экрана (побочном фокусе линзы). Если же экран расположен достаточно далеко, то особой необходимости в линзе нет — лучи, приходящие в данную точку экрана от различных щелей, будут и так почти параллельны друг другу.

Рассмотрим вторичные волны, отклоняющиеся на угол φ. Разность хода между двумя волнами, идущими от соседних щелей, равна маленькому катету прямоугольного треугольника с гипотенузой d; или, что то же самое, эта разность хода равна катету AB треугольника ABC. Но угол ACB равен углу φ, поскольку это острые углы со взаимно перпендикулярными сторонами. Следовательно, наша разность хода равна $d\sin\varphi$.

Интерференционные максимумы наблюдаются в тех случаях, когда разность хода равна целому числу длин волн:

$$d\sin\varphi = k\lambda \quad (k = 0, 1, 2, \ldots). \tag{4.43}$$

При выполнении этого условия все волны, приходящие в точку P от различных щелей, будут складываться в фазе и усиливать друг друга[19].

Формула (4.43) позволяет найти углы, задающие направления на максимумы:

$$\sin\varphi_k = \frac{k\lambda}{d} \quad (k = 0, 1, 2, \ldots). \tag{4.44}$$

При $k = 0$ получаем $\varphi = 0$. Это *центральный максимум*, или максимум *нулевого порядка*. Разность хода всех вторичных волн, идущих без отклонения, равна нулю, и в центральном максимуме они складываются с нулевым сдвигом фаз. Центральный максимум — это центр дифракционной картины, самый яркий из максимумов. Дифракционная картина на экране симметрична относительно центрального максимума.

[19] Линза при этом не вносит дополнительной разности хода — несмотря на то, что разные лучи проходят через линзу разными путями. Почему так получается? Мы не будем вдаваться в этот вопрос, поскольку его обсуждение выходит за рамки ЕГЭ по физике.

При $k = 1$ получаем угол:

$$\varphi_1 = \arcsin \frac{\lambda}{d}\,.$$

Этот угол задаёт направления на *максимумы первого порядка*. Их два, и расположены они симметрично относительно центрального максимума. Яркость в максимумах первого порядка несколько меньше, чем в центральном максимуме.

Аналогично, при $k = 2$ имеем угол:

$$\varphi_2 = \arcsin \frac{2\lambda}{d}\,.$$

Он задаёт направления на *максимумы второго порядка*. Их тоже два, и они также расположены симметрично относительно центрального максимума. Яркость в максимумах второго порядка несколько меньше, чем в максимумах первого порядка.

Направления на максимумы первых двух порядков показаны на рис. 4.93.

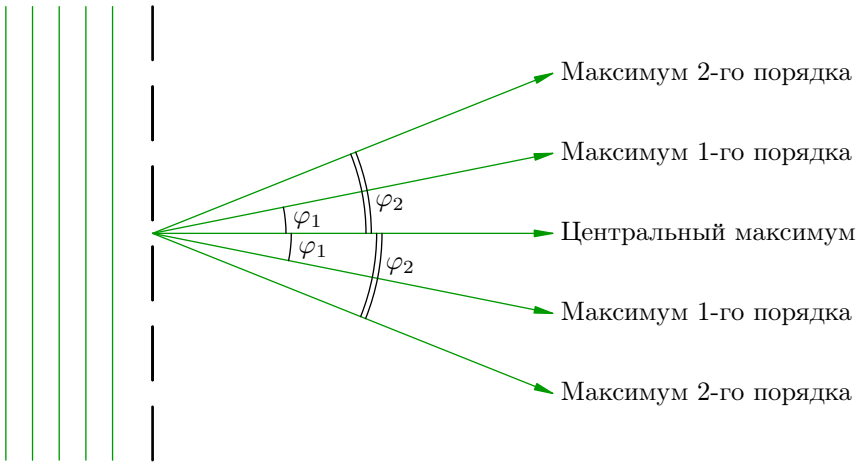


Рис. 4.93. Максимумы первых двух порядков

Вообще, два симметричных *максимума k-го порядка* определяются углом:

$$\varphi_k = \arcsin \frac{k\lambda}{d}\,. \tag{4.45}$$

При небольших k соответствующие углы обычно невелики. Например, при $\lambda = 0{,}65$ мкм и $d = 10$ мкм максимумы первого порядка расположены под углом $\varphi_1 = \arcsin(0{,}65/10) = 3{,}7°$.

Яркость максимумов k-го порядка постепенно убывает с ростом k. Сколько всего максимумов можно увидеть? На этот вопрос легко ответить с помощью формулы (4.44). Ведь синус не может быть больше единицы, поэтому:

$$k \leqslant \frac{d}{\lambda}\,.$$

Используя те же числовые данные, что и выше, получим: $k \leqslant 15{,}4$. Следовательно, наибольший возможный порядок максимума для данной решётки равен 15.

Посмотрите ещё раз на рис. 4.91. На экране видны 11 максимумов. Это центральный максимум, а также по два максимума первого, второго, третьего, четвёртого и пятого порядков.

С помощью дифракционной решётки можно измерить неизвестную длину волны. Направляем пучок света на решётку (период которой мы знаем), измеряем угол φ_1 на максимум первого порядка, пользуемся формулой (4.43) и получаем:

$$\lambda = d \sin \varphi_1.$$

4.12.4 Дифракционная решётка как спектральный прибор

Выше мы рассматривали дифракцию монохроматического света, каковым является лазерный луч. Часто приходится иметь дело с *немонохроматическим* излучением. Оно является смесью различных монохроматических волн, которые составляют *спектр* данного излучения. Например, белый свет — это смесь волн всего видимого диапазона, от красного до фиолетового.

Оптический прибор называется *спектральным*, если он позволяет раскладывать свет на монохроматические компоненты и тем самым исследовать спектральный состав излучения. Простейший спектральный прибор вам хорошо известен — это стеклянная призма. К числу спектральных приборов относится также и дифракционная решётка.

Предположим, что на дифракционную решётку падает белый свет. Давайте вернёмся к формуле (4.44) и подумаем, какие выводы из неё можно сделать.

Положение центрального максимума ($\varphi = 0$) не зависит от длины волны. В центре дифракционной картины сойдутся с нулевой разностью хода *все* монохроматические составляющие белого света. Поэтому в центральном максимуме мы увидим яркую белую полосу.

А вот положения максимумов порядка $k \geqslant 1$ определяются длиной волны. Чем меньше λ, тем меньше угол φ_k для данного k. Поэтому в максимуме k-го порядка монохроматические волны разделяются в пространстве: самой близкой к к центральному максимуму окажется фиолетовая полоса, самой далёкой — красная.

Следовательно, в каждом порядке $k \geqslant 1$ белый свет раскладывается решёткой в спектр. Максимумы первого порядка всех монохроматических компонент образуют *спектр первого порядка*; затем идут спектры второго, третьего и так далее порядков. Спектр каждого порядка имеет вид цветной полосы, в которой присутствуют все цвета радуги — от фиолетового до красного.

Дифракция белого света показана[20] на рис. 4.94. Мы видим белую полосу в центральном максимуме, а по бокам — два спектра первого порядка. По мере возрастания угла отклонения цвет полос меняется от фиолетового к красному.

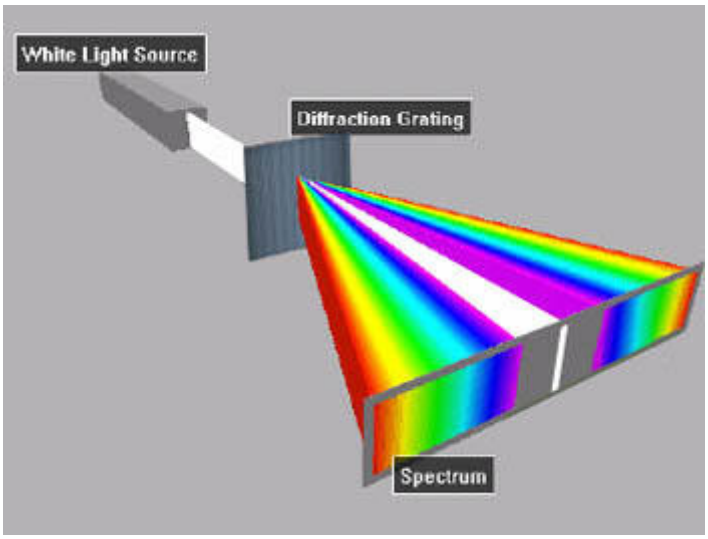


Рис. 4.94. Дифракция белого света на решётке

Но дифракционная решётка не только позволяет наблюдать спектры, т. е. проводить качественный анализ спектрального состава излучения. Важнейшим достоинством дифракционной решётки является возможность количественного анализа — как уже говорилось выше, мы с её помощью можем *измерять* длины волн. При этом измерительная процедура весьма проста: фактически она сводится к измерению угла направления на максимум.

[20]Изображение с сайта h2physics.org.

Естественными примерами дифракционных решёток, встречающихся в природе, являются перья птиц, крылья бабочек, перламутровая поверхность морской раковины. Если, прищурившись, посмотреть на солнечный свет, то можно увидеть радужную окраску вокруг ресниц. Наши ресницы действуют в данном случае как прозрачная дифракционная решётка на рис. 4.92, а в качестве линзы выступает оптическая система роговицы и хрусталика.

Спектральное разложение белого света, даваемое дифракционной решёткой, проще всего наблюдать, глядя на обычный компакт-диск (рис. 4.95)[21]. Оказывается, дорожки на поверхности диска образуют отражательную дифракционную решётку!

Рис. 4.95. Компакт-диск как отражательная решётка

[21]Изображение с сайта en.wikipedia.org.

4.13 Дисперсия света

Пусть солнечный луч переходит из воздуха в прозрачную среду (например, воду или стекло). Если угол падения α не равен нулю, то, как вы помните, угол преломления β определяется из закона преломления:

$$\sin\beta = \frac{\sin\alpha}{n}\,.$$

Величина n, называемая показателем преломления, характеризует среду и от угла падения не зависит.

Оказывается, однако, что среда по-разному реагирует на прохождение электромагнитных волн различных частот. Имеет место *дисперсия — зависимость показателя преломления среды от частоты света.*

4.13.1 Опыт Ньютона

Классический опыт по наблюдению дисперсии был поставлен Ньютоном. Узкий луч солнечного света направлялся на треугольную стеклянную призму (рис. 4.96).

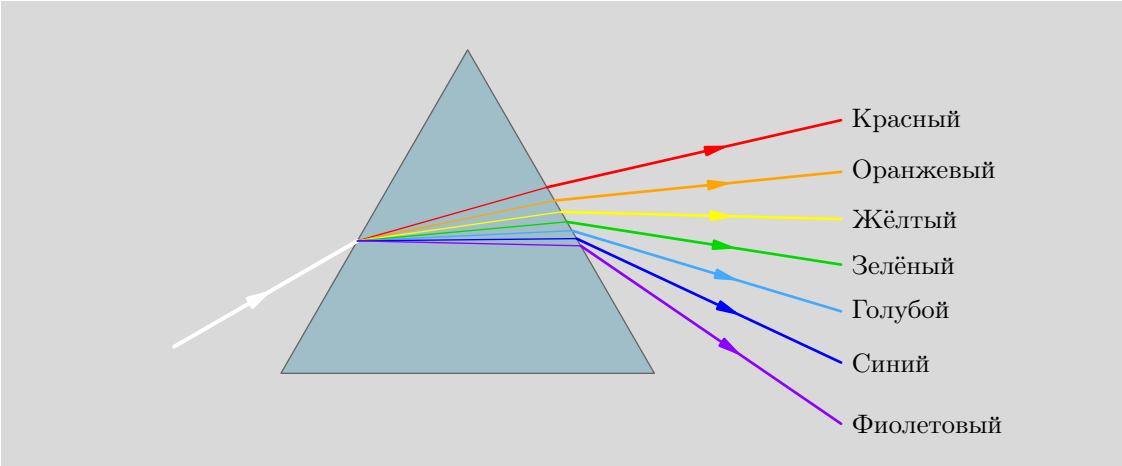


Рис. 4.96. Разложение белого света в спектр

На экране за призмой появлялся *спектр* — радужная полоса. Один край спектра оказался красным, другой — фиолетовым, а цвета внутри спектра непрерывно переходили друг в друга.

Выделяя луч какого-либо цвета (например, красного или синего) и запуская его в другую призму, мы уже не увидим изменения цвета преломлённого луча. Стало быть, компоненты радуги являются простейшими цветами, не разложимыми далее. Их можно собрать обратно с помощью второй призмы, и тогда снова получится белый свет. Следовательно, белый свет является смесью световых пучков различных цветов, непрерывно заполняющих диапазон видимого света от красного до фиолетового.

Мы видим, таким образом, что стеклянная призма является простейшим *спектральным прибором* — она позволяет исследовать спектральный состав белого света. С действием более сложного спектрального прибора — дифракционной решётки — мы познакомились в предыдущем разделе.

Как показывает опыт Ньютона, слабее всего преломляется красный свет, а сильнее всего — фиолетовый. В видимом диапазоне красный свет имеет наименьшую частоту, а фиолетовый — наибольшую. Коль скоро показатель преломления становится всё больше по мере движения от красного конца спектра к фиолетовому, мы делаем вывод, что *показатель преломления стекла увеличивается с возрастанием частоты света.*

Но показатель преломления есть отношение скорости света в воздухе к скорости света в среде: $n = c/v$. Значит, *чем больше частота света, тем с меньшей скоростью свет распространяется в стекле*. Наибольшую скорость внутри стеклянной призмы имеет красный свет, наименьшую — фиолетовый.

Различие в скоростях света для разных частот проявляется только при наличии среды. В вакууме скорость распространения электромагнитных волн не зависит от частоты и равна c.

Открытая и исследованная Ньютоном, дисперсия света больше двухсот лет ждала своего объяснения — нужны были соответствующие сведения о строении вещества. Классическая теория дисперсии была предложена Лоренцем лишь в конце XIX века. Более точная квантовая теория дисперсии появилась в первой половине прошлого столетия.

4.13.2 Хроматическая аберрация

Предположим, что на собирающую линзу параллельно главной оптической оси падает пучок белого света. Преломляясь в линзе, он, казалось бы, должен собраться в её фокусе. Однако вследствие дисперсии возникает *хроматическая аберрация* — некоторая расфокусировка пучка, вызванная различной преломляемостью разных компонент белого света.

Явление хроматической аберрации показано на рис. 4.97.

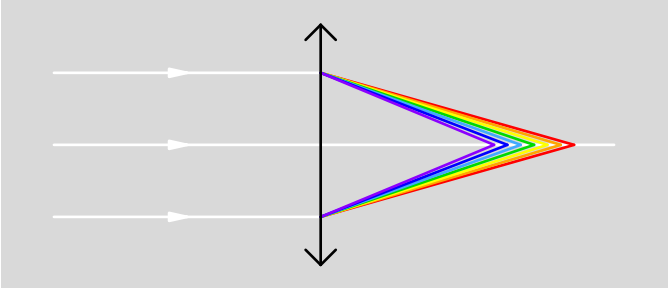


Рис. 4.97. Хроматическая аберрация

Показатель преломления материала линзы принимает наименьшее значение для красного света, и потому красный свет преломляется слабее всего. Красные лучи собираются на главной оптической оси в наиболее удалённой от линзы точке. Жёлтые лучи собираются ближе к линзе, зелёные — ещё ближе, и, наконец, в ближайшей к линзе точке сойдутся фиолетовые лучи.

Хроматическая аберрация ухудшает качество изображений — снижает чёткость, даёт лишние цветные полосы. Но с хроматической аберрацией можно бороться. Для этого в оптической технике применяют так называемые ахроматические линзы, получаемые накладыванием на собирающую линзу дополнительной рассеивающей линзы. Догадайтесь — зачем нужна рассеивающая линза?

Глава 5

Теория относительности

Классическая механика Ньютона отлично работает, когда скорости движения тел малы по сравнению со скоростью света.

Однако по мере приближения скорости движения тела к скорости света всё отчётливее проявляются новые, так называемые *релятивистские* эффекты: сокращение длины, замедление времени и т. д. Эти эффекты не описываются классической механикой. Здесь работает более общая физическая теория — *специальная теория относительности* (СТО).

Теория относительности остаётся справедливой и при небольших скоростях движения, однако в этом случае релятивистские эффекты становятся исчезающе малыми и не оказывают существенного влияния. Вот и получается, что «в обычной жизни» теория относительности переходит в классическую механику. Можно сказать, что классическая механика является предельным случаем СТО для малых скоростей движения.

Теория относительности изложена в следующих разделах пособия:

→ Принцип относительности Галилея
→ Принципы СТО
→ Релятивистская кинематика
→ Релятивистская динамика

5.1 Принцип относительности Галилея

Изучение теории относительности Эйнштейна мы начинаем с более глубокого рассмотрения принципа относительности Галилея. Это позволит нам лучше понять, каковы были предпосылки создания теории относительности.

Рассмотрим мысленный эксперимент с кораблём, предложенный Галилеем.

5.1.1 Наблюдатель на корабле

Представьте себе, что вы находитесь в каюте корабля. Никакого движения в пространстве вы не ощущаете — вам кажется, что корабль стоит на месте. Но вас всё же интересует, покоится ли корабль или движется равномерно и прямолинейно. Можете ли вы установить это, не выглядывая в иллюминатор?

Допустим, что с данной целью вы производите всевозможные эксперименты, наблюдая различные *механические* явления в вашей каюте. Вы исследуете свободное падение тел, соскальзывание тела с наклонной плоскости, вращательное движение, колебания маятников, распространение звуковых волн... Вам детально известен ход этих явлений в неподвижной лаборатории на земле, и теперь вы пытаетесь найти какие-либо отклонения в их протекании, вызванные равномерным прямолинейным движением судна.

Никаких отклонений обнаружить не удастся! Поставив в каюте корабля любой механический эксперимент и сопоставив его с аналогичным экспериментом на земле, вы увидите, что полученные результаты не отличаются друг от друга. Например, вы бросаете мячик со скоростью 5 м/с под углом 60° к горизонту относительно палубы. Оказывается, мячик на корабле опишет ровно ту же самую траекторию, что и на берегу при тех же начальных условиях (скорость и угол броска).

Равномерное прямолинейное движение корабля никак не сказывается на протекании механических явлений на этом корабле. Поэтому никакой опыт из механики, проведённый в лаборатории корабля, не в состоянии определить, покоится ли корабль или движется равномерно и прямолинейно.

Систему отсчёта, связанную с землёй, во многих ситуациях можно считать инерциальной[1]. Система отсчёта корабля, движущаяся относительно земной системы отсчёта равномерно и прямолинейно, также будет инерциальной. Мы приходим к выводу, что с точки зрения механических явлений инерциальные системы отсчёта совершенно равноправны: никакой механический эксперимент не в состоянии выделить и сделать привилегированной какую-то одну инерциальную систему отсчёта по сравнению с остальными.

Это и есть принцип относительности, открытый Галилеем.

Принцип относительности Галилея. Всякое механическое явление при одних и тех же начальных условиях протекает одинаково в любой инерциальной системе отсчёта.

5.1.2 Инвариантность законов механики

Принцип относительности Галилея означает, что законы механики одинаковы во всех инерциальных системах отсчёта. А именно, математическая форма второго и третьего законов Ньютона не меняется при переходе от одной инерциальной системы отсчёта к другой. Давайте убедимся в этом непосредственно на следующем простом примере.

Рассмотрим две системы отсчёта: K и K'. Координатные оси этих систем сонаправлены. Систему K будем считать неподвижной. Система K' движется относительно неё с постоянной

[1] Конечно, она не инерциальна. Земля совершает суточное вращение и движется вокруг Солнца, поэтому земная лаборатория будет иметь ускорение. Но во многих задачах этим ускорением можно пренебречь.

скоростью $\vec{v}$ вдоль общего направления осей X и X' (рис. 5.1).

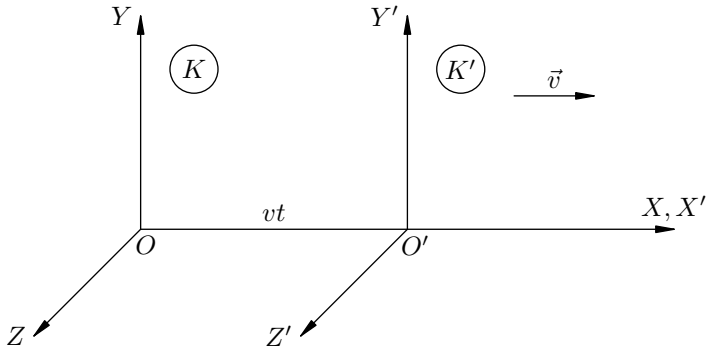


Рис. 5.1. Система K' движется относительно системы K

В тот момент, когда начала координат O и O' совпадали, часы обеих систем были выставлены на ноль и запущены. Стало быть, часы в системах K и K' идут синхронно, показывая одно и то же время t. В момент времени t расстояние OO' равно vt.

Нас интересует, как описывается движение тела (для определённости называемого далее *частицей*) в системах отсчёта K и K'.

Прежде всего, выясним, как связаны друг другом координаты частицы и моменты времени в обеих системах отсчёта.

Пусть в момент времени t по часам K частица имеет в системе K координаты x, y, z. Вообще, четвёрка чисел (x, y, z, t) называется *событием*. Событие состоит в том, что в данной точке пространства в данный момент времени что-то происходит — вот, например, в точке с координатами x, y, z в момент времени t оказывается наша частица.

В системе K' это же событие описывается четвёркой чисел (x', y', z', t'). А именно, местонахождение частицы в системе K' описывается координатами x', y', z', а часы K' показывают при этом время t'.

Глядя на рис. 5.1, совершенно ясно, что x' будет меньше x на величину vt, координата y' совпадает с y, а z' совпадает с z. Кроме того, как уже было сказано, время на часах K' и K одно и то же: $t' = t$.

Итак, имеем:

$$x' = x - vt, \quad y' = y, \quad z' = z, \quad t' = t. \tag{5.1}$$

Формулы (5.1) называются *преобразованиями Галилея*. Они связывают координаты и время одного и того же события, измеренные в разных инерциальных системах отсчёта: в движущейся системе K' и неподвижной системе K.

Таким образом, *преобразования Галилея в механике служат математическим описанием перехода от одной инерциальной системы отсчёта к другой.* Рассмотрим некоторые следствия, вытекающие из преобразований Галилея.

Пусть наша частица имеет в системе K скорость $\vec{u}$, а в системе K' — скорость $\vec{u}\,'$. Как связаны между собой эти скорости? Дифференцируем первые три равенства (5.1) по времени (которое одинаково в обеих системах отсчёта):

$$\dot{x}' = \dot{x} - v, \quad \dot{y}' = \dot{y}, \quad \dot{z}' = \dot{z}.$$

Производные координат по времени — это проекции скоростей:

$$u'_x = u_x - v, \quad u'_y = u_y, \quad u'_z = u_z. \tag{5.2}$$

Три равенства (5.2) можно записать в виде одной векторной формулы:

$$\vec{u}\,' = \vec{u} - \vec{v},$$

или

$$\vec{u} = \vec{u}\,' + \vec{v}.$$

Получился хорошо известный нам закон сложения скоростей: скорость тела относительно неподвижной системы отсчёта есть скорость тела относительно движущейся системы отсчёта плюс скорость движущейся системы относительно неподвижной. Мы видим, таким образом, что *закон сложения скоростей в механике является следствием преобразований Галилея.*

Дифференцируем по времени ещё раз — на сей раз соотношения (5.2). Производная постоянной величины v обращается в нуль, и мы получаем равенство ускорений:

$$a'_x = a_x, \quad a'_y = a_y, \quad a'_z = a_z,$$

или

$$\vec{a}\,' = \vec{a}.$$

Итак, *ускорение частицы одинаково во всех инерциальных системах отсчёта.* Это ещё одно следствие преобразований Галилея.

Теперь запишем второй закон Ньютона для нашей частицы в системе K:

$$m\vec{a} = \vec{F}. \tag{5.3}$$

При переходе в систему K' ускорение частицы $\vec{a}$, как мы выяснили, остаётся прежним. А что можно сказать об остальных двух величинах, входящих в (5.3), — массе и силе?

Масса есть мера инертности тела; масса показывает, в какой степени тело «сопротивляется» изменению скорости. Но приращение скорости Δu нашей частицы будет одним и тем же в любой инерциальной системе отсчёта. Следовательно, *масса частицы m во всех инерциальных системах отсчёта одинакова.*

Силы в механике зависят от расстояний между телами и, быть может, скоростей тел друг относительно друга. Но расстояние между двумя точками пространства одинаково во всех инерциальных системах отсчёта. Скорость одной частицы относительно другой также не зависит от того, в какой инерциальной системе отсчёта рассматривается движение. Стало быть, *сила $\vec{F}$ одинакова во всех инерциальных системах отсчёта.*

Величины и соотношения, не меняющиеся при определённых условиях, часто называются *инвариантными.* Так, ускорение, масса и сила инвариантны относительно выбора инерциальной системы отсчёта. Поэтому второй и третий законы Ньютона во всех системах отсчёта имеют одинаковый вид, т. е. инвариантны относительно преобразований Галилея.

Законы механики инвариантны относительно преобразований Галилея — такова альтернативная формулировка принципа относительности Галилея. Подчеркнём, что речь идёт об инвариантности математической формы законов механики. В результате этой инвариантности одно и то же механическое явление, наблюдаемое при одних и тех же начальных условиях, будет протекать одинаково во всех инерциальных системах отсчёта.

5.2 Принципы СТО

Принцип относительности Галилея, подробно рассмотренный в предыдущем разделе, говорит о том, что никакие лабораторные опыты механики не помогут определить, покоится ли лаборатория или движется равномерно и прямолинейно.

Но возникает закономерный вопрос: а кто заставляет нас ограничиваться лишь механическими явлениями? Давайте перейдём в другие области физики: будем наблюдать в движущейся лаборатории распространение тепла или света, ставить опыты с электромагнитными колебаниями, изучать ядерные процессы... Раз уж механика нам не помощник, то, быть может, где-нибудь в молекулярной физике, электродинамике, оптике, атомной или ядерной физике найдутся явления, на протекании которых скажется равномерно-прямолинейное движение лаборатории? Тогда, сопоставив ход таких явлений в неподвижной и в движущейся системах отсчёта, мы зафиксируем факт движения и сможем измерить его скорость.

С развитием электродинамики поначалу казалось, что так оно и есть. Дело в том, что, в отличие от законов механики Ньютона, уравнения Максвелла оказались не инвариантными относительно преобразований Галилея.

5.2.1 Гипотеза о мировом эфире

Из уравнений Максвелла следует, например, что свет в вакууме распространяется со скоростью $c = 300000$ км/с в любом направлении, причём эта скорость не зависит от того, покоится ли источник света или движется.

Физиков данный факт ничуть не удивлял: свет рассматривался как колебания особой всепроникающей среды — неподвижного мирового эфира. Считалось, что электромагнитные волны распространяются в эфире аналогично звуковым волнам в воздухе, а со звуком ведь дело обстоит точно так же: сигнал от бибикнувшего автомобиля бежит в воздухе во все стороны со скоростью примерно 340 м/с вне зависимости от скорости, с которой движется автомобиль.

А теперь представьте себе, что вы находитесь в звездолёте, который мчится в космическом вакууме со скоростью $v = 50000$ км/с относительно удалённых звёзд. Вы сидите лицом по ходу движения звездолёта и смотрите на лампочку, которая находится в его носовой части.

Свет от лампочки, не обращая внимания на её движение, перемещается относительно звёзд со скоростью c. Вы движетесь навстречу свету со скоростью v; стало быть, относительно вас свет имеет скорость $c + v = 350000$ км/с. Вы измеряете эту скорость, сопоставляете её с известным значением c и приходите к выводу, что двигаетесь со скоростью 50000 км/с! Таким образом, электромагнитные явления вроде бы позволяют отличить покой от равномерного прямолинейного движения.

У вас, кстати, может возникнуть вопрос: а чем плох аналогичный эксперимент со звуком? Давайте бибикнем в носовой части длинного движущегося лимузина, измерим скорость звука относительно нас и опровергнем принцип относительности Галилея! Ничего не выйдет: если лимузин замкнутый (как и должно быть), то он увлекает свой воздух вместе с собой, и вы ничего не заметите[2]. А вот в звездолёте вам никуда не деться от всепроникающего «эфирного ветра», который несётся вам в лицо и увеличивает тем самым скорость света в описанном выше эксперименте с лампочкой[3].

Соответственно, многие учёные (в том числе выдающийся голландский физик Х. Лоренц) считали, что инерциальные системы отсчёта, будучи равноправными с точки зрения механики, в

[2]Конечно, в открытом автомобиле звук будет нестись на вас вместе с окружающим воздухом, и скорость звука действительно окажется больше; но открытый автомобиль является нарушением правил — это всё равно, что выглянуть наружу.

[3]Пытаясь спасти принцип относительности Галилея применительно к электродинамике, Герц предположил, что эфир также увлекается движущимися телами — как воздух лимузином. Из этой гипотезы следовало, однако, что струя воды, увлекая эфир, должна увлекать и луч света — а в экспериментах такого не наблюдалось.

электродинамике перестают быть таковыми. Имеется выделенная, привилегированная система отсчёта, связанная с неподвижным мировым эфиром. Остальные системы отсчёта движутся относительно неё, и возникающий «эфирный ветер» меняет в них величину скорости света.

С целью обнаружения эфирного ветра в 1881 году был поставлен один из самых знаменитых физических экспериментов — опыт Майкельсона. С помощью чувствительного интерферометра производились попытки измерить скорость Земли относительно эфира. А именно, исследовалась интерференционная картина, даваемая двумя когерентными пучками света, имеющими перпендикулярные направления. Интерферометр движется относительно эфира вместе с Землёй; при вращении интерферометра меняется направление эфирного ветра относительно интерферометра, что должно сказываться на скоростях пучков и давать сдвиги интерференционной картины.

Однако никаких сдвигов обнаружено не было! Наблюдения проводились в разное время года (когда скорость Земли ощутимо меняла направление) и неизменно давали отрицательный результат. Интерферометр был настолько точный, что списать отсутствие эфирного ветра на недостаточную чувствительность прибора было нельзя.

Почему же движение Земли относительно эфира не удаётся зафиксировать? Не сомневаясь в существовании эфира, Лоренц заметил, что результаты опыта Майкельсона полностью объясняются, если сделать невероятное предположение: размеры движущегося предмета сокращаются в направлении движения! Так, если стержень длины l_0 начинает двигаться вдоль своей оси со скоростью v, то его длина становится равной

$$l = l_0\sqrt{1 - \frac{v^2}{c^2}}\,. \tag{5.4}$$

Эта гипотеза, названная *лоренцевым сокращением*, не вытекала на тот момент из каких-либо физических принципов и стояла особняком, будучи призвана лишь справиться с отрицательным результатом опыта. Но тем не менее формула (5.4) действительно оказалась верна! Её объяснение пришло позже, уже в рамках теории относительности Эйнштейна.

5.2.2 Постулаты Эйнштейна

Альберт Эйнштейн — величайшая фигура в истории физики. Для разрешения трудностей, описанных выше, он отказался от некоторых сложившихся в физике устоев и предпринял весьма радикальные шаги. Сформулируем ещё раз те проблемы, с которыми столкнулась физика, и их решения, предложенные Эйнштейном.

1. Не удаётся обнаружить привилегированную систему отсчёта, связанную с неподвижным мировым эфиром.

Так её и нет вовсе. Никакого эфира не существует. Все инерциальные системы отсчёта полностью равноправны между собой, и никакими физическими опытами нельзя выделить одну из них среди остальных.

Таким образом, Эйнштейн обобщил принцип относительности Галилея с механических на вообще все физические явления.

Принцип относительности Эйнштейна. Всякое физическое явление при одних и тех же начальных условиях протекает одинаково в любой инерциальной системе отсчёта.

Следовательно, если ваша лаборатория находится внутри корабля, то не только механический, но и *вообще никакой* эксперимент не даст вам ответа на вопрос, покоится ли корабль или движется равномерно и прямолинейно. Вы можете ставить опыты с газами, изучать тепловые явления, наблюдать за распространением электромагнитных волн, следить за атомными и ядерными процессами, анализировать взаимодействия элементарных частиц — и нигде вам

не удастся обнаружить каких-либо отклонений в протекании этих явлений, вызванных фактом равномерно-прямолинейного движения корабля.

В предыдущем разделе мы убедились в том, что законы механики имеют одинаковую математическую форму во всех инерциальных системах отсчёта: уравнения, выражающие эти законы, инвариантны относительно преобразований Галилея. Таков смысл принципа относительности Галилея. Обобщающий его принцип относительности Эйнштейна утверждает, что *любой физический закон имеет одинаковую математическую форму во всех инерциальных системах отсчёта.* Все уравнения, выражающие законы физики, должны быть инвариантны относительно перехода из одной инерциальной системы отсчёта в другую.

В частности, основные уравнения электродинамики — уравнения Максвелла — должны сохранять свою форму при таком переходе. Как же тогда быть со следующей трудностью?

2. Электродинамика противоречит механике в том, что уравнения Максвелла не инвариантны относительно преобразований Галилея.

Что ж, это проблема механики, а не электродинамики. Уравнения Максвелла блестяще работают в области электромагнитных явлений. Если преобразования Галилея не вяжутся с уравнениями Максвелла, то неверны преобразования Галилея, а не уравнения Максвелла.

Но легко сказать — преобразования Галилея неверны! Во-первых, они, казалось бы, совершенно очевидны — вам ведь не составило никакого труда в них разобраться. Чему там, собственно говоря, быть неверным?

А во-вторых — следствием преобразований Галилея, как мы видели, является закон сложения скоростей. Вы неоднократно пользовались им при решении задач. Что же получается — и закон сложения скоростей объявляется неверным?

Да, именно так! — гласил ответ Эйнштейна. Классическая механика Ньютона нуждается в глубоком, коренном пересмотре своих основных принципов. И слабый пункт классической механики состоит в том, что механические законы предполагают *мгновенность* распространения взаимодействий между телами.

Рассмотрим, например, гравитационное притяжение двух тел. Если одно из тел сместить в сторону, то, согласно закону всемирного тяготения, второе тело «почувствует» этот факт мгновенно, как только изменится расстояние от него до первого тела. Получается, что взаимодействие передаётся от одного тела к другому с бесконечной скоростью.

Эксперименты, однако, показывают, что механизм передачи взаимодействий состоит в следующем: изменение состояния тела меняет поле около него; возникшее возмущение поля начинает бежать во все стороны с некоторой *конечной* скоростью и лишь спустя определённый промежуток времени достигает другого тела. Мгновенно передающихся взаимодействий ни в каких опытах не наблюдается.

Но если взаимодействия не могут передаваться в бесконечной скоростью, то *в природе существует предельная, максимальная скорость распространения взаимодействий.* Изменённые законы механики должны учитывать наличие этой предельной скорости и, соответственно, конечность времени передачи взаимодействий между телами.

Второй постулат[4] Эйнштейна отводит исключительную роль скорости света.

Принцип инвариантности скорости света. В каждой инерциальной системе отсчёта свет движется в вакууме с одной и той же скоростью; величина этой скорости не зависит от того, покоится или движется источник света.

Таким образом, вышеописанный опыт с лампочкой в носовой части звездолёта нам провести не удастся: скорость света относительно наблюдателя в звездолёте будет равна c, а не $c + v$,

[4] *Постулат* в физике — это утверждение, которое служит обобщением опытных фактов и принимается без доказательства. Постулат формулирует свойства природы: «Наш мир таков, что...» Постулаты аналогичны аксиомам в геометрии. В основе любой физической теории лежат некоторые постулаты — они играют роль первичных утверждений, из которых остальные утверждения теории выводятся в качестве следствий.

и наблюдатель не сможет заметить факт движения звездолёта. Классический закон сложения скоростей применительно к скорости света не работает.

Мы увидим далее, что максимальная скорость распространения взаимодействий, присущая нашему миру, оказывается равной как раз скорости света в вакууме. Никакой сигнал, никакое тело, никакой вообще материальный объект в природе не может двигаться со скоростью, превышающей c. Величина c является фундаментальной константой, отражающей свойства мира, в котором мы живём.

Оба постулата Эйнштейна — принцип относительности и принцип инвариантности скорости света — легли в основу *специальной теории относительности (СТО)*. Эта теория затрагивает глубокие свойства пространства-времени, радикально меняя наши представления об окружающем мире[5]. Механика, построенная Эйнштейном на постулатах СТО, получила название *релятивистской* (от англ. relativity — относительность).

Новые и удивительные свойства пространства-времени и новые законы, устанавливаемые в СТО, проявляются при больших скоростях движения — и тем ярче, чем ближе мы подходим к скорости света. В повседневной жизни мы не замечаем этих релятивистских эффектов — по той простой причине, что привычные нам скорости чрезвычайно малы по сравнению со скоростью света. Во многих практических задачах можно считать скорость света бесконечной — и тогда прекрасно работает классическая механика.

Итак, классическая механика оказывается приближённой теорией и годится для небольших скоростей. Релятивистская механика используется тогда, когда скорости тел достаточно близки к скорости света — в таких ситуациях классическая механика отказывает совершенно. Классическая механика является предельным случаем релятивистской механики: формулы классической механики получаются из релятивистских формул предельным переходом $c \to \infty$.

Какие же новые свойства пространства-времени и новые физические законы открыла теория относительности? Мы будем рассказывать о них в двух следующих разделах. Здесь мы покажем лишь, что из постулатов СТО следуют весьма неожиданные и, казалось бы, парадоксальные выводы.

Рассмотрим системы отсчёта K и K' — те же, что и в предыдущем листке (рис. 5.2). В момент времени $t = 0$, когда их начала O и O' находятся в одной точке, в этой точке происходит световая вспышка.

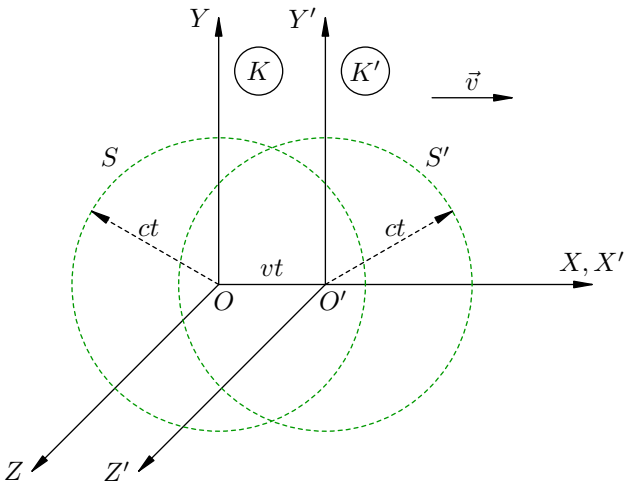


Рис. 5.2. Кажущийся парадокс со световой вспышкой

Где окажется волновой фронт вспышки к моменту времени t?

[5]Впоследствии Эйнштейн развил *общую теорию относительности (ОТО)*, которая описывает свойства пространства-времени при наличии гравитационного поля. Оказывается, гравитационное поле приводит к искривлению пространства-времени. ОТО настолько сложна, что рассмотрение даже самых её начал выходит далеко за рамки школьной программы. В СТО влиянием гравитации пренебрегают.

В системе K свет распространяется во все стороны со скоростью c. Поэтому в системе K вспышка достигнет сферы S радиуса ct с центром в точке O.

В системе K' скорость света также равна c. Значит, в системе K' вспышка достигнет сферы S' того же радиуса ct, но с центром в точке O'.

Однако точки O и O' к моменту t разойдутся на расстояние vt. Получается, что волновой фронт в один и тот же момент времени находится на двух разных сферах S и S'. Противоречие?

Противоречия на самом деле нет. Причина кажущегося парадокса кроется в понятии *одновременности*. На место нашего интуитивного понимания одновременности приходит чёткое определение этого термина, даваемое в СТО. Читаем следующий раздел!

5.3 Релятивистская кинематика

Напомним, что основной задачей механики является описание движения, а именно — выяснение того, как меняются координаты тела с течением времени.

Для описания движения нужно иметь систему отсчёта. В классической механике, как мы знаем, в систему отсчёта входят три объекта: тело отсчёта (относительно которого рассматривается движение), жёстко связанная с телом отсчёта система координат, а также часы для измерения времени. Наблюдатель, находящийся в данной системе отсчёта, имеет возможность измерять координаты тела и сопоставлять эти координаты с показаниями часов. В результате наблюдатель получает зависимость координат тела от времени; располагая такой зависимостью, он может найти скорость тела и другие кинематические величины.

5.3.1 Одновременность событий

Сопоставление координат тела и показаний часов — ключевой момент. Здесь мы подходим к важнейшему понятию одновременности событий. Прежде всего, процитируем Эйнштейна.

> Мы должны обратить внимание на то, что все наши суждения, в которых время играет какую-либо роль, всегда являются суждениями об *одновременных событиях*. Если я, например, говорю: «Этот поезд прибывает сюда в 7 часов», — то это означает примерно следующее: «Указания маленькой стрелки моих часов на 7 часов и прибытие поезда суть одновременные события».
>
> (А. Эйнштейн. «К электродинамике движущихся тел».)

Что такое одновременность в классической механике? Вопрос, казалось бы, ясен: события являются одновременными, если они происходят в один и тот же момент времени по часам наблюдателя. Отметим здесь два существенных момента.

- Неважно, происходят ли данные события в одной точке пространства или в различных точках. В классической механике мы спокойно говорим об одновременности пространственно разделённых событий.

- Понятие одновременности имеет абсолютный смысл: два события, одновременные в одной системе отсчёта, будут одновременными и в любой другой системе отсчёта. Во всех инерциальных системах отсчёта время течёт одинаково — это выражается равенством $t' = t$ в преобразованиях Галилея.

Такое понимание одновременности, однако, носит интуитивный характер. И, что совсем плохо, оно базируется на предположении о *мгновенности* передачи взаимодействий.

В самом деле, если сигналы от событий, происходящих в разных точках пространства, достигают наблюдателя мгновенно, то какая ему разница, насколько велико расстояние между этими событиями? Никакой задержки в приходе сигналов ведь не будет. Точно так же несущественно и то, покоится ли наблюдатель или движется — раз сигналы распространяются с бесконечной скоростью, события будут казаться наблюдателю одновременными независимо от факта его движения.

Но в действительности скорость сигнала является конечной и не может превышать скорость света в вакууме. Тем самым наше интуитивное понимание одновременности пространственно разделённых событий оказывается некорректным. Ведь если мы, держа в руках секундомер, фиксируем по нему время наступления окружающих событий и пытаемся судить об их одновременности, то нам придётся считаться с задержками прихода сигналов из различных точек пространства. Более того, эти задержки могут оказываться разными в зависимости от того, находимся ли мы в покоящейся системе отсчёта или в движущейся.

Что же получается — понятие одновременности вообще теряет смысл? Оказывается, нет! Эйнштейн предложил чёткую программу преодоления указанных трудностей. Суть её состоит в следующем: раз уж всё оказывается так плохо при измерении времени по одним-единственным часам наблюдателя, то давайте использовать *много синхронно идущих часов*, расставленных в разных точках пространства. Два события будут считаться одновременными, если совпадают показания часов, расположенных в тех точках, где произошли события.

А теперь — подробнее и по пунктам.

1. Пусть в некоторой точке пространства имеются часы. Если *в этой точке* происходит событие, то наши часы показывают время данного события. Таким образом, если в этой самой точке происходят два события, то мы всегда можем сказать, одновременны они или нет — просто сравнив показания наших часов в моменты наступления событий.

   Итак, с определением одновременности событий, происходящих в одной точке пространства, проблем нет.

2. Для определения понятия одновременности пространственно разделённых событий нам понадобится много одинаковых часов, расставленных в пространстве достаточно часто. Каждые часы показывают время событий, происходящих в той точке, где эти часы расположены.

   Чтобы была возможность судить об одновременности событий, происходящих в различных точках пространства, все эти часы должны идти *синхронно*, т. е. показывать одно и то же время. Но возникает естественный вопрос: а как этого добиться? Каким образом можно произвести синхронизацию часов?

3. Чтобы синхронизировать часы, расположенные в различных точках пространства, Эйнштейн предложил использовать световые сигналы.

   Пусть в точках A и B имеются часы. Предположим, что из точки A в точку B посылается световой сигнал, который отражается в точке B и возвращается назад в A.

   Пусть в момент отправления сигнала часы A показывали t_1, а в момент возвращения сигнала показания тех же часов A равны t_2.

   Правило Эйнштейна. По определению, часы A и B идут синхронно, если в момент прихода сигнала в точку B показания часов B равны $(t_1 + t_2)/2$.

   Иными словами, часы B должны показывать ровно середину промежутка между t_1 и t_2. Правило Эйнштейна иллюстрируется на рис. 5.3.

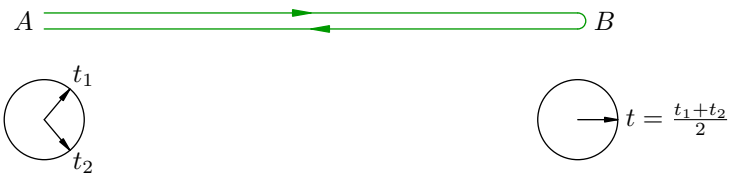


Рис. 5.3. Синхронизация часов по правилу Эйнштейна

   Можно дать и другую, равносильную формулировку правила Эйнштейна.

   Произведём в середине отрезка AB вспышку света. По определению, часы A и B идут синхронно, если в моменты прихода света в точки A и B показания часов совпадают.

4. Правило Эйнштейна основано на том, что скорость света в вакууме не зависит от направления распространения света. В самом деле, ведь при синхронизации часов мы считаем,

что световой сигнал идёт с одной и той же скоростью в обоих направлениях: как от A к B, так и обратно от B к A.

Одинаковость скорости света по всем направлениям — это факт, подтверждаемый многочисленными опытами.

5. Может возникнуть следующий вопрос: а зачем вообще использовать какие-то световые сигналы? Давайте сначала поместим двое часов в точку A, поставим их одинаково, а затем перенесём одни из этих часов в точку B. Вот и получится пара синхронизированных часов в двух различных точках A и B!

   Беда заключается в том, что такой способ не согласуется с правилом Эйнштейна. Если в точке B уже имеются часы, синхронизированные по правилу Эйнштейна с часами A, то перенесённые из A часы покажут в точке B время *меньшее*, чем первые. При этом перенесённые часы будут отставать тем больше, чем с большей скоростью они двигались! Об этом свидетельствует опыт, и мы скоро поймём, почему так получается.

   Так что альтернативы правилу Эйнштейна нет: оно является простым, естественным и приводит к стройной теории, прекрасно согласующейся с экспериментом.

6. По правилу Эйнштейна мы можем синхронизировать любую пару часов. Но является ли это правило непротиворечивым? А именно, если мы синхронизировали указанным способом сначала часы A и B, а затем часы B и C, то окажутся ли при этом синхронизированными часы A и C?

   Хотелось бы думать, что да, однако из правила Эйнштейна это логически не следует (как, впрочем, не следует и ответ «нет»). Эйнштейн *постулировал* непротиворечивость своего правила: *да, часы A и C окажутся при этом синхронизированными.* Данный постулат согласуется с экспериментом; принятие этого постулата не ведёт в дальнейшем к противоречиям в теории.

7. Итак, мы получили релятивистскую систему отсчёта с большим количеством часов. Все часы идут согласованно, они синхронизированы по правилу Эйнштейна. Время каждого события (*местное время*) измеряется по часам, расположенным в том месте, где событие совершилось.

   Теперь можно дать определение одновременности событий.

   Два пространственно разделённых события в данной системе отсчёта считаются одновременными, если при наступлении этих событий совпадают показания часов, расположенных в тех точках, где события произошли.

   Можно запомнить более короткую формулировку: *события одновременны, если их местные времена совпадают.*

Как видим, существование максимальной скорости распространения сигналов ведёт к коренному пересмотру наших обыденных представлений о пространстве и времени. Оказалось, например, что понятие одновременности событий нуждается в строгом определении. Данное определение является непротиворечивым, согласуется с опытом и приводит к следствиям, весьма неожиданным с повседневной точки зрения.

Так, понятие одновременности, а также величины промежутков времени и расстояний между точками теряют свой абсолютный характер и становятся относительными, то есть зависящими от выбора той или иной системы отсчёта.

5.3.2 Относительность одновременности

Давайте ещё раз осмыслим определение одновременности событий. Мы смогли его дать, введя предварительно единое время для нашей системы отсчёта. Это единое время задаётся множеством синхронно идущих часов, расставленных в различных точках пространства.

Тем самым понятие одновременности пространственно разделённых событий оказывается «привязанным» к данной системе отсчёта. Выясняется, что *два события, одновременные в одной системе отсчёта, могут оказаться не одновременными в другой системе отсчёта.* В этом нетрудно убедиться на следующем простом примере.

Рассмотрим вагон, который движется вправо со скоростью $\vec{v}$ (рис. 5.4). В точке S, находящейся в центре вагона, происходит световая вспышка. Одновременно ли свет достигнет точек A и B, расположенных соответственно на задней и передней стенке вагона?

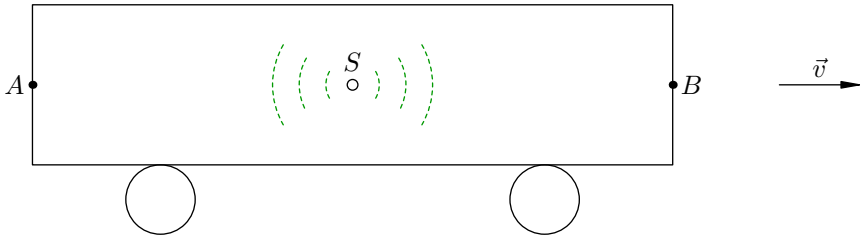


Рис. 5.4. Два события, одновременные в вагоне и не одновременные на земле

В системе отсчёта, связанной с вагоном, всё происходит точно так же, как в неподвижной лаборатории. По вагонным часам свет придёт в точки A и B одновременно.

Но в системе отсчёта, связанной с землёй, картина окажется иной. Точка A движется навстречу сигналу, а точка B удаляется от него; поэтому для достижения точки A свету потребуется пройти меньшее расстояние, чем для достижения точки B. Но в земной системе отсчёта скорость света будет одинакова в обоих направлениях — ведь согласно второму постулату СТО скорость света не зависит от факта движения источника. Стало быть, по земным часам свет придёт в точку A раньше, чем в точку B.

Таким образом, два события — приход сигнала от источника S в точки A и B — являются одновременными в системе отсчёта вагона и не одновременными в системе отсчёта земли.

В конце предыдущего листка было приведено противоречивое на первый взгляд рассуждение, в котором волновой фронт световой вспышки оказывался одновременно на двух различных сферах. Перечитайте это рассуждение ещё раз. Теперь становится ясно, что разрешение возникшего противоречия состоит в относительности понятия одновременности. Действительно, свет одновременно достигает поверхности сферы с центром в точке O только в системе K, но не в системе K'. И наоборот, свет одновременно достигает сферической поверхности с центром в точке O' только в системе K', но не в системе K.

5.3.3 Относительность промежутков времени

Снова рассмотрим вагон, который движется со скоростью $\vec{v}$. Предположим, что пассажир вагона подбрасывает яблоко; оно летит вертикально вверх, возвращается, и пассажир ловит его.

В системе отсчёта вагона события «яблоко брошено» и «яблоко поймано» происходят в одной точке. Промежуток времени между этими событиями, т. е. время полёта яблока в системе отсчёта вагона, измеряется по *одним и тем же* часам, расположенным в точке броска-ловли.

Но в системе отсчёта земли наши события происходят в различных пространственных точках. Момент броска яблока фиксируется по часам, расположенным в исходной точке, а момент ловли — по *другим* часам, расположенным в той точке, куда переместится вагон за время полёта яблока. Эти двое часов синхронизированы по правилу Эйнштейна. Время полёта яблока в

системе отсчёта земли — это разность показаний вторых часов в момент ловли и первых часов в момент броска.

И вот оказывается, что время полёта яблока, измеренное по вагонным часам, будет *меньше* времени полёта, измеренного по часам на земле!

Разобраться в этом не сложно. Давайте только заменим яблоко на световой сигнал, который бегает между горизонтальными зеркалами, расположенными внутри вагона (рис. 5.5). Тем самым мы максимально упростим себе задачу.

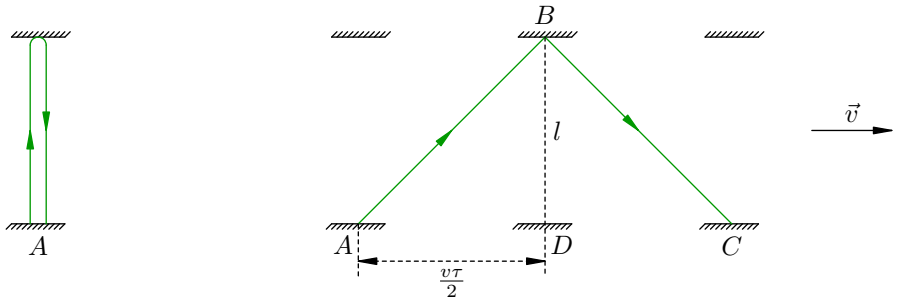


Рис. 5.5. Промежуток времени относителен

Сначала рассмотрим ход сигнала в системе отсчёта вагона. Сигнал выходит из точки A, идёт вертикально вверх, отражается от зеркала и возвращается назад в точку A (рис. 5.5, слева).

Время распространения сигнала от нижнего зеркала к верхнему и обратно, измеренное по часам A, обозначим τ_0. Если l — расстояние между зеркалами, то выполнено соотношение

$$2l = c\tau_0. \tag{5.5}$$

Теперь перейдём в систему отсчёта земли. Здесь сигнал будет двигаться между зеркалами по ломаной ABC (рис. 5.5, справа).

Время распространения сигнала от нижнего зеркала к верхнему и обратно есть разность показаний синхронизированных часов: в точке C в момент прихода сигнала и в точке A в момент его отправления. Обозначим это время через τ.

За время τ сигнал проходит путь $AB+BC$, равный $c\tau$, а вагон — путь $AC = v\tau$. По теореме Пифагора имеем: $AB^2 = AD^2 + BD^2$ или

$$\left(\frac{c\tau}{2}\right)^2 = \left(\frac{v\tau}{2}\right)^2 + l^2.$$

Умножаем это равенство на 4:

$$(c\tau)^2 = (v\tau)^2 + (2l)^2.$$

С учётом (5.5) получим:

$$(c\tau)^2 = (v\tau)^2 + (c\tau_0)^2.$$

Отсюда можно выразить τ через τ_0:

$$\tau = \frac{\tau_0}{\sqrt{1 - \dfrac{v^2}{c^2}}}\,. \tag{5.6}$$

Полученная формула (5.6) носит совершенно общий характер. Пусть имеются две системы отсчёта K и K', причём система K' движется относительно K со скоростью v. Рассмотрим два события, которые в системе K' происходят в одной точке пространства. Время τ_0 между этими событиями в системе K' называется *собственным временем*. По часам системы K между этими событиями проходит время τ, которое связано с собственным временем соотношением (5.6).

Обратите внимание, что в формуле (5.6) собственное время τ_0 делится на величину, меньшую единицы; поэтому всегда выполнено неравенство $\tau > \tau_0$. При этом время τ оказывается тем больше, чем с большей скоростью система K' движется относительно K.

Данный эффект — так называемое *релятивистское замедление времени* — оказывается весьма существенным при скоростях, близких к скорости света. Давайте вернёмся к нашему примеру с пассажиром, подбрасывающим яблоко в вагоне. Предположим, что вагон движется со скоростью $v = 0{,}999c$, а время полёта яблока по часам пассажира равно $\tau_0 = 1$ с. Тогда на земле между подбрасыванием яблока и его ловлей пройдёт время

$$\tau = \frac{1}{\sqrt{1 - 0{,}999^2}} \approx 22{,}4 \text{ с}.$$

Ну а теперь представьте, что с Земли отправляется звездолёт со скоростью $0{,}999c$. Космическое путешествие длится 10 лет по часам астронавтов. Вернувшись на Землю, астронавты обнаруживают, что попали в совершенно иную эпоху — ведь по земным часам прошло 224 года!

Звучит фантастично, но никакой фантастики тут нет — мы имеем дело со строгими фактами теории относительности. Ни с чем подобным нам сталкиваться не доводилось по той причине, что мы не умеем пока перемещаться с околосветовыми скоростями. Но как только эти скорости будут достигнуты, станет реальностью и машина времени для путешествий в будущее :-)

5.3.4 Относительность расстояний

При выводе формулы (5.6) мы неявно предполагали, что расстояние l между зеркалами одинаково как в вагоне, так и на земле. Но так ли это на самом деле? Да, это действительно так: вертикальные размеры предметов являются одними и те же как в вагонной, так и в земной системе отсчёта.

Чтобы убедиться в этом, давайте возьмём два одинаковых вертикальных стержня; один из них поместим в вагон, а другой оставим на земле. Оба стержня пусть будут на одной и той же высоте над землёй. Когда стержни поравняются друг с другом, концы одного стержня сделают засечки на другом стержне. Так вот, из принципа относительности следует, что эти засечки должны прийтись в точности на концы другого стержня.

В самом деле, пусть по засечкам оказывается, например, что вагонный стержень короче земного, т. е. движущийся стержень короче покоящегося. Но по принципу относительности инерциальные системы отсчёта полностью равноправны. Давайте перейдём в систему отсчёта вагона: там вагонный стержень будет покоиться, а земной — двигаться. Тогда получится, что движущийся стержень длиннее покоящегося. Противоречие!

Итак, *поперечные* размеры предметов одинаковы как в покоящейся, так и в движущейся системе отсчёта. Иначе обстоит дело с *продольными* размерами.

Вновь вернёмся к нашему вагону и рассмотрим стержень AB, расположенный вдоль вектора скорости вагона (рис. 5.6; изображать вагон надобности уже нет). Стержень, таким образом, двигается со скоростью v параллельно оси X.

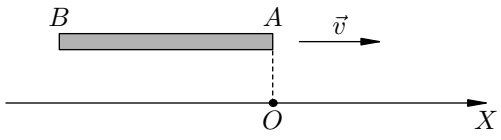


Рис. 5.6. Длина стержня относительна

Пусть l_0 — длина неподвижного стержня, измеренная в вагоне. Она называется *собственной длиной* стержня. Через l обозначим длину движущегося стержня, измеренную на земле.

Для нахождения соотношения между l и l_0 рассмотрим два события: 1) прохождение точки A мимо фиксированной точки O на оси X; 2) прохождение точки B мимо точки O.

В земной системе отсчёта наши события происходят в одной точке O. Промежуток времени между этими событиями по земным часам пусть равен τ_0 (это собственное время, разделяющее данные события). Очевидно, что

$$v = \frac{l}{\tau_0}\,. \tag{5.7}$$

В системе отсчёта вагона указанные события происходят в двух различных точках A и B. Промежуток времени между этими событиями по вагонным часам равен τ. Аналогично имеем:

$$v = \frac{l_0}{\tau}\,. \tag{5.8}$$

Приравнивая правые части формул (5.7) и (5.8), получим:

$$l = l_0\frac{\tau_0}{\tau}\,.$$

Но в силу формулы (5.6) имеем:

$$\frac{\tau_0}{\tau} = \sqrt{1 - \frac{v^2}{c^2}}\,.$$

Отсюда получаем окончательную формулу:

$$l = l_0\sqrt{1 - \frac{v^2}{c^2}}\,. \tag{5.9}$$

Как видим, собственная длина l_0 умножается на величину, меньшую единицы; стало быть, длина движущегося стержня будет меньше длины покоящегося стержня. Это так называемое *лоренцево сокращение* — все тела сокращают размеры в направлении своего движения.

Подчеркнём ещё раз: *длина стержня в системе отсчёта, относительно которой стержень движется, меньше длины этого же стержня в системе отсчёта, относительно которой он покоится.* Данный эффект связан лишь с особенностями измерительных процедур, свойственных теории относительности. Никаких реальных «сжатий» в движущемся стержне, разумеется, не происходит.

5.3.5 Преобразования Лоренца

Теперь мы можем вывести формулы, связывающие координаты и время фиксированного события в двух различных инерциальных системах отсчёта.

Пусть снова имеются две системы отсчёта: система K и движущаяся относительно неё система K' (рис. 5.7). При $t = 0$ начала O и O' этих систем совпадают.

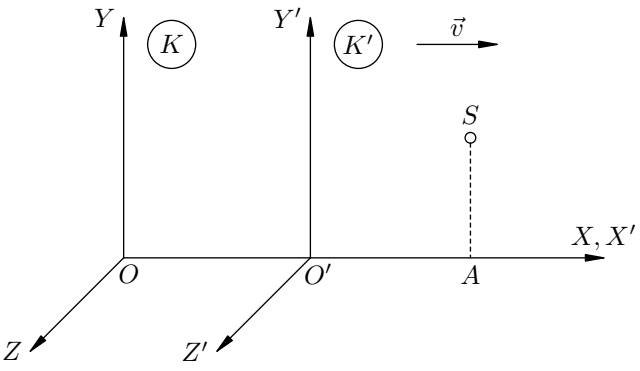


Рис. 5.7. К выводу преобразований Лоренца

Рассмотрим некоторое событие S (например, вспышку света). В системе K это событие происходит в точке с координатами (x, y, z) в момент времени t. В системе K' это же событие происходит в точке с координатами (x', y', z') в момент времени t'.

Как мы уже выяснили, поперечные размеры тел в обеих системах отсчёта одни и те же. Поэтому имеем: $y' = y$, $z' = z$.

Пусть A — проекция точки S на общую ось абсцисс. Найдём длину отрезка OA в системах K и K'.

В системе K отрезок OA покоится. Его длина равна x — это собственная длина данного отрезка. В системе K' отрезок OA движется со скоростью v, и его длина в силу формулы (5.9) равна $x\sqrt{1 - v^2/c^2}$. Но с другой стороны, в системе K' длина OA равна $x' + vt'$. Следовательно,

$$x' + vt' = x\sqrt{1 - \frac{v^2}{c^2}}\,. \tag{5.10}$$

Теперь аналогично найдём длину отрезка $O'A$ в системах K и K'.

В системе K' отрезок $O'A$ покоится, его собственная длина равна x'. В системе K отрезок $O'A$ движется со скоростью v, и его длина равна $x'\sqrt{1 - v^2/c^2}$. С другой стороны, длина $O'A$ в системе K равна $x - vt$. Поэтому

$$x - vt = x'\sqrt{1 - \frac{v^2}{c^2}}\,. \tag{5.11}$$

Из формулы (5.11) выразим x'. Полученное выражение подставим в (5.10) и выразим оттуда t'. В результате получим:

$$x' = \frac{x - vt}{\sqrt{1 - \dfrac{v^2}{c^2}}}\,, \quad y' = y, \quad z' = z, \quad t' = \frac{t - \dfrac{vx}{c^2}}{\sqrt{1 - \dfrac{v^2}{c^2}}}\,. \tag{5.12}$$

Формулы (5.12) называются *преобразованиями Лоренца*. Они дают искомую связь координат и времени события S в инерциальных системах отсчёта K и K'. Эти релятивистские формулы, вытекающие из принципов СТО, служат заменой классическим преобразованиям Галилея, опирающимся на представления о мгновенности распространения взаимодействий.

При малых скоростях движения, т. е. при $v \ll c$, мы можем считать отношение v/c равным нулю. Тогда преобразования Лоренца переходят в соотношения:

$$x' = x - vt, \quad y' = y, \quad z' = z, \quad t' = t. \tag{5.13}$$

Эти формулы есть не что иное, как преобразования Галилея. Мы видим, что преобразования Галилея служат предельным случаем преобразований Лоренца, когда скорости тел малы по сравнению со скоростью света. Поэтому при малых скоростях движения релятивистская механика Эйнштейна переходит в классическую механику Ньютона.

С системой равеств (5.10) и (5.11) можно поступить иначе. Выразим x из (5.10), подставим в (5.11) и выразим оттуда t. В результате придём к другому варианту записи преобразований Лоренца:

$$x = \frac{x' + vt'}{\sqrt{1 - \dfrac{v^2}{c^2}}}\,, \quad y = y', \quad z = z', \quad t = \frac{t' + \dfrac{vx'}{c^2}}{\sqrt{1 - \dfrac{v^2}{c^2}}}\,. \tag{5.14}$$

Формулы (5.12) задают переход из системы K в систему K'. Формулы (5.14) задают обратный переход из системы K' в систему K.

В предельном случае $v \ll c$ преобразования Лоренца (5.14) также переходят в преобразования Галилея:

$$x = x' + vt', \quad y = y', \quad z = z', \quad t = t'. \tag{5.15}$$

Эти формулы, как легко видеть, полностью совпадают с формулами (5.13).

5.3.6 Релятивистский закон сложения скоростей

Опять рассмотрим наши системы отсчёта K и K'. Пусть точка M движется вдоль общего направления осей X и X' (рис. 5.8).

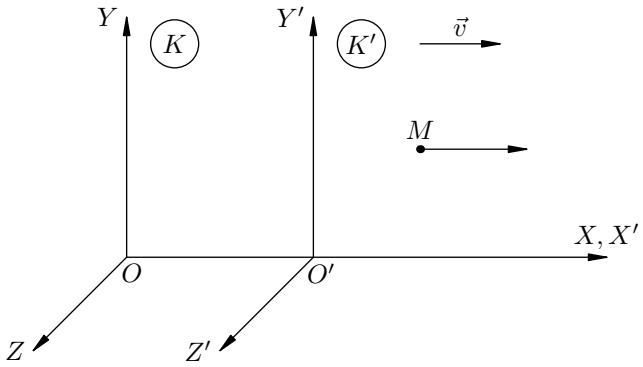


Рис. 5.8. К выводу закона сложения скоростей

Пусть u — скорость точки M в системе K; в системе K' скорость этой точки пусть будет u'. Как связаны друг с другом u и u'?

Давайте вспомним, как выводится соответствующая формула в классической механике. Берём первое из равенств (5.15) с заменой t' на t:

$$x = x' + vt.$$

Переходим к бесконечно малым приращениям координат и времени:

$$dx = dx' + vdt.$$

Делим обе части на dt:

$$\frac{dx}{dt} = \frac{dx'}{dt} + v.$$

Остаётся заметить, что $\frac{dx}{dt} = u$, $\frac{dx'}{dt} = u'$:

$$u = u' + v. \tag{5.16}$$

Вот мы и получили классический закон сложения скоростей, которым неоднократно пользовались при решении задач механики.

Однако данный закон не может быть верным в теории относительности. В самом деле, рассмотрим вместо точки M световой сигнал в вакууме, мчащийся в системе K' со скоростью $u' = c$. Согласно закону (5.16) получится, что скорость нашего сигнала в системе K будет равна $u = c + v$. Но это противоречит принципу относительности, в силу которого скорость света в вакууме имеет одно и то же значение во всех инерциальных системах отсчёта.

Возникновение данного противоречия не удивительно: ведь вывод формулы (5.16) базируется на преобразованиях Галилея, которые в теории относительности уступают место преобразованиям Лоренца. Поэтому правильный закон сложения скоростей нужно выводить теперь из преобразований Лоренца.

Идея вывода — та же самая, что и для формулы (5.16). Мы исходим из того, что

$$u = \frac{dx}{dt}\,, \quad u' = \frac{dx'}{dt'}\,. \tag{5.17}$$

В соотношениях (5.14) переходим к бесконечно малым приращениям координат и времени:

$$dx = \frac{dx' + vdt'}{\sqrt{1 - \dfrac{v^2}{c^2}}}\,, \quad dt = \frac{dt' + \dfrac{vdx'}{c^2}}{\sqrt{1 - \dfrac{v^2}{c^2}}}\,.$$

Делим первое из данных равенств на второе:

$$\frac{dx}{dt} = \frac{dx' + vdt'}{dt' + \dfrac{vdx'}{c^2}}\,.$$

Разделим числитель и знаменатель правой части на dt':

$$\frac{dx}{dt} = \frac{\dfrac{dx'}{dt'} + v}{1 + \dfrac{v}{c^2}\dfrac{dx'}{dt'}}\,.$$

Остаётся учесть соотношения (5.17) и написать:

$$u = \frac{u' + v}{1 + \dfrac{vu'}{c^2}}\,. \tag{5.18}$$

Это и есть *релятивистский закон сложения скоростей*, который приходит на смену классическому.

Теперь уже никакого противоречия не возникает: если скорость сигнала $u' = c$ в системе K', то в системе K его скорость равна:

$$u = \frac{c + v}{1 + \dfrac{vc}{c^2}} = \frac{c + v}{1 + \dfrac{v}{c}} = c,$$

как того и требует принцип относительности.

При $v \ll c$ формулы (5.18), как нетрудно видеть, переходят в формулы (5.16). Иными словами, *при малых скоростях движения релятивистский закон сложения скоростей переходит в классический закон.*

5.4 Релятивистская динамика

В классической динамике мы начали с законов Ньютона, потом перешли к импульсу, а после него — к энергии. Здесь мы ради простоты изложения поступим ровно наоборот: начнём с энергии, затем перейдём к импульсу и закончим релятивистским уравнением движения — модификацией второго закона Ньютона для теории относительности.

5.4.1 Релятивистская энергия

Предположим, что изолированное тело массы m покоится в данной системе отсчёта. Одно из самых впечатляющих достижений теории относительности — это знаменитая *формула Эйнштейна*:

$$E = mc^2. \tag{5.19}$$

Здесь E — энергия тела, c — скорость света в вакууме. Поскольку тело покоится, энергия E, вычиляемая по формуле (5.19), называется *энергией покоя*.

Формула (5.19) утверждает, что каждое тело само по себе обладает энергией — просто потому, что оно существует в природе. Образно говоря, природа затратила определённые усилия на то, чтобы «собрать» данное тело из мельчайших частиц вещества, и мерой этих усилий служит энергия покоя тела. Энергия эта весьма велика; так, в одном килограмме вещества заключена энергия

$$E = 1 \cdot (3 \cdot 10^8)^2 = 9 \cdot 10^{16} \text{ Дж}.$$

Интересно, какое количество топлива нужно сжечь, чтобы выделилось столько энергии? Возьмём, например, дерево. Его удельная теплота сгорания равна $q = 10^7$ Дж/кг, поэтому находим: $m = E/q = 9 \cdot 10^9$ кг. Это девять миллионов тонн!

Ещё для сравнения: такую энергию единая энергосистема России вырабатывает примерно за десять дней.

Почему столь грандиозная энергия, содержащаяся в теле, до сих пор оставалась нами незамеченной? Почему в нерелятивистских задачах, связанных с сохранением и превращением энергии, мы не учитывали энергию покоя? Скоро мы ответим на этот вопрос.

Поскольку энергия покоя тела прямо пропорциональна его массе, изменение энергии покоя на величину ΔE приводит к изменению массы тела на

$$\Delta m = \frac{\Delta E}{c^2}.$$

Так, при нагревании тела возрастает его внутренняя энергия, и, стало быть, масса тела увеличивается! В повседневной жизни мы не замечаем этого эффекта ввиду его чрезвычайной малости. Например, для нагревания воды массой $m = 1$ кг на $\Delta t = 100°\text{C}$ (удельная теплоёмкость воды равна $c_\text{в} = 4200$ Дж/(кг $\cdot$ °C)) ей нужно передать количество теплоты:

$$Q = c_\text{в} m \Delta t = 4{,}2 \cdot 10^5 \text{ Дж}.$$

Увеличение массы воды будет равно:

$$\Delta m = \frac{Q}{c^2} = \frac{4{,}2 \cdot 10^5}{9 \cdot 10^{16}} \approx 4{,}7 \cdot 10^{-12} \text{ кг}.$$

Столь ничтожное изменение массы невозможно заметить на фоне погрешностей измерительных приборов.

Формула (5.19) даёт энергию покоящегося тела. Что изменится, если тело движется?

Снова рассмотрим неподвижную систему отсчёта K и систему K', движущуюся относительно K со скоростью v. Пусть тело массы m покоится в системе K'; тогда энергия тела в системе

K' есть энергия покоя, вычисляемая по формуле (5.19). Оказывается, при переходе в систему K энергия преобразуется так же, как и время — а именно, энергия тела в системе K, в которой тело движется со скоростью v, равна:

$$E = \frac{mc^2}{\sqrt{1 - \dfrac{v^2}{c^2}}}\,. \tag{5.20}$$

Формула (5.20) была также установлена Эйнштейном. Величина E — это *полная энергия* движущегося тела. Поскольку в данной формуле mc^2 делится на «релятивистский корень», меньший единицы, полная энергия движущегося тела превышает энергию покоя. Полная энергия будет равна энергии покоя только при $v = 0$.

Выражение для полной энергии (5.20) позволяет сделать важные выводы о возможных скоростях движения объектов в природе.

1. Каждое массивное тело обладает определённой энергией, поэтому необходимо выполнение неравенства
$$1 - \frac{v^2}{c^2} > 0.$$
Оно означает, что $v < c$: *скорость массивного тела всегда меньше скорости света.*

2. В природе существуют безмассовые частицы (например, фотоны), несущие энергию. При подстановке $m = 0$ в формулу (5.20) её числитель обращается в нуль. Но энергия-то фотона ненулевая!

   Единственный способ избежать здесь противоречия — это принять, что *безмассовая частица обязана двигаться со скоростью света.* Тогда и знаменатель нашей формулы обратится в нуль, так что формула (5.20) попросту откажет. Нахождение формул для энергии безмассовых частиц не входит в компетенцию теории относительности. Так, выражение для энергии фотона устанавливается в квантовой физике.

Интуитивно чувствуется, что полная энергия (5.20) состоит из энергии покоя и собственно «энергии движения», т. е. кинетической энергии тела. При малых скоростях движения это показывается явным образом. Используем приближённые формулы, справедливые при $\alpha \ll 1$:

$$\sqrt{1-\alpha} \approx 1 - \frac{\alpha}{2}\,, \tag{5.21}$$

$$\frac{1}{1-\alpha} \approx 1 + \alpha. \tag{5.22}$$

С помощью этих формул последовательно получаем из (5.20):

$$E = \frac{mc^2}{\sqrt{1 - \dfrac{v^2}{c^2}}} \approx \frac{mc^2}{1 - \dfrac{1}{2}\dfrac{v^2}{c^2}} \approx mc^2\left(1 + \frac{1}{2}\frac{v^2}{c^2}\right) = mc^2 + \frac{mv^2}{2}\,. \tag{5.23}$$

Таким образом, при малых скоростях движения полная энергия сводится просто к сумме энергия покоя и кинетической энергии. Это служит мотивировкой для определения понятия кинетической энергии в теории относительности:

$$E_{\text{кин}} = \frac{mc^2}{\sqrt{1 - \dfrac{v^2}{c^2}}} - mc^2. \tag{5.24}$$

При $v \ll c$ формула (5.24) переходит в нерелятивистское выражение $E_{\text{кин}} = mv^2/2$.

Теперь мы можем ответить на заданный выше вопрос о том, почему до сих пор не учитывалась энергия покоя в нерелятивистских энергетических соотношениях.

Как видно из (5.23), при малых скоростях движения энергия покоя входит в полную энергию в качестве слагаемого. В задачах, например, механики и термодинамики изменения энергии тел составляют максимум несколько миллионов джоулей; эти изменения столь незначительны по сравнению с энергиями покоя рассматриваемых тел, что приводят к микроскопическим изменениям их масс. Поэтому с высокой точностью можно считать, что суммарная масса тел не меняется в ходе механических или тепловых процессов. В результате суммы энергий покоя тел в начале и в конце процесса попросту сокращаются в обеих частях закона сохранения энергии!

Но такое бывает не всегда. В других физических ситуациях изменения энергии тел могут приводить к более заметным изменениям суммарной массы. Мы увидим, например, что в ядерных реакциях отличия масс исходных и конечных продуктов обычно составляют доли процента. Скажем, при распаде ядра урана $^{235}_{92}\mathrm{U}$ суммарная масса продуктов распада примерно на 0,1% меньше массы исходного ядра. Эта одна тысячная доля массы ядра высвобождается в виде энергии, которая при взрыве атомной бомбы способна уничтожить город.

При неупругом столкновении часть кинетической энергии тел переходит в их внутренюю энергию. Релятивистский закон сохранения полной энергии учитывает этот факт: суммарная масса тел после столкновения увеличивается!

Рассмотрим в качестве примера два тела массы m, летящих навстречу друг другу с одинаковой скоростью $3c/5$. В результате неупругого столкновения образуется тело массы M, скорость которого равна нулю по закону сохранения импульса (об этом законе речь впереди). Согласно закону сохранения энергии получаем:

$$\frac{mc^2}{\sqrt{1-\dfrac{(3c/5)^2}{c^2}}}+\frac{mc^2}{\sqrt{1-\dfrac{(3c/5)^2}{c^2}}}=Mc^2,$$

$$2\cdot\frac{mc^2}{\sqrt{1-\left(\dfrac{3}{5}\right)^2}}=Mc^2,$$

$$\frac{2m}{4/5}=M,$$

$$M=\frac{5}{2}m.$$

Мы видим, что, $M > 2m$ — масса образовавшегося тела превышает сумму масс тел до столкновения. Избыток массы, равный $m/2$, возник за счёт перехода кинетической энергии сталкивающихся тел во внутреннюю энергию.

5.4.2 Релятивистский импульс

Классическое выражение для импульса $\vec{p}=m\vec{v}$ не годится в теории относительности — оно, в частности, не согласуется с релятивистским законом сложения скоростей. Давайте убедимся в этом на следующем простом примере.

Пусть система K' движется относительно системы K со скоростью $v = c/2$ (рис. 5.9). Два тела массы m в системе K' летят навстречу друг другу с одинаковой скоростью $u' = c/2$. Происходит неупругое столкновение.

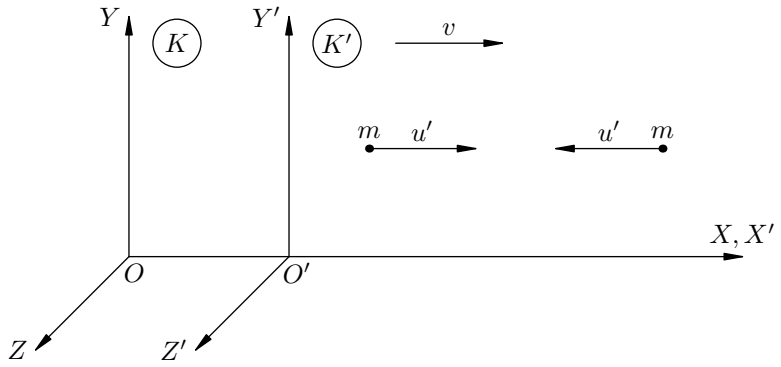


Рис. 5.9. К закону сохранения импульса

В системе K' тела после столкновения останавливаются. Давайте, как и выше, найдём массу M образовавшегося тела:

$$Mc^2 = 2\frac{mc^2}{\sqrt{1-\dfrac{(c/2)^2}{c^2}}} = \frac{2mc^2}{\sqrt{1-\left(\dfrac{1}{4}\right)^2}} = \frac{4mc^2}{\sqrt{3}},$$

откуда

$$M = \frac{4m}{\sqrt{3}}\,.$$

Теперь посмотрим на процесс столкновения с точки зрения системы K. До столкновения левое тело имеет скорость:

$$u_1 = \frac{v+u'}{1+\dfrac{vu'}{c^2}} = \frac{c/2+c/2}{1+\dfrac{(c/2)(c/2)}{c^2}} = \frac{4c}{5}\,.$$

Правое тело имеет скорость:

$$u_2 = \frac{v-u'}{1-\dfrac{vu'}{c^2}} = \frac{c/2-c/2}{1-\dfrac{(c/2)(c/2)}{c^2}} = 0.$$

Нерелятивистский импульс нашей системы до столкновения равен:

$$mu_1 - mu_2 = \frac{4mc}{5}\,.$$

После столкновения получившееся тело массы M двигается со скоростью $v = c/2$. Его нерелятивистский импульс равен:

$$Mv = \frac{4m}{\sqrt{3}}\frac{c}{2} = \frac{2m}{\sqrt{3}}\,.$$

Как видим, $mu_1 - mu_2 \neq Mv$, то есть нерелятивистский импульс не сохраняется.

Оказывается, правильное выражение для импульса в теории относительности получается делением классического выражения на «релятивистский корень»: импульс тела массы m, двигающегося со скоростью $\vec{v}$, равен:

$$\vec{p} = \frac{m\vec{v}}{\sqrt{1-\dfrac{v^2}{c^2}}}\,. \tag{5.25}$$

Давайте вернёмся к только что рассмотренному примеру и убедимся, что теперь с законом сохранения импульса всё будет в порядке.

Импульс системы до столкновения:

$$p_{\text{до}} = \frac{mu_1}{\sqrt{1-\dfrac{u_1^2}{c^2}}} - \frac{mu_2}{\sqrt{1-\dfrac{u_2^2}{c^2}}} = \frac{m(4c/5)}{\sqrt{1-\dfrac{(4c/5)^2}{c^2}}} - 0 = \frac{4mc/5}{3/5} = \frac{4mc}{3}\,.$$

Импульс после столкновения:

$$p_{\text{после}} = \frac{Mv}{\sqrt{1-\dfrac{v^2}{c^2}}} = \frac{Mc/2}{\sqrt{1-\dfrac{(c/2)^2}{c^2}}} = \frac{(4m/\sqrt{3})(c/2)}{\sqrt{3}/2} = \frac{4mc}{3}\,.$$

Вот теперь всё правильно: $p_{\text{до}} = p_{\text{после}}$!

5.4.3 Связь энергии и импульса

Из формул (5.20) и (5.25) можно получить замечательное соотношение между энергией и импульсом в теории относительности. Возводим обе части этих формул в квадрат:

$$E^2 = \frac{m^2c^4}{1-\dfrac{v^2}{c^2}}\,, \quad p^2 = \frac{m^2v^2}{1-\dfrac{v^2}{c^2}}\,.$$

Преобразуем разность:

$$E^2 - p^2c^2 = \frac{m^2c^4}{1-\dfrac{v^2}{c^2}} - \frac{m^2v^2c^2}{1-\dfrac{v^2}{c^2}} = \frac{m^2c^2(c^2-v^2)}{c^2-v^2} = m^2c^4.$$

Это и есть искомое соотношение:

$$E^2 - p^2c^2 = m^2c^4. \tag{5.26}$$

Данная формула позволяет выявить простую связь между энергией и импульсом фотона. Фотон имеет нулевую массу и движется со скоростью света. Как уже было замечено выше, сами по себе энергия и импульс фотона в СТО найдены быть не могут: при подстановке в формулы (5.20) и (5.25) значений $m = 0$ и $v = c$ мы получим нули в числителе и знаменателе. Но зато с помощью (5.26) легко находим: $E^2 - p^2c^2 = 0$, или

$$E = pc. \tag{5.27}$$

В квантовой физике устанавливается выражение для энергии фотона, после чего с помощью формулы (5.27) находится его импульс.

5.4.4 Релятивистское уравнение движения

Рассмотрим тело массы m, движущееся вдоль оси X под действием силы F. Уравнение движения тела в классической механике — это второй закон Ньютона: $ma = F$. Если за бесконечно малое время dt приращение скорости тела равно dv, то $a = dv/dt$, и уравнение движения запишется в виде:

$$m\frac{dv}{dt} = F. \tag{5.28}$$

Теперь заметим, что $mdv = d(mv) = dp$ — изменение нерелятивистского импульса тела. В результате получим «импульсную» форму записи второго закона Ньютона — производная импульса тела по времени равна силе, приложенной к телу:

$$\frac{dp}{dt} = F. \tag{5.29}$$

Все эти вещи вам знакомы, но повторить никогда не помешает ;-)

Классическое уравнение движения — второй закон Ньютона — является инвариантным относительно преобразований Галилея, которые в классической механике описывают переход из одной инерциальной системы отсчёта в другую (это означает, напомним, что при указанном переходе второй закон Ньютона сохраняет свой вид). Однако в СТО переход между инерциальными системами отсчёта описывается преобразованиями Лоренца, а относительно них второй закон Ньютона уже не является инвариантным. Следовательно, классическое уравнение движения должно быть заменено релятивистским, которое сохраняет свой вид под действием преобразований Лоренца.

То, что второй закон Ньютона (5.28) не может быть верным в СТО, хорошо видно на следующем простом примере. Допустим, что к телу приложена постоянная сила. Тогда согласно классической механике тело будет двигаться с постоянным ускорением; скорость тела будет линейно возрастать и с течением времени превысит скорость света. Но мы знаем, что на самом деле это невозможно.

Правильное уравнение движения в теории относительности оказывается совсем не сложным. *Релятивистское уравнение движения имеет вид* (5.29)*, где* p *— релятивистский импульс:*

$$\frac{d\left(\dfrac{mv}{\sqrt{1 - v^2/c^2}}\right)}{dt} = F. \tag{5.30}$$

Производная релятивистского импульса по времени равна силе, приложенной к телу.

В теории относительности уравнение (5.30) приходит на смену второму закону Ньютона.

Давайте выясним, как же в действительности будет двигаться тело массы m под действием постоянной силы F. При условии $F = \text{const}$ из формулы (5.30) получаем:

$$\frac{mv}{\sqrt{1 - \dfrac{v^2}{c^2}}} = Ft.$$

Остаётся выразить отсюда скорость:

$$v = \frac{cFt}{\sqrt{F^2t^2 + m^2c^2}}\,. \tag{5.31}$$

Посмотрим, что даёт эта формула при малых и при больших временах движения. Пользуемся приближёнными соотношениями при $\alpha \ll 1$:

$$\sqrt{1 + \alpha} \approx 1 + \frac{\alpha}{2}\,, \tag{5.32}$$

$$\frac{1}{1 + \alpha} \approx 1 - \alpha. \tag{5.33}$$

Формулы (5.32) и (5.33) отличаются от формул (5.21) и (5.22) только лишь знаком в левых частях. Постарайтесь запомнить все эти четыре приближённых равенства — они часто используются в физике.

Итак, начинаем с малых времён движения. Преобразуем выражение (5.31) следующим образом:

$$v = \frac{cFt}{mc\sqrt{1+\dfrac{F^2t^2}{m^2c^2}}}\,.$$

При малых t имеем:

$$\frac{F^2t^2}{m^2c^2} \ll 1.$$

Последовательно пользуясь нашими приближёнными формулами, получим:

$$v \approx \frac{cFt}{mc\left(1+\dfrac{1}{2}\dfrac{F^2t^2}{m^2c^2}\right)} \approx \frac{Ft}{m}\left(1-\frac{F^2t^2}{2m^2c^2}\right).$$

Выражение в скобках почти не отличается от единицы, поэтому при малых t имеем:

$$v \approx \frac{Ft}{m} = at\,.$$

Здесь $a = F/m$ — ускорение тела. Мы получили результат, хорошо известный нам из классической механики: скорость тела линейно растёт со временем. Это и не удивительно — при малых временах движения скорость тела также невелика, поэтому мы можем пренебречь релятивистскими эффектами и пользоваться обычной механикой Ньютона.

Теперь переходим к большим временам. Преобразуем формулу (5.31) по-другому:

$$v \approx \frac{cFt}{Ft\sqrt{1+\dfrac{m^2c^2}{F^2t^2}}} = \frac{c}{\sqrt{1+\dfrac{m^2c^2}{F^2t^2}}}\,.$$

При больших значениях t имеем:

$$\frac{m^2c^2}{F^2t^2} \ll 1,$$

и тогда:

$$v \approx \frac{c}{1+\dfrac{1}{2}\dfrac{m^2c^2}{F^2t^2}} \approx c\left(1-\frac{m^2c^2}{2F^2t^2}\right).$$

Хорошо видно, что при $t \to \infty$ скорость тела v неуклонно приближается к скорости света c, но всегда остаётся меньше c — как того и требует теория относительности.

Зависимость скорости тела от времени, даваемая формулой (5.31), графически представлена на рис. 5.10.

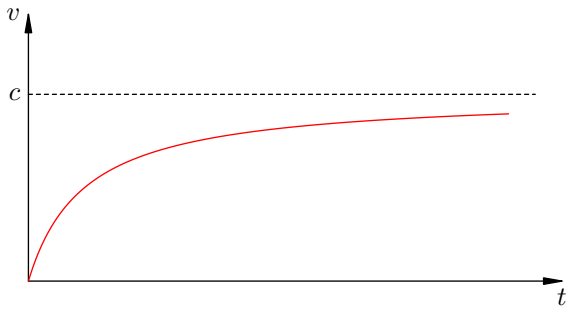


Рис. 5.10. Разгон тела под действием постоянной силы

Начальный участок графика — почти линейный; здесь пока работает классическая механика. Впоследствии сказываются релятивистские поправки, график искривляется, и при больших временах наша кривая асимптотически приближается к прямой $v = c$.

Глава 6

Квантовая физика

Классическая электродинамика, прекрасно описывающая широкий круг электромагнитных явлений, «дала сбой» в процессах взаимодействия света с веществом. Оказалось, что экспериментально установленные законы фотоэффекта нельзя объяснить в рамках электродинамики.

Планетарная модель атома, которую получил Резерфорд в ходе своих знаменитых опытов по рассеянию α-частиц, также находится в глубоком противоречии с электродинамикой Максвелла. Электрон, двигаясь вокруг ядра, должен был бы (согласно законам электродинамики) непрерывно излучать электромагнитные волны, расходовать на это свою энергию и в конце концов упасть на ядро. Атомы были бы неустойчивыми и «схлопывались», чего в действительности не наблюдается.

Выяснилось, что законы классической физики — механики Ньютона и электродинамики Максвелла — не работают внутри атома. Микромир живёт по совершенно иным правилам, которые непривычны и удивительны для нас. Для описания явлений микромира потребовалось создать совершенно новое направление — *квантовую физику*.

Квантовой физике посвящены следующие разделы пособия:

→ Фотоэффект
→ Фотоны
→ Корпускулярно-волновой дуализм
→ Линейчатые спектры
→ Строение атома
→ Атом Бора
→ Лазер
→ Строение ядра
→ Радиоактивность
→ Энергия связи ядра
→ Ядерные реакции

6.1 Фотоэффект

Фотоэффект — это выбивание электронов из вещества падающим светом. Явление фотоэффекта было открыто Генрихом Герцем в 1887 году в ходе его знаменитых экспериментов по излучению электромагнитных волн.

Напомним, что Герц использовал специальный разрядник (вибратор Герца) — разрезанный пополам стержень с парой металлических шариков на концах разреза. На стержень подавалось высокое напряжение, и в промежутке между шариками проскакивала искра. Так вот, Герц обнаружил, что при облучении шариков ультрафиолетовым светом проскакивание искры облегчалось!

Герц, однако, был поглощён исследованием электромагнитных волн и не принял данный факт во внимание. Год спустя фотоэффект был независимо открыт русским физиком Александром Григорьевичем Столетовым. Тщательные экспериментальные исследования, проведённые Столетовым в течение двух лет, позволили сформулировать основные законы фотоэффекта.

6.1.1 Опыты Столетова

В своих знаменитых экспериментах Столетов использовал фотоэлемент[1] собственной конструкции. Его схема изображена на рис. 6.1.

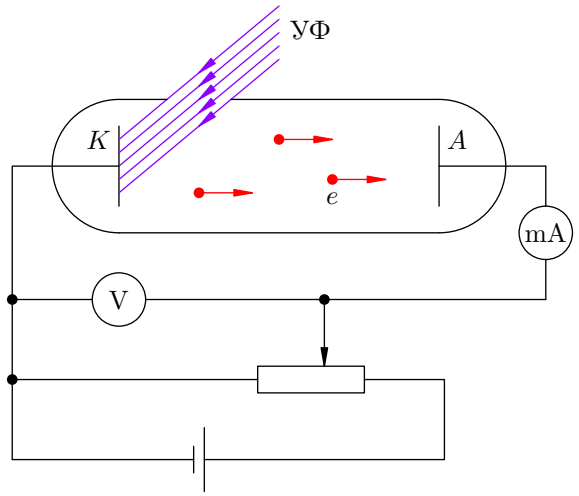


Рис. 6.1. Фотоэлемент Столетова

В стеклянную колбу, из которой выкачан воздух (чтобы не мешать лететь электронам), введены два электрода: цинковый катод K и анод A. На катод и анод подаётся напряжение, величину U которого можно менять с помощью потенциометра и измерять вольтметром V.

Сейчас на катод подан «минус», а на анод — «плюс», но можно сделать и наоборот (и эта перемена знака — существенная часть опытов Столетова). Напряжению на электродах приписывается тот знак, который подан на анод[2]. В данном случае, например, напряжение U положительно.

Катод освещается ультрафиолетовыми лучами УФ через специальное кварцевое окошко, сделанное в колбе (стекло поглощает ультрафиолет, а кварц пропускает). Ультрафиолетовое излучение выбивает с катода электроны e, которые разгоняются напряжением U и летят на анод. Включённый в цепь миллиамперметр mA регистрирует электрический ток. Этот ток называется *фототоком*, а выбитые электроны, его создающие, называются *фотоэлектронами*.

[1] *Фотоэлементом* называется любое устройство, позволяющее наблюдать фотоэффект.

[2] Поэтому поданное на электроды напряжение U часто называют *анодным напряжением*.

В опытах Столетова можно независимо варьировать три величины: анодное напряжение, интенсивность света и его частоту. Начнём с напряжения.

6.1.2 Зависимость фототока от напряжения

Меняя величину и знак анодного напряжения, можно проследить, как меняется фототок. График этой зависимости, называемый *характеристикой фотоэлемента*, представлен на рис. 6.2.

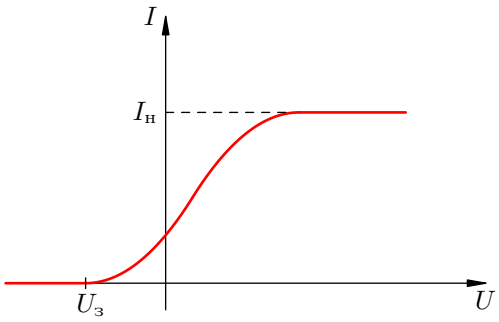


Рис. 6.2. Характеристика фотоэлемента

Давайте обсудим ход полученной кривой. Прежде всего заметим, что электроны вылетают из катода с различными скоростями и в разных направлениях; максимальную скорость, которую имеют фотоэлектроны в условиях опыта, обозначим v.

Если напряжение U отрицательно и велико по модулю, то фототок отсутствует. Это легко понять: электрическое поле, действующее на электроны со стороны катода и анода, является тормозящим (на катоде «плюс», на аноде «минус») и обладает столь большой величиной, что электроны не в состоянии долететь до анода. Начального запаса кинетической энергии не хватает — электроны теряют свою скорость на подступах к аноду и разворачиваются обратно на катод. Максимальная кинетическая энергия вылетевших электронов оказывается меньше, чем модуль работы поля при перемещении электрона с катода на анод:

$$\frac{mv^2}{2} < eU.$$

Здесь $m = 9{,}1 \cdot 10^{-31}$ кг — масса электрона, $e = -1{,}6 \cdot 10^{-19}$ Кл — его заряд.

Будем постепенно увеличивать напряжение, т. е. двигаться слева направо вдоль оси U из далёких отрицательных значений.

Поначалу тока по-прежнему нет, но точка разворота электронов становится всё ближе к аноду. Наконец, при достижении напряжения $U_з$, которое называется *задерживающим напряжением*, электроны разворачиваются назад в момент достижения анода (иначе говоря, электроны прибывают на анод с нулевой скоростью). Имеем:

$$\frac{mv^2}{2} = eU_з. \tag{6.1}$$

Таким образом, *величина задерживающего напряжения позволяет определить максимальную кинетическую энергию фотоэлектронов.*

При небольшом превышении задерживающего напряжения появляется слабый фототок. Его формируют электроны, вылетевшие с максимальной кинетической энергией почти точно вдоль оси колбы (т. е. почти перпендикулярно катоду): теперь электронам хватает этой энергии, чтобы добраться до анода с ненулевой скоростью и замкнуть цепь. Остальные электроны, которые имеют меньшие скорости или полетели в сторону от анода, на анод не попадают.

При повышении напряжения фототок увеличивается. Анода достигает большее количество электронов, вылетающих из катода под всё бóльшими углами к оси колбы. Обратите внимание, что фототок присутствует при нулевом напряжении!

Когда напряжение выходит в область положительных значений, фототок продолжает возрастать. Оно и понятно: электрическое поле теперь разгоняет электроны, поэтому всё большее их число получают шанс оказаться на аноде. Однако достигают анода пока ещё не все фотоэлектроны. Например, электрон, вылетевший с максимальной скоростью перпендикулярно оси колбы (т. е. вдоль катода), хоть и развернётся полем в нужном направлении, но не настолько сильно, чтобы попасть на анод.

Наконец, при достаточно больших положительных значениях напряжения ток достигает своей предельной величины $I_н$, называемой *током насыщения*, и дальше возрастать перестаёт. Почему? Дело в том, что напряжение, ускоряющее электроны, становится настолько велико, что анод захватывает вообще все электроны, выбитые из катода — в каком бы направлении и с какими бы скоростями они не начинали движение. Стало быть, дальнейших возможностей увеличиваться у фототока попросту нет — ресурс, так сказать, исчерпан.

6.1.3 Законы фотоэффекта

Величина $I_н$ тока насыщения — это, по существу, количество электронов, выбиваемых из катода за одну секунду. Будем менять интенсивность света, не трогая частоту. Опыт показывает, что ток насыщения меняется пропорционально интенсивности света.

Первый закон фотоэффекта. Число электронов, выбиваемых из катода за секунду, пропорционально интенсивности падающего на катод излучения (при его неизменной частоте).

Ничего неожиданного в этом нет: чем больше энергии несёт излучение, тем ощутимее наблюдаемый результат. Загадки начинаются дальше.

А именно, будем изучать зависимость максимальной кинетической энергии фотоэлектронов от частоты и интенсивности падающего света. Сделать это несложно: ведь в силу формулы (6.1) нахождение максимальной кинетической энергии выбитых электронов фактически сводится к измерению задерживающего напряжения.

Сначала меняем частоту излучения ν при фиксированной интенсивности. Получается такой график (рис. 6.3):

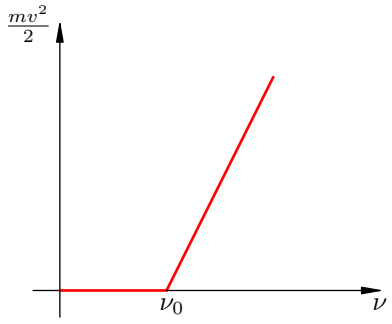


Рис. 6.3. Зависимость энергии фотоэлектронов от частоты света

Как видим, существует некоторая частота ν_0, называемая *красной границей фотоэффекта*, разделяющая две принципиально разные области графика. Если $\nu < \nu_0$, то фотоэффекта нет. Если же $\nu > \nu_0$, то максимальная кинетическая энергия фотоэлектронов линейно растёт с частотой.

Теперь, наоборот, фиксируем частоту и меняем интенсивность света. Если при этом $\nu < \nu_0$, то фотоэффект не возникает, какова бы ни была интенсивность! Не менее удивительный факт

обнаруживается и при $\nu > \nu_0$: максимальная кинетическая энергия фотоэлектронов от интенсивности света не зависит.

Все эти факты нашли отражение во втором и третьем законах фотоэффекта.

Второй закон фотоэффекта. Максимальная кинетическая энергия фотоэлектронов линейно возрастает с частотой света и не зависит от его интенсивности.

Третий закон фотоэффекта. Для каждого вещества существует красная граница фотоэффекта — наименьшая частота света ν_0, при которой фотоэффект ещё возможен. При $\nu < \nu_0$ фотоэффект не наблюдается ни при какой интенсивности света.

6.1.4 Трудности классического объяснения фотоэффекта

Как можно было бы объяснить фотоэффект с точки зрения классической электродинамики и волновых представлений о свете?

Известно, что для вырывания электрона из вещества требуется сообщить ему некоторую энергию A, называемую *работой выхода* электрона. В случае свободного электрона в металле это работа по преодолению поля положительных ионов кристаллической решётки, удерживающего электрон на границе металла. В случае электрона, находящегося в атоме, работа выхода есть работа по разрыву связи электрона с ядром.

В переменном электрическом поле световой волны электрон начинает совершать колебания. И если энергия колебаний превысит работу выхода, то электрон будет вырван из вещества.

Однако в рамках таких представлений невозможно понять второй и третий законы фотоэффекта. Действительно, почему кинетическая энергия выбитых электронов не зависит от интенсивности излучения? Ведь чем больше интенсивность, тем больше напряжённость электрического поля в электромагнитной волне, тем больше сила, действующая на электрон, тем больше энергия его колебаний и с тем большей кинетической энергией электрон вылетит из катода. Логично? Логично. Но эксперимент показывает иное.

Далее, откуда берётся красная граница фотоэффекта? Чем «провинились» низкие частоты? Казалось бы, с ростом интенсивности света растёт и сила, действующая на электроны; поэтому даже при низкой частоте света электрон рано или поздно будет вырван из вещества — когда интенсивность достигнет достаточно большого значения. Однако красная граница ставит жёсткий запрет на вылет электронов при низких частотах падающего излучения.

Кроме того, неясна *безынерционность* фотоэффекта. Именно, при освещении катода излучением сколь угодно слабой интенсивности (с частотой выше красной границы) фотоэффект начинается *мгновенно* — в момент включения освещения. Между тем, казалось бы, электронам требуется некоторое время для «расшатывания» связей, удерживающих их в веществе, и это время «раскачки» должно быть тем больше, чем слабее падающий свет. Аналогия такая: чем слабее вы толкаете качели, тем дольше придётся их раскачивать до заданной амплитуды. Выглядит опять-таки логично, но опыт — единственный критерий истины в физике! — этим доводам противоречит.

Так на рубеже XIX и XX столетий в физике возникла тупиковая ситуация: электродинамика, предсказавшая существование электромагнитных волн и великолепно работающая в диапазоне радиоволн, отказалась объяснять явление фотоэффекта.

Выход из этого тупика был найден Альбертом Эйнштейном в 1905 году. Он предложил простое уравнение, описывающее фотоэффект. Все три закона фотоэффекта оказались следствиями уравнения Эйнштейна.

Главная заслуга Эйнштейна состояла в отказе от попыток истолковать фотоэффект с позиций классической электродинамики. Эйнштейн привлёк к делу смелую гипотезу о квантах, высказанную Максом Планком пятью годами ранее.

6.1.5 Гипотеза Планка о квантах

Классическая электродинамика отказалась работать не только в области фотоэффекта. Она также дала серьёзный сбой, когда её попытались использовать для описания излучения нагретого тела (так называемого теплового излучения).

Суть проблемы состояла в том, что простая и естественная электродинамическая модель теплового излучения приводила к бессмысленному выводу: любое нагретое тело, непрерывно излучая, должно постепенно потерять всю свою энергию и остыть до абсолютного нуля. Как мы прекрасно знаем, ничего подобного не наблюдается.

В ходе решения этой проблемы Макс Планк высказал свою знаменитую гипотезу.

Гипотеза о квантах. Электромагнитная энергия излучается и поглощается не непрерывно, а отдельными неделимыми порциями — квантами. Энергия кванта пропорциональна частоте излучения:

$$E = h\nu. \tag{6.2}$$

Соотношение (6.2) называется *формулой Планка*, а коэффициент пропорциональности h — *постоянной Планка*.

Принятие этой гипотезы позволило Планку построить теорию теплового излучения, прекрасно согласующуся с экспериментом. Располагая известными из опыта спектрами теплового излучения, Планк вычислил значение своей постоянной:

$$h = 6{,}63 \cdot 10^{-34}\ \text{Дж} \cdot \text{с}. \tag{6.3}$$

Успешность гипотезы Планка наводила на мысль, что законы классической физики неприменимы к малым частицам вроде атомов или электронов, а также к явлениям взаимодействия света и вещества. Подтверждением данной мысли как раз и послужило явление фотоэффекта.

6.1.6 Уравнение Эйнштейна для фотоэффекта

Гипотеза Планка говорила о дискретности *излучения* и *поглощения* электромагнитных волн, то есть о прерывистом характере взаимодействия света с веществом. При этом Планк считал, что *распространение* света — это непрерывный процесс, происходящий в полном соответствии с законами классической электродинамики.

Эйнштейн пошёл ещё дальше: он предположил, что *свет в принципе обладает прерывистой структурой: не только излучение и поглощение, но также и распространение света происходит отдельными порциями — квантами, обладающими энергией* $E = h\nu$.

Планк рассматривал свою гипотезу лишь как математический трюк и не решился опровергнуть электродинамику применительно к микромиру. Физической реальностью кванты стали благодаря Эйнштейну.

Кванты электромагнитного излучения (в частности, кванты света) стали впоследствии называться *фотонами*. Таким образом, свет состоит из особых частиц — фотонов, движущихся в вакууме со скоростью c. Каждый фотон монохроматического света, имеющего частоту ν, несёт энергию $h\nu$.

Фотоны могут обмениваться энергией и импульсом[3] с частицами вещества; в таком случае мы говорим о *столкновении* фотона и частицы. В частности, происходит столкновение фотонов с электронами металла катода.

Поглощение света — это поглощение фотонов, то есть *неупругое* столкновение фотонов с частицами (атомами, электронами). Поглощаясь при столкновении с электроном, фотон передаёт ему свою энергию. В результате электрон получает кинетическую энергию мгновенно, а не постепенно, и именно этим объясняется безынерционность фотоэффекта.

[3] Об импульсе фотона речь пойдёт в следующем разделе.

Уравнение Эйнштейна для фотоэффекта есть не что иное, как закон сохранения энергии. На что идёт энергия фотона $h\nu$ при его неупругом столкновении с электроном? Она расходуется на совершение работы выхода A по извлечению электрона из вещества и на придание электрону кинетической энергии $mv^2/2$:

$$h\nu = A + \frac{mv^2}{2}\,. \tag{6.4}$$

Слагаемое $mv^2/2$ оказывается *максимальной* кинетической энергией фотоэлектронов. Почему максимальной? Этот вопрос требует небольшого пояснения.

Электроны в металле могут быть свободными и связанными. Свободные электроны «гуляют» по всему металлу, связанные электроны «сидят» внутри своих атомов. Кроме того, электрон может находиться как вблизи поверхности металла, так и в его глубине.

Ясно, что максимальная кинетическая энергия фотоэлектрона получится в том случае, когда фотон попадёт на свободный электрон в поверхностном слое металла — тогда для выбивания электрона достаточно одной лишь работы выхода.

Во всех других случаях придётся затрачивать дополнительную энергию — на вырывание связанного электрона из атома или на «протаскивание» глубинного электрона к поверхности. Эти лишние затраты приведут к тому, что кинетическая энергия вылетевшего электрона окажется меньше.

Замечательное по простоте и физической ясности уравнение (6.4) содержит в себе всю теорию фотоэффекта. Давайте посмотрим, какое объяснение получают законы фотоэффекта с точки зрения уравнения Эйнштейна.

1. Число выбиваемых электронов пропорционально числу поглощённых фотонов. С увеличением интенсивности света количество фотонов, падающих на катод за секунду, возрастает. Стало быть, пропорционально возрастает число поглощённых фотонов и, соответственно, число выбитых за секунду электронов.

2. Выразим из формулы (6.4) кинетическую энергию:

$$\frac{mv^2}{2} = h\nu - A.$$

   Действительно, кинетическая энергия выбитых электронов линейно растёт с частотой и не зависит от интенсивности света.

   Зависимость кинетической энергии от частоты имеет вид уравнения прямой, проходящей через точку $(A/h, 0)$. Этим полностью объясняется ход графика на рис. 6.3.

3. Для того, чтобы начался фотоэффект, энергии фотона должно хватить как минимум на совершение работы выхода: $h\nu \geqslant A$. Наименьшая частота ν_0, определяемая равенством

$$h\nu_0 = A,$$

   как раз и будет красной границей фотоэффекта. Как видим, красная граница фотоэффекта $\nu_0 = A/h$ определяется только работой выхода, т. е. зависит лишь от вещества облучаемой поверхности катода.

   Если $\nu < \nu_0$, то фотоэффекта не будет — сколько бы фотонов за секунду не падало на катод. Следовательно, интенсивность света роли не играет; главное — хватает ли отдельному фотону энергии, чтобы выбить электрон.

Уравнение Эйнштейна (6.4) даёт возможность экспериментального нахождения постоянной Планка. Для этого надо предварительно определить частоту излучения и работу выхода материала катода, а также измерить кинетическую энергию фотоэлектронов.

В ходе таких опытов было получено значение h, в точности совпадающее с (6.3). Такое совпадение результатов двух независимых экспериментов — на основе спектров теплового излучения и уравнения Эйнштейна для фотоэффекта — означало, что обнаружены совершенно новые «правила игры», по которым происходит взаимодействие света и вещества. В этой области классическая физика в лице механики Ньютона и электродинамики Максвелла уступает место *квантовой физике* — теории микромира, построение которой продолжается и сегодня.

6.2 Фотоны

В результате исследования явлений, связанных с взаимодействием света и вещества (тепловое излучение и фотоэффект), физики пришли к выводу, что свет состоит из отдельных порций энергии — *фотонов*. Излучение света, его распространение и поглощение происходит строго этими порциями.

Фотоны обладают энергией и импульсом и могут обмениваться ими с частицами вещества (скажем, с электронами или атомами). Как вы помните, такой обмен импульсом и энергией мы называем *столкновением* фотона и частицы.

При упругом столкновении фотон меняет направление движения — свет рассеивается. При неупругом столкновении фотон поглощается отдельной частицей или совокупностью частиц вещества — так происходит поглощение света.

Словом, фотон ведёт себя как частица и поэтому — наряду с электроном, протоном и некоторыми другими частицами — причислен к разряду *элементарных частиц*.

6.2.1 Энергия фотона

Выражение для энергии фотона с частотой ν мы уже знаем:

$$E = h\nu. \tag{6.5}$$

Часто бывает удобно работать не с обычной частотой ν, а с циклической частотой $\omega = 2\pi\nu$. Тогда вводят другую постоянную Планка «аш с чертой»:

$$\hbar = \frac{h}{2\pi} = 1{,}05 \cdot 10^{-34}\ \text{Дж} \cdot \text{с}.$$

Выражение (6.5) для энергии фотона примет вид:

$$E = \hbar\omega.$$

Фотон движется в вакууме со скоростью света c и потому является *релятивистской* частицей: описывая фотон, мы должны привлекать формулы теории относительности. А там имеется такая формула для энергии тела массы m, движущегося со скоростью v:

$$E = \frac{mc^2}{\sqrt{1 - \dfrac{v^2}{c^2}}}\,. \tag{6.6}$$

Если предположить, что $m \neq 0$, то формула (6.6) приводит к бессмысленному заключению: энергия фотона должна быть бесконечной. Чтобы избежать этого противоречия, остаётся признать, что *масса фотона равна нулю*. Формула (6.6) позволяет сделать и более общий вывод: *только безмассовая частица может двигаться со скоростью света*.

6.2.2 Импульс фотона

Обладая энергией, фотон должен обладать и импульсом. Действительно, важнейшая формула теории относительности даёт связь энергии и импульса частицы:

$$E^2 = p^2c^2 + m^2c^4. \tag{6.7}$$

Для фотона, имеющего нулевую массу, эта формула сводится к простому соотношению:

$$E = pc.$$

Отсюда для импульса фотона получаем:

$$p = \frac{E}{c} = \frac{h\nu}{c}\,. \tag{6.8}$$

Направление импульса фотона совпадает с направлением светового луча.

Учитывая, что отношение c/ν есть длина волны λ, формулу (6.8) можно переписать так:

$$p = \frac{h}{\lambda}\,. \tag{6.9}$$

В видимом диапазоне наименьшими значениями энергии и импульса обладают фотоны красного света — у них самая маленькая частота (и самая большая длина волны). При движении в сторону фиолетового участка спектра энергия и импульс фотона линейно возрастают с частотой.

6.2.3 Давление света

Свет оказывает давление на освещаемую поверхность. Такой вывод был сделан Максвеллом из теоретических соображений и получил экспериментальное подтверждение в знаменитых опытах П. Н. Лебедева. Если понимать свет как поток фотонов, обладающих импульсом, то можно легко объяснить давление света и вывести формулу Максвелла.

Предположим, что на некоторое тело падает свет частоты ν. Лучи направлены перпендикулярно поверхности тела; площадь освещаемой поверхности равна S (рис. 6.4).

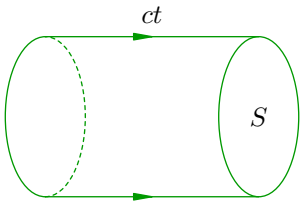


Рис. 6.4. Давление света

Пусть n — концентрация фотонов падающего света, то есть число фотонов в единице объёма. За время t на нашу поверхность попадают фотоны, находящиеся внутри цилиндра высотой ct. Их число равно:

$$N = nV = nSct.$$

При падении света на поверхность тела часть световой энергии отражается, а часть — поглощается. Пусть r — *коэффициент отражения* света; величина $r < 1$ показывает, какая часть световой энергии отражается от поверхности. Соответственно, величина $1 - r$ — это доля падающей энергии, поглощаемая телом.

Как мы теперь знаем, энергия света пропорциональна числу фотонов. Поэтому можно написать, какое количество фотонов (из общего числа N) отразится от поверхности, а какое — поглотится ею:

$$N_{\text{отр}} = rN, \quad N_{\text{погл}} = (1 - r)N.$$

Импульс каждого падающего фотона равен $p = h\nu/c$. Поглощённый фотон испытывает неупругое столкновение с телом и передаёт ему импульс p. Отражённый фотон после упругого столкновения меняет направление своего импульса на противоположное, и поэтому импульс, переданный телу отражённым фотоном, равен $2p$.

Таким образом, от каждого фотона, входящего в световой поток, тело получает некоторый импульс. Вот простая и очевидная причина того, что свет оказывает давление на освещаемую поверхность.

Суммарный импульс, полученный телом от N падающих фотонов, равен:

$$P = 2p \cdot N_{\text{отр}} + p \cdot N_{\text{погл}} = 2prN + p(1-r)N = (1+r)pN.$$

На нашу поверхность S действует сила F, равная импульсу, полученному телом в единицу времени:

$$F = \frac{P}{t} = (1+r)p\frac{N}{t} = (1+r)\frac{h\nu}{c}\frac{nSct}{t} = (1+r)h\nu nS.$$

Давление света есть отношение этой силы к площади освещаемой поверхности:

$$p_{\text{света}} = \frac{F}{S} = (1+r)h\nu n. \tag{6.10}$$

Выражение $h\nu n$ имеет простой физический смысл: будучи произведением энергии фотона на число фотонов в единице объёма, оно равно энергии света в единице объёма, то есть объёмной плотности энергии w. Тогда соотношение (6.10) приобретает вид:

$$p_{\text{света}} = (1+r)w.$$

Это и есть формула для давления света, теоретически выведенная Максвеллом (в рамках классической электродинамики) и экспериментально проверенная в опытах Лебедева.

6.2.4 Двойственная природа света

В результате рассмотрения всей совокупности оптических явлений возникает естественный вопрос: что же такое свет? Непрерывно распределённая в пространстве электромагнитная волна или поток отдельных частиц — фотонов? Теория и эксперименты приводят к заключению, что оба ответа должны быть утвердительными.

1. Явления интерференции и дифракции света, характерные для любых волновых процессов, не оставляют сомнений в том, что свет есть форма волнового движения материи. Таким образом, мы должны признать: да, *свет имеет волновую природу*, свет — это электромагнитная волна.

2. Однако явления взаимодействия света и вещества (например, фотоэффект) указывают на то, что свет ведёт себя как поток отдельных частиц. Эти частицы — фотоны — ведут, так сказать, самостоятельный образ жизни, обладают энергией и импульсом, участвуют во взаимодействиях с атомами и электронами. Излучение света — это рождение фотонов. Распространение света — это движение фотонов в пространстве. Отражение и поглощение света — это соответственно упругие и неупругие столкновения фотонов с частицами вещества.

   Все попытки истолковать указанные явления излучения и поглощения света в рамках волновых представлений классической физики окончились неудачей. Оставалось лишь согласиться с тем, что *свет имеет корпускулярную*[4] *природу*, свет — это совокупность фотонов, мчащихся в пространстве.

Таким образом, свет имеет двойственную, *корпускулярно-волновую* природу — он может проявлять себя то так, то эдак. В одних явлениях (интерференция, дифракция) на передний план выходит волновая природа, и свет ведёт себя в точности как волна. Но в других явлениях (фотоэффект) доминирует корпускулярная природа, и свет ведёт себя подобно потоку частиц.

[4] От латинского слова *corpusculum* — маленькое тельце, частица.

Странно всё это, не правда ли? Но что поделать — так устроена природа. Мы, люди, живём среди макроскопических тел, и наше воображение оказалось не способным полноценно представить себе явления микромира.

Природа, однако, неизмеримо шире и богаче того, что может вместить в себя человеческое воображение. Признав это и руководствуясь не столько собственным воображением, сколько наблюдениями, результатами экспериментов и весьма изощрённой математикой, люди начали успешно создавать квантовую теорию микроскопических явлений и процессов.

О некоторых парадоксальных на первый взгляд — но тем не менее подтверждённых экспериментально! — выводах квантовой теории мы поговорим в следующем разделе.

6.3 Корпускулярно-волновой дуализм

Корпускулярно-волновой дуализм[5] — это физический принцип, утверждающий, что любой объект природы может вести себя и как частица, и как волна.

С первым проявлением этого принципа мы столкнулись в предыдущем листке, когда говорили о двойственной, корпускулярно-волновой природе света. В явлениях интерференции и дифракции свет демонстрирует свою волновую природу. В явлении фотоэффекта свет выступает как дискретный поток частиц — фотонов.

Является ли свет каким-то особым объектом нашего мира, таким, что подобный дуализм присущ только ему? Или, быть может, корпускулярно-волновой дуализм — это свойство вообще всех материальных объектов, просто впервые обнаружен он был для света?

6.3.1 Гипотеза де Бройля

Идея об универсальной двойственности корпускулярных и волновых свойств всех объектов природы была впервые высказана Луи де Бройлем (в 1924 году) в качестве гипотезы о волновых свойствах частиц.

Итак, мы знаем, что свету с частотой ν и длиной волны λ соответствуют частицы — фотоны, обладающие энергией $E = h\nu$ и импульсом $p = h/\lambda$. Де Бройль, в сущности, постулировал обратное.

Гипотеза де Бройля. Движению каждой частицы соответствует распространение некоторой волны. Частота и длина этой волны определяются энергией и импульсом частицы:

$$\nu = \frac{E}{h}\,, \quad \lambda = \frac{h}{p}\,. \tag{6.11}$$

Точно так же, любой волне с частотой ν и длиной волны λ отвечают частицы с энергией $E = h\nu$ и импульсом $p = h/\lambda$.

Чтобы лучше осмыслить гипотезу де Бройля, давайте обсудим дуализм «волна–частица» на примере электромагнитного излучения.

В случае электромагнитных волн мы имеем следующую закономерность. По мере увеличения длины волны всё легче наблюдать волновые свойства излучения и всё труднее — корпускулярные. И наоборот, чем меньше длина волны, тем ярче выражены корпускулярные свойства излучения и тем труднее наблюдать его волновые свойства. Изменение соотношения корпускулярных и волновых свойств хорошо прослеживается при движении по известной вам шкале электромагнитных волн.

- *Радиоволны.* Длины волн здесь настолько велики, что корпускулярные свойства излучения практически не проявляются. Волновые свойства в этом диапазоне абсолютно доминируют.

  Длины волн могут составлять несколько метров или даже километров, так что волновая природа проявляется «сама собой» — радиоволны в процессе дифракции запросто огибают дома или горы. Излучение радиоволн и их взаимодействие с материальными объектами отлично описывается в рамках классической электродинамики.

- *Видимый свет и ультрафиолет.* Это своего рода «переходная область»: в оптике мы можем наблюдать как волновые свойства света, так и корпускулярные.

  Однако в обоих случаях надо постараться. Так, длины волн видимого света много меньше размеров окружающих нас тел, поэтому в опытах по интерференции или дифракции

[5] Слово *дуализм* означает *двойственность*.

света нужно создавать специальные условия (малость щелей или отверстий, удалённость экрана). В свою очередь, термин «красная граница фотоэффекта» также подчёркивает пограничность данного диапазона: фотоэффект начинается лишь при переходе через красную границу.

- *Рентгеновское и гамма-излучение.* Длины волн очень малы, и наблюдать волновые свойства излучения весьма затруднительно. Так, верхняя граница длин волн рентгеновского излучения составляет 10 нм; это лишь на два порядка превышает размер атома. Ясно, что дифракцию на «обычных» препятствиях при такой длине волны наблюдать невозможно.

  Однако в рентгеновский диапазон входят длины волн порядка размера атома и межатомных расстояний в кристалле (0,1 нм). Поэтому дифракция рентгеновских лучей наблюдается на «естественных» дифракционных решётках — а именно, на кристаллических решётках твёрдых тел[6].

  Энергия квантов в рентгеновском и гамма-диапазоне настолько велика, что излучение ведёт себя почти стопроцентно как поток частиц.

Рассуждая по аналогии с электромагнитными волнами, можно заключить, что и частица будет проявлять волновые свойства тем лучше, чем больше её длина волны де Бройля (в масштабах данной ситуации).

Так, мы совсем не наблюдаем волновых свойств у окружающих нас тел. (Видели вы, например, интерференцию движущихся автомобилей?) А почему? Давайте посчитаем длину дебройлевской волны объекта массой 1 кг, движущегося со скоростью 1 м/с:

$$\lambda = \frac{6{,}63 \cdot 10^{-34}}{1 \cdot 1} = 6{,}63 \cdot 10^{-34}\,\text{м}.$$

Это на 25 порядков меньше размера атома. Воображение отказывается представить себе столь малую величину. Разумеется, никакого волнового поведения у нашего объекта при таких условиях не обнаруживается — он стопроцентно ведёт себя как «частица», то есть как материальная точка классической механики.

6.3.2 Дифракция электронов

Совсем другое дело — электрон. Масса электрона равна $9{,}1 \cdot 10^{-31}$ кг, и столь малое значение массы (а стало быть, и импульса в формуле $\lambda = h/p$) может дать длину волны де Бройля, достаточную для экспериментального обнаружения волновых свойств.

И вот оказывается, что электроны с энергией 100 эВ (при такой энергии становится несущественным хаотическое тепловое движение электронов, и электронный пучок можно считать когерентным) имеют дебройлевскую длину волны примерно 0,1 нм — это как раз порядка размера атома и расстояний между атомами в кристаллической решётке! Опыт по наблюдению дифракции рентгеновских лучей на кристаллических структурах уже имелся, поэтому оставалось направить на кристаллическую решётку пучок электронов.

Впервые это было сделано в знаменитом эксперименте американских физиков Дэвиссона и Джермера (1927 год). Дифракция электронов на кристаллах была обнаружена! Как и ожидалось, полученная дифракционная картина имела тот же характер, что и при дифракции на кристаллической решётке рентгеновских лучей.

Впоследствии волновые свойства были обнаружены и у более крупных частиц: протонов, нейтронов, атомов и молекул. Гипотеза де Бройля, таким образом, получила надёжное опытное подтверждение.

[6] Эта идея была высказана немецким физиком Лауэ в 1912 году.

6.3.3 Соотношение неопределённостей

Обнаружение корпускулярных свойств электромагнитных волн и волновых свойств частиц показало, что объекты микромира подчиняются необычным законам. Эти законы совершенно неожиданны для нас, привыкших наблюдать за макроскопическими телами.

Наше сознание выработало некоторые образы частицы и волны, вполне пригодные для описания объектов классической физики. Частица — это маленький, локализованный в пространстве сгусток вещества. Волна — это распределённый (не локализованный) в пространстве колебательный процесс. Как же эти понятия могут совмещаться в одном объекте (например, в электроне)?

Вообразить такое действительно получается с трудом. Но что поделать — это факт. Природа оказывается намного богаче нашего воображения. В своей повседневной жизни мы находимся очень далеко от микромира, и в привычном нам диапазоне макроскопических тел природа демонстрирует свои «крайние» проявления — в виде «только частиц» или «только волн». Вот почему корпускулярные и волновые свойства представляются нам несовместимыми друг с другом. Но на самом деле это не так: в микромире оказывается, что один и тот же объект (например, электрон) легко может обладать обоими свойствами одновременно — словно человек, обладающий разными, несовместимыми на первый взгляд чертами характера.

Так, будучи частицей, электрон локализован в пространстве; но, будучи волной, локализован не в точке, а «размазан» по некоторой области. Координаты и скорость электрона не могут быть измерены одновременно сколь угодно точно. Неопределённость координаты Δx и неопределённость соответствующей проекции импульса Δp_x оказываются связанными *соотношением неопределённостей* Гейзенберга:

$$\Delta x \cdot \Delta p_x \geqslant \hbar. \tag{6.12}$$

Соотношение неопределённостей (6.12) имеет фундаментальный характер — оно применимо к любым объектам природы. Чем точнее мы знаем координаты объекта (то есть чем в меньшей пространственной области он локализован), тем больше получается разброс значений его импульса (то есть тем с большей скоростью объект «готов вылететь» из этой области). И наоборот, чем точнее мы знаем импульс объекта, тем меньше у нас информации о том, где этот объект находится.

Но коль скоро нет возможности одновременно точно измерить координаты и скорость, то теряет смысл понятие траектории движения объекта. Механика Ньютона перестаёт работать в микромире и уступает место *квантовой механике*.

6.4 Линейчатые спектры

Если пропустить солнечный свет через стеклянную призму или дифракционную решётку, то возникнет хорошо известный вам *непрерывный спектр* (рис. 6.5)[7]:

Рис. 6.5. Непрерывный спектр

Спектр называется непрерывным потому, что в нём присутствуют все длины волн видимого диапазона — от красной границы до фиолетовой. Мы наблюдаем непрерывный спектр в виде сплошной полосы, состоящей из разных цветов.

Непрерывным спектром обладает не только солнечный свет, но и, например, свет электрической лампочки. Вообще, оказывается, что любые твёрдые и жидкие тела (а также весьма плотные газы), нагретые до высокой температуры, дают излучение с непрерывным спектром.

Ситуация качественно меняется, когда мы наблюдаем свечение разреженных газов. Спектр перестаёт быть непрерывным: в нём появляются разрывы, увеличивающиеся по мере разрежения газа. В предельном случае чрезвычайно разреженного атомарного газа спектр становится *линейчатым* — состоящим из отдельных достаточно тонких линий.

Мы рассмотрим два типа линейчатых спектров: спектр испускания и спектр поглощения.

6.4.1 Спектр испускания

Предположим, что газ состоит из *атомов* некоторого химического элемента и разрежен настолько, что атомы почти не взаимодействуют друг с другом. Раскладывая в спектр излучение такого газа (нагретого до достаточно высокой температуры), мы увидим примерно следующую картину (рис. 6.6):

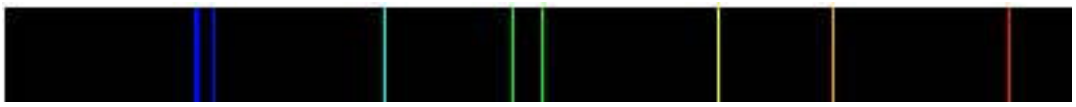


Рис. 6.6. Линейчатый спектр испускания

Этот линейчатый спектр, образованный тонкими изолированными разноцветными линиями, называется *спектром испускания*.

Любой атомарный разреженный газ излучает свет с линейчатым спектром. Более того, для каждого химического элемента спектр испускания оказывается уникальным, играя роль «удостоверения личности» этого элемента. По набору линий спектра испускания можно однозначно сказать, с каким химическим элементом мы имеем дело.

Поскольку газ разрежен и атомы мало взаимодействуют друг с другом, мы можем заключить, что свет излучают атомы *сами по себе*. Таким образом, *атом характеризуется дискретным, строго определённым набором длин волн излучаемого света*. У каждого химического элемента, как мы уже сказали, этот набор свой.

[7]Изображения на рис. 6.5, 6.6 и 6.7 взяты с сайта www.nanospectrum.ru.

6.4.2 Спектр поглощения

Атомы излучают свет, переходя из возбуждённого состояния в основное. Но вещество может не только излучать, но и поглощать свет. Атом, поглощая свет, совершает обратный процесс — переходит из основного состояния в возбуждённое.

Снова рассмотрим разреженный атомарный газ, но на сей раз в холодном состоянии (при достаточно низкой температуре). Свечения газа мы не увидим; не будучи нагретым, газ не излучает — атомов в возбуждённом состоянии оказывается для этого слишком мало.

Если сквозь наш холодный газ пропустить свет с непрерывным спектром, то можно увидеть что-то вроде этого (рис. 6.7):

Рис. 6.7. Линейчатый спектр поглощения

На фоне непрерывного спектра падающего света появляются тёмные линии, которые образуют так называемый *спектр поглощения*. Откуда берутся эти линии?

Под действием падающего света атомы газа переходят в возбуждённое состояние. При этом оказывается, что для возбуждения атомов годятся не любые длины волн, а лишь некоторые, строго определённые для данного сорта газа. Вот именно эти длины волн газ и «забирает себе» из проходящего света.

Более того, газ изымает из непрерывного спектра ровно те самые длины волн, которые излучает сам! Тёмные линии в спектре поглощения газа в точности соответствуют ярким линиям его спектра испускания. На рис. 6.8 сопоставлены спектры испускания и поглощения разреженных паров натрия[8]:

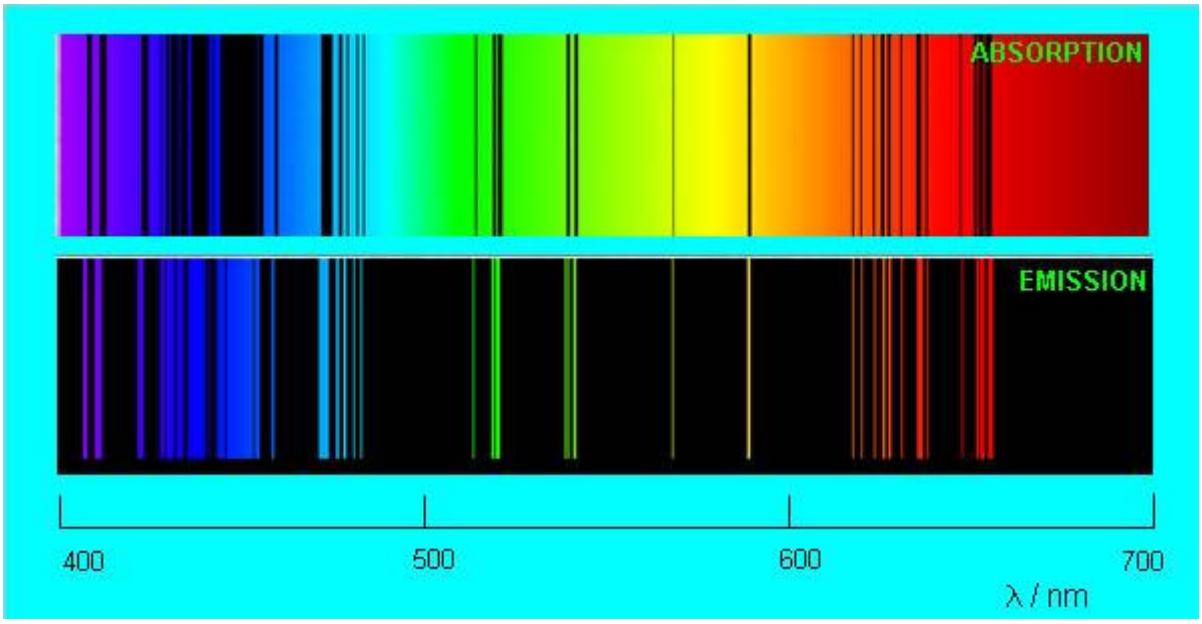


Рис. 6.8. Спектры поглощения и испускания для натрия

Впечатляющее совпадение линий, не правда ли?

Глядя на спектры испускания и поглощения, физики XIX века пришли к выводу, что атом не является неделимой частицей и обладает некоторой внутренней структурой. В самом деле, что-то ведь внутри атома должно обеспечивать механизм излучения и поглощения света!

[8]Изображение с сайта www.nt.ntnu.no.

Кроме того, уникальность атомных спектров говорит о том, что этот механизм различен у атомов разных химических элементов; стало быть, атомы разных химических элементов должны отличаться по своему внутреннему устройству.

Строению атома будет посвящён следующий раздел.

6.4.3 Спектральный анализ

Использование линейчатых спектров в качестве уникальных «паспортов» химических элементов лежит в основе *спектрального анализа* — метода исследования химического состава вещества по его спектру.

Идея спектрального анализа проста: спектр излучения исследуемого вещества сопоставляется с эталонными спектрами химических элементов, после чего делается вывод о присутствии или отсутствии того или иного химического элемента в данном веществе. При определённых условиях методом спектрального анализа можно определить химический состав не только качественно, но и количественно.

В результате наблюдения различных спектров были открыты новые химические элементы. Первыми из таких элементов были цезий и рубидий; они получили название по цвету линий своего спектра[9].

В 1868 году в спектре Солнца были обнаружены линии, не соответствующие ни одному из известных химических элементов. Новый элемент был назван *гелием* (от греческого *гелиос* — солнце). Впоследствии гелий был обнаружен в атмосфере Земли.

Вообще, спектральный анализ излучения Солнца и звёзд показал, что все входящие в их состав входят элементы имеются и на Земле. Таким образом, оказалось, что все объекты Вселенной собраны из одного и того же «набора кирпичиков».

[9] В спектре цезия наиболее выражены две линии небесно-синего цвета, по-латыни называемого *caesius*. Рубидий же даёт две характерные линии рубинового цвета.

6.5 Строение атома

Атомы каждого химического элемента имеют строго индивидуальный линейчатый спектр, присущий только данному элементу и не меняющийся от опыта к опыту. Как это можно объяснить? Как вывести формулу, дающую весь набор частот атомного спектра? Чтобы сделать это, нужно узнать, как устроен атом.

6.5.1 Модель Томсона

Первую модель строения атома придумал английский физик Джозеф Джон Томсон (удостоенный Нобелевской премии за открытие электрона). В конечном счёте она оказалась неверной, но сыграла важную роль, будучи стимулом последующих экспериментальных исследований Резерфорда. Физики называли модель Томсона «пудинг с изюмом».

Согласно Томсону атом представляет собой шар размером порядка 10^{-8} см. По этому шару некоторым образом распределён положительный заряд, а внутри шара, подобно изюминкам, находятся электроны (рис. 6.9).

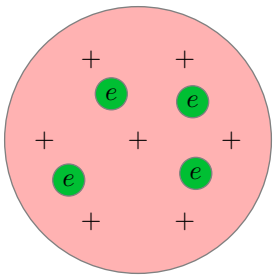


Рис. 6.9. Модель атома Томсона

Суммарный заряд электронов в точности равен положительному заряду шара, поэтому атом в целом электрически нейтрален.

Излучение атомов объясняется колебаниями электронов около положений равновесия (как вы помните, любой ускоренно движущийся заряд излучает электромагнитные волны). Однако вся совокупность экспериментальных данных по атомным спектрам не укладывалась в модель Томсона. Например, были подобраны формулы, хорошо описывающие спектр излучения атома водорода, но эти формулы из модели Томсона никак не следовали.

6.5.2 Опыты Резерфорда

Верна ли модель Томсона? Как в действительности распределены положительные и отрицательные заряды внутри атома? Чтобы ответить на эти вопросы, нужен был эксперимент, позволяющий проникнуть внутрь атома. Ученик Томсона, знаменитый английский физик Эрнест Резерфорд предложил с этой целью бомбардировать атом высокоэнергетичными α-частицами и смотреть, как они будут отклоняться положительным зарядом атома.

Что такое α-частицы? Потоки этих частиц — так называемые «альфа-лучи» — были обнаружены при радиоактивном распаде некоторых элементов (например, радия). В результате тщательных исследований, проведённых опять-таки Резерфордом, было установлено, что каждая α-частица имеет положительный заряд, равный по модулю удвоенному заряду электрона, и массу, превышающую массу электрона примерно в 8000 раз. То есть, α-частица оказалась полностью ионизованным (лишённым электронов) атомом гелия.

Резерфорд говорил об α-частицах как об ионах гелия; сейчас мы знаем, что это *ядра* гелия. Но в те времена об атомных ядрах ещё ничего не знали — о них Резерфорду лишь предстояло догадаться, глядя на результаты своих знаменитых опытов!

Энергия α-частиц очень велика — достаточно сказать, что скорость их вылета из радиоактивного образца составляет примерно $1/15$ скорости света. Поэтому интересно было выяснить, на какие углы будут отклоняться столь мощные «снаряды» при рассеянии на отдельных атомах, а точнее — на их положительных зарядах.

Пучок α-частиц направлялся на тончайшую золотую фольгу. Как гласит история, Резерфорд не сомневался в том, что углы отклонения должны быть весьма малы: имея столь огромную энергию, α-частицы должны проходить сквозь фольгу как нож сквозь масло. Только «для очистки совести», на всякий случай, он попросил учеников посмотреть, не возникает ли рассеяния α-частиц на большие углы.

Каково же было всеобщее удивление, когда такие частицы обнаружились! Действительно, как и следовало ожидать, подавляющая доля α-частиц отклонялась несущественно. Но совсем небольшая их часть (примерно одна частица из нескольких тысяч) отклонялась на угол, больший 90° (рис. 6.10).

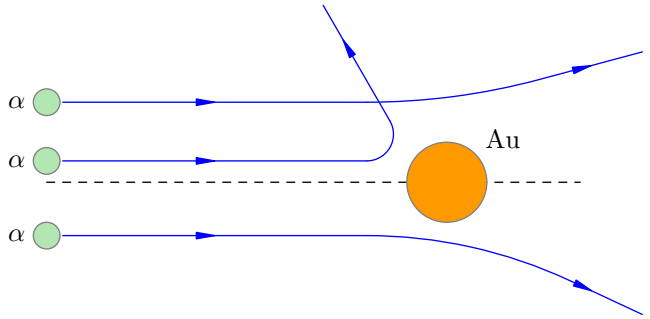


Рис. 6.10. Рассеяние α-частиц на атомах

Эти отклонения казались совершенно невероятными. По словам Резерфорда, *дело выглядело так, словно артиллерийский снаряд налетел на кусок бумаги и от удара повернул назад.*

А «бумагой» в образном сравнении Резерфорда служил атом, устроенный согласно модели Томсона. Действительно, допустим, что положительный заряд атома q «размазан» по всему атому, то есть шару радиусом $R \sim 10^{-8}$ см. Этот положительный заряд создаёт электрическое поле, тормозящее и отклоняющее α-частицы. Вблизи атома потенциал данного поля:

$$\varphi \sim \frac{kq}{R}\,.$$

Расчёты, однако, показывают, что такое поле оказывается слишком слабым — его тормозящего действия никак не хватит для того, чтобы остановить α-частицу и отбросить её назад!

Таким образом, наличие α-частиц, отброшенных фольгой, опровергло модель Томсона. Что же было предложено взамен?

6.5.3 Планетарная модель атома

Чтобы отбросить α-частицу, положительный заряд атома должен создавать куда более сильное электрическое поле, чем то, которое получается в модели Томсона. А чтобы создать такое поле, положительный заряд должен быть сосредоточен в области, гораздо меньшей размера атома.

Размер этой области можно вычислить. Если положительный заряд q занимает область размером r, то вблизи заряда создаётся электрическое поле с потенциалом

$$\varphi \sim \frac{kq}{r}\,.$$

Зная кинетическую энергию α-частицы, можно найти величину тормозящего потенциала φ, а затем и размер r положительно заряженной области. Вычисления, проведённые Резерфордом,

дали следующий результат:

$$r \sim 10^{-13}\,\text{см}.$$

Эта величина на пять порядков (в сто тысяч раз!) меньше размера атома. Так на смену модели Томсона пришла *планетарная модель атома* (рис. 6.11).

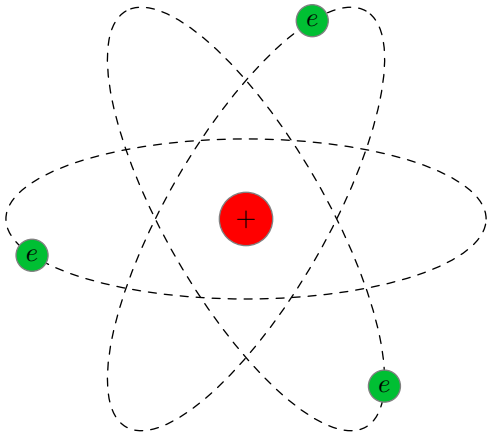


Рис. 6.11. Планетарная модель атома

В центре атома находится крошечное положительно заряженное *ядро*, вокруг которого, словно планеты вокруг Солнца, движутся электроны. Между ядром и электронами действуют силы кулоновского притяжения, но упасть на ядро электроны не могут за счёт своего движения — точно так же, как и планеты не падают на Солнце, хоть и притягиваются к нему.

Заряд ядра по модулю равен суммарному заряду электронов, так что атом в целом электрически нейтрален. Однако электроны могут быть выбиты из своих орбит и покинуть атом — тогда атом превращается в положительно заряженный ион.

Масса электронов составляет очень малую часть общей массы атома. Например, в атоме водорода всего один электрон, и его масса в 1836 раз меньше массы ядра. Следовательно, почти вся масса атома сосредоточена в ядре — и это при том, что ядро в сто тысяч раз меньше самого атома.

Чтобы лучше почувствовать соотношение масштабов атома и ядра, представьте себе, что атом стал размером с Останкинскую телебашню (500 м). Тогда ядро окажется горошиной размером 5 мм, лежащей у вас на ладони. И тем не менее, почти вся масса атома заключена в этой горошине!

Вот таким удивительным объектом оказался атом. Однако планетарная модель атома, объяснив результаты опытов Резерфорда по рассеянию α-частиц, стала лишь первым шагом на пути к пониманию внутриатомных процессов. А именно, планетарная модель приводила к одному серьёзному противоречию, и преодоление этого противоречия Нильсом Бором положило начало физике атома. Читаем следующий раздел!

6.6 Атом Бора

Планетарная модель атома, успешно истолковав результаты опытов по рассеянию α-частиц, в свою очередь столкнулась с очень серьёзными трудностями.

Как мы знаем, любой заряд, движущийся с ускорением, излучает электромагнитные волны. Это — неоспоримый факт классической электродинамики Максвелла, подтверждаемый многочисленными наблюдениями.

Нам также хорошо известно, что электромагнитные волны несут энергию. Стало быть, ускоренно движущийся заряд, излучая, теряет энергию, которая этим излучением уносится.

А теперь давайте возьмём произвольный электрон в планетарной модели. Он двигается вокруг ядра, так что направление его скорости постоянно меняется. Следовательно, электрон всё время имеет некоторое ускорение (например, при равномерном движении по окружности это будет центростремительное ускорение), и поэтому должен непрерывно излучать электромагнитные волны. Расходуя свою энергию на излучение, электрон будет постепенно приближаться к ядру; в конце концов, исчерпав запас своей энергии полностью, электрон упадёт на ядро.

Если исходить из того, что механика Ньютона и электродинамика Максвелла работают внутри атома, и провести соответствующие вычисления, то получается весьма озадачивающий результат: расход энергии электрона на излучение (с последующим падением электрона на ядро) потребует совсем малого времени — порядка 10^{-8} секунды. За это время атом должен полностью «коллапсировать» и прекратить своё существование.

Таким образом, классическая физика предрекает неустойчивость атомов, устроенных согласно планетарной модели. Этот вывод находится в глубоком противоречии с опытом: ведь на самом деле ничего такого не наблюдается. Предметы нашего мира вполне устойчивы и не коллапсируют на глазах! Атом может сколь угодно долго пребывать в невозбуждённом состоянии, не излучая при этом электромагнитные волны.

6.6.1 Постулаты Бора

Оставалось признать, что внутри атомов перестают действовать известные законы классической физики. Микромир подчиняется совсем другим законам.

Первый прорыв в познании законов микромира принадлежит великому датскому физику Нильсу Бору. Он предложил три постулата[10], резко расходящиеся с механикой и электродинамикой, но тем не менее позволяющих правильно описать простейший из атомов — атом водорода.

Классическая физика хорошо описывает *непрерывные* процессы — движение материальной точки, изменение состояния идеального газа, распространение электромагнитных волн... Энергия объекта, подчиняющегося механике или электродинамике, в принципе может принимать *любые* значения. Однако линейчатые спектры указывают на *дискретность* процессов, происходящих внутри атомов. Эта дискретность должна фигурировать в законах новой теории.

Первый постулат Бора. Всякий атом (и вообще, всякая атомная система) может находиться не во всех состояниях с любым, наперёд заданным значением энергии. Возможен лишь дискретный набор избранных состояний, называемых стационарными, в которых энергия атома принимает значения $E_1, E_2, \ldots, E_n, \ldots$ Находясь в стационарном состоянии, атом не излучает электромагнитные волны.

Как видим, первый постулат Бора вопиющим образом противоречит классической физике: налагается запрет на любые значения энергии, кроме избранного прерывистого набора, и признаётся, что электроны, вроде бы движущиеся ускоренно, на самом деле не излучают.

[10]Напомним, что *постулат* — это базовое, первичное утверждение физической теории, обобщающее опытные данные. Постулат не следует из каких-то других утверждений, он просто констатирует факт: так, мол, ведёт себя природа.

Выглядит фантастически, не правда ли? Однако в том же 1913 году, когда Бор предложил свои постулаты, существование стационарных состояний было подтверждено экспериментально — в специально поставленном опыте немецких физиков Франка и Герца. Таким образом, стационарные состояния — это не выдумка, а объективная реальность.

Значения разрешённого набора $E_1, E_2, \ldots, E_n, \ldots$ называются *уровнями энергии* атома. Что происходит при переходе с одного уровня энергии на другой?

Второй постулат Бора. Если атом переходит из стационарного состояния с большей энергией E_n в стационарное состояние с меньшей энергией E_k, то разность этих энергий может высвободиться в виде излучения. В таком случае излучается фотон с энергией

$$h\nu = E_n - E_k. \tag{6.13}$$

Эта же формула работает и при поглощении света: в результате столкновения с фотоном атом переходит из состояния E_k в состояние с большей энергией E_n, а фотон при этом исчезает.

Для примера на рис. 6.12 показано излучение фотона при переходе атома с энергетического уровня E_3 на уровень E_1. Переход заключается в том, что электрон «соскакивает» с одной орбиты на другую, расположенную ближе к ядру.

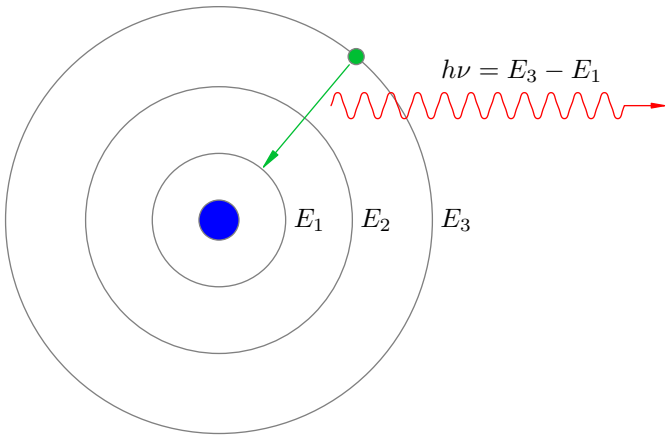


Рис. 6.12. Излучение фотона атомом

Формула (6.13) даёт качественное представление о том, почему атомные спектры испускания и поглощения являются линейчатыми.

В самом деле, атом может излучать волны лишь тех частот, которые соответствуют разностям значений энергии разрешённого дискретного набора $E_1, E_2, \ldots, E_n, \ldots$; соответственно, набор этих частот также получается дискретным. Вот почему спектр излучения атомов состоит из отдельно расположенных резких ярких линий.

Вместе с тем, атом может поглотить не любой фотон, а только тот, энергия $h\nu$ которого в точности равна разности $E_n - E_k$ каких-то двух разрешённых значений энергии E_n и E_k. Переходя в состояние с более высокой энергией E_n, атомы поглощают ровно те самые фотоны, которые способны излучить при обратном переходе в исходное состояние E_k. Попросту говоря, атомы забирают из непрерывного спектра те линии, которые сами же и излучают; вот почему тёмные линии спектра поглощения холодного атомарного газа находятся как раз в тех местах, где расположены яркие линии спектра испускания этого же газа в нагретом состоянии.

Качественного объяснения характера атомных спектров, однако, недостаточно. Хотелось бы иметь теорию, позволяющую *вычислить* частоты наблюдаемых спектров. Бору удалось это сделать в самом простом случае — для атома водорода.

6.6.2 Атом водорода

Атом водорода состоит из ядра с зарядом $+e$, которое называется *протоном*, и одного электрона с зарядом $-e$ (через e обозначена абсолютная величина заряда электрона). При построении своей теории атома водорода Бор сделал три дополнительных предположения.

1. Прежде всего, мы ограничиваемся рассмотрением только *круговых* орбит электрона[11]. Таким образом, электрон движется вокруг протона по окружности радиуса r с постоянной по модулю скоростью v (рис. 6.13).

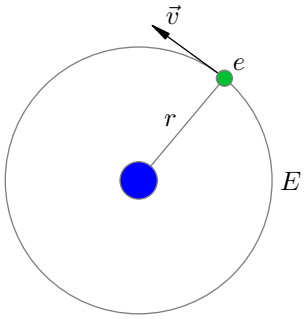


Рис. 6.13. Модель атома водорода

2. Величина mvr, равная произведению импульса электрона mv на радиус орбиты r, называется *моментом импульса* электрона. В каких единицах измеряется момент импульса? Смотрим:
$$[mvr] = \text{кг} \cdot \frac{\text{м}}{\text{с}} \cdot \text{м} = \left(\text{кг} \cdot \frac{\text{м}}{\text{с}^2}\right) \cdot \text{м} \cdot \text{с} = \text{Н} \cdot \text{м} \cdot \text{с} = \text{Дж} \cdot \text{с}.$$
Это в точности размерность постоянной Планка! Именно здесь Бор увидел появление дискретности, необходимой для квантового описания атома водорода.

   Правило квантования (третий постулат Бора). Момент импульса электрона может принимать лишь дискретный набор значений, кратных «перечёркнутой» постоянной Планка:
$$mvr = n\hbar \quad (n = 1, 2, 3, \ldots). \tag{6.14}$$

3. Выше мы говорили, что классическая физика перестаёт работать внутри атома. Так оно в действительности и есть, но вопреки этому мы предполагаем, что электрон притягивается к протону с силой, вычисляемой по закону Кулона, а движение электрона подчиняется второму закону Ньютона:
$$\frac{mv^2}{r} = \frac{ke^2}{r^2}\,. \tag{6.15}$$

Эти три предположения позволяют довольно просто получить формулы для уровней энергии атома водорода. Переписываем соотношение (6.15) в виде:
$$mv^2 = \frac{ke^2}{r}\,. \tag{6.16}$$

Из правила квантования (6.14) выражаем v:
$$v = \frac{n\hbar}{mr}\,,$$

[11] Впоследствии теория Бора была распространена на общий случай эллиптических орбит.

и подставляем это в (6.16):

$$\frac{n^2\hbar^2}{mr^2} = \frac{ke^2}{r}\,.$$

Отсюда получаем формулу для допустимых радиусов орбит электрона:

$$r_n = \frac{\hbar^2 n^2}{kme^2}\,. \tag{6.17}$$

Теперь перейдём к нахождению энергии электрона. Потенциальная энергия кулоновского взаимодействия электрона с ядром равна:

$$W = \frac{k(+e)(-e)}{r} = -\frac{ke^2}{r}\,.$$

(Она отрицательна, так как отсчитывается от бесконечно удалённой точки, в которой достигает максимального — а именно, нулевого — значения.)

Полная энергия E электрона равна сумме его кинетической и потенциальной энергий:

$$E = \frac{mv^2}{2} - \frac{ke^2}{r}\,.$$

Вместо mv^2 подставим правую часть выражения (6.16):

$$E = \frac{ke^2}{2r} - \frac{ke^2}{r} = -\frac{ke^2}{2r}\,. \tag{6.18}$$

Полная энергия, как видим, отрицательна. Если на радиус орбиты никаких ограничений не накладывается, как это имеет место в классической физике, то энергия может принимать любые по модулю значения. Но согласно (6.17) существует лишь дискретный набор возможных значений радиуса; подставляя их в (6.18), получаем соответствующий набор допустимых значений энергии атома водорода:

$$E_n = -\frac{ke^2}{2r_n} = -\frac{k^2me^4}{2\hbar^2n^2} \quad (n = 1, 2, 3, \ldots). \tag{6.19}$$

Основное состояние атома водорода — это состояние с наименьшей энергией E_1. В основном состоянии атом может находиться неограниченно долго. Вычисление даёт:

$$E_1 = -2{,}18 \cdot 10^{-18}\,\text{Дж} = -13{,}6\,\text{эВ}.$$

Мы видим, что если атом находится в основном состоянии, то для выбивания электрона нужно сообщить атому энергию, равную как минимум 13,6 эВ. Эта величина носит название *энергии ионизации* атома водорода.

По формуле (6.17) легко вычислить радиус орбиты основного состояния:

$$r_1 = \frac{\hbar^2}{kme^2} = 5 \cdot 10^{-9}\,\text{см}.$$

То есть, диаметр атома оказывается равным как раз 10^{-8} см — величине, известной из опыта. Таким образом, теория Бора впервые смогла объяснить размер атома!

Кроме того, в рамках теории Бора удаётся получить формулы для вычисления частот (или длин волн) спектра атома водорода. Так, согласно второму постулату Бора и формуле (6.19) имеем:

$$\nu = \frac{E_n - E_k}{h} = \frac{k^2me^4}{4\pi\hbar^3}\left(\frac{1}{k^2} - \frac{1}{n^2}\right). \tag{6.20}$$

На практике чаще имеют дело с длинами волн. Учитывая, что $\nu = c/\lambda$, формулу (6.20) можно переписать так:

$$\frac{1}{\lambda} = \frac{k^2 m e^4}{4\pi\hbar^3 c}\left(\frac{1}{k^2} - \frac{1}{n^2}\right) = R\left(\frac{1}{k^2} - \frac{1}{n^2}\right). \tag{6.21}$$

Константа $R = 1{,}1 \cdot 10^7\,\text{м}^{-1}$ называется *постоянной Ридберга*. Теория Бора даёт значение этой постоянной, очень хорошо согласующееся с экспериментом.

Длины волн спектра атома водорода образуют *серии*, характеризующиеся фиксированным значением k в формуле (6.21). Все длины волн данной серии излучаются при переходах на уровень E_k с вышележащих энергетических уровней E_n ($n = k+1, k+2, k+3, \ldots$).

Переходы в основное состояние:

$$E_2 \to E_1, \quad E_3 \to E_1, \quad E_4 \to E_1, \quad \ldots$$

образуют *серию Лаймана*. Длины волн этой серии описываются формулой (6.21) при $k = 1$:

$$\frac{1}{\lambda} = R\left(1 - \frac{1}{n^2}\right).$$

Линии серии Лаймана лежат в ультрафиолетовом диапазоне.

Переходы на второй уровень:

$$E_3 \to E_2, \quad E_4 \to E_2, \quad E_5 \to E_2, \quad \ldots$$

образуют *серию Бальмера*. Длины волн этой серии подчиняются формуле (6.21) при $k = 2$:

$$\frac{1}{\lambda} = R\left(\frac{1}{4} - \frac{1}{n^2}\right).$$

Первые четыре линии серии Бальмера лежат в видимом диапазоне (рис. 6.14)[12], остальные — в ультрафиолетовом.

Рис. 6.14. Видимый спектр атома водорода (серия Бальмера)

Переходы на третий уровень:

$$E_4 \to E_3, \quad E_5 \to E_3, \quad E_6 \to E_3, \quad \ldots$$

образуют *серию Пашена*. Длины волн этой серии описываются формулой (6.21) при $k = 3$:

$$\frac{1}{\lambda} = R\left(\frac{1}{9} - \frac{1}{n^2}\right).$$

Все линии серии Пашена лежат в инфракрасном диапазоне.

Имеются ещё три «именованных» серии: это *серия Брэккета* (переходы на уровень $k = 4$), *серия Пфунда* (переходы на уровень $k = 5$) и *серия Хэмпфри* (переходы на уровень $k = 6$). Все линии этих серий лежат в далёкой инфракрасной области.

[12] Изображение с сайта dic.academic.ru.

6.6.3 Достоинства и недостатки теории Бора

О достоинствах модели атома водорода, предложенной Бором, мы так или иначе уже сказали. Резюмируем их.

- Теория Бора продемонстрировала, что для описания атомных объектов принципиально недостаточно представлений классической физики. В микромире работают другие, совершенно новые законы.

  Для микромира характерно *квантование* — дискретность изменения величин, описывающих состояние объекта. В качестве меры квантования, как показала теория Бора, может выступать постоянная Планка $\hbar$, которая является универсальной константой и играет фундаментальную роль во всей физике микромира (а не только в явлениях излучения и поглощения света).

- Теория Бора впервые и совершенно точно указала на факт наличия стационарных энергетических состояний атома, образующих дискретный набор. Этот факт оказался общим свойством объектов микромира.

- В рамках модели Бора удалось получить формулы для вычисления частот спектра атома водорода и объяснить размер атома. Классическая физика была не в состоянии решить эти проблемы.

Однако теория Бора, разумеется, не могла претендовать на роль общей теории, описывающей микромир. Модель Бора обладала рядом существенных недостатков.

- Теория Бора *непоследовательна*. С одной стороны, она отвергает описание атома на основе классической физики, так как постулирует наличие стационарных состояний и правила квантования, непонятных с точки зрения механики и электродинамики. С другой стороны, классические законы — второй закон Ньютона и закон Кулона — используются для записи уравнения движения электрона по круговой орбите[13].

- Теория Бора не смогла дать адекватное описание самого простого после водорода атома гелия. Подавно не могло быть и речи о распространении теории Бора на более сложные атомы[14].

- Даже в самом атоме водорода теория Бора смогла описать не всё. Например, дав выражения для частот спектральных линий, модель Бора не объясняла различие в их интенсивностях. Кроме того, неясен оставался механизм образования молекулы водорода из двух атомов.

Несмотря на свои недостатки, теория Бора стала важнейшим этапом развития физики микромира. Полуклассическая-полуквантовая модель Бора послужила промежуточным звеном между классической физикой и последовательной *квантовой механикой*, построенной десятилетием позже — в 1920-х годах.

[13] Как шутливо замечал известный физик Брэгг, в теории Бора по понедельникам, средам и пятницам надо применять классические законы, а по вторникам, четвергам и субботам — квантовые.

[14] Как опять-таки шутили физики, атом Бора — это не атом бора, а атом водорода.

6.7 Лазер

Лазер (оптический квантовый генератор) — выдающееся достижение физики XX века. Уникальные свойства лазерного луча находят применение в самых разнообразных сферах — от мощных промышленных агрегатов до мелкой бытовой техники.

Слово *laser* образовано первыми буквами фразы *Light Amplification by Stimulated Emission of Radiation*. Дословный перевод таков: «усиление света с помощью вынужденной эмиссии излучения». В русском языке используется более короткий термин *вынужденное излучение* или *индуцированное излучение* (слово «эмиссия» опускается). Явление индуцированного излучения, которое мы сейчас обсудим, составляет физическую основу действия любого лазера.

6.7.1 Индуцированное излучение

Рассмотрим два энергетических состояния атома: основное состояние с энергией E_1 и возбуждённое состояние с энергией $E_2 > E_1$. Переходя из возбуждённого состояния в основное, атом излучает фотон с энергией $h\nu = E_2 - E_1$.

Если возбуждённый атом предоставлен сам себе, то он переходит в основное состояние самопроизвольно. Момент перехода предсказать невозможно, и потому излучение в этом случае называется *спонтанным*. Спонтанное излучение различных атомов является некогерентным, поскольку атомы излучают совершенно независимо друг от друга.

Оказывается, однако, что атом можно *заставить* излучить фотон, *вынудив* соскок электрона с верхнего уровня на нижний. Это произойдёт под действием внешнего электромагнитного поля, частота которого совпадает с частотой перехода $\nu = (E_2 - E_1)/h$. Эффект аналогичен явлению резонанса при вынужденных колебаниях; в качестве вынуждающей силы выступает при этом электромагнитное поле. Излучение атома, вызванное внешним электромагнитным полем, называется *вынужденным* или *индуцированным*[15] (рис. 6.15).

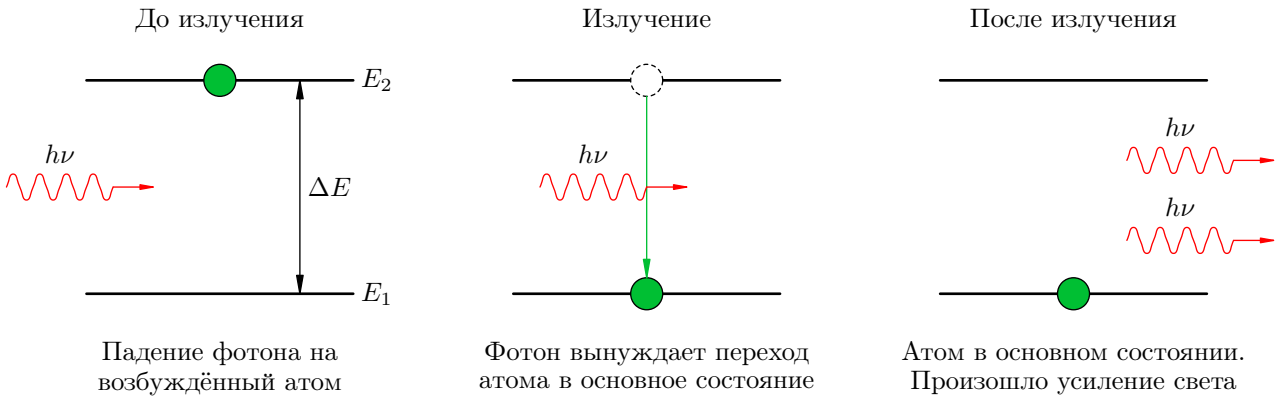


Рис. 6.15. Индуцированное излучение: $h\nu = \Delta E = E_2 - E_1$

Левая часть рисунка соответствует начальной ситуации, когда на возбуждённый атом, находящийся в состоянии с энергией E_2, падает фотон, энергия $h\nu$ которого в точности равна разности $\Delta E = E_2 - E_1$ энергий возбуждённого и основного состояний.

Средняя часть рисунка показывает вынужденный переход атома в основное состояние: под воздействием падающего фотона электрон соскакивает с верхнего уровня на нижний.

В правой части рисунка мы видим результат этого соскока. Падающий фотон никуда не делся, а атом, перейдя в основное состояние, излучил ещё один фотон, *тождественный падающему фотону*. Таким образом, вместо одного падающего фотона «на входе» возникло два фотона-близнеца «на выходе».

[15] Возможность индуцированного излучения света атомами предсказал А. Эйнштейн в 1916 году.

Итак, в результате индуцированного излучения атом испускает монохроматическую волну, совпадающую с падающей волной по частоте и фазе. Эти две волны, интерферируя, *усиливают* друг друга! Именно этот эффект усиления света используется в лазере.

6.7.2 Инверсная населённость

Для усиления света, проходящего через некоторую среду, одного только индуцированного излучения атомов среды недостаточно. Нужно ещё создать *инверсную населённость* энергетических уровней атомов. Давайте разберёмся, что это такое.

Когда свет идёт через вещество, возникают два противоположных эффекта.

1. Свет поглощается атомами среды, находящимися в основном состоянии. Эти атомы переходят в возбуждённое состояние.

2. Свет усиливается за счёт индуцированного излучения возбуждённых атомов. Испуская фотоны, эти атомы возвращаются в основное состояние.

Возникает вопрос: какой эффект перевешивает?

В обычных условиях большинство атомов вещества находятся в основном состоянии, меньшая их часть — в возбуждённом. В таком случае мы имеем *нормальную населённость* энергетических уровней (рис. 6.16, слева).

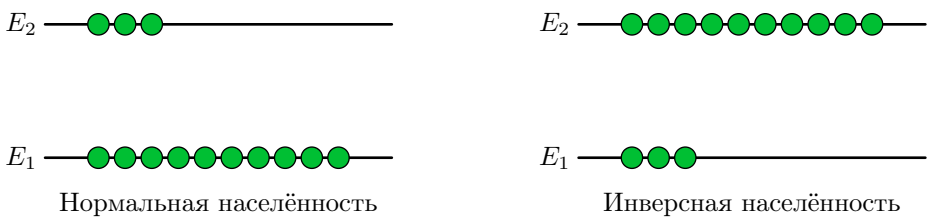


Рис. 6.16. Два типа населённости уровней

Если свет проходит через среду с нормальной населённостью, то бóльшая часть атомов поглощает свет; индуцированное излучение создаётся малым количеством атомов. В результате число фотонов уменьшается со временем, и световой поток ослабляется.

Можно, однако, создать условия, когда большинство атомов среды находится в возбуждённом состоянии. В таком случае населённость уровней называется *инверсной* (рис. 6.16, справа).

Если свет подходящей частоты распространяется в среде с инверсной населённостью, то фотоны чаще налетают на возбуждённые атомы, чем на невозбуждённые, так что индуцированное излучение фотонов преобладает над их поглощением. По мере прохождения света количество фотонов нарастает, и свет усиливается.

Как же создать в среде инверсную населённость? Нельзя ли просто направить на вещество мощный поток света и освещать до тех пор, пока больше половины атомов не перейдёт в возбуждённое состояние?

Добиться этого и в самом деле можно, но проблема состоит в том, что двумя энергетическими уровнями тут не обойдёшься. Действительно, поглощая фотоны, атомы будут переходить из основного состояния E_1 в возбуждённое состояние E_2; но те же самые фотоны падающего света будут вынуждать и обратные переходы с уровня E_2 на уровень E_1, сопровождающиеся индуцированным излучением. Следовательно, число возбуждённых атомов с энергией E_2 не может превысить числа атомов в основном состоянии E_1.

6.7.3 Трёхуровневая система рубина

Классической схемой создания в среде инверсной населённости является система из *трёх* энергетических уровней E_1, E_2 и E_3, в которой состояние E_1 является основным, состояние E_3 — «короткоживущим», а состояние E_2 — «долгоживущим». Такие уровни имеются в кристаллах рубина.

Работа трёхуровневой системы показана на рис. 6.17. В начальной ситуации большинство атомов находится в основном состоянии E_1.

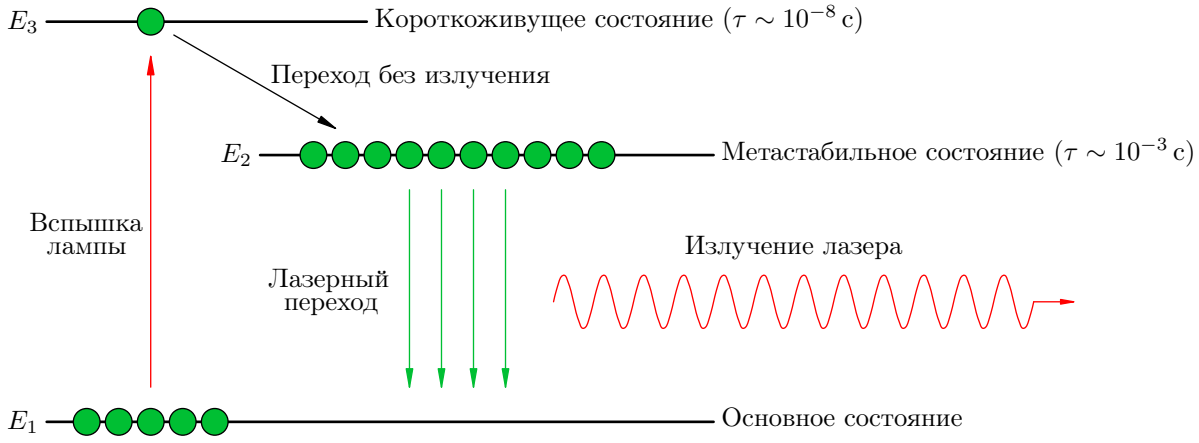


Рис. 6.17. Инверсная населённость в трёхуровневой системе

Мощная вспышка лампы переводит бóльшую часть атомов среды в возбуждённое состояние с энергией E_3. Но электроны атомов не задерживаются на этом энергетическом уровне. Состояние E_3 обладает малым *временем жизни* τ, равным примерно 10^{-8} с; по истечении данного промежутка времени электроны соскакивают с уровня E_3 вниз — но не назад на уровень E_1, а на промежуточный уровень E_2.

Переход $E_3 \to E_2$ не сопровождается излучением — энергия этого перехода передаётся тепловым колебаниям кристаллической решётки. Но самое главное заключается в том, что состояние E_2 имеет огромное по атомным масштабам время жизни — порядка 10^{-3} с. Эта величина, как видим, в 100000 раз больше времени жизни состояния E_3; по этой причине энергетическое состояние E_2 называется *метастабильным*.

Итак, в результате вспышки лампы атомы из основного состояния E_1 транзитом через уровень E_3 переводятся на метастабильный уровень E_2, и благодаря большому времени жизни этого уровня начинают на нём накапливаться. Если переходы $E_1 \to E_3$ совершаются достаточно быстро, то на уровне E_2 окажется больше половины атомов среды — возникает инверсная населённость данного уровня!

Ну а затем достаточно небольшого числа спонтанных переходов $E_2 \to E_1$, и излучённые фотоны вызовут лавину таких же, но теперь уже индуцированных *лазерных переходов* $E_2 \to E_1$. Число фотонов, порождённых лазерными переходами, стремительно нарастает и создаёт импульс лазерного излучения.

Таким образом, луч лазера генерируется в ходе *согласованного* «сброса» многих атомов с уровня E_2 на уровень E_1. Атомы при этом излучают *синхронно*, «в такт», и волны, излучённые атомами, идентичны друг другу по частоте и фазе. Излучение лазера, будучи результатом наложения таких идентичных волн, получается *когерентным* — в отличие, например, от некогерентного света электрической лампочки, порождаемого спонтанным излучением различных атомов в случайные моменты времени.

6.7.4 Устройство лазера

В общих чертах схема устройства лазера выглядит следующим образом (рис. 6.18)[16].

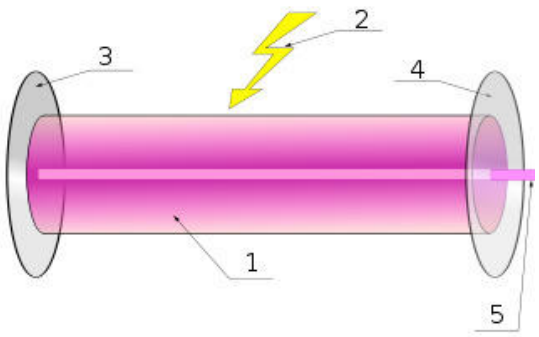


Рис. 6.18. Устройство лазера

В атомах *активной среды* 1 создаётся инверсная населённость в результате *вспышки* 2. Активной средой может служить, например, кристалл рубина или какое-либо другое подходящее вещество.

На торцах активной среды стоят *непрозрачное зеркало* 3 и *полупрозрачное зеркало* 4. Эти зеркала обеспечивают, как говорят, *обратную связь*. Они нужны для более эффективного усиления *лазерного луча* 5.

А именно, происходит вот что. После вспышки, как мы уже говорили, создаётся инверсная населённость, и достаточно нескольких спонтанно излучённых фотонов при лазерном переходе $E_2 \to E_1$, чтобы началось их лавинообразное размножение в результате индуцированного излучения. Но при отсутствии зеркал эта лавина не успеет затронуть весьма большое количество возбуждённых атомов — возникший импульс уже покинет пределы активной среды.

Чтобы этого не произошло, световой импульс разумно завернуть назад и погонять некоторое время внутри активной среды, заставляя высвечиваться всё большее и большее количество атомов — и тем самым всё более усиливая лазерный луч. Вот для этого и нужны зеркала на торцах. Одно зеркало, естественно, должно быть полупрозрачным и частично выпускать излучение наружу — мы ведь хотим использовать луч лазера для каких-то целей :-)

Зеркала выполняют ещё одну важную функцию: благодаря им лазерный луч имеет *очень малую расходимость*. Дело в том, что лучи, отклоняющиеся от оси цилиндра, рано или поздно выйдут через боковую цилиндрическую поверхность, а лучи, идущие параллельно оси, будут циркулировать туда-сюда сколько угодно и максимально усиливаться. Вот почему лазерный луч даёт маленькое световое пятно даже на удалённых предметах.

[16]Изображение с сайта en.wikipedia.org.

6.8 Строение ядра

После опытов Резерфорда, прояснивших устройство атома, возник естественный вопрос: из чего состоит атомное ядро? Ответа пришлось ждать двадцать лет — до открытия нейтрона.

Ядро самого простого атома водорода, как вы помните, было названо *протоном*. Протон имеет положительный заряд $e = 1{,}6 \cdot 10^{-19}$ Кл (равный по модулю заряду электрона) и массу $m_p = 1{,}6726 \cdot 10^{-27}$ кг. Масса протона примерно в 1836 раз больше массы электрона.

Нейтрон был открыт в 1932 году английским физиком Джеймсом Чедвиком. Масса нейтрона $m_n = 1{,}6749 \cdot 10^{-27}$ кг оказалась очень близка к массе протона. Однако, в отличие от протона, нейтрон не имеет электрического заряда.

Открытие нейтрона послужило ключом к пониманию устройства атомного ядра.

6.8.1 Нуклонная модель ядра

Сразу после открытия нейтрона несколько физиков одновременно высказали идею *протонно-нейтронной*, или *нуклонной*, модели ядра. Согласно этой модели *ядро состоит из протонов и нейтронов*. Будучи «кирпичиками», из которых строится ядро, протоны и нейтроны получили общее название *нуклонов*[17].

Модель атомного ядра показана[18] на рис. 6.19. Красным цветом условно изображены протоны, чёрным — нейтроны.

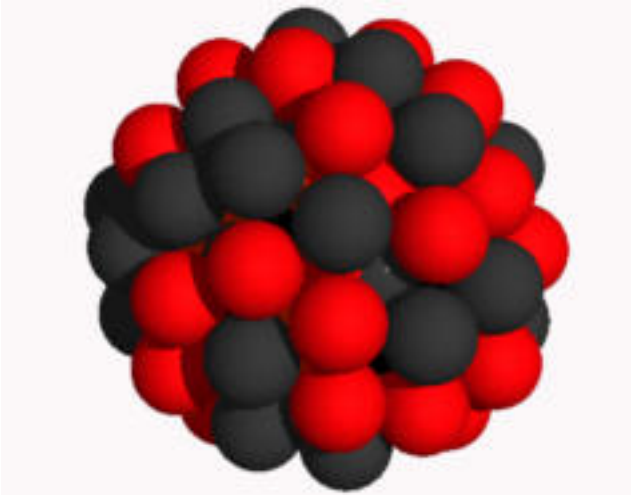


Рис. 6.19. Модель ядра атома

Число протонов в ядре называется *зарядовым числом* и обозначается Z. *Заряд ядра*, следовательно, равен Ze. Поскольку атом в целом электрически нейтрален, величина Z совпадает с числом электронов в атоме. Зарядовое число, таким образом, есть не что иное, как порядковый номер химического элемента в таблице Менделеева.

Общее число нуклонов в ядре называется *массовым числом* и обозначается A. Число нейтронов в ядре тогда будет равно $A - Z$.

Запись ${}^{A}_{Z}\mathrm{X}$ означает, что в ядре элемента X содержится A нуклонов, из которых Z являются протонами. Например, ядро алюминия ${}^{27}_{13}\mathrm{Al}$ состоит из 27 нуклонов, а именно из 13 протонов и 14 нейтронов. Ядро гелия ${}^{4}_{2}\mathrm{He}$ — так называемая α-частица — состоит из двух протонов и двух нейтронов.

[17] От лат. *nucleus* — ядро.

[18] Изображение с сайта alternativephysics.org.

6.8.2 Изотопы

Что будет, если изменить число нейтронов ядре? Какие-то свойства вещества в результате должны поменяться — например, плотность. Однако все химические свойства при этом останутся прежними — ведь за них отвечает зарядовое число Z, а оно-то не менялось!

Изотопы — это разновидности одного и того же химического элемента, различающиеся числом нейтронов в ядре.

Например, у водорода три изотопа: обычный ${}^{1}_{1}\mathrm{H}$, *дейтерий* ${}^{2}_{1}\mathrm{H}$ и *тритий* ${}^{3}_{1}\mathrm{H}$. А химический элемент уран имеет 26 изотопов! В природе наиболее распространён уран ${}^{238}_{92}\mathrm{U}$, а в атомной энергетике и ядерном оружии используется уран ${}^{235}_{92}\mathrm{U}$.

Изотопы совершенно идентичны в отношении химических свойств, и их невозможно разделить никакими химическими методами. Оказывается, почти любой элемент таблицы Менделеева представляет собой смесь изотопов в различных пропорциях — вот почему атомные массы химических элементов не равны целым числам. Как правило, атомная масса всё же достаточно близка к целому числу, поскольку в природе доминирует изотоп именно с такой атомной массой (например, в природном уране доля изотопа ${}^{238}_{92}\mathrm{U}$ составляет 93%; соответственно, в таблице Менделеева мы видим атомную массу урана, равную 238,03). Но бывают и исключения: так, атомная масса хлора равна 35,5.

Изотопы могут различаться также своими радиоактивными свойствами: у одного и того же химического элемента могут быть как стабильные изотопы, так и подверженные радиоактивному распаду (например, углерод ${}^{12}_{6}\mathrm{C}$ стабилен, а изотоп ${}^{14}_{6}\mathrm{C}$ — радиоактивен). Собственно, именно это наблюдение — что вроде бы одно и то же вещество бывает то радиоактивным, то нет — и навело в своё время на мысль о существовании изотопов. Радиоактивность будет темой следующего раздела.

6.9 Радиоактивность

Явление радиоактивности обнаружил французский физик Анри Антуан Беккерель, и произошло это совершенно случайно.

В начале 1896 года всё научное сообщество было охвачено интересом к недавно открытым всепроникающим рентгеновским лучам. Беккерель решил выяснить, не появляются ли рентгеновские лучи при освещении солнечным светом некоторых минералов, и выбрал для своих экспериментов весьма редкую соль урана.

Опыт Беккереля был чрезвычайно прост. Кристаллы соли выставлялись на солнце и лежали при этом на фотопластинке. Разумеется, фотопластинка заворачивалась в чёрную бумагу, чтобы её не засветил солнечный свет. Но чёрная бумага — не помеха рентгеновским лучам, и если они действительно возникают, то засветят фотопластинку.

Итак, Беккерель положил завёрнутую фотопластинку с насыпанной поверх урановой солью на солнечный свет, подержал несколько часов и затем проявил фотопластинку. Ожидания подтвердились! После проявления на фотопластинке проступили очертания кристаллов соли урана.

Полагая, что и впрямь обнаружись рентгеновские лучи, испускаемые урановой солью под действием солнечного света, Беккерель доложил об этом на заседании Французской академии наук. Доклад вызвал большой интерес, и было решено, что на следующем заседании, то есть через неделю, Беккерель расскажет о результатах новых опытов.

А погода тем временем испортилась, и солнце на всю неделю скрылось за облаками. Медный крест, покрытый урановой солью и приготовленный для опытов, в ожидании солнца несколько дней пролежал в ящике письменного стола — поверх фотопластинки, завёрнутой в чёрную бумагу.

Накануне нового доклада облачность так и не рассеялась, и докладывать Беккерелю было нечего. Однако отчаяние и удачу порой разделяет лишь один шаг. Неизвестно почему, но Беккерель решил проявить фотопластинку, лежавшую в столе. Каково же было его удивление, когда он увидел проступившие на ней почернения в виде отчётливой тени креста!

Таким образом, солнце оказалось совершенно ни при чём. Было обнаружено новое явление природы: урановая соль без каких-либо внешних факторов, сама по себе испускает некоторое излучение, пронизывающее чёрную бумагу.

На следующий день Беккерель рассказал об этом на заседании Французской академии и затем приступил к интенсивным исследованиям. В ходе своих экспериментов он обнаружил следующие черты нового явления.

- Новые лучи могут проникать сквозь предметы и ионизировать воздух.
- Засвечивают фотопластинку только те вещества, которые содержат уран.
- Интенсивность излучения зависит только от количества урана в веществе. Само химическое соединение при этом роли не играет. *Максимально интенсивным является излучение чистого урана.*

Новое явление было впоследствии названо *радиоактивностью*. Из опытов Беккереля следовало, что радиоактивность есть свойство химического элемента урана самого по себе — то есть свойство, которым обладают *атомы* урана.

Уран оказался не единственным радиоактивным элементом. Мария Склодовская-Кюри спустя два года после открытия Беккереля обнаружила аналогичное излучение *тория*. Вместе с мужем, Пьером Кюри, они открыли новый радиоактивный химический элемент — *полоний*. Наконец, вручную переработав 11 тонн руды, Мария Склодовская-Кюри получила маленькую капельку чистого *радия*, который излучал в три миллиона раз интенсивнее урана.

6.9.1 Виды радиоактивных излучений

Каков состав радиоактивного излучения? Оказалось, что радиоактивные вещества испускают три типа лучей, различающихся по своим физическим свойствам.

Эти три компоненты обнаруживаются в результате пропускания радиоактивного излучения солей урана через сильное магнитное поле (рис. 6.20).

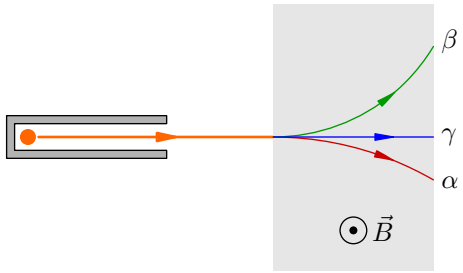


Рис. 6.20. Виды радиоактивных излучений

А именно, излучение радиоактивного препарата, находящегося внутри свинцового контейнера с узким каналом, направляется на фотопластинку. В отсутствии магнитного поля на фотопластинке наблюдается одно тёмное пятно. Но если пропустить излучение сквозь область магнитного поля, то пятен становится *три* — одно на прежнем месте и два по бокам от него на разных расстояниях. Это означает, что радиоактивное излучение в магнитном поле распалось на три существенно различные части.

То, что две компоненты отклонились в разные стороны, означает, что они являются соответственно потоками положительных и отрицательных зарядов. Третья компонента, не отклоняющаяся магнитным полем, электрического заряда не несёт.

Положительно заряженной компоненте была присвоена буква α; её называли *α-излучением*, *α-лучами* или потоком *α-частиц*. Альфа-лучи достаточно слабо отклонялись магнитным полем. Тщательные исследования Резерфорда показали, что α-частицы — это полностью ионизованные атомы гелия, то есть *ядра гелия*.

Отрицательно заряженная компонента была названа *β-излучением* (или *β-лучами*). Они отклонялись магнитным полем значительно сильнее, чем α-частицы. Бета-лучи оказались потоком *электронов*, мчащихся со скоростями, близкими к скорости света.

Нейтральная компонента получила название *γ-излучения* (или *γ-лучей*). Гамма-лучи оказались электромагнитными волнами чрезвычайно высокой частоты — выше, чем у рентгеновского излучения[19]. Соответственно, проникающая способность гамма-лучей также больше, чем у рентгеновских лучей.

Среди трёх компонент радиоактивного излучения наибольшей проникающей способностью также обладают гамма-лучи — они могут пробиться сквозь слой свинца толщиной в несколько сантиметров. Сильнее поглощаются веществом бета-лучи: тут хватит нескольких миллиметров свинца, чтобы поглотить их полностью. Слабее всего проникают сквозь вещество α-частицы: они не могут, например, пройти через лист бумаги.

6.9.2 Радиоактивные превращения

Многочисленные эксперименты с радиоактивными веществами показали, что *радиоактивность сопровождается изменениями атомов, и в результате этих изменений одни химические элементы превращаются в другие.*

[19] Электромагнитная природа гамма-излучения была установлена экспериментально: обнаружилась дифракция гамма-лучей на кристаллических решётках. Эти же опыты позволили измерить и длину волны гамма-излучения.

Положение химического элемента в таблице Менделеева определяется числом электронов в нейтральном атоме, или, что то же самое — зарядом ядра атома. Поэтому превращения химических элементов означают, что *в результате радиоактивных процессов изменения претерпевают атомные ядра*.

Ядра атомов радиоактивных элементов являются *нестабильными*. Каждое такое ядро в некоторый момент распадается, поэтому явление радиоактивности называют ещё *радиоактивным распадом*.

В процессе радиоактивного распада исходное вещество постепенно исчезает. Новые вещества, являющиеся продуктами распада, также могут быть нестабильными и распадаться дальше. Наблюдаются целые цепочки радиоактивных распадов — вплоть до образования стабильных элементов.

Самой известной такой цепочкой является *радиоактивное семейство урана*. Начинается эта цепочка с *альфа-распада* ядра ${}^{238}_{92}\mathrm{U}$, в результате которого образуется ядро тория ${}^{234}_{90}\mathrm{Th}$ и вылетает α-частица:

$$ {}^{238}_{92}\mathrm{U} \to {}^{234}_{90}\mathrm{Th} + {}^{4}_{2}\mathrm{He}. \tag{6.22} $$

Затем родившееся ядро тория испытывает *бета-распад*, испуская электрон и превращаясь в ядро протактиния ${}^{234}_{91}\mathrm{Pa}$:

$$ {}^{234}_{90}\mathrm{Th} \to {}^{234}_{91}\mathrm{Pa} + {}^{0}_{-1}e. \tag{6.23} $$

Обратите внимание, что электрону приписывается зарядовое число -1 (так как заряд электрона равен $-e$) и массовое число 0 (так как электрон не содержит нуклонов).

В обеих формулах (6.22) и (6.23) мы наблюдаем два важных момента.

- *Сумма массовых чисел продуктов распада равна массовому числу исходного ядра.* Этот баланс массовых чисел отражает неизменность общего числа нуклонов до и после распада.
- *Сумма зарядовых чисел продуктов распада равна зарядовому числу исходного ядра.* Этот факт служит одним из многочисленных экспериментальных подтверждений закона сохранения заряда.

Поскольку α-частица уносит заряд $+2e$, а электрон уносит заряд $-e$, то возникает следующая закономерность превращения химических элементов при α- и β-распадах.

Правило смещения. После α-распада элемент смещается на две клетки назад, то есть к началу периодической системы. После β-распада элемент смещается на одну клетку вперёд, то есть к концу периодической системы.

Общие формулы, выражающие правило смещения при альфа- и бета-распадах, выглядят следующим образом:

$$ {}^{A}_{Z}\mathrm{X} \to {}^{A-4}_{Z-2}\mathrm{Y} + {}^{4}_{2}\mathrm{He}, $$
$$ {}^{A}_{Z}\mathrm{X} \to {}^{A}_{Z+1}\mathrm{Y} + {}^{0}_{-1}e. $$

Формулы (6.22) и (6.23) — это самое начало радиоактивного семейства урана. Всего в этой цепочке происходит восемь α-распадов и шесть β-распадов (при каждом β-распаде вдобавок излучается γ-квант), пока в самом конце цепочки не образуется стабильное ядро свинца ${}^{206}_{82}\mathrm{Pb}$.

Излучение всех элементов радиоактивного семейства урана как раз и засветило фотопластинку Беккереля, и именно эта смесь излучений была впервые разложена на компоненты в магнитном поле (рис. 6.20).

6.9.3 Закон радиоактивного распада

Нестабильное ядро распадается *самопроизвольно* (или, как ещё говорят, *спонтанно*). Происходит это в случайный момент времени, так что невозможно предсказать, когда именно распадётся каждое конкретное ядро. Тем не менее, ядра каждого элемента обладают определённым *средним временем жизни*, характерным для данного элемента.

А именно, опыт показывает, что *распад радиоактивного элемента происходит со строго определённой, присущей именно этому элементу скоростью*. Скорость распада у разных элементов различна; она является такой же неотъемлемой характеристикой радиоактивного элемента, как зарядовое или массовое число. Вне зависимости от условий опыта можно точно сказать, спустя какой промежуток времени интенсивность излучения данного элемента уменьшится, например, в два раза.

Период полураспада — это время, в течение которого распадается половина имеющихся радиоактивных атомов. Период полураспада как раз и является количественной характеристикой скорости радиоактивного распада.

Величина периода полураспада может быть очень разной. Например, период полураспада урана $^{238}_{92}\mathrm{U}$ равен 4,5 млрд. лет, радия $^{226}_{88}\mathrm{Ra}$ — 1600 лет, полония $^{210}_{84}\mathrm{Po}$ — 138 дней, а у инертного газа радона $^{222}_{86}\mathrm{Rn}$ он составляет всего 3,8 суток.

Выведем теперь закон радиоактивного распада, а именно — найдём, как зависит от времени количество N атомов, не претерпевших пока радиоактивный распад. Начальное число радиоактивных атомов равно N_0, период полураспада равен T.

Имеем следующую простую цепочку рассуждений.

Спустя время $t_1 = T$ количество оставшихся атомов будет равно

$$N_1 = \frac{N_0}{2} = N_0 \cdot 2^{-1}.$$

Спустя время $t_2 = 2T$ атомов останется

$$N_2 = \frac{N_1}{2} = \frac{N_0}{4} = N_0 \cdot 2^{-2}.$$

Спустя время $t_3 = 3T$ атомов останется

$$N_3 = \frac{N_2}{2} = \frac{N_0}{8} = N_0 \cdot 2^{-3}.$$

Становится ясно, что спустя время $t_k = kT$ атомов останется

$$N_k = N_0 \cdot 2^{-k}.$$

Поставляя сюда $k = t_k/T$, получим:

$$N_k = N_0 \cdot 2^{-t_k/T}.$$

Отбрасывая индекс k, находим число оставшихся атомов в зависимости от времени:

$$N = N_0 \cdot 2^{-t/T}. \tag{6.24}$$

Мы получили *закон радиоактивного распада*. Количество нераспавшихся атомов оказывается показательной функцией, убывающей с течением времени.

Ещё одной характеристикой скорости распада радиоактивного элемента является *активность* — число радиоактивных распадов, происходящих в единицу времени. Активность A есть производная по времени от числа $N_0 - N$ распавшихся атомов:

$$A = \frac{d(N_0 - N)}{dt} = -\frac{dN}{dt} = N_0 \frac{\ln 2}{T} \cdot 2^{-t/T}.$$

Обозначая множитель перед показательной функцией через A_0 (это будет активность в начальный момент времени), получим:

$$A = A_0 \cdot 2^{-t/T}.$$

Мы видим, что зависимость активности от времени имеет точно такой же вид, как и закон радиоактивного распада (6.24). График зависимости активности от времени приведён на рис. 6.21.

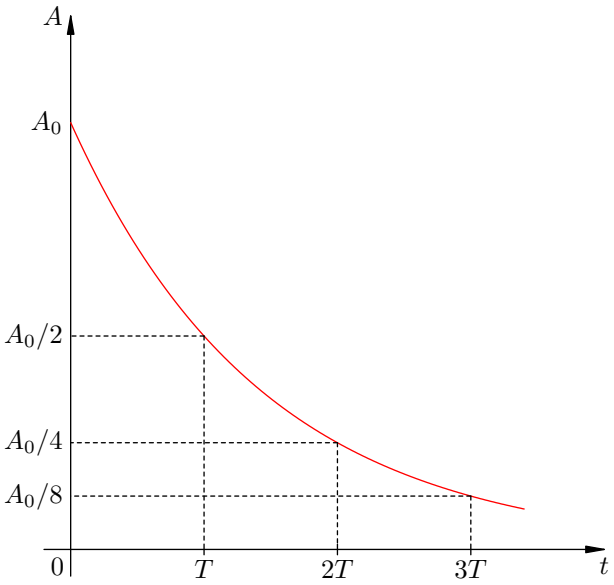


Рис. 6.21. Зависимость активности от времени

Ясно, что активность убывает тем быстрее, чем меньше период полураспада. И наоборот, при большом периоде полураспада активность меняется медленно. Например, активность радона ($T = 3{,}8$ суток) уменьшается буквально на глазах, а активность солей урана ($T = 4{,}5$ млрд. лет) остаётся практически неизменной на протяжении человеческой жизни.

6.10 Энергия связи ядра

Атомное ядро, согласно нуклонной модели, состоит из нуклонов — протонов и нейтронов. Но какие силы удерживают нуклоны внутри ядра?

За счёт чего, например, держатся вместе два протона и два нейтрона внутри ядра атома гелия? Ведь протоны, отталкиваясь друг от друга электрическими силами, должны были бы разлететься в разные стороны! Может быть, это гравитационное притяжение нуклонов друг к другу не даёт ядру распасться?

Давайте проверим. Пусть два протона находятся на некотором расстоянии r друг от друга. Найдём отношение силы F_{el} их электрического отталкивания к силе F_{gr} их гравитационного притяжения:

$$\frac{F_{\text{el}}}{F_{\text{gr}}} = \frac{ke^2/r^2}{Gm^2/r^2} = \frac{ke^2}{Gm^2}\,.$$

Заряд протона $e = 1{,}6 \cdot 10^{-19}$ Кл, масса протона $m \approx 1{,}7 \cdot 10^{-27}$ кг, поэтому имеем:

$$\frac{F_{\text{el}}}{F_{\text{gr}}} = \frac{9 \cdot 10^9 \cdot 1{,}6^2 \cdot 10^{-38}}{6{,}67 \cdot 10^{-11} \cdot 1{,}7^2 \cdot 10^{-54}} \sim 10^{36}.$$

Какое чудовищное превосходство электрической силы! Гравитационное притяжение протонов не то что не обеспечивает устойчивость ядра — оно вообще не заметно на фоне их взаимного электрического отталкивания.

Следовательно, существуют иные силы притяжения, которые скрепляют нуклоны внутри ядра и превосходят по величине силу электрического отталкивания протонов. Это — так называемые *ядерные силы*.

6.10.1 Ядерные силы

До сих пор мы знали два типа взаимодействий в природе — гравитационные и электромагнитные. Ядерные силы служат проявлением нового, третьего по счёту типа взаимодействий — *сильного взаимодействия*. Мы не будем вдаваться в механизм возникновения ядерных сил, а лишь перечислим их наиболее важные свойства.

1. Ядерные силы действуют между любыми двумя нуклонами: протоном и протоном, протоном и нейтроном, нейтроном и нейтроном.

2. Ядерные силы притяжения протонов внутри ядра примерно в 100 раз превосходят силу электрического отталкивания протонов. Более мощных сил, чем ядерные, в природе не наблюдается.

3. Ядерные силы притяжения являются *короткодействующими*: радиус их действия составляет около 10^{-15} м. Это и есть размер ядра — именно на таком расстоянии друг от друга нуклоны удерживаются ядерными силами. При увеличении расстояния ядерные силы очень быстро убывают; если расстояние между нуклонами станет равным $2 \cdot 10^{-15}$ м, ядерные силы почти полностью исчезнут.

   На расстояниях, меньших 10^{-15} м, ядерные силы становятся силами отталкивания.

Сильное взаимодействие относится к числу фундаментальных — его нельзя объяснить на основе каких-то других типов взаимодействий. Способность к сильным взаимодействиям оказалась свойственной не только протонам и нейтронам, но и некоторым другим элементарным частицам; все такие частицы получили название *адронов*. Электроны и фотоны к адронам не относятся — они в сильных взаимодействиях не участвуют.

6.10.2 Атомная единица массы

Массы атомов и элементарных частиц чрезвычайно малы, и измерять их в килограммах неудобно. Поэтому в атомной и ядерной физике часто применяется куда более мелкая единица — так называемая *атомная единица массы* (сокращённо *а. е. м.*).

По определению, *атомная единица массы есть 1/12 массы атома углерода* $^{12}_{6}\mathrm{C}$. Вот её значение с точностью до пяти знаков после запятой в стандартной записи:

$$1\text{ а. е. м.} = 1{,}66054 \cdot 10^{-27}\,\text{кг} = 1{,}66054 \cdot 10^{-24}\,\text{г}.$$

(Такая точность нам впоследствии понадобится для вычисления одной очень важной величины, постоянно применяющейся в расчётах энергии ядер и ядерных реакций.)

Оказывается, что 1 а. е. м., выраженная в граммах, численно равна величине, обратной к постоянной Авогадро $N_\text{A} = 6{,}02214 \cdot 10^{23}\,\text{моль}^{-1}$:

$$\frac{1}{N_\text{A}} = \frac{1}{6{,}02214 \cdot 10^{23}} = 1{,}66054 \cdot 10^{-24}\,\text{моль}.$$

Почему так получается? Вспомним, что число Авогадро есть число атомов в 12 г углерода. Кроме того, масса m_C атома углерода равна 12 а. е. м. Отсюда имеем:

$$12\,\text{г} = N_\text{A} m_\text{C} = N_\text{A} \cdot 12\text{ а. е. м.},$$

поэтому $N_\text{A} \cdot 1\text{ а. е. м.} = 1\,\text{г}$, что и требовалось.

Как вы помните, любое тело массы m обладает энергией покоя E, которая выражается формулой Эйнштейна:

$$E = mc^2. \tag{6.25}$$

Выясним, какая энергия заключена в одной атомной единице массы. Нам надо будет провести вычисления с достаточно высокой точностью, поэтому берём скорость света с пятью знаками после запятой:

$$c = 2{,}99792 \cdot 10^8\,\frac{\text{м}}{\text{с}}.$$

Итак, для массы $m_1 = 1$ а. е. м. имеем соответствующую энергию покоя E_1:

$$E_1 = m_1 c^2 = 1{,}66054 \cdot 10^{-27} \cdot 2{,}99792^2 \cdot 10^{16} = 1{,}49241 \cdot 10^{-10}\,\text{Дж}. \tag{6.26}$$

В случае малых частиц пользоваться джоулями неудобно — по той же причине, что и килограммами. Существует гораздо более мелкая единица измерения энергии — *электронвольт* (сокращённо *эВ*).

По определению, *1 эВ есть энергия, приобретаемая электроном при прохождении ускоряющей разности потенциалов 1 вольт*:

$$1\,\text{эВ} = eV = 1{,}60218 \cdot 10^{-19}\,\text{Кл} \cdot 1\,\text{В} = 1{,}60218 \cdot 10^{-19}\,\text{Дж}. \tag{6.27}$$

(вы привыкли использовать в задачах величину элементарного заряда в виде $e = 1{,}6 \cdot 10^{-19}\,\text{Кл}$, но здесь нам нужны более точные вычисления).

Теперь, наконец, мы готовы вычислить обещанную величину — *энергетический эквивалент атомной единицы массы*, выраженный в МэВ. Из (6.26) и (6.27) получаем:

$$E_1 = \frac{1{,}49241 \cdot 10^{-10}}{1{,}60218 \cdot 10^{-19}} = 0{,}93149 \cdot 10^9\,\text{эВ} = 931{,}5\,\text{МэВ}. \tag{6.28}$$

Итак, запоминаем: **энергия покоя одной а. е. м. равна 931,5 МэВ**. Данное соотношение вы будете неоднократно использовать в задачах. Ведь обычные единицы измерения массы и

энергии в ядерной физике — это атомная единица массы и электронвольты с различными приставками (кэВ, МэВ, ГэВ и сейчас уже ТэВ).

В дальнейшем нам понадобятся массы и энергии покоя протона, нейтрона и электрона. Приведём их с точностью, достаточной для решения задач. Энергия получена из массы умножением на только что вычисленный переводной коэффициент 931,5 МэВ/а. е. м.

$$\begin{aligned} m_p &= 1{,}00728 \text{ а. е. м.}, & E_p &= 938{,}3 \text{ МэВ}; \\ m_n &= 1{,}00867 \text{ а. е. м.}, & E_n &= 939{,}6 \text{ МэВ}; \\ m_e &= 5{,}486 \cdot 10^{-4} \text{ а. е. м.}, & E_e &= 0{,}511 \text{ МэВ}. \end{aligned}$$

Дефект массы и энергия связи

Мы привыкли, что масса тела равна сумме масс частей, из которых оно состоит. В ядерной физике от этой простой мысли приходится отвыкать.

Давайте начнём с примера и возьмём хорошо знакомую нам α-частицу — ядро $^4_2\mathrm{He}$. В таблице (например, в задачнике Рымкевича) имеется значение массы *нейтрального атома* гелия: она равна 4,00260 а. е. м. Для нахождения массы M ядра гелия нужно из массы нейтрального атома вычесть массу двух электронов, находящихся в атоме:

$$M = 4{,}00260 - 2 \cdot 0{,}0005486 = 4{,}00150 \text{ а. е. м.}$$

В то же время, суммарная масса двух протонов и двух нейтронов, из которых состоит ядро гелия, равна:

$$2m_p + 2m_n = 2 \cdot 1{,}00728 + 2 \cdot 1{,}00867 = 4{,}03190 \text{ а. е. м.}$$

Мы видим, что сумма масс нуклонов, составляющих ядро, превышает массу ядра на

$$\Delta m = 2m_p + 2m_n - M = 4{,}03190 - 4{,}00150 = 0{,}0304 \text{ а. е. м.}$$

Величина Δm называется *дефектом массы*. В силу формулы Эйнштейна (6.25) дефекту массы отвечает изменение энергии:

$$\Delta E = \Delta m c^2 = 0{,}0304 \cdot 931{,}5 \approx 28 \text{ МэВ}.$$

Величина ΔE обозначается также $E_{\text{св}}$ и называется *энергией связи* ядра $^4_2\mathrm{He}$. Таким образом, энергия связи α-частицы составляет приблизительно 28 МэВ.

Каков же физический смысл энергии связи (и, стало быть, дефекта масс)?

Чтобы расщепить ядро на составляющие его протоны и нейтроны, нужно *совершить работу* против действия ядерных сил. Эта работа не меньше определённой величины $A_{\min}$; *минимальная* работа $A_{\min}$ по разрушению ядра совершается в случае, когда высвободившиеся протоны и нейтроны *покоятся*.

Ну а если над системой совершается работа, то энергия системы *возрастает* на величину совершённой работы. Поэтому суммарная энергия покоя нуклонов, составляющих ядро и взятых по отдельности, оказывается *больше* энергии покоя ядра на величину $A_{\min}$.

Следовательно, и суммарная масса нуклонов, из которых состоит ядро, будет больше массы самого ядра. Вот почему возникает дефект массы.

В нашем примере с α-частицей суммарная энергия покоя двух протонов и двух нейтронов больше энергии покоя ядра гелия на 28 МэВ. Это значит, что для расщепления ядра $^4_2\mathrm{He}$ на составляющие его нуклоны нужно совершить работу, равную как минимум 28 МэВ. Эту величину мы и назвали энергией связи ядра.

Итак, *энергия связи ядра — это минимальная работа, которую необходимо совершить для расщепления ядра на составляющие его нуклоны.*

Энергия связи ядра есть разность энергий покоя нуклонов ядра, взятых по отдельности, и энергии покоя самого ядра. Если ядро массы M состоит из Z протонов и N нейтронов, то для энергии связи $E_{\text{св}}$ имеем:

$$E_{\text{св}} = (Zm_p + Nm_n)c^2 - Mc^2 = (Zm_p + Nm_n - M)c^2.$$

Величина $\Delta m = Zm_p + Nm_n - M$, как мы уже знаем, называется дефектом массы.

6.10.3 Удельная энергия связи

Важной характеристикой прочности ядра является его *удельная энергия связи*, равная отношению энергии связи к числу нуклонов:

$$\varepsilon = \frac{E_{\text{св}}}{A}\,.$$

Удельная энергия связи есть энергия связи, приходящаяся на один нуклон, и имеет смысл средней работы, которую необходимо совершить для удаления нуклона из ядра.

На рис. 6.22 представлена зависимость удельной энергии связи естественных (то есть встречающихся в природе[20]) изотопов химических элементов от массового числа A.

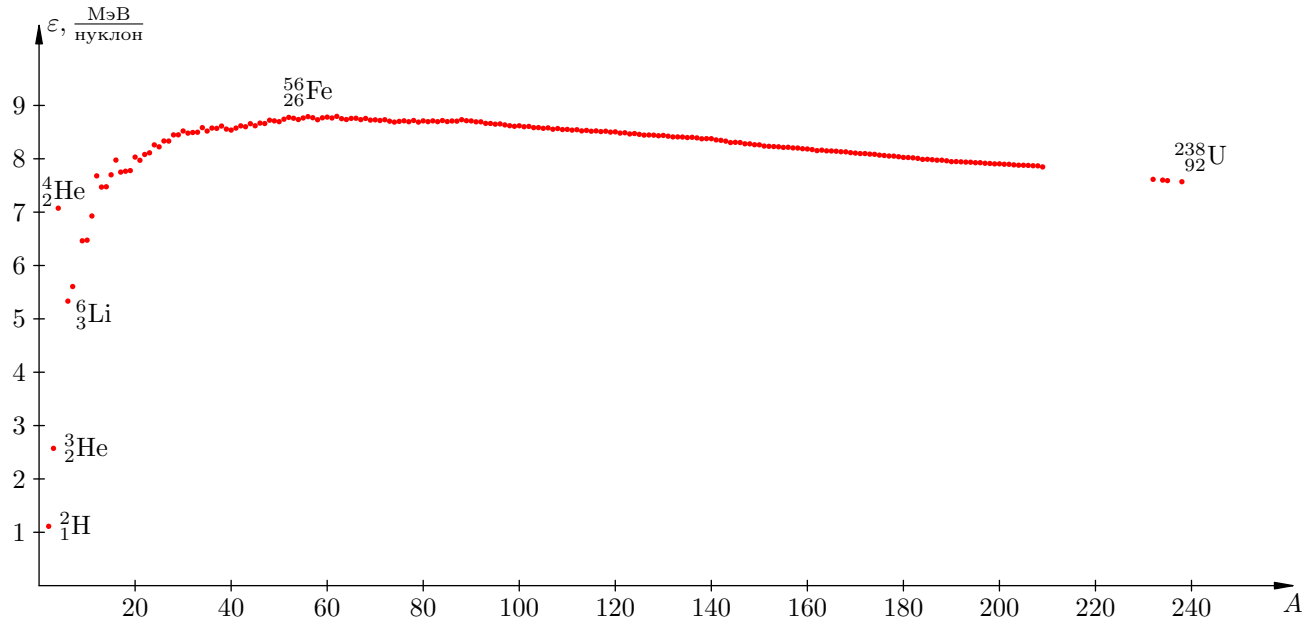


Рис. 6.22. Удельная энергия связи естественных изотопов

У лёгких элементов удельная энергия связи возрастает с ростом A, достигая максимального значения 8,8 МэВ/нуклон в окрестности железа $^{56}_{26}$Fe (то есть в диапазоне изменения A примерно от 50 до 65). Затем она плавно убывает до величины 7,6 МэВ/нуклон у урана $^{238}_{92}$U.

Такой характер зависимости удельной энергии связи от числа нуклонов объясняется совместным действием двух разнонаправленных факторов.

Первый фактор — поверхностные эффекты. Если нуклонов в ядре мало, то значительная их часть находится *на поверхности* ядра. Эти поверхностные нуклоны окружены меньшим числом соседей, чем внутренние нуклоны, и, соответственно, взаимодействуют с меньшим числом соседних нуклонов. При увеличении A доля внутренних нуклонов растёт, а доля поверхностных

[20] Элементы с массовыми числами 210–231, 233, 236, 237 в естественных условиях не встречаются. Этим объясняются пробелы в конце графика.

нуклонов — падает; поэтому работа, которую нужно совершить для удаления одного нуклона из ядра, в среднем должна увеличиваться с ростом A.

Однако с возрастанием числа нуклонов начинает проявляться *второй фактор — кулоновское отталкивание протонов*. Ведь чем больше протонов в ядре, тем большие электрические силы отталкивания стремятся разорвать ядро; иными словами, тем сильнее каждый протон отталкивается от остальных протонов. Поэтому работа, необходимая для удаления нуклона из ядра, в среднем должна уменьшаться с ростом A.

Пока нуклонов мало, первый фактор доминирует над вторым, и потому удельная энергия связи возрастает.

В окрестности железа ($50 \leqslant A \leqslant 65$) действия обоих факторов сравниваются друг с другом, в результате чего удельная энергия связи выходит на максимум. Это область наиболее устойчивых, прочных ядер.

Затем второй фактор начинает перевешивать, и под действием всё возрастающих сил кулоновского отталкивания, распирающих ядро, удельная энергия связи убывает.

6.10.4 Насыщение ядерных сил

Тот факт, что второй фактор доминирует у тяжёлых ядер, говорит об одной интересной особенности ядерных сил: они обладают свойством *насыщения*. Это означает, что каждый нуклон в большом ядре связан ядерными силами не со всеми остальными нуклонами, а лишь с небольшим числом своих соседей, и число это не зависит от размеров ядра.

Действительно, если бы такого насыщения не было, удельная энергия связи продолжала бы возрастать с увеличением A — ведь тогда каждый нуклон скреплялся бы ядерными силами со всё большим числом нуклонов ядра, так что первый фактор неизменно доминировал бы над вторым. У кулоновских сил отталкивания не было бы никаких шансов переломить ситуацию в свою пользу!

6.11 Ядерные реакции

В предыдущем листке мы неоднократно говорили о расщеплении атомного ядра на составные части. Но как этого добиться в действительности? В результате каких физических процессов можно разбить ядро?

Наблюдения радиоактивного распада в изменяющихся внешних условиях — а именно, при различных давлениях и температурах, в электрических и магнитных полях — показали, что скорость радиоактивного распада от этих условий не зависит. Никаких превращений химических элементов друг в друга все эти факторы вызвать не способны. Очевидно, изменения энергии тут слишком малы, чтобы повлиять на атомное ядро — так ветер, обдувающий кирпичный дом, не в состоянии его разрушить.

Но разрушить дом можно артиллерийским снарядом. И Резерфорд в 1919 году решил воспользоваться наиболее мощными «снарядами», которые имелись тогда в распоряжении. Это были α-частицы, вылетающие с энергией около 5 МэВ при радиоактивном распаде урана. (Как вы помните, это те самые снаряды, которыми он восемь лет назад бомбардировал лист золотой фольги в своих знаменитых опытах, породивших планетарную модель атома.)

Правда, превращений золота в другие химические элементы в тех экспериментах не наблюдалось. Ядро золота $^{197}_{79}\mathrm{Au}$ само по себе весьма прочное, да и к тому же содержит довольно много протонов; они создают сильное кулоновское поле, отталкивающее α-частицу и не подпускающее её слишком близко к ядру. А ведь для разбивания ядра α-снаряд должен сблизиться с ядром настолько, чтобы включились ядерные силы! Что ж, раз большое количество протонов мешает — может, взять ядро полегче, где протонов мало?

Резерфорд подверг бомбардировке ядра азота $^{14}_{7}\mathrm{N}$ и в результате осуществил первую в истории физики *ядерную реакцию*:

$$^{14}_{7}\mathrm{N} + {}^{4}_{2}\mathrm{He} \to {}^{17}_{8}\mathrm{O} + {}^{1}_{1}\mathrm{H}. \tag{6.29}$$

В правой части (6.29) мы видим *продукты реакции* — изотоп кислорода и протон.

Стало ясно, что для изучения ядерных реакций нужно располагать частицами-снарядами высоких энергий. Такую возможность дают *ускорители* элементарных частиц. Ускорители имеют два серьёзных преимущества перед естественными «радиоактивными пушками».

1. В ускорителях можно разгонять любые заряженные частицы. В особенности это касается протонов, которые при естественном распаде ядер не появляются. Протоны хороши тем, что несут минимальный заряд, а значит — испытывают наименьшее кулоновское отталкивание со стороны ядер-мишеней.

2. Ускорители позволяют достичь энергий, на несколько порядков превышающие энергию α-частиц при радиоактивном распаде. Например, в Большом адронном коллайдере протоны разгоняются до энергий в несколько ТэВ; это в миллион раз больше, чем 5 МэВ у α-частиц в реакции (6.29), осуществлённой Резерфордом.

Так, с помощью протонов, прошедших через ускоритель, в 1932 году удалось разбить ядро лития (получив при этом две α-частицы):

$$^{7}_{3}\mathrm{Li} + {}^{1}_{1}\mathrm{H} \to {}^{4}_{2}\mathrm{He} + {}^{4}_{2}\mathrm{He}. \tag{6.30}$$

Ядерные реакции дали возможность искусственного превращения химических элементов. Кроме того, в продуктах реакций стали обнаруживаться новые, не известные ранее частицы. Например, при облучении бериллия α-частицами в том же 1932 году был открыт нейтрон:

$$^{9}_{4}\mathrm{Be} + {}^{4}_{2}\mathrm{He} \to {}^{12}_{6}\mathrm{O} + {}^{1}_{0}n.$$

Нейтроны замечательно подходят для раскалывания ядер: не имея электрического заряда, они беспрепятственно проникают внутрь ядра[21]. Так, при облучении азота нейтронами протекает следующая реакция:

$$^{14}_{7}\mathrm{N} + ^{1}_{0}n \to ^{11}_{5}\mathrm{B} + ^{4}_{2}\mathrm{He}.$$

В ходе экспериментов с ядерными реакциями была открыта *искусственная радиоактивность* — получены радиоактивные изотопы, не встречающиеся в естественных условиях. Например, в реакции

$$^{27}_{13}\mathrm{Al} + ^{4}_{2}\mathrm{He} \to ^{30}_{15}\mathrm{P} + ^{1}_{0}n$$

получился радиоактивный изотоп фосфора $^{30}_{15}\mathrm{P}$, которого нет в природе (ядро «природного» фосфора $^{31}_{15}\mathrm{P}$ стабильно). Этот радиоактивный фосфор распадается, испуская позитрон и превращаясь в кремний:

$$^{30}_{15}\mathrm{P} \to ^{30}_{14}\mathrm{Si} + ^{0}_{1}e.$$

6.11.1 Энергетический выход ядерной реакции

Обсуждая энергию связи, мы видели, что в результате ядерных процессов масса системы частиц не остаётся постоянной. Это, в свою очередь, приводит к тому, что кинетическая энергия продуктов ядерной реакции отличается от кинетической энергии исходных частиц.

Прежде всего напомним, что полная энергия E частицы массы m складывается из её энергии покоя mc^2 и кинетической энергии K:

$$E = mc^2 + K.$$

Пусть в результате столкновения частиц A и B происходит ядерная реакция, продуктами которой служат частицы X и Y:

$$A + B \to X + Y. \tag{6.31}$$

Полная энергия системы частиц сохраняется:

$$E_A + E_B = E_X + E_Y,$$

то есть

$$(m_Ac^2 + K_A) + (m_Bc^2 + K_B) = (m_Xc^2 + K_X) + (m_Yc^2 + K_Y). \tag{6.32}$$

Кинетическая энергия исходных частиц равна $K_A + K_B$. Кинетическая энергия продуктов реакции равна $K_X + K_Y$. *Энергетический выход* Q ядерной реакции — это разность кинетических энергий продуктов реакции и исходных частиц:

$$Q = (K_X + K_Y) - (K_A + K_B).$$

Из (6.32) легко получаем:

$$Q = (m_A + m_B - m_X - m_Y)c^2. \tag{6.33}$$

Если $Q > 0$, то говорят, что реакция идёт *с выделением энергии*: кинетическая энергия продуктов реакции *больше* кинетической энергии исходных частиц. Из (6.33) мы видим, что в этом случае суммарная масса продуктов реакции *меньше* суммарной массы исходных частиц.

Если же $Q < 0$, то реакция идёт *с поглощением энергии*: кинетическая энергия продуктов реакции *меньше* кинетической энергии исходных частиц. Суммарная масса продуктов реакции в этом случае *больше* суммарной массы исходных частиц.

[21] При этом ускорять нейтроны не надо — медленные нейтроны легче проникают в ядра. Нейтроны, оказывается, нужно даже замедлять, и делается это пропусканием нейтронов через обычную воду.

Таким образом, термины «выделение» и «поглощение» энергии не должны вызывать недоумение: они относятся только к *кинетической* энергии частиц. Полная энергия системы частиц, разумеется, в любой реакции остаётся неизменной.

Чтобы посчитать энергетический выход Q ядерной реакции (6.31), действуем по следующему алгоритму.

1. С помощью таблицы масс нейтральных атомов находим m_A, m_B, m_X и m_Y, выраженные в а. е. м. (для нахождения массы ядра не забываем вычесть из массы нейтрального атома массу электронов).

2. Вычисляем массу $m_1 = m_A + m_B$ исходных частиц, массу $m_2 = m_X + m_Y$ продуктов реакции и находим разность масс $\Delta m = m_1 - m_2$.

3. Умножаем Δm на 931,5 и получаем величину Q, выраженную в МэВ.

Мы сейчас подробно рассмотрим вычисление энергетического выхода Q на двух примерах бомбардировки ядер лития ${}^7_3\mathrm{Li}$: сначала — протонами, затем — α-частицами.

В первом случае имеем уже упоминавшуюся выше реакцию (6.30):

$$ {}^7_3\mathrm{Li} + {}^1_1\mathrm{H} \to {}^4_2\mathrm{He} + {}^4_2\mathrm{He}. $$

Масса атома лития ${}^7_3\mathrm{Li}$ равна 7,01601 а. е. м. Масса электрона равна 0,000548 а. е. м. Вычитая из массы атома массу трёх его электронов, получаем *массу ядра лития* ${}^7_3\mathrm{Li}$:

$$ 7{,}01601 - 3 \cdot 0{,}000548 = 7{,}01437 \text{ а. е. м.} $$

Масса протона равна 1,00728 а. е. м., так что масса исходных частиц:

$$ m_1 = 7{,}01437 + 1{,}00728 = 8{,}02165 \text{ а. е. м.} $$

Переходим к продуктам реакции. Масса атома гелия равна 4,00260 а. е. м. Вычитаем массу электронов и находим *массу ядра гелия* ${}^4_2\mathrm{He}$:

$$ 4{,}00260 - 2 \cdot 0{,}000548 = 4{,}00150 \text{ а. е. м.} $$

Умножая на 2, получаем массу продуктов реакции:

$$ m_2 = 2 \cdot 4{,}00150 = 8{,}00300 \text{ а. е. м.} $$

Масса, как видим, уменьшилась ($m_2 < m_1$); это означает, что наша реакция идёт с выделением энергии. Разность масс:

$$ \Delta m = m_1 - m_2 = 8{,}02165 - 8{,}00300 = 0{,}01865 \text{ а. е. м.} $$

Выделившаяся энергия:

$$ Q = 0{,}01865 \cdot 931{,}5 = 17{,}4 \text{ МэВ}. $$

Теперь рассмотрим второй пример. При бомбардировке ядер лития α-частицами происходит реакция:

$$ {}^7_3\mathrm{Li} + {}^4_2\mathrm{He} \to {}^{10}_5\mathrm{B} + {}^1_0 n. \tag{6.34} $$

Массы исходных ядер нам уже известны; остаётся сосчитать их суммарную массу:

$$ m_1 = 7{,}01437 + 4{,}00150 = 11{,}01587 \text{ а. е. м.} $$

Из таблицы берём массу атома бора $^{10}_{5}\mathrm{B}$ (она равна 10,01294 а. е. м.); вычитаем массу пяти электронов и получаем массу ядра атома бора:

$$10{,}01294 - 5 \cdot 0{,}000548 = 10{,}01020 \text{ а. е. м.}$$

Масса нейтрона равна 1,00867 а. е. м. Находим массу продуктов реакции:

$$m_2 = 10{,}01020 + 1{,}00867 = 11{,}01887 \text{ а. е. м.}$$

На сей раз масса увеличилась ($m_2 > m_1$), то есть реакция идёт с поглощением энергии. Разность масс равна:

$$\Delta m = m_1 - m_2 = -0{,}0030 \text{ а. е. м.}$$

Энергетический выход реакции:

$$Q = -0{,}0030 \cdot 931{,}5 = -2{,}8 \text{ МэВ.}$$

Таким образом, в реакции (6.34) поглощается энергия 2,8 МэВ. Это означает, что суммарная кинетическая энергия продуктов реакции (ядра бора и нейтрона) на 2,8 МэВ меньше, чем суммарная кинетическая энергия исходных частиц (ядра лития и α-частицы). Поэтому, чтобы данная реакция в принципе осуществилась, энергия исходных частиц должна быть не меньше величины 2,8 МэВ.

6.11.2 Деление ядер

Бомбардируя ядра урана медленным нейтронами, немецкие физики Ган и Штрассман обнаружили появление элементов средней части периодической системы: бария, криптона, стронция, рубидия, цезия и т. д. Так было открыто *деление ядер* урана.

На рис. 6.23 мы видим процесс деления ядра[22]. Захватывая нейтрон, ядро урана делится на два *осколка*, и при этом освобождаются два-три нейтрона.

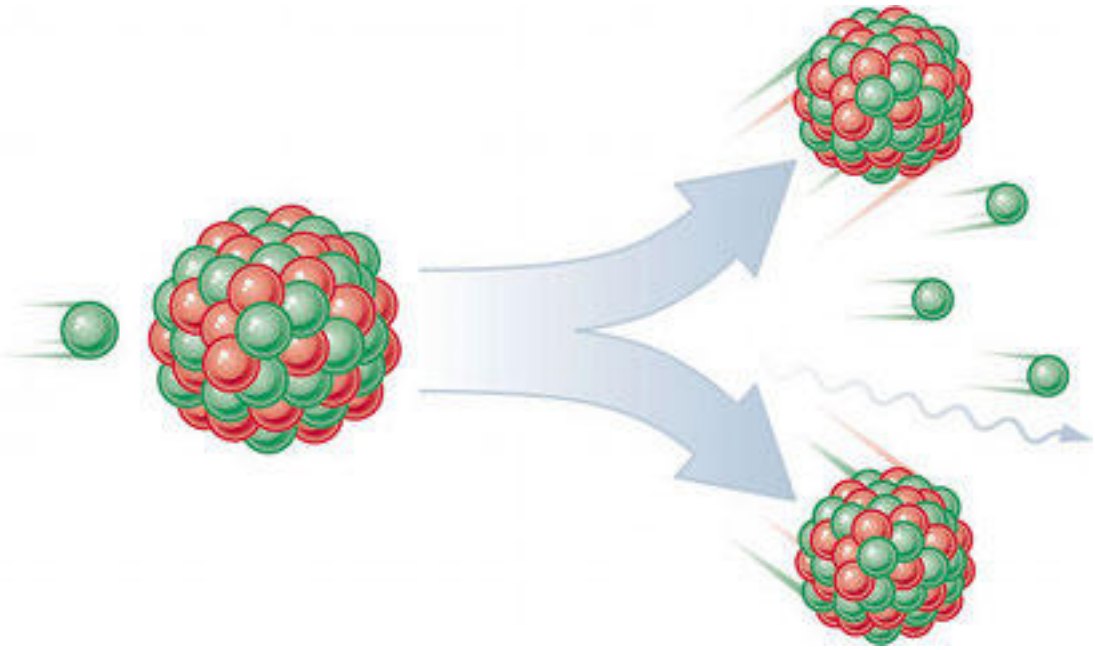


Рис. 6.23. Деление ядра урана

[22]Изображение с сайта oup.co.uk.

Осколки являются ядрами радиоактивных изотопов элементов середины таблицы Менделеева. Один из осколков всегда больше другого. Например, при бомбардировке урана $^{235}_{92}\mathrm{U}$ могут встречаться такие комбинации осколков (как говорят, реакция идёт по следующим *каналам*).

- Барий и криптон: $^{235}_{92}\mathrm{U} + {}^{1}_{0}n \to {}^{144}_{56}\mathrm{Ba} + {}^{89}_{36}\mathrm{Kr} + 3\,{}^{1}_{0}n$.
- Цезий и рубидий: $^{235}_{92}\mathrm{U} + {}^{1}_{0}n \to {}^{140}_{55}\mathrm{Cs} + {}^{94}_{37}\mathrm{Rb} + 2\,{}^{1}_{0}n$.
- Ксенон и стронций: $^{235}_{92}\mathrm{U} + {}^{1}_{0}n \to {}^{140}_{54}\mathrm{Xe} + {}^{94}_{38}\mathrm{Sr} + 2\,{}^{1}_{0}n$.

Вообще, известно около 50 каналов распада ядра урана $^{235}_{92}\mathrm{U}$. В каждой из этих реакций выделяется очень большая энергия — порядка 200 МэВ. Сравните эту величину с найденным выше энергетическим выходом реакции (6.30), равным 17,4 МэВ! Откуда берётся такое количество энергии?

Основная причина состоит в том, что из-за наличия большого числа протонов — а их в ядре урана упаковано 92 штуки — кулоновские силы отталкивания, распирающие ядро, очень велики. Ядерные силы пока ещё в состоянии удерживать ядро от распада, но могучий кулоновский фактор готов сказать своё слово в любой момент.

И такой момент настаёт, когда в ядре застревает нейтрон (рис. 6.24)[23].

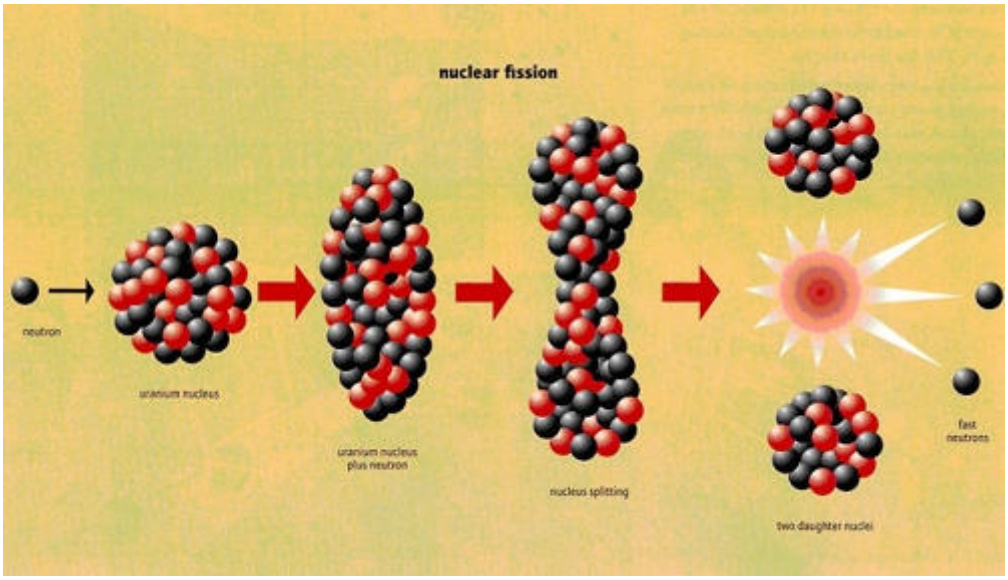


Рис. 6.24. Деформация, колебания и разрыв ядра

Застрявший нейтрон вызывает деформацию ядра и колебания его формы. Эти колебания могут стать столь интенсивными, что ядро вытянется в «гантельку». Между половинками гантельки образуется тонкий перешеек из небольшого числа нуклонов.

Получается, что силам электрического отталкивания половинок гантельки противостоят короткодействующие ядерные силы, скрепляющие нуклоны перешейка. Чем тоньше становится перешеек, тем меньше в нём оказывается нуколонов, и тем слабее будет действие ядерных сил. В конце концов перешеек не выдержит, и ядро разорвётся.

Осколки ядра разлетаюттся с огромной скоростью — около 1/30 скорости света. Они и уносят бóльшую часть энергии, высвобождающейся при делении ядра[24].

[23] Изображение с сайта investingreenenergy.com.

[24] Из 200 МэВ энергетического выхода примерно 168 МэВ приходится на кинетическую энергию осколков. Остальная энергия распределяется между нейтронами и другими частицами, появляющимися в процессе деления: гамма-квантами, электронами и антинейтрино.

Появление двух-трёх свободных нейтронов — характерная черта деления ядра урана. Чем объясняется их неизменное присутствие в каждом канале деления?

Дело в том, что по мере движения от конца таблицы Менделеева к её началу отношение N/A числа нейтронов к числу нуклонов в стабильных ядрах *уменьшается*. Возьмём в качестве примера уран и продукты его деления — барий и криптон (стабильные ядра).

- Для урана $^{235}_{92}\mathrm{U}$ имеем $N/A = 143/235 = 0{,}61$;
- Для бария $^{137}_{56}\mathrm{Ba}$ имеем $N/A = 81/137 = 0{,}59$;
- Для криптона $^{84}_{36}\mathrm{Kr}$ имеем $N/A = 48/84 = 0{,}57$.

Поэтому при делении тяжёлого ядра возникают «лишние» нейтроны, часть которых и освобождается в качестве продуктов реакции. Другая часть лишних нейтронов остаётся в ядрах-осколках, делая их радиоактивными. Осколки дают начало целым цепочкам β-распадов, последовательно уменьшающих отношение N/A. Например:

$$^{140}_{54}\mathrm{Xe} \xrightarrow{\beta} {}^{140}_{55}\mathrm{Cs} \xrightarrow{\beta} {}^{140}_{56}\mathrm{Ba} \xrightarrow{\beta} {}^{140}_{57}\mathrm{La} \xrightarrow{\beta} {}^{140}_{58}\mathrm{Ce}.$$

(концом этой цепочки является стабильное ядро изотопа церия). В ядерной энергетике такие цепочки распадов продуктов деления тяжёлых ядер создают серьёзную экологическую проблему ликвидации радиоактивных отходов.

Деление тяжёлых ядер можно истолковать с точки зрения уже известного нам графика зависимости удельной энергии связи ядра от его массового числа (рис. 6.25).

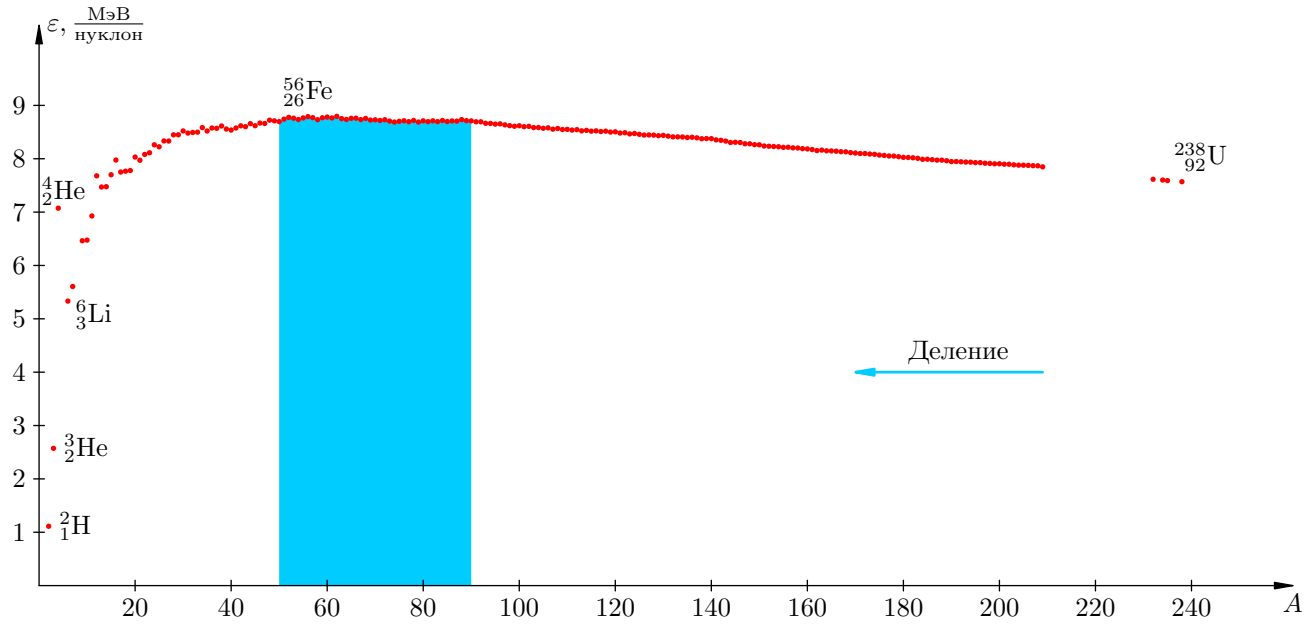


Рис. 6.25. Деление тяжёлых ядер энергетически выгодно

Цветом выделена область $50 \leqslant A \leqslant 90$, в которой удельная энергия связи достигает наибольшего значения 8,7 МэВ/нуклон. Это область наиболее устойчивых ядер. Справа от этой области удельная энергия связи плавно уменьшается до 7,6 МэВ/нуклон у ядра урана.

Процесс превращения менее устойчивых ядер в более устойчивые является энергетически выгодным и сопровождается выделением энергии. При делении ядра урана, как видим, удельная энергия связи повышается примерно на 1 МэВ/нуклон; эта энергия как раз и выделяется в процессе деления. Умножив это на число нуклонов в ядре урана, получим приблизительно те самые 200 МэВ энергетического выхода, о которых говорилось выше.

6.11.3 Цепная ядерная реакция

Появление двух-трёх нейтронов в процессе деления ядра урана — важнейший факт. Эти нейтроны «первого поколения» могут попасть в новые ядра и вызвать их деление; в результате деления новых ядер возникнут нейтроны «второго поколения», которые попадут в следующие ядра и вызовут их деление; возникнут нейтроны «третьего поколения», которые приведут к делению очередных ядер и т. д. Так идёт *цепная ядерная реакция*, в ходе которой высвобождается колоссальное количество энергии.

Для протекания цепной ядерной реакции необходимо, чтобы число N_i высвободившихся нейтронов в очередном поколении было не меньше числа N_{i-1} нейтронов в предыдущем поколении. Величина

$$k = \frac{N_i}{N_{i-1}}$$

называется *коэффициентом размножения нейтронов*. Таким образом, цепная реакция идёт при условии $k \geqslant 1$. Если $k < 1$, то цепная реакция не возникает.

В случае $k > 1$ происходит лавинообразное нарастание числа освобождающихся нейтронов, и цепная реакция становится *неуправляемой*. Так происходит взрыв атомной бомбы.

В ядерных реакторах происходит *управляемая* цепная реакция деления с коэффициентом размножения $k = 1$. Стационарное течение управляемой цепной реакции обеспечивается введением в активную зону реактора (то есть в ту область, где протекает реакция) специальных управляющих стержней, поглощающих нейтроны. При полностью введённых стержнях поглощение ими нейтронов настолько велико, что $k < 1$ и реакция не идёт. В процессе запуска реактора стержни постепенно выводят из активной зоны, пока выделяемая мощность не достигнет требуемого уровня. Этот уровень тщательно контролируется, и при его превышении включаются устройства, вводящие управляющие стержни назад в активную зону.

6.11.4 Термоядерная реакция

Наряду с реакцией деления тяжёлых ядер энергетически выгодным оказывается и обратный в некотором смысле процесс — *синтез лёгких ядер*, то есть слияние ядер лёгких элементов (расположенных в начале периодической таблицы) с образованием более тяжёлого ядра.

Чтобы началось слияние ядер, их нужно сблизить вплотную — чтобы вступили в действие ядерные силы. Для такого сближения нужно преодолеть кулоновское отталкивание ядер, резко возрастающее с уменьшением расстояния между ними. Это возможно лишь при очень большой кинетической энергии ядер, а значит — при очень высокой температуре (в десятки и сотни миллионов градусов). Поэтому реакция ядерного синтеза называется *термоядерной реакцией*.

В качестве примера термоядерной реакции приведём реакцию слияния ядер дейтерия и трития (тяжёлого и сверхтяжёлого изотопов водорода), в результате которой образуется ядро гелия и нейтрон:

$$^{2}_{1}\mathrm{H} + {}^{3}_{1}\mathrm{H} \to {}^{4}_{2}\mathrm{He} + {}^{1}_{0}n. \tag{6.35}$$

Эта реакция идёт с выделением энергии, равной 17,6 МэВ (попробуйте сами провести расчёты и получить данную величину). Это очень много, если учесть, что в реакции участвуют всего 5 нуклонов! В самом деле, в расчёте на один нуклон в реакции (6.35) выделяется энергия примерно 3,5 МэВ, в то время как при делении ядра урана выделяется «всего» 1 МэВ на нуклон.

Таким образом, термоядерные реакции служат источником ещё большего количества энергии, чем реакции деления ядер. С физической точки зрения это понятно: энергия реакции ядерного деления есть в основном кинетическая энергия осколков, разогнанных *электрическими* силами отталкивания, а при ядерном синтезе энергия высвобождается в результате разгона нуклонов навстречу друг другу под действием куда более мощных *ядерных* сил притяжения.

Проще говоря, при делении ядер высвобождается энергия электрического взаимодействия, а при синтезе ядер — энергия сильного (ядерного) взаимодействия.

В недрах звёзд достигаются температуры, подходящие для синтеза ядер. Свет Солнца и далёких звёзд несёт энергию, выделяющуюся в термоядерных реакциях — при слиянии ядер водорода в ядра гелия и последующем слиянии ядер гелия в ядра более тяжёлых элементов, расположенных в средней части периодической системы. Направление термоядерного синтеза показано на рис. 6.26; синтез лёгких ядер энергетически выгоден, так как направлен в сторону увеличения удельной энергии связи ядра.

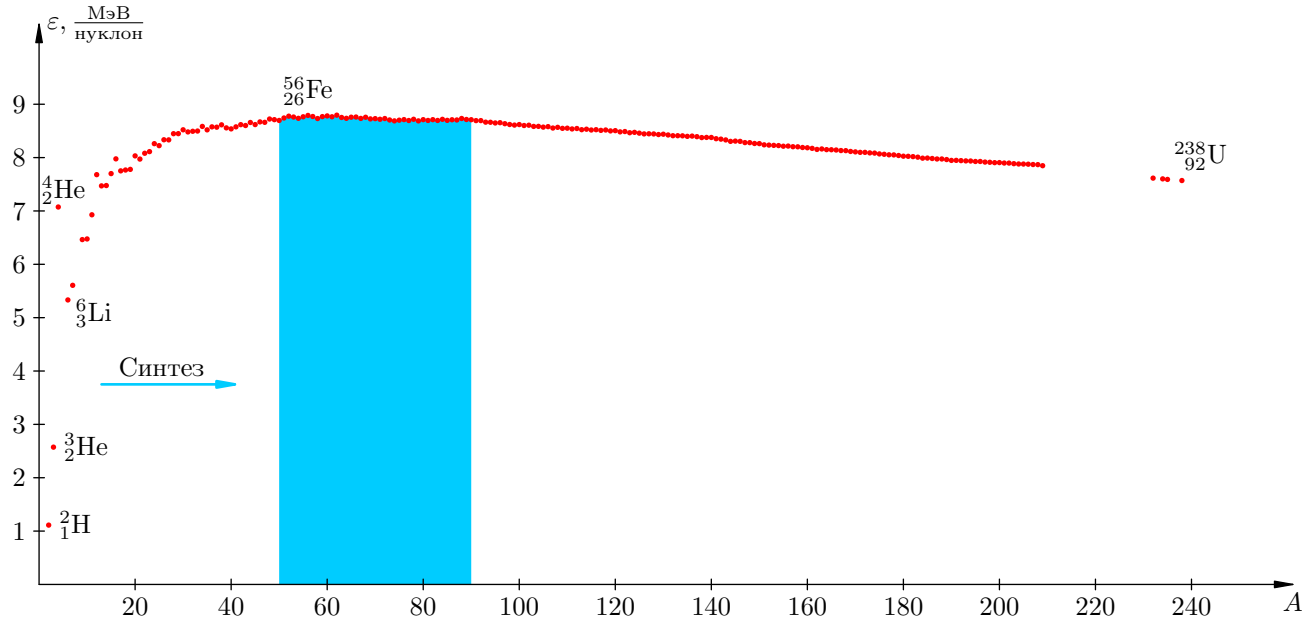


Рис. 6.26. Синтез лёгких ядер энергетически выгоден

Неуправляемая термоядерная реакция осуществляется при взрыве водородной бомбы. Сначала взрывается встроенная атомная бомба — это нужно для создания высокой температуры и давления на первой ступени термоядерного взрыва. Термоядерное горючее сжимается давлением излучения и разогревается до такой степени, что начинаются реакции ядерного синтеза, и происходит взрыв собственно водородной бомбы.

Осуществление управляемой термоядерной реакции остаётся пока нерешённой проблемой, над которой физики работают уже более полувека. Если удастся добиться управляемого течения термоядерного синтеза, то человечество получит в своё распоряжение фактически неограниченный источник энергии. Это чрезвычайно важная задача, стоящая перед нынешним и будущими поколениями — в свете угрожающей перспективы истощения нефтегазовых ресурсов нашей планеты.

Глава 7

Приложение. Векторы в физике

Векторы — мощный инструмент математики и физики. На языке векторов формулируются основные законы механики и электродинамики. Чтобы понимать физику, нужно научиться работать с векторами.

Данная глава содержит подробное изложение материала, необходимого для того, чтобы приступить к изучению механики:

→ Скалярные и векторные величины
→ Сложение векторов
→ Умножение скаляра на вектор
→ Угол между векторами
→ Проекция вектора на ось
→ Векторы и координаты на плоскости
→ Векторы и координаты в пространстве
→ Скалярное произведение векторов

Нет особого смысла читать всю эту главу сразу. Гораздо полезнее обращаться к ней по мере необходимости в процессе чтения остальных глав пособия.

К тексту данного приложения полезно будет вернуться на первом курсе при изучении аналитической геометрии и линейной алгебры — чтобы осознать, например, откуда берутся аксиомы линейного и евклидова пространства.

7.1 Скалярные и векторные величины

В процессе изучения физики мы встречаем два типа величин — скалярные и векторные.

Определение. Скалярная величина, или скаляр — это физическая величина, для задания которой (в подходящих единицах измерения) достаточно одного числа.

Скаляров очень много в физике. Масса тела равна 3 кг, температура воздуха равна $-10°$C, напряжение в сети равно 220 В... Во всех этих случаях интересующая нас величина задаётся одним-единственным числом. Следовательно, масса, температура и электрическое напряжение являются скалярами.

Но скаляр в физике — это не просто число. *Скаляр есть число, снабжённое размерностью*[1]. Так, задавая массу, мы не можем написать $m = 3$; надо указать единицу измерения — например, $m = 3$ кг. И если в математике мы можем сложить числа 3 и 220, то в физике сложить 3 килограмма и 220 вольт не получится: мы имеем право складывать лишь те скаляры, которые обладают одинаковой размерностью (массу с массой, напряжение с напряжением и т. д.).

Определение. Векторная величина, или вектор — это физическая величина, характеризуемая: 1) неотрицательным скаляром; 2) направлением в пространстве. При этом скаляр называется модулем вектора, или его абсолютной величиной.

Предположим, что автомобиль движется со скоростью 60 км/ч. Но ведь это неполная информация о движении, не так ли? Может оказаться важным и то, *куда* едет автомобиль, в каком именно *направлении*. Поэтому важно знать не только модуль (абсолютную величину) скорости автомобиля — в данном случае это 60 км/ч — но и её направление в пространстве. Значит, скорость является вектором.

Другой пример. Допустим, на полу лежит кирпич массой 1 кг. На кирпич действует сила 100 Н (это модуль силы, или её абсолютная величина). Как будет двигаться кирпич? Вопрос лишён смысла до тех пор, пока не указано направление действия силы. Если сила действует вверх, то и кирпич будет двигаться вверх. Если сила действует горизонтально, то и кирпич поедет горизонтально. А если сила действует вертикально вниз, то кирпич вообще не сдвинется с места — он будет только вжиматься в пол. Мы видим, таким образом, что сила также является вектором.

Векторная величина в физике также обладает размерностью. *Размерность вектора — это размерность его модуля.*

Мы будем обозначать векторы буквами со стрелкой. Так, вектор скорости можно обозначить через $\vec{v}$, а вектор силы — через $\vec{F}$. Собственно, *вектор — это и есть стрелка* или, как ещё говорят, *направленный отрезок* (рис. 7.1).

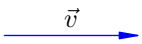


Рис. 7.1. Вектор $\vec{v}$

Начальная точка стрелки называется *началом* вектора, а конечная точка (остриё) стрелки — *концом* вектора. В математике вектор с началом в точке A и концом в точке B обозначается также $\overrightarrow{AB}$; нам такое обозначение тоже иногда понадобится.

Вектор, начало и конец которого совпадают, называется *нулевым* вектором (или *нулём*) и обозначается $\vec{0}$. Нулевой вектор есть попросту точка; он не имеет определённого направления. Длина нулевого вектора, разумеется, равна нулю.

[1] Попадаются и безразмерные скаляры: коэффициент трения, коэффициент полезного действия, показатель преломления среды... Так, показатель преломления воды равен 1,33 — это исчерпывающая информация, никакой размерностью данное число не обладает.

Рисование стрелок полностью решает задачу графического представления векторных величин. *Направление стрелки указывает направление данного вектора, а длина стрелки в подходящем масштабе есть модуль этого вектора.*

Предположим, например, что два автомобиля двигаются навстречу друг другу со скоростями $u = 30$ км/ч и $v = 60$ км/ч. Тогда векторы $\vec{u}$ и $\vec{v}$ скоростей автомобилей будут иметь противоположные направления, причём длина вектора $\vec{v}$ в два раза больше (рис. 7.2).

Рис. 7.2. Вектор $\vec{v}$ вдвое длиннее

Как вы уже поняли, буква без стрелки (например, u или v в предыдущем абзаце) обозначает модуль соответствующего вектора. В математике модуль вектора $\vec{v}$ обычно обозначается $|\vec{v}|$, но физики, если ситуация позволяет, предпочтут именно v — букву без стрелки.

Векторы называются *коллинеарными*, если они расположены на одной прямой или на параллельных прямых.

Пусть имеются два коллинеарных вектора. Если их направления совпадают, то векторы называются *сонаправленными*; если же их направления различны, то векторы называются *противоположно направленными*. Так, выше на рис. 7.2 векторы $\vec{u}$ и $\vec{v}$ являются противоположно направленными.

Два вектора называются *равными*, если они сонаправлены и имеют равные модули (рис. 7.3).

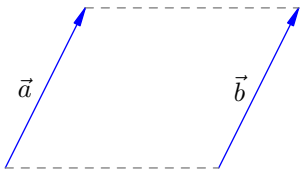


Рис. 7.3. Векторы $\vec{a}$ и $\vec{b}$ равны: $\vec{a} = \vec{b}$

Таким образом, равенство векторов отнюдь не означает непременного совпадения их начал и концов: *мы можем переносить вектор параллельно самому себе, и при этом получится вектор, равный исходному.* Такой перенос постоянно применяется в тех случаях, когда желательно свести начала векторов в одну точку — например, при нахождении суммы или разности векторов. К рассмотрению операций над векторами мы и переходим.

7.2 Сложение векторов

В физике можно складывать только векторы, обладающие одинаковой размерностью. Мы можем складывать скорость со скоростью, силу с силой, но не имеем права сложить вектор скорости с вектором силы.

Правила сложения векторов можно объяснить на двух характерных примерах: сложении перемещений и сложении сил.

7.2.1 Правило треугольника

Начнём с перемещений. *Перемещением* называется вектор, соединяющий начальное и конечное положения тела.

Если, например, тело находилось в точке A и после этого оказалось в точке B, то перемещением тела будет вектор $\vec{s} = \overrightarrow{AB}$. Перемещение тела не зависит от формы траектории; оно определяется лишь начальной и конечной точками движения. На рис. 7.4 изображено перемещение тела $\vec{s}$ и для сравнения пунктиром показана траектория тела.

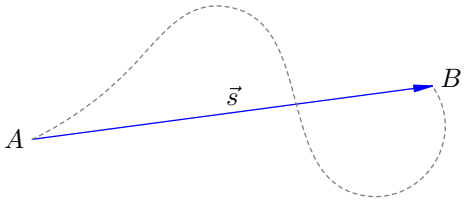


Рис. 7.4. Вектор перемещения

Предположим, что тело совершило перемещение $\vec{s}_1$ из точки A в точку B, а затем — перемещение $\vec{s}_2$ из точки B в точку C (рис. 7.5). Итоговое перемещение есть вектор $\vec{s}$, соединяющий начальную точку A с конечной точкой C.

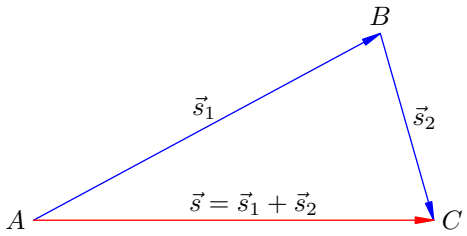


Рис. 7.5. Сложение перемещений

Перемещение $\vec{s}$ есть результат двух последовательно совершённых перемещений $\vec{s}_1$ и $\vec{s}_2$, и поэтому естественно считать, что оно является их суммой: $\vec{s} = \vec{s}_1 + \vec{s}_2$. Это приводит нас к *правилу треугольника* для сложения произвольных векторов (рис. 7.6).

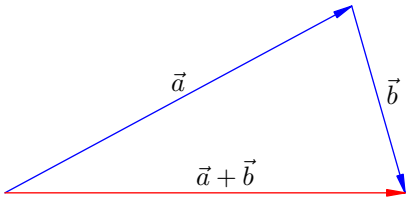


Рис. 7.6. Правило треугольника

Правило треугольника. Поместим начало вектора $\vec{b}$ в конец вектора $\vec{a}$. Тогда вектор $\vec{a} + \vec{b}$ соединяет начало вектора $\vec{a}$ с концом вектора $\vec{b}$.

7.2.2 Правило параллелограмма

Несколько иная картина возникает при сложении сил. Допустим, что в точке O находится небольшое тело и к нему приложены две силы: $\vec{F}_1$ и $\vec{F}_2$.

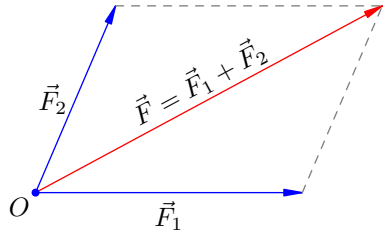


Рис. 7.7. Сложение сил

Опыт показывает, что совместное действие этих сил равноценно действию *одной* силы $\vec{F}$, которая служит диагональю параллелограмма, построенного на векторах $\vec{F}_1$ и $\vec{F}_2$ (рис. 7.7).

Иными словами, движение нашего тела не претерпит никаких изменений, если убрать силы $\vec{F}_1$ и $\vec{F}_2$ и заменить их силой $\vec{F}$. Эта сила $\vec{F}$ называется *равнодействующей* (или *результирующей*) двух сил $\vec{F}_1$ и $\vec{F}_2$; она является результатом их совместного применения, и потому естественно считать, что она будет их суммой: $\vec{F} = \vec{F}_1 + \vec{F}_2$.

Данное соображение приводит нас к *правилу параллелограмма* для сложения двух произвольных векторов (рис. 7.8).

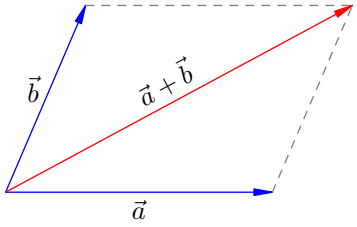


Рис. 7.8. Правило параллелограмма

Правило параллелограмма. Поместим начала векторов $\vec{a}$ и $\vec{b}$ в одну точку. Тогда вектор $\vec{a} + \vec{b}$, имея начало в той же точке, является диагональю параллелограмма, построенного на векторах $\vec{a}$ и $\vec{b}$.

Итак, имеются два естественных способа складывать векторы: правило треугольника и правило параллелограмма. Если бы эти правила приводили к разным результатам, было бы очень скверно. Но, к счастью, результат получается один и тот же!

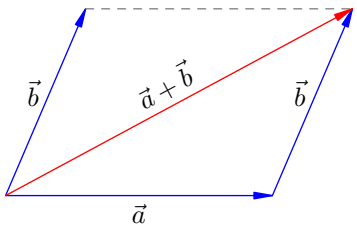


Рис. 7.9. Правило треугольника = Правило параллелограмма

Посмотрите на рис. 7.9. Сначала мы сложили векторы $\vec{a}$ и $\vec{b}$ по правилу параллелограмма. Затем перенесли вектор $\vec{b}$ параллельно самому себе так, чтобы его начало совпало с концом

вектора $\vec{a}$ (перенесённый вектор $\vec{b}$ изображён на рисунке пунктиром). Тем самым возникла возможность сложить наши векторы по правилу треугольника, и в результате мы получаем тот же суммарный вектор $\vec{a}+\vec{b}$, что и в первом случае — а именно, диагональ параллелограмма.

Таким образом, правила треугольника и параллелограмма легко сводятся друг к другу, и между ними нет никакой разницы. В физике мы чаще пользуемся правилом параллелограмма (складывая силы, скорости, ускорения, напряжённости поля и т. п.), поскольку складываемые векторы обычно приложены в одной точке.

Единственная загвоздка с нашими правилами состоит в том, что при сложении *коллинеарных* векторов не возникает ни треугольника, ни параллелограмма. Но правило треугольника — в том виде, как оно было сформулировано — продолжает работать (рис. 7.10).

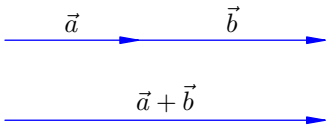


Рис. 7.10. Сложение коллинеарных векторов

А именно, мы помещаем начало вектора $\vec{b}$ в конец вектора $\vec{a}$ и соединяем начало вектора $\vec{a}$ с концом вектора $\vec{b}$. Получится вектор $\vec{a}+\vec{b}$, который на рисунке расположен ниже.

7.2.3 Свойства сложения векторов

Операция сложения векторов обладает всеми хорошими алгебраическими свойствами, которые присущи сложению чисел и привычны для нас.

1. От перестановки слагаемых сумма не меняется (математики называют это *коммутативностью* сложения):
$$\vec{a}+\vec{b}=\vec{b}+\vec{a}. \tag{7.1}$$
Это легко следует из правила параллелограмма (рис. 7.8). Действительно, какая разница, в каком порядке суммировать векторы $\vec{a}$ и $\vec{b}$, если диагональ параллелограмма всё равно одна и та же?

2. Возникает интересный вопрос: а как сложить *три* вектора? Можно ли определить сумму $\vec{a}+\vec{b}+\vec{c}$? Давайте сделаем это двумя способами: найдём векторы $(\vec{a}+\vec{b})+\vec{c}$ и $\vec{a}+(\vec{b}+\vec{c})$, а затем сравним результаты.

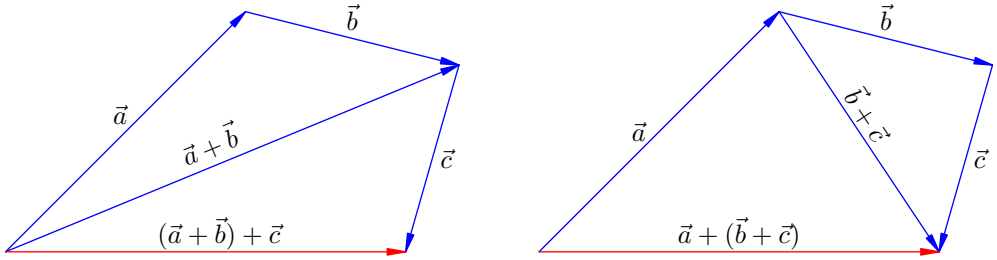


Рис. 7.11. $(\vec{a}+\vec{b})+\vec{c}=\vec{a}+(\vec{b}+\vec{c})$

Как мы видим из рис. 7.11, результаты совпадают! Имеем следующий закон (математики называют его *ассоциативностью*):
$$(\vec{a}+\vec{b})+\vec{c}=\vec{a}+(\vec{b}+\vec{c}). \tag{7.2}$$

Вместе с коммутативностью (7.1) это означает, что сумма $\vec{a}+\vec{b}+\vec{c}$ корректно определена: мы можем складывать данные векторы, комбинируя их как угодно, и результат всегда будет получаться одним и тем же. Например, можно найти нашу сумму так (рис. 7.12):

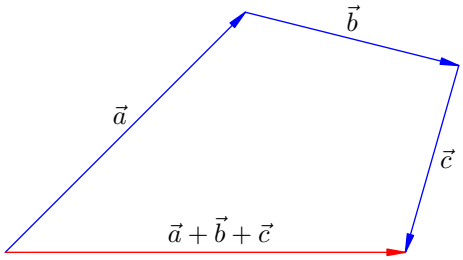


Рис. 7.12. Сумма трёх векторов

Вообще, можно показать, что сумма любого конечного числа векторов не зависит от того, в каком порядке мы складываем векторы. Например, для нахождения суммарного вектора можно воспользоваться *правилом многоугольника* (рис. 7.13; пример для пяти векторов):

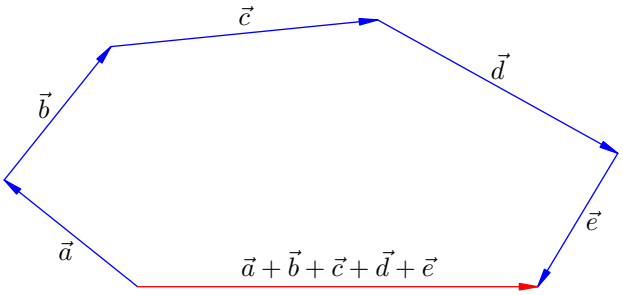


Рис. 7.13. Правило многоугольника

В физических задачах бывает важно углядеть, каким именно образом лучше просуммировать векторы. Вот стандартная ситуация. Пусть длины векторов $\vec{a}$ и $\vec{c}$ равны 1, длина вектора $\vec{b}$ равна 2; угол между $\vec{a}$ и $\vec{b}$ равен 60°, угол между $\vec{b}$ и $\vec{c}$ тоже равен 60° (рис. 7.14). Требуется найти длину вектора $\vec{a}+\vec{b}+\vec{c}$.

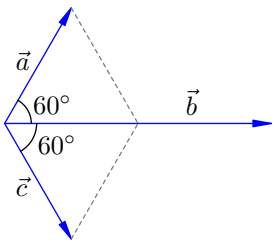


Рис. 7.14. Найти длину вектора $\vec{a}+\vec{b}+\vec{c}$

Искать сначала сумму $\vec{a}+\vec{b}$ и прибавлять потом к ней $\vec{c}$ — можно, но это не самая лучшая идея. Давайте начнём с того, что сложим $\vec{a}$ и $\vec{c}$! Доведите сами до конца это решение, осталось совсем немного. Ответ: 3.

3. Прибавление к вектору нулевого вектора ничего не меняет:

$$\vec{a}+\vec{0}=\vec{a}. \tag{7.3}$$

Это совершенно очевидно, если представить себе такое сложение с точки зрения правила треугольника.

4. Для каждого вектора $\vec{a}$ существует *противоположный* вектор, обозначаемый $-\vec{a}$; сумма вектора и его противоположного равна нулю:

$$\vec{a} + (-\vec{a}) = \vec{0}. \tag{7.4}$$

Противоположный вектор $-\vec{a}$ равен по длине вектору $\vec{a}$ и противоположен ему по направлению (рис. 7.15).

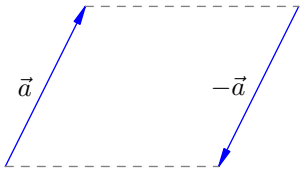


Рис. 7.15. Противоположный вектор

Понятие противоположного вектора вплотную подводит нас к операции вычитания векторов. Эта операция настолько важна в физике, что мы обсудим её отдельно.

7.2.4 Вычитание векторов

Вычитание вектора — это прибавление противоположного вектора. Иными словами, *разностью векторов* $\vec{a}$ *и* $\vec{b}$ *называется сумма* $\vec{a} + (-\vec{b})$.

Такое формальное определение не слишком годится для нас. Мы подойдём к вычитанию векторов с несколько иной стороны.

Рассмотрим три вектора $\vec{a}$, $\vec{b}$, $\vec{c}$ такие, что $\vec{b} + \vec{c} = \vec{a}$ (рис. 7.16).

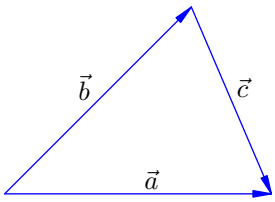


Рис. 7.16. К определению разности векторов

Хорошо было бы перенести вектор $\vec{b}$ вправо со знаком минус, написав $\vec{c} = \vec{a} - \vec{b}$, и сказать при этом, что вектор $\vec{c}$ есть *разность* векторов $\vec{a}$ и $\vec{b}$. Так и делают! Рисунок 7.17 дублирует рис. 7.16 — с тем лишь отличием, что вместо $\vec{c}$ стоит $\vec{a} - \vec{b}$.

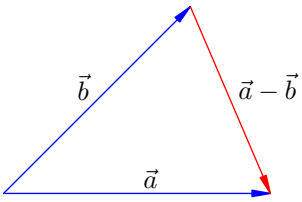


Рис. 7.17. Разность векторов

Давайте будем считать рис. 7.17 определением разности векторов. Итак, чтобы найти векторную разность $\vec{a} - \vec{b}$, мы последовательно делаем следующие шаги.

1. Если начала векторов $\vec{a}$ и $\vec{b}$ находятся в разных точках, то приводим эти векторы к одному началу, параллельно перенося один из векторов.

2. Соединяем концы векторов и «укалываем» тот вектор, *из которого производится вычитание*[2]. В данном случае стрелка направляется к вектору $\vec{a}$.

Разумеется, наглядное определение с помощью рис. 7.17 даёт в результате тот же самый вектор, что и упомянутое выше формальное определение разности $\vec{a}-\vec{b}$ как суммы $\vec{a}+(-\vec{b})$. Попробуйте сами понять, почему так получается!

Разность векторов в физике встречается часто, особенно в механике. Например, ускорение определяется следующим образом:

$$\vec{a}=\frac{\vec{v}-\vec{v}_0}{t}\,.$$

Здесь $\vec{v}_0$ — начальная скорость тела, $\vec{v}$ — конечная скорость, t — время, за которое скорость изменилась от $\vec{v}_0$ до $\vec{v}$. Разность $\Delta\vec{v}=\vec{v}-\vec{v}_0$ называется *изменением скорости*[3].

Таким образом, ускорение есть изменение скорости, делённое на время, за которое это изменение произошло. Об умножении (и тем самым о делении) вектора на скаляр мы поговорим чуть ниже, а пока давайте разберём несложную задачу.

Задача. Тело движется по окружности со скоростью v. Найти модуль изменения скорости тела за четверть периода.

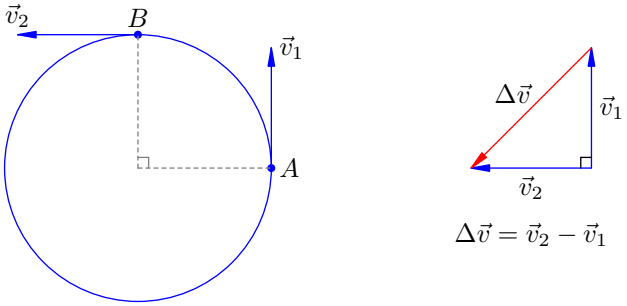


Рис. 7.18. К задаче про изменение скорости

Решение. Пусть в некоторой точке A окружности скорость тела равна $\vec{v}_1$. За четверть периода тело пройдёт четверть окружности и окажется в точке B; пусть скорость тела в этой точке равна $\vec{v}_2$ (рис. 7.18).

Конечно, $|\vec{v}_1|=|\vec{v}_2|=v$, но $\vec{v}_1$ и $\vec{v}_2$ — *разные* векторы (их направления различны), и потому изменение скорости не равно нулю. Смотрим на равнобедренный прямоугольный треугольник, изображённый на рис. 7.18 справа, и по теореме Пифагора заключаем, что $|\Delta\vec{v}|=v\sqrt{2}$.

[2]Можно запомнить это как *правило УУ* — «Уколоть Уменьшаемое».

[3]Вообще, *изменение* какой-либо физической величины — это всегда разность её конечного и начального значений.

7.3 Умножение скаляра на вектор

Векторы можно не только складывать друг с другом, но и умножать на скаляры. Между выражениями «умножение скаляра на вектор» и «умножение вектора на скаляр» никакой принципиальной разницы нет.

При умножении скаляра на вектор получается вектор. Размерность вектора-произведения равна произведению размерностей скаляра и исходного вектора.

Перемножение скаляра и вектора встречается в физике везде, где фигурируют сами векторы. Например, при движении с постоянной скоростью $\vec{v}$ перемещение тела за время t выражается формулой:

$$\vec{s} = \vec{v}t.$$

Импульс тела определяется как произведение массы на скорость:

$$\vec{p} = m\vec{v}.$$

Кстати, импульс не обладает собственной единицей измерения. Размерность импульса есть просто произведение размерностей массы и скорости: кг $\cdot$ м/с.

Произведение массы тела на вектор ускорения присутствует в фундаментальном законе механики — втором законе Ньютона:

$$m\vec{a} = \vec{F}$$

(здесь $\vec{F}$ есть сумма векторов всех сил, приложенных к телу).

Скаляр, умножаемый на вектор, не обязан быть положительным. Например, электрическое поле характеризуется вектором напряжённости $\vec{E}$, который задан в каждой точке поля. Если в данную точку помещён заряд q, то сила, действующая на этот заряд со стороны электрического поля, равна:

$$\vec{F} = q\vec{E}.$$

При этом заряд q может быть как положительным, так и отрицательным.

7.3.1 Что такое умножение скаляра на вектор?

Давайте начнём с примеров и посмотрим на рис. 7.19.

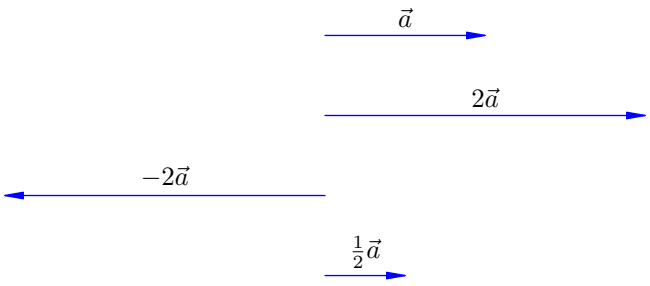


Рис. 7.19. Умножение разных скаляров на вектор $\vec{a}$

В самом верху рисунка расположен вектор $\vec{a}$. Ниже находится вектор $2\vec{a}$: он в два раза длиннее вектора $\vec{a}$ и сонаправлен с ним.

Ещё ниже мы видим вектор $-2\vec{a}$. Он также вдвое длиннее вектора $\vec{a}$, но имеет противоположное направление.

Наконец, в самом низу рисунка расположен вектор $\frac{1}{2}\vec{a}$. Он сонаправлен с вектором $\vec{a}$ и в два раза короче него. Этот вектор можно обозначить также $\vec{a}/2$.

После этого примера операция умножения вектора на скаляр ясна без всяких определений, но строгое определение мы всё же дадим.

Определение. Пусть λ — скаляр. Запись $\vec{b} = \lambda\vec{a}$ означает, что: 1) $|\vec{b}| = |\lambda||\vec{a}|$; 2) вектор $\vec{b}$ сонаправлен с вектором $\vec{a}$ при $\lambda > 0$ и направлен противоположно вектору $\vec{a}$ при $\lambda < 0$ (если же $\lambda = 0$, то из первого пункта следует, что $\vec{b} = \vec{0}$).

Разделить вектор на скаляр $\lambda \neq 0$ — это значит умножить этот вектор на скаляр $1/\lambda$.

Операция умножения скаляра на вектор также обладает хорошими свойствами — привычными с точки зрения алгебры. Рассмотрим их.

7.3.2 Свойства умножения скаляра на вектор

Называть рассматриваемые свойства мы будем так, как это принято в математике. Запоминать эту терминологию сейчас не обязательно, но на первом курсе вы всё равно никуда от неё не денетесь :-)

1. Умножение вектора на единицу не меняет этого вектора:

$$1 \cdot \vec{a} = \vec{a}. \tag{7.5}$$

   Это непосредственно следует из определения.

2. Ассоциативность умножения:

$$\lambda(\mu\vec{a}) = (\lambda\mu)\vec{a}. \tag{7.6}$$

   Иными словами, если сначала умножить вектор $\vec{a}$ на μ, а потом полученный вектор умножить на λ, то это всё равно, что сразу умножить $\vec{a}$ на скаляр $\lambda\mu$.

   Ассоциативность умножения скаляра на вектор иллюстрируется на рис. 7.20. Сначала вектор $\vec{a}$ умножили на 2 и получили вектор $2\vec{a}$. Затем полученный вектор умножили на 3. Получился вектор $3(2\vec{a}) = (3 \cdot 2)\vec{a} = 6\vec{a}$.

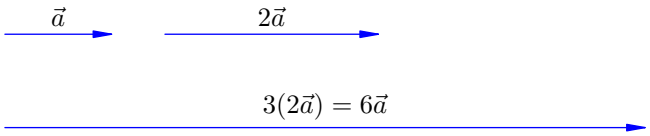


Рис. 7.20. Ассоциативность умножения скаляра на вектор

   Давайте возьмём пример из физики. Если тело массы m, движущееся со скоростью $\vec{v}$, налетает на покоящееся тело массы M и слипается с ним (так называемый неупругий удар), то из закона сохранения импульса легко следует, что после удара слипшиеся тела будут двигаться со скоростью

$$\frac{m\vec{v}}{m+M}.$$

   Благодаря ассоциативности умножения мы можем понимать эту запись как угодно:

   - либо сначала умножили вектор $\vec{v}$ на скаляр m и затем поделили полученный вектор $m\vec{v}$ на скаляр $m + M$;
   - либо сначала поделили m на $m+M$ и потом умножили полученное число $m/(m+M)$ на вектор $\vec{v}$.

   Результат в обоих случаях будет одним и тем же.

3. Дистрибутивность умножения относительно сложения скаляров:

$$(\lambda + \mu)\vec{a} = \lambda\vec{a} + \mu\vec{a}. \tag{7.7}$$

Попросту говоря, мы можем раскрывать скобки (если читать данное равенство слева направо) или выносить за скобки общий векторный множитель (если — справа налево).

Рисунок 7.21 иллюстрирует это свойство. Пусть дан вектор $\vec{a}$. Берём вектор $2\vec{a}$ и в его конец помещаем начало вектора $3\vec{a}$. Складывая их, получаем вектор $2\vec{a}+3\vec{a} = 5\vec{a} = (2+3)\vec{a}$.

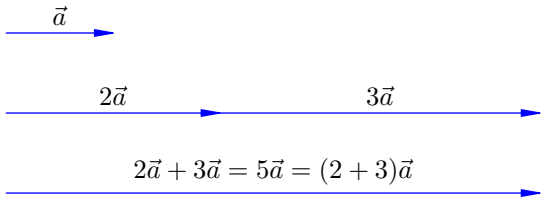


Рис. 7.21. Дистрибутивность относительно сложения скаляров

Вот соответствующий физический пример. Пусть два тела массами m_1 и m_2 движутся с одинаковой скоростью $\vec{v}$. Импульс первого тела равен $\vec{p}_1 = m_1\vec{v}$, импульс второго тела равен $\vec{p}_2 = m_2\vec{v}$. Импульс системы этих тел равен векторной сумме импульсов каждого тела:

$$\vec{p} = \vec{p}_1 + \vec{p}_2 = m_1\vec{v} + m_2\vec{v}.$$

Согласно свойству (7.7) общий векторный множитель можно вынести за скобки:

$$\vec{p} = (m_1 + m_2)\vec{v}.$$

4. Дистрибутивность относительно сложения векторов:

$$\lambda(\vec{a} + \vec{b}) = \lambda\vec{a} + \lambda\vec{b}. \tag{7.8}$$

И в таком случае, как видим, мы можем раскрывать скобки или, наоборот, выносить за скобки общий скалярный множитель.

Иллюстрацией этого свойства служит рис. 7.22. Пусть $\vec{c} = \vec{a} + \vec{b}$ (левая часть рисунка). Тогда $2\vec{c} = 2(\vec{a}+\vec{b})$. Но из правой части рисунка мы видим, что $2\vec{c} = 2\vec{a}+2\vec{b}$. Следовательно, $2(\vec{a} + \vec{b}) = 2\vec{a} + 2\vec{b}$.

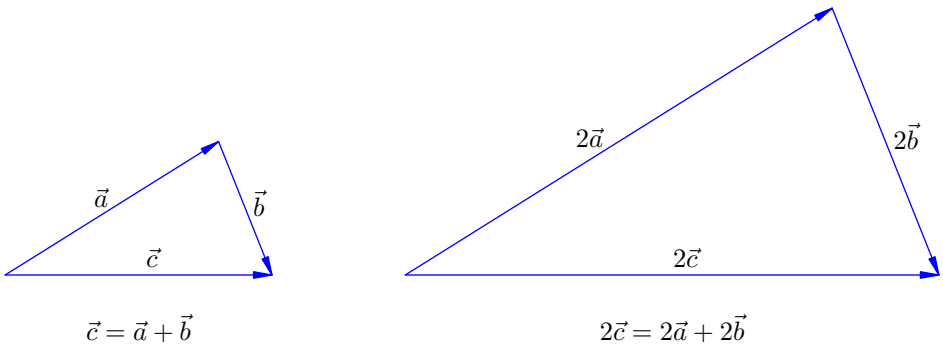


Рис. 7.22. Дистрибутивность относительно сложения векторов

И здесь рассмотрим пример из физики. Пусть имеется заряд q. Расположенные неподалёку заряды q_1 и q_2 создают в точке нахождения заряда q электрические поля, напряжённости которых равны $\vec{E}_1$ и $\vec{E}_2$ соответственно. Какая сила будет действовать на заряд q?

Со стороны заряда q_1 на заряд q действует сила $\vec{F}_1 = q\vec{E}_1$. Со стороны заряда q_2 на заряд q действует сила $\vec{F}_2 = q\vec{E}_2$. Искомая сила $\vec{F}$ является равнодействующей сил $\vec{F}_1$ и $\vec{F}_2$:

$$\vec{F} = \vec{F}_1 + \vec{F}_2 = q\vec{E}_1 + q\vec{E}_2.$$

Согласно свойству (7.8) общий скалярный множитель выносится за скобки:

$$\vec{F} = q(\vec{E}_1 + \vec{E}_2).$$

Свойства (7.1)–(7.8) позволяют обращаться с векторными выражениями по хорошо знакомым алгебраическим правилам: раскрывать скобки, приводить подобные, переносить слагаемые в другую часть равенства с противоположным знаком... Вы наверняка это знаете и используете; но вы могли не задумываться о том, что эти вещи *не так уж очевидны*. Одна из целей данного приложения — приоткрыть занавес и прояснить, почему к выражениям с векторами во многом применимы те же правила алгебры, что и к обычным буквенным выражениям.

7.4 Угол между векторами

Выше мы рассмотрели две операции над векторами: сложение векторов и умножение скаляра на вектор. Если бы при работе с векторами нам были доступны лишь эти две операции, наши возможности в физике и геометрии оказались бы весьма ограниченными.

Но, к счастью, это не так. Мы можем вдобавок ввести понятие угла между векторами и с помощью него определить новые операции, которые хорошо согласуются с уже изученными. Благодаря столь богатому набору операций векторы становятся мощным инструментом исследования физического мира.

7.4.1 Что такое угол между векторами?

Можно сказать так: *угол между векторами — это угол между их направлениями.* Это, конечно, не очень строго, но зато интуитивно понятно.

Но мы не будем гнаться за строгим определением, а просто сделаем рисунок — он скажет лучше всяких слов.

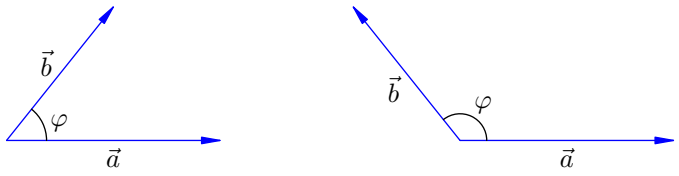


Рис. 7.23. Угол между векторами

Как видим из рис. 7.23, угол φ между векторами $\vec{a}$ и $\vec{b}$ — это угол, образованный лучами, идущими вдоль этих векторов из общего начала. Угол между векторами принимает значения от 0 до 180°.

7.4.2 Угол между вектором и осью

Ключевую роль в дальнейшем будет играть понятие оси. *Ось — это прямая, снабжённая направлением* (рис. 7.24).

Рис. 7.24. Ось

Вы давно привыкли к координатным осям, поэтому данное понятие вам хорошо знакомо.

Угол между вектором и осью определяется точно так же, как и угол между векторами. *Угол между вектором и осью — это угол между их направлениями* (рис. 7.25). Так, вектор $\vec{a}$ образует с осью X острый угол α, а вектор $\vec{b}$ образует с осью X тупой угол β.

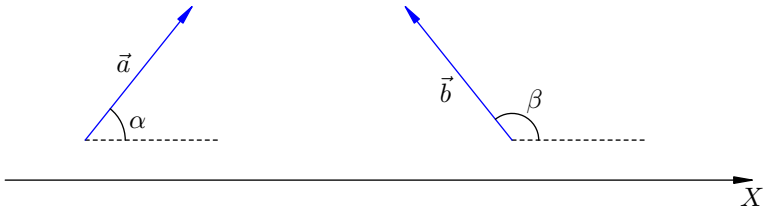


Рис. 7.25. Угол между вектором и осью

Угол между вектором и осью также принимает значения от 0 до 180°.

7.5 Проекция вектора на ось

Теперь мы готовы ввести важнейшее понятие проекции вектора на ось. Оно постоянно используется при решении физических задач.

7.5.1 Что такое проекция вектора на ось?

Пусть даны вектор $\vec{a}$ и ось X. Предполагается, что на оси X имеется масштаб, позволяющий измерять длины отрезков и присваивать им размерность вектора $\vec{a}$.

Из начала и конца вектора $\vec{a}$ опустим перпендикуляры на ось X; пусть A и B — основания этих перпендикуляров (рис. 7.26). Длину отрезка AB обозначим $|AB|$.

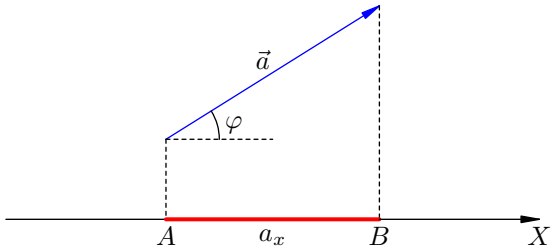


Рис. 7.26. Проекция вектора на ось

Определение. Проекция a_x вектора $\vec{a}$ на ось X равна длине отрезка AB, взятой со знаком плюс, если угол φ между вектором $\vec{a}$ и осью X является острым, и взятой соответственно со знаком минус, если φ тупой (или развёрнутый). Если угол φ прямой, то $a_x = 0$.

Короче говоря, имеем следующую формулу:

$$a_x = \begin{cases} |AB|, & \text{если } \varphi < 90^\circ; \\ -|AB|, & \text{если } \varphi > 90^\circ; \\ 0, & \text{если } \varphi = 90^\circ. \end{cases} \tag{7.9}$$

Рисунок 7.27 иллюстрирует все эти возможности.

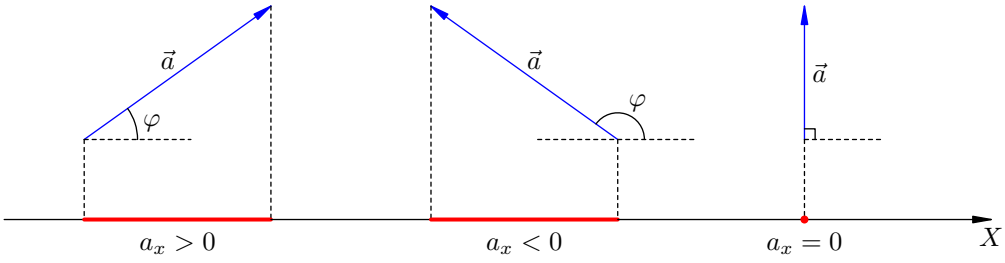


Рис. 7.27. Проекция вектора на ось. Примеры

Глядя на рис. 7.27, нетрудно сообразить, что все эти три случая — два красных отрезка и точка — охватываются одной-единственной формулой, и очень простой!

Следствие 1. Если угол между вектором $\vec{a}$ и осью X равен φ, то проекция a_x вычисляется по формуле:

$$a_x = a\cos\varphi. \tag{7.10}$$

Здесь, как обычно, $a = |\vec{a}|$ — модуль вектора $\vec{a}$.

Действительно, если $\varphi < 90^\circ$, то формула (7.10) даёт длину левого красного отрезка на рис. 7.27.

Если $\varphi > 90°$, то, переходя в средней части рис. 7.27 к углу, смежному с углом φ, мы видим, что формула (7.10) даёт длину среднего красного отрезка со знаком минус (за счёт отрицательности косинуса), что нам как раз и нужно.

Наконец, если $\varphi = 90°$, то формула (7.10) даёт $a_x = 0$, поскольку косинус прямого угла равен нулю. Именно так и должно быть (правая часть рисунка).

Предположим теперь, что на оси X задано вдобавок начало отсчёта, так что она является привычной координатной осью. Тогда имеем ещё одну формулу для проекции a_x, которая также содержит в «заархивированном» виде все три случая рисунка 7.27.

Следствие 2. Пусть x_1 и x_2 — координаты соответственно начала и конца вектора $\vec{a}$. Тогда проекция a_x вычисляется по формуле:

$$a_x = x_2 - x_1. \tag{7.11}$$

Действительно, посмотрим на рис. 7.28. Это случай положительной проекции. Из рисунка очевидно, что разность $x_2 - x_1$ равна длине красного отрезка, а эта длина в данном случае как раз и есть проекция a_x.

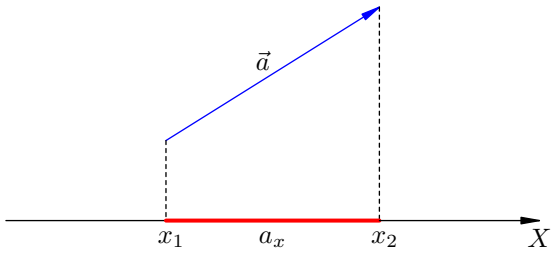


Рис. 7.28. Проекция вектора на ось. К следствию 2

Что будет в оставшихся двух случаях ($a_x < 0$ и $a_x = 0$)? Убедитесь, пожалуйста, самостоятельно, что формула (7.11) и для них остаётся справедливой.

7.5.2 Свойства проектирования вектора на ось

Операция проектирования вектора на ось замечательным образом согласована с операциями сложения векторов и умножения скаляра на вектор. А именно, какова бы ни была ось X, имеют место следующие два свойства проектирования.

1. *Проекция вектора $\vec{a} + \vec{b}$ на ось X равна $a_x + b_x$.*

   Краткая словесная формулировка: *проекция суммы векторов равна сумме их проекций.* Это справедливо для суммы любого числа векторов, не только двух.

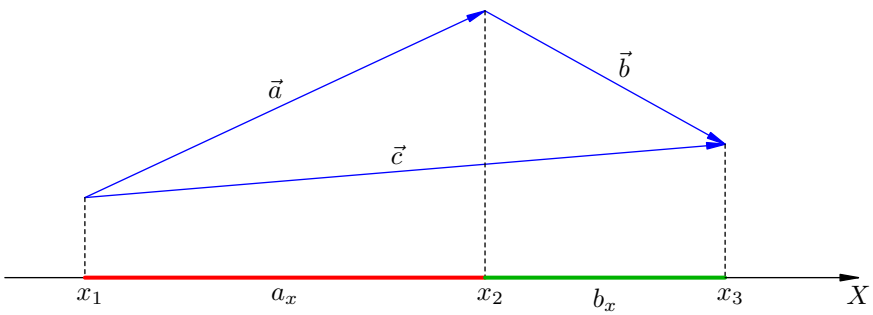


Рис. 7.29. $\vec{c} = \vec{a} + \vec{b} \Rightarrow c_x = a_x + b_x$

   Прежде всего проиллюстрируем данное утверждение на рисунке. Поместим начало вектора $\vec{b}$ в конец вектора $\vec{a}$, и пусть $\vec{c} = \vec{a} + \vec{b}$ (рис. 7.29).

На данном рисунке хорошо видно, что проекция c_x равна сумме длин красного и зелёного отрезков, то есть как раз $a_x + b_x$.

Правда, рис. 7.29 сделан для случая $a_x > 0$ и $b_x > 0$. Чтобы доказать наше утверждение сразу для всех возможных значений проекций a_x и b_x, мы проведём следующее универсальное рассуждение, опирающееся на формулу (7.11).

Итак, пусть векторы $\vec{a}$ и $\vec{b}$ расположены произвольным образом. Снова совместим начало вектора $\vec{b}$ с концом вектора $\vec{a}$ и обозначим $\vec{c} = \vec{a} + \vec{b}$. Пусть:

- x_1 — координата начала вектора $\vec{a}$ и одновременно начала вектора $\vec{c}$;
- x_2 — координата конца вектора $\vec{a}$ и одновременно начала вектора $\vec{b}$;
- x_3 — координата конца вектора $\vec{b}$ и одновременно конца вектора $\vec{c}$.

Эти обозначения также присутствуют на рис. 7.29.

В силу формулы (7.11) имеем: $a_x = x_2 - x_1$, $b_x = x_3 - x_2$, $c_x = x_3 - x_1$. Теперь легко видеть, что:

$$a_x + b_x = (x_2 - x_1) + (x_3 - x_2) = x_3 - x_1 = c_x.$$

Наше первое свойство проектирования тем самым доказано.

2. *Проекция вектора $\lambda\vec{a}$ на ось X равна λa_x.*

   Словесная формулировка: *проекция произведения скаляра на вектор равна произведению скаляра на проекцию вектора.*

   Снова начнём с иллюстрации. В левой части рисунка 7.30 изображён вектор $\vec{a}$ с положительной проекцией a_x.

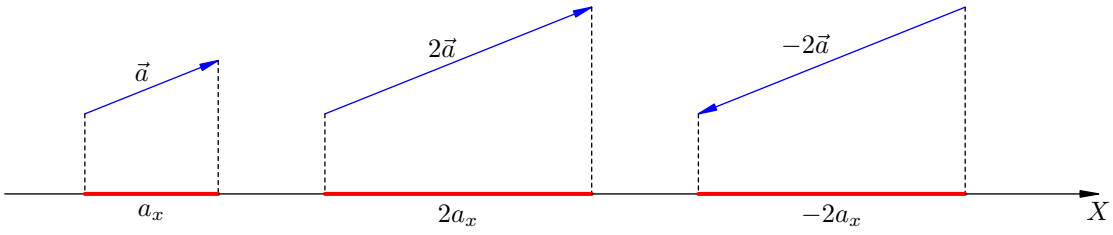


Рис. 7.30. Проекция вектора $\lambda\vec{a}$ равна λa_x

Если умножить вектор $\vec{a}$ на 2, то его длина увеличится в два раза, проекция вектора также увеличится вдвое (сохраняя знак) и станет равна $2a_x$.

Если умножить вектор $\vec{a}$ на -2, то его длина опять-таки увеличится в два раза, но направление изменится на противоположное. Проекция изменит знак и станет равна $-2a_x$.

Тем самым суть второго свойства ясна, и теперь можно дать строгое доказательство. Итак, пусть $\vec{b} = \lambda\vec{a}$. Мы ходим доказать, что $b_x = \lambda a_x$.

Воспользуемся для этого формулой (7.10). Имеем:

$$a_x = a\cos\varphi, \quad b_x = b\cos\theta,$$

где φ — угол между вектором $\vec{a}$ и осью X, а θ — угол между вектором $\vec{b}$ и осью X. Кроме того, в силу определения умножения скаляра на вектор:

$$b = |\lambda|a.$$

Таким образом:

$$b_x = |\lambda| a \cos\theta.$$

Если $\lambda > 0$, то $|\lambda| = \lambda$; в этом случае вектор $\vec{b}$ сонаправлен с вектором $\vec{a}$, и потому $\theta = \varphi$. Имеем:

$$b_x = \lambda a \cos\varphi = \lambda a_x.$$

Если $\lambda < 0$, то $|\lambda| = -\lambda$; в этом случае вектор $\vec{b}$ противоположен по направлению вектору $\vec{a}$. Нетрудно сообразить при этом, что $\theta = \pi - \varphi$ (например, если φ острый, то θ есть смежный с ним тупой, и наоборот). Имеем тогда:

$$b_x = (-\lambda) a \cos(\pi - \varphi) = (-\lambda) a (-\cos\varphi) = \lambda a \cos\varphi = \lambda a_x.$$

Ну а в тривиальном случае $\lambda = 0$ доказывать нечего: тогда $\vec{b} = \vec{0}$ и $b_x = 0 = 0 \cdot a_x = \lambda a_x$. Итак, во всех случаях получается нужное соотношение, и тем самым второе свойство проектирования полностью доказано.

7.5.3 Операция проектирования в физике

Доказанные свойства операции проектирования очень важны для нас. В механике, например, мы будем пользоваться ими на каждом шагу.

Так, решение многих задач по динамике начинается с записи второго закона Ньютона в векторной форме. Возьмём, к примеру, маятник массы m, подвешенный на нити. Для маятника второй закон Ньютона будет иметь вид:

$$m\vec{a} = m\vec{g} + \vec{T} + \vec{f}, \tag{7.12}$$

где $\vec{T}$ — сила упругости нити, $\vec{f}$ — сила сопротивления воздуха.

Записав второй закон Ньютона в векторной форме, мы переходим к его проектированию на подходящие оси. Берём равенство (7.12) и проектируем на ось X:

$$ma_x = mg_x + T_x + f_x. \tag{7.13}$$

При переходе от векторного равенства (7.12) к скалярному равенству (7.13) использованы оба свойства проектирования! А именно, благодаря свойству 1 мы записали проекцию суммы векторов как сумму их проекций; в силу же свойства 2 мы смогли записать проекции векторов $m\vec{a}$ и $m\vec{g}$ в виде ma_x и mg_x.

Таким образом, оба свойства операции проектирования обеспечивают переход от векторных равенств к скалярным, и переход этот можно выполнять формально и не задумываясь: *отбрасываем стрелки в обозначениях векторов и ставим вместо них индексы проекций.* Именно так выглядит переход от уравнения (7.12) к уравнению (7.13).

7.6 Векторы и координаты на плоскости

Чрезвычайно важным применением трёх рассмотренных нами операций над векторами (сложения, умножения на скаляр и проектирования на ось) является *разложение вектора по базису* декартовой прямоугольной системы координат.

7.6.1 Разложение вектора по базису

Рассмотрим систему координат OXY (рис. 7.31). Оси X и Y снабжены *единичными векторами* $\vec{i}$ и $\vec{j}$ — длины этих векторов равны единице, причём этой самой единице не приписывается никакая размерность. Векторы $\vec{i}$ и $\vec{j}$ называются *базисом* системы координат OXY (или *базисом на плоскости*).

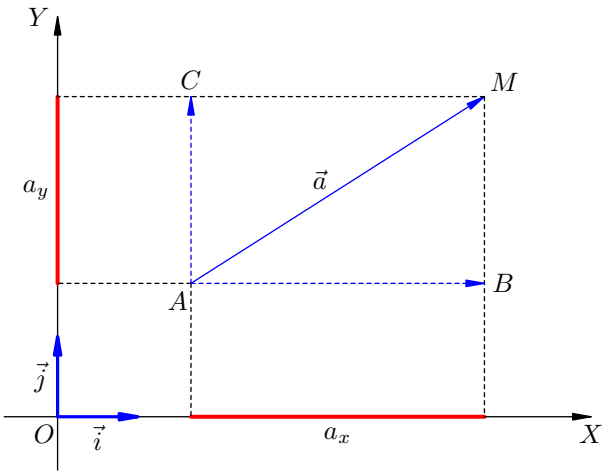


Рис. 7.31. Разложение вектора по базису на плоскости

Пусть вектор $\vec{a} = \overrightarrow{AM}$ имеет начало в точке A. Опустим из начала и конца вектора $\vec{a}$ перпендикуляры на координатные оси и проведём векторы $\overrightarrow{AB}$ и $\overrightarrow{AC}$ параллельно осям OX и OY соответственно. Тогда $\vec{a} = \overrightarrow{AB} + \overrightarrow{AC}$.

Ясно, что $\overrightarrow{AB} = a_x\vec{i}$ при любом знаке проекции a_x. Аналогично, $\overrightarrow{AC} = a_y\vec{j}$ при любом знаке проекции a_y. Следовательно,

$$\vec{a} = a_x\vec{i} + a_y\vec{j}. \tag{7.14}$$

Это и есть разложение вектора $\vec{a}$ по базису $\vec{i}$, $\vec{j}$ системы координат OXY. Проекции a_x и a_y называются также *координатами* вектора $\vec{a}$ в базисе $\vec{i}$, $\vec{j}$.

7.6.2 Нахождение модуля вектора по его проекциям

В физике, как правило, мы находим проекции интересующего нас вектора по отдельности, решая для этих проекций соответствующие уравнения. Дальнейший процесс «сборки» вектора по его проекциям никакого труда не представляет.

Из треугольника ABM (рис. 7.31) по теореме Пифагора имеем:

$$AM = \sqrt{AB^2 + BM^2}.$$

Но $AM = a$ — это модуль вектора $\vec{a}$. Кроме того, при любых знаках проекций a_x и a_y справедливы равенства $AB^2 = a_x^2$ и $BM^2 = a_y^2$. Следовательно:

$$a = \sqrt{a_x^2 + a_y^2}\,. \tag{7.15}$$

Формула (7.15) часто используется в физических задачах. Вот типичный пример.

Задача. Тело брошено горизонтально со скоростью $\vec{v}_0$. Найти скорость тела спустя время t. Под каким углом к горизонту направлена эта скорость?

Решение. Имеем: $\vec{v} = \vec{v}_0 + \vec{g}t$. Направляя ось X горизонтально, а ось Y — вертикально вверх, для проекций скорости получим:

$$v_x = v_0, \quad v_y = -gt.$$

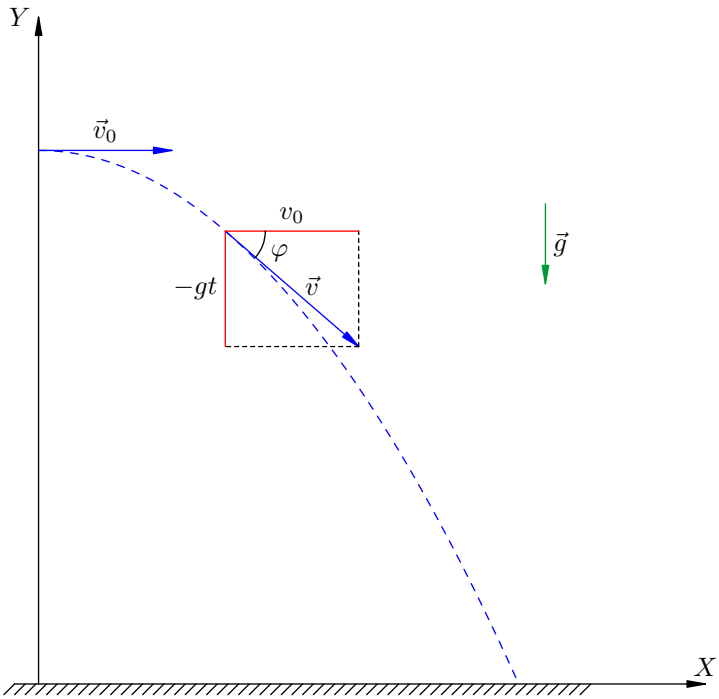


Рис. 7.32. К задаче о горизонтальном броске

Теперь формула (7.15) даёт:

$$v = \sqrt{v_0^2 + g^2t^2}\,.$$

Это, впрочем, очевидно из рис. 7.32 и непосредственно по теореме Пифагора.

Для искомого угла φ имеем:

$$\operatorname{tg}\varphi = \frac{|v_y|}{v_x} = \frac{gt}{v_0}\,.$$

Отсюда

$$\varphi = \operatorname{arctg}\left(\frac{gt}{v_0}\right).$$

7.7 Векторы и координаты в пространстве

Мы продолжаем изучать координатное представление векторов — но теперь переходим от плоскости к трёхмерному пространству.

7.7.1 Разложение вектора по базису

Рассмотрим систему координат $OXYZ$ (рис. 7.33). Оси X, Y и Z снабжены единичными векторами $\vec{i}$, $\vec{j}$ и $\vec{k}$, которые составляют *базис* данной системы координат (или, как ещё говорят, *базис в пространстве*).

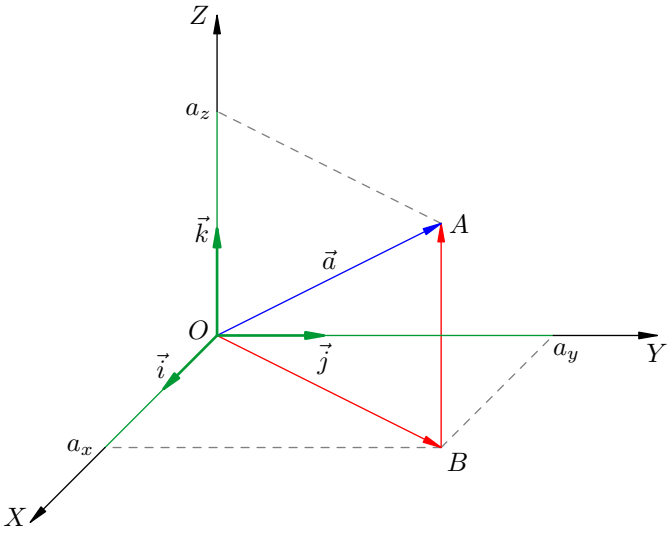


Рис. 7.33. Разложение вектора по базису в пространстве

Рассмотрим вектор $\vec{a} = \overrightarrow{OA}$. Пусть точка B служит основанием перпендикуляра, опущенного из точки A на плоскость OXY.

Из точки A опустим перпендикуляр на ось Z (он будет параллелен OB); получим зелёный отрезок a_z, представляющий собой проекцию вектора $\vec{a}$ на ось Z.

Из точки B опустим перпендикуляры на оси X и Y; получим зелёные отрезки a_x и a_y, представляющие собой проекции вектора $\vec{a}$ на оси X и Y.

Имеем: $\vec{a} = \overrightarrow{OB} + \overrightarrow{BA}$. Вектор $\overrightarrow{OB}$ лежит в плоскости OXY и раскладывается по базису на плоскости: $\overrightarrow{OB} = a_x\vec{i} + a_y\vec{j}$. Вектор $\overrightarrow{BA}$ параллелен оси Z, так что $\overrightarrow{BA} = a_z\vec{k}$. Всё это вместе даёт:

$$\vec{a} = a_x\vec{i} + a_y\vec{j} + a_z\vec{k}. \tag{7.16}$$

Данное равенство и есть разложение вектора $\vec{a}$ по базису $\vec{i}$, $\vec{j}$, $\vec{k}$. Оно служит пространственным аналогом формулы (7.14).

Числа a_x, a_y, a_z называются *координатами* вектора $\vec{a}$ в базисе $\vec{i}$, $\vec{j}$, $\vec{k}$.

7.7.2 Нахождение модуля вектора по его проекциям

Из треугольника OAB по теореме Пифагора имеем:

$$a = OA = \sqrt{OB^2 + BA^2}\,.$$

Заметим теперь, что $OB^2 = a_x^2 + a_y^2$ и $BA^2 = a_z^2$. Получаем:

$$a = \sqrt{a_x^2 + a_y^2 + a_z^2}\,. \tag{7.17}$$

Данная формула является пространственным аналогом формулы (7.15).

7.8 Скалярное произведение векторов

Для начала давайте вспомним, как в механике определяется понятие работы силы. Рассмотрим тело, находящееся на горизонтальной поверхности (рис. 7.34). Пусть на тело действует сила $\vec{F}$ под углом α к горизонту, и под действием этой силы тело совершило перемещение $\vec{s}$.

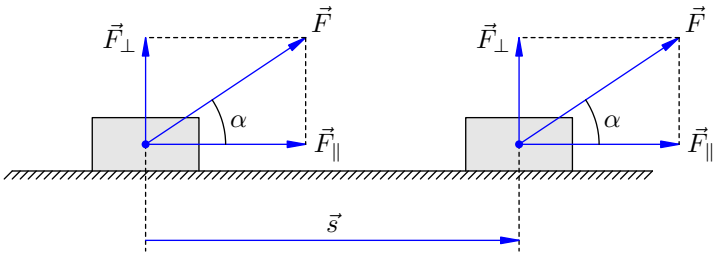


Рис. 7.34. К определению работы силы

Разложим силу $\vec{F}$ на две составляющих: $\vec{F} = \vec{F}_\parallel + \vec{F}_\perp$; сила $\vec{F}_\parallel$ параллельна вектору перемещения, а сила $\vec{F}_\perp$ перпендикулярна ему. Работой A силы $\vec{F}$ называется в данном случае произведение модуля параллельной составляющей на модуль перемещения:

$$A = F_\parallel s.$$

Но $F_\parallel = F\cos\alpha$, поэтому

$$A = Fs\cos\alpha. \tag{7.18}$$

Формула (7.18) как раз и является определением физической величины, называемой работой. Это определение справедливо для любого угла α между силой и перемещением. Если, например, $\alpha > 90°$, то работа отрицательна (за счёт отрицательности косинуса). Если сила перпендикулярна перемещению, то работа этой силы равна нулю.

Заметим ещё, что в силу формулы (7.10) величина $F\cos\alpha$ есть F_s — проекция вектора $\vec{F}$ на ось вектора[4] $\vec{s}$. Поэтому

$$A = F_s s.$$

Эта формула также справедлива для любого угла между векторами $\vec{F}$ и $\vec{s}$.

7.8.1 Что такое скалярное произведение?

Пусть даны векторы $\vec{a}$ и $\vec{b}$. Угол между этими векторами обозначим φ (рис. 7.35).

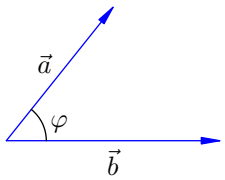


Рис. 7.35. К определению скалярного произведения

Определение. Скалярное произведение векторов $\vec{a}$ и $\vec{b}$ (обозначается $\vec{a}\cdot\vec{b}$) — это скаляр, равный произведению модулей векторов на косинус угла между ними:

$$\vec{a}\cdot\vec{b} = ab\cos\varphi. \tag{7.19}$$

В силу формулы (7.10) величина $a\cos\varphi$ есть a_b — проекция вектора $\vec{a}$ на ось вектора $\vec{b}$. Поэтому имеем:

$$\vec{a}\cdot\vec{b} = a_b b. \tag{7.20}$$

[4] *Осью вектора* называется ось, направление которой совпадает с направлением данного вектора.

7.8.2 Свойства скалярного произведения

1. Скалярное умножение коммутативно:

$$\vec{a} \cdot \vec{b} = \vec{b} \cdot \vec{a}. \tag{7.21}$$

Это очевидно из формулы (7.19). Ведь если поменять местами векторы-сомножители, то угол между ними не изменится.

2. При скалярном умножении вектора на самого себя получается квадрат его модуля:

$$\vec{a} \cdot \vec{a} = a^2. \tag{7.22}$$

Это также очевидно из формулы (7.19) — вектор $\vec{a}$ образует сам с собой нулевой угол, и потому $\vec{a} \cdot \vec{a} = a \cdot a \cos 0 = a^2$.

Кстати, величина $\vec{a} \cdot \vec{a}$ называется *скалярным квадратом* вектора $\vec{a}$ и обозначается $\vec{a}^2$. Таким образом, $\vec{a}^2 = a^2$.

3. Скалярное произведение ненулевых векторов равно нулю тогда и только тогда, когда угол между векторами прямой.

Это очевидно. Раз $a, b \neq 0$, то

$$\vec{a} \cdot \vec{b} = 0 \Leftrightarrow \cos\varphi = 0 \Leftrightarrow \varphi = 90^\circ.$$

4. Скалярное произведение ассоциативно при умножении на скаляр:

$$(\lambda\vec{a}) \cdot \vec{b} = \lambda(\vec{a} \cdot \vec{b}). \tag{7.23}$$

Таким образом, не играет роли, в какой последовательности выполнять указанные операции.

Для доказательства используем формулу (7.20). Согласно этой формуле $\lambda(\vec{a} \cdot \vec{b}) = \lambda a_b b$.

Обозначим $\vec{c} = \lambda\vec{a}$. Тогда в силу свойства 2 операции проектирования имеем: $c_b = \lambda a_b$. Ну а теперь снова используем формулу (7.20):

$$(\lambda\vec{a}) \cdot \vec{b} = \vec{c} \cdot \vec{b} = c_b b = \lambda a_b b.$$

Итак, обе части доказываемого соотношения равны одной и той же величине $\lambda a_b b$, так что наша ассоциативность действительно имеет место. Поэтому скобки в таких выражениях можно опускать и писать просто $\lambda\vec{a} \cdot \vec{b}$.

Обратите внимание, что скалярное произведение не обладает «полноценной» ассоциативностью: $(\vec{a} \cdot \vec{b}) \cdot \vec{c} \neq \vec{a} \cdot (\vec{b} \cdot \vec{c})$. (Сможете сами придумать пример?) Следовательно, нельзя записать выражение $\vec{a} \cdot \vec{b} \cdot \vec{c}$ — оно не является корректно определённым, поскольку его значение зависит от порядка выполнения умножений.

5. Скалярное произведение дистрибутивно:

$$(\vec{a} + \vec{b}) \cdot \vec{c} = \vec{a} \cdot \vec{c} + \vec{b} \cdot \vec{c}. \tag{7.24}$$

Для доказательства обозначим $\vec{u} = \vec{a} + \vec{b}$. Согласно свойству 1 проектирования вектора на ось проекция вектора $\vec{u}$ на ось вектора $\vec{c}$ равна сумме проекций: $u_c = a_c + b_c$. Тогда имеем:

$$(\vec{a} + \vec{b}) \cdot \vec{c} = \vec{u} \cdot \vec{c} = u_c c = (a_c + b_c)c = a_c c + b_c c = \vec{a} \cdot \vec{c} + \vec{b} \cdot \vec{c},$$

что и требовалось. Следовательно, в таких ситуациях мы можем обычным образом раскрывать скобки и выносить за скобки общий векторный множитель.

7.8.3 Скалярное произведение в физике

Подчеркнём ещё раз, что скалярное произведение — это не вектор, а скаляр. Иными словами, в физике скалярное произведение есть число, обладающее размерностью. Размерность скалярного произведения равно произведению размерностей векторов-сомножителей.

Из определения работы — формулы (7.18) — мы видим теперь, что работа есть скалярное произведение векторов силы и перемещения:

$$A = \vec{F} \cdot \vec{s}. \tag{7.25}$$

Если тело движется равномерно и прямолинейно, то есть с постоянной скоростью $\vec{v}$, то $\vec{s} = \vec{v}t$. Подставляя это в формулу (7.25), получим:

$$A = \vec{F} \cdot \vec{v}t. \tag{7.26}$$

Благодаря ассоциативности (7.23) при умножении на скаляр нам всё равно, в каком порядке перемножаются эти множители. Удобно воспринять формулу (7.26) как $A = (\vec{F} \cdot \vec{v})t$ и поделить обе части на t. Получим формулу для мощности:

$$P = \frac{A}{t} = \vec{F} \cdot \vec{v}.$$

Далее, пусть на тело действуют две силы: $\vec{F}_1$ и $\vec{F}_2$. Эти силы совершают соответственно работы:

$$A_1 = \vec{F}_1 \cdot \vec{s}, \quad A_2 = \vec{F}_2 \cdot \vec{s}.$$

Какую работу совершает равнодействующая $\vec{F}$ этих сил? Пользуемся дистрибутивностью скалярного произведения(7.24):

$$A = \vec{F} \cdot \vec{s} = (\vec{F}_1 + \vec{F}_2) \cdot \vec{s} = \vec{F}_1 \cdot \vec{s} + \vec{F}_2 \cdot \vec{s} = A_1 + A_2.$$

Вывод: работа равнодействующей силы равна сумме работ каждой из сил в отдельности. Иными словами, приложенные к телу силы складываются векторно, а их работы — алгебраически[5].

7.8.4 Вычисление скалярного произведения в координатах

Если на плоскости или в пространстве имеется прямоугольная система координат, то возникает замечательно простая формула для нахождения скалярного произведения векторов через их координаты.

1. Плоскость. Предположим, что на плоскости задана прямоугольная система координат OXY (как показано на рис. 7.31). Векторы $\vec{i}$ и $\vec{j}$ — единичные векторы координатных осей.

Векторы $\vec{a}$ и $\vec{b}$ расположены на этой плоскости. Пусть, как обычно, a_x и a_y — проекции вектора $\vec{a}$ на координатные оси (или, что то же самое, координаты вектора $\vec{a}$ в базисе $\vec{i}$, $\vec{j}$). Аналогичный смысл имеют обозначения b_x и b_y.

Теорема. Скалярное произведение векторов $\vec{a}$ и $\vec{b}$, расположенных на плоскости, вычисляется через их координаты следующим образом:

$$\vec{a} \cdot \vec{b} = a_x b_x + a_y b_y. \tag{7.27}$$

Для доказательства используем формулу (7.14) разложения вектора по базису:

$$\vec{a} = a_x\vec{i} + a_y\vec{j}, \quad \vec{b} = b_x\vec{i} + b_y\vec{j}.$$

[5]Проявление этого факта мы встречаем в электростатике: напряжённости полей, создаваемых в данной точке разными зарядами, складываются векторно, а потенциалы этих полей — алгебраически.

Подставляем эти разложения в качестве сомножителей в скалярное произведение векторов $\vec{a}$ и $\vec{b}$, после чего пользуемся дистрибутивностью (7.24), обычным образом раскрывая скобки:

$$\vec{a}\cdot\vec{b}=(a_x\vec{i}+a_y\vec{j})\cdot(b_x\vec{i}+b_y\vec{j})=a_xb_x\vec{i}\cdot\vec{i}+a_xb_y\vec{i}\cdot\vec{j}+a_yb_x\vec{j}\cdot\vec{i}+a_yb_y\vec{j}\cdot\vec{j}.$$

Остаётся заметить, что $\vec{i}\cdot\vec{i}=\vec{j}\cdot\vec{j}=1$, $\vec{i}\cdot\vec{j}=0$, и потому

$$\vec{a}\cdot\vec{b}=a_xb_x+a_yb_y.$$

Теорема доказана.

Если в формуле (7.27) положить $\vec{b}=\vec{a}$, то получим:

$$\vec{a}\cdot\vec{a}=a_x^2+a_y^2.$$

Но $\vec{a}\cdot\vec{a}$, как мы знаем из свойства (7.22), равно a^2. Поэтому

$$a^2=a_x^2+a_y^2,$$

или

$$a=\sqrt{a_x^2+a_y^2}\,.$$

Мы снова получили формулу (7.15), но на сей раз вышли на неё со стороны скалярного произведения.

2. Пространство. Предположим, что в пространстве задана прямоугольная система координат $OXYZ$ с базисом $\vec{i}$, $\vec{j}$, $\vec{k}$ (как показано на рис. 7.33).

Возьмём два произвольных вектора $\vec{a}$ и $\vec{b}$, расположенных в пространстве. Вектор $\vec{a}$ имеет координаты a_x, a_y, a_z. Вектор $\vec{b}$ имеет координаты b_x, b_y, b_z.

Теорема. Скалярное произведение векторов $\vec{a}$ и $\vec{b}$, расположенных в пространстве, вычисляется через их координаты следующим образом:

$$\vec{a}\cdot\vec{b}=a_xb_x+a_yb_y+a_zb_z. \tag{7.28}$$

Доказательство совершенно аналогично. Используем формулу (7.16) разложения вектора по базису:

$$\vec{a}=a_x\vec{i}+a_y\vec{j}+a_z\vec{k},\quad \vec{b}=b_x\vec{i}+b_y\vec{j}+a_z\vec{k}.$$

Получаем:

$$\vec{a}\cdot\vec{b}=(a_x\vec{i}+a_y\vec{j}+a_z\vec{k})(b_x\vec{i}+b_y\vec{j}+b_z\vec{k})=$$
$$=a_xb_x\vec{i}\cdot\vec{i}+a_xb_y\vec{i}\cdot\vec{j}+a_xb_z\vec{i}\cdot\vec{k}+a_yb_x\vec{j}\cdot\vec{i}+a_yb_y\vec{j}\cdot\vec{j}+a_yb_z\vec{j}\cdot\vec{k}+a_zb_x\vec{k}\cdot\vec{i}+a_zb_y\vec{k}\cdot\vec{j}+a_zb_z\vec{k}\cdot\vec{k}.$$

Шесть слагаемых с «перекрёстными» произведениями базисных векторов обращаются в нуль, а остальные три слагаемых дают требуемую формулу:

$$\vec{a}\cdot\vec{b}=a_xb_x+a_yb_y+a_zb_z.$$

Теорема доказана.

Если, как и выше, в формуле (7.28) положить $\vec{b}=\vec{a}$, то получим:

$$a^2=\vec{a}\cdot\vec{a}=a_x^2+a_y^2+a_z^2,$$

откуда

$$a=\sqrt{a_x^2+a_y^2+a_z^2}\,.$$

Мы снова пришли к формуле (7.17).