Міністерство освіти і науки України
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

І. Ф. Скіцько, О. І. Скіцько

Фізика для інженерів

Підручник

2-ге видання, перероблене і доповнене

Затверджено Вченою радою КПІ ім. Ігоря Сікорського як підручник для здобувачів ступеня бакалавра за спеціальностями F5 «Кібербезпека та захист інформації», G5 «Електронні комунікації, приладобудування та радіотехніка»

Електронне мережеве навчальне видання

Київ
КПІ ім. Ігоря Сікорського
2026

УДК 534.1– 535.8; 537.1–539.18(075.8)
С42

Затверджено Вченою радою КПІ ім. Ігоря Сікорського
(протокол № 11 від 10.11.2025 р.)

Рецензенти: *Гаврюшенко Д. А.*, д-р фіз.-мат. наук, проф.,
Київський національний університет імені Тараса Шевченка
Зачек І. Р., д-р фіз.-мат. наук, проф.,
Національний університет «Львівська політехніка»

Відповідальний редактор: *Лінчевський І. В.*, д-р фіз.-мат. наук, проф.,
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Скіцько І. Ф.

С42 Фізика для інженерів : підручник / І. Ф. Скіцько, О. І. Скіцько. 2-ге вид., перероб. і допов. – Електрон. текстові дані 1 файл: 21,3 Мб). – Київ : КПІ ім. Ігоря Сікорського, Вид-во «Політехніка», 2026. – 715 с.

ISBN 978-966-990-248-1 (Ел.)

Викладено фізичні основи електростатики, постійного струму, електромагнетизму, коливань і хвиль, хвильової оптики, квантової природи випромінювання, будови атома та атомного ядра, квантової механіки, фізики твердого тіла і контактних явищ, термоелектронної емісії, квантової електроніки, поширення світла оптичними волокнами.

Особливо акцентовано увагу на матеріалі, що використовується у техніці зв'язку, – це фізика металів і напівпровідників, фізика контактних явищ металів і напівпровідників, а також фізичні основи передавання інформації оптичними волокнами.

Для здобувачів ступеня бакалавра за спеціальностями F5 «Кібербезпека та захист інформації», G5 «Електронні комунікації та радіотехніка» відповідно до програми з фізики для здобувачів закладів вищої освіти технічного профілю.

УДК 534.1– 535.8; 537.1–539.18(075.8)

ISBN 978-966-990-248-1 (Ел.)

© І. Ф. Скіцько, О. І. Скіцько, 2026
© КПІ ім. Ігоря Сікорського (ФМФ), 2026

Зміст

ПЕРЕДМОВА ……..…….... 10
Розділ 1. Електростатичне поле..... 12
1.1. Електричний заряд……..... 12
1.2. Закон Кулона……..…..... 14
1.3. Напруженість електричного поля.… 16
1.4. Принципи суперпозиції електричних полів……..…….…..... 17
1.5. Графічне зображення електричних полів…….……... 18
1.6. Потік вектора напруженості електричного поля……..…… 19
1.7. Теорема Остроградського–Гауса.... 20
1.8. Застосування теореми Остроградського–Гауса до розрахунку електричних полів…… 24
1.9. Робота сил електростатичного поля…….… 23
1.10. Різниця потенціалів. Потенціал. Еквіпотенційні поверхні…….. 29
1.11. Циркуляція вектора напруженості електростатичного поля…..……. 31
§ 1.12. Зв'язок між напруженістю та потенціалом електростатичного поля… 32
1.13. Рівняння Пуассона (теорема Остроградського–Гауса в диференційній формі)…….. 33
Контрольні запитання……..……. 35
Розділ 2. Поляризація діелектриків та поле всередині діелектрика……. 36
2.1. Провідники, напівпровідники та діелектрики……..…… 36
2.2. Електричний диполь в однорідному і неоднорідному електричних полях …. 38
2.3. Поляризація діелектриків та її типи. вектор поляризації (поляризованість), діелектрична сприйнятливість……. 40
2.4. Електричне поле всередині діелектрика. Діелектрична проникність та її фізичний зміст. Зв'язок між діелектричною сприйнятливістю та діелектричною проникністю…..……. 46
2.5. Електричне поле на межі двох діелектриків…….. 48
2.6. Особливі властивості деяких діелектриків.Сегнетоелектрики. П'єзоелектрики. Електрети…..…….... 50
Контрольні запитання…..…….. 58
Розділ 3. Провідники в електричному полі……. 59
3.1. Розподіл зарядів, напруженість та потенціал електричного поля у провіднику……. 59
3.2. Електроємність відокремленого провідника……. 66
3.3. Конденсатори……..…….. 68
3.4. Енергія зарядженого відокремленого провідника та конденсатора. Енергія електростатичного поля……...… 73
Контрольні запитання…….. 75
Розділ 4. Постійний електричний струм……. 75
4.1. Електричний струм та його види. Струм провідності та умови його існування…… 75
4.2. Закон Ома та Джоуля–Ленца в інтегральній та диференційній формах.. 78

Зміст

4.3. Електрорушійна сила. Закон Ома для неоднорідної ділянки кола…… 84
4.4. Розрахунок розгалужених електричних кіл за допомогою правил Кірхгофа…… 90
Контрольні запитання…….. 92
Розділ 5. Магнітне поле…….. 92
5.1. Взаємодія струмів. Закон Ампера. Магнітне поле …….. 92
5.2. Взаємодія елементів струму…..... 95
5.3. Індукція магнітного поля. Закон Біо–Савара–Лапласа ……..…. 99
5.4. Застосування закону Біо–Савара–Лапласа до розрахунку магнітних полів..... 102
5.4.1. Магнітне поле прямолінійного провідника зі струмом…….…….. 102
5.4.2. Поле колового витка зі струмом…….…… 103
5.4.3. Магнітне поле соленоїда…… 105
5.4.4. Магнітне поле електричного заряду, який рухається……..…….. 106
5.5. Циркуляція вектора індукції магнітного поля або закон повного струму……..……..... 108
5.6. Теорема Остроградського–Гауса для магнітного поля……. 110
5.7. Магнітне поле довгого соленоїда, тороїда та прямого проводу…… 111
5.8. Контур зі струмом у магнітному полі…….……..… 115
5.9. Робота під час переміщення провідника зі струмом у магнітному полі……..... 118
5.10. Дія електричного і магнітного полів на рухомий електричний заряд… 121
5.10.1. Сила Лоренца…….…… 121
5.10.2. Ефект Холла……..….. 123
5.10.3. Рух електронів у схрещених електричному і магнітному полях. Визначення питомого заряду лектрона методом магнетрона……..... 124
Контрольні запитання……. 133
Розділ 6. Постійне магнітне поле в речовині……... 134
6.1. Намагнічування середовищ…..….. 134
6.2. Магнітомеханічні явища……...136
6.3. Закон повного струму для магнітного поля в речовині. Напруженість магнітного поля. Магнітна сприйнятливість та проникність…….. 138
6.4. Магнітне поле на межі двох магнетиків……... 140
6.5. Класифікація магнетиків…….. 143
6.6. Діа - і парамагнетизм…… 144
6.7. Феромагнетизм……. 149
6.8. Магнітні матеріали. Ферити……. 156
6.9. Пояснення феромагнетизму……. 157
6.10. Процеси намагнічування феромагнетика……. 162
Контрольні запитання……. 164
Розділ 7. Електромагнітна індукція…….. 165
7.1. Явище електромагнітної індукції……. 165

7.2. Закон Ленца……..... 169
7.3. Основний закон електромагнітної індукції…….……... 170
7.4. Пояснення закону електромагнітної індукції……... 171
7.5. Самоіндукція. Індуктивність…..... 177
7.6. Взаємна індукція…….. 180
7.7. Власна енергія струму…… 182
7.8. Енергія магнітного поля…….. 184
7.9. Взаємна енергія двох струмів……. 185
7.10. Вихрові струми……..…... 186
7.11. Витіснення змінного струму (скін-ефект)……. 188
7.12. Методи вимірювання магнітної індукції…… 190
Контрольні запитання..... 191
Розділ 8. Рівняння Максвелла…….... 192
8.1. Перше рівняння Максвелла…… 192
8.2. Струм зміщення. Друге рівняння Максвелла……... 193
8.3. Рівняння Максвелла……...…. 197
8.4. Рівняння Максвелла в диференціальній формі…….. 199
8.5. Значення теорії Максвелла……..... 201
Контрольні запитання……..... 203
Розділ 9. Коливання…….. 204
9.1. Види коливань. Загальні поняття…….…. 204
9.2. Власні коливання……..…... 206
9.3. Рівняння власних коливань. Коливання у разі відсутності згасання …… 208
9.4. Геометричне зображення гармонічних коливань. Додавання гармонічних коливань однакового напрямку й однакової частоти……... 210
9.5. Додавання гармонічних коливань однакового напрямку з різним.и частотами. Биття…….……... 212
9.6. Додавання взаємно перпендикулярних коливань……...… 213
9.7. Диференціальне рівняння загасаючих коливань і його розв'язання …….. 216
9.8. Вимушені електричні коливання. Резонанс……...… 221
9.9. Параметричний резонанс……. 228
9.10. Змінний струм……..... 230
9.11. Автоколивання ……...…….232
Контрольні запитання…… 235
Розділ 10. Пружні хвилі……... 236
10.1. Утворення хвиль у пружному середовищі. Поздовжні і поперечні хвилі…….……..... 236
10.2. Рівняння плоскої та сферичної хвиль……..... 239
10.3. Хвильове рівняння……..…. 241
10.4. Енергія пружної хвилі ……..... 242
10.5. Групова швидкість поширення хвиль…… 244
10.6. Інтерференція хвиль …….. 247
10.7. Звук…….…..…… 253
10.8. Ефект Доплера для звукової хвилі……..... 257
Контрольні запитання…… 260

Розділ 11. Електромагнітні хвилі….. 261
11.1. Загальні відомості про електромагнітні хвилі…….. 261
11.2. Доведення необхідності існування електромагнітних хвиль….. 262
11.3. Властивості електромагнітних хвиль……. 265
11.4. Енергія електромагнітних хвиль. Потік енергії. Вектор Пойтінга…….. 267
11.5. Генерування й експериментальне дослідження електромагнітних хвиль (радіохвиль)……. 269
11.6. Випромінювання елементарного диполя…….. 272
Контрольні запитання…… 278
Розділ 12. Інтерференція електромагнітних хвиль…….. 279
12.1. Основні відомості про хвилі оптичного діапазону. Інтенсивність електромагнітних хвиль……. 279
12.2. Інтерференція світла ……..... 281
12.3. Розрахунок інтерференційної картини від двох когерентних джерел .284
12.4. Способи спостереження інтерференції світла…… 287
12.5. Інтерференція квазімонохроматичного світла. Часова когерентність …… 293
12.6. Роль кінцевих розмірів джерела світла. Просторова когерентність…….. 295
12.7. Практичне застосування інтерференції світла…… 300
Контрольні запитання……. 304
Розділ 13. Дифракція електромагнітних хвиль……..... 305
13.1. Дифракція хвиль…….. 305
13.2. Принцип Гюйгенса–Френеля …..... 306
13.3. Метод зон Френеля. Прямолінійне поширення світла…….…… 308
13.4. Дифракція Френеля на круглому отворі та диску…….…… 313
13.5. Дифракція Френеля на прямолінійному краю напівплощини… 317
13.6. Дифракція Фраунгофера …..…… 322
13.7. Дифракційна ґратка ……..…. 326
13.8. Дифракція рентгенівського випромінювання …….... 332
Контрольні запитання……. 334
Розділ 14. Поляризація світла……. 335
14.1. Природне і поляризоване світло. Закон Малюса ……..... 335
14.2. Поляризація під час відбивання та заломлення світла…… 341
14.3. Подвійне променезаломлення ….. 346
14.4. Поляризаційні призми та поляроїди…… 351
14.5. Штучна оптична анізотропія ……. 354
14.6. Обертання площини поляризації ……..…… 358
Контрольні запитання……. 362
Розділ 15. Взаємодія електромагнітних хвиль з речовиною……..... 362
15.1. Дисперсія світла. Області нормальної й аномальної дисперсії. Класична електронна теорія дисперсії …….. 362

15.2. Групова швидкість світла. Хвильовий пакет (груда хвиль)…….. 370
15.3. Вплив магнітного поля на рух електронів у атомі. Пояснення ефекту Фарадея …… 375
15.4. Поглинання світла …….. 379
15.5. Розсіяння світла…….……. 381
15.6. Ефект Доплера в оптичних явищах ……...…….. 384
Контрольні запитання……. ……386
Розділ 16. Квантова природа випромінювання……...…. 386
16.1. Теплове випромінювання. Його характеристики. Абсолютно чорне тіло. Закон Кірхгофа……...……. 386
16.2. Закон Стефана–Больцмана. Закон Віна. Формула Релея–Джінса … 393
16.3. Квантова гіпотеза та формула Планка……..... 398
16.4. Оптична пірометрія……...…… 402
16.5. Зовнішній фотоефект та його закони. Рівняння Ейнштейна для зовнішнього фотоефекту……..... 406
16.6. Маса та імпульс фотона…….. 411
16.7. Тиск світла……... 425
16.8. Ефект Комптона…….. 418
Контрольні запитання……. 420
Розділ 17. Будова атома…… 421
17.1. Дослід Резерфорда. Планетарна модель атома……...…. 421
17.2. Атомні спектри……...…… 427
17.3. Постулати Бора. Досліди Франка та Герца. Теорія атома водню……...…. 430
17.4. Подальший розвиток теорії Бора. Квантування Бора–Вільсона–Зоммерфельда…...…. 437
17.5. Просторове квантування ……..…. 441
17.6. Ефект Зеємана …….……. 443
17.7. Дослід Штерна і Герлаха …...…… 445
17.8. Власний момент електрона …...… 447
17.9. Квантові стани електронів в атомі. Квантові числа……. 449
17.10. Електронна оболонка атома. Принцип Паулі…… 451
17.11. Періодична система хімічних елементів ……. 453
17.12. Рентгенівські спектри ..……...… 455
Контрольні запитання …… ……..…. 460
Розділ 18. Елементи квантової механіки ……. 460
18.1. Протиріччя хвиля – частинка …… 461
18.2. Хвилі де Бройля …….. 462
18.3. Експериментальне підтвердження гіпотези де Бройля ……..… 464
18.4. Співвідношення невизначеностей як прояв корпускулярно-хвильового дуалізму властивостей матерії ……467
18.5. Статистичне тлумачення хвиль де Бройля…….. 475
18.6. Ймовірність місцезнаходження мікрочастинки ……. 478
18.7. Рівняння Шредінгера. Принцип причинності в квантовій механіці ……. 480
18.8. Рух вільної частинки. Частинка в прямокутній потенціальній ямі…….... 482
18.9. Розв'зок рівняння Шредінгера

для потенціального бар'єра…… 485
18.9.1. Постановка проблеми …….. 485
18.9.2. Розв'язок рівняння Шредінгера для потенціального бар'єру нескінченної ширини ……...… 487
18.9.3. Розв'язок рівняння Шредінгера для потенціальногобар'єра кінцевої ширини. Тунельний ефект ……..…….... 492
18.10. Холодна емісія електронів з металу……..…… 496
18.11. Ефект Рамзауера ……..…… 497
Контрольні запитання……. 506

Розділ 19. Основи зонної теорії твердих тіл……..…... 507
19.1. Тверде тіло. Аморфні та кристалічні тверді тіла …….... 507
19.2. Структура кристалів ……..……508
19.3. Основні типи міжатомної взаємодії у твердих тілах ……..……..… 511
19.4. Рідкі кристали. Сили Ван дер Ваальса……..... 515
19.5. Класифікація твердих тіл за величиною електропровідності та її залежністю від температури ……...… 519
19.6. Енергетичні зони у кристалах … 523
19.7. Статистика електронів у кристалах…….. 531
19.8. Електропровідність металів та її залежність від температури. Надпровідність …….. 538
19.9. Статистика електронів і дірок у напівпровідниках ……..…. 549
19.10. Електропровідність напівпровідників та її залежність від температури .. 552
19.11. Внутрішній фотоефект ……..…. 560
19.12. Дрейф і дифузія носіїв заряду 569
19.13. Ефект Ганна ……..……... 577
Контрольні запитання……. 584

Розділ 20. Термоелектронна емісія… 586
20.1. Історія відкриття термоелектронної емісії……. 586
20.2. Робота виходу електрона з металу. 587
20.3. Закони термоелектронної емісії … 590
20.4. Контакт двох металів ……..…….. 595
Контрольні запитання……. 597

Розділ 21. Контактні явища ……597
21.1. Проникнення зовнішнього електричного поля у метал, діелектрик, напівпровідник…… 598
21.2. Розподіл поля в бар'єрі. Ширина бар'єра ……..…. 600
21.3. *p–n* - перехід та його властивості ……..605:
21.3.1. Способи отримання *p–n* - переходів …… 605
21.3.2. Потенціальний бар'єр в області контакту *p*- і *n*-напівпровідників …….. 610
21.3.3. Розподіл електричного поля в бар'єрі. Ширина бар'єра ……... 614
21.3.4. Зворотне зміщення…… 616
21.3.5. Пряме зміщення……..…… 618
21.4. Ємність електронно-діркового переходу……..…… 621
21.4.1. Варікапи……..…. 624

21.5. Фотодіоди і світлодіоди …..…… 625
21.6. Інші види контактів ……..…… 632
Контрольні запитання ……..……. 636

Розділ 22. Елементи квантової електроніки…… 637
22.1. Фізичні основи квантової електроніки…….. 637
22.2. Лазери ……..……...… 644
22.3. Властивості лазерного випромінювання……..…… 650
22.4. Особливості конструкції та роботи деяких типів лазерів ……..… 652
22.5. Фізичні основи волоконно – оптичних ліній зв'язку …. 658
Контрольні запитання……. 670

Розділ 23. Атомні ядра…….. 671
23.1. Атомні ядра та їх характеристики 671
23.2. Ядерні сили ……...… 673
23.3. Енергія зв'язку ядра ……....… 675
23.4. Моделі атомного ядра …… 678
23.5. Радіоактивність. Основний закон радіоактивного перетворення атомних ядер ……..……... 680
23.6. Ядерні реакції …….…….. 685
23.7. Поділ ядер …….……... 687
23.8. Ланцюгова реакція поділу. Ядерний реактор……. 690
23.9. Реакції синтезу легких ядер …... 694
Контрольні запитання…… 697
Список використаної літератури…….. 698
Додаток 1 ……..……..... 699
Предметний покажчик ……... 702

Передмова

Курс фізики у закладі вищої освіти належить до фундаментальної дисципліни, яка становить основу теоретичної підготовки студентів інженерних спеціальностей, виконує роль тієї бази, без якої неможлива успішна діяльність інженера в будь-якій галузі сучасного виробництва та яка формує його світогляд. Запропонований курс фізики розрахований на двосеместрове навчання студентів, які спеціалізуються у технічних науках. Автори намагались створити такий курс, в якому особливу увагу приділено основам фізики, що використовуються у техніці зв'язку.

Нині швидкими темпами змінюються техніка і технології. Використовуються нові матеріали і технології для створення надсучасних технічних пристроїв, відбувається перехід від мікроелектроніки до наноелектроніки. Цей перехід надто помітний у сфері сучасного зв'язку. З'явилися нові спеціальності, як, наприклад, «Кібербезпека та захист інформації» та інші. Усе це вимагає нових підходів у викладанні й наповнюваності змісту загального курсу фізики.

Знання основ загальної фізики, особливо фізики металів і напівпровідників, фізики контактних явищ металів і напівпровідників, фізичних основ передачі інформації оптичними волокнами врай важливі для оволодіння сучасними принципами побудови й функціонування засобів захисту інформації та телекомунікаційного обладнання, тому такі розділи курсу фізики включені до відповідних модулів освітньо-професійних програм підготовки бакалаврів за спеціальностями F5 «Кібербезпека та захист інформації» та G5 «Електроніка, електронні комунікації, приладобудування та радіотехніка». Усе це зумовило написання цього підручника.

Для написання підручника використано літературу, список якої наведено наприкінці книги. Головним у відборі матеріалу підручника визначено чинну нині програму з курсу фізики для зазначених вище спеціальностей, а також матеріали курсу лекцій, які протягом багатьох років викладають слухачам у Державному закладі «Інститут спеціального зв'язку та захисту інформації КПІ ім. Ігоря Сікорського». Підручник складається із двадцяти трьох розділів, в яких викладено фізичні основи електростатики, постійного струму,

електромагнетизму, коливань і хвиль, хвильової оптики, квантової природи випромінювання, будови атома та атомного ядра, квантової механіки, фізики твердого тіла і контактних явищ, термоелектронної емісії, квантової електроніки, поширення світла оптичними волокнами. Матеріал підручника максимально забезпечує теоретичні основи спеціальних курсів для вищезазначених спеціальностей. Особливу увагу приділено визначенню вихідних понять та їх походження.

Автори намагались дотримуватись лаконічного стилю викладу матеріалу, що одночасно поєднував би основи класичної й сучасної фізики, зробити акценти у тексті на основних поняттях і законах, які мають сприяти кращому засвоєнню матеріалу. Вирішальну роль у роботі над курсом мало постійне активне спілкування зі студентами на лекціях, практичних і лабораторних заняттях, консультаціях, заліках та іспитах. Неможливо написати підручник і не мати справ з тими, кому його призначено, спілкуватись з ними тільки в односторонньому порядку на лекціях. У викладанні, як і в будь-якій іншій справі, потрібний експеримент. Із декількох можливих способів викладання того чи іншого питання треба надати перевагу тому, який у процесі навчання дає кращий результат.

Під час написання підручника автори ставили за мету логічно підвести матеріал курсу до такого об'єму інформації, щоб здобувач зрозумів природу виникнення особливих властивостей контакту двох напівпровідників, контакту «метал – напівпровідник», механізм поширення та способи введення й виведення оптичного сигналу в оптоволокнах, принципів роботи оптичних квантових підсилювачів і генераторів, фотоприймачів (пікселів). Розуміння цих питань дуже важливе для становлення інженера-зв'язківця. Автори хотіли подати фізику не як деякий об'єм інформації, який треба засвоїти, а як розумну, логічну й красиву науку. Фізику слід вивчати під гаслом «Жодного зазубрювання!». Замість пам'яті формальної краще використовувати пам'ять логічну – запам'ятовування має досягатись через глибоке розуміння.

Використаний математичний апарат забезпечує достатній рівень строгості викладення матеріалу курсу й відповідає математичній підготовці здобувачів перших курсів закладів вищої технічної освіти.

Розділ 1. Електростатичне поле

1.1. Електричний заряд

У сучасній фізиці відомі чотири види фундаментальні взаємодії: гравітаційна, електромагнітна, сильна, слабка. ***Сильна*** взаємодія проявляється між частинками в ядерних масштабах (10^{-15} м). ***Слабка*** ще більш короткодіча (приблизно $2 \cdot 10^{-18}$ м), проявляється між елементарними частинками і відіграє важливу роль під час їх взаємних перетворень. ***Гравітаційна взаємодія*** (сила тяжіння) – універсальна взаємодія між будь-якими тілами. Якщо взаємодія відносно слабка і тіла рухаються повільно (порівняно зі швидкістю світла: $c = 3 \cdot 10^8$ м/с), то справедливий закон всесвітнього тяжіння Ньютона (1867 р.): *дві будь-які матеріальні частинки з масами* m_1 і m_2 *притягуються у напрямку одна до одної із силою F, яка прямо пропорційна добутку мас і обернено пропорційна квадрату відстані r між ними:*

$$F = G\frac{m_1 m_2}{r^2}. \qquad (1.1)$$

(під матеріальними частинками розуміють будь-які тіла за умови, що їхні лінійні розміри набагато менші за відстань між ними). Коефіцієнт пропорційності G називають ***гравітаційною сталою***, числове значення якої вперше визначив Г. Кавендіш (1798 р.). За сучасними даними $G = 6{,}6745 \cdot 10^{-11}$ м^3/(кг $\cdot$ с^2). Згідно із законом (1.1) сила F залежить тільки від положення частинок у цей момент часу, тобто гравітаційна взаємодія поширюється миттєво. Сила (1.1) завжди є силою тяжіння.

Електромагнітна взаємодія на багато порядків більш інтенсивна від гравітаційної і слабкої та є далекодійною, на противагу сильній взаємодії. Вона найбільш часто трапляється і найбільше проявляється в мега -, макро - і мікросвіті. Така взаємодія обумовлена участю електромагнітного поля, яке або поглинається, або випромінюється, що й обумовлює взаємодію між тілами. *Тому, якщо частинки (тіла) взаємодіють між собою із силами, які повільно зменшуються зі збільшенням відстані між ними й набагато перевищують силу всесвітнього тяжіння (1.1), то говорять, що ці тіла мають електричний заряд.* Можна сказати, що електричний заряд – це фізична величина, яка визначає інтенсивність електромагнітної взаємодії, подібно тому, як маса визначає інтенсивність гравітаційної взаємодії.

Електричний заряд пов'язаний з матеріальним носієм і є джерелом електромагнітного поля. Електричний заряд є внутрішньою характеристикою

елементарних частинок і визначає їх електромагнітну взаємодію. *Вся сукупність електричних і магнітних явищ є проявом існування, руху та взаємодії електричних зарядів.*

Електромагнітна взаємодія є далекодійною і може приводити як до притягання, так і до відштовхування між тілами, на відміну від гравітаційної взаємодії. Це обумовлює наявність.двох різнойменних електричних зарядів: позитивного і негативного. Різнойменні заряди притягуються, а однойменні – відштовхуються. Позитивними важають ті заряди, які виникають на скляній паличці під час натирання її шовком, а негативними – заряди, які виникають на ебонітовій паличці (або бурштиновій (янтарній)) під час натирання її хутром. У природі найчастіше трапляється взаємна компенсація позитивних і негативних зарядів (через силу тяжіння) на макроскопічних тілах, як результат, тіла стають електрично нейтральними і не проявляють електричної взаємодії.

Численними експериментами було встановлено закон збереження електричних зарядів: ***сумарний заряд електрично ізольованої системи залишається незмінним. Заряди можуть лише передаватись від одного тіла системи до іншого або зміщуватись у середині цього тіла.***

Примітка. Систему називають електрично ізольованою, якщо через поверхню, яка обмежує систему, не можуть проникати електричні заряди.

Електричні заряди можуть зникати й виникати знову, але завжди зникають або виникають два електричні заряди протилежних знаків. Величина заряду, яку вимірюють у різних інерційних системах відліку, є однаковою. Це означає, що електричний заряд є релятивістськи інваріантним, тобто його величина не залежить від того, рухається він, чи перебуває у стані спокою.

Думку про те, що електричний заряд дискретний, вперше висловив Б. Франклін у 1749 р. Перші обґрунтовані припущення про те, що електрика не є неперервною « рідиною », а складається з особливих частинок, були зроблені на основі праць М. Фарадея з дослідження явища електролізу. Цю найменшу порцію електрики, яку часто називають електричним зарядом, або «атомом електрики», ірландський фізик Г. Стоней у 1891 р. назвав електроном. Пряме експериментальне вимірювання елементарного заряду виконав у 1909 р. американський фізик Р. Міллікен. Він встановив, що заряд q є цілим кратним однієї й тієї самої величини $|e|$, тобто

$q = n|e|$, де $n = \pm 1, \pm 2, \ldots$. Експериментально було визначено, що $|e| = 1{,}6 \cdot 10^{-19}$ Кл. Ще в 1897 р. Дж. Томсон визначив питомий заряд електрона: $e/m_e = 1{,}7588 \cdot 10^{11}$ Кл/кг. Звідси маса електрона $m_e = 9{,}0971 \times \times 10^{-31}$ кг. Отже, електрон є матеріальним носієм найменшої маси й найменшого електричного заряду загалом. Електричний заряд електрона домовились вважати негативним відповідно до більш ранньої угоди, коли вважали негативним заряд наелектризованого бурштину (янтаря) (грецьке electron – бурштин (янтар)).

Носієм позитивного елементарного електричного заряду є протон. За сучасними експериментальними даними позитивний заряд протона за абсолютним значенням дорівнює негативному елементарному заряду електрона з відносною точністю 10^{-21}. Важко пояснити, чому заряди електрона і протона однакові з такою фантастичною точністю. *Мабуть, квантування заряду є загадковим і універсальним законом природи*.

Дослідження показують, що заряди всіх елементарних частинок є точно однаковими. Явище квантування заряду перебуває за межами класичного електромагнетизму. Зазвичай квантування ми не враховуємо й чинимо так, начебто наші точкові заряди q можуть мати будь-яку величину. Це не викликає труднощів. Утім корисно згадати, що класична теорія не спроможна пояснити структуру елементарних частинок (не можна, між іншим, стверджувати, що сучасна квантова теорія може це зробити!). Що утримує електрон від розпаду – така ж сама таємниця, як і те, що визначає точну величину його заряду. Тут мають існувати якісь сили, відмінні від електричних, оскільки електростатичні сили, які діють між різними частинами електрона, спричиняють відштовхування.

1.2. Закон Кулона

Електростатика вивчає електричні поля нерухомих зарядів. Спочатку розглянемо електростатичні поля у вакуумі. Основний кількісний закон електростатики відкрив Кулон у 1785 р.: *сила взаємодії F двох точкових зарядів у вакуумі направлена вздовж прямої, яка з'єднує ці заряди, пропорційна добутку їх величин q_1 та q_2 й обернено пропорційна квадрату відстані між ними r*. Вона є силою притягання, якщо знаки зарядів різні, та силою відштовхування, якщо знаки зарядів однакові. Математично

$$F = k\frac{q_1 q_2}{r^2}, \qquad (1.2)$$

де k – коефіцієнт пропорційності.

Точковість зарядів у законі Кулона означає, що лінійні розміри тіл, на яких зосереджені ці заряди, мізерно малі порівняно з відстанню між ними. Силу Кулона називають центральною силою, тому що вона напрямлена вздовж прямої, що з'єднує заряди.

Закон Кулона можна записати у векторній формі:

$$\vec{F} = k\frac{q_1 q_2}{r^2}\vec{e}_r, \qquad (1.3)$$

де $\vec{e}_r$ – одиничний вектор, що має напрямок від заряду q_1 до заряду q_2, якщо визначають силу, яка діє на заряд q_2 зі сторони заряду q_1 (рис. 1.1, *а*), і навпаки, $\vec{e}_r$ має напрямок від заряду q_2 до заряду q_1, якщо визначають силу, яка діє на заряд q_1 зі сторони заряду q_2 (рис. 1.1, *б*).

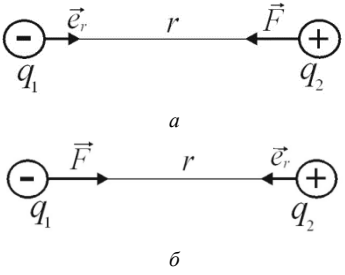


Рис. 1.1. Визначення напрямку сили взаємодії двох зарядів

Суттєвим фізичним змістом закону (1.2) є твердження про обернену залежність сили взаємодії зарядів від квадрату відстані між ними. Щодо цього, закон (1.2) дуже корелює із законом (1.1). Дослідні дані показують, що закон Кулона справедливий для відстаней від 10^{-15} м до, принаймні, декількох кілометрів. Є всі підстави вважати, що для відстаней, які менші 10^{-16} м, закон Кулона перестає бути справедливим.

У СІ для зарядів у вакуумі коефіцієнт k у формулі (1.2) записують у формі

$$k = 1/(4\pi\varepsilon_0), \qquad (1.4)$$

де $\varepsilon_0 = 8{,}85 \cdot 10^{-12}$ Кл^2/(Н $\cdot$ м^2,

електрична стала і, отже, $k = 9 \cdot 10^9$ Н $\cdot$ м^2/Кл^2. Зауважимо, що Кл^2/(Н $\cdot$ м^2), ще записують, як Ф/м. Множник 4π у виразі (1.4) відображає сферичну симетрію електростатичного поля точкового заряду, оскільки величина 4π дорівнює повному тілесному куту в стерадіанах.

Закон Кулона ($F \sim 1/r^2$) піддавався експериментальній перевірці багато разів, аж до нині. Це пов'язано з тим, що було побудовано квантову електродинаміку, яка принципово по-іншому, ніж класична теорія, пояснювала взаємодію зарядів. Утім, це не означає, що закони класичної електродинаміки було повністю відкинуто. Взаємовідносини квантових і класичних законів визначають принципом відповідності: класичні закони є граничним випадком квантових, а отже, між ними має існувати зв'язок. Одним із проявів цього зв'язку є таке твердження:

якщо закон Кулона не є точним законом «обернених квадратів» ($F \sim 1/r^2$), то маса спокою фотона m_λ має відрізнятися від нуля. Слід зауважити, що питання про рівність нулю m_λ – це питання експериментальне.

Оскільки низка важливих положень сучасної фізики грунтується на припущенні про те, що $m_\lambda = 0$, то проблему експериментального підтвердження цього припущення слід розглядати як принципово важливу. Дослідна перевірка закону Кулона – один зі способів її розв'язку. Про експериментальну перевірку того, що маса спокою фотона $m_\lambda = 0$, розглянуто в підрозд. 16.6. Досліди, проведені в земних умовах, а також оцінки, зроблені за допомогою аналізу астрофізичних даних, не дають підстав вважати масу спокою фотона відмінною від нуля, і тому, закон взаємодії точкових електричних зарядів зберігає статус точного закону «обернених квадратів».

1.3. Напруженість електричного поля

За сучасними поглядами, матеріальним носієм взаємодії нерухомих зарядів є електричне поле. Основною ознакою наявності електричного поля є те, що на будь-який заряд, внесений у це поле, діє сила. Розкриття властивостей поля – одне з найважливіших завдань фізики. Далі буде розглянуто стаціонарні електричні поля, тобто такі поля, що не змінюються з часом і створюються нерухомими електричними зарядами. Такі поля називають *електростатичними*.

Під час вивчення змінного електромагнітного поля й, особливо, електромагнітних хвиль можна переконатися, що поле може існувати без зарядів (електромагнітні хвилі) і що поняття поля має глибокий фізичний зміст: електромагнітне поле є об'єктивною реальністю. Поле реальне так само, як і речовина, і як речовина є одним із видів матерії.

Властивості електричного поля вивчають за допомогою пробного точкового позитивного електричного заряду $q_{\text{пр}}$. Пробний заряд має бути досить малим, щоб запобігати спотворенню початкового поля. Його наявність не повинна змінити розподіл зарядів, а також не має суттєво вплинути на результати вимірювань.

Якщо в поле, що створюється зарядом q, помістити пробний заряд $q_{\text{пр}}$, то на нього буде діяти сила (1.3), яка пропорційна величині $q_{\text{пр}}$, тому ця сила не може бути характеристикою самого поля. Утім, величина, яка дорівнює відношенню $\vec{F}/q_{\text{пр}}$ (питома сила, сила,

що діє на одиничний позитивний заряд), може бутӣ *силовою характеристикою* поля.

Отже, векторну величину

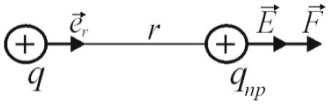
$$\vec{E} = \vec{F}/q_{\text{пр}} \qquad (1.5)$$

називають *напруженістю електричного поля.* Враховуючи закон Кулона (1.3) і відношення (1.5), отримуємо формулу для напруженості поля точкового заряду q на відстані r:

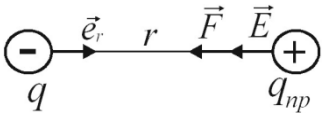
$$\vec{E} = \frac{kq}{r^2}\vec{e}_r. \qquad (1.6)$$

Звідси видно, що поле точкового заряду є центрально-симетричним.

За напрямок вектора напруженості $\vec{E}$ беруть напрямок сили, з якою поле діє на пробний позитивний заряд, вміщений у певну точку поля (рис. 1.2).



а



б

Рис. 1.2. Визначення напрямку вектора напруженості електричного поля

У СІ одиниця напруженості електричного поля 1 Н/Кл – це напруженість такого поля, яке діє з силою 1 Н на точковий заряд 1 Кл. Частіше напруженість поля вимірюють у вольтах на метр (В/м).

1.4. Принцип суперпозиції електричних полів

Суттєвим фізичним змістом закону Кулона є твердження про адитивну дію електричних зарядів. Згідно з цим твердженням *сила взаємодії двох зарядів не змінюється за наявності третього заряду.* Незалежно від кількості зарядів, які входять у деяку систему, закон Кулона (1.2) можна використовувати для розрахунку взаємодії кожної пари зарядів. Це твердження є основою ***принципу суперпозиції***, який використовують для знаходження величини і напрямку вектора напруженості $\vec{E}$ в кожній точці поля за заданим розподілом у просторі та величиною зарядів. Суть ***принципу суперпозиції*** полягає в тому, що сила, яка діє на пробний заряд $q_{\text{пр}}$, що розміщений у будь-якій точці поля системи n зарядів, буде векторною сумою сил, які створюються кожною парою (пробний заряд $q_{\text{пр}}$ і заряд системи) та окремо діють на $q_{\text{пр}}$ у цій точці.

Математично це можна записати так:

$$\vec{E} = \frac{\vec{F}}{q_{\text{пр}}} = \frac{\sum_i^n \vec{F}_i}{q_{\text{пр}}} = \frac{\vec{F}_1}{q_{\text{пр}}} + \ldots + \frac{\vec{F}_n}{q_{\text{пр}}};$$

$$\vec{E} = \vec{E}_1 + \vec{E}_2 + \cdots + \vec{E}_n = \sum_i^n \vec{E}_i, \qquad (1.7)$$

де $\vec{E}_1, \vec{E}_2, \ldots, \vec{E}_n$ – напруженості полів, які створюються окремими

зарядами у цій точці простору, де перебуває $q_{\text{пр}}$.

Принцип суперпозиції дає можливість обчислювати напруженість поля будь-якої системи зарядів. Подумки поділяючи, наприклад, заряджене тіло скінченних розмірів на точкові заряди, знаходимо складові напруженості в певній точці, створені окремими елементами зарядженого тіла. Потім, згідно з принципом суперпозиції, визначаємо результуючу напруженість.

1.5. Графічне зображення електричних полів

Електричні поля зображують за допомогою ***ліній напруженості*** (силових ліній), які проводять так, щоб дотичні до цих ліній у кожній точці збігалися з лінією вектора $\vec{E}$ (рис. 1.3).

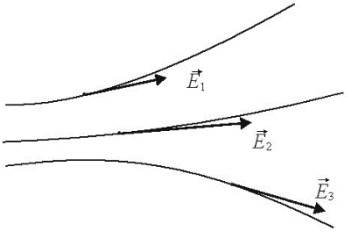


Рис. 1.3. Зображення електричних полів за допомогою силових ліній

Лінії напруженості мають початок і кінець або прямують у нескінченність чи з нескінченності. Згідно з формулою (1.6) і рис. 1.2 вони починаються на позитивному заряді й закінчуються на негативному заряді. ***Лінії напруженості ніколи не перетинаються***, інакше в точці перетину силових ліній були б дві різні напруженості електричного поля. Ці лінії проводять з такою густиною, щоб кількість ліній, які пронизують одиничну площину, перпендикулярну до вектора $\vec{E}$, чисельно дорівнювала величині вектора $\vec{E}$ електричного поля в місці розміщення площини. На рис. 1.4 наведено приклади графічного зображення електричних полів за допомогою ліній напруженості для *а* і *б* – позитивного і негативного зарядів, *в* – системи позитивного і негативного зарядів, *г* – системи двох негативних зарядів, *д* – зарядженої негативно площини, *е* – системи двох різнойменно заряджених однаковими за величиною зарядами та однакових за розмірами площини.

Поле, у всіх точках якого величина і напрямок вектора напруженості незмінні, називають ***однорідним***. Воно утворюється між зарядженими площинами, якщо вони паралельні й нескінченно великі (рис. 1.4, *е*). Однорідне поле зображують паралельними лініями напруженості з однаковою густиною.

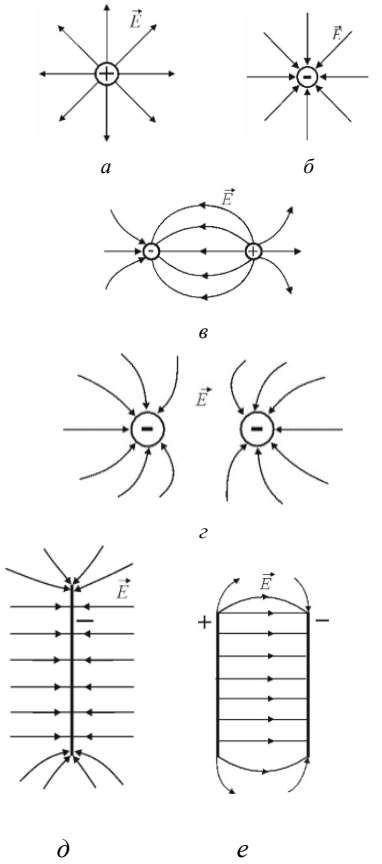


Рис. 1.4. приклади графічного зображення електричних полів за допомогою ліній напруженості

1.6. Потік вектора напруженості електричного поля

Зв'язок між електричним полем і його джерелами можна виразити простим способом, який буде мати багато застосувань. Розпочнемо із визначення величини, яку називають ***потоком*** вектора напруженості електричного поля.

Нехай в однорідному електричному полі ($\vec{E} = \text{const}$) перебуває плоска поверхня S, нормаль $\vec{n}$ якої утворює кут α з вектором $\vec{E}$ (рис. 1.5). Тоді величину

$$\Phi_E = \vec{E}\vec{n}S = |\vec{E}||\vec{n}|S\cos\alpha = \\ = ES\cos\alpha = E_n S \qquad (1.8)$$

будемо називати потоком вектора $\vec{E}$ через поверхню S (E_n – проєкція вектора $\vec{E}$ на нормаль $\vec{n}$). Утім, у загальному випадку електричні поля неоднорідні $\vec{E} = \\ = \vec{E}(x, y, z)$, а поверхня S не є плоскою (рис. 1.6). Таку поверхню S поділяємо на елементарні (нескінченно малі) площини dS, які можна вважати плоскими, поле в їх межах – однорідним, а $\vec{n}$ – нормаль до dS. Вибір додатного напрямку $\vec{n}$ умовний (вважають додатною зовнішню нормаль до поверхні S).

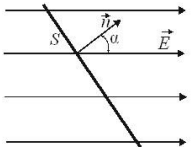


Рис. 1.5. Плоска поверхня в однорідному електричному полі

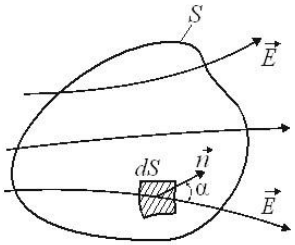


Рис. 1.6. Елементарна поверхня в неоднорідному електричному полі

Тоді елементарний потік вектора $\vec{E}$ через dS,

$$d\Phi_E = \vec{E}d\vec{S} = \vec{E}\vec{n}dS = \\ = |\vec{E}||\vec{n}|dS\cos\alpha = EdS\cos\alpha = E_n dS, \quad (1.9)$$

де формально $d\vec{S} = \vec{n}dS$.

Тепер складемо потоки через всі елементарні площини dS. Отримаємо скалярну величину Φ_E, яка є потоком вектора $\vec{E}$ через всю поверхню S:

$$\Phi_E = \int_S \vec{E}\, d\vec{S} = \int_S E_n dS. \quad (1.10)$$

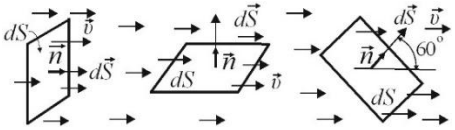
Знак потоку залежить від вибору напрямку нормалі, тобто кута α між $\vec{E}$ та $\vec{n}$.

Для замкнутих поверхонь

$$\Phi_E = \oint_S \vec{E}d\vec{S}, \quad (1.11)$$

де знак $\oint_S$ означає, що додавання елементарних потоків $d\Phi_E$ відбувається по замкнутій поверхні S.

Щоб зрозуміти походження назви «потік», виберемо векторну функцію, яка є швидкістю руху рідини, наприклад у річці, де швидкість змінюється від одного місця до іншого, але є постійною з часом у будь-якій точці. Позначимо це векторне поле через $\vec{v}$, м/с. Тоді, якщо dS (м^2)означає елемент площини, зануреної у воду, то добуток $\vec{v}d\vec{S} = |\vec{v}| \cdot |\vec{n}| \cdot \cos\alpha \cdot dS$ є *потоком* води через елемент площини, м^3/с (рис. 1.7).



потік $= vdS$ потік $= 0$ потік $= v\,dS\cos 60^\circ = 0{,}5vdS$

Рис. 1.7. Походження назви «потік»

Слід зауважити, що таке визначення потоку можна застосовувати до будь-якої векторної функції, якою б фізичною величиною вона не була.

1.7. Теорема Остроградського—Гауса

Розглянемо найбільш простий випадок розрахунку Φ_E. Припустимо, що поле створене ізольованим позитивним електричним зарядом q і що поверхнею S_1 є поверхня сфери, радіус якої r, у центрі якої розміщений цей точковий заряд (рис. 1.8). Чому дорівнює потік Φ_E через таку поверхню? Відповісти на це запитання легко, оскільки величина напруженості поля $|\vec{E}|$ у кожній точці сферичної поверхні S_1 згідно з (1.6) дорівнює $\frac{q}{4\pi\varepsilon_0 r^2}$ і напрямок $\vec{E}$ збігається із зовнішньою нормаллю в цій точці.

Тоді знаходимо:

$$\Phi_E = \oint_{S_1} \vec{E}\, d\vec{S}_1 = \oint_{S_1} E dS_1 =$$

$$= \oint_{S_1} \frac{q}{4\pi\varepsilon_0 r^2} dS_1 = \frac{q}{4\pi\varepsilon_0 r^2} \oint_{S_1} dS_1 =$$

$$= \frac{q}{4\pi\varepsilon_0 r^2} 4\pi r^2 = q/\varepsilon_0. \quad (1.12)$$

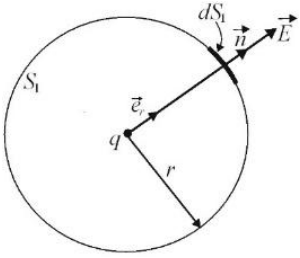


Рис. 1.8. Точковий заряд в центрі сферичної повехні

Це означає, що потік Φ_E не залежить від розмірів сфери і що силові лінії вектора $\vec{E}$ неперервні.

Тепер уявімо іншу поверхню S_2, яка охоплює поверхню S_1, але не має сферичної форми (рис. 1.9).

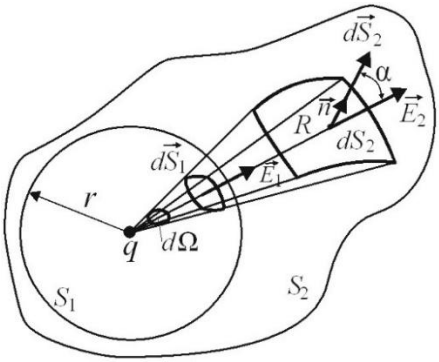


Рис. 1.9. Точковий заряд оточений замкнутою поверхнею не сферичної форми

Доведемо, що потік вектора $\vec{E}$ через поверхню S_2 дорівнює потоку $\vec{E}$ через сферичну поверхню S_1. Для доведення розглянемо конус, який виходить із q й вирізає елементарну ділянку dS_1 на поверхні сфери S_1, продовжується до зовнішньої поверхні S_2 й вирізає на ній елемент площею dS_2 на відстані R від заряду q. Знайдемо потоки через dS_1 та dS_2:

$$d\Phi_{E_1} = \vec{E}_1 d\vec{S}_1 = \vec{E}_1 \vec{n}\, dS_1 = |\vec{E}_1||\vec{n}| \times$$

$$\times \cos 0° \cdot dS_1 = \frac{q}{4\pi\varepsilon_0 r^2} \cdot dS_1 =$$

$$= \frac{q}{4\pi\varepsilon_0} d\Omega, \quad (1.13)$$

де $d\Omega = \frac{dS_1}{r^2}$ – тілесний кут при вершині конуса;

$$d\Phi_{E_2} = \vec{E}_2 d\vec{S}_2 = \vec{E}_2 \vec{n}\, dS_2 = |\vec{E}_2||\vec{n}| \times$$
$$\times \cos\alpha \cdot dS_2 = \frac{q}{4\pi\varepsilon_0 R^2} dS_{2n} = \frac{q}{4\pi\varepsilon_0} d\Omega, \quad (1.14)$$

де $dS_{2n} = dS_2 \cdot \cos\alpha$ – проєкція dS_2 на поверхню сфери, радіус якої R. У свою чергу, $dS_{2n} / R^2 = d\Omega$ – тілесний кут при вершині конуса, тобто кут, під яким із точки, де перебуває q, видно елемент поверхні dS_2.

Порівнявши (1.13) та (1.14) бачимо, що $d\Phi_{E_1} = d\Phi_{E_2}$. Кожну частину зовнішньої поверхні S_2 можна сумістити з відповідною частиною внутрішньої сферичної поверхні S_1 таким чином, що повний потік $\vec{E}$ буде однаковий через обидві поверхні S_1 і S_2. Інакше кажучи, повний потік через поверхню S_2 також буде дорівнювати q/ε_0, але поверхня S_2

має ***довільну*** форму. *Отже, потік вектора $\vec{E}$ через довільну поверхню, яка охоплює заряд q, дорівнює q/ε_0.* Можна довести, що повний потік вектора $\vec{E}$ через замкнуту поверхню S дорівнює нулю, якщо заряд q розміщений зовні поверхні. Це твердження очевидне для точкового заряду і сферичної поверхні (рис. 1.10). Потік вектора $\vec{E}_1$ через елементарну площину dS_1 становить

$$d\Phi_1 = \vec{E}_1\vec{n}\,dS_1 = |\vec{E}_1||\vec{n}|\cos\alpha_1 \cdot dS_1 = \\ = \frac{q}{4\pi\varepsilon_0}\cdot\left(\frac{-dS_{1n}}{r_1^2}\right) = -\frac{q}{4\pi\varepsilon_0}d\Omega,$$

де $dS_1 \cdot \cos\alpha_1 = -dS_{1n}$ − проєкція площини dS_1 на поверхню сфери радіусом r_1 (r_1 – відстань від заряду q до площини dS_1); $d\Omega$ – тілесний кут при вершині конуса, яка перебуває в точці розміщення заряду q ($\cos\alpha_1 < 0$).

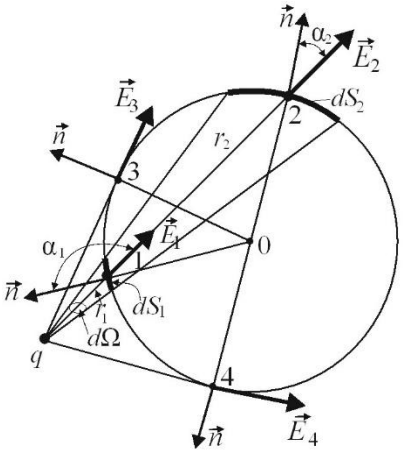


Рис. 1.10. Точковий заряд знаходиться поза сферичною поверхнею

Аналогічно потік вектора $\vec{E}_2$ через елементарну площину dS_2 становитиме

$$d\Phi_2 = \vec{E}_2\vec{n}\,dS_2 = |\vec{E}_2||\vec{n}|\cos\alpha_2 \cdot dS_2 = \\ = \frac{q}{4\pi\varepsilon_0}\cdot\left(\frac{dS_{2n}}{r_2^2}\right) = \frac{q}{4\pi\varepsilon_0}d\Omega,$$

де $dS_2 \cdot \cos\alpha_2 = dS_{2n}$ − проєкція ділянки dS_2 на поверхню сфери радіусом r_2 (r_2 – відстань від заряду q до ділянки dS_2); $d\Omega$ – тілесний кут при вершині конуса (точка розміщення заряду q). Сума $d\Phi_1 + d\Phi_2 = 0$.

Перебираючи попарно відповідні площини на поверхні сфери, доводимо, що сумарний потік вектора $\vec{E}$ у цьому випадку дорівнює нулю. Звертаємо увагу на те, що в дотичних точках 3 та 4 сфери скалярний добуток $\vec{E}\cdot\vec{n} = 0$, тому що вектори $\vec{E}$ і $\vec{n}$ взаємно перпендикулярні. Саме в цих точках потік вектора $\vec{E}$ змінює знак із від'ємного на додатний.

Із вищесказаного можна зроби-ти такий висновок: потік вектора $\vec{E}$ через замкнуту поверхню не залежить від її розмірів і форми. Це зв'язано із законом «обернених квадратів» у формулі (1.6).

Тепер застосуємо принцип суперпозиції. Будь-яке електричне поле є сумою полів окремих джерел (1.7). Зрозуміло, що потік є адитивною величиною в тому розумінні, що якщо матимемо деяку кількість джерел поля $q_1, q_2, \ldots, q_n$, поля кожного із яких мали б

напруженість $\vec{E}_1,\ \vec{E}_2, \dots, \vec{E}_n$ відповідно, то потік Φ_E через деяку замкнуту поверхню S у реально існуючому полі можна виразити рівністю

$$\Phi_E = \oint_S \vec{E} d\vec{S} = \oint_S (\vec{E}_1 + \vec{E}_2 + \dots + \vec{E}_n) d\vec{S}.$$

Вище було встановлено (формула (1.12)), що $\oint_S \vec{E}_i\, d\vec{S} = q_i / \varepsilon_0$, якщо заряд q_i розміщений всередині поверхні *S*, а $\oint_S \vec{E}_i\, d\vec{S}$ дорівнює нулю, якщо заряд q_i, розміщений зовні поверхні *S*. Таким чином, кожний заряд q всередині поверхні S вносить у потік q/ε_0, а всі заряди зовні поверхні нічого не вносять у сумарний потік.

Таким чином отримуємо закон (теорему) Остроградського−Гауса: ***потік вектора напруженості $\vec{E}$ електричного поля у вакуумі через довільну замкнуту поверхню S дорівнює алгебраїчній сумі тих зарядів q, які охоплюються поверхнею S, поділеною на електричну сталу вакууму ε_0:***

$$\oint_S \vec{E} d\vec{S} = \oint_S E_n dS = \frac{1}{\varepsilon_0} \sum_{i=1}^{n} q_i. \quad (1.15)$$

Теорему Остроградського−Гауса можна називати законом, оскільки вона еквівалентна закону Кулона, її так само можна вважати ***основним законом електростатичної взаємодії*** після визначення заряду і поля. Закони Остроградського−Гауса і Кулона не є двома незалежними фізичними законами, а є одним і тим самим законом, який виражений у різних формах.

Повертаючись до доведення теореми можна побачити, що воно ґрунтується на оберненій пропорційності взаємодії зарядів квадрату відстані між ними, і, звичайно, на адитивності взаємодій, тобто на принципі суперпозиції. Теорему Остроградського − Гауса можна застосовувати до будь-якого фізичного поля, в якому діє закон «обернених квадратів», наприклад до гравітаційного поля, в якому діє сила (1.1).

1.8. Застосування теореми Остроградського−Гауса до розрахунку електричних полів

Розглянемо застосування теореми Остроградського − Гауса до розрахунку електричних полів певної симетрії.

1. *Електричне поле рівномірно зарядженої нескінченної площини.* Площину можна вважати нескінченною, якщо відстанню від точки, в якій визначається напруженість поля, до площини можна знехтувати порівняно з геометричними розмірами площини. Поверхнева густина заряду σ − це кількість заряду, що припадає на одиничну площу площини ($[\sigma] = = \text{Кл}/\text{м}^2$). Електричне поле по обидва боки площини, якою б не була його

величина, має бути направлене перпендикулярно площині; у системі не може бути іншого виділеного напрямку (рис. 1.11). У точках P і P', які розміщені на однаковій відстані по обидва боки від площини, вектори напруженості $\vec{E}_P$ і $\vec{E}_{P'}$ поля через симетрію повинні мати однакову величину, але протилежні напрямки. Намалюємо циліндр (рис. 1.11) з поперечним перерізом S, на одному з кінців якого розміщено точку P, а на іншому – P'. Потік вектора $\vec{E}$ через бічну поверхню циліндра дорівнює нулю, оскільки лінії напруженості поля паралельні твірним і не перетинають його бічної поверхні. Через основи циліндра потік дорівнює $SE_P+SE_{P'} = 2SE_P$.

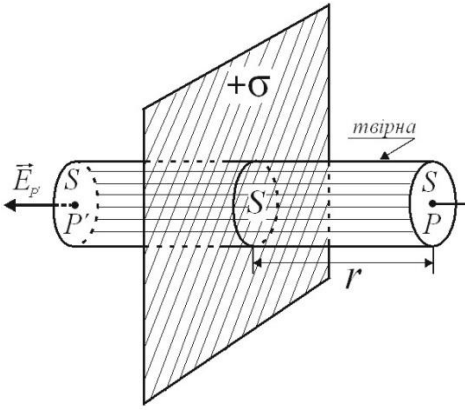


Рис. 1.11. Вибір циліндричної замкнутої повехні для розрахунку електричного поля площини

Вибрана циліндрична поверхня охоплює заряд, який розміщено на поверхні S площини, величина якого дорівнює σS. Тоді за теоремою Остроградського−Гауса (1.15) $2ES = \sigma S/\varepsilon_0$. Звідси

$$E = \sigma/(2\varepsilon_0). \qquad (1.16)$$

Незалежність $\vec{E}$ від відстані r свідчить про те, що поле нескінченної рівномірно зарядженої площини в усьому просторі поза нею однорідне (граничні ефекти не враховуємо). Співвідношення (1.16) можна отримати більш складним способом, використовуючи принцип суперпозиції (1.7).

2. *Електричне поле між двома різнойменно зарядженими паралельними нескінченними площинами.* Будемо вважати, що поверхнева густина зарядів площин однакові за величиною $|+\sigma| = |-\sigma|$. За принципом суперпозиції можна стверджувати, що в областях за межами площин поля немає (у кожній точці простору за межами площин значення $\vec{E}$ від обох заряджених площин однакові за величиною і протилежні за знаком, тобто $\vec{E}^+ + \vec{E}^- = 0$), а в областях між площинами напрямки $\vec{E}$ від обох заряджених площин мають однаковий напрямок (рис. 1.12), й отже,

$$E = E^+ + E^- = \frac{\sigma}{2\varepsilon_0} + \frac{\sigma}{2\varepsilon_0} = \frac{\sigma}{\varepsilon_0}. \qquad (1.17)$$

Електричне поле двох різнойменних заряджених площин локалізова-

не в об'ємі між цими двома площинами і є однорідним.

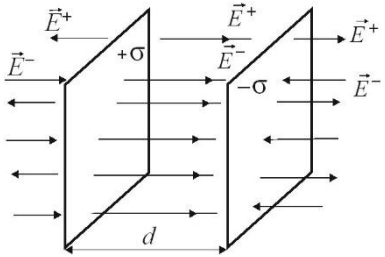


Рис. 1.12. Електричне поле між двома різнойменно зарядженими паралельними нескінченними площинами

3. *Електричне поле лінійного заряду*. Довгий прямий заряджений провідник, якщо знехтувати його товщиною, можна характеризувати кількістю електрики (заряду) на одиницю довжини. Позначимо цю *лінійну густину заряду*, яку вимірюють у СІ в кулонах на метр (Кл/м), літерою λ. З міркувань симетрії випливає, що напруженість поля у будь-якій точці має бути направлена вздовж радіальної прямої, яка перпендикулярна до проводу, а величина напруженості може залежати від відстані r точки до проводу (рис. 1.13). Уявімо циліндр, віссю якого є заряджений провід, висота циліндра h, а радіус основи r (рис. 1.13).

Потік вектора $\vec{E}$ через торці (основи) циліндра дорівнює нулю. Потік через бічну поверхню циліндра дорівнює площі $2\pi rh$, помноженій на напруженість поля на поверхні $E(r)$. Утім, заряд, який охоплює циліндрична поверхня, дорівнює λh. Таким чином, згідно з теоремою (1.15) можемо записати, що $2\pi rhE(r) = \lambda\, h/\varepsilon_0$. Звідси

$$E(r) = \frac{\lambda}{2\pi\varepsilon_0 r}. \qquad (1.18)$$

Поле лінійного заряду є неоднорідним.

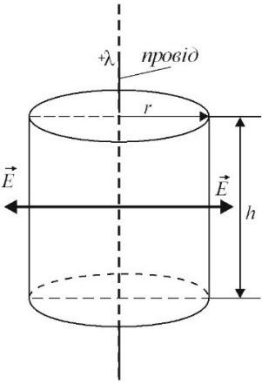


Рис. 1.13. Заряджений прямий провід і замкнута циліндрична повехня

4. *Електричне поле рівномірно зарядженої нескінченної циліндричної поверхні.* Якщо радіус зарядженої циліндричної поверхні R, то для відстані $r > R$ від осі циліндра отримаємо випадок, що розглядався у пункті 3. Якщо $r < R$, то допоміжна циліндрична поверхня не містить зарядів всередині, унаслідок чого $E(r < R) = 0$. Таким чином, для зарядженого циліндра, щоб визначити напруженість поля за $r > R$, можна

використовувати формулу (1.18). Поле циліндра є неоднорідним.

5. *Електричне поле сферичного розподілу заряду.* Розглянемо поле сферично-симетричного розподілу заряду, тобто розподілу, в якому об'ємна густина заряду ρ ($\rho = \Delta q/\Delta V$ – заряд одиниці об'єму, який вимірюють у кулонах на метро кубічний (Кл/м^3), *залежить тільки від відстані до центральної точки.* На рис. 1.14 зображено поперечний переріз такого розподілу.

Густина заряду в цьому розподілі нерівномірна: у центрі вона більша, ніж на деякій відстані від нього, потім зменшується, потім знову збільшується, і на відстані, більшій за R, дорівнює нулю. Слід визначити напруженість електричного поля у деякій точці P_1, яка розміщена за межами розподілу заряду, і в точці P_2 всередині нього (рис. 1.15).

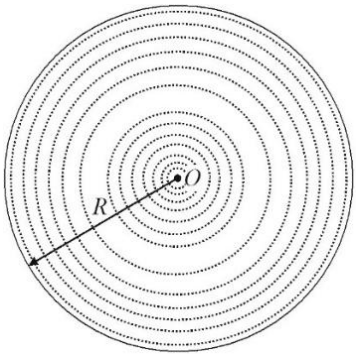


Рис. 1.14. Сферичний розподіл заряду

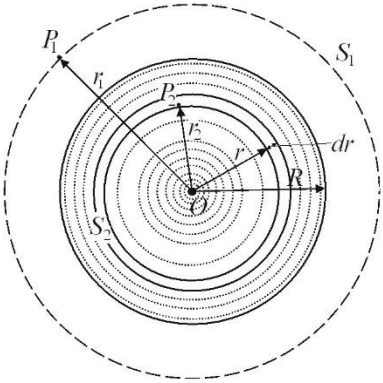


Рис. 1.15. Густина заряду залежить тільки від відстані до центральної точки

Через сферичну симетрію електричне поле в будь-якій точці має бути направлене по радіусу – другого напрямку бути не може. Величина поля E має бути однаковою у всіх точках сферичної поверхні S_1, радіус якої r_1, оскільки всі такі точки є еквівалентними.

Позначимо величину цього поля E_1. Тоді потік цього поля через поверхню S_1 дорівнює $4\pi r_1^2 E_1$ і за теоремою Остроградського−Гауса має дорівнювати заряду, що охоплює ця поверхня, поділеному на ε_0. Таким чином,

$$E_1 = \frac{\text{заряд всередині } S_1}{4\pi\varepsilon_0 r_1^2}. \qquad (1.19)$$

Порівнюючи цю формулу з полем точкового заряду (1.6), видно, що поле у всіх точках поверхні S_1 є таким самим, якби весь заряд всередині S_1 був зосереджений у центрі O. Аналогічне твердження можна застосувати до будь-

якої сфери, проведеної всередині поділу заряду, центр якої міститься в точці O. Поле у будь-якій точці поверхні S_2 є таким, якби весь заряд всередині S_2 був зосереджений у центрі O, а зовні S_2 зарядів не було б.

Розрахувати заряд всередині поверхонь S_1 або S_2 можна так. Загальна формула для розрахунку заряду в деякому об'ємі V така:

$$q = \int_V \rho dV,$$

де dV – елементарний об'єм, у межах якого $\rho = \text{const}$. На рис. 1.15 виберемо сфери, радіуси яких r і $r + dr$, де dr – нескінченно малий приріст радіуса r сфери. Тоді за об'єм dV візьмемо простір між цими двома сферами. Цей об'єм $dV = 4\pi r^2 dr$, тому формула для розрахунку заряду всередині сфери, радіус якої r, для сферично-симетричного розподілу заряду така:

$$q = \int_0^r 4\pi\rho(r) r^2\, dr. \qquad (1.20)$$

Таким чином заряд всередині поверхні S_1 становить

$$q = \int_0^{r_1} 4\pi\rho(r) r^2\, dr,$$

а заряд всередині поверхні S_2 –

$$q = \int_0^{r_2} 4\pi\rho(r) r^2\, dr.$$

Розглянемо окремі випадки.

6. *Електричне поле рівномірно зарядженої сфери, радіус якої R* (рис. 1.16).

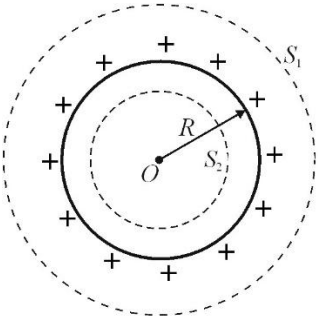


Рис. 1.16. Рівномірно заряджена сфера

Оскільки всередині поверхні S_2 зарядів немає, то напруженість поля всередині сфери дорівнює нулю. Для поверхні S_1 заряд всередині неї дорівнює заряду сфери, а напруженість поля для відстаней $r > R$ можна розрахувати за формулою (1.19).

7. *Електричне поле однорідної зарядженої кулі, радіус якої R.* Для цього випадку $\rho = \text{const}$ (рис. 1.17). Тоді заряд всередині поверхні S_1 становить

$$q = \int_0^R 4\pi\rho r^2 dr = \frac{4\pi\rho R^3}{3},$$

а напруженість поля у точках поверхні S_1 визначають за формулою (1.19):

$$E_1 = \frac{\rho R^3}{3\varepsilon_0 r_1^2} = \frac{q}{4\pi\varepsilon_0 r_1^2}.$$

Заряд всередині поверхні S_2 за формулою (1.20) становить

$$q_2 = 4\pi\rho \int_0^{r_2} r^2 dr = \frac{4\pi\rho r_2^3}{3}.$$

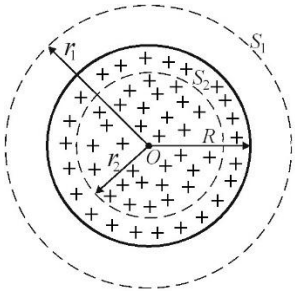


Рис. 1.17. Однорідно заряджена куля

Тоді за формулою (1.19) напруженість поля у точках на поверхні сфери S_2 буде

$$E_2 = \frac{q_2}{4\pi\varepsilon_0 r_2^2} = \frac{\rho}{3\varepsilon_0} r_2. \qquad (1.21)$$

Це означає, що напруженість поля всередині кулі лінійно зростає з відстанню від центра кулі O від нуля до $\rho R/(3\varepsilon_0)$. На рис. 1.18 наведено графік залежності E від r для рівномірно зарядженої кулі.

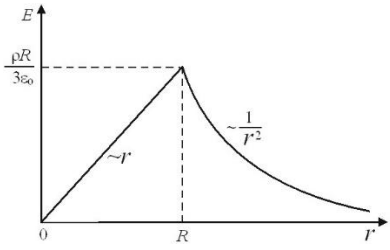


Рис. 1.18. Графік залежності E від r для рівномірно зарядженої кулі

1.9. Робота сил електростатичного поля

Обчислимо роботу сил електростатичного поля заряду q під час переміщення пробного заряду $q_{\text{пр}}$ в цьому полі з точки *1*, що перебуває на відстані r_1 від заряду q, у точку *2* на відстані r_2 від нього (рис. 1.19).

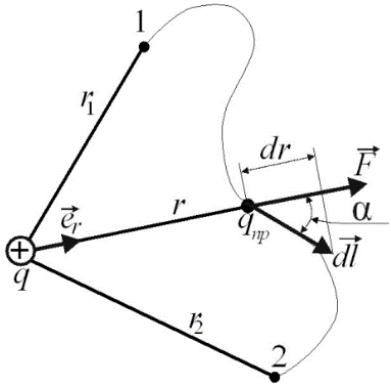


Рис. 1.19. Переміщення заряду в електричному полі

У будь якій точці цього поля на заряд $q_{пр}$ діє сила Кулона (1.3):

$$\vec{F} = \frac{1}{4\pi\varepsilon_0}\frac{qq_{\text{пр}}}{r^2}\vec{e}_r.$$

Робота сили $\vec{F}$ на елементарному переміщенні $d\vec{l}$ становить

$$dA = \vec{F}d\vec{l} = Fdl\cos\alpha = \\ = \frac{1}{4\pi\varepsilon_0}\frac{qq_{\text{пр}}}{r^2}dl\cos\alpha = \frac{qq_{\text{пр}}dr}{4\pi\varepsilon_0 r^2},$$

де dr – проєкція вектора $d\vec{l}$ на напрямок дії сили $\vec{F}$ (рис. 1.19). Робота під час переміщення заряду $q_{\text{пр}}$ із точки *1* в точку *2* становить

$$A_{12} = \int_{r_1}^{r_2} dA = \frac{qq_{\text{пр}}}{4\pi\varepsilon_0}\int_{r_1}^{r_2}\frac{dr}{r^2} = \\ = \frac{qq_{\text{пр}}}{4\pi\varepsilon_0}\left(-\frac{1}{r}\Big|_{r_1}^{r_2}\right),$$

звідки

$$A_{12} = \frac{qq_{\text{пр}}}{4\pi\varepsilon_0 r_1} - \frac{qq_{\text{пр}}}{4\pi\varepsilon_0 r_2}. \quad (1.22)$$

Із (1.22) випливає, що робота сил електростатичного поля не залежить від форми способу переміщення заряду $q_{\text{пр}}$ між точками *1–2*, її визначають лише розміщенням початкової *1* (r_1) і кінцевої *2* (r_2) точок. Силові поля, які задовольняють таку умову, називають *потенціальними*, або *консервативними*. Отже, електростатичне поле точкового заряду є потенціальним. Цей висновок можна поширити на випадок електричних полів будь-якої системи нерухомих точкових зарядів.

1.10. Різниця потенціалів. Потенціал. Еквіпотенціальні поверхні

Роботу сил консервативного поля (1.22) виконують за рахунок зменшення потенціальної енергії, тобто

$$A_{12} = W_{p_1} - W_{p_2} = -(W_{p_2} - W_{p_1}), \quad (1.23)$$

де

$$W_{p_1} = \frac{qq_{\text{пр}}}{4\pi\varepsilon_0 r_1} + \text{const},$$

а

$$W_{p_2} = \frac{qq_{\text{пр}}}{4\pi\varepsilon_0 r_2} + \text{const}$$

потенціальна енергія заряду $q_{\text{пр}}$ у полі заряду q на відстанях r_1 і r_2 від нього відповідно. У загальному випадку

$$W_p = \frac{qq_{\text{пр}}}{4\pi\varepsilon_0 r} + \text{const}. \quad (1.24)$$

Значення константи у виразі (1.24) вибирають таким чином, щоб у разі видалення заряду $q_{\text{пр}}$ від заряду q на нескінченність (тобто $r \to \infty$) потенціальна енергія дорівнювала нулю. За такої умови (1.24) набуває вигляду

$$W_p = \frac{qq_{\text{пр}}}{4\pi\varepsilon_0 r}. \quad (1.25)$$

Якщо заряди q і $q_{\text{пр}}$ однойменні, то потенціальна енергія їхньої взаємодії (відштовхування) додатна і зростає у разі зближення цих зарядів (рис. 1.20).

У випадку взаємного притягання різнойменних зарядів потенціальна енергія їхньої взаємодії від'ємна і зменшується під час зближення зарядів.

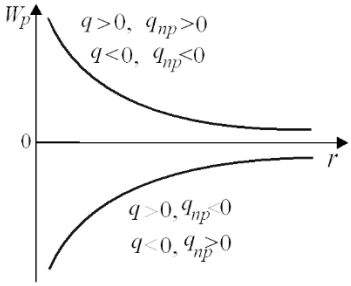


Рис. 1.20. Потенціальна енергія двох зарядів

Для різних пробних зарядів $q'_{\text{пр}}$, $q''_{\text{пр}}$ і т. д. будуть виконуватись різні роботи (1.22), однак відношення $A_{12}/q_{\text{пр}}$ буде однаковим для всіх пробних зарядів. Величину

$$\varphi_1 - \varphi_2 = A_{12}/q_{\text{пр}} \quad (1.26)$$

називають різницею потенціалів електростатичного поля між точками *1* і *2*. Вираз $\varphi_1 - \varphi_2$ однозначно визначають через роботу (1.22):

$$\varphi_1 - \varphi_2 = \frac{q}{4\pi\varepsilon_0 r_1} - \frac{q}{4\pi\varepsilon_0 r_2}, \quad (1.27)$$

де

$$\varphi_1 = \frac{q}{4\pi\varepsilon_0 r_1} + \text{const},$$

а

$$\varphi_2 = \frac{q}{4\pi\varepsilon_0 r_2} + \text{const}$$

називають потенціалами поля у точках *1* і *2*. Потенціал, як і потенціальна енергія (1.24), визначають з точністю до константи, тобто

$$\varphi = \frac{q}{4\pi\varepsilon_0 r} + \text{const}. \quad (1.28)$$

Якщо вважати, що на нескінченно великій відстані r ($r \to \infty$) від заряду q потенціал дорівнює нулю, то константу у (1.28) беруть за нуль і

$$\varphi = \frac{q}{4\pi\varepsilon_0 r}. \quad (1.29)$$

Якщо порівняти формули (1.25) і (1.29), то

$$\varphi = W_p / q_{\text{пр}}. \quad (1.30)$$

Із (1.30) випливає, що потенціалом φ будь-якої точки електростатичного поля називають фізичну величину, яка чисельно дорівнює потенціальній енергії одиничного позитивного заряду, поміщеного в цю точку.

Із формули (1.26) ще можемо записати, що

$$\varphi_1 = A_{1\infty} / q_{\text{пр}},$$

де враховано, що $q_{\text{пр}}$ із точки *1* переміщено на нескінченність ($r_2 \to \infty$).

Отже, ***потенціал*** цієї точки електростатичного поля – це така фізична величина, яка чисельно дорівнює роботі, яку виконують зовнішні сили (проти сил електростатичного поля) під час переміщення одиничного позитивного заряду з нескінченності в цю точку поля.

Одиниця потенціалу – вольт, 1 В – це потенціал такої точки поля, в якій заряд величиною 1 Кл має потенціальну енергію 1 Дж.

Робота є адитивною величиною. Нехай поле створюється системою n точкових зарядів $q_1, q_2, \ldots, q_n$. Робота, яку виконують сили цього поля над зарядом $q_{\text{пр}}$, буде дорівнювати алгебраїчній сумі робіт сил, обумовлених кожним зарядом окремо: $A_{12} = \sum_{i=1}^{n} A_i$. Згідно з (1.22) кожна з робіт A_i становить

$$A_i = \frac{1}{4\pi\varepsilon_0}\left(\frac{q_i q_{\text{пр}}}{r_{i1}} - \frac{q_i q_{\text{пр}}}{r_{i2}}\right),$$

де r_{i1} – відстань від заряду q_i до початкового положення заряду $q_{\text{пр}}$; r_{i2} – відстань від q_i до кінцевого положення заряду $q_{\text{пр}}$. Отже,

$$A_{12} = \frac{1}{4\pi\varepsilon_0}\sum_{i=1}^{n}\frac{q_i q_{\text{пр}}}{r_{i1}} - \\ - \frac{1}{4\pi\varepsilon_0}\sum_{i=1}^{n}\frac{q_i q_{\text{пр}}}{r_{i2}}.$$

Зіставивши цей вираз із співвідношенням (1.23), отримаємо для потенціальної енергії заряду $q_{\text{пр}}$ у полі системи зарядів вираз

$$W_p = \frac{1}{4\pi\varepsilon_0}\sum_{i=1}^{n}\frac{q_i q_{\text{пр}}}{r_i},$$

з якого випливає, що

$$\varphi = \frac{W_p}{q_{\text{пр}}} = \frac{1}{4\pi\varepsilon_0}\sum_{i=1}^{n}\frac{q_i}{r_i}. \qquad (1.31)$$

Зіставляючи формули (1.31) і (1.29), можна зробити висновок, що потенціал поля, яке створене системою зарядів, дорівнює алгебраїчній сумі потенціалів, які створюються кожним із зарядів окремо. Якщо значення напруженості полів складаються під час накладання полів векторно, потенціали складаються алгебраїчно. Через це розрахунок потенціалів є більш простим, ніж розрахунок напруженості електричного поля:

Геометричне місце точок з однаковим потенціалом називають еквіпотенціальною поверхнею. Її рівняння має такий вигляд:

$$\varphi(x, y, z) = \text{const}.$$

Під час переміщення $q_{\text{пр}}$ по еквіпотенціальній поверхні на елементарний відрізок $d\vec{l}$ потенціал не змінюється, а отже, робота

$$A_{12} = q_{\text{пр}}(\varphi_1 - \varphi_2) = \int_1^2 \vec{F}d\vec{l} =$$
$$= \int_1^2 q_{\text{пр}}\vec{E}d\vec{l} = 0.$$

Звідси скалярний добуток $\vec{E}d\vec{l} = 0$. Оскільки $|\vec{E}| \neq 0$, $|d\vec{l}| \neq 0$, то $\vec{E}d\vec{l} = 0$ за умови, що кут між $\vec{E}$ і $d\vec{l}$ дорівнює $\pi/2$. Це означає, що вектор $\vec{E}$ напруженості електричного поля в кожній точці напрямлений перпендикулярно до еквіпотенціальної поверхні.

Еквіпотенціальні поверхні точкового заряду – це сферичні оболонки навколо нього (рис. 1.21) ($\varphi_1 > \varphi_2 > \varphi_3$).

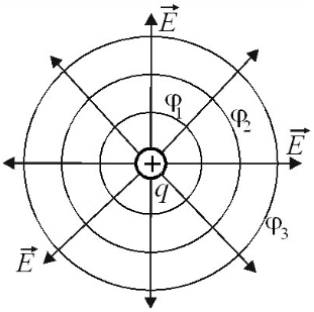


Рис. 1.21. Еквіпотенціальні поверхні точкового заряду

1.11. Циркуляція вектора напруженості електричного поля

З формули (1.22) також випливає, що під час переміщення точкового заряду в електростатичному полі по довільному замкнутому контуру Γ ($r_1 = r_2$) робота дорівнює нулю. Математично цю умову записують так:

$$A = \oint_\Gamma \vec{F}\,d\vec{l} = \oint_\Gamma q_{\text{пр}}\vec{E}d\vec{l} =$$
$$= q_{\text{пр}}\oint_\Gamma \vec{E}d\vec{l} = 0.$$

Оскільки $q_{\text{пр}} \neq 0$, то

$$\oint_{\Gamma} \vec{E} d\vec{l} = 0. \qquad (1.32)$$

Лінійний інтеграл $\oint_{\Gamma} \vec{E} d\vec{l}$, обчислений за довільним замкнутим контуром Г, називають *циркуляцією вектора* $\vec{E}$.

Циркуляція вектора $\vec{E}$ по контуру Г – це сума скалярних добутків $\vec{E}_1 d\vec{l}_1 + \vec{E}_2 d\vec{l}_2 + \vec{E}_3 d\vec{l}_3 + \cdots$, обчислених у кожній точці контуру Г (рис. 1.22).

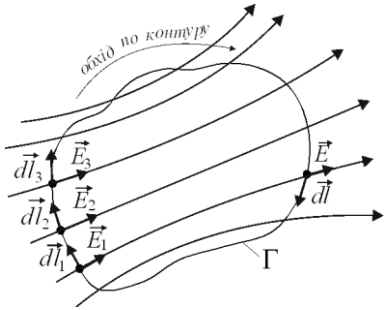


Рис. 1.22. Циркуляція вектора $\vec{E}$ *по контуру Г*

Співвідношення (1.32) є одним з фундаментальних рівнянь електростатики, яке відображає те, що силові лінії електростатичного поля незамкнуті – вони починаються на позитивних зарядах і закінчуються на негативних або прямують у нескінченність.

Під час переміщення пробного точкового заряду у такому полі по замкнутому контуру на одних ділянках шляху робота буде додатною, на інших – від'ємною, а повна робота завжди дорівнює нулю. Рівняння (1.32) є інтегральною формою запису потенціальності електростатичного поля.

1.12. Зв'язок між напруженістю та потенціалом електростатичного поля

Оскільки напруженість $\vec{E}$ і потенціал φ є різними за фізичним змістом характеристиками тих самих точок електростатичного поля, то між ними має існувати взаємозв'язок.

Нехай в електричному полі перебуває заряд $q_{\text{пр}}$. Робота під час переміщення цього заряду вздовж осі *OX* між двома нескінченно близькими точками буде такою:

$$dA = F_x dx = q_{\text{пр}} E_x dx.$$

І навпаки, елементарна робота під час переміщення заряду $q_{\text{пр}}$ в електростатичному полі виражають через різницю потенціалів між цими точками:

$$dA = q_{\text{пр}}(\varphi_1 - \varphi_2) = -q_{\text{пр}}(\varphi_2 - \varphi_1) = \\ = -q_{\text{пр}}(\varphi + d\varphi - \varphi) = -q_{\text{пр}} d\varphi,$$

де $\varphi_1 = \varphi$, а $\varphi_2 = \varphi + d\varphi$.

Тоді, прирівнявши елементарні роботи, отримуємо

$$E_x dx = -d\varphi, \text{ або } E_x = -\partial\varphi/\partial x.$$

Знак «мінус» означає, що під дією сил електричного поля додатний заряд переміщується у бік зменшення потенціалу. Похідна $\partial\varphi/\partial x$ означає, що

$\varphi = \varphi(x, y, z)$, але в цьому випадку похідну береть тільки за змінною x.

Аналогічні міркування можна поширити і на напрямки переміщень вздовж осей *OY* та *OZ*:

$$E_y = -\frac{\partial \varphi}{\partial y}; \qquad E_z = -\frac{\partial \varphi}{\partial z}.$$

Таким чином знаходимо E_x, E_y, E_z – компоненти вектора напруженості $\vec{E}$. Вираз для $\vec{E}$ можемо записати так:

$$\vec{E} = E_x\vec{\imath} + E_y\vec{\jmath} + E_z\vec{k} = $$
$$= -\left(\frac{\partial \varphi}{\partial x}\vec{\imath} + \frac{\partial \varphi}{\partial y}\vec{\jmath} + \frac{\partial \varphi}{\partial z}\vec{k}\right).$$

У векторному аналізі градієнт скалярної величини φ називають таку векторну величину, для якої справедливий такий запис:

$$\text{grad}\varphi = \frac{\partial \varphi}{\partial x}\vec{\imath} + \frac{\partial \varphi}{\partial y}\vec{\jmath} + \frac{\partial \varphi}{\partial z}\vec{k}.$$

Цей вектор направлений у бік найшвидшого зростання потенціалу φ.

Отже,

$$\vec{E} = -\text{grad}\varphi. \qquad (1.33)$$

Знак «–» вказує на те, що вектор $\vec{E}$ напруженості поля направлений у бік найшвидшого зменшення потенціалу. Напруженість у будь-якій точці електростатичного поля дорівнює градієнту потенціалу в цій точці поля, взятому з оберненим знаком.

Знаючи потенціал φ в кожній точці поля, за формулою (1.33) можемо обчислити напруженість у кожній точці поля.

Можна розв'язати й обернену задачу, тобто знаючи напруженість поля у кожній точці поля, можна знайти різницю потенціалів між довільними точками. Робота під час переміщеня заряду $q_{\text{пр}}$ з точки *1* у точку *2* становить

$$A_{12} = \int_1^2 \vec{F}d\vec{l} = \int_1^2 q_{\text{пр}}\vec{E}d\vec{l} = $$
$$= q_{\text{пр}}(\varphi_1 - \varphi_2).$$

Звідси отримуємо, що

$$\varphi_1 - \varphi_2 = \int_1^2 \vec{E}d\vec{l}. \qquad (1.34)$$

Інтеграл можна брати вздовж довільної лінії, яка з'єднує точки *1* і *2*, оскільки електричне поле є потенціальним.

1.13. Рівняння Пуассона (теорема Остроградського–Гауса в диференціальній формі)

Співвідношення (1.15) виражає теорему Остроградського–Гауса в інтегральній формі. Величини, що входять у вираз теореми, характеризують різні точки електричного поля. Так, якщо вектор $\vec{E}$ характеризує точки деякої замкнутої поверхні, то електричний заряд стосується точок об'єму, обмеженого цією поверхнею.

Отримаємо аналогічний вираз цієї теореми в диференціальній формі, тобто знайдемо співвідношення між фізичними

величинами, які характеризують якусь точку поля. Для цього знайдемо потік вектора напруженості поля $\vec{E}$ через поверхню нескінченно малого прямокутного паралелепіпеда з ребрами $dx,\ dy,\ dz,$ який охоплює заряд, розподілений з об'ємною густиною ρ (рис. 1.23). Будемо вважати, що ρ = $= \rho(x, y, z)$ є неперервною функцією координат.

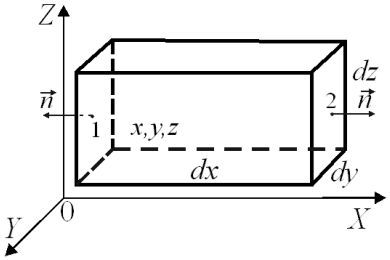


Рис. 1.23. Нескінченно малий прямокутний паралепіпед

Уявлення про неперервний розподіл електрики у просторі є такою ж ідеалізацію, як і уявлення про неперервний розподіл речовини. На грані *1* зовнішня нормаль направлена проти позитивного напрямку осі *OX*, тому потік вектора $\vec{E}$ через цю грань становитиме $-E_x(x)dydz$. На протилежній грані *2*, навпаки, напрямок зовнішньої нормалі $\vec{n}$ збігається з позитивним напрямком осі *OX* і для потоку вектора $\vec{E}$ через цю грань слід писати $+E_x(x + dx)dydz$. Сума обох потоків буде

$$(E_x(x + dx) - E_x(x))dydz = \\ = \frac{\partial E_x}{\partial x} dxdydz = \frac{\partial E_x}{\partial x} dV,$$

де $dV = dxdydz$ – об'єм елементарного паралепіпеда. Враховано, що приріст функції

$$E_x(x + dx) - E_x(x) = \frac{\partial E_x}{\partial x} dx,$$

де $\frac{\partial E_x}{\partial x}$ – похідна від $E_x(x, y, z)$ тільки за змінною *x*.

Аналогічно знаходимо потоки через дві пари останніх граней:

$$(E_y(y + dy) - E_y(y))dxdz = \frac{\partial E_y}{\partial y} dV;$$

$$(E_z(z + dz) - E_z(z))dxdy = \frac{\partial E_z}{\partial z} dV.$$

Повний потік через всю поверхню паралелепіпеда становитиме

$$d\Phi_E = (\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z})dV = \ d\text{iv}\vec{E}\ dV, \tag{1.35}$$

де введено позначення

$$\text{div}\vec{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}. \cdot \tag{1.36}$$

За теоремою Остроградського–Гауса

$$d\Phi_E = dq/\varepsilon_0 = \rho dV/\varepsilon_0, \tag{1.37}$$

де dq – заряд, що міститься всередині паралепіпеда. Прирівнюючи (1.35) і (1.37), отримаємо

$$\text{div}\vec{E} = \rho/\varepsilon_0. \tag{1.38}$$

Ця формула виражає теорему Остроградського–Гауса в диференціальній формі.

Величина, яку визначає вираз (1.36), зберігає зміст незалежно від

конкретної фізичної або геометричної природи ,вектора $\vec{E}$. Її називають ***дивергенцією*** вектора $\vec{E}$. Інтегральна і диференціальна форми теореми Остроградського–Гауса повністю еквівалентні.

Ці теореми справедливі не тільки в електростатиці, але і в електродинаміці, де розглядаються змінні з часом електромагнітні поля. Вся сукупність дослідних даних підтверджує це твердження. Отже, рівняння (1.15) і математично еквівалентне рівняння (1.38) перестають бути лише наслідками закону Кулона, натомість надають великого значення ***основним постулатам*** теорії електрики.

Контрольні запитання

1. Поясніть зміст закону Кулона та його фундаментальність для електростатичного поля.

2. Який фізичний зміст вектора напруженості електричного поля?

3. Дайте визначення силовій лінії електричного поля. Які її властивості та фізичний зміст?

4. Дайте визначення потоку вектора напруженості електричного поля.

5. Який зв'язок теореми Остроградського–Гауса для електростатичного поля із законом Кулона?

6. Яке застосування теореми Остроградського–Гауса до розрахунку електричного поля: точкового заряду, безмежної однорідної зарядженої площини, безмежно довгої однорідно зарядженої прямої нитки, однорідної зарядженої кулі?

7. Які є властивості та методи обчислення роботи з переміщення заряду в електричному полі?

8. Як визначити різницю потенціалів? Який її фізичний зміст?

9. У чому суть принципу суперпозиції електричних полів? Дайте визначення принципів для напруженості та потенціалу. Яка відмінність між цими принципами?

10. Який зв'язок між вектором напруженості $\vec{E}$ і потенціалом поля φ?

11. Яке взаємне розміщення силових ліній вектора $\vec{E}$ та еквіпотенціальних ліній (поверхонь) потенціалу φ електричного поля? Довести цю властивість їх взаємного розміщення.

12. Дайте визначення й запишіть вираз для рівняння Пуассона. Який його фізичний зміст і зв'язок з теоремою Остроградського–Гауса?

13. Що визначає циркуляція вектора напруженості поля $\vec{E}$ по

замкнутому контуру? Який її фізичний зміст?

Розділ 2. Поляризація діелектриків і поле всередині діелектрика

2.1. Провідники, напівпровідники та діелектрики

За здатністю проводити електричний струм усі тверді тіла поділяють на три великі групи (табл. 2.1): метали, напівпровідники і діелектрики (або ізолятори). Сюди можна додати й надпровідники – низку металів та їх сполуки, які за дуже низьких температур проводять струм без втрат енергії – їх електричний опір дорівнює нулю.

Таблиця 2.1

Провідники, напівпровідники та діелектрики

Матеріал	Питомий опір ρ, Ом·м	Провідні властивості
Провідники	$10^{-8}...10^{-5}$	Є "вільні" носії заряду (електрони), кількість яких практично не залежить від температури
Напівпровідники	$10^{-5}...10^{8}$	Є "вільні" носії заряду, кількість яких залежить (експоненціально) від температури
Діелектрики	$10^{8}...10^{16}$	При нормальних умовах відсутні "вільні" носії заряду

Найкращими провідниками є метали. Питомий опір ρ металів порядку $10^{-8}…10^{-5}$ Ом·м. У діелектриків питомий опір більший за 10^{8} Ом·м.

Діелектрики (англ. dielectric, від грец. diá – через, наскрізь та англ. electric – електричний) – це речовини, які погано проводять електричний струм. Термін «діелектрик» увів Фарадей для позначення речовин, в які проникає електричне поле. Діелектриками є гази (неіонізовані), деякі рідини і тверді тіла. Електропровідність діелектриків порівняно з металами дуже мала. Напівпровідники перебувають між провідниками та діелектриками.

На рис. 2.1 наведено значення питомого опору деяких металів, напівпровідників і діелектриків. Видно, що питомий опір у срібла менший, ніж у тефлону, в 10^{24} разів. Жодна фізична характеристика не змінюється так сильно під час переходу від речовини до речовини, як питомий опір.

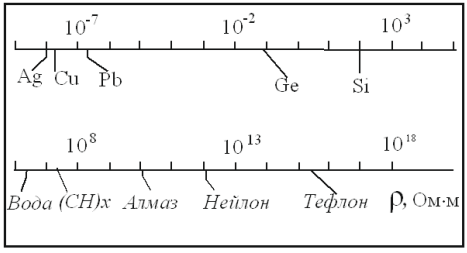


Рис. 2.1 Опір деяких металів, напівпровідників і діелектриків

Останнім часом у техніці почали використовувати і надпровідники. Основна перевага надпровідників у тому, що вони проводять електричний струм без втрат енергії. Недоліком наявних надпровідників, що їх треба охолоджувати до дуже низьких температур (один із високотемпературних сплавів Nb_3Ge переходить у надпровідний стан тільки за 21,3 K (−251,7 °C).

І діелектрики, і напівпровідники, і метали широко використовують у техніці – діелектрики, наприклад, як ізолятори. Назвемо деякі з них. До твердих діелектриків належать скло, гума, фарфор, слюда, пластмаси і полімерні матеріали, папір, дерево, алмаз, кварц та інші. Рідкими діелектриками є хімічно чиста (дистильована) вода, ацетон, гліцерин, спирт метиловий, хлорбензол, мінеральні нафтові мастила. Діелектриками є неіонізовані гази і пара речовини, що перебувають у не дуже сильних електричних полях, повітря, азот та інші.

Напівпровідниками є здебільшого багато як твердих речовин (наприклад, германій, кремній), так і хімічних сполук (наприклад, GaAs, CdTe, PbS). Напівпровідникові прилади використовують у багатьох пристроях, починаючи з побутової радіоапаратури і закінчуючи сучасними комп'ютерами, телефонами і под., створення яких неможливе без напівпровідникових елементів.

Із металів виготовляють з'єднувальні проводи, а також різні електросилові установки (генератори струму, електромотори, трансформатори і под.). На рис. 2.1 наведено значення питомого опору за $T = 293$ K і постійної напруги для електротехнічних матеріалів, а також їх основні відмінності з погляду провідних властивостей. Кількісну різницю в електропровідності діелектриків і металів класична фізика пояснює тим, що в металах є «вільні» електрони, а в діелектриках усі електрони міцно зв'язані з атомами.

Варто зауважити, що різкої межі між групами матеріалів немає, особливо умовної межі між діелектриками і напівпровідниками. У разі підвищення температури опір кристалічного діелектрика різко зменшується, й останній стає напівпровідником, і навпаки, у разі зниження температури напівпровідник стає діелектриком. За певних умов напівпровідники можуть поводитися як метал.

2.2. Електричний диполь в однорідному і неоднорідному електричних полях

Під час внесення діелектрика в електричне поле відбувається його поляризація, яка призводить до появи на його поверхнях, які перетинаються силовими лініями поля, нескомпенсованих зв'язаних електричних зарядів протилежних знаків. Інакше кажучи, дія електричного поля на діелектрик зводиться до зміни розподілу електричних зарядів, і в цьому суть явища поляризації діелектриків.

Для розуміння явища поляризації діелектриків треба розглянути поведінку електричного диполя в електричному полі. ***Електричний диполь*** – система двох однакових за величиною і протилежних за знаком точкових зарядів $+q$ і $-q$, які перебувають один від одного на деякій відстані l. Дипольний момент

$$\vec{p} = q\vec{l}, \qquad (2.1)$$

де вектор $\vec{l}$ має напрямок від заряду $-q$ до заряду $+q$ (напрямок осі диполя показано на рис. 2.2).

Сумарний заряд диполя дорівнює нулю, проте він створює електричне поле, яке має осьову симетрію, тому поле у будь-якій площині, яка проходить через вісь диполя, буде одним і тим самим, причому вектор $\vec{E}$ перебуває в цій площині. Положення точки відносно диполя будемо характеризувати за допомогою радіуса-вектора $\vec{r} = r\vec{e}_r$ або за допомогою полярних координат r і θ (рис. 2.2).

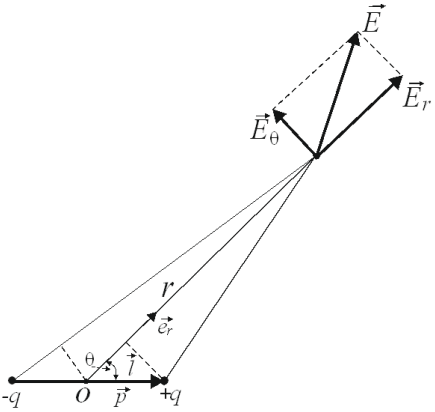


Рис. 2.2. Електричне поле диполя

Потенціал і напруженість поля диполя визначають за такими формулами:

$$\varphi(r,\theta) = \frac{1}{4\pi\varepsilon_0}\cdot\frac{p\cos\theta}{r^2}; \qquad (2.2)$$

$$E(r,\theta) = \frac{1}{4\pi\varepsilon_0}\cdot\frac{p}{r^3}\sqrt{1+3\cos^2\theta}. \qquad (2.3)$$

Порівняння формули (2.2) з (1.29), а (2.3) з (1.6) свідчить про те, що потенціал і напруженість поля диполя зменшуються з відстанню швидше, ніж потенціал і напруженість поля точкового заряду ($\varphi \sim 1/r$, $E \sim 1/r^2$). На рис. 1.4, *в* показано лінії вектора $\vec{E}$ для диполя.

Диполь в однорідному електричному полі. Результуюча сила

$\vec{F} = \vec{F}_1 + \vec{F}_2$, яка діє на диполь, дорівнює нулю, оскільки сили $\vec{F}_1$ і $\vec{F}_2$, які діють на позитивний і негативний заряди диполя, однакові за величиною і протилежні за напрямком (рис. 2.3). Ці сили утворюють пару сил. Момент пари сил M становить

$$M = qEl \sin \alpha, \qquad (2.4)$$

де α – кут між векторами $\vec{l}$ і напруженістю поля $\vec{E}$. Користуючись поняттям моменту диполя (2.1), вираз для моменту пари сил (2.4) можна записати у вигляді

$$M = pE \sin(\vec{p}, \vec{E}). \qquad (2.5)$$

Напрямок моменту цієї пари сил збігається з напрямком осі обертання диполя, тобто перпендикулярний до $\vec{p}$ та $\vec{E}$, тому вектор моменту пари сил $\vec{M}$, який діє на диполь, можна виразити формулою

$$\vec{M} = \vec{p} \times \vec{E}. \qquad (2.6)$$

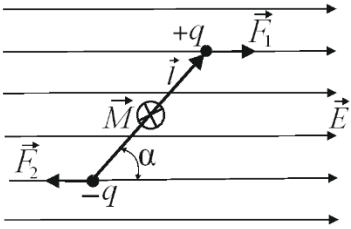


Рис. 2.3. Електричний диполь в однорідному зовнішньому електричному полі

Бачимо, що в однорідному полі на диполь діє тільки пара сил, яка прагне повернути диполь так, щоб $\vec{p}$ та $\vec{E}$ були паралельними ($\alpha = 0$).

Щоб повернути диполь в електричному полі на деякий кут, слід виконати певні дії. Через те, що ці дії дорівнюють збільшенню потенціальної енергії диполя, то звідси можна знайти вираз для енергії диполя в електричному полі. Візьмемо за нуль відліку енергію положення диполя, за якого його вісь перпендикулярна до напрямку поля ($\alpha = \pi/2$). Тоді енергія диполя, момент $\vec{p}$ якого утворює кут α з напрямком поля, становить

$$W_P = \int_{\frac{\pi}{2}}^{\alpha} pE \sin \alpha d\alpha = \\ = -pE \cos \alpha = -\vec{p}\,\vec{E}. \qquad (2.7)$$

Зауважимо, що вираз (2.7) не враховує енергії взаємодії зарядів, які утворюють диполь. Розміри диполя невеликі, тому ця формула справедлива і для неоднорідного поля.

Диполь в неоднорідному полі

Для спрощення будемо вважати, що момент $\vec{p}$ диполя паралельний напрямку поля ($\alpha = 0$) (рис. 2.4).

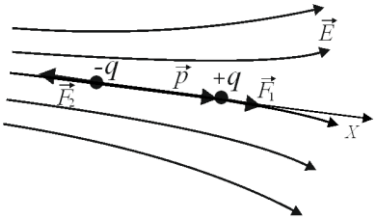


Рис. 2.4. Електричний диполь в неоднорідному електричному полі $\vec{E}$ у випадку, коли $\vec{p}$ направлений по полю $\vec{E}$. Сила $\vec{F}_2$ більша за силу $\vec{F}_1$, і результуюча сила $\vec{F}_2 - \vec{F}_1$ намагається перемістити диполь в область більшої напруженості зовнішнього електричного поля

У цьому випадку сили, які діють на кінці диполя, вже неоднакові, тому їх результуюча сила не дорівнює нулю. На диполь у неоднорідному полі діє сила, яка прагне пересунути диполь в область поля з більшою напруженістю. Знайдемо величину цієї сили. Направимо координатну вісь X уздовж моменту диполя і вважатимемо, що довжина диполя дуже мала (близько розмірів молекули) Сила, яка діє на негативний кінець диполя, становить $-qE$, де E – напруженість поля у точці із зарядом $-q$. Сила, яка діє на позитивний кінець диполя, дорівнює $+q\left(E-\frac{\partial E}{\partial x}l\right)$, де l – довжина диполя, $\partial E/\partial x$ – градієнт поля. Тому повна сила F_x становить

$$F_x = F_2 - F_1 = qE - q\left(E-\frac{\partial E}{\partial x}l\right) = \\ = ql\frac{\partial E}{\partial x} = p\frac{\partial E}{\partial x}.$$

Для випадку, коли дипольний момент $\vec{p}$ утворює кут α з напрямком поля $\vec{E}$, сила

$$F_x = p\frac{\partial E}{\partial x}\cos\alpha. \qquad (2.8)$$

В однорідному полі градієнт поля $\partial E/\partial x = 0$ і результуюча сила F_x дорівнює нулю.

Якщо диполь міститься в неоднорідному полі і непаралельний полю, то на нього діє і пара сил, яка прагне повернути диполь паралельно полю, і сила, яка втягує диполь в область сильнішого поля. Зрештою, диполь опиняється в області сильнішого поля так, щоб вектор $\vec{p}$ був зорієнтований по полю (за напрямком вектора $\vec{E}$).

2.3. Поляризація діелектриків та її типи. Вектор поляризації (поляризованість), діелектрична сприйнятливість

Явище обмеженого зміщення зарядів в атомах і молекулах або напрямленої орієнтації дипольних моментів жорстких молекул у зовнішньому електричному полі, внаслідок якого на поверхні діелектрика виникають зв'язані електричні заряди, називають поляризацією діелектриків.

Механізми поляризації діелектриків різні й залежать від характеру хімічного зв'язку. Наприклад, в іонних кристалах (NaCl і подібних) поляризація є результатом зсуву одних іонів відносно інших (іонна поляризація (рис. 2.5, *а*)) і деформації електронних оболонок окремих іонів (електронна поляризація). У кристалах з ковалентним зв'язком (наприклад, алмаз) поляризація обумовлена головним чином зміщенням електронів, які здійснюють хімічний зв'язок (рис. 2.5, *б*).

У полярних діелектриках (наприклад, твердий H_2S) молекули мають електричні дипольні моменти, які

без зовнішнього електричного поля $\vec{E}_0$ орієнтовані хаотично, а в полі набувають переважну орієнтацію за полем $\vec{E}_0$ (рис. 2.5, *в*). Така орієнтаційна поляризація типова для багатьох рідин і газів. Є ще один важливий з практичного погляду вид поляризації, який називають спонтанною.

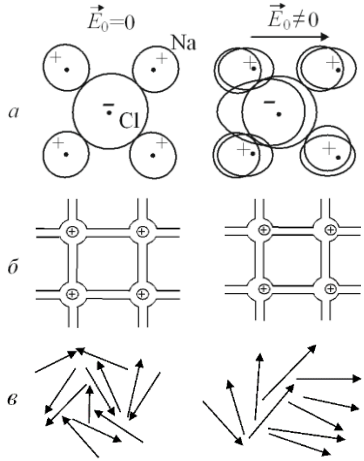


Рис. 2.5. Поляризація діелектриків: *а* – іонна («+» – іони Na^+, «–» – іони Cl^-); *б* – електронна («+» – іони C (вуглецю) кристалічних ґраток, суцільні хвилясті лінії – ковалентні (електронні) зв'язки); *в* – орієнтаційна (стрілки – дипольні моменти молекул H_2S, $\vec{E}_0$ – вектор напруженості зовнішнього електричного поля)

Спонтанна поляризація наявна у кристалічних речовинах – сегнетоелектриках. У деякому інтервалі температур у сегнетоелектричних кристалах є окремі, сильно поляризовані області – **домени** («домен» від лат. dominicus – господарський, панський). У середньовіччі «доменом» називали особисте володіння великого феодала. У фізиці слово «домен» використовують для позначення області, властивості якої різко різняться від властивостей матеріалу, який її оточує.

Діелектрики поділяють на такі типи:

1. ***Неполярні діелектрики*** – це діелектрики, які складаються з молекул, дипольний момент яких дорівнює нулю. Це можливо тоді, коли «центри мас» позитивних і негативних зарядів молекул збігаються ($l = 0$) без впливу електричного поля.

Прикладом неполярних діелектриків є гази: N_2, H_2, O_2, CO_2, CH_4. В електричному полі сили, які діють на позитивні й негативні заряди молекули, направлені в різні сторони і тому розсувають їх (рис. 2.5, *б*). Центр позитивних і центр негативних зарядів молекули зміщуються на відстань l, тим більшу, чим більша напруженість поля $\vec{E}_0$. Виникає дипольний момент молекули $p = ql$, і таким чином молекула перетворюється в полярну. Оскільки $l \sim E$, а $p = ql$, то $p \sim E$, або $p = \beta\varepsilon_0 E$, де E – напруженість поля всередині діелектрика, β – поляризованість окремої молекули діелектрика. Величина β має неоднакові значення для атомів і молекул різних речовин. Поляризованість β

характеризує здатність електронів в атомі або в молекулі зміщуватись під дією сил електричного поля. Дипольні моменти $\vec{p}$ молекул неполярних діелектриків називають індукованими, або квазіпружними.

Під час внесення неполярного діелектрика в електричне поле всі індуковані дипольні моменти розміщуються ланцюжками вздовж лінії напруженості $\vec{E}_0$ (рис. 2.6). Унаслідок цього грані діелектрика набувають різнойменних зарядів – діелектрик поляризується. Таку поляризацію називають ***електронною***.

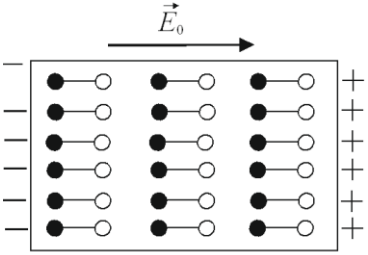


Рис. 2.6. Розміщення дипольних моментів молекул в електричному полі (світлі кружечки – позитивні заряди, темні кружечки – негативні заряди)

2. ***Полярні діелектрики – це речовини, в яких «центри мас» позитивних і негативних зарядів*** не збігаються. До полярних діелектриків належать гази (SO_2, H_2S, NH_3 та ін.) й рідини (вода H_2O, соляна кислота HCl, бензол C_6H_6 тощо).

Молекули таких діелектриків називають ***полярними***. Ці молекули без зовнішнього поля мають дипольні моменти $\vec{p}_0 = q\vec{l}$. Їх називають ***жорсткими диполями***. Величина p_0 має порядок $10^{-29} \ldots 10^{-30}$ Кл·м. Якщо такий діелектрик не перебуває у зовнішньому електричному полі, то внаслідок хаотичного теплового руху молекул вектори їхніх дипольних моментів орієнтовані хаотично (рис. 2.7, *а*).

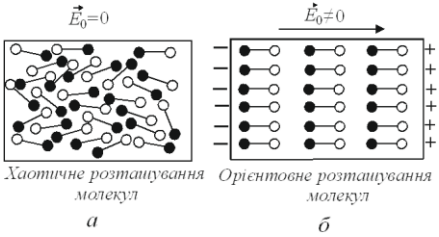


Рис. 2.7. Випадки розташування молекул: а – хаотичне; б – орієнтовне

Якщо діелектрик внести в електричне поле, то під дією поля полярні молекули діелектрика намагаються повернутись так, щоб вектори їх дипольних моментів збігались з напрямком вектора напруженості поля $\vec{E}_0$ (рис. 2.5, *в*; рис. 2.7, *б*). Утім, тепловий рух молекул хаотично розкидає диполі й заважає орієнтації всіх векторів $\vec{p}$ (дипольних моментів) уздовж поля. Орієнтація дипольних моментів буде тим повнішою, чим сильніше електричне поле і чим слабший тепловий рух

молекул, тобто нижча температура. Цей процес називають ***орієнтаційною поляризацією*** діелектрика.

Зауважимо, що молекули полярних діелектриків набувають додаткових, індукованих зовнішнім полем, дипольних моментів (як у випадку неполярних діелектриків), і це також впливає на їх поляризацію. Утім, наведений дипольний момент $p = \beta\varepsilon_0 E$ навіть за $E = 10^6$ В/м має порядок 10^{-34} Кл·м, тобто в середньому на чотири порядки менший від p_0, тому електронною поляризацією для полярних діелектриків можна знехтувати.

3. ***Іонні діелектрики – це речовини, молекули яких мають іонну будову.*** Прикладом таких діелектриків є NaCl, NaI та інші. Іонні кристали є просторовими решітками з правильним чергуванням іонів різних знаків (рис. 2.8).

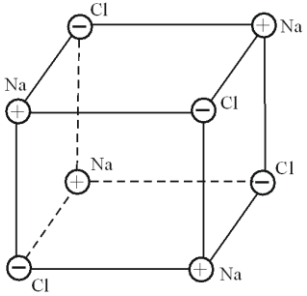


Рис. 2.8. Елементарна комірка кристалу NaCl

У цих кристалах не можна виділити окремі молекули. Іонні кристали розглядають як систему вміщених одна в одну іонних підграток. У цих діелектриках кожна пара сусідніх різнойменних іонів подібна до диполя. У зовнішньому електричному полі позитивні іони кристалічної ґратки зміщуються за напрямком напруженості поля $\vec{E}$, негативні – у протилежному напрямку. В результаті відбувається певна деформація ґратки або, як зазвичай говорять, – зміщення позитивної і негативної підграток, що приводить до виникнення дипольних моментів, тобто до поляризації діелектриків (рис. 2.5, *а*). ***Таку поляризацію називають іонною.*** Заряди, які виникають на гранях діелектрика, не вільні, вони пов'язані з атомами та молекулами речовини.

Таким чином, механізм поляризації діелектрика такий: під дією зовнішнього електричного поля молекули речовини набувають дипольних моментів (якщо їх раніше в них не було), які потім орієнтуються за полем. Орієнтація молекул є причиною появи на протилежних поверхнях діелектрика, які перетинаються силовими лініями поля, зарядів протилежних знаків.

Зауваження. Явище поляризації здебільшого супроводжується діелектричними втратами. Так називають потужність, яка витрачається у непровідній речовині у разі її

переполяризації у змінних електричних полях. Ця потужність розсіюється у вигляді тепла, яке підвищує температуру діелектрика.

Діелектричні втрати є важливою характеристикою високочастотних матеріалів. У радіопристроях діелектричні втрати збільшують активний опір електричних кіл і, як наслідок, згасання коливань у контурах. Великі втрати викликають недопустиме нагрівання діелектрика в полі високої частоти, що може призвести до його теплового руйнування.

Явище поляризації діелектриків визначає таке важливе їх застосування, як заповнення конденсаторів з метою збільшення ємності останніх.

Поляризацію діелектрика характеризують вектором поляризації – електричним дипольним моментом одиниці об'єму діелектрика:

$$\vec{P} = \sum_{i=1}^{m} \frac{\vec{p}_i}{\Delta V}, \qquad (2.9)$$

де $\vec{p}_i$ – дипольний момент i-ї молекули, яка входить в об'єм ΔV; m – кількість молекул у цьому об'ємі; ΔV – такий малий об'єм, у межах якого можна вважати постійним напруженість електричного поля, густину, температуру, але водночас у ньому міститься достатньо велика для усереднення кількість молекул. Для однорідного діелектрика, який перебуває в однорідному електричному полі, справедлива така рівність:

$$\vec{P} = n\vec{p}, \qquad (2.10)$$

де $n = m/\Delta V$ – кількість молекул в одиниці об'єму; $\vec{p}$ – дипольний момент молекули. Оскільки $\vec{p} = \beta\varepsilon_0\vec{E}$, то $\vec{P} = n\beta\varepsilon_0\vec{E}$. Позначимо $n\beta = \chi$, тоді

$$\vec{P} = \chi\varepsilon_0\vec{E}. \qquad (2.11)$$

Коефіцієнт χ називають *діелектричною сприйнятливістю* речовини, або *поляризованістю одиниці об'єму діелектрика*. Коефіцієнт χ – величина безрозмірна ($\chi = 80$ для води, $\chi = 25$ для спиртів). Діелектрична сприйнятливість – це макроскопічний параметр, який характеризує здатність структурних одиниць діелектриків до поляризації, тобто до зміщення зарядів, наприклад в атомах чи молекулах.

Розглянемо однорідний діелектрик, який перебуває у зовнішньому однорідному електричному полі, напруженість якого $\vec{E}_0$. Для простоти розгляду будемо вважати, що діелектрик – це прямокутна пластинка, і лінії вектора $\vec{E}_0$ перпендикулярні до її поверхні (рис. 2.9).

Під дією поля діелектрик поляризується, тобто відбувається зміщення зарядів. Унаслідок цього на правій грані діелектрика утвориться

надлишок позитивного заряду з поверхневою густиною $+\sigma'$, на лівій – від’ємного заряду з поверхневою густиною $-\sigma'$. Ці заряди називають зв’язаними. Зв’язані заряди на гранях пластинки (як дві заряджені площини) створюють електричне поле, напруженість якого $\vec{E}'$ (рис. 2.9). Оскільки вектори $\vec{E}_0$ і $\vec{E}'$ мають протилежні напрямки, то напруженість поля в діелектрику становитиме

$$E = E_0 - E' < E_0. \qquad (2.12)$$

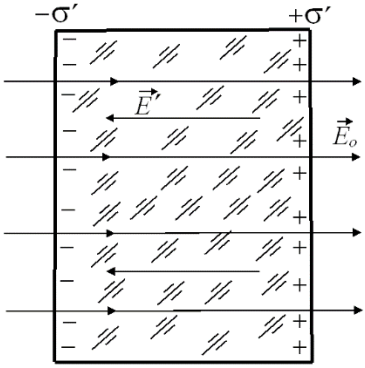


Рис. 2.9. Діелектрик у зовнішньому однорідному електричному полі

Таким чином, поляризація діелектрика призводить до послаблення зовнішнього поля $\vec{E}_0$ всередині пластини.

Визначимо поверхневу густину зв’язаних зарядів σ' у загальному випадку, якщо зовнішнє електричне поле $\vec{E}_0$ утворює кут α з нормаллю до грані діелектрика (рис. 2.10).

Розглянемо циліндр, площі основи якого ΔS перебувають на гранях пластини, а твірна паралельна вектору поляризації $\vec{P}$ й дорівнює якійсь величині l. Повний дипольний момент такого циліндричного діелектрика становить

$$P_{\Delta V} = P \cdot \Delta V,$$

де P – величина вектора поляризації; $\Delta V = \Delta S \cdot l \cdot \cos\alpha$ – об’єм циліндра. І навпаки, вважаючи циліндр за диполь, можемо записати, що

$$P_{\Delta V} = \sigma' \cdot \Delta S \cdot l.$$

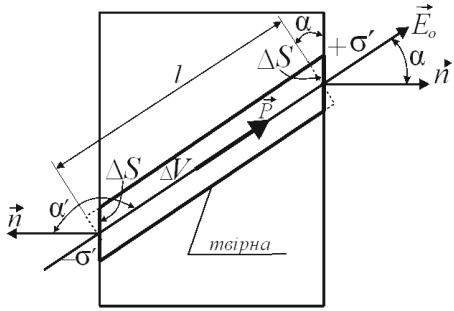


Рис. 2.10. Зовнішнє електричне поле $\vec{E}_0$ утворює кут α з нормаллю до грані діелектрика

Тоді $P \cdot \Delta S \cdot l \cdot \cos\alpha = \sigma' \cdot \Delta S \cdot l$ і

$$\sigma' = P\cos\alpha = P_n = \vec{P} \cdot \vec{n}, \qquad (2.13)$$

де P_n – проєкція вектора поляризації на напрямок нормалі до поверхні діелектрика. Для правої поверхні діелектрика на рис. 2.10 кут α – гострий, $\cos\alpha > 0$, $P_n > 0$ і, відповідно, величина σ' для неї позитивна; для лівої поверхні кут α' – тупий, $\cos\alpha' < 0$, $P_n < 0$ і,

відповідно, величина σ' для неї негативна.

2.4. Електричне поле всередині діелектрика. Діелектрична проникність та її фізичний зміст. Зв'язок між діелектричною сприйнятливістю та діелектричною проникністю

Розглянемо діелектрик, який перебуває у зовнішньому електроста тичному полі з напруженістю $\vec{E}_0$, яке створене вільними зарядами q_i. Поле всередині діелектрика є суперпозицією зовнішнього поля $\vec{E}_0$ і внутрішнього поля $\vec{E}'$, яке створюється зв'язаними зарядами, тобто

$$\vec{E} = \vec{E}_0 + \vec{E}'. \qquad (2.14)$$

Слід уточнити, що під $\vec{E}'$ розуміють деяке усереднене поле зв'язаних зарядів. Властивості вільних і зв'язаних зарядів тотожні, тому для зв'язаних зарядів також справедлива теорема Остроградського–Гауса. Згідно з (1.15) можемо записати, що

$$\oint_S \vec{E} d\vec{s} = \oint_S \vec{E}_0 d\vec{s} + \oint_S \vec{E}' d\vec{s} = \frac{1}{\varepsilon_0}(\sum q_i + q'), \qquad (2.15)$$

де $\sum q_i$ – сума вільних зарядів, яку охоплює замкнута поверхня S; $\sum q'$ – сума зв'язаних зарядів, з якими пов'язана напруженість поля $\vec{E}'$, що також охоплює поверхня S. Оскільки поле $\vec{E}$ всередині діелектрика завжди менше від зовнішнього поля $\vec{E}_0$, то потік поля $\vec{E}'$ через замкнуту поверхню S становить

$$\oint_S \vec{E}' d\vec{s} = \frac{1}{\varepsilon_0}\sum q' < 0.$$

Можна показати, що величина зв'язаних зарядів

$$\sum q' = -\oint_S \vec{P} d\vec{s}. \qquad (2.16)$$

У результаті (2.15) перепишемо як

$$\oint_S (\varepsilon_0 \vec{E} + \vec{P}) d\vec{s} = \sum q_i, \qquad (2.17)$$

оскільки інтегрування ведеться за однією і тією самою замкнутою поверхнею S.

Для характеристики електричного поля в діелектрику введемо допоміжну величину

$$\vec{D} = \varepsilon_0 \vec{E} + \vec{P}, \qquad (2.18)$$

яку будемо називати *вектором електричного зміщення (вектором електричної індукції).*

У результаті рівняння, яке виражає теорему Остроградського–Гауса для електричного поля у середовищі, можна записати у такому вигляді:

$$\Phi_D = \oint_S \vec{D} \cdot d\vec{S} = \sum q_i, \qquad (2.19)$$

де Φ_D – потік електричного зміщення.

Згідно з цією теоремою ***потік електричного зміщення електричного поля через довільну замкнуту поверхню дорівнює алгебраїчній сумі вільних зарядів, які охоплюються цією поверхнею.***

Вводячи допоміжну величину (вектор $\vec{D}$), ми позбулися труднощів, які пов'язані з наявністю в (2.15) невідомої величини $\sum q'$. Для вакууму $\vec{P} = 0$, тому $\vec{D} = \varepsilon_0 \vec{E}$. Згідно з (2.11) вектор поляризації $\vec{P}$ пропорційний до напруженості поля в діелектрику. Отже,

$$\vec{D} = \varepsilon_0 \vec{E} + \chi \varepsilon_0 \vec{E} = \varepsilon \varepsilon_0 \vec{E}, \qquad (2.20)$$

де $\varepsilon = 1 + \chi$. Тоді

$$\vec{D} = \varepsilon \varepsilon_0 \vec{E}. \qquad (2.21)$$

Так само, як і χ, ε – безрозмірна величина. Цю величину називають відносною діелектричною проникністю, або просто діелектричною проникністю і вона є важливою константою діелектрика. Для сили взаємодії двох точкових зарядів, занурених в однорідний безмежний діелектрик, закон Кулона (1.2) записують як

$$F = \frac{k q_1 q_2}{\varepsilon r^2}.$$

Слід зауважити, що ця формула справедлива лише для рідких і газоподібних діелектриків. Можна сказати, що діелектрична проникність ε показує, у скільки разів сила взаємодії двох точкових зарядів у діелектрику менша від цієї самої сили взаємодії зарядів у вакуумі за однакових відстаней між ними.

Далі буде описано, що ε показує, у скільки разів напруженість поля в діелектрику менша від напруженості поля поза ним, якщо силові лінії вектора $\vec{E}$ перпендикулярні до поверхні діелектрика.

Зауважимо, що формула (2.21) справедлива для ізотропних діелектриків. Для анізотропних діелектриків напрямки $\vec{D}$ і $\vec{E}$ не збігаються.

Висновок 2.1. Результуюче поле в діелектрику визначає вектор напруженості $\vec{E}$ (формула (2.14)), і тому воно залежить від властивостей діелектрика. Вектор $\vec{D}$ від середовища не залежить й описує електростатичне поле, що створюється вільними зарядами. ***Зв'язані заряди, що виникають в діелектрику, викликають перерозподіл вільних зарядів.*** Отже, вектор $\vec{D}$ характеризує електростатичне поле, що створюється вільними зарядами (тобто у вакуумі), але за такого їх розподілу у просторі, який є за наявності діелектрика.

Висновок 2.2. Вектор $\vec{D}$ не залежить від властивостей середовища тоді, коли поле створюється зарядженими тілами в однорідному ізотропному середовищі.

Висновок 2.3. Лінії вектора $\vec{E}$ можуть починатися і закінчуватись як на вільних зарядах, так і на зв'язаних зарядах, а лінії вектора $\vec{D}$ – лише на вільних зарядах. Через області поля, де

перебувають зв'язані заряди, лінії вектора $\vec{D}$,проходять так, ніби зв'язаних зарядів немає.

2.5. Електричне поле на межі двох діелектриків

На межі двох середовищ з різними діелектричними властивостями вектори електричного поля змінюються за величиною і напрямком. Умови поведінки векторів $\vec{D}$ і $\vec{E}$ на межі поділу середовищ називають граничними. Ці умови отримують з основних рівнянь поля. Спочатку встановимо граничні умови для нормальної складової вектора $\vec{D}$ (рис. 2.11). Нехай лінія MN – межа поділу двох однорідних ізотропних діелектриків *1* і *2*, проникності яких ε_1 і ε_2 відповідно; $\vec{D}_1$ – вектор зміщення поля у середовищі *1*, а $\vec{D}_2$ – вектор зміщення в середовищі *2*. На межі поділу візьмемо уявну циліндричну поверхню, висота якої h, а площа основи S. Вважатимемо площу основи настільки малою, що в її межах поле можна вважати однорідним; $\vec{n}_1$ і $\vec{n}_2$ – нормалі до площини основи в першому і другому середовищах (нормалі направлені назовні поверхні).

Використаємо теорему Остроградського–Гауса (2.19). Якщо вільних зарядів на межі між діелектриками немає, то $\Phi_D = 0$. Потік через основу S у першому середовищі становить

$$\Phi_{D_{1n}} = \vec{D}_1\vec{n}_1 S = -D_{1n}S,$$

де D_{1n} – проєкція вектора $\vec{D}_1$ на нормаль $\vec{n}_1$. Аналогічно потік через основу S удругому середовищі

$$\Phi_{D_{2n}} = \vec{D}_2\vec{n}_2 S = D_{2n}S,$$

де D_{2n} – проєкція вектора $\vec{D}_2$ на нормаль $\vec{n}_2$. Потік через бічну поверхню можна подати у вигляді $\Phi'_{D_n} = \langle D_n\rangle S_{\text{біч}}$, де $\langle D_n\rangle$ – значення D_n, усереднене по всій бічній поверхні, $S_{\text{біч}}$ – бічна поверхня циліндра. Таким чином, згідно з (2.19) можемо записати

$$\Phi_D = \Phi_{D_{1n}} + \Phi_{D_{2n}} + \Phi'_{D_n} = 0 \text{ або}$$

$$-D_{1n}S + D_{2n}S + \langle D_n\rangle S_{\text{біч}} = 0. \quad (2.22)$$

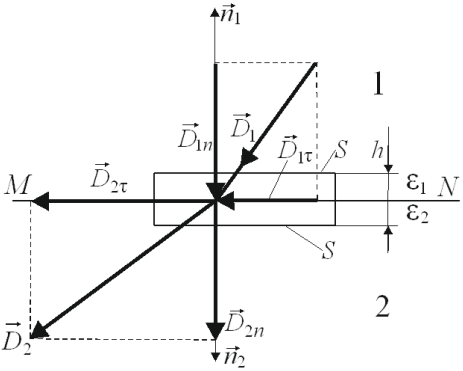


Рис. 2.11. Електричне поле на межі двох діелектриків (вектор $\vec{D}$)

Якщо висота циліндра h прямує до нуля, то $S_{\text{біч}}$ буде прямувати до нуля, і з (2.22) отримаємо

$$D_{1n} = D_{2n}. \quad (2.23)$$

Скориставшись формулою (2.21), можемо записати, що

$$\varepsilon_1\varepsilon_0 E_{1n} = \varepsilon_2\varepsilon_0 E_{2n},$$

звідки отримаємо таке співвідношення:

$$\frac{E_{1n}}{E_{2n}} = \frac{\varepsilon_2}{\varepsilon_1}. \qquad (2.24)$$

Виберемо на межі поділу двох діелектриків MN невеликий прямокутний контур Г, довжина якого a, а ширина b. Контур частково проходить у першому діелектрику, частково – у другому (рис. 2.12). Межа поділу MN (вісь x) проходить через середини сторін контура, ширина яких b.

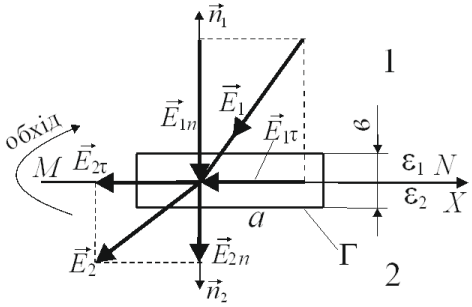


Рис. 2.12. Електричне поле на межі двох діелектриків (вектор $\vec{E}$)

Нехай у діелектриках створені поля: напруженість поля у першому діелектрику дорівнює $\vec{E}_1$, а в другому – $\vec{E}_2$. Кожний із векторів $\vec{E}_1$ і $\vec{E}_2$ можна подати у вигляді суми нормальної і тангенційної складових:

$$\vec{E}_1 = \vec{E}_{1n} + \vec{E}_{1\tau};\ \vec{E}_2 = \vec{E}_{2n} + \vec{E}_{2\tau},$$

де $\vec{E}_{1\tau}$ і $\vec{E}_{2\tau}$ – проєкції векторів $\vec{E}_1$ і $\vec{E}_2$ на орт $\vec{\tau}$, який направлений вздовж лінії перетину площини, що розділяє діелектрики, із площиною, в якій містяться вектори $\vec{E}_1$ і $\vec{E}_2$.

Циркуляція вектора $\vec{E}$ по вибраному замкнутому контуру Г згідно з теоремою про циркуляцію (1.32) має дорівнювати нулю. У разі малих розмірів контура Г і згідно із вказаним на рис. 2.12 обходом циркуляцію вектора $\vec{E}$ можна записати у вигляді

$$\oint_\Gamma \vec{E}\, d\vec{l} = -E_{1\tau}a + E_{2\tau}a + \langle E_b\rangle 2b = 0,$$

де $\langle E_b\rangle$ – середнє значення напруженості поля E_b на перпендикулярних до межі MN ділянках (завдовжки $2b$) контура Г. Якщо ширина контура b прямує до нуля, то отримуємо таку рівність:

$$E_{1\tau} = E_{2\tau}. \qquad (2.25)$$

Замінимо згідно з (2.21) проєкцією вектора $\vec{E}$ проєкціями вектора $\vec{D}$, які розділимо на $\varepsilon\varepsilon_0$. Тоді отримаємо співвідношення

$$\frac{D_{1\tau}}{\varepsilon_1} = \frac{D_{2\tau}}{\varepsilon_2}. \qquad (2.26)$$

Отримані результати означають, що під час переходу крізь межу поділу двох діелектриків нормальна складова вектора $\vec{D}$ (2.23) і тангенційна складова вектора $\vec{E}$ (2.25) змінюються неперервно. Тангенційна складова вектора $\vec{D}$ (2.26) і нормальна складова вектора $\vec{E}$ (2.24) під

час переходу крізь межу поділу зазнають розриву. Співвідношення (2.23), (2.24), (2.25) і (2.26) визначають умови, яким мають задовольняти вектори $\vec{E}$ і $\vec{D}$ на межі двох діелектриків (у тому випадку, якщо на цій межі немає вільних зарядів). Ці співвідношення отримані для електростатичного поля, однак вони справедливі і для полів, які змінюються з часом.

Знайдені умови справедливі і для межі діелектрика з вакуумом. У цьому випадку одну з діелектричних проникностей слід взяти за одиницю.

2.6. Особливі властивості деяких діелектриків

Розглянуті раніше неполярні й полярні діелектрики мають ту особливість, що вони поляризуються тільки під дією зовнішнього електричного поля $\vec{E}_0$, а після припинення його дії знову повертаються в неполяризований стан.

Утім, у природі є велика група діелектриків, переважно у твердому кристалічному стані, які виявляють макроскопічну поляризованість, навіть якщо зовнішнього поля немає. Це спонтанно (лат. spontaneus – самодовільний) поляризовані діелектрики, тобто такі, поляризація яких викликана не зовнішніми, а внутрішніми причинами. До спонтанно поляризованих діелектриків належать ***піроелектрики*** і ***сегнетоелектрики***. Спонтанна поляризація є наслідком зміщення «центру ваги» позитивних і негативних зарядів кристалічної ґратки речовини.

Для піроелектриків спостерігається не сама спонтанна поляризація $\vec{P}$ (вона компенсується полями вільних електричних зарядів, що адсорбуються поверхнею із повітря або компенсується внутрішніми іонами), а її зміна $\Delta\vec{P}$ у разі швидкої зміни температури ΔT (піроелектричний ефект). Густина поверхневих зарядів, які при цьому виникають,

$$\sigma = p \cdot \Delta T,$$

де p – піроелектрична стала, величина якої має порядок 10^{-3} Кл/(м^2·К). Відомий піроелектрик – турмалін (боратосилікатний мінерал змінного складу). В ньому у разі зміни температури на 1 ℃ виникає електричне поле $E \sim 400$ В/см. Крім турмаліну, до піроелектриків належать сульфат літію, цукор, окремі види кераміки та ін. Характерною особливістю піроелектриків є наявність поляризації в усій області температур їхнього існування. Піроелектрики застосовують як індикатори і приймальні випромінювання та пристрої теплового бачення.

Сегнетоелектрики

Екєпериментальні дослідження піроелектриків показали, що їх за поведінкою в зовнішньому електричному полі можна поділити на лінійні й нелінійні діелектрики. Цей поділ визначається характером залежності вектора поляризації $\vec{P}(\vec{E})$. Серед нелінійних твердих діелектриків найбільш поширеними є ***сегнетоелектрики***.

Сегнетоелектрики, кристалічні діелектрики, що мають у певному інтервалі температур, яку називають полярною фазою, спонтанну поляризацією, що суттєво змінюється під дією зовнішніх впливів (електричних полів, змін температури тощо).

Сегнетоелектричні властивості були вперше виявлені у кристалів ***сегнетової солі*** $KNaC_4H_4O_6{\times}4H_2O$ у 1920 р., потім у дигідрофосфата калія ($(KDP)KH_2PO_4$). Відомо декілька сотень сегнетоелектриків. Особливість їх у тому, що в них вектор поляризації $\vec{P}$ порівняно легко змінюється під дією електричних полів, пружних деформацій, змін температури та ін. Сегнетоелектрики неоднорідно поляризовані. Вони складаються із ***доменів*** – областей із різними напрямками поляризації (рис. 2.13). У результаті сумарного електричного дипольного моменту зразка фактично не спостерігається.

Рівноважна доменна структура сегнетоелектрика відповідає мінімуму вільної енергії кристала. Доменну структура реального кристала визначають природою і характером розподілу його дефектів, а також історією зразка. Кількість різних доменів, взаємна орієнтація їхньої спонтанної поляризації залежить від ***симетрії кристала***. ***Сегнетоелектриками є кристали, в яких немає центра симетрії.*** Властивості сегнетоелектриків максимально виражені в певному напрямку у кристалі, який називають ***полярною віссю***. Інакше кажучи, сегнетоелектрики мають різку анізотропію діелектричних властивостей. ***Анізотропія*** означає, що фізичні властивості тіла різні в різних напрямках. Модифікацію кристалів, в яких сегнетоелектрики спонтанно поляризовані, називають ***полярною фазою*** (рис. 2.15).

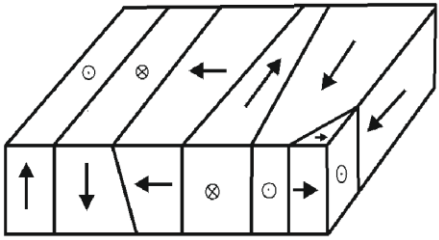


Рис. 2.13. Схематичне зображення доменів у сегнетоелектрику ($BaTiO_3$); стрілки і познаки ⊙ та ⊗ вказують на напрямок вектора поляризації $\vec{P}$ домена

З підвищенням температури вище деякого значення T_K, характерного для кожного сегнетоелектрика, відбувається фазовий перехід, у результаті якого сегнетоелектричні властивості зникають і кристал переходить у звичайний неполярний діелектрик. Такий стан кристала називають параелектричним, а фазу неполярною. Температуру T_K називають температурою Кюрі. У табл. 2.2 наведено деякі характеристики частини сегнетоелектриків.

Таблиця 2.2

Характеристики деяких сегнетоелектриків

Речовина	Хімічна формула	Темпер. фазового переходу T_K, °С	Максимальна спонтанна поляризац. P, мкКл/см^2
Титанат барію	$BaTiO_3$	120	25
Сегнетова сіль	$KNaC_4H_4O_6 \cdot 4H_2O$	−18... −24	0,25
Дигідрофосфат калію	KH_2PO_4	−150	5,1
Ніобат літію	$LiNbO_3$	1210	50
Титанат вісмуту	$Bi_4Ti_3O_{12}$	675	-

Фактично важливим сегнетоелектриком є титанат барію, який використовують для виготовлення конденсаторів з ємністю, яка змінюється залежно від температури і величини задіяної напруги.

Залежність діелектричної проникності титанату барію від напруженості електричного поля і температури показано на рис. 2.14 і 2.15. Титанат барію має високу механічну міцність, велике значення ε, найбільш просту кристалічну структуру порівняно зі всіма відомими сегнетоелектриками, три точки Кюрі (−90 °С, 5 °С і 120 °С).

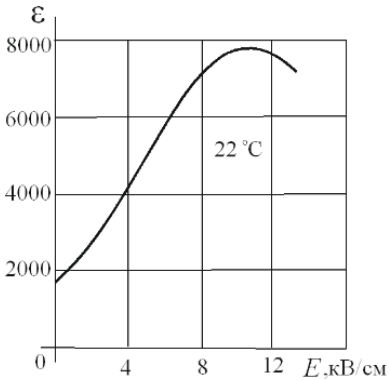


Рис. 2.14. Залежність діелектричної проникності титанату барію від напруженості електричного поля

Пояснимо залежності $\varepsilon(E)$ і $\varepsilon(T)$ для сегнетоелектрика. Під час додавання сегнетоелектрика в електричне поле він поляризується. Електрична індукція сегнетоелектрика утворюється із трьох складових:

$$\vec{D} = \varepsilon_0 \vec{E} + \vec{P} + \vec{P}_s, \qquad (2.27)$$

де $\vec{E}$ – середнє макроскопічне поле в сегнетоелектрику; $\vec{P}$ – середній електричний момент одиниці об'єму, обумовлений електронною та іонною

поляризацією; $\vec{P}_s$ – електричний момент спонтанної поляризації. Вектори $\vec{E}$ і $\vec{P}$ мають однаковий напрямок, а вектор $\vec{P}_s$ у загальному випадку має інший напрямок. Повна діелектрична проникність сегнетоелектрика, яка обумовлена поляризацією, що викликана зовнішнім полем і спонтанною поляризацією, дорівнює відношенню $\vec{D}/(\varepsilon_0\vec{E})$. Із (2.27) отримаємо

$$\varepsilon = 1 + \frac{P}{\varepsilon_0 E} + \frac{P_{sE}}{\varepsilon_0 E}, \qquad (2.28)$$

де P_{sE} – проєкція вектора $\vec{P}_s$ на напрямок поля $\vec{E}$. Зі збільшенням поля $\vec{E}$ повний електричний момент одиниці об'єму (вектор поляризації) може збільшуватись із двох причин: по-перше, внаслідок зростання електричного моменту поляризації зміщення $\vec{P}$, по-друге, внаслідок зростання кількості областей спонтанної поляризації (доменів), напрямки електричних моментів яких збігаються з напрямком зовнішнього поля або утворюють з ним гострі кути. Так можна пояснити залежність $\varepsilon(E)$.

Спонтанна поляризація сегнетоелектрика залежить від зміни температури, оскільки при цьому змінюється інтенсивність теплового руху, який суттєво зменшує її, особливо на межах полярної області. В неполярній області сегнетоелектрик поводиться як звичайний діелектрик, в якому поляризація пропорційна напруженості електричного поля (доменів у цій області температур немає). Утім, у всіх випадках діелектрична проникність змінюється з температурою. Так, вище точки Кюрі ($T > T_K$) ця зміна відбувається за законом Кюрі–Вейса:

$$\varepsilon = \frac{C}{T - T_K},$$

де C і T_K – сталі, з яких T_K називають температурою (точкою) Кюрі. В області точки Кюрі поляризованість сегнетоелектрика аномально велика (рис. 2.15). Це пов'язано з тим, що за температури Кюрі відбувається перебудова структури кристалічних ґраток сегнетоелектрика. Титанат барію в неполярній фазі за температури вище 120 °С має кубічну структуру, нижче 120 °С – тетрагональну.

Отже, властивості сегнетоелектриків пояснюються існуванням областей спонтанної поляризації (доменів). Утворення доменів у титанаті барію відбувається таким чином. У неполярній фазі (рис. 2.15) (вище точки Кюрі) титанат барію має кубічну структуру (рис. 2.16, *а*): у вершинах куба – іони Ba^{2+}, у центрах граней – іони O^{2-}, у центрі куба – іон Ti^{4+}. Через наявність центра симетрії дипольний момент елементарної комірки дорівнює нулю, і титанат барію в неполярній фазі є

звичайним діелектриком з великою діелектричною проникністю.

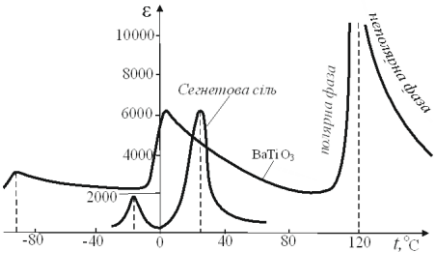
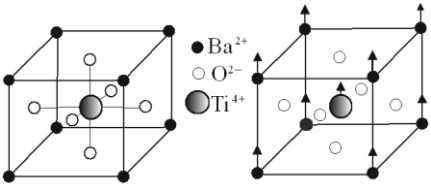


Рис. 2.15. Залежність діелектричної проникності сегнетової солі та титанату барію від температури



а *б*

Рис. 2.16. Елементарна комірка титаната барію. а – неполярна фаза, б – полярна фаза

За температури Кюрі відбувається фазовий перехід, який призводить до деформації кубічної структури (рис. 2.16, *б*): іони Ba^{2+} і Ti^{4+} зміщуються відносно іонів O^{2-} і таким чином створюють дипольний момент елементарної комірки, структура якої в полярній фазі тетрагональна. Близьке розміщення окремих диполів один відносно одного обумовлює їх сильну електричну взаємодію між собою. Унаслідок цього ціла група диполів орієнтується в одному напрямку (така орієнтація диполів відповідає мінімуму енергії), тобто утворюються області (домени) з дипольним моментом, який дорівнює сумі дипольних моментів молекул, що входять у цю область.

Накладання зовнішнього електричного поля створює переважну орієнтацію векторів поляризації доменів у напрямку поля, у зв'язку з чим поляризованість та електрична проникність сегнетоелектрика стають досить великими (рис. 2.15).

Підвищення температури в полярній області полегшує орієнтацію векторів поляризації окремих доменів у напрямку поля. За температури Кюрі крива $\varepsilon(T)$ сегнетоелектрика (рис. 2.15) має різко виражений максимум. У разі підвищення температури вище точки Кюрі тетрагональна решітка титанату барію перебудовується у правильну кубічну, доменні групи розпадаються і величина ε різко зменшується.

З аналізу залежності $\varepsilon(T)$ на рис. 2.15 випливає, що допускати нагрівання конденсаторів, які містять діелектрик титанату барію, до 120 °С не рекомендовано, оскільки значно змінюється ε, а отже, будуть значно змінюватися характеристики конденсаторів.

Зауважимо, що сегнетоелектрики використовують для виготовлення

малогабаритних конденсаторів ***великої ємності*** та як нелінійні елементи схем (помножувачів частоти, діелектричних підсилювачів і под.).

Для сегнетоелектриків характерне явище *діелектричного гістерезису* (від грец. hysteresis – відставання, запізнення), яке полягає у відставанні змін вектора поляризації $\vec{P}$ від змін вектора напруженості $\vec{E}$ електричного поля (рис. 2.17).

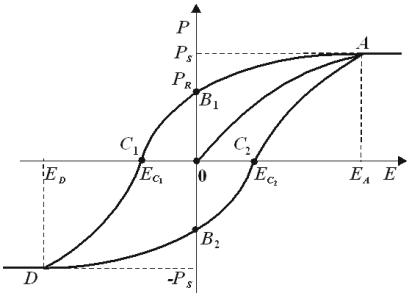


Рис. 2.17. Петля гістерезису для сегнетоелектрика

Якщо неполяризований сегнетоелектрик помістити в електричне поле і збільшувати напруженість поля E, то викликана цим полем поляризація зростає вздовж кривої OA, яку називають ***основною кривою*** поляризації сегнетоелектрика; E_A – напруженість поля, за якої досягається величина спонтанної поляризації P_S. У разі зменшення поля E поляризація зменшується не вздовж кривої OA, а вздовж кривої AB_1, причому в точці B_1 зовнішнього поля немає, $E = 0$, але сегнетоелектрик поляризований. У разі зміни напрямку зовнішнього поля вектор поляризації зменшується до нуля в точці C_1, а потім, змінивши свій напрямок, зростає за величиною вздовж кривої C_1D. Значення напруженості поля E_{C_1} у точці C_1 називають ***коерцитивною силою***. Наступна зміна зовнішнього поля від E_D до E_A супроводжується зміною вектора поляризації вздовж кривої DB_2C_2A. У точках B_1 і B_2 відбувається ***залишкова поляризація*** P_R. Величину напруженості E_{C_2} у точці C_2 також називають коерцитивною силою. Звужуючи інтервал зміни E, можна отримати нескінченну множину петель гістерезису (рис. 2.17), вкладених одна в одну. Таким чином, вектори поляризації $\vec{P}$ і електричного зміщення $\vec{D}$ є багатозначними функціями E. У зв'язку з цим можна говорити про діелектричну сприйнятливість і діелектричну проникність сегнетоелектрика тільки для основної кривої OA (рис. 2.17).

У разі періодичної зміни поляризації сегнетоелектрика на подолання тертя під час повертання електричних моментів витрачається електрична енергія, яка використовується на нагрівання сегнетоелектрика. Площа петлі гістерезису пропорційна електричній енергії, що перетворюється у

внутрішню енергію одиниці об’єму сегнетоелектрика за один цикл.

П’єзоелектрики

Явище виникнення на гранях кристалів зв’язаних електричних зарядів протилежних знаків під час їхньої механічної деформації називають прямим ***п’єзоелектричним ефектом*** (грец. piezo – тиснути).

Це явище відкрили й дослідили у 1880 р. французькі фізики Жан і П’єр Кюрі на кристалі кварцу (SiO_2). Дослідження показали, що п’єзоелектричні властивості мають майже 1500 різних кристалічних речовин, зокрема турмалін, сегнетова сіль, сульфат літію ($Li_2SO_4{\cdot}H_2O$), титанат барію ТВ-1, титанат барію кальцію ТБК-3, цинкова обманка, хлорат натрію, тростинний цукор, напівпровідники CdS, InSb, GaAs і багато інших кристалів.

П’єзоелектричний ефект притаманний лише іонним кристалам, які не мають центра симетрії. Якщо кристал має центр симетрії, то під дією однорідної механічної деформації центр симетрії зберігається, «центри ваги» позитивних і негативних іонів кристала збігаються, а отже, в них не може виникати п’єзополяризації.

Знак зв’язаних зарядів на гранях кристала без центра симетрії під дією деформацій однозначно визначається напрямком і характером деформації. Якщо під час стискання на певній грані виникає позитивний заряд, то під час розтягу ця грань заряджається негативно. Після припинення деформації кристалічний діелектрик повертається у неполярний стан.

Механізм п’єзоефекту пояснимо на прикладі кристала кварцу, який широко застосовують завдяки його механічним і електричним властивостям. Кристалічні ґратки кварцу складається з іонів кремнію Si^+ і негативних іонів кисню O^-, має три осі симетрії (на рис. 2.18, *а* їх позначено X_1, X_2, X_3), які є полярними осями кристала.

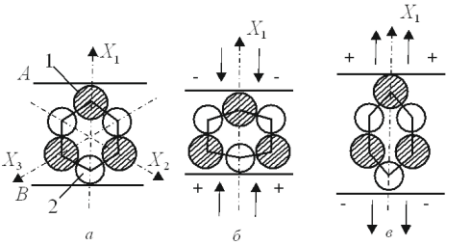


Рис. 2.18. Структурна схема кварцу. Заштриховані кружечки відповідають іонам Si^+, світлі – парі іонів O^-; *а* – недеформований стан; *б* – стиснення вздовж осі X_1; *в* – розтяг вздовж осі X_1. Стрілками показано напрямок деформації

На рис. 2.18, *б* і *в* показано проєкції іонів Si^+ і O^- на площину, яка перпендикулярна осі третього порядку X_3. Під час стиснення вздовж осі X_1 позитивні іони 1(Si^+) і негативні іони

2(O^-) зміщуються у глибину комірки кристала (рис. 2.18, *б*), у результати чого на площинах *A* і *B* з’являються заряди. Під час розтягу (рис. 2.18, *в*) на площинах *A* і *B* виникають заряди протилежних знаків.

Максимальний п’єзоелектричний ефект спостерігається тоді, коли деформацію здійснюють у напрямку однієї з полярних осей кристала, тому кварцеві пластинки вирізають так, щоб пара площин, що утворюються під час розрізання, була перпендикулярна до однієї з полярних осей.

П’єзоелектричний ефект є оборотним. Накладання різниці потенціалів до граней кристала (створення всередині напруженості електричного поля) призводить до механічної деформації (розтягу або стискуванню залежно від полярності напруги). Якщо подати змінну електричну напругу до граней кристала, тобто створити змінну поляризацію кристала, то пластинка буде поперечно розтягуватись і стискуватись вздовж, наприклад, осі X_1 (рис. 2.18, *б, в*). Таким чином у пластинці збуджуються механічні коливання. Ці коливання максимально інтенсивні, якщо частота змінної прикладеної до пластини напруги збігається із власною частотою коливань пластинки.

Таким чином, ***явище деформації п’єзоелектричного*** кристала під дією прикладеної до його граней різниці потенціалів електричного поля називають ***оберненим п’єзоелектричним ефектом***.

Використовують п’єзоелектричні матеріали для виготовлення п’єзодатчиків, призначених перетворювати механічні коливання в електричні, акустичних генераторів для перетворення електричних змін у звукові коливання. Виготовляють адаптери для відтворення звуку, різні прилади для вимірювання частоти й амплітуди вібрації двигунів, турбін, літаків тощо; п’єзоелектричні сейсмографи для реєстрації коливань земної кори, прилади для вимірювання прискорень – акселерометри; п’єзоелек-тричні манометри (датчики тиску), стабілізатори частоти радіостанцій та інше.

Електрети

Електрети – діелектрики, які довгий час зберігають поляризований стан після припинення зовнішніх дій, які викликали поляризацію, і створюють електричне поле в оточуючому просторі (електричний аналог постійних магнітів).

Якщо речовину, молекули якої мають дипольний момент, розплавити і помістити в сильне електричне поле, то її молекули частково вишикуються по

полю. У разі охолодження розплаву в електричному полі з наступним виключенням поля у затверділій речовині поворот молекул утруднений і вони довгий час зберігають переважну орієнтацію (від декількох днів до багатьох років). Перший електрет таким способом виготовив у 1922 р. японський фізик М. Йогучи.

Електрети отримують з аморфних воску і смол (бджолиний віск, парафін і т. д.), полімерів (полі метилметакрилат, полівінілхлорид та ін.), неорганічних полікристалічних діелектриків (титанати лужноземельних металів, ультрафарфорів та інших керамічних діелектриків), монокристалічних неорганічних діелектриків (лужногалоїдні монокристали, наприклад, LiF, корунд, рутил), стекла та ін.

Стабільні електрети отримують нагріванням з наступним охолодженням у сильному електричному полі (*термоелектрети*), освітленням у сильному електричному полі (*фотоелектрети*), опроміненням радіоактивним випромінюванням (*радіоелектрети*), поляризацією в сильному електричному полі без нагрівання (*електроелектрети*) або в магнітному полі (*магнітоелектрети*), під час застигання органічних розчинів в електричному полі (*кріоелектрети*), механічною деформацією полімерів (*механоелектрети*), тертям (*трибоелектрети*), дією поля коронного розряду (*короноелектрети*). Усі електрети мають стабільний поверхневий заряд.

Електрети використовують як джерела постійного електричного поля (електретні мікрофони й телефони, вібродатчики, генератори слабих змінних сигналів, електрометри, електростатичні вольтметри та ін.), а також як чутливі датчики у пристроях дозиметрії, електричної пам'яті; для виготовлення барометрів, гігрометрів і газових фільтрів, п'єзодатчиків та ін. Фотоелектрети застосовують в електрофотографії.

Контрольні запитання

1. Дайте визначення вектора поляризації та поясніть його фізичний зміст.

2. Яким чином можна визначити поляризаційні (зв'язані) заряди за допомогою вектора поляризації?

3. Дайте визначення вектора електричного зміщення (індукції) та поясніть його фізичний зміст.

4. Сформулюйте теорему Остроградського–Гауса для електричного поля у діелектриках.

5. Який зв'язок між векторами електричного зміщення і вектором напруженості електричного поля?

6. У чому особливість електричного поля на межі двох діелектриків?

7. Які основні властивості сегнетоелектриків? Яке практичне застосування сегнетоелектриків?

8. Які речовини є п'єзоелектриками та електретами? Яке їх практичне застосування?

Розділ 3. Провідники в електричному полі

3.1. Розподіл зарядів, напруженість і потенціал електричного поля у провіднику

Характерною особливістю провідників є наявність у них вільних носіїв заряду (див. підрозд. 2.1). У металах це електрони провідності («вільні» електрони). Концентрацією «вільних» електронів, наприклад міді, є величина $n = 8{,}5 \cdot 10^{28}$ м$^{-3}$. Ці електрони можуть рухатись у металах вільно, але не можуть самостійно покидати його поверхню.

Якщо провідник помістити у зовнішнє електростатичне поле $\vec{E}_{\text{зовн}}$, то на кожний «вільний» електрон буде діяти сила $\vec{F}_1 = e\vec{E}_{\text{зовн}}$ (рис. 3.1). Під дією сили $\vec{F}_1$ відбувається переміщення «вільних» електронів, і внаслідок цього електричні заряди перерозподіляються – на одній стороні провідника буде надлишок «вільних» електронів, які заряджають її негативно, на іншій виникає їх нестача, і ця сторона заряджається позитивно.

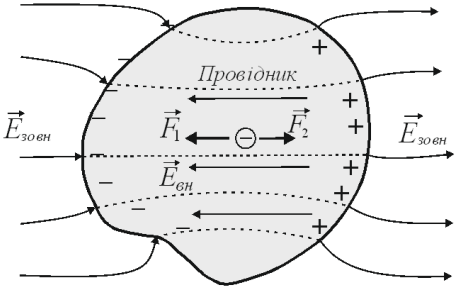


Рис. 3.1. Провідник в електричному полі

Явище перерозподілу вільних носіїв заряду у провіднику під дією зовнішнього електричного поля, внаслідок чого виникає електризація, називають електростатичною індукцією, або електризацією через вплив.

Індуктивні заряди створюють всередині провідника внутрішнє електростатичне поле $\vec{E}_{вн}$, яке діє на «вільні» електрони із силою $\vec{F}_2 = e\vec{E}_{вн}$. Переміщення «вільних» зарядів у провіднику припиняється, коли $\vec{F}_1 + \vec{F}_2 = 0$. Тоді $e(\vec{E}_{зовн} + \vec{E}_{вн}) = 0$. Оскільки $e \neq 0$, то $\vec{E}_{зовн} + \vec{E}_{вн} = 0$ і $\vec{E}_{вн} = -\vec{E}_{зовн}$. Внутрішнє наведене

електричне поле $\vec{E}_{вн}$ буде компенсувати зовнішнє $\vec{E}_{зовн}$, і рух «вільних» зарядів припиниться. Якщо б всередині провідника залишилось хоч яке-небудь електричне поле, воно змусило б рухатись ще якісь електрони.

Таким чином, ***результуюча напруженість електричного поля всередині провідника дорівнює нулю***.

Тепер розглянемо середину провідника. Оскільки всередині провідника напруженість електричного поля дорівнює нулю, то це значить, що і градієнт потенціалу φ дорівнює нулю: $\vec{E} = -\text{grad}\varphi$. Це означає, що φ від точки до точки не змінюється ($\varphi = \text{const}$). Таким чином, *будь-який провідник – це еквіпотенціальна область, і його поверхня також еквіпотенціальна*.

Якщо у провідному матеріалі електричне поле скрізь дорівнює нулю, то і дивергенція $\vec{E}$ також дорівнює нулю ($\text{div}\vec{E} = \rho/\varepsilon_0$, формула (1.38)), тому густина заряду у внутрішній частині провідника дорівнює нулю ($\rho = 0$).

Утім, якщо всередині провідника не може бути надлишкових зарядів, то як тоді він може бути зарядженим? Що маємо на увазі, коли говоримо, що провідник «заряджений»? Де ці заряди перебувають? Мабуть, вони можуть міститися на поверхні провідника, де є сили, які утримують їх. Тож вони у провідниках не зовсім «вільні».

Слід сказати, що спеціально незаряджений провідник є електронейтральним. Сумарний заряд усіх «вільних» електронів у 1 м^3 міді $Q = 8{,}5\cdot10^{28}\cdot1{,}6\cdot10^{-19}$ Кл $\cong 1{,}36\cdot10^{10}$ Кл. Цей негативний заряд електронів компенсується позитивним зарядом іонів міді. Якщо провіднику надати певну кількість електронів, то він зарядиться негативно, а якщо забрати в нього частину електронів, то він зарядиться позитивно. Пробійна напруженість електричного поля для повітря $E_i \sim 10^7$ В/м. Це означає, що провідну кулю (сферу), радіус якої $r = 1$ м, можна зарядити зарядом $q \cong 4\pi\varepsilon_0 r^2 E_i \cong$ $\cong 10^{-3}$ Кл. Цей заряд $q << Q$. Утім, величина заряду q є помітною тому, що незаряджений провідник загалом є електронейтральним. Цей надмірний заряд у будь-якому провіднику перебуває в тонкому шарі біля поверхні, товщина якого зазвичай становить один-два атоми. Для наших нинішніх цілей достатньо правильно буде говорити, що будь-який заряд, який потрапив на (або в) провідник, збирається на його поверхні. Це зрозуміло ще і з того, що однойменні заряди відштовхуються, і вони будуть віддалятися один від одного на

максимальну відстань у провіднику, тобто на поверхню.

Знайдемо зв'язок між напруженістю поля E біля поверхні зарядженого провідника, де поверхнева густина заряду σ (рис. 3.2).

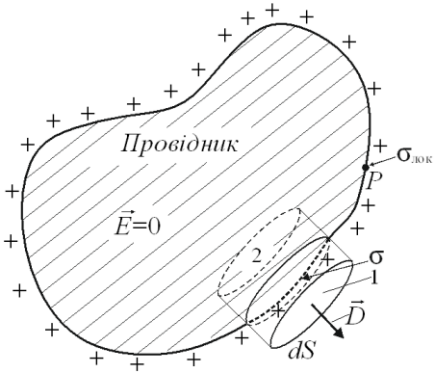


Рис. 3.2. Електричне поле зарядженого провідника

Для цього розглянемо замкнуту поверхню у вигляді циліндра з основою dS, вісь якого орієнтована вздовж вектора $\vec{D}$ (рис. 3.2). Оскільки поверхня провідника еквіпотенціальна, то вектор $\vec{D}$, як і $\vec{E}$, перпендикулярні до поверхні провідника. Основа циліндра *1* перебуває зовні провідника, а основа *2* – всередині нього.

Використаємо теорему Остроградського–Гауса (2.19). Оскільки поля всередині провідника немає, то потік вектора $\vec{D}$ через замкнуту циліндричну поверхню можна визначити лише потоком через зовнішню основу циліндра *1*:

$$d\Phi_D = D\cdot dS = q = \sigma dS.$$

Звідси

$$D = \sigma = \varepsilon\varepsilon_0 E.$$

Отже, напруженість електричного поля біля поверхні зовні провідника довільної форми становить

$$E = \sigma/(\varepsilon\varepsilon_0), \qquad (3.1)$$

де ε – відносна діелектрична провідність середовища, в якому перебуває провідник.

Виникає питання, чому шар зарядів на провіднику створює не таке поле, як шар зарядів сам по собі? Інакше кажучи, чому у формулі (3.1) (для повітря $\varepsilon = 1$) величина напруженості вдвічі більша, ніж у формулі (1.16). Справа у тому, щоб у провіднику вектор $\vec{E}$ дорівнював нулю, в ньому обов'язково мають бути якісь заряди. Безпосередньо близько біля точки P на поверхні (рис. 3.2) заряди справді створюють поле $E_{лок} = \sigma_{лок}/(2\varepsilon_0)$ як всередині, так і зовні провідника. Утім, усі інші заряди провідника створюють у точці P додаткове поле, яке за величиною дорівнює $E_{лок}$. Напруженість сумарного внутрішнього поля дорівнює нулю, а зовнішнього – подвоюється: $2E_{лок} = \sigma_{лок}/\varepsilon_0$, (випадок, коли зовнішнє середовище вакуум (повітря)).

Розподіл зарядів по поверхні провідника

Поверхня зарядженого провідника під час рівноваги є еквіпотенціальною, напруженість поля E всередині дорівнює нулю, однак поверхнева густина заряду залежить від кривизни поверхні. Доведемо це на прикладі двох заряджених куль, радіуси яких R_1 і R_2, що з'єднані між собою тонким провідником. Після встановлення рівноваги заряди на кулях розподіляються так, що потенціали куль будуть однаковими (дві кулі з'єднані провідником – це один провідник), тобто $\varphi_1 = \varphi_2$. Оскільки згідно з формулою (1.19)

$$\varphi_1 = q_1/(4\pi\varepsilon_0 R_1);$$
$$\varphi_2 = q_2/(4\pi\varepsilon_0 R_2),$$

а

$$q_1 = 4\pi\sigma_1 R_1^2 \text{ і } q_2 = 4\pi\sigma_2 R_2^2,$$

то

$$\frac{\sigma_1}{\sigma_2} = \frac{R_2}{R_1}. \qquad (3.2)$$

Отже, поверхнева густина зарядів на кулях обернено пропорційна їхнім радіусам кривизни, тобто де кривизна поверхні (кривизна кола, радіус якого R, скрізь дорівнює $1/R$) зарядженого провідника більша, там і густина заряду більша. Оскільки біля поверхні зарядженого провідника $E = \sigma/(\varepsilon\varepsilon_0)$, то у місцях з великою кривизною (вістря) σ і, відповідно, E набувають великих значень.

На рис. 3.3 зображено лінії напруженості (штрихові лінії) та еквіпотенціальних поверхонь (суцільні лінії) поля позитивно зарядженого провідника циліндричної форми з конічним виступом на одному кінці та конічною западиною на іншому.

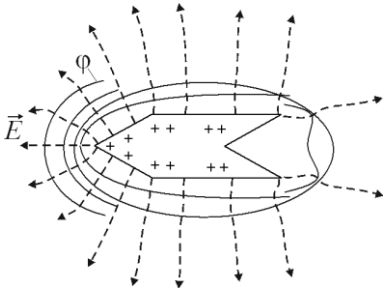


Рис. 3.3. лінії напруженості (штрихові лінії) та еквіпотенціальні поверхні (суцільні лінії) поля позитивно зарядженого провідника циліндричної форми

З рис. 3.3 видно, що поблизу вістря і виступів еквіпотенціальні поверхні найсильніше зближені, відповідно і поверхнева густина вільних зарядів на цих ділянках провідника більша, ніж на інших поверхнях тіла. В області конічної западини напруженість поля мінімальна.

Значна величина напруженості поля поблизу гострого виступу на зарядженому провіднику приводить до явища «електричного вітру». В атмосферному повітрі завжди є невелика

кількість позитивних іонів і вільних електронів, що виникають під впливом космічних променів, випромінювання радіоактивних речовин тощо. В сильному електричному полі поблизу вістря електрони рухаються з великою швидкістю і, стикаючись з молекулами повітря, іонізують їх. Отже, виникає все більша кількість позитивних рухомих зарядів, які утворюють «електричний вітер». Існування «електричного вітру» супроводжується розряджанням наелектризованого тіла, оскільки до нього рухаються під дією електричного поля заряди протилежного знака відносно зарядів наелектризованого тіла. Це приводить до стікання електричних зарядів з металевого вістря, яке використовують для побудови блискавковідводів. У високовольтних лініях електропередачі стікання зарядів із загострених частин проводів призводить до значних втрат енергії.

Поле всередині порожнини провідника

Всередині провідника поля немає, а чи є воно в порожнині? Покажемо, що якщо порожнина порожня, то поля у ній бути не може, яка б не була форма провідника або порожнини (рис. 3.4). Розглянемо замкнуту поверхню S на рис. 3.4, яка оточує порожнину, але залишається скрізь у речовині провідника. На поверхні S у кожній її точці напруженість поля дорівнює нулю, тому потік через S дорівнює нулю, і ***сумарний*** заряд всередині S також має дорівнювати нулю.

У загальному випадку можна сказати, що на внутрішній поверхні провідника S_1 є однакова кількість позитивного і негативного зарядів, що розмістились у різних місцях цієї поверхні (рис. 3.4). Утім, однакові за величиною і протилежні за знаком заряди мають рухатись назустріч один одному і «знищити» один одного.

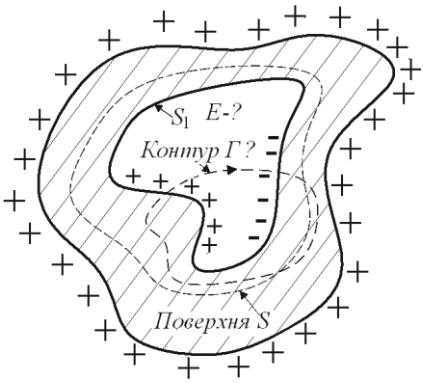


Рис. 3.4. Поле всередині порожнини провідника

Уцьому можна переконатися, застосувавши закон про циркуляцію вектора $\vec{E}$ (формула (1.32)). Нехай на деяких частинах внутрішньої поверхні з'явилися заряди. Відомо, що десь ще має бути така сама кількість протилежних за знаком зарядів. Утім, будь-які лінії вектора $\vec{E}$ починаються на позитивних

зарядах і закінчуються на негативних (розглянемо випадок, коли в порожнині вільних зарядів немає). Уявімо контур Г, який перетинає порожнину вздовж силової лінії від якогось позитивного заряду до якогось негативного і повертається до вихідної точки по тілу провідника (рис. 3.4). Інтеграл $\int \vec{E} d\vec{l}$ вздовж такої силової лінії у порожнині в межах від позитивного до негативного заряду був би відмінній від нуля, а такий самий інтеграл по шляху контуру Г у металі дорівнював би нулю, оскільки у провіднику $\vec{E} = 0$. У результаті отримали б, що $\oint_\Gamma \vec{E} d\vec{l} \neq 0$. Водночас згідно з теоремою про циркуляцію вектора $\vec{E}$ інтеграл за будь-яким замкнутим контуром Г в електростатичному полі $\oint_\Gamma \vec{E} d\vec{l}$ завжди дорівнює нулю. Значить, всередині порожнини не може бути жодних полів, що рівнозначно тому, що не може бути жодних зарядів на внутрішній поверхні провідника S_1.

Отже, якщо порожнина повністю оточена провідником, то жодний статичний розподіл зарядів зовні ніколи не створить поля всередині. Це пояснює принцип «захисту» електронного обладнання, яке поміщають у металеву коробку. В електростатиці (але не у змінних полях) поля по обидва боки суцільної провідної оболонки повністю не залежать одне від одного.

Властивість зарядів розміщуватись на зовнішній поверхні провідника використовують для створення *електростатичних генераторів*, які призначені для нагромадження великих зарядів і досягнення різниці потенціалів у декілька мільйонів вольтів. Винахідником такого генератора є американський фізик Ван де Грааф. Використовують ці генератори у високовольтних прискорювачах заряджених частинок.

Теорема єдиності в електростатиці

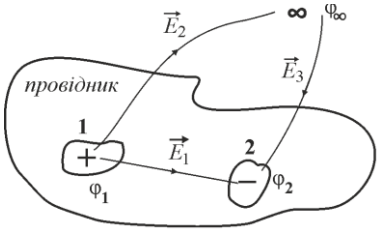
Надаймо провіднику деякий позитивний заряд q. Виникає запитання: *чи єдиним чином весь цей заряд може розподілитися по поверхні провідника?* Відповідь на це запитання і є змістом теореми єдиності в електростатиці.

Доведення теореми. Припустимо, що заряд q може розподілятись по провіднику двома способами, тобто є два різних розподіли заряду q по поверхні провідника. Позначимо ці розподіли D_1 і D_2. Якщо цьому самому провіднику (незарядженому) надати негативний заряд $-q$, то серед його рівноважних розподілів по поверхні провідника будуть розподіли D_1' і D_2', які відрізнятимуться від попереднього розподілу D_1 та D_2 тільки тим, що знаки відповідних зарядів, у цьому місці провідника, мають

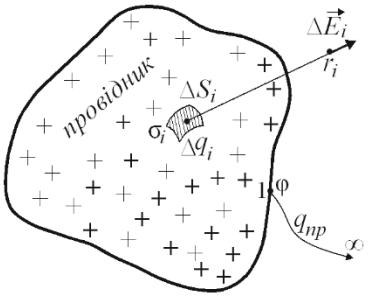
бути змінені на протилежні. Справді, силами взаємодій між зарядами в конфігурації D_1 (D_2) є силами взаємного відштовхування зарядів, які перебувають на поверхні. Від зміни знаків цих зарядів сили взаємодії залишаються незмінними, тому кожний елементарний поверхневий заряд буде у стані рівноваги. Утім, електричне поле в кожній точці простору зовні провідника змінить напрямок на протилежний. Таким чином, конфігурація D_1' (D_2') також буде рівноважною.

Тепер припустимо, що наданий провіднику заряд q розподілився по конфігурації D_1. «Закріпимо» цей розподіл і надамо провіднику заряд $-q$ так, щоб він розподілився по конфігурації D_2'. Тоді сумарний заряд провідника дорівнює нулю згідно з принципом суперпозиції, система буде перебувати у рівновазі й жодного перетікання зарядів по поверхні провідника не буде. Таким чином, сумарний заряд провідника дорівнює нулю, але на поверхні провідника з'являться області, які будуть різнойменно заряджені – одна позитивно *1*, а інша негативно *2* (рис. 3.5, *а*). Розглянемо силову лінію $\vec{E}$ поля, яка виходить з позитивної області *1* (рис. 3,5, *а*).

1. Оскільки провідник вілокремлений, то ця силова лінія або закінчується на ньому ($\vec{E}_1$), або прямує до нескінченності ($\vec{E}_2$). Якщо силова лінія закінчується на провіднику, то між точками її початку *1* і кінця *2* має бути різниця потенціалів $\varphi_1 - \varphi_2$. Утім, цього бути не може. Поверхня провідника у стані рівноваги еквіпотенціальна.



а



б

Рис. 3.5.Розподіл заряда у провіднику

2. Якщо ж силова лінія $\vec{E}_3$ (рис. 3.5, *а*) приходить до негативно зарядженої області *2* із нескінченності, а на нескінченності є який-небудь фіксований потенціал φ_∞, то він вищий за потенціал негативно зарядженої області, тобто $\varphi_\infty > \varphi_2$. Якщо силова лінія $\vec{E}_2$ з позитивно зарядженої області *2* закінчується на нескінченності, то її

потенціал φ_1 вищий за потенціал φ_∞ на нескінченності, який є вищий від потенціалу φ_2 негативно зарядженої області *2*. Тоді це означає, що між позитивною *1* та негативною *2* областями провідника і в цьому випадку має бути різниця потенціалів $\varphi_1 - \varphi_2 \neq 0$. Утім, це протирічить умові рівноважності зарядів на поверхні провідника.

Отже, припущення про можливість двох різних рівноважних розподілів заряду по провіднику неможливе. Заряд на провіднику розподіляється єдиним способом. Це означає, що якщо надати провіднику заряд q, то на його поверхні виникне густина заряду σ, різна в кожній точці його поверхні. Якщо провіднику надати ще один такий заряд q, то густина заряду в кожній точці на його поверхні збільшиться у два рази і т. д.

Це дало можливість ввести таке поняття, як електроємність провідника, оскільки відношення заряду q до потенціалу φ провідника буде величиною сталою, яка й отримала назву електроємності.

3.2. Електроємність відокремленого провідника

Відокремленим називають провідник, який перебуває настільки далеко від інших тіл, що впливом їхніх електричних полів можна знехтувати.

Якщо надати відокремленому провіднику заряд q, то електричне поле, яке створює провідник у будь-якій точці простору, визначають суперпозицією полів, що створюються окремими зарядженими ділянками провідника. Кожна така ділянка має заряд $\Delta q_i = \sigma_i \Delta S_i$, де ΔS_i – площа ділянки, σ_i – поверхнева густина заряду на цій ділянці провідника (рис. 3.5, *б*). З'ясуємо, що відбудеться, якщо провіднику надати заряд у n разів більший (або менший), ніж q. Густина заряду в кожній точці провідника згідно з ***теоремою єдиності*** зміниться у n разів.

Напруженість поля $\Delta E_i = k\Delta q_i / r_i^2$, яка створюється кожною ділянкою провідника в будь-якій точці простору, зміниться в n разів, а отже і сумарна напруженість поля у кожній точці зовні провідника зміниться в n разів. Це означає, що елементарна робота, яку здійснює електричне поле провідника над одиничним позитивним пробним зарядом $q_{\text{пр}}$ у разі його переміщення на будь-якій ділянці, зміниться у n разів. У стільки ж разів зміниться робота електростатичних сил, які діють на одиничний позитивний пробний заряд $q_{\text{пр}}$, у разі переміщення його від будь-якої точки поверхні провідника (поверхня провідника еквіпотенціальна) на нескінченність.

Оскільки робота $A_{1\infty} = q_{\text{пр}}\varphi$, де φ – потенціал провідника, то це означає, що φ також зміниться в n разів.

Таким чином, під час зміни заряду провідника в n разів у стільки ж разів зміниться потенціал провідника, а отже, відношення q/φ залишається незмінним. Величину

$$C = q/\varphi \qquad (3.3)$$

називають електроємністю (або просто ємністю) відокремленого провідника. Чим більшого заряду можна надати провіднику за заданого потенціалу, тим більша його ємність. Можна сказати, що *електроємність характеризує здатність провідника накопичувати заряд за цього потенціалу.* Для цього провідника (провідника з ємністю C) є єдине значення заряду q, якого треба надати провіднику, щоб його потенціал був φ.

Електроємність однозначно визначають формою провідника, його розмірами електричними властивостями навколишнього середовища. Електроємність Не залежить від агрегатного стану речовини, хімічної природи матеріалу, наявності всередині порожнин тощо.

Якщо розміри провідника *потрібно* зменшити в *n* разів, то і ємність провідника зменшиться в n разів, і навпаки.

Електроємність не залежить ні від матеріалу провідника, ні від його агрегатного стану, ні від форми й розмірів порожнин всередині нього. Це пов'язано з тим, що надлишкові заряди розподіляються тільки на зовнішній поверхні провідника. Електроємність не залежить також від заряду провідника та його потенціалу.

Одиниця ємності – фарада (Ф).

Один фарада – це ємність такого провідника, потенціал якого змінюється на 1 В у разі надання йому заряду в 1 Кл (формула (3.3)).

Для тіл сферичної форми $\varphi = q/(4\pi\varepsilon_0 R)$, тому електроємність відокремленого сферичного провідника $C = q/\varphi = 4\pi\varepsilon_0 R$, де R – радіус сфери (кулі). Якщо вважати, що Земля має форму кулі, радіус якої $R_3 = 6370$ км, то електроємність Землі

$$C = 4\pi \cdot 8{,}85 \cdot 10^{-12} \cdot 6{,}37 \cdot \text{Ф} =$$
$$= 709 \cdot 10^{-6}\text{Ф} = 709 \text{ мкФ}.$$

Щоб сферичний провідник мав ємність 1 Ф, його радіус має бути

$$R = 1/(4\pi\varepsilon_0) \cong 9 \cdot 10^9 \text{м} = 1411 R_3.$$

На практиці, однак, потрібні пристрої, які мають здатність за малих розмірів і невеликих відносно навколишніх тіл потенціалів нагромаджувати значні за величиною заряди. Для цього використовують

невідокремлені провідники у пристроях, які називають *конденсаторами*.

Покажемо, що ємність невідокремленого провідника завжди більша від ємності відокремленого провідника. Нехай провідник *1* (рис. 3.6) має заряд q_1, і потенціал його при цьому дорівнює φ_1. Тоді ємність провідника

$$C_1 = q_1/\varphi_1.$$

Припустимо, що досить близько від зарядженого провідника *1* з'явився провідник *2* (рис. 3.6). Неважливо, чи провідник *2* заряджений, чи незаряджений. Провідник *2* зазнає впливу електричного поля провідника *1*. На поверхні провідника *2*, яка ближче до провідника *1*, з'являться негативні заряди (протилежні за знаком заряду провідника *1*), а на протилежній поверхні – позитивні заряди (рис. 3.6). Негативні заряди провідника *2* ближчі до провідника *1*, ніж його позитивні заряди.

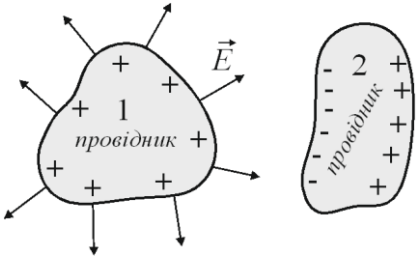


Рис. 3.6. Два невідокремлені провідники

Це означає, що негативні заряди провідника *2* сильніше впливають на потенціал провідника *1* (знижують φ_1), ніж позитивні заряди провідника *2* (вони підвищують φ_1). Результатом цього впливу є те, що потенціал провідника *1* φ_1 знижується і стає $\varphi_1' < \varphi_1$. Оскільки заряд q_1 не змінюється, то ємність провідника *1* стає $C_1' = \frac{q_1}{\varphi_1'} > C_1$. Чим ближче перебуває провідник *2* до провідника *1*, тим більша буде його ємність C_1. Зауважимо, що замість провідника *2* може бути використаний і діелектрик. Результатом такої зміни буде також збільшення ємності провідника *1*.

3.3. Конденсатори

Система, яку зображено на рис. 3.6, дає можливість накопичувати (конденсувати) на провіднику *1* за одного й того самого потенціалу більший заряд, ніж тоді, коли провідник *1* відокремлений. Це дуже важливо для технічного застосування. Вказану систему покладено в основу пристроїв, які називають конденсаторами.

Конденсатор – система із двох (або більше) провідників (які називають обкладинками) що розділені діелектриком, товщина якого менша порівняно з лінійними розмірами обкладинок. Мала відстань між обкладинками порівняно з їх лінійними розмірами дає можливість досягати більшої ємності системи і фактично повністю зосередити поле конденсатора між обкладинками, тобто

уникнути впливу цього поля на навколишні тіла і забезпечити незалежність ємності конденсатора від їх оберненого впливу. Оскільки обкладинки – провідники, то вони екранують зовнішні поля і, таким чином, усувається і їх вплив на ємність конденсатора.

Конденсатори бувають:

а) плоскі – обкладинки є пара паралельних металічних пластин;

б) сферичні – обкладинки є дві концентричні провідні сфери;

в) циліндричні – обкладинки є два коаксіальні (спільна вісь) провідні циліндри.

Обкладинки конденсатора заряджають зарядами однакової величини, але протилежних знаків. У цьому випадку потенціал провідника *1* найбільше знижується зарядженим провідником *2* (рис. 3.6). Ємність такого конденсатора визначають як

$$C = \frac{|q|}{\varphi_1 - \varphi_2} = \frac{q}{U}, \qquad (3.4)$$

де $|q|$ – заряд однієї із обкладинок конденсатора, зазвичай позитивно зарядженої; $U = \varphi_1 - \varphi_2$ – різниця потенціалів, або напруга між обкладинками конденсатора.

Плоский конденсатор

Схему плоского конденсатора показано на рис. 3.7. Простір між обкладинками (пластинами) заповнений діелектриком з відносною діелектричною проникністю ε. Площа пластин *S*, і якщо лінійні розміри пластин великі порівняно з відстанню між ними *d*, то електричне поле між пластинами можна вважити еквівалентним полю між двома нескінченними площинами (див. формулу (1.16)), які заряджені різнойменно і поверхневі густини яких дорівнюють $+\sigma$ і $-\sigma$.

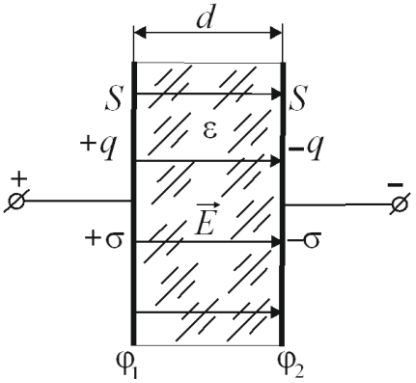


Рис. 3.7. Плоский конденсатор з однорідним діелектриком

Крім того, відстань *d* має бути настільки малою, щоб порушення однорідності поля поблизу його країв можна було не брати до уваги.

Напруженість електричного поля і різниця потенціалів між обкладинками конденсатора в цьому випадку становлять (згідно з (1.16)):

$$E = \frac{\sigma}{\varepsilon\varepsilon_0} = \frac{q}{\varepsilon\varepsilon_0 S};$$

$$\varphi_1 - \varphi_2 = Ed = \frac{q}{\varepsilon\varepsilon_0 S} d.$$

Згідно з визначенням (3.4) ємність плоского конденсатора

$$C = \frac{q}{\varphi_1 - \varphi_2} = \frac{\varepsilon\varepsilon_0 S}{d}. \qquad (3.5)$$

Ємність конденсатора, який має шаруватий діелектрик (рис. 3.8), визначають за такою формулою:

$$C = \frac{\varepsilon_0 S}{\frac{d_1}{\varepsilon_1} + \frac{d_2}{\varepsilon_2}}. \qquad (3.6)$$

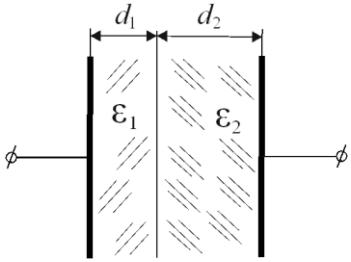


Рис. 3.8. Конденсатора з шаруватим діелектриком

Циліндричний конденсатор

Циліндричний конденсатор – це система із двох коаксіальних циліндрів з радіусами r_1 і r_2 (рис. 3.9).

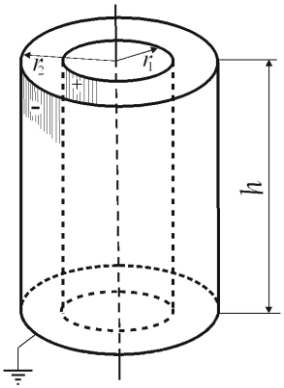


Рис. 3.9. Циліндричний конденсатор

Простір між циліндрами заповнений діелектриком з діелектричною проникністю ε. Якщо висота циліндрів значно перевищує їх радіуси $h >> r_1$ і $h >> r_2$, то можна знехтувати крайовими ефектами і для визначення різниці потенціалів між циліндричними обкладинками конденсатора скористатися формулою (1.18), враховуючи наявність діелектрика. Інакше кажучи, $E = \lambda/(2\pi\varepsilon\varepsilon_0 r)$. Оскільки $E = -d\varphi/dr$, то

$$-d\varphi = Edr = \lambda dr/(2\pi\varepsilon\varepsilon_0 r).$$

Інтегруючи цей вираз у межах від r_1 до r_2, знаходимо:

$$-\int_{\varphi_1}^{\varphi_2} d\varphi = \frac{\lambda}{2\pi\varepsilon\varepsilon_0}\int_{r_1}^{r_2}\frac{dr}{r};$$

$$-(\varphi_2 - \varphi_1) = \frac{\lambda}{2\pi\varepsilon\varepsilon_0}\ln r\Big|_{r_1}^{r_2};$$

$$\varphi_1 - \varphi_2 = \frac{q}{2\pi\varepsilon\varepsilon_0 h}\ln\frac{r_2}{r_1},$$

де $\lambda = q/h$ – лінійна густина заряду.

Тоді ємність циліндричного конденсатора

$$C = \frac{q}{\varphi_1 - \varphi_2} = \frac{2\pi\varepsilon\varepsilon_0 h}{\ln\frac{r_2}{r_1}}. \qquad (3.7)$$

Якщо зазор d між обкладками конденсатора малий порівняно з r_1 ($d << r_1$), то $\ln\frac{r_2}{r_1} = \ln\frac{r_1+d}{r_1} = \ln\left(1+\frac{d}{r_1}\right)$ можна розкласти в ряд і обмежитись лише членами першого порядку малості (формули наближеного обчислення), а саме:

$$\ln\left(1+\frac{d}{r_1}\right)\approx\frac{d}{r_1}.$$

Тоді формула (3.7) набуває такого вигляду (3.5):

$$C=\frac{2\pi\varepsilon\varepsilon_0 r_1 h}{d}=\frac{\varepsilon\varepsilon_0 S}{d},$$

де $S=2\pi r_1 h$ – бічна поверхня циліндричної поверхні. Інакше кажучи, за таких умов циліндричний конденсатор можна наближено вважати плоским.

Сферичний конденсатор

Обкладинки такого конденсатора – це дві концентричні провідні сфери, радіуси яких r_1 і r_2, які розділені тонким шаром діелектрика завтовшки $d=r_2-r_1$ (рис. 3.10). Поля поза конденсатором, що створені внутрішньою і зовнішньою обкладками, взаємно компенсується.

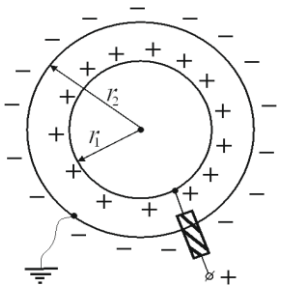


Рис. 3.10. Сферичний конденсатор

Поле між обкладками створюється зарядом q внутрішньої кулі, а заряд $-q$ зовнішньої кулі всередині цієї кулі не створює електричного поля. Напруженість поля між обкладками конденсатора визначають за формулою (1.19). Тоді згідно з (1.34) різниця потенціалів між обкладками конденсатора буде

$$-\int_{\varphi_1}^{\varphi_2} d\varphi=\frac{q}{4\pi\varepsilon\varepsilon_0}\int_{r_1}^{r_2}\frac{dr}{r^2}.$$

Після інтегрування отримаємо:

$$\varphi_1-\varphi_2=\frac{q}{4\pi\varepsilon\varepsilon_0}\left(\frac{1}{r_1}-\frac{1}{r_2}\right).$$

Звідси електроємність сферичного конденсатора

$$C=\frac{q}{\varphi_1-\varphi_2}=\frac{4\pi\varepsilon\varepsilon_0 r_1 r_2}{r_2-r_1}. \quad (3.8)$$

Якщо товщина шару діелек-трика d мала, то можна вважати, що $r_2=r_1+d\approx r_1$.Тоді

$$C=\frac{4\pi\varepsilon\varepsilon_0 r_1^2}{d}=\frac{\varepsilon\varepsilon_0 4\pi r_1^2}{d}=\frac{\varepsilon\varepsilon_0 S}{d},$$

де $S=4\pi r_1^2$ – площа обкладинки конденсатора. Коли сферичний конденсатор має дуже тонкий шар діелектрика, його електроємність можна обчислювати за формулою ємності плоского конденсатора.

Конденсатор характеризується не лише електроємністю, а й «пробивною напругою» – різницею потенціалів між його обкладками, за якої може відбуватися пробій, тобто електричний розряд через шар діелектрика в конденсаторі. Величина пробивної напруги залежить від властивостей діелектрика, його товщини і форми обкладок.

Для отримання потрібної ємності за відповідної напруги в електро- і

радіотехніці конденсатори часто з'єднують між собою в батареї. З'єднання може бути паралельним і послідовним.

У разі паралельного з'єднання (рис. 3.11) напруга U на кожному з конденсаторів однакова. Тоді заряди на кожному з конденсаторів $q_1 = C_1U; q_2 = C_2U; \dots q_n = C_nU$.

Заряд батареї конденсаторів

$$q = q_1 + q_2 + \cdots + q_n = (C_1 + C_2 + \cdots + C_n)U.$$

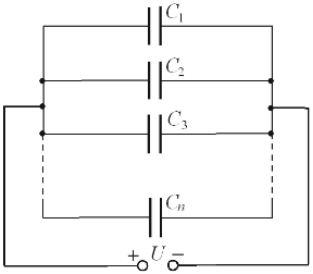


Рис. 3.11. Паралельного з'єднання конденсаторів

Електрична ємність батареї

$$C = \frac{q}{U} = C_1 + C_2 + \cdots + C_n. \quad (3.9)$$

Інакше кажучи, у разі паралельного з'єднання конденсаторів результуюча ємність батареї дорівнює сумі ємностей окремих конденсаторів.

У разі послідовного з'єднання негативно заряджену обкладку першого конденсатора з'єднують з позитивно зарядженою обкладкою другого і т. д. (рис. 3.12).

Якщо на батарею подають напругу $U = \varphi_A - \varphi_B$, то до такої різниці потенціалів заряджаються тільки крайні обкладки першого й останнього конденсатора, причому

$$U=\varphi_A - \varphi_B=(\varphi_1-\varphi_2)+ +(\varphi_2-\varphi_3)+\cdots+(\varphi_{n-1}-\varphi_n), \quad (3.10)$$

де $\varphi_1, \varphi_2, \dots, \varphi_n$ – потенціали пластин конденсаторів (рис. 3.12).

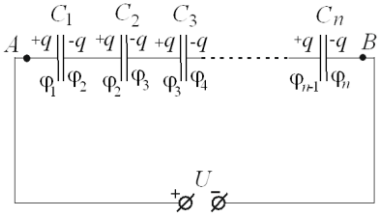


Рис. 3.12. Послідовне з'єднання конденсаторів

Провідник, що з'єднує негативно заряджену обкладку першого та позитивно заряджену обкладку другого конденсатора, можна розглядати разом з обкладками як один провідник (одне провідне тіло). Вільні заряди цього тіла під дією поля прикладеної напруги перерозподіляються так, що на одній обкладці з'являється негативний заряд $-q$, а на іншій – позитивний $+q$, тому можна зробити висновок, що заряди на всіх послідовно з'єднаних конденсаторах незалежно від їхньої ємності однакові й дорівнюють заряду всієї батареї, тобто

$$q = q_1 = q_2 = \dots = q_n.$$

Ємність батареї послідовно з'єднаних конденсаторів

$$C = \frac{q}{\varphi_A - \varphi_B},$$

звідки

$$\varphi_A - \varphi_B = q/C.$$

Для першого конденсатора $\varphi_1 - \varphi_2 = \frac{q}{C_1}$, для другого $\varphi_2 - \varphi_3 = \frac{q}{C_2}$ і т. д. Тоді (3.10) можна переписати так:

$$\frac{q}{C} = \frac{q}{C_1} + \frac{q}{C_2} + \ldots + \frac{q}{C_n}.$$

Після скорочення на q отримуємо формулу для обчислення електроємності батареї послідовно з'єднаних конденсаторів:

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \ldots + \frac{1}{C_n}. \qquad (3.11)$$

Якщо в батарею з'єднують n конденсаторів з однаковою ємністю C_0, то під час паралельного їх з'єднання $C = nC_0$, а під час послідовного з'єднання $C = C_0/n$. У разі послідовного з'єднання конденсаторів результуюча ємність батареї завжди менша від найменшої ємності конденсатора, що входить до батареї.

3.4. Енергія зарядженого відокремленого провідника та конденсатора. Енергія електростатичного поля

Щоб зарядити провідник, слід виконувати роботу проти кулонівських сил електричного відштовхування між однойменними зарядженими частинками. Заряджання провідника можна здійснювати різними способами. Нехай на провідник з нескінченності послідовно переносяться заряди величиною dq. У разі накопичення заряду на провіднику наступне перенесення на нього зарядів dq все більш утруднюватиметься внаслідок зростання кулонівських сил відштовхування між однойменними зарядами. Якщо в який-небудь момент часу потенціал поверхні провідника дорівнює φ, то під час перенесення з нескінченності на провідник заряду dq треба виконати роботу (формула (1.26))

$$dA = dq(\varphi_\infty - \varphi) = -\varphi dq,$$

де $\varphi_\infty = 0$.

Оскільки потенціал провідника пов'язаний з його зарядом і ємністю співвідношенням (3.3), то

$$dA = -\frac{1}{C} q dq.$$

Повна робота у разі зміни заряду провідника від 0 до q становить

$$A = \int_0^q dA = -\frac{1}{C}\int_0^q q dq = -\frac{q^2}{2C}.$$

Ця робота спрямовується на надання енергії зарядженому провіднику, тобто

$$W = -A = \frac{q^2}{2C} = \frac{C\varphi^2}{2} = \frac{q\varphi}{2}. \qquad (3.12)$$

За допомогою формули (3.12) можна знайти енергію зарядженого конденсатора як енергію двох заряджених обкладок (провідників):

$$W = \frac{1}{2}(+q)\varphi_1 + \frac{1}{2}(-q)\varphi_2 = \frac{1}{2}q(\varphi_1 - \varphi_2) = \frac{1}{2}qU.$$

Отже, енергію зарядженого конденсатора з урахуванням (3.4) можна записати так:

$$W = \frac{qU}{2} = \frac{CU^2}{2} = \frac{q^2}{2C}, \qquad (3.13)$$

де $U = \varphi_1 - \varphi_2$ – різниця потенціалів між обкладками конденсатора.

Формулу (3.13) перепишемо в дещо іншому вигляді. Будемо вважати електричне поле всередині конденсатора однорідним, тобто знехтуємо крайовими ефектами. Тоді $U = Ed, C = \varepsilon\varepsilon_0 S/d$ і вираз (3.13) набуває такого вигляду:

$$W = \frac{1}{2}\frac{\varepsilon\varepsilon_0 S}{d}E^2d^2 = \frac{\varepsilon\varepsilon_0 E^2}{2}V, \qquad (3.14)$$

де $V = Sd$ – об'єм, обмежений пластинами (обкладками) конденсатора. Тому можемо вважати, що відношення

$$\omega = \frac{W}{V} = \frac{\varepsilon\varepsilon_0 E^2}{2} \qquad (3.15)$$

є енергією, що припадає на одиницю об'єму конденсатора, тобто об'ємну густину енергії електричного поля. Якщо поле неоднорідне, тобто $E = E(x, y, z)$, то енергія електричного поля в елементарному об'ємі поля становить

$$dW = \omega(x, y, z)dV = \frac{\varepsilon\varepsilon_0 E^2}{2}dV.$$

Тоді формулу (3.14) можна записати як

$$W = \frac{1}{2}\int_V \varepsilon\varepsilon_0 E^2 dV. \qquad (3.16)$$

З математичного погляду співвідношення (3.14) і (3.16) є іншими формами запису співвідношення (3.13), однак формальне математичне перетворення рівнянь часто відкриває можливості зовсім нового фізичного інтегрування виражених ними співвідношень. Так, формула (3.16) виражає електричну енергію у вигляді нескінченної суми доданків, кожний з яких дорівнює $\varepsilon\varepsilon_0 E^2 dV/2$ і належить певному об'єму dV. Це відповідає уявленням теорії поля (теорії близькодії), де вважають, що енергія, подібно до речовини, розподілена у просторі з об'ємною густиною (3.15), яку можна записати ще й так:

$$\omega = \frac{\varepsilon\varepsilon_0 E^2}{2} = \frac{ED}{2} = \frac{D^2}{2\varepsilon\varepsilon_0}. \qquad (3.17)$$

Для ізотропних діелектриків напрямки векторів $\vec{E}$ і $\vec{D}$ збігаються, тому $\omega = \vec{E}\vec{D}/2$.

Враховуючи співвідношення (2.18), отримаємо для ω такий вираз:

$$\omega = \frac{\vec{E}(\varepsilon_0\vec{E}+\vec{P})}{2} = \frac{\varepsilon_0 E^2}{2} + \frac{\vec{E}\vec{P}}{2}. \qquad (3.18)$$

Перший доданок у (3.18) відповідає густині енергії поля $\vec{E}$ у вакуумі. Другий доданок – це густина енергії, яка витрачається на поляризацію діелектрика.

Формула (3.13) відповідає положенню теорії далекодії, де W розглядають як потенціальну енергію заряджених тіл, що притягуються або відштовхуються одне від одного. Інакше кажучи, у концепції далекодії вважають,

що електрична енергія локалізована на заряджених провідниках.

Якому з цих двох уявлень про локалізацію електричної енергії надати перевагу? В межах електростатики принципово неможливо виконати жодного досліду, який дав би можливістьзробити вибір одного з цих уявлень. Це пояснюється тим, що в електростатиці електричне поле невіддільне від зарядів, джерел поля. Величиною і розміщенням зарядів однозначно визначають електростатичне поле. Навпаки, знаючи напруженість електростатичного поля у будь-яких точках простору, можна однозначно вказати розподіл зарядів. Зовсім інша ситуація у разі змінних полів. Змінні електромагнітні поля можуть існувати самостійно, незалежно від джерел, які їх збудили. Електромагнітне поле при цьому існуватиме у вигляді електромагнітних хвиль, які, поширюючись у просторі, переносять енергію. Ця енергія не може бути потенціальною енергією взаємодійних зарядів, оскільки їх немає.

Якщо електростатику розглядати як частину електродинаміки, то слід надавати перевагу теорії поля з її уявленнями про локалізацію електричної енергії у просторі.

Контрольні запитання

1. Які основні властивості провідників в електричному полі?
2. Як визначити напруженість електричного поля біля поверхні зарядженого провідника?
3. Які властивості провідника дають можливість ввести поняття його ємності?
4. Що таке конденсатори? Назвіть їх властивості та вимоги до конструкцій. Які є типи конденсаторів?
5. Як визначити енергію зарядженого провідника та конденсатора? Як визначаються об'ємна густина енергії та загальна формула для розрахунку енергії електричного поля?
6. Де локалізована енергія електричного поля, на заряджених провідниках, чи в просторі, де є електричне поле?

Розділ 4. Постійний електричний струм

4.1. Електричний струм і його види. Струм провідності та умови його існування

Електричний струм – упорядкований (направлений) рух електрично заряджених частинок або заряджених макроскопічних тіл.

За напрямок струму беруть напрямок руху позитивно заряджених

частинок; якщо струм створюється негативно‚ зарядженими частинками (наприклад, електронами), то напрямком струму вважають протилежний напрямок руху частинок. Розрізняють такі види електричного струму: 1) струм провідності; 2) конвекційний струм; 3) дифузійний струм.

1. *Струм провідності – це направлене переміщення вільних електричних зарядів, які перебувають у провідному середовищі (твердому, рідкому чи газоподібному) відносно інших частинок, що входять до складу цього середовища, унаслідок дії на вільні заряди електричного поля.* Про наявність електричного струму *у провідниках* можна з'ясувати за діями, які він викликає: нагрівання провідників, зміна хімічного складу, *створення магнітного поля.* Магнітна дія струму проявляється у всіх без винятку провідниках; у *надпровідниках* не відбувається виділення теплоти, а хімічна дія струму спостерігається переважно в електролітах. Магнітне поле створюється не тільки струмом провідності або конвекційним струмом, але і змінним електричним полем у діелектриках і вакуумі. Величину, яка пропорційна швидкості зміни електричного поля з часом, Дж. Максвелл назвав струмом зміщення (розглянуто далі). Струм зміщення входить у рівняння Максвелла на рівних правах із струмом, який обумовлений рухом зарядів, тому повний електричний струм, який дорівнює сумі струму провідності та струму зміщення, визначає створюване ним магнітне поле.

У металах електричний струм створюють «вільні» електрони, у напівпровідниках – негативно заряджені електрони і позитивно заряджені дірки, у провідних розчинах (електролітах) – позитивно й негативно заряджені іони, у газах – електрони, позитивно і негативно заряджені іони. Струмом провідності є також електричний струм у вакуумі (наприклад, рух електронів в електронних лампах, електронно-променевих трубках).

2. *Конвекційний струм* – рух макроскопічних заряджених тіл як цілого (наприклад, заряджених крапель дощу).

3. *Дифузійний струм* – направлений рух електрично заряджених частинок унаслідок їх дифузії *без дії на них електричного поля*. У напівпровідниках можна створити в певному місці надлишкову концентрацію носіїв заряду порівняно із сусіднім місцем. Утворюється градієнт концентрації (швидкість зміни концентрації носіїв заряду з відстанню), що і призводить до направленого руху заряджених частинок.

У цьому розділі буде розглянуто струм провідності. Для появи й існування струму провідності треба, щоб виконувались дві умови:

Перша умова – наявність у цьому середовищі вільних електричних зарядів – носіїв заряду.

Друга умова – треба, щоб на електричні заряди діяла сила $\vec{F} = q\vec{E}$. Отже, у цьому середовищі має існувати електричне поле, енергія якого витрачалась би на переміщення електричних зарядів. Зрозуміло, що *провідник (метал) зі струмом вже перестає бути еквіпотенціальним.* Порушити еквіпотенціальність провідника можна за допомогою пристрою (джерела енергії), який називають джерелом *електрорушійної сили* (надалі е.р.с.).

До характеристик струму провідності може належати середня швидкість направленого руху носіїв заряду, яку ще називають швидкістю дрейфу $\vec{v}_{\text{др}}$. Відомо, що носії заряду і за відсутності, і за наявності електричного поля у провідниках, беруть участь у хаотичному тепловому русі, середня швидкість якого є набагато більшою, ніж $\vec{v}_{\text{др}}$. З огляду на рівноправність усіх напрямків теплового руху і великої кількості носіїв заряду хаотичний рух не може створити електричного струму. Струм провідності виникає під час створення у провіднику електричного поля і накладання на хаотичний рух направленого руху, який обумовлений дією поля. Якщо $\vec{v}_T$ – вектор швидкості теплового (хаотичного) руху, а $\vec{v}$ – вектор швидкості направленого руху конкретного носія заряду, то усереднення сумарної швидкості носія заряду $\vec{v}_T + \vec{v}$ за всіма носіями заряду дає, що

$$< \vec{v}_T + \vec{v} > = < \vec{v}_T > + < \vec{v} > = < \vec{v} > = \vec{v}_{\text{др}},$$

де середнє значення *вектора* швидкості теплового руху дорівнює нулю, а середнім значенням *вектора швидкості направленого* руху є вектор дрейфової швидкості.

Основною кількісною характеристикою електричного струму є *сила струму I, тобто скалярна величина, яку вимірюють кількістю електрики (величиною заряду), що проходить крізь поперечний переріз провідника за одиницю часу.* Якщо за нескінченно малий проміжок часу dt крізь поперечний переріз провідника проходить заряд величиною dq, то сила струму

$$I = \frac{dq}{dt}. \qquad (4.1)$$

Співвідношення (4.1) виражає миттєве значення сили струму. Якщо сила струму і його напрямок з часом не змінюється, то такий струм називають

стаціонарним, або *постійним*. Силу постійного струму визначають співвідношенням

$$I = \frac{q}{t}, \qquad (4.2)$$

де q – електричний заряд, який проходить крізь поперечний переріз провідника за час t. Силу струму в СІ вимірюють в амперах (А). *Одиниця сили струму «ампер» є основною в* СІ.

У загальному випадку електричний струм можна розподілити по перерізу провідника, через який він проходить, нерівномірно, тому для характеристики розподілу електричного струму по перерізу провідника вводять поняття *густини струму*. Під густиною струму у кожній цій точці провідника треба розуміти границю, до якої прямує відношення сили струму ΔI до елемента площі перерізу ΔS, розміщеного перпендикулярно до напрямку струму, за $\Delta S \to 0$, тобто

$$j = \lim_{\Delta S \to 0} \frac{\Delta I}{\Delta S} = \frac{dI}{dS}. \qquad (4.3)$$

Якщо електричний струм розподілений рівномірно по перерізу провідника, через який він проходить, то густина струму

$$j = \frac{I}{S}, \qquad (4.4)$$

де S – площа поперечного перерізу.

Одиницею густини струму в СІ є ампер на метр квадратний (А/м^2).

Припустімо, що у провіднику заряд окремого носія заряду дорівнює заряду електрона e, їхня концентрація – n, а середня швидкість направленого (дрейфуючого) руху $\vec{v}_{\text{др}}$, то згідно з визначенням густини струму

$$\vec{j} = en\vec{v}_{\text{др}}. \qquad (4.5)$$

Швидкість $\vec{v}_{\text{др}}$ характеризує рух носіїв заряду у цій точці провідника, тому вектор $\vec{j}$ визначає електричний струм у тій самій точці.

Якщо струм створюється рухом позитивних і негативних зарядів, то у загальному випадку вираз для вектора густини струму записують так:

$$\vec{j} = e^+n^+\vec{v}^+_{\text{др}} + e^-n^-\vec{v}^-_{\text{др}}, \qquad (4,6)$$

де n^+і n^- – концентрації позитивних і негативних носіїв заряду; e^+і e^-, $\vec{v}^+_{\text{др}}$ і $\vec{v}^-_{\text{др}}$ – їхні заряди і дрейфова швидкість відповідно.

4.2. Закон Ома та Джоуля–Ленца в інтегральній і диференціальній формах

Нехай у провіднику створене електричне поле $\vec{E}$ внаслідок того, що торець *1* провідника має потенціал φ_1, а торець *2* – потенціал φ_2, причому $\varphi_2 < \varphi_1$ (рис. 4.1).

Будемо вважати, що це електричне поле з напруженістю $\vec{E}$ стаціонарне (незмінне з часом). На вільні

заряди діє сила $\vec{F} = e\vec{E}$, яка зумовлює стаціонарний рух електричних зарядів провідника – постійний струм. Зрозуміло, що це можливо, коли різниця потенціалів $\varphi_1 - \varphi_2$ підтримується постійною. Цю різницю потенціалів називають напругою і позначають буквою U, тобто $\varphi_1 - \varphi_2 = U$.

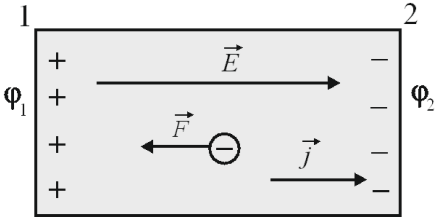


Рис. 4.1. Струм у провіднику внаслідок різниці потенціалів $\varphi_1 - \varphi_2$

Стаціонарне електричне поле за фізичними властивостями не відрізняється від електростатичного поля, оскільки воно є полем електричних зарядів, розміщених у певних ділянках кола струму з незмінною в часі густиною.

Якщо струм постійний, то густина електричних зарядів у будь-якій точці простору не змінюється з часом, хоча відбувається потік електричних зарядів. Тут на місце тих зарядів, що переходять в інше місце, надходять нові. Такі заряди, як підтверджує дослід, створюють у просторі таке саме кулонівське електричне поле, яке б створили нерухомі заряди тієї самої густини. З цього випливає, що електричне поле стаціонарних струмів також є потенціальним. Утім, не слід вважати, що електричне поле постійних струмів і кулонівське поле зарядів одне й те саме – між ними є істотні відмінності. Під час рівноваги зарядів напруженість кулонівського поля у провіднику дорівнює нулю. Електричне поле постійних струмів є також кулонівським полем, але заряди, що його збуджують, перебувають у русі, оскільки поле постійних струмів є і всередині провідника. Якщо його немає, у провіднику струму не було б. Силові лінії електростатичного поля завжди перпендикулярні до поверхні провідника, коли у провіднику немає струму. За наявності стаціонарного струму у провіднику його поверхня не є еквіпотенціальною і лінії напруженості поля, напрямок яких збігається з напрямком струму, можуть бути направлені під гострим кутом до поверхні провідника.

Оскільки поле постійних струмів має потенціальний характер, то можна говорити про потенціал окремих точок кола постійного струму і про різницю потенціалів між ними.

Німецький фізик Г. Ом у 1827 р. експериментально довів, що сила струму

I у провіднику прямо пропорційна напрузі U на його кінцях:

$$I = kU, \qquad (4.7)$$

де $k-$ коефіцієнт пропорційності, який залежить від розмірів провідника, матеріалу, з якого він виготовлений, та його фізичного стану. Його називають ***електричною провідністю провідника***. На практиці частіше користуються оберненою до електропровідності величиною, яку Ом назвав опором провідника, $R = 1/k$.

Тоді формулу (4.7) можна записати так:

$$I = \frac{U}{R}. \qquad (4.8)$$

Опір провідників у СІ вимірюють в омах (Ом), провідність – у сименсах (См).

На основі (4.8) визначають одиницю опору в 1 Ом. Це опір такого провідника, в якому за напруги між його кінцями зі значенням 1 В є струм силою 1 А. *Електричний опір* металів пов'язаний з розсіюванням електронів провідності на теплових коливаннях кристалічних ґраток і її структурних неоднорідностях.

Формула (4.8) виражає в *інтегральній формі закон Ома для однорідної ділянки кола. Ділянку кола називають однорідною, якщо на ній немає джерел струму*. Закон Ома справджується з великою точністю для металевих провідників та електролітів. Відхилення від закону Ома становить близько 1 % для струмів дуже великої густини – кількох мільйонів амперів на 1 см^2. Закон порушується і в тих випадках, коли у провідниках не вистачає носіїв електричних зарядів і настає насичення струму (несамостійний розряд у газах), а також за досить високих напруг та ін. Залежність $I = f(U)$ називають вольт-амперною характеристикою (ВАХ) провідника.

Зауважимо, що опір провідника в першому наближенні не залежить від величини струму, який проходить по провіднику, а залежить від його геометричних розмірів, форми та від матеріалу провідника і його стану. Для однорідних провідників циліндричної форми

$$R = \rho\frac{l}{S}, \qquad (4.9)$$

де $l-$ довжина провідника; $S-$ площа його поперечного перерізу; $\rho-$ коефіцієнт пропорційності, який залежить від матеріалу та фізичного стану провідника. Його називають *питомим опором* матеріалу. Одиницею питомого опору в СІ є Ом на метр (Ом·м). В електротехніці часто довжину провідників вимірюють у метрах, а площу поперечного перерізу – у квадратних міліметрах. Тоді питомий опір доцільно подавати в Ом·мм^2/м.

Питомий опір провідників значною мірою залежить від домішок, а також від способу їх виготовлення. Найменший питомий опір мають срібло і мідь. На практиці для передавання електричної енергії найчастіше використовують дріт із міді або алюмінію. Для виготовлення реостатів, котушок опору використовують сплави з великим питомим опором (ніхром, нікелін). Як уже зазначалося, питомий опір провідників залежить не тільки від хімічного складу речовини, але й від їх стану, зокрема від температури. Залежність питомого опору однорідної речовини від температури характеризують температурним коефіцієнтом опору:

$$\alpha = \frac{1}{\rho}\frac{d\rho}{dT}. \qquad (4.10)$$

Температурний коефіцієнт опору чисельно дорівнює відносній зміні опору провідника за зміни його температури на 1 К. Величина α для різних речовин різна. Для всіх чистих металів температурний коефіцієнт опору близький до 1/273 $К^{-1}$. Температурний коефіцієнт опору дещо змінюється у разі зміни температури цієї речовини. Деякі сплави мають досить малі значення температурного коефіцієнта опору α. Так, для константану $\alpha = 0{,}1\cdot10^{-4}$ $К^{-1}$. Дріт із таких сплавів використовують для виготовлення еталонів опору, котушок вимірювальних місткових схем тощо.

Як показують досліди, для кожного хімічного металевого провідника є певний інтервал температур, в якому залежність питомого опору від температури має лінійний характер, тобто

$$\rho = \rho_0(1 + \alpha t), \qquad (4.11)$$

де ρ_0 – питомий опір, що відповідає початку температурного інтервалу. Здебільшого за початок цього інтервалу беруть температуру 0 °С. Залежність опору металів від температури покладено в основу роботи приладів для вимірювання температури (термометри опору), приладів для вимірювання енергії випромінювання (болометри) та автоматичних пристроїв для термостатування, терморегулювання будь-яких процесів з великою точністю.

Користуючись поняттям густини струму, отримаємо формулу *закону Ома в диференціальній формі*, яка встановлює зв'язок між характеристиками струму в певній точці провідника. Виділимо подумки в однорідному провіднику прямий циліндр завдовжки Δl з площею основи ΔS так, щоб його твірні збігалися з напрямком ліній струму (рис. 4.2). Вважатимемо, що різниця потенціалів у

перерізах *1* і *2* $\Delta\varphi = \varphi_1 - \varphi_2$, $\varphi_1 > \varphi_2$, а опір виділеного циліндра $R = \rho\frac{\Delta l}{\Delta S}$.

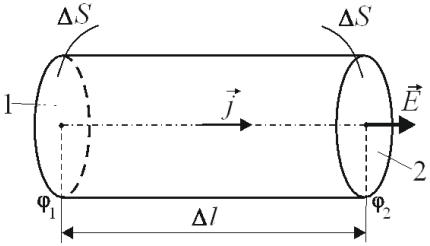


Рис. 4.2. Достатньо малий циліндричний провідник в однорідному провідному середовищі

Згідно із законом Ома (4.8) для цього елемента провідника струм

$$I = \frac{\Delta\varphi}{R} = \frac{\Delta\varphi}{\rho\frac{\Delta l}{\Delta S}} = \frac{1}{\rho}\frac{\Delta\varphi}{\Delta l}\Delta S.$$

Звідси отримаємо, що

$$j = \frac{I}{\Delta S} = \frac{1}{\rho}\frac{\Delta\varphi}{\Delta l}. \qquad (4.12)$$

Позначимо $1/\rho = \sigma$. Величину σ називають ***питомою провідністю*** (у СІ вимірюють в $Ом^{-1} \cdot м^{-1}$). Оскільки $\lim\limits_{\Delta l \to 0}\frac{\Delta\varphi}{\Delta l} = E$ – напруженість електричного поля в цій точці провідника, то вираз (4.12) перепишемо так:

$$j = \sigma E. \qquad (4.13)$$

Оскільки напрямки $\vec{E}$ і $\vec{j}$ збігаються, то рівняння (4.13) можна записати у векторній формі:

$$\vec{j} = \sigma\vec{E}. \qquad (4.14)$$

Співвідношення (4.13) і (4.14) виражають *закон Ома у диференціальній формі для однорідної ділянки кола.*

Закон Ома в диференціальній формі зв'язує густину струму в кожній точці всередині провідника з напруженістю електричного поля у тій самій точці. Закон (4.14) справедливий і для змінних струмів і є одним з основних рівнянь електродинаміки (рівнянь Максвелла).

Оскільки направлений рух носіїв заряду створює електричне поле у провіднику, то можна вважати, що дрейфова швидкість $\vec{v}_{др}$ носіїв заряду пропорційна напруженості поля $\vec{E}$ у провіднику:

$$\vec{v}_{др} \sim \vec{E}, \text{ або } \vec{v}_{др} = u\vec{E}, \qquad (4.15)$$

де u – рухливість носіїв заряду.

Рухливість носіїв заряду чисельно дорівнює швидкості напрямленого руху зарядів, якого вони набувають під дією електричного поля у провіднику, напруженість якого дорівнює 1 В/м.

Тоді формулу (4.5) для густини струму можна записати у вигляді

$$\vec{j} = enu\vec{E}. \qquad (4.16)$$

Порівнюючи (4.16) і (4.14), отримаємо, що

$$\sigma = enu. \qquad (4.17)$$

Отже, питома електропровідність провідників (металів) прямо пропорційна концентрації вільних електронів та їх рухливості.

Робота електричного струму

Якщо по ділянці провідника з опором R, на кінцях якої є напруга U, проходить струм I (рис. 4.1), то через його поперечний переріз за час dt буде перенесено електричний заряд $dq = Idt$. Стандартно напруга $U = \varphi_1 - \varphi_2$ між довільними точками провідника дорівнює роботі, яка виконується при перенесенні одиниці заряду між цими точками провідника. Під час перенесення заряду dq виконується робота

$$dA = Udq = IUdt.$$

У випадку постійного струму

$$A = IUt. \qquad (4.18)$$

На основі закону Ома (4.8) можна записати:

$$A = I^2Rt = \frac{U^2}{R}t. \qquad (4.19)$$

Закон Джоуля–Ленца

Якщо постійний струм проходить по нерухомому провіднику, в якому немає хімічних реакцій, робота струму перетворюється у внутрішню енергію провідника і він нагрівається. Вивчаючи теплову дію електричного струму, вчені Д. Джоуль і Е. Ленц відкрили закон, який сформулювали так: кількість теплоти Q, що виділяється у провіднику під час проходження у ньому постійного струму, прямо пропорційна квадрату сили струму I, опору провідника R і часу проходження струму, тобто

$$Q = I^2Rt. \qquad (4.20)$$

Формула (4.20) виражає закон Джоуля−Ленца в інтегральній формі.

Оскільки кількість теплоти Q і робота A є кількісними мірами перетворення енергії електричного струму у внутрішню енергію, то всі формули роботи можна використати для визначення кількості теплоти.

Запишемо закон Джоуля−Ленца у диференціальній формі через локальні характеристики провідника і електричного поля у ньому. Користуючись співвідношеннями

$$I = jS = \sigma ES \text{ і } \sigma = 1/\rho$$

вираз (4.20) запишемо так: $Q = \sigma^2E^2S^2 \times \times \rho\frac{l}{S}t = \sigma E^2Slt = \sigma E^2Vt,$ де $V = lS -$ об'єм провідника.

Кількість теплоти, що виділяється в одиниці об'єму провідника за одиницю часу (питома теплова потужність),

$$Q_{пит} = \frac{Q}{Vt} = \sigma E^2 = \rho j^2. \quad (4.21)$$

На основі закону Ома в диференціальній формі формулу (4.14) перепишемо (4.21) так:

$$Q_{пит} = \vec{j}\vec{E}. \qquad (4.22)$$

Співвідношення (4.21) і (4.22) виражають закон Джоуля–Ленца у диференціальній формі.

4.3. Електрорушійна сила

Щоб у провіднику існував струм (рис. 4.1), слід підтримувати у ньому електричне поле, тобто підтримувати різницю потенціалів $\varphi_1 - \varphi_2$. Якщо цього не робити, то електрони, рухаючись до торця *1*, будуть понижувати його потенціал φ_1, підвищувати потенціал торця *2* і з часом різниця $\varphi_1 - \varphi_2$ стане дорівнювати нулю, поле всередині провідника зникне і струм перестане існувати. Щоб підтримувати сталою різницю потенціалів $\varphi_1 - \varphi_2 \neq 0$, електрони від торця *1* необхідно переносити до торця *2* за напрямком напруженості поля $\vec{E}$. Інакше кажучи, електрони треба переміщувати від позитивних зарядів до негативних, тобто виконувати роботу переміщення, витрачаючи при цьому енергію. Таке переміщення можна здійснювати тільки за допомогою *сторонніх сил*. Ці сили можуть бути обумовлені хімічними процесами у гальванічних елементах і акумуляторах; дифузією носіїв струму в неоднорідному середовищі; електричними полями, які створюються магнітними полями, що змінюються з часом. Сторонні сили виникають у генераторах за рахунок механічної енергії обертання ротора генератора. *Сторонні сили не можуть бути кулонівськими силами.* Кулонівські сили є потенціальними, і робота таких сил з переміщення зарядів по замкнутому контуру (колу) дорівнює нулю. Під дією кулонівських сил заряди можуть тільки перерозподілятись поверхнею провідника, яка одразу стає еквіпотенціальною. Отже, на певних ділянках кола (ділянка *1–2* на рис. 4.3) сили, що діють з боку електричного поля $\vec{E}$ на носії заряду, не є єдиною причиною струму.

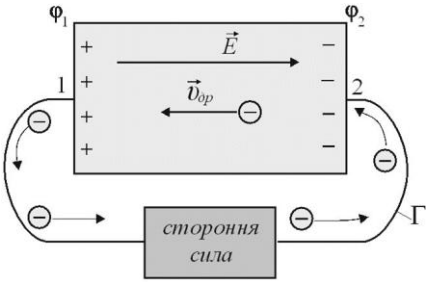


Рис. 4.3. Стороння сила в колі постійного струму

Сторонні сили, які діють на заряди, разом із силами поля $\vec{E}$ зумовлюють постійний струм. Щоб урахувати дію сторонніх сил, вводять поняття вектора *напруженості поля сторонніх сил* $\vec{E}_{ст}$, який за визначенням чисельно дорівнює силі, яка діє на одиничний позитивний заряд.

Сторонні сили, переміщуючи електричні заряди, виконують роботу, тому *фізичну величину, що чисельно дорівнює роботі, яку виконують*

сторонні сили під час переміщення одиничного позитивного заряду, називають електрорушійною силою (е.р.с), що діє в колі:

$$\mathcal{E} = A_{ст}/q. \qquad (4.23)$$

Стороння сила $\vec{F}_{ст}$, що діє на заряд q, становить

$$\vec{F}_{ст} = q\vec{E}_{ст}.$$

На електричних схемах сторонню силу зображують так, як показано на рис. 4.4. Напруженість поля сторонніх сил $\vec{E}_{ст}$ діє в межах пристрою (джерела струму) і направлена від «мінуса» до «плюса».

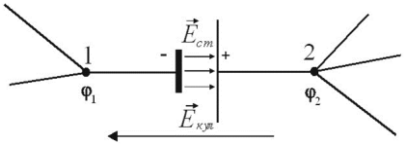


Рис. 4.4. Стороння $\vec{E}_{ст}$ та кулонівська $\vec{E}_{кул}$ сили на ділянці кола

Робота сторонніх сил з переміщення у напрямку струму заряду q уздовж усього замкнутого кола Г (рис. 4.3) становить

$$A_0 = \oint_Г \vec{F}_{ст} d\vec{l} = q \oint_Г \vec{E}_{ст} d\vec{l}.$$

Тоді

$$\mathcal{E} = \frac{A_0}{q} = \oint_Г \vec{E}_{ст} d\vec{l}. \qquad (4.24)$$

Електрорушійну силу $\mathcal{E}$, що діє в замкнутому колі, визначають циркуляцією вектора напруженості сторонніх сил $\vec{E}_{ст}$.

Електрорушійна сила, яка діє на якій-небудь ділянці кола *1–2* (рис. 4.4), становить

$$\mathcal{E} = \int_1^2 \vec{E}_{ст}\, d\vec{l}. \qquad (4.25)$$

Результуюча сила, що діє в колі на заряд,

$$\vec{F} = \vec{F}_{ст} + \vec{F}_{кул} = q\cdot(\vec{E}_{ст} + \vec{E}_{кул}). \qquad (4.26)$$

Робота, яку виконує результуюча сила над зарядом q на ділянці кола *1–2* (рис. 4.5), становить

$$A_{12} = \int_1^2 \vec{F}\, d\vec{l} = q \int_1^2 (\vec{E}_{ст} + \vec{E}_{кул}) d\vec{l} = q \int_1^2 \vec{E}_{ст}\, d\vec{l} + q \int_1^2 \vec{E}_{кул}\, d\vec{l} = q\mathcal{E} + q(\varphi_1 - \varphi_2). \qquad (4.27)$$

Водночас враховано співвідношення (1.34).

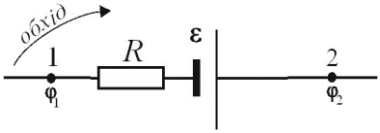


Рис. 4.5. Неоднорідна ділянка кола

Напругою U_{12} на ділянці кола 1–2 називають фізичну величину, що визначається роботою, яку виконує сумарне поле електростатичних і сторонніх сил під час переміщення одиничного позитивного заряду на цій ділянці кола. Отже,

$$U_{12} = \frac{A_{12}}{q} = \varphi_1 - \varphi_2 + \mathcal{E}. \qquad (4.28)$$

Поняття напруги є узагальненим поняттям різниці потенціалів: напруга на кінцях ділянки кола дорівнює різниці

потенціалів у тому випадку, якщо на цій ділянці не прикладена е.р.с.

Ділянку кола, на якій на носії струму діють сторонні сили, називають неоднорідною. Ділянку кола, на якій не діють сторонні сили, називають однорідною. Для однорідної ділянки кола

$$U_{12} = \varphi_1 - \varphi_2. \qquad (4.29)$$

На неоднорідній ділянці кола на носії струму діє сила (4.26), тому дрейфова швидкість $\vec{v}_{\text{др}}$ буде пропорційна силі (4.26). Відповідно, густина струму буде пропорційна сумі напруженостей $\vec{E}_{кул} + \vec{E}_{ст}$:

$$\vec{j} = \sigma(\vec{E}_{кул} + \vec{E}_{ст}) = \frac{1}{\rho}(\vec{E}_{кул} + \vec{E}_{ст}). \qquad (4.30)$$

Цей вираз є *узагальненим законом Ома в диференціальній формі* для неоднорідної ділянки кола. Із закону (4.30) можна отримати й інтегральну форму закону. Для цього помножимо рівняння (4.30) скалярно на елемент $d\vec{l}$ осі провідника, взятому у напрямку вектора $\vec{j}$, і проінтегруємо від перерізу *1* до перерізу *2* нерозгалуженої ділянки кола (рис. 4.5):

$$\int_1^2 \frac{\vec{j}d\vec{l}}{\sigma} = \int_1^2 \vec{E}_{кул}d\vec{l} + \int_1^2 \vec{E}_{ст}d\vec{l}. \qquad (4.31)$$

Оскільки

$$\frac{\vec{j}d\vec{l}}{\sigma} = \frac{jdl\cos 0^\circ}{\sigma} = I\frac{\rho dl}{S} = IdR,$$

де $I-$ сила струму; $S-$ площа поперечного перерізу провідника; $dR-$ опір елемента провідника, довжина якого dl, $\vec{j}||d\vec{l}$ і струм I однаковий в усіх перерізах кола. Тоді

$$\int_1^2 \frac{\vec{j}d\vec{l}}{\sigma} = \int_1^2 IdR = I\int_1^2 dR = IR_{12}, \qquad (4.32)$$

де $R_{12}-$ опір ділянці кола *1–2*.

Враховуючи співвідношення (1.34), (4.25) і (4.32), рівняння (4.31) перепишемо так:

$$IR_{12} = \varphi_1 - \varphi_2 + \mathcal{E}_{12}, \qquad (4.33)$$

де $\mathcal{E}_{12}$ – сумарна е.р.с., яка діє на ділянці кола *1–2*.

Поділивши (4.33) на R_{12}, отримаємо математичний *вираз закону Ома в інтегральній формі для неоднорідної ділянки кола*:

$$I = \frac{\varphi_1 - \varphi_2 + \mathcal{E}_{12}}{R_{12}}. \qquad (4.34)$$

Використовуючи закон (4.34), слід дотримуватись певних правил:

1. Вибрати довільно обхід на ділянці кола *1–2* (рис. 4.5).

2. Вибрати довільно напрямок струму на цій ділянці. Якщо напрямок струму і напрямок обходу збігаються, то алгебраїчну величину струму треба брати як позитивну величину ($I > 0$). Якщо напрямок струму і напрямок обходу на ділянці *1–2* протилежні, то $I < 0$.

3. Електрорушійна сила $\mathcal{E}$ є також величиною алгебраїчною. Вважають, що дія $\mathcal{E}$ має напрямок, що збігається з напрямком поля $\vec{E}_{ст}$ (рис.

4.6). Якщо напрямок $\vec{E}_{ст}$ збігається із напрямком обходу на ділянці *1–2*, то $\mathcal{E} > 0$. Якщо напрямок поля $\vec{E}_{ст}$ і напрямок обходу протилежні, то $\mathcal{E} < 0$ (рис. 4.6).

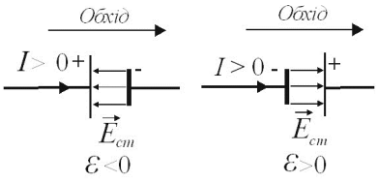


Рис. 4.6. Визначення алгебраїчного знаку величини $\mathcal{E}$

Якщо електричне коло замкнене, то точки *1* і *2* ділянки кола (рис. 4.5) збігаються. Це означає, що $\varphi_1 = \varphi_2$ і $R_{12} = R$, де R – загальний опір кола. Тому закон Ома для замкненого кола має такий вигляд:

$$I = \frac{\mathcal{E}}{R}, \qquad (4.35)$$

де $\mathcal{E}$ – алгебраїчна сума всіх е.р.с., які діють у цьому колі.

Гальванічний елемент

У звичайних батареях (акумуляторах) переміщення носіїв заряду в області, де електричне поле протилежне струму, відбувається за допомогою хімічної енергії. За рахунок цієї енергії позитивний заряд може рухатись до місця з більш високим електричним потенціалом, якщо він при цьому бере участь у хімічній реакції, яка дає йому більше енергії, ніж потрібно для того, щоб подолати силу кулонівського відштовхування. За рахунок цієї енергії негативні заряди рухаються до місця з меншим електричним потенціалом, якщо вони беруть участь у хімічній реакції.

Щоб розглянути, як цей механізм працює, дослідимо це на прикладі нормального елемента Вестона, який має велике значення в лабораторії, оскільки є еталоном під час точного вимірювання напруги (рис. 4.7).

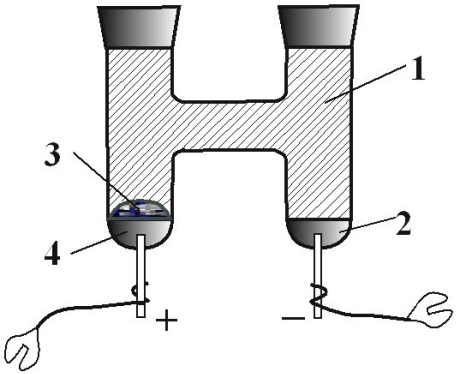


Рис. 4.7. 1 – Водний розчин сульфату кадмію $Cd^{++}SO_4^{--}$ *(електроліт); 2 – амальгама кадмію (кадмій, розчинений у ртуті); 3 – сульфат ртуті* $Hg_2SO_4^{--}$ *(дуже слабо розчинний); 4 – чиста ртуть. Це «ненасичений» елемент*

Він складається із Н-подібної скляної посудини, яка наповнена розчином сульфату кадмію $CdSO_4$. У дно кожного плеча впаяні зовнішні виходи, які здійснюють контакти із внутрішніми електродами. Лівим внутрішнім електродом є рідка чиста ртуть *4*, а

правим – та сама ртуть, в якій розчинений металічний кадмій *2* (у ртуті розчиняється багато металів: такий розчин називають *амальгамою*). Зліва над поверхнею ртуті містяться декілька кристалів сульфату ртуті Hg_2SO_4 (*3*) – сполука, яка дуже слабо розчиняється у воді. Між зовнішніми виходами є різниця потенціалів, причому лівий вихід позитивний відносно правого (абсолютна величина потенціалу не має значення; тут важлива тільки різниця).

Ось як працює такий елемент: частина іонів елемента кадмію переходить з амальгами *2* у водний розчин *1* (рис. 4.8, *б*), причому кожний іон кадмію залишає після себе два електрони й амальгамний електрод *2* набуває великого негативного заряду.

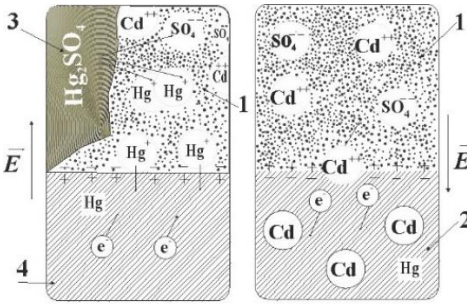


Рис. 4.8. Тут схематично показано, що відбувається на поверхні позитивного електрода (а) і на поверхні негативного електрода (б), коли елемент Вестона підключають до зовнішнього навантаження

Утім, цей процес зупиняється, як тільки в електроді накопичується стільки електронів, що їхня сила притягання іонів Cd^{++} перешкоджає подальшому їхньому виходу у водний розчин *1*. Якщо тепер з'єднати виводи елемента провідним опором, то електрони потечуть від негативного електрода до позитивного. Це дозволить новій порції іонів Cd^{++} перейти у розчин *1*, а залишені ними електрони просто поповнять негативний заряд цього електрода. В колі буде проходити стаціонарний струм, який супроводжується міграцією іонів, які замикають коло всередині водного розчину *1*. Водночас дещо відбувається і на другому електроді (рис. 4.8, *а*). Іони ртуті Hg^+ залишають розчин *3*, зустрічаються з електронами, які надходять зовні, і стають нейтральними атомами ртуті. У розчині нові іони Hg^+ виникають у разі розчинення Hg_2SO_4, причому одночасно в електроліт *1* додаються нові сульфатні іони Hg^+. Водночас, як показано на рис. 4.8, *б*, неперервно відбувається дисоціація атомів кадмію, які надходять в електроліт *1* уже у вигляді іонів Cd^{++}.

У результаті, власне, відбувається вилучення електронів від атомів кадмію і приєднання їх до іонів ртуті. Хімік сказав би, що кадмій окислюється, а ртуть відновлюється.

Елемент працює тому, що цей обмін енергетично вигідний. Відносна сила зв'язку електронів у структурі атома кадмію й атома ртуті така, що спроби атомів ртуті приєднати електрони переважають бажання атомів кадмію утримувати їх. Зауважимо, що на кожній із поверхонь поділу іони Cd^{++} (рис. 4.8, *б*) та іони Hg^{+} (рис. 4.8, *а*) рухаються проти сил $q\vec{E}$ електричного поля.

Розглянемо тепер зміну електричного потенціалу у всій системі як під час проходження струму, так і коли його немає.

На рис. 4.9, *б* по вертикалі відкладено зміну потенціалу вздовж кола, яке розімкнуте в одній точці. Різниця потенціалів на затискачах при розімкнутому колі є е.р.с. елемента, яку позначено як $\mathcal{E}$. Електричне поле дорівнює градієнту потенціалу з оберненим знаком. Як і в будь-якому електростатичному полі, лінійний інтеграл $\oint_\Gamma \vec{E}d\vec{l} = 0$.

На рис. 4.10, *б* показано розподіл потенціалу в тому випадку, коли по зовнішньому опору проходить струм.

Всередині електроліту є електричне поле, яке направлене у той самий бік, що і струм. Розчин сульфату кадмію поводить себе як звичайний омічний опір. Тепер різниця потенціалів на затискачах менша за $\mathcal{E}$ через наявність внутрішнього спаду напруги на електроліті, а також, можливо, через додатковий опір перехідних шарів. Лінійний інтеграл $\oint_\Gamma \vec{E}d\vec{l}$ по всьому колу, як і раніше, дорівнює нулю. Нехай у результаті проходження струму через кожний переріз кола пройшов заряд q, тоді $\mathcal{E}q$ є енергією, яка розсіяна у зовнішньому колі та всередині елемента.

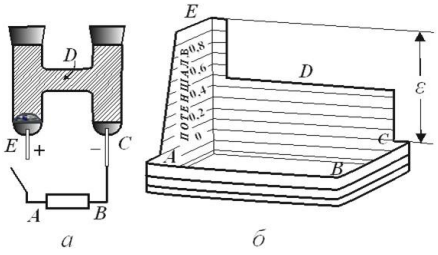


Рис. 4.9. Розподіл в розімкнутому колі (а) потенціалу (б). Потенціал правого затискача C взятий за нульове значення

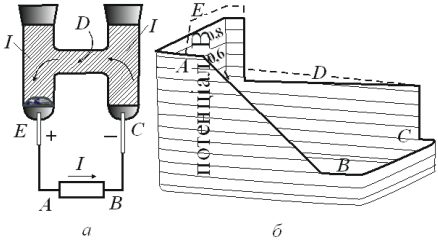


Рис. 4.10. Розподіл потенціалу (б) у випадку, коли по зовнішньому опору проходить струм (а). Зверніть увагу на спад потенціалу всередені електроліту. Цей розгляд непридатний для елемента Вестона, який використовують як еталон. Еталонний елемент застосовують в умовах, коли проходить дуже малий струм

Ця енергія отримана з хімічної енергії елемента, яка є джерелом існування сторонньої сили $\vec{E}_{ст}$.

Ланцюжок реакцій в елементі оборотний. Це означає, що якщо включити в коло назустріч елементу інше джерело з більшою е.р.с., то струм проходитиме в оберненому напрямку, й описані процеси будуть проходити в оберненому порядку. Саме це відбувається під час зарядки батареї акумуляторів.

Електрорушійна сила елемента залежить від властивостей атомів. Її значення близьке до 1 В, оскільки енергія зв'язку зовнішніх електронів в атомі перебуває в діапазоні декількох електронвольт, а в е.р.с. проявляється, власне, різниця цих енергій зв'язку.

Сам елемент Вестона використовують не як джерело електричної енергії, а швидше як еталон різниці потенціалів, тому ситуація, яку показано на рис. 4.10, коли проходить досить значний струм і вихідна напруга падає приблизно на 10 %, є грубим зловживанням до такого елемента.

Електрорушійна сила елемента Вестона відтворювана у високому степені. Якщо водний розчин *1* (рис. 4.7) насичений надлишком сульфату кадмію в області обох електродів, то е.р.с. за 20 °С дорівнює 1,018800...1,019600 В з похибкою до 0,002 % і стабільністю (за 1 рік до 20 мкВ). Водночас допускається проходження струму 0,5...10 мкА. Використовуючи як еталон елемент Вестона і відповідний потенціометр, можна надійно вимірювати різницю потенціалів з точністю до 10^{-6} В.

4.4. Розрахунок розгалужених електричних кіл за допомогою правил Кірхгофа

За допомогою закону збереження заряду і закону Ома (4.33) для неоднорідної ділянки кола можна розрахувати будь-яке складне електричне коло. Такий розрахунок значно полегшується, якщо застосувати два правила Кірхгофа.

Перше правило Кірхгофа встановлює взаємозв'язок між струмами у провідниках, що сходяться у вузлі. **Вузлом розгалуженого кола називають точку, в якій сходяться три і більше провідників** (рис. 4.11).

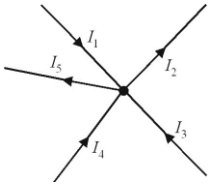


Рис. 4.11. Вузол електричного кола

Нехай постійні струми I_1, I_3, I_4 входять у вузол, а струми I_2, I_5 виходять із нього. Оскільки у вузлах заряди не

накопичуються, то кількість заряду, що приносять струми I_1, I_3, I_4 до вузла, дорівнює кількості заряду, що виносять струми I_2, I_5 з вузла, тому на основі *закону збереження заряду* можна записати, що

$$I_1 + I_3 + I_4 - I_2 - I_5 = 0, \quad (4.36)$$

або

$$-I_1 - I_3 - I_4 + I_2 + I_5 = 0. \quad (4.37)$$

Рівняння (4.36) або (4.37) виражають перше правило Кірхгофа. Для зручності користування цим рівнянням струмам, що входять у вузол, надають один знак, а струмам, що виходять – протилежний. Так, якщо вважати струм, що входить у вузол, додатним, а струм, що виходить з вузла, – від'ємним, то перше правило Кірхгофа можна сформулювати так: ***алгебраїчна сума усіх сил струмів, що сходяться у будь-якому вузлі розгалуженого кола, дорівнює нулю***:

$$\sum_{i=1}^{m} I_i = 0. \quad (4.38)$$

Друге правило Кірхгофа. Виділимо в розгалуженому колі простий замкнутий контур, наприклад *ABCDA* (рис. 4.12).

Застосуємо закон Ома (4.33) до кожної ділянки кола цього контура:

$$I_1 R_1 + I_1 r_1 = \varphi_A - \varphi_B - \mathcal{E}_1;$$

$$I_2 R_2 + I_2 r_2 = \varphi_B - \varphi_C + \mathcal{E}_2;$$

$$-I_3 r_3 = \varphi_C - \varphi_D + \mathcal{E}_3;$$

$$-I_4 R_4 = \varphi_D - \varphi_A.$$

Додамо почленно ліві та праві частини цих рівностей. Тоді отримаємо, що

$$I_1 R_1 + I_1 r_1 + I_2 R_2 + I_2 r_2 - I_3 r_3 + \\ - I_4 R_4 = -\mathcal{E}_1 + \mathcal{E}_2 + \mathcal{E}_3. \quad (4.39)$$

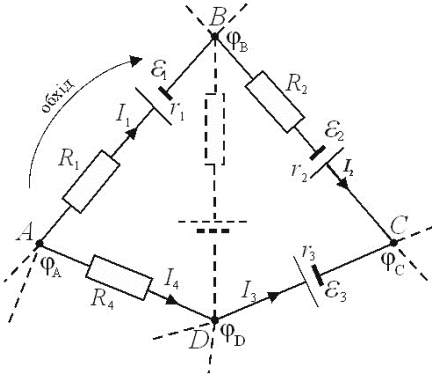


Рис. 4.12. Простий замкнутий контур в розгалуженому колі

Аналогічні співвідношення отримаємо для будь-якого простого замкнутого контура. Узагальнюючи співвідношення (4.39), запишемо, що

$$\sum IR + \sum Ir = \sum \mathcal{E}. \quad (4.40)$$

Співвідношення (4.40) виражає ***друге правило Кірхгофа: у будь-якому простому замкнутому контурі, довільно обраному у розгалуженому електричному колі, алгебраїчна сума добутків сил струмів на опори відповідних ділянок цього контура дорівнює алгебраїчній сумі е.р.с., що діють у контурі.***

У сумі (4.40) знаки струмів і е.р.с. $\mathcal{E}$ (див. прąвила на стор. 86) беруть згідно з вибраним обходом по контуру (рис. 4.12). Обхід для всіх замкнутих контурів розгалуженого кола треба здійснювати в одному напрямку. Найчастіше обхід контура вибирають за стрілкою годинника.

Якщо під час розв'язування системи рівнянь, складених на основі (4.38) і (4.40), сили струмів виявились додатними, то це означає, що напрямок їх обрано правильно. Інакше дійсний напрямок струму протилежний до обраного нами.

Контрольні запитання

1. Дайте визначення сили струму. Які умови існування струму провідності? Наведіть характеристики струму.

2. Які причини існування електричного опору металів?

3 Дайте визначення електрорушійної сили сторонніх сил. У чому відмінність між е.р.с. та різницею потенціалів?

4. Напишіть закон Ома в інтегральній та диференціальній формах для неоднорідної ділянки кола.

5. Напишіть закон Ома в інтегральній та диференціальній формах для однорідної ділянки кола.

6. Напишіть закон Ома для замкнутого кола.

7. Дайте визначення вектору густини струму. Яка швидкість носіїв заряду є дрейфовою? Який її фізичний зміст?

8. Сформулюйте перше правило Кірхгофа. На основі якого фізичного закону воно встановлене?

9. Сформулюйте друге правило Кірхгофа. На основі якого фізичного закону воно ґрунтується?

10. Напишіть і поясніть закон Ленца–Джоуля в інтегральній та диференціальній формах.

Розділ 5. Магнітне поле

5.1. Взаємодія струмів. Закон Ампера. Магнітне поле

Нерухомі точкові заряди (рис. 5.1) взаємодіють за законом Кулона (формула (1.2)). Ця взаємодія, за теорією близькодії, здійснюється через електростатичне поле, яке існує навколо нерухомих заряджених частинок. Разом з тим на основі закону Кулона не можна пояснити характеру взаємодії тих самих зарядів, які рухаються.

Два позитивні нерухомі заряди відштовхуються із силою Кулона (рис. 5.1). З початком руху зарядів виникає нова якість взаємодії (рис. 5.2). Експериментально досліджувати взаємодію двох зарядів, які рухаються,

складно. Утім, коли рухається багато зарядів, то експериментальні дослідження реальні. Направлений рух зарядів – це струми.

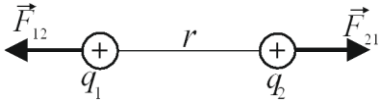


Рис. 5.1. Два нерухомі заряди

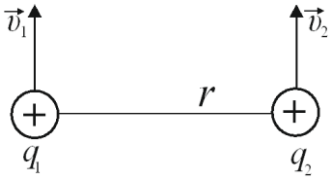


Рис. 5.2. Два заряди, що рухаються

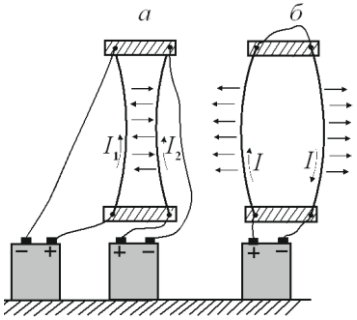


Рис. 5.3. Взаємодія провідників зі струмом

Ампер у 1820 р. відкрив, що два тонкі прямолінійні металічні провідники, якими проходить електричний струм, взаємно притягуються (рис. 5.3, *а*), якщо струми мають однаковий напрямок (позитивні заряди рухаються в одному напрямку (рис. 5.2) хоча між ними діє кулонівська сила відштовхування). Якщо ж струми мають протилежні напрямки (позитивний і негативний заряди рухаються в одному напрямку і між ними діє кулонівська сила притягання), то провідники відштовхуються (рис. 5.3, *б*).

Зауважимо, що електрична взаємодія самих проводів дорівнює нулю, оскільки загалом вони електронейтральні.

Дослідним способом Ампер встановив закон взаємодії двох тонких прямолінійних паралельних провідників, якщо відстань між ними r набагато менша за їх довжину:

$$F_l = 2kI_1I_2/r, \qquad (5.1)$$

де I_1 та I_2 – сила струму в першому і другому провідниках відповідно; F_l – сила їхньої взаємодії, що припадає на одиницю довжини проводу; $2k$ – коефіцієнт пропорційності.

На основі формули (5.1) встановлюють одиницю сили струму в СІ: ампер – це сила постійного струму, який, проходячи двома паралельними прямолінійними дуже тонкими провідниками нескінченної довжини, які розміщені на відстані $r = 1$ м один від одного у вакуумі, викликав би між цими провідниками силу, що дорівнювала б $2 \cdot 10^{-7}$ ньютонам на кожний метр довжини.

У раціональному вигляді формулу (5.1) записують так:

$$F_l = \frac{\mu_0}{4\pi} \cdot \frac{2I_1I_2}{r}, \qquad (5.2)$$

де μ_0 – *магнітна стала*. Якщо $I_1 = I_2 =$ $= 1$ А, $r = 1$ м, $F_l = 2 \cdot 10^{-7}$ Н/м, то із (5.2) отримуємо, що

$$\mu_0 = 4\pi \cdot 10^{-7} \text{ Гн/м}. \qquad (5.3)$$

Взаємодія струмів здійснюється через поле, яке називають магнітним. Магнітне поле отримало назву у зв'язку з тим, що Ерстед (майже одночасно з відкриттям Ампера) відкрив дію електричного струму на магнітну стрілку (рис. 5.4).

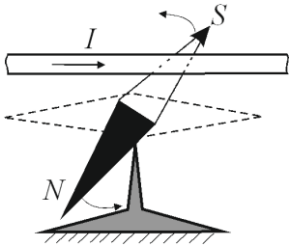


Рис. 5.4. Дія електричного струму на магнітну стрілку

Магнітну стрілку розміщували під проводом. Під час пропускання по провіднику струму стрілка орієнтувалась перпендикулярно до проводу. У разі зміни напрямку струму стрілка поверталась на 180°. Те саме відбувалося під час переносу стрілки із положення під проводом у положення над проводом. Коли віддаляли магнітну стрілку від провідника, орієнтуюча дія струму зменшувалась.

У цих дослідах було зроблено декілька фундаментальних відкриттів. *По-перше*, показано, що електричний струм створює магнітне поле (на магнітну стрілку може діяти тільки магнітне поле), *по-друге*, що величина магнітного поля у різних точках простору навколо провідника зі струмом різна (зменшується з відстанню від провідника), і, *по-третє*, що магнітне поле має напрямок у просторі, тобто є полем векторним. Після відкриття Ерстеда почалися інтенсивні дослідження магнітних полів і струмів (рис. 5.5).

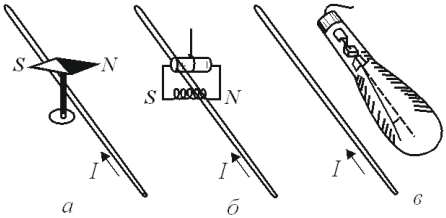


Рис. 5.5. Взаємодії магнітного поля і струмів

Дія магнітного поля провідника зі струмом на магнітну стрілку (рис. 5.5, *а*) нічим не відрізняється від його дії на котушку, якою проходить струм від батареї (рис. 5.5, *б*), або дії на рухомий пучок електронів в електронно-променевій трубці (рис. 5.5, *в*). Останні досліди свідчать не тільки про те, що магнітне поле створюється рухомими електричними зарядами, але й про те, що магнітне поле, у свою чергу, діє лише на рухомі електричні заряди і цією дією інтегрально визначається взаємодія

провідників зі струмом або провідника зі струмом і магнітної стрілки, або взагалі магнітна взаємодія. Отже, навколо будь-якого рухомого заряду, чи то буде електрон, іон або заряджене тіло, крім електричного поля існує також і магнітне поле. Електричне поле діє як на рухомі, так і на нерухомі електричні заряди. Магнітне поле діє лише на рухомі електричні заряди.

5.2. Взаємодія елементів струму

У дослідах Ампера насамперед було встановлено, що сила взаємодії двох провідників пропорційна величині струму в кожному із провідників. Далі досліди показали, що коли провід зі струмом зігнути, як показано на рис. 5.6, *а*, то він не чинить магнітної дії. І навпаки, такий провідник не зазнає дії сили з боку інших провідників зі струмом. Магнітна дія не спостерігається і тоді, коли одну частину проводу (і притому довільно) обмотати навколо другої (рис. 5.6, *б*).

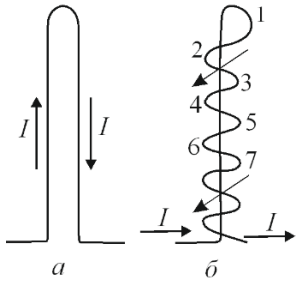


Рис. 5.6. Зігнуті провідники (а та б) не чинять магнітної дії

Із цих результатів випливає висновок, що які-небудь елементи провідника $d\vec{l}_1$, $d\vec{l}_2$ і $d\vec{l}_3$ разом чинять таку саму магнітну дію, як один елемент $d\vec{l}$ (рис. 5.7), що замикає ці відрізки. Зокрема, дія зігнутих відрізків *12* і *23* провідника *б* з рис. 5.6 така, як коли замість них був би прямолінійний відрізок, що сполучає точки *1* та *3*, дія *34* і *45* дорівнює дії *35* і т. д., і тому дія всього провідника *б* така сама, як і провідника *а* (рис. 5.6), тобто дорівнює нулю.

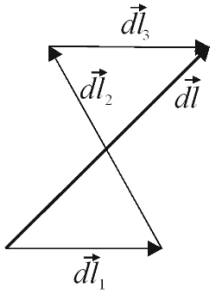


Рис. 5.7. До поняття елемента струму

Із сказаного випливає, що магнітна дія нескінченно малого відрізка проводу залежить від добутка $Id\vec{l}$, де I – величина струму, а $d\vec{l}$ – вектор, що має довжину відрізка dl і направлений вздовж струму. ***Цей добуток називають елементом струму.***

Поняття елемента струму в законах магнітної взаємодії відіграє те саме значення, що й поняття точкового заряду в законах електричної взаємодії.

Сила $d\vec{F}_{12}$, з якою діє який-небудь елемент струму $I_1 d\vec{l}_1$ на інший елемент струму $I_2 d\vec{l}_2$, пропорційна величині кожного з елементів струму, тобто їх добутку:

$$dF_{12} \sim I_1 d\vec{l}_1 \cdot I_2 d\vec{l}_2.$$

Ця сила залежить також від відстані r_{12} між елементами струму. Ампер припустив, що вона обернено пропорційна відстані в деякому степені n і з дослідів зробив висновок, що $n = 2$:

Величина і напрямок сили взаємодії двох елементів струму залежать ще й від їх взаємного ***орієнтування***. Визначення цієї залежності було дуже складним і потребувало низки досліджень різних учених. Спочатку розглянемо випадок, коли відрізки провідників (елементи струму) $d\vec{l}_1$ і $d\vec{l}_2$ перебувають в одній площині, і позначимо через $\vec{r}_{12}$ радіус-вектор, проведений від першого відрізка до другого (рис. 5.8).

Згідно зі сказаним вище, дію відрізка $d\vec{l}_1$ можна замінити дією двох його складових $d\vec{l}_{1\tau}$ і $d\vec{l}_{1n}$, одна з яких направлена вздовж радіус-вектора $\vec{r}_{12}$, друга – перпендикулярна до нього.

З даних досліду зроблено висновок, що відрізок $d\vec{l}_{1\tau}$ не чинить жодної дії на $d\vec{l}_2$ за будь-якого орієнтування останнього. Тому сила, яка діє на $d\vec{l}_2$, пропорційна $dl_{1n} = dl_1 \cdot \sin\alpha_1$, де α_1 – кут між векторами $d\vec{l}_1$ і $\vec{r}_{12}$:

$$dF_{12} \sim \sin\alpha_1 = \sin(d\vec{l}_1, \vec{r}_{12}).$$

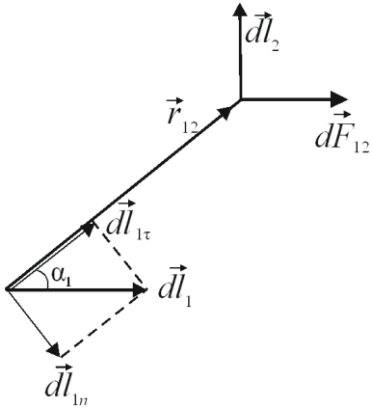


Рис. 5.8. Взаємодія двох елементів струму, які перебувають в одній площині

У загальному випадку відрізки провідників $d\vec{l}_1$ і $d\vec{l}_2$ можуть не перебувати в одній площині. Тоді відрізок $d\vec{l}_2$ можна розкласти на дві складові, одна з яких, наприклад $d\vec{l}_{2\tau}$, перебуває у площині П, що містить $d\vec{l}_1$ і $\vec{r}_{12}$ (рис. 5.9), а друга, наприклад $d\vec{l}_{2n}$, перпендикулярна до цієї площини.

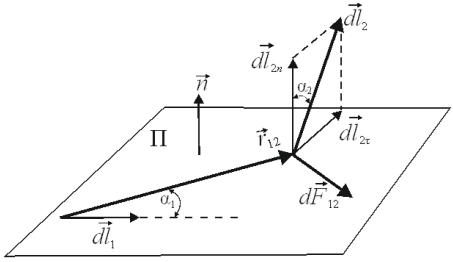


Рис. 5.9. Взаємодія двох елементів струму в загальному випадку

Досліди показують, що сила, яка діє на $d\vec{l}_{2n}$, завжди дорівнює нулю, тому сила, яка діє на $d\vec{l}_2$, пропорційна складовій $dl_{2\tau} = dl_2 \sin\alpha_2$, де α_2 – кут між вектором $d\vec{l}_2$ і нормаллю $\vec{n}$ до площини П:

$$dF_{12} \sim \sin\alpha_2 = \sin(d\vec{l}_2, \vec{n}).$$

Об'єднуючи всі знайдені дослідні результати в одну формулу, знаходимо:

$$dF_{12} = k\frac{I_1 dl_1 \cdot I_2 dl_2 \cdot \sin\alpha_1 \cdot \sin\alpha_2}{r^2{}_{12}}. \quad (5.4)$$

Тут k означає коефіцієнт пропорційності, який залежить від вибору системи одиниць (у СІ $k = \frac{\mu_0}{4\pi}$).

Формула (5.4) виражає величину сили dF_{12}. Напрямок цієї сили визначають за таким правилом: сила, яка діє на елемент струму $I_2 d\vec{l}_2$, перпендикулярна до цього елемента і перебуває у площині, що містить елемент струму $I_1 d\vec{l}_1$ й радіус-вектор $\vec{r}_{12}$ (рис. 5.9). Водночас напрямок сили визначають ще ***правилом правого свердлика***: якщо обертати свердлик з правою нарізкою так, щоб його руків'я переміщувалась від елемента струму $I_2 d\vec{l}_2$ до нормалі $\vec{n}$, то поступальний рух свердлика збігається з напрямком $d\vec{F}_{12}$. Напрямок нормалі $\vec{n}$ також визначають за цим правилом: він збігається з поступальним рухом свердлика, якщо його обертати від елемента струму $I_1 d\vec{l}_1$ до радіус-вектора $\vec{r}_{12}$ (рис. 5.9). Величину і напрямок сили магнітної взаємодії можна подати однією формулою, якщо скористатися поняттям векторного добутку двох векторів. Векторний добуток $d\vec{l}_1 \times \vec{r}_{12}$ направлений уздовж нормалі $\vec{n}$ до площини П, а векторний добуток цього вектора на вектор $d\vec{l}_2$, тобто подвійний векторний добуток $d\vec{l}_2 \times d\vec{l}_1 \times \vec{r}_{12}$, перебуває у площині П і перпендикулярний до $d\vec{l}_2$, тобто направлений так само, як і сила $d\vec{F}_{12}$. Отже, можемо записати, що

$$d\vec{F}_{12} = \frac{\mu_0}{4\pi}\frac{I_2 d\vec{l}_2 \times I_1 d\vec{l}_1 \times \vec{r}_{12}}{r_{12}^3}. \quad (5.5)$$

Закон Ампера, виражений формулами (5.4) і (5.5), є основним у вченні про магнетизм і відграє таке саме значення, що і закон Кулона в електростатиці. Користуючись цим законом, можна обчислити силу взаємодії між елементами струму, на які можна поділити будь-які замкнені провідники, а підсумовуючи потім ці сили для всіх елементів струму, можна обчислити силу взаємодії між замкненими контурами скінченних розмірів будь-якої форми.

На останок зазначимо, що закон Ампера у формі (5.4) і (5.5), на перший погляд не задовольняє третій закон Ньютона. Так, наприклад, якщо розглянути два елементи струму *1* і *2*, що зображено на рис. 5.10, то із закону

Ампера (5.5) випливає, що сила F_{12}, яка діє на елемент *2*, дорівнює нулю; але сила F_{21}, що діє на елемент *1*, не дорівнює нулю.



Рис. 5.10. Взаємодія двох елементів струму

Такий результат отримуємо тому, що на досліді можна дослідити тільки взаємодію замкнених контурів скінченної величини і не можна реалізувати взаємодію окремих елементів струму. Отже, з досліду можна вивести закон взаємодії елементів струму тільки з точністю до деякого доданка, який перетворюється в нуль під час підсумовування по замкненому контуру. Такий доданок опущено в (5.4) і (5.5), і це є причиною начебто порушення третього закону Ньютона. Утім, цей доданок не відграє жодного значення, оскільки закон Ампера ми завжди застосовуємо тільки до замкнених контурів, а в цьому випадку він завжди дає результати, які узгоджуються з третім законом Ньютона.

Магнітне поле

Під час пояснення магнітної взаємодії струмів ми стикаємося з тими самими питаннями, що й під час пояснення електричної взаємодії зарядів. І тут можна запитати, чому виникають сили, які діють на контур зі струмом у присутності іншого контура, і як ці сили передаються від одного провідника до іншого? Чи відбуваються які-небудь зміни у просторі навколо проводу зі струмом, коли іншого проводу немає і магнітна взаємодія не проявляється?

З тих самих причин, які викладено в розд. 1, сучасна фізика спростовує можливість далекодії в магнітних явищах, так само як і в електричних. Причину виникнення сил магнітної взаємодії ми бачимо в появі навколо провідників зі струмом *магнітного поля.* Далі ми побачимо, що магнітне поле є носієм низки фізичних властивостей. Основна властивість магнітного поля полягає в тому, що на провідники зі струмом, які перебувають у ньому, діють сили.

Магнітне поле виникає навколо проводу зі струмом завжди, навіть у разі відсутності інших провідників, коли магнітної взаємодії не спостерігається. І в цьому випадку у просторі, який оточує провідник зі струмом, відбуваються певні фізичні зміни. Основне завдання дослідження магнітних явищ полягає у вивченні властивостей магнітного поля і законів, яким воно підпорядковується.

Термін «магнітне поле» увів у 1845 р. англійський фізик М. Фарадей, який вважав, що як електрична, так і

магнітна взаємодія здійснюється через єдине матеріальне (електромагнітне) поле. Англійський фізик Дж. Максвелл у 1873 р. створив класичну теорію електромагнітного поля, а у 20-х роках XX століття було сформульовано квантову теорію.

5.3. Індукція магнітного поля. Закон Біо−Савара−Лапласа

Кількісною характеристикою магнітного поля є спеціальна фізична величина – індукція магнітного поля, яку ми визначаємо аналогічно до напруженості електричного поля (див. розд. 1, формули (1.5) і (1.6)). Основним законом магнітної взаємодії струмів є закон Ампера (5.5), який виражає силу взаємодії двох елементів струму. У ньому можна виділити

$$d\vec{B}_1 = \frac{\mu_0}{4\pi}\frac{I_1 d\vec{l}_1 \times \vec{r}_{12}}{r^3{}_{12}}, \qquad (5.6)$$

тобто частину, яка не містить другого елемента струму $I_2 d\vec{l}_2$. Ця величина $d\vec{B}_1$ залежить тільки від елемента струму $I_1 d\vec{l}_1$ і положення тієї точки, де перебуває елемент струму $I_2 d\vec{l}_2$, і тому характеризує магнітне поле елемента струму $I_1 d\vec{l}_1$ в точці перебування елемента струму $I_2 d\vec{l}_2$. Вона й отримала назву ***індукції магнітного поля.*** Вводячи поняття індукції магнітного поля в ***закон Ампера*** (5.5), можна подати силу, яка діє на елемент струму $I_2 d\vec{l}_2$, у такому вигляді:

$$d\vec{F}_{12} = I_2 d\vec{l}_2 \times d\vec{B}_1. \qquad (5.7)$$

Отже, ми отримуємо таке визначення індукції магнітного поля: *індукція магнітного поля $d\vec{B}$, яка створюється в цій точці елементом струму $Id\vec{l}$, становить*

$$d\vec{B} = \frac{\mu_0}{4\pi}\frac{Id\vec{l}\times\vec{r}}{r^3}, \qquad (5.8)$$

де $\vec{r}$ – радіус-вектор, проведений з елемента провідника $d\vec{l}$ у розглядувану точку. Через те, що векторний добуток двох векторів є також вектором, то очевидно, що ***індукція магнітного поля є вектором***.

З (5.8) випливає, що *величина* індукції магнітного поля у точці, віддаленій на відстань r від елемента струму, становить

$$dB = \frac{\mu_0}{4\pi}\frac{Idl\sin\alpha}{r^2}, \qquad (5.9)$$

де α – кут між $d\vec{l}$ і радіусом-вектором $\vec{r}$, проведеним від елемента струму в розглядувану точку (рис. 5.11).

Напрямок вектора індукції магнітного поля перпендикулярний до $d\vec{l}$ і $\vec{r}$, тобто перпендикулярний до площини, що містить обидва ці вектори. Цей напрямок підлягає правилу правого свердлика: напрямок магнітного поля збігається з напрямком руху кінця руків'я

свердлика з правою нарізкою, що рухається поступально в напрямку струму. Так, наприклад, якщо струм проходить вертикально згори вниз (рис. 5.11), то правий свердлик треба обертати за годинниковою стрілкою (якщо дивитися зверху), а кінець його руків'я, що перебуває в точці a, рухатиметься при цьому від рисунка до читача; напрям магнітного поля в точці a показаний на рис. 5.11. У точці $б$ поле направлене протилежно від читача за рисунок.

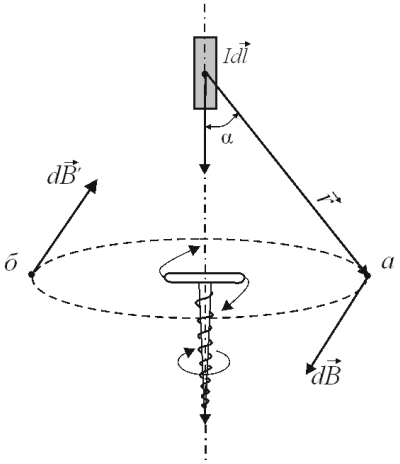


Рис. 5.11. Індукція магнітного поля елемента струму

Формули (5.8) і (5.9), які виражають індукцію магнітного поля елемента струму, називають ***законом Біо–Савара–Лапласа***. Закон Біо–Савара–Лапласа є одним з основних експериментальних законів електромагнітних явищ, і він, подібно до закону Кулона, лежить в основі класичної електродинаміки. Цей закон дає змогу розраховувати індукцію магнітних полів струмів.

Для магнітного поля, як і для електричного, справедливий ***принцип накладання, або суперпозиції.*** Якщо є кілька елементів струму або кілька контурів зі струмом, кожний з яких створює магнітне поле, то у разі одночасної дії всіх струмів індукція $\vec{B}$ результуючого поля дорівнює векторній сумі індукцій окремих полів:

$$\vec{B} = \vec{B}_1 + \vec{B}_2 + \vec{B}_3 + ... = \sum \vec{B}_k. \quad (5.10)$$

Користуючись виразом (5.8) або (5.9) для індукції магнітного поля елемента струму і принципом накладання (5.10), можна обчислити магнітне поле будь-якого провідника зі струмом.

Слід зазначити, що таку характеристику магнітного поля, як індукція $\vec{B}$, часто вводять таким чином.

Розглянемо замкнений плоский контур зі струмом, розміри якого малі порівняно з відстанню до струмів, що утворюють поле. За позитивний напрямок нормалі $\vec{n}$ беруть напрямок поступального руху свердлика, руків'я якого обертається за напрямком струму в контурі (рис. 5.12).

Контур зі струмом характеризується ***магнітним моментом*** $\vec{p}_m$, величина якого дорівнює добутку сили

струму I, що проходить у контурі, на площу поверхні контуру S: $\vec{p}_m = IS\vec{n}$, де $\vec{n}$ – одиничний вектор нормалі до поверхні рамки. Напрямок вектора $\vec{p}_m$ збігається з напрямком позитивної нормалі рамки.

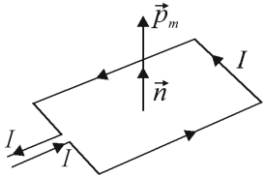


Рис. 5.12. Контур зі струмом, як магнітний диполь

Контур зі струмом у магнітному полі повертається, набуваючи рівноважного положення, і його позитивна нормаль розміщується вздовж осі магнітної стрілки в напрямку від її магнітного полюса (рис. 5.13). Контуром зі струмом можна скористатись і для кількісного опису магнітного поля. На контур у магнітному полі діє пара сил. Обертальний момент пари сил M залежить від властивостей контура:

$$M \sim p_m.$$

Якщо контур зі струмом повернути на 90° від рівноважного положення, то на нього буде діяти максимальний обертальний момент M_{max}.

Якщо у це місце магнітного поля поміщати контури з різними магнітними моментами, то на них діятимуть різні обертальні моменти, але відношення M_{max}/p_m для всіх контурів однакове і служить кількісною характеристикою магнітного поля:

$$B = M_{max}/p_m.$$

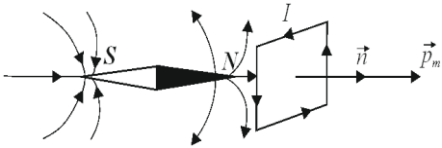


Рис. 5.13. Рівноважне положення контура зі струмом у магнітному полі

Магнітну індукцію $\vec{B}$ у цьому місці магнітного поля визначають максимальним обертальним моментом, що діє на контур з одиничним магнітним моментом.

Одиниця магнітної індукції – Тесла:

1 Тл – магнітна індукція такого магнітного поля, в якому на рамку з магнітним моментом 1 А·м^2, діє максимальний момент сили 1 Н·м.

За напрямок магнітної індукції беруть напрямок магнітного моменту контура $\vec{p}_m$, який перебуває в рівноважному положенні у цьому полі.

Для графічного зображення магнітних полів зручно користувались лініями магнітної індукції.

Лініями магнітної індукції називають такі лінії, дотичні до яких у кожній точці збігаються з напрямком вектора $\vec{B}$ у цих точках поля.

Напрямок лінії індукції магнітного поля струму визначають за правилом свердлика: якщо вкручувати свердлик за напрямком руху струму у провіднику, то напрямок руху його руків'я покаже напрямок ліній магнітної індукції.

Лінії магнітної індукції можна спостерігати за допомогою дрібних металевих ошурків, які в магнітному полі поводять себе як маленькі магнітні стрілки. Вигляд ліній магнітної індукції простих магнітних полів показано на рис. 5.14.

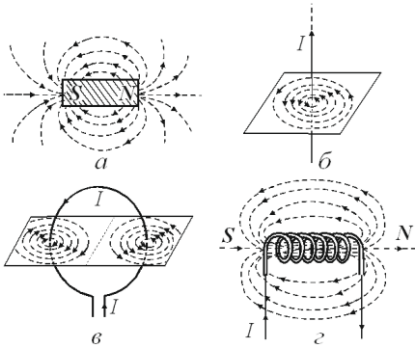


Рис. 5.14. Вигляд ліній магнітної індукції простих магнітних полів

Лінії магнітної індукції охоплюють провідник зі струмом, який створює поле. Поблизу провідника лінії перебувають у площині, що перпендикулярна до провідника (рис. 5.14, *б*). Лінії індукції магнітного поля не можуть обриватися в жодній точці поля, тобто вони завжди замкнені. ***Лінії індукції постійного магніту*** (рис. 5.14, *а*) ***виходять з його північного полюса N і входять у південний полюс S.***

Магнітне поле соленоїда, тобто довгої котушки зі струмом (рис. 5.14, *г*), подібне до магнітного поля штабового магніту. Північний полюс магніту збігається з тим кінцем соленоїда, з якого струм у витках проходить проти годинникової стрілки. Магнітне поле колового струму (рис. 5.14, *в*), який є одним витком соленоїда, подібне на поле дуже короткого штабового магніту, що розміщений у центрі витка так, щоб його вісь була перпендикулярна до площини.

5.4. Застосування закону Біо–Савара–Лапласа до розрахунку магнітних полів

5.4.1. Магнітне поле прямолінійного провідника зі струмом

Знайдемо індукцію магнітного поля, яка створюється прямим проводом зі струмом у точці A (рис. 5.15), що перебуває на відстані r від осі провода.

Поділимо провідник на нескінченно малі елементи Idl (наприклад, елемент BC). За формулою (5.9) знайдемо величину індукції $d\vec{B}$, яку створює елемент $Id\vec{l}$ у точці A:

$$dB = \frac{\mu_0}{4\pi}\frac{Idl\sin\alpha}{AC^2}.$$

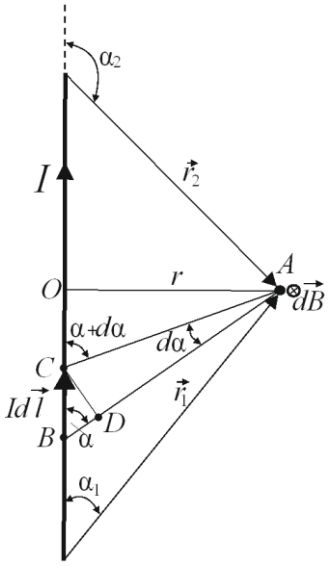


Рис. 5.15. До розрахунку магнітного поля прямолінійного провідника зі струмом

На основі формули (5.8) неважко переконатись, що для заданого напрямку струму I всі елементарні значення індукції магнітного поля напрямлені в точці A в один бік по одній прямій перпендикулярно до площини рисунка від нас, тому результуючу індукцію магнітного поля всіх елементів струму в точці A за принципом суперпозиції (5.10) можна визначити так:

$$dB = \frac{\mu_0}{4\pi}\int_l \frac{Idl\sin\alpha}{AC^2}, \qquad (5.11)$$

де інтегрування ведеться за всією довжиною провода. Щоб виконати інтегрування, перейдемо до однієї змінної α. Будемо вважати, що в трикутнику ABC кут CAB дорівнює $d\alpha$, який є нескінченно малим. Тоді із трикутника ACD можемо знайти, що $CD = AC \cdot d\alpha$. Із трикутника BCD знайдемо, що $BC{=}dl{=}\frac{CD}{\sin\alpha} = \frac{AC d\alpha}{\sin\alpha}$.·Із трикутника ACO знаходимо, що $AC = OA\,/\sin(\alpha + d\alpha) = r/\sin(\alpha + d\alpha)$;

$$\sin(\alpha + d\alpha) = \sin\alpha\cdot\cos d\alpha + \\ +\cos\alpha\cdot\sin d\alpha \cong \sin\alpha.$$

Оскільки $\cos d\alpha \approx 1$, а $\sin d\alpha \approx 0$, тоді

$$BC = dl = \frac{r d\alpha}{\sin^2\alpha}.$$

Тепер вираз (5.11) перепишемо так:

$$B = \frac{\mu_0 I}{4\pi r}\int_{\alpha_1}^{\alpha_2}\sin\alpha\, d\alpha,$$

звідки після інтегрування отримаємо, що

$$B = \frac{\mu_0 I}{4\pi r}(\cos\alpha_1 - \cos\alpha_2), \qquad (5.12)$$

де α_1 та α_2 – кути, що утворюються радіусами-векторами $\vec{r}_1$ і $\vec{r}_2$, проведеними в точку A з початку і кінця провідника відповідно, з напрямком струму. У випадку нескінченно довгого прямого провідника $\alpha_1 \to 0$, $\alpha_2 \to \pi$, і формула (5.12) набуде вигляду

$$B = \frac{\mu_0 I}{2\pi r}, \qquad (5.13)$$

де $2\pi r$ – довжина силової лінії, радіус якої r.

5.4.2. Поле колового витка зі струмом

Нитку зі струмом у вигляді кільця, радіус якого R, показано на рис. 5.16, *а*.

Можемо здогадуватись без будь-яких розрахунків, що магнітне поле повинно мати вигляд, як показано на рис. 5.16, *б*, де показано декілька силових ліній у площині *ZOX*. Поле має обертальну симетрію відносно осі *Z*, а самі силові лінії симетричні відносно площини колового витка *YOX*. Дуже близько до нитки поле нагадує поле довгого прямого провода.

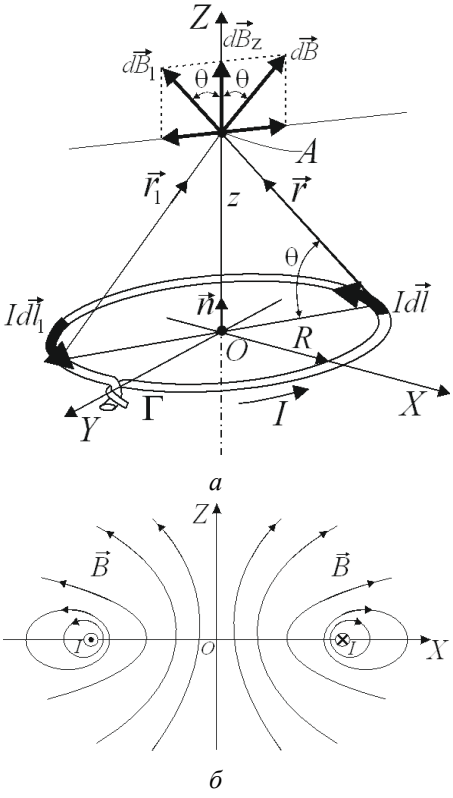


Рис. 5.16. Магнітне поле колового витка зі струмом

На осі *Z* вирахуємо поле, використовуючи співвідношення (5.8). Кожний елемент кільця $Id\vec{l}$ у точці *A* створює вектор $d\vec{B}$, який перпендикулярний до $\vec{r}$ і до $Id\vec{l}$. Розглянемо діаметрально протилежний до $Id\vec{l}$ елемент струму $Id\vec{l}_1$, який у точці *A* створює вектор $d\vec{B}_1$ (рис. 5.16, *а*).

Проєкції векторів $d\vec{B}$ та $d\vec{B}_1$ на будь-яку вісь, яка перпендикулярна до осі *Z*, однакові за величиною і протилежні за напрямком. Це означає, що складатись будуть *Z*-компоненти вектора $d\vec{B}$. Згідно з (5.9)

$$dB = \frac{\mu_0 Idl \sin(\pi/2)}{4\pi r^2},$$

а

$$dB_z = \frac{\mu_0}{4\pi}\frac{Idl}{r^2}\cos\theta = \frac{\mu_0}{4\pi}\frac{IRdl}{r^3},$$

де враховано, що $\cos\theta = R/r$.

Інтегруючи по всьому кільцю, отримуємо

$$B_z = \frac{\mu_0 IR}{4\pi r^3}\oint_l dl = \frac{\mu_0 IR}{4\pi r^3}2\pi R =$$

$$= \frac{\mu_0 IR^2}{2r^3} = \frac{\mu_0 IR^2}{2(R^2+z^2)^{3/2}}. \qquad (5.14)$$

Водночас враховано, що $r = \sqrt{R^2+z^2}$, де z – координата точки *A* на осі *Z*. Якщо врахувати, що магнітний момент колового витка зі струмом $\vec{p}_m = \pi R^2 I\vec{n}$, то формулу (5.14) перепишемо так:

$$\vec{B}_z = \frac{\mu_0 \vec{p}_m}{2\pi(R^2+z^2)^{3/2}}. \qquad (5.15)$$

У центрі кільця $z = 0$, і величина індукції поля згідно з (5.14) буде

$$B_0 = \frac{\mu_0 I}{2R}. \qquad (5.16)$$

5.4.3. Магнітне поле соленоїда

Циліндричну провідну котушку, яку показано на рис. 5.17, *а*, називають соленоїдом. Будемо вважати, що провід розподілений щільно й рівномірно так, що кількість витків на одиницю довжини вздовж циліндра є величиною постійною, яка дорівнює n. Таким чином, насправді струм проходить по спіралі, але якщо витків багато і вони розміщені щільно один до одного, то можна цим знехтувати і розглядати соленоїд як сукупність кілець зі струмом.

Тоді ми можемо використати співвідношення (5.14) як основу для розрахунку індукції в будь-якій точці на осі циліндра z, наприклад у точці A. Розглянемо спочатку внесок кільця зі струмом, яке розміщене між радіусами, що проведені з точки A й утворюють з віссю Z кути від α до $\alpha + d\alpha$ (рис. 5.17, *б*). Ширина цього кільця dz. Із трикутника ABC знайдемо, що $AC = BC / \sin\alpha = R/\sin\alpha$. Тоді $dz = \frac{AC \cdot d\alpha}{\sin\alpha} = \frac{R d\alpha}{\sin^2\alpha}$.·Сила струму, який проходить кільцем, ширина якого dz і містить dN витків, буде

$$dI = IdN = Indz = \frac{nIRd\alpha}{\sin^2\alpha}.$$

Оскільки $r = AC = R/\sin\alpha$, $AB = z = AC\cos\alpha = R\cos\alpha/\sin\alpha$, то за формулою (5.14) отримаємо

$$dB_z = \frac{\mu_0 R^2 dI}{2R^3}\sin^3\alpha = \frac{\mu_0 nI}{2}\sin\alpha\, d\alpha.$$

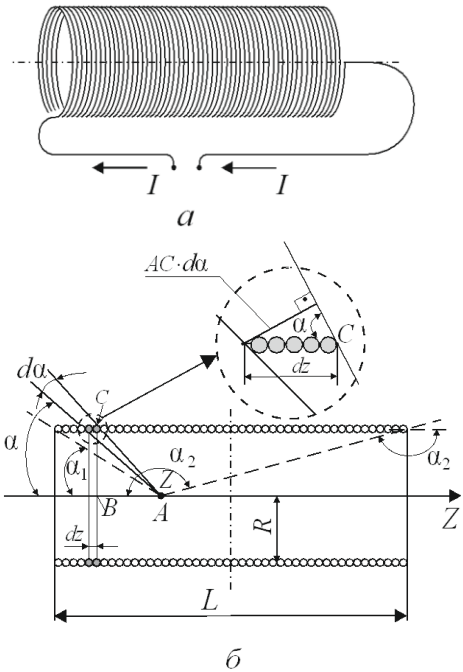


Рис. 5.17. До розрахунку магнітного поля соленоїда

Інтегрування у межах від α_1 до α_2 дає

$$B_z = \frac{\mu_0 nI}{2}\int_{\alpha_1}^{\alpha_2}\sin\alpha\ d\alpha = \frac{\mu_0 nI}{2}(\cos\alpha_1 - \cos\alpha_2), \qquad (5.17)$$

де кути α_1 й α_2 показані на рис. 5.17, *б*.

Скористаємось рівнянням (5.17) і побудуємо графік залежності індукції в

точці A від її положення на осі відносно середини соленоїда (рис. 5.18).

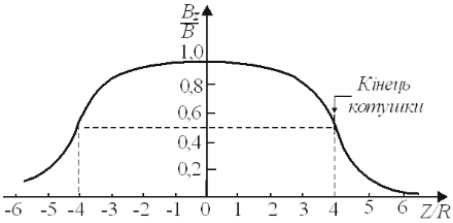


Рис. 5.18. Розподіл магнітного по осі соленоїда

Вважатимемо, що довжина котушки L у чотири рази більша за її діаметр $2R$. По осі ординат відкладено відношення величини індукції B_z на осі соленоїда до величини поля на осі соленоїда нескінченно великої довжини з тією самою кількістю витків n на одиницю довжини і з тією самою силою струму в кожному витку. Для нескінченно довгої котушки $\alpha_1 \to 0, \alpha_2 \to \to \pi$ і формула (5.17) набуде такого вигляду:

$$B = \mu_0 nI. \qquad (5.18)$$

У центрі котушки «чотири до одного» індукція поля досить близька до величини, що визначає формула (5.18), і залишається майже сталою, поки не наблизимось до одного з кінців котушки. На рис. 5.19 зображено лінії індукції магнітного поля всередині та зовні котушки з такими пропорціями. Зауважимо, що деякі силові лінії проходять через обмотку. Циліндричний шар струму є поверхнею розриву для магнітного поля.

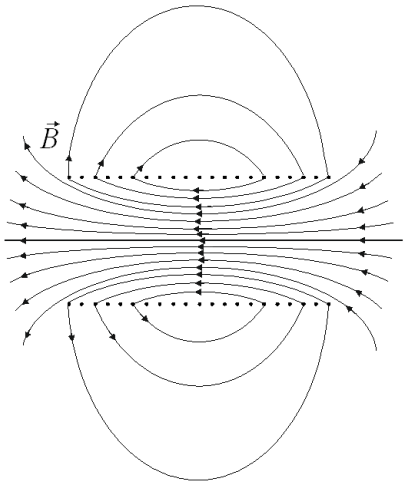


Рис. 5.19. Силові лінії магнітного поля соленоїда. Точками позначено переріз проводів котушки

5.4.4. Магнітне поле електричного заряду, який рухається

За законом Біо–Савара–Лапласа (5.8) магнітна індукція пропорційна силі струму, тобто кількості елементарних зарядів, які проходять за одиницю часу через переріз провідника. Знайдемо індукцію магнітного поля, яка пов'язана з одним елементарним зарядом, який рухається. Для цього в законі (5.8) елемент струму $Id\vec{l}$ (рис. 5.20) перепишемо так:

$$Id\vec{l} = jSd\vec{l} = \vec{j}Sdl = en\vec{v}_{др}dV = e\vec{v}_{др}dN,$$

де $\vec{j}$ – вектор густини струму; n –

концентрація носів струму; $\vec{\upsilon}_{др}$ – швидкість їх направленого руху (дрейфова швидкість); S – переріз проводу; dN – кількість носіїв заряду в об'ємі елемента струму $Id\vec{l}$.

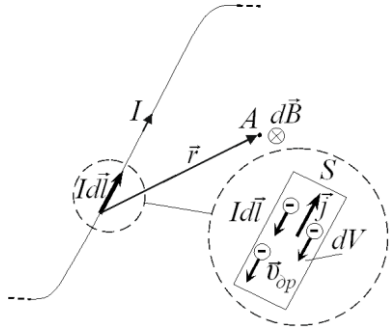


Рис. 5.20. Макроскопічний вигляд елемента струму

Тоді закон (5.8) перепишемо так:

$$d\vec{B} = \frac{\mu_0}{4\pi}\frac{e\vec{\upsilon}_{др}\times\vec{r}}{r^3}dN.$$

Оскільки всі носії заряду (електрони) в середньому мають однакову направлену швидкість руху, то індукція магнітного поля, що створюється одним елементарним зарядом, буде

$$\vec{B}_e = \frac{d\vec{B}}{dN} = \frac{\mu_0}{4\pi}\frac{e\vec{\upsilon}_{др}\times\vec{r}}{r^3},$$

або в загальному випадку

$$d\vec{B} = \frac{\mu_0}{4\pi}\frac{e\vec{\upsilon}_{др}\times\vec{r}}{r^3}dN.$$

Оскільки всі носії заряду (електрони) приблизно мають однакову направлену швидкість руху, то індукція магнітного поля, що створюється одним елементарним зарядом, буде

$$\vec{B}_e = \frac{d\vec{B}}{dN} = \frac{\mu_0}{4\pi}\frac{e\vec{\upsilon}_{др}\times\vec{r}}{r^3},$$

або в загальному випадку

$$\vec{B}_q = \frac{\mu_0}{4\pi}\frac{q\vec{v}\times\vec{r}}{r^3}, \qquad (5.19)$$

де q – заряд, що рухається із швидкістю $\vec{v}$. Із формули (5.19) випливає, що вектор $\vec{B}$ у кожній точці A направлений перпендикулярно до площини, яка проходить через вектори $\vec{v}$ та $\vec{r}$ (рис. 5.21).

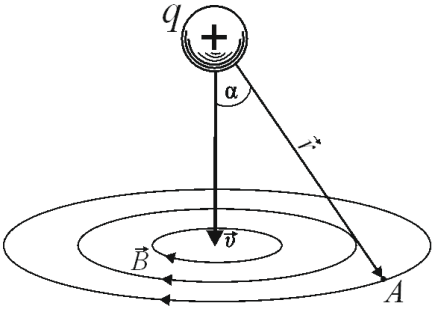


Рис. 5.21. Магнітне поле електричного заряду, який рухається

Напрямок вектора $\vec{B}$ для руху позитивно заряджених частинок найпростіше визначати правилом «лівої руки». Чотири пальці лівої руки розмістити за напрямком вектора $\vec{v}$, у долоню спрямувати вектор $\vec{r}$, і великий палець покаже напрямок вектора $\vec{B}$. Визначення напрямку $\vec{B}$ для негативно

заряджених частинок здійснюють за правилом «правої руки».

5.5. Циркуляція вектора індукції магнітного поля або закон повного струму

В електростатиці було встановлено, що робота під час переміщення одиничного пробного заряду в електричному полі нерухомого заряду не залежить від форми шляху. Інакше кажучи, робота під час переміщення одиничного пробного заряду в електричному полі за довільним замкнутим контуром дорівнює нулю. Такі поля називають потенціальними.

Математично умову потенціальності поля записували у вигляді рівності (1.32). Ця умова вказувала на те, що силові лінії електростатичного поля незамкнені: початок вони беруть на позитивних зарядах і закінчуються на негативних або починаються на нескінченності і прямують у нескінченність.

Дотепер магнітних зарядів у природі не виявлено. Магнітні поля створюються рухомими електричними зарядами, тобто струмами. Спостереження спектрів магнітних полів (рис. 5.14) показують, що лінії індукції магнітного поля замкненні. Поля, які мають замкнені силові лінії, називають ***вихровими***. Це відповідає тому, що у природі немає магнітних зарядів.

Обчислимо циркуляцію вектора $\vec{B}$ по замкнутому контуру Г. Стандартно циркуляція (формула (1.32)) дорівнює інтегралу

$$\oint_{\Gamma} \vec{B} d\vec{l}. \qquad (5.20)$$

Найпростіше обчислити цей інтеграл для поля прямого нескінченно довгого струму. Нехай замкнутий контур Г перебуває у площині, яка перпендикулярна струму (рис. 5.22). Струм перпендикулярний до площини рисунка і направлений на нас.

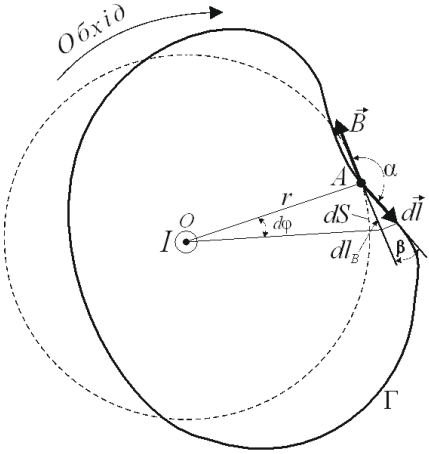


Рис. 5.22. Контур Г охоплює провідник зі струмом

У кожній точці A контура Г вектор $\vec{B}$, направлений по дотичній до кола, радіус якого $OA = r$. Коло перебуває у площині рисунка, а центр O – на осі струму, тому

$$\vec{B}d\vec{l} = Bdl\cos\alpha = Bdl\cos(\pi - \beta) = \\ = -Bdl\cos\beta = -Bdl_B.$$

($dl_B = dl \cdot \cos\beta -$ проєкція елемента $d\vec{l}$ на напрямок вектора $\vec{B}$). Із рис. 5.22 видно, що $dl_B = rd\varphi$, де $r -$ відстань від проводу зі струмом до dl, $d\varphi -$ кут, на який повертається радіальна пряма *OA* під час переміщення уздовж контура на відрізок dl.

Таким чином, з урахуванням (5.13) отримуємо

$$\vec{B}d\vec{l} = -Bdl_B = -Brd\varphi = -\frac{\mu_0 I}{2\pi}d\varphi.$$

Тоді (5.20) запишемо так:

$$\oint_\Gamma \vec{B}\,d\vec{l} = -\frac{\mu_0}{2\pi}\oint_\Gamma d\varphi. \qquad (5.21)$$

Під час обходу по контуру, який охоплює струм, радіальна пряма *OA* (рис. 5.22) увесь час повертається в одному напрямку, тому $\oint_\Gamma d\varphi = 2\pi$. Зовсім інша ситуація, коли контур Г не охоплює струм (рис. 5.23). У цьому випадку радіальна пряма *OA′* повертається за обходом спочатку проти годинникової стрілки (ділянка *1–2*), а потім у протилежному напрямку (ділянка *2–1*, пряма *OA″*), унаслідок чого $\oint_\Gamma d\varphi = 0$. Враховуючи цей результат, (5.21) можемо записати так:

$$\oint_\Gamma \vec{B}\,d\vec{l} = -\mu_0 I, \qquad (5.22)$$

де *I* – струм, що охоплює контур Г.

Знак у виразі (5.22) залежить від напрямку обходу по контуру Г і напрямку струму. Якщо напрямок обходу утворює з напрямком струму лівогвинтову систему (як на рис. 5.22), то $\oint_\Gamma \vec{B}d\vec{l} < 0$, інакше $\oint_\Gamma \vec{B}d\vec{l} > 0$. Знак циркуляції вектора $\vec{B}$ можна врахувати, вважаючи струм *I* алгебраїчною величиною. Струм *I* вважають позитивним, якщо його напрямок пов'язаний з обходом по контуру Г правилом правого свердлика; струм протилежного напрямку буде негативним. Припустимо, що деякий контур Г охоплює декілька проводів зі струмом (рис. 5.24). Згідно з принципом суперпозиції (5.10)

$$\oint_\Gamma \vec{B}d\vec{l} = \oint_\Gamma(\sum_k \vec{B}_k)d\vec{l} = \sum_k \oint_\Gamma \vec{B}_k d\vec{l}.$$

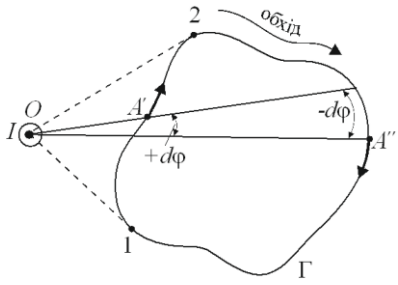


Рис. 5.23. Контур Г не охоплює провідник зі струмом

Кожний з інтегралів у цій сумі дорівнює $\mu_0 I_k$, якщо струм I_k охоплює контур. *Кожний струм враховується стільки разів, скільки його охоплює контур*. Наприклад, на рис. 5.25 струм охоплюється контуром два рази.

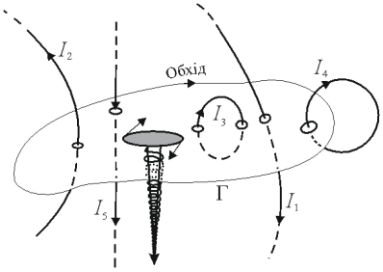


Рис. 5.24. До теореми про циркуляція вектора індукції магнітного поля

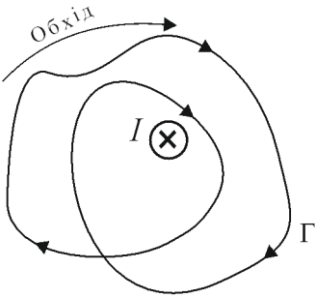


Рис. 5.25. Струм охоплюється контуром два рази

Отже,

$$\oint_{\Gamma} \vec{B}\, d\vec{l} = \mu_0 \sum_k I_k. \quad (5.23)$$

Нагадаємо, що I_k – алгебраїчна величина. Для струмів на рис. 5.24 можемо записати, що

$$\sum_k I_k = I_1 - I_2 - I_4 + I_5.$$

Закон, який виражається рівністю (5.23), називають ***теоремою про циркуляцію вектора індукції магнітного поля, або законом повного струму для струмів провідності: циркуляція вектора $\vec{B}$ за довільним замкнутим контуром дорівнює добутку магнітної сталої на алгебраїчну суму струмів, що охоплює цей контур***.

Отриманий вираз закону ***повного струму справедливий лише*** для постійного магнітного поля у вакуумі.

5.6. Теорема Остроградського–Гауса для магнітного поля

Доведемо, що магнітний потік вектора індукції $\vec{B}$ через довільну замкнуту поверхню S дорівнює нулю. Потік вектора $\vec{B}$ через поверхню S визначають аналогічно потоку вектора напруженості електричного поля $\vec{E}$ (формули (1.10) і (1.11)). Для магнітних полів, які створюють рухомі заряди, справедливий принцип суперпозиції. Крім того, потік геометричної суми декількох векторів $\vec{B}_k$ через будь-яку поверхню дорівнює алгебраїчній сумі потоків цих векторів через ту саму поверхню, тому під час розрахунку магнітного потоку достатньо обмежитись окремим випадком, коли поле $\vec{B}$ створюється окремим точковим зарядом, який рівномірно рухається зі швидкістю $\vec{v}$. Для простоти будемо вважати, що заряд рівномірно рухається перпендикулярно до площини рис. 5.26. Згідно з формулою (5.19) і рис. 5.21 магнітними силовими лініями будуть коаксіальні кола (мають спільний центр),

площини яких паралельні площині рисунка, а центри розміщені на прямій, уздовж якої рухається заряд.

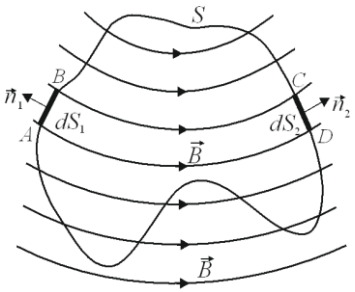


Рис. 5.26. До розрахунку магнітного потоку через замкнуту поверхню

Візьмемо нескінченно тонку кільцеву трубку $ABCD$, яка утворена магнітними силовими лініями. Через осьову симетрію магнітний потік через поперечний переріз трубки буде залишатись сталим за всією її довжиною. Трубка перетне замкнену поверхню S парну кількість разів, наприклад два. Магнітні потоки через ділянки $d\vec{S}_1$ і $d\vec{S}_2$, які трубка вирізає на поверхні S, однакові за величиною, але протилежні за знаком (нормаль $\vec{n}_1$ протилежна до напрямку індукції $\vec{B}$, а нормаль $\vec{n}_2$ збігається з напрямком вектора $\vec{B}$). Сума потоків $d\Phi_1 = \vec{B}_1 d\vec{S}_1$ і $d\Phi_2 = \vec{B}_2 d\vec{S}_2$ через такі ділянки дорівнює нулю. Це справедливо і в тому випадку, коли трубка перетинає поверхню S довільну кількість разів, оскільки ця кількість завжди парна. Утім, весь простір можна розділити на нескінченно тонкі кільцеві трубки, і кожна з них не буде вносити жодного вкладу в магнітний потік через замкнену поверхню S. Сумарний магнітний потік через таку поверхню буде дорівнювати нулю. Таким чином, теорему доведено, тобто

$$\oint_S \vec{B} d\vec{S} = 0. \qquad (5.24)$$

Рівняння (5.24) відображає те, що у природі не має джерел магнітного поля у вигляді магнітних зарядів, подібних до електричних, а силові лінії магнітного поля постійних струмів замкнуті, а отже, їх потік через довільну замкнуту поверхню дорівнює нулю. Таке поле називають ***вихровим*** або ***соленоїдальним***.

5.7. Магнітне поле довгого соленоїда, тороїда та прямого проводу

Соленоїд. Якщо довжина котушки $L \gg 2R$ (див. підрозд. 5.4), де R − радіус перерізу котушки, то таку котушку можна вважати довгим (нескінченно довгим) соленоїдом. Магнітне поле $\vec{B}$ всередині такого соленоїда є однорідним (рис. 5.18; 5.19), його визначають за формулою (5.18). Магнітне поле $\vec{B}'$ зовні соленоїда також будемо вважати однорідним. За допомогою закону повного струму (5.23)

отримуємо формулу (5.18). Спочатку покажемо, що для довгого соленоїда індукція магнітного поля зовні соленоїда дорівнює нулю. Для цього візьмемо площину, перпендикулярну до осі соленоїда (рис. 5.27). Унаслідок замкнутості силових ліній вектора $\vec{B}$ магнітні потоки через внутрішню частину S цієї площини і через зовнішню частину S' мають бути однаковими за величиною і протилежними за знаком, щоб виконувалась теорема (5.24).

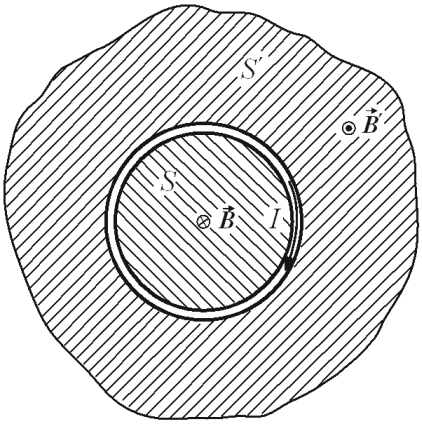


Рис. 5.27. Вид магнітного поля в площині, перпендикулярній до осі соленоїда

Оскільки поля однорідні й вектори $\vec{B}$ і $\vec{B}'$ перпендикулярні до площини, то кожний з потоків дорівнює добутку BS і $B'S'$. Таким чином, отримуємо співвідношення

$$BS = B'S'.$$

Ліва частина цієї рівності – кінцева величина. У правій частині цієї рівності площина S' може бути нескінченно великою, що можливо, коли $B' = 0$. Так ми довели, що зовні нескінченно довгого соленоїда магнітна індукція дорівнює нулю.

Розглянемо соленоїд у площині, яка містить вісь соленоїда (рис. 5.28).

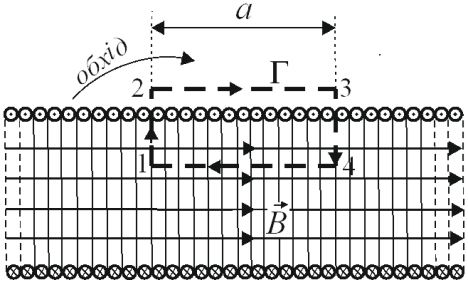


Рис. 5.28. Вид соленоїда у площині, яка містить вісь соленоїда

Для використання закону (5.23) виберемо замкнутий контур Г *1-2-3-4-1*, як показано на рис. 5.28. Ширина контура дорівнює *a,* обхід по контуру здійснюється за годинниковою стрілкою. Тоді циркуляція вектора $\vec{B}$ по такому контуру буде дорівнювати $-Ba$, а алгебраїчна сума струмів, що охоплює цей контур, буде дорівнювати $-naI$, де n – густина намотки соленоїда. Згідно з (5.23) отримуємо рівність

$$-B \cdot a = -\mu_0 nI \cdot a.$$

Звідси

$$B = \mu_0 nI, \qquad (5.18')$$

що збігається з формулою (5.18).

Тороїдом називають кільцеву котушку, яку можна отримати із соленоїда, зігнувши його так, як показано на рис. 5.29. Лінії магнітної індукції поля тороїда повинні мати форму кіл, центри яких перебувають на прямій, що проходить через центр тороїда O і перпендикулярна до площини рисунка. За законом повного струму (5.23) циркуляція вектора $\vec{B}$ уздовж кола, радіус якого r (рис. 5.29), буде становити

$$\oint_\Gamma \vec{B}d\vec{l} = \oint_\Gamma Bdl = B\oint_\Gamma dl = B\cdot 2\pi r. \quad (5.25)$$

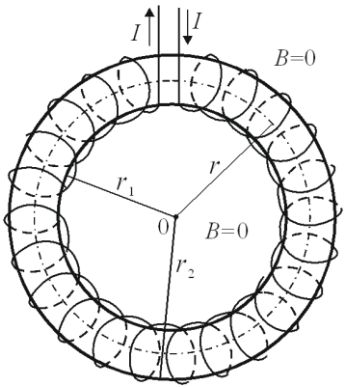


Рис. 5.29. Тороїд

Позначимо кількість витків обмотки тороїда N, а струм в обмотці $-I$. Якщо $r < r_1$, то коло радіусом r не охоплює провідники зі струмом і у формулі (5.23) $\sum_k I_k = 0$. Це означає, що індукція $B = 0$. Якщо $r > r_2$, то коло радіусом r охоплює 2 N провідників зі струмом I. В N із них струм проходить в одному напрямку, а в інших N провідниках – у зворотному напрямку, тому знову $\sum_k I_k = 0$ та $B = 0$. Отже, поза тороїдом магнітного поля немає.

Інакше кажучи, магнітне поле тороїда локалізується всередині об'єму тороїда (об'єму, який охоплюється витками тороїда і для якого $r_1 < r < r_2$). Коло радіусом r, яке перебуває всередині тороїда, охоплює N провідників, струми в яких дорівнюють I й однаково напрямлені, тому $\sum_k I_k = NI$.

Згідно з (5.23) і (5.25) запишемо, що $2\pi rB = \mu_0 NI$. Звідси отримаємо

$$B = \mu_0 \frac{NI}{2\pi r}. \quad (5.26)$$

Магнітна індукція поля всередині тороїда зменшується зі збільшенням відстані від його центра O:

$$B_{max} = \mu_0 \frac{NI}{2\pi r_1},\ B_{min} = \mu_0 \frac{NI}{2\pi r_2}.$$

Індукція на основній лінії тороїда $r = (r_1 + r_2)/2$ становить

$$B = \mu_0 \frac{N}{2\pi r} I = \mu_0 nI, \quad (5.27)$$

де n – кількість витків на одиницю довжини середньої лінії тороїда. Бачимо, що формули для індукції поля для нескінченно довгого соленоїда (5.18), (5.18') і для тороїда (формула (5.27)) однакові, тому дуже часто використовують таке поняття, як тонкий тороїд. Тороїд вважають тонким, коли діаметр його перерізу $2(r_2 - r_1)$ є набагато меншим за довжину середньої лінії $2\pi r$. Для такого тороїда, як і для нескінченно

довгого соленоїда, магнітне поле всередині ϧитків вважають однорідним, а індукцію розраховують за формулою (5.27).

Прямий провід. Розглянемо розрахунок магнітного поля довгого прямого проводу в точці, яка перебуває поза проводом на відстані r від його осі (рис. 5.30). Використаємо для обчислення закон повного струму (5.23). За контур Г зручно взяти силову лінію (має вид кола), радіус якого *r*. Площина кола перпендикулярна до струму й має центр на осі струму. Оскільки $\vec{B}$ і $d\vec{l}$ мають однаковий напрямок і $|\vec{B}|$ є однаковою в кожній точці контура Г, то

$$\oint_\Gamma \vec{B}d\vec{l} = \oint_\Gamma Bdl = B\oint_\Gamma dl = B2\pi r.$$

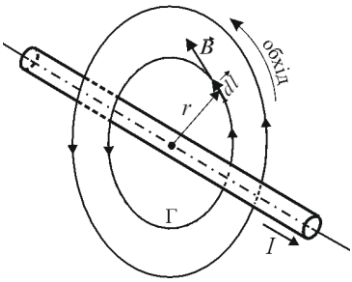


Рис. 5.30. Магнітне поле прямого проводу

Тоді згідно з (5.23) $2\pi rB = \mu_0 I$.

Звідси

$$B = \frac{\mu_0 I}{2\pi r}. \qquad (5.28)$$

Цей результат ми отримали вже раніше (див. формулу (5.13)). Бачимо, що розрахунок за допомогою закону (5.23) є простіший, ніж безпосереднє сумування полів окремих елементів струму. Обчислимо тепер індукцію поля в якій-небудь точці *A* ***всередині*** проводу, яка перебуває на відстані r від осі провода (рис. 5.31, *а*).

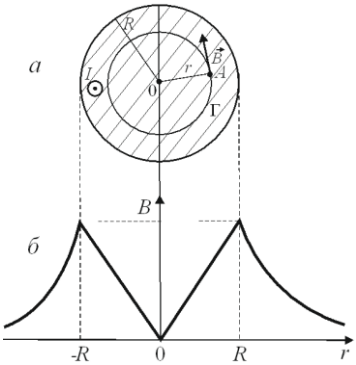


Рис. 5.31. Розподіл магнітного поля в площині перпендикулярній осі проводу

Замкнутий контур Г виберемо знову у вигляді кола, що проходить через точку *A*, а центр перебуває на осі проводу. Тоді за формулою (5.23), як і в попередньому випадку, можемо записати, що

$$\oint_\Gamma \vec{B}d\vec{l} = 2\pi rB = \mu_0 \sum_k I_k = \mu_0 \pi r^2 j,$$

де $j = I/(\pi R^2)$ – густина струму (стала у всіх точках провідника). Звідси отримуємо, що

$$B = \frac{\mu_0 I}{2\pi R^2} r, \qquad (5.29)$$

де I – повна величина сили струму через весь переріз провода; R – радіус провода.

Отже, індукція магнітного поля всередині провода збільшується з відстанню від осі за лінійним законом (формула (5.29)), а в зовнішньому просторі зменшується за гіперболічним законом (формула (5.28)). Цю залежність зображено графічно на рис. 5.31, *б*.

5.8. Контур зі струмом у магнітному полі

З'ясуємо, як поводить себе контур Г зі струмом у магнітному полі. Розглянемо випадок, коли поле однорідне ($\vec{B} = \text{const}$) (рис. 5.32). Нехай контур Г орієнтований так, що позитивна нормаль $\vec{n}$ перпендикулярна до вектора $\vec{B}$ (рис. 5.32).

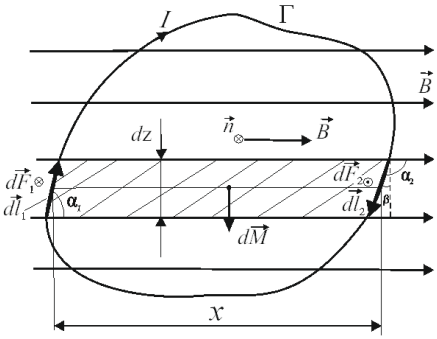


Рис. 5.32. Контур зі струмом в магнітному полі

Поділимо площу контура на вузькі паралельні вектору $\vec{B}$ смужки, ширина яких dz. Зліва і справа смужка обмежена елементами контура $d\vec{l}_1$ і $d\vec{l}_2$. Згідно із законом Ампера (5.7) на елемент струму $Id\vec{l}_1$ буде діяти сила $d\vec{F}_1 = Id\vec{l}_1 \times$ $\times \vec{B}$, а на $Id\vec{l}_2$ – сила $d\vec{F}_2 = Id\vec{l}_2 \times \vec{B}$. Сила $d\vec{F}_1$ направлена за рисунок (від нас). Модуль цієї сили становить $dF_1 =$ $= IBdl_1 \sin\alpha_1 = IBdz$. елемент струму $Id\vec{l}_2$ діє сила $d\vec{F}_2$, яка направлена на нас. Модуль цієї сили становить

$$dF_2 = IBdl_2 \sin\alpha_2 = IBdl_2 \sin(\frac{\pi}{2} + \beta) =$$

$$= IBdl_2 \cos\beta = IBdz.$$

Тому $d\vec{F}_1 = -d\vec{F}_2$.

Отриманий результат означає, що сили, які прикладені до протилежних елементів контура $d\vec{l}_1$ і $d\vec{l}_2$, утворюють пару, обертальний момент якої становить

$$dM = F_1 \cdot x = IBxdz = IBdS,$$

де dS – площа смужки; x – середня лінія, а dz – висота трапеції (смужки).

Якщо обертати свердлик за напрямками дії сил $d\vec{F}_1$ і $d\vec{F}_2$, то він буде рухатись у напрямку вектора $d\vec{M}$. Це означає, що вектор $d\vec{M}$ буде перпендикулярний до векторів $\vec{n}$ і $\vec{B}$, тобто можемо записати, що

$$d\vec{M} = I\vec{n} \times \vec{B}dS.$$

Просумуємо цей вираз по всіх смужках і отримаємо обертальний момент, який діє на контур:

$$\vec{M} = \int_S I\,\vec{n} \times \vec{B}d\vec{S} =$$

$$= I\vec{n} \times \vec{B} \int_S dS = I\vec{n} \times \vec{B}S. \quad (5.30)$$

Під час отримання формули (5.30) враховано, що поле однорідне, тому добуток $\vec{n} \times \vec{B}$ для всіх смужок є однаковим і його можна винести за знак інтеграла. Величина S у виразі (5.30) є площею контура Г.

Вираз (5.30) можна подати ще й у такому вигляді:

$$\vec{M} = IS\vec{n} \times \vec{B}.$$

Якщо врахувати, що $\vec{p}_m = IS\vec{n}$ (див. підрозд. 5.3) є дипольним магнітним моментом контура зі струмом, то

$$\vec{M} = \vec{p}_m \times \vec{B}. \qquad (5.31)$$

Модуль вектора $\vec{M}$ становить

$$M = p_m B \sin\alpha, \qquad (5.32)$$

де α – кут між векторами $\vec{p}_m$ і $\vec{B}$.

Із механіки відомо, що якщо кут α збільшити на $d\alpha$, треба виконати роботу проти сил, які діють на контур у магнітному полі,

$$dA = Md\alpha = p_m B \sin\alpha \, d\alpha. \quad (5.33)$$

Повертаючись у початкове положення, контур може повернути затрачену на його поворот механічну роботу, виконавши її над яким-небудь тілом. Отже, робота (5.33) витрачається на збільшення потенціальної енергії $W_{p_{мех}}$, яку має контур зі струмом у магнітному полі,

$$dW_{p_{мех}} = p_m B \sin\alpha d\alpha.$$

Інтегруючи цей вираз, знаходимо, що

$$W_{p_{мех}} = \int dW_{p_{мех}} = p_m B \int \sin\alpha d\alpha =$$
$$= -p_m B \cos\alpha + \text{const}.$$

Якщо вважати, що const = 0, то

$$W_{p_{мех}} = -p_m B \cos\alpha = -\vec{p}_m \vec{B}. \quad (5.34)$$

Якщо магнітний момент $\vec{p}_m$ контура паралельний до напрямку зовнішнього поля $\vec{B}$ ($\alpha = 0$), то $M = 0$, а $W_{p_{min}} = -p_m B$, і це положення контура відповідає стійкій рівновазі. Якщо $\vec{p}_m$ антипаралельний $\vec{B}$ ($\alpha = \pi$), то $M = 0$, а $W_{p_{max}} = p_m B$, і контур Г перебуває в нестійкій рівновазі.

Розглянемо тепер виток зі струмом у неоднорідному магнітному полі (рис. 5.33), при цьому вважатимемо для простоти, що магнітне поле в центрі витка перпендикулярне до площини витка. Сили $d\vec{F}$, які діють на окремі ділянки витка, будуть, як і раніше, перпендикулярні до струму і до магнітного поля.

Утім тепер через те, що силові лінії вже не паралельні, ці сили утворюватимуть деякий кут із площиною витка. Складові цих сил $d\vec{F}_\tau$, паралельні витку, створять зусилля, які розтягують або стискають виток, а складові $d\vec{F}_n$, які перпендикулярні до площини витка, в сумі дадуть результуючу силу $\vec{F}$, яка прагне переміщати виток у магнітному полі.

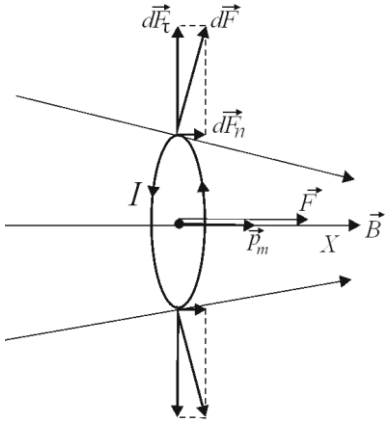


Рис. 5.33. Контур зі струмом у неоднорідному магнітному полі

Застосовуючи правило лівої руки, легко бачити, що коли момент струму $\vec{p}_m$ паралельний магнітному полю (як зображено на рис. 5.33, то виток втягуватиметься в область сильнішого поля, а коли вектор $\vec{p}_m$ антипаралельний полю, то виток виштовхуватиметься і переміщатиметься в область слабшого поля.

Легко зрозуміти, що в загальному випадку неоднорідного поля, індукція $\vec{B}$ якого не перпендикулярна до площини витка, діятимуть і пара сил, яка прагне повернути виток, і сила, яка спричиняє поступальне переміщення витка.

За допомогою виразу (5.34) для енергії контура в магнітному полі легко знайти вираз для сили $\vec{F}$. Якщо орієнтація магнітного моменту $\vec{p}_m$ відносно поля (вектора $\vec{B}$) залишається незмінною ($\alpha = \text{const}$), то $W_{p_{\text{мех}}}$ буде залежати тільки від x (через $B(x)$). Із механіки відомо, що

$$F_x = -\frac{\partial W_{p_{\text{мех}}}}{\partial x} = p_m \frac{\partial B}{\partial x} \cos \alpha.$$

Якщо вважати, що в інших напрямках поле змінюється слабо, то проєкціями сили на інші осі можна знехтувати і вважати, що $F = F_x$, тому

$$F = p_m \frac{\partial B}{\partial x} \cos \alpha. \qquad (5.35)$$

Формула (5.35) повністю відображає поведінку контура зі струмом у магнітному полі. Наприклад: якщо $\vec{p}_m$ і $\vec{B}$ мають один напрямок ($\alpha = 0$), сила $F > 0$, тобто направлена в бік зростання $\vec{B}$ ($\frac{\partial B}{\partial x} > 0$); для випадку $\frac{\partial B}{\partial x} < 0$ знак і напрямок сили зміниться на протилежний, але сила однаково буде втягувати контур в область сильного поля. Якщо $\vec{p}_m$ і $\vec{B}$ антипаралельні $\alpha = \pi$, сила негативна, тобто направлена в бік зменшення $\vec{B}$.

Підсумовуючи сказане, видно, що й сили, які діють на контур зі струмом у магнітному полі, і магнітне поле, яке створюється контурами, визначаються їх магнітними моментами.

Виникнення пари сил, яка діє на рамку зі струмом у магнітному полі, використовують для побудови дуже поширеного типу електровимірювальних

приладів – ***магнітоелектричних приладів***. Вони мають легку рамку із дротом, закріплену на наріжному підвісі тієї чи іншої конструкції, здатну повертатися у магнітному полі. Зрозуміло, що повертання рамки зі струмом у магнітному полі здебільшого використовують у побудові електродвигунів.

5.9. Робота під час переміщення провідника зі струмом у магнітному полі

Розглянемо контур зі струмом, який утворений нерухомими проводами і перемичкою $KL = l$, яка може ковзати по них (рис. 5.34).

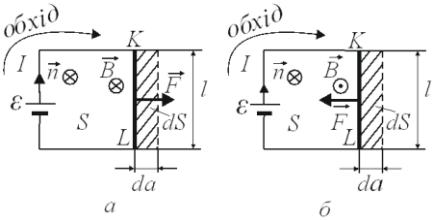


Рис. 5.34. Робота під час переміщення провідника зі струмом у магнітному полі

Нехай контур перебуває у зовнішньому магнітному полі, яке будемо вважати однорідним і перпендикулярним до площини контура. На перемичку KL зі струмом у магнітному полі буде діяти сила Ампера (5.7), яка для нашого випадку, показаного на рис. 5.34, становить

$$F = ILB.$$

У разі переміщення перемички вправо на відстань da ця сила виконує позитивну роботу:

$$dA = Fda = IBlda = IBdS, \quad (5.36)$$

де $dS = lda -$ заштрихована площа на рис. 5.34, *а*.

З'ясуємо, як змінюється під час переміщення перемички KL потік магнітної індукції через площу контура. У загальному випадку потік вектора індукції через площу контура зі струмом становить

$$\Phi = \int \vec{B}\,\vec{n}dS,$$

де $\vec{n}$ – позитивна нормаль, тобто нормаль, яка утворює з напрямком струму в контурі (з обходом по контуру) правогвинтову систему (див. підрозд. 5.3). Тоді у випадку, показаному на рис. 5.34, *а*, потік буде позитивним і дорівнюватиме BS ($S -$ площа контура). У разі переміщення перемички вправо площа контура отримає позитивний приріст dS. У результаті потік також отримає позитивний приріст $d\Phi = BdS$, тому вираз (5.36) можна подати у вигляді

$$dA = Id\Phi. \quad (5.37)$$

За умови напрямку магнітного поля на читача (рис. 5.34, *б*), яке діє на перемичку, сила $\vec{F}$ буде направлена вліво, тому у разі переміщення перемички вправо на da магнітна сила виконує негативну роботу:

$$dA = -IBlda = -IBdS. \quad (5.38)$$

У цьому випадку потік через контур становить $\vec{B}\vec{n}S = BS\cos\pi = -BS$. У разі зміни площі контура на dS потік отримує приріст $d\Phi = \vec{B}\vec{n}dS = BdS\cos\pi = -BdS$. Значить, вираз (5.38) також можна записати у вигляді (5.37).

Величину $d\Phi$ у виразі (5.37) можна трактувати як потік через площу, яку описує перемичка під час руху, тому можна сказати, що робота, яку виконує магнітна сила над ділянкою контура зі струмом, дорівнює добутку сили струму на величину магнітного потоку через поверхню, яку описує ця ділянка під час руху.

Формули (5.36) і (5.38) можна об'єднати в один векторний вираз. Розглянемо перемичку як вектор $\vec{l}$, що має напрямок струму (рис. 5.35). Незалежно від напрямку вектора $\vec{B}$ (від нас або на нас) силу Ампера, яка діє на перемичку, можна подати у вигляді $\vec{F} = I\vec{l}\times\vec{B}$.

Під час переміщення перемички на $d\vec{a}$ сила виконує роботу

$$dA = \vec{F}d\vec{a} = I\vec{l}\times\vec{B}d\vec{a}.$$

Виконаємо у змішаному добутку векторів циклічну перестановку множників. У результаті отримаємо, що

$$dA = I\vec{B}d\vec{a}\times\vec{l}. \qquad (5.39)$$

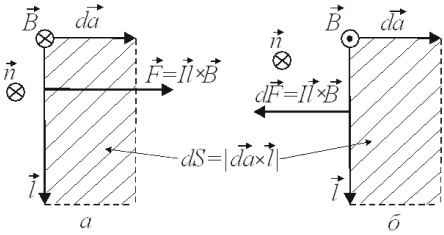


Рис. 5.35. До розрахунку роботи при переміщенні провідника зі струмом в магнітному полі

Із рис. 5.35 видно, що векторний добуток $d\vec{a}\times\vec{l}$ за величиною дорівнює ділянці dS, яку описує перемичка під час її руху, і має напрямок позитивної нормалі $\vec{n}$, тобто $d\vec{a}\times\vec{l} = \vec{n}dS$. Отже, (5.39) перепишемо так:

$$dA = I\vec{B}\vec{n}dS. \qquad (5.40)$$

У випадку, який показано на рис. 5.35, *а*, $\vec{B}\vec{n} = B$, і ми отримуємо формулу (5.36). У випадку, показаному на рис. 5.35, *б*, $\vec{B}\vec{n} = -B$, і ми отримуємо формулу (5.38).

Вираз $\vec{B}\vec{n}dS$ визначає приріст магнітного потоку через контур, обумовлений переміщенням перемички. Таким чином формулу (5.40) можна записати у вигляді (5.37). Утім, формула (5.40) має перевагу перед (5.37), оскільки із неї «автоматично» отримується знак $d\Phi$, а значить, і знак dA.

Розглянемо жорсткий довільної форми контур Г зі струмом, що перебуває в довільному магнітному полі. Знайдемо

роботу, яка виконується під час довільного нескінченно малого переміщення контура. Нехай елемент контура $d\vec{l}$ переміщається на $d\vec{a}$ (рис. 5.36), при цьому над ним магнітна сила виконує таку роботу:

$$dA_{ел} = Id\vec{l} \times \vec{B}d\vec{a}, \qquad (5.41)$$

де $\vec{B}$ – магнітна індукція в тому місці, де перебуває елемент контура $d\vec{l}$.

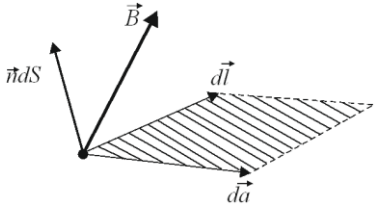


Рис. 5.36. До розрахунку роботи, яка виконується під час довільного нескінченно малого переміщення контура

Здійснивши в (5.41) циклічну перестановку множників, отримаємо

$$dA_{ел} = I\vec{B}d\vec{a} \times d\vec{l}. \qquad (5.42)$$

Векторний добуток $d\vec{a} \times d\vec{l}$ за модулем дорівнює площі паралелограма, побудованого на векторах $d\vec{a}$ і $d\vec{l}$, тобто площі dS, яку описує елемент під час його переміщення на $d\vec{a}$. Тому

$$\vec{B}d\vec{a} \times d\vec{l} = \vec{B}\vec{n}dS = d\Phi_{ел}, \qquad (5.43)$$

де $\vec{n}$ – позитивна нормаль до ділянки dS; $d\Phi_{ел}$ – приріст магнітного потоку через контур, обумовлений переміщенням елемента контура $d\vec{l}$. Взявши до уваги рівність (5.43), запишемо (5.42) у вигляді

$$dA_{ел} = I\, d\Phi_{ел}. \qquad (5.44)$$

Просумувавши вираз (5.44) за всіма елементами контура, отримаємо вираз для роботи магнітних сил під час довільного нескінченно малого переміщення контура Г:

$$dA = \oint_{\Gamma} dA_{ел} = \oint_{\Gamma} Id\Phi_{ел} = I\oint_{\Gamma} d\Phi_{ел} = Id\Phi, \qquad (5.45)$$

де $d\Phi$ – повний приріст потоку через весь контур Г під час його довільного нескінченно малого переміщення.

Щоб знайти роботу, яка виконується під час кінцевого довільного переміщення контура з початкового положення в кінцеве, проінтегруємо вираз (5.45):

$$A_{12} = \int_{\Phi_1}^{\Phi_2} dA = I\int_{\Phi_1}^{\Phi_2} d\Phi = I(\Phi_2 - \Phi_1), \qquad (5.46)$$

де Φ_1 і Φ_2 – значення магнітного потоку через контур Г у початковому і кінцевому положеннях. Таким чином, робота, яку виконують магнітні сили над контуром, дорівнює добутку сили струму в контурі на приріст магнітного потоку через контур.

Зауважимо, що робота (5.46) виконується не за рахунок енергії зовнішнього магнітного поля, а за рахунок джерела, яке підтримує незмінним струм у контурі.

5.10. Дія електричного і магнітного полів на електричний заряд, який рухається

5.10.1. Сила Лоренца

Виникнення сили Ампера (5.7), що діє на елемент провідника зі струмом у магнітному полі, можна пояснити так. Під час проходження струму носії заряду у провіднику рухаються напрямлено, тому магнітне поле відхиляє їх в один бік. Водночас вони стикаються з кристалічною ґраткою металу і передають їй певний імпульс, якого набули під дією магнітного поля. Макроскопічним результатом елементарних процесів зіткнення окремих носіїв заряду з кристалічною граткою провідника є виникнення сили Ампера.

Магнітне поле діє на вільні електрони в металі й за відсутності електричною струму у провіднику, та електрони в цьому випадку рухаються тільки хаотично, і сумарний імпульс, який вони надають кристалічній гратці провідника, дорівнює нулю, а провідник залишається нерухомим.

Для обчислення сили, що діє на рухомий заряд у магнітному полі, використаємо формулу (5.7) для випадку магнітного поля з індукцією $\vec{B}$:

$$d\vec{F} = Id\vec{l} \times \vec{B}.$$

Елемент струму, як і в підрозд. 5.4, можна подати так:

$$Id\vec{l} = e\vec{v}_{др}dN.$$

Оскільки електрони рухаються направлено, створюючи струм, то сила, яка припадає на один електрон, становить

$$\vec{F}_л = \frac{d\vec{F}}{dN} = e\vec{v}_{др} \times \vec{B}.$$

У загальному випадку

$$\vec{F}_л = \frac{d\vec{F}}{dN} = e\vec{v} \times \vec{B}. \qquad (5.47)$$

Напрямок сили Лоренца визначають за правилом векторного добутку або правилом лівої руки: якщо долоню лівої руки розмістити так, щоб у неї входив вектор $\vec{B}$, чотири витягнуті пальці спрямувати вздовж вектора швидкості $\vec{v}$ руху позитивних зарядів, то відігнутий на 90° великий палець покаже напрямок сили, що діє на позитивний заряд. На негативний заряд сила діє у протилежному напрямку (рис. 5.37).

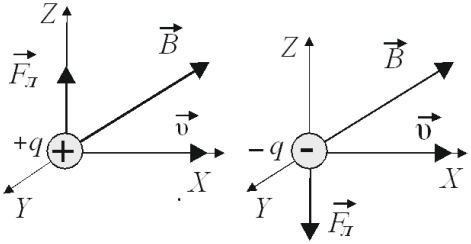


Рис. 5.37. Дія магнітного поля на заряди, які рухаються

Отже, магнітне поле не діє на електричні заряди, які не рухаються.

Сила Лоренца завжди перпендикулярна до швидкості руху зарядженої частинки, тому вона змінює

лише напрямок цієї швидкості, не змінюючи її модуля. Отже, сила Лоренца роботи не виконує, і кінетична енергія частинки під час руху в магнітному полі не змінюється.

Якщо на рухомий електричний заряд, крім магнітного поля з індукцією $\vec{B}$, діє також електричне поле з напруженістю $\vec{E}$, то результуюча сила $\vec{F}$, яка прикладена до заряду, становить

$$\vec{F} = q\vec{E} + q\vec{v} \times \vec{B}. \qquad (5.48)$$

Це формула Лоренца.

Якщо заряджена частинка рухається у магнітному полі зі швидкістю $\vec{v}$ уздовж ліній магнітної індукції або у протилежний бік до напрямку магнітної індукції, то $\alpha = 0$, або $\alpha = \pi$. У такому разі $F_л = 0$, магнітне поле на частинку не діє, і вона рухається рівномірно й прямолінійно.

Якщо заряджена частинка рухається у магнітному полі зі швидкістю $\vec{v}$ перпендикулярно до вектора $\vec{B}$, то сила Лоренца є сталою за модулем і нормальною до траєкторії частинки. Частинка рухатиметься по колу, оскільки сила Лоренца за другим законом Ньютона буде створювати доцентрове прискорення. Отже, $qvB = \frac{mv^2}{r}$. Звідси

$$r = \frac{m}{q}\frac{v}{B},$$

де r – радіус кола.

Використавши зв'язок $v = \omega r$, знайдемо циклічну частоту ω та період обертання T частинки навколо ліній індукції в магнітному полі:

$$\omega = \frac{v}{r} = \frac{q}{m}B,\ T = \frac{2\pi r}{v} = \frac{2\pi m}{Bq}. \qquad (5.49)$$

Період обертання частинки в однорідному магнітному полі не залежить від її швидкості (за $v << c$, де c – швидкість світла у вакуумі). На цьому ґрунтується дія циклічних прискорювачів заряджених частинок. Якщо швидкість зарядженої частинки $\vec{v}$ напрямлена під кутом α до вектора $\vec{B}$ (рис. 5.38), то її рух можна подати у вигляді суперпозиції:

1) рівномірного прямолінійного руху вздовж поля зі швидкістю $|v| = v\cos\alpha$;

2) рівномірного руху зі швидкістю $v_\perp = v\sin\alpha$ уздовж кола, площина якого перпендикулярна до поля $\vec{B}$. Радіус кола $r = \frac{mv\sin\alpha}{qB}$.

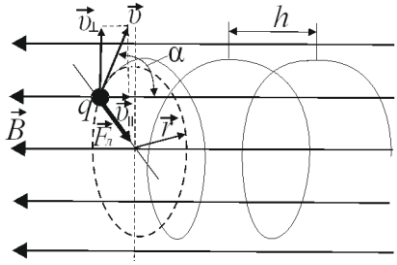


Рис. 5.38. Рух зарядженої частинки в магнітному полі

У результаті складання обох рухів виникає рух уздовж спіралі, вісь якої

паралельна до магнітного поля. Крок гвинтової лінії

$$h = \upsilon_{||}T = \upsilon T\cos\alpha = \frac{2\pi m\upsilon\cos\alpha}{qB}.$$

Напрямок, в якому закручується спіраль, залежить від знака заряду частинки.

5.10.2. Ефект Холла

У 1879 р. Е. Холл здійснив такий експеримент. Він пропускав електричний струм I через золоту пластинку у вигляді паралелепіпеда і вимірював різницю потенціалів $\Delta\varphi$ між точками C і D на верхній і нижній гранях (рис. 5.39).

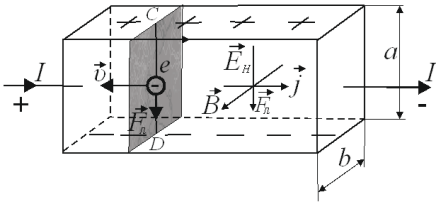


Рис. 5.39. Ефект Холла

Ці точки перебувають у одному й тому самому поперечному перерізі пластинки, тому виявилось, що $\Delta\varphi = 0$. Коли пластинку зі струмом Холл помістив в однорідне магнітне поле, лінії магнітної індукції якого перпендикулярні до бічних граней пластинки, то було вставлено, що різниця потенціалів $\Delta\varphi_{\text{H}} = \varphi_{\text{C}} - \varphi_{\text{D}} \neq 0$, і вона прямо пропорційна силі струму I, магнітній індукції поля B й обернено пропорційна ширині пластинки, тобто

$$\Delta\varphi_{\text{H}} = \varphi_{\text{C}} - \varphi_{\text{D}} = R_H\frac{IB}{b}, \quad (5.50)$$

де R_H – коефіцієнт пропорційності, названий ***сталою*** Холла. Явище, яке експериментально виявив Холл, отримало назву ***ефекту Холла***.

Наступні дослідження показали, що ефект Холла спостерігається в усіх провідниках і напівпровідниках. Зміна напрямку струму або напрямку магнітного поля на протилежний викликає зміну знака різниці потенціалів $\Delta\varphi$.

Розглянемо, як можна пояснити ефект Холла. Помістимо металеву пластинку зі струмом густиною $\vec{j}$ у магнітне поле з індукцію $\vec{B}$, лінії індукції якого перпендикулярні до $\vec{j}$ (рис. 5.39). На електрони діє сила Лоренца, яка направлена вниз і дорівнює $F_л = Be\upsilon$, тому на нижній грані пластинки збиратиметься нескомпенсований негативний заряд, і вона заряджатиметься негативно, а на верхній грані пластинки виникатиме нестача негативних зарядів, і вона заряджатиметься позитивно. Унаслідок цього між гранями пластинки виникає додаткове поперечне електричне поле, направлене згори вниз. Розглянемо момент динамічної рівноваги, коли сили $F_л$ і F_H, з якими діють на електрони

магнітне поле та холлівське електричне поле, стануть рівними (ця рівність настає вже через 10^{-12} с після замикання кола):

$$F_л = F_H, \quad Bev = eE_H.$$

Звідси

$$E_H = vB.$$

Невідому швидкість v напрямленого руху виразимо через густину струму j:

$$j = env, v = j/(e\,n),$$

де n – концентрація електронів у металевій пластинці; v – їх швидкість направленого руху.

Тоді

$$E_H = \frac{Bj}{ne}.$$

Помножимо ліву частину виразу на ab, а праву на $S = ab$:

$$E_H ab = \frac{Bj}{ne} S.$$

Оскільки $E_H \cdot a = \Delta\varphi_H$, а $jS = I$, то

$$\Delta\varphi_H = \frac{1}{en}\frac{BI}{b}. \qquad (5.51)$$

Порівнюючи вираз (5.51) з виразом (5.50), який отриманий на основі експерименту, отримуємо, що стала Холла R_H обернено пропорційна до добутку заряду електрона на їх концентрацію n:

$$R_H = \frac{1}{ne}. \qquad (5.52)$$

За виміряними значеннями сталої Холла можна:

1) *визначити концентрацію носіїв струму*, якщо характер провідності і заряд q носіїв струму відомі, а саме

$$n = 1/(qR_H).$$

Так, для одновалентних металів виявилося, що концентрація електронів провідності збігається з концентрацією атомів. Оскільки питому електропровідність речовини обчислюють за формулою (див. підрозд. 4.2) $\sigma = enu$, де u – рухливість носіїв заряду, то $\sigma = u/R_H$

та $u = R_H \cdot \sigma$, тобто добуток $R_H \cdot \sigma$, визначає рухливість носіїв заряду;

2) *зробити висновок про природу провідності,* оскільки знак сталої Холла збігається зі знаком заряду q.

5.10.3. Рух електронів у схрещених електричному і магнітному полях. Визначення питомого заряду електрона методом магнетрона

Вивчення руху різних заряджених частинок в електричному і магнітному полях дає можливість визначити таку важливу характеристику, як відношення їх заряду до маси q/m, а отже, отримати дані про природу цих частинок і їх властивості.

Діючи на потоки електронів та іонів електричними і магнітними полями, можна керувати цими потоками, тобто змінювати їхню інтенсивність і напрямок руху; це лежить в основі роботи різних

електронних приладів (електронно-променевих трубок, електронних мікроскопів, прискорювачів заряджених частинок, телевізійних трубок, фотоелектричних помножувачів, електронно-оптичних перетворювачів і под.).

Велике практичне значення має рух електронів у магнетроні. Магнетрон – це багато резонансний прилад для генерації електромагнітних коливань надвисоких частот. В основу роботи магнетрона покладено взаємодію електронів, що рухаються у магнітному полі по криволінійних траєкторіях із збуджуючим електричним полем. Анод магнетрона – масивний порожнистий циліндр, у внутрішній частині якого вирізані об'ємні резонатори із щілинами, які виходять на внутрішню поверхню циліндра (рис. 5.40). По осі циліндра розміщений катод *К*. Під дією магнітного поля $\vec{B}$, яке направлене вздовж осі циліндра, траєкторії електронів, що вилітають з катода, викривляються.

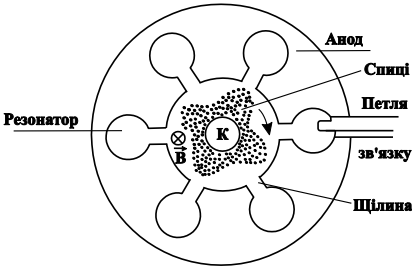


Рис. 5.40. Принципова схема магнетрона

Коли в резонаторах збуджуються коливання, то біля щілин виникає змінне електричне поле. Під дією поля надвисокої частоти (НВЧ) і схрещених статичних електричного і магнітного полів електрони, що вилітають з катода, утворюють згустки – «спиці» (точки на рис. 5.40). Електрони у згустках під час взаємодії з гальмівним полем НВЧ віддають полю потенціальну енергію і наближаються до анода. На анод вони потрапляють, віддавши електромагнітному полю майже всю енергію, що обумовлює високий (до 90 %) ККД. Магнетрони здатні генерувати коливання аж до міліметрового діапазону. Зовнішній вигляд магнетрона побутової мікрохвильової печі показаний на рис. 5.41.

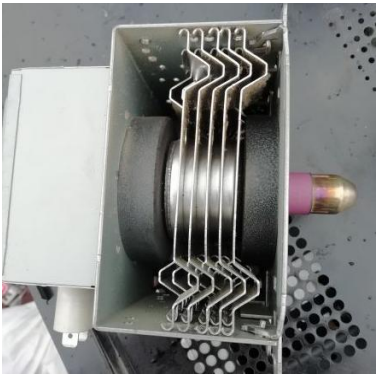


Рис. 5.41. Зовнішній вигляд магнетрона побутової мікрохвильової печі

Як магнетрон використовують

найпростішу його модель – двохелектродну радіолампу з холодним анодом і катодом, який підігрівають (рис. 5.42). Будова лампи показана на рис. 5.43. Лампа перебуває в однорідному магнітному полі довгого соленоїда. Вектор індукції магнітного поля соленоїда направлений вздовж осі катода й анода. Зазвичай ниткоподібний катод – це джерело термоелектронів, оточений коаксіальним циліндричним анодом (рис. 5.43).

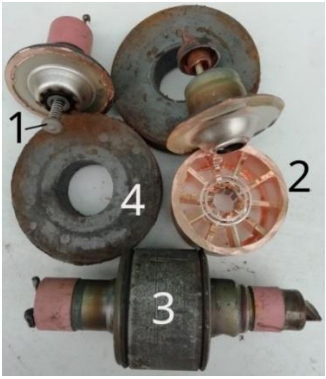


Рис. 5.42. Електронна лампа, яку використовують у магнетроні, що показаний на рис. 5.41: 1 – катод; 2 – анод; 3 – загальний вигляд лампи; 4 – магніти

Розглянемо характер руху електронів у лампі. Якщо струму I_c у соленоїді немає ($\vec{B} = 0$), електрони рухаються від катода до анода в радіальному напрямку вздовж силових ліній електричного поля між катодом і анодом (рис. 5.44, *а*). За заданого значення анодної напруги між анодом і катодом встановлюється анодний струм I_a = const. Пропустимо струм, сила якого I_c, витками соленоїда, і створимо, таким чином, однорідне магнітне поле вздовж катода (рис. 5.44). На рухомі електрони діятиме сила Лоренца $\vec{F_л} = e\vec{v} \times \vec{B}$, яка перпендикулярна до напрямку швидкості руху електронів. Траєкторія електронів викривляється (рис. 5.44, *б*), але всі вони будуть потрапляти на анод, а отже, анодний струм I_a не змінюється. З ростом I_c збільшується $\vec{B}$, і кривизна траєкторії руху електронів також зростатиме. За певного, критичного, значення індукції поля електрони почнуть рухатись по замкнутій траєкторії (рис. 5.44, *в*) і не будуть потрапляти на анод. Анодний струм I_a різко спадає.

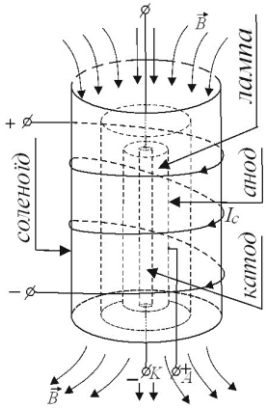


Рис. 5.43. Схема розташування електронної лампи в магнетроні

Критичні значення сили струму соленоїда $I_{кр}$, за яких різко зменшується

анодний струм, визначають за скидовими характеристиками лампи, під якими розуміють графіки залежності анодного струму I_a від струму в соленоїді I_c за заданих значень анодної напруги U_a.

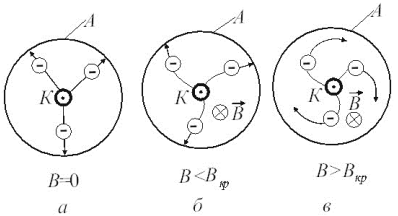


Рис. 5.44. Траєкторія руху електронів у лампі за різних значень індукції магнітного поля B

Після включення постійного струму в соленоїді його магнітне поле починає діяти на електрони із силою $F_л$, яка завжди направлена перпендикулярно до швидкості $\vec{v}$ їхнього руху. Під дією цієї сили рух електронів набуває більш складного характеру і їхні траєкторії стають криволінійними; вони починаються на катоді й закінчуються на аноді лампи (рис. 5.44, *а*), випадок $B < B_{кр}$. За деякої індукції поля $B = B_{кр}$, яку називають *критичною*, траєкторії електронів перетворюються у криволінійні, які торкаються поверхні анода (рис. 5.44, *в*). Під час індукції $B > B_{кр}$ електрони рухаються траєкторіями, які перебувають між анодом і катодом. Це означає, що за певного співвідношення між швидкостями електронів v і значенням індукції B електрони перестануть досягати анода й анодний струм I_a в лампі має дорівнювати нулеві. У цьому проявляється ефект дії магнітного поля на електрони, які рухаються в електричному полі, що створене різницею потенціалів U_a між анодом і катодом лампи.

Якби всі електрони мали однакову початкову швидкість v_0 під час емісії з катода, то величина $B_{кр}$ для них була б також однаковою й у разі досягнення сили струму в соленоїді, яка дорівнює $I_{кр}$, сила анодного струму I_a різко упала б до нуля, як показано на рис. 5.45, *а*.

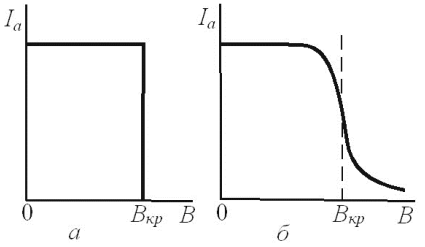


Рис. 5.45. Залежність анодного струму лампи магнетрона від індукції магнітного поля в соленоїді

Утім, не існує різкого однозначного спаду анодного струму I_a до нуля у разі досягнення величини магнітного поля $B > B_{кр}$ В області значень індукції, яка близька до $B_{кр}$, відбувається монотонний спад струму, як показано на рис. 5.45, *б*. Такий спад струму пояснюється здебільшого тим, що електрони вилітають з катода з різними

швидкостями, і тому критичні умови для різних електронів досягаються за різних значень індукції B магнітного поля.

Зауважимо, що на залежність анодного струму I_a від індукції магнітного поля B також впливають некоаксіальність між анодом і катодом, нахил вектора індукції магнітного поля відносно катода, просторовий заряд між анодом і катодом та ін., а в наших умовах ще й конструкція електронної лампи, яка не зовсім придатна для таких експериментів.

Експериментальне визначення e/m_e цим методом полягає у тому, що в лампі магнетрона встановлюється постійний струм розжарення катода і постійна різниця потенціалів між анодом і катодом, тобто створюються умови для стійкості анодного струму. Потім, змінюючи від нуля значення сили струму в соленоїді I_c і таким чином збільшуючи індукцію магнітного поля, отримують суттєве зменшення анодного струму I_a в лампі.

Для електрона, який вилетів з катода і має радіальну складову швидкості, яка дорівнює нулю, а також знаючи геометрію лампи, тобто радіус анода r_a і радіус катода r_k, величину U_a, а також $B_{\text{кр}}$, за формулою (5.35) визначають величину e/m_e:

$$\frac{e}{m_e} = \frac{8U_a}{B_{\text{кр}}^2 r_a^2 \left(1-\frac{r_k^2}{r_a^2}\right)^2}, \qquad (5.53)$$

де e – заряд електрона; m_e – маса спокою електрона; r_a – радіус анода; r_k – радіус катода; U_a – напруга між анодом і катодом (анодна напруга).

Характерною особливістю цього методу є те, що він виключає необхідність вивчення траєкторії руху електрона в магнітному полі. Навпаки, зміною значення величини B (за постійного U_a) досягається наперед задана траєкторія руху електронів, за якої вони не можуть потрапити на анод лампи всупереч дії електричного поля. Таким чином, дослід зводиться до зняття залежності анодного струму I_a (за $U_a =$ $=$ const) від індукції магнітного поля соленоїда B (або струму соленоїда I_c).

Виведення формули (5.53) для розрахунку e/m_e. Розглянемо рух електрона від катода до анода в лампі, яку показано на рис. 5.43, а її переріз зображено на рис. 5.46. Магнітне поле $\vec{B}$ направлене вздовж осі лампи і перпендикулярне електричному полю $\vec{E}$, яке направлене від анода до катода. У загальному випадку згідно з другим законом Ньютона рівняння руху електрона від катода до анода в такій лампі запишемо так:

$$m_e \frac{d\vec{v}}{dt} = e\vec{E} + e\vec{v} \times \vec{B}, \qquad (5.54)$$

де $e\vec{E}$ – сила з боку електричного поля, а $e\vec{v} \times \vec{B}$ – сила Лоренца, яка діє на рухомий електричний заряд з боку магнітного поля (рис. 5.46) (вона пропорційна індукції $\vec{B}$ і швидкості $\vec{v}$).

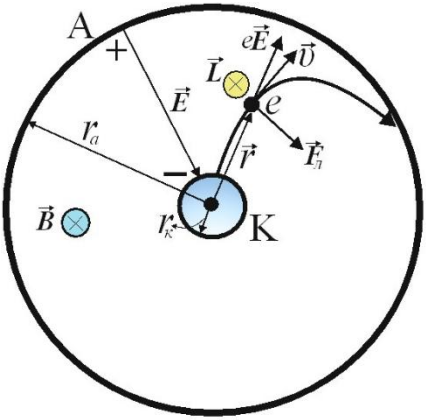


Рис. 5.46. Рух електрона в лампі від катода до анода вмагнітному полі

Якщо рівняння (5.54) векторно помножити на радіус-вектор $\vec{r}$, який визначає положення електрона в лампі, то (5.54) можна переписати так:

$$\vec{r} \times m_e \frac{d\vec{v}}{dt} = \vec{r} \times e\vec{E} + \vec{r} \times e\vec{v} \times \vec{B}. \qquad (5.55)$$

Це вже буде рівняння моментів сил. Врахуємо, що вектор $\vec{E}$ направлений по лінії вектора $\vec{r}$, тоді $\vec{r} \times e\vec{E} = 0$. Скористаємось формулою векторної алгебри, що $\vec{a} \times \vec{b} \times \vec{c} = \vec{b}(\vec{a}\vec{c}) - \vec{c}(\vec{a}\vec{b})$. Тоді вираз для момента сили Лоренца можна переписати так:

$$\vec{r} \times m_e \frac{d\vec{v}}{dt} = e\vec{v}(\vec{r}\vec{B}) - \vec{B}(\vec{r} \cdot e\vec{v}). \qquad (5.56)$$

Оскільки поле $\vec{B}$ перпендикулярне $\vec{r}$, то $\vec{r}\vec{B} = 0$. Ліву частину виразу (5.55) перепишемо так:

$$\vec{r} \times m_e \frac{d\vec{v}}{dt} = \vec{r} \times \frac{d(m_e\vec{v})}{dt} = \frac{d\vec{L}}{dt},$$

де $\vec{L} = \vec{r} \times m_e\vec{v}$ – момент кількості руху електрона відносно центральної осі катода. Справді,

$$\frac{d\vec{L}}{dt} = \frac{d\vec{r} \times m_e\vec{v}}{dt} = \frac{d\vec{r}}{dt} \times m_e\vec{v} + \vec{r} \times m_e \frac{d\vec{v}}{dt}. \qquad (5.57)$$

Перший доданок у виразі (5.57) дорівнює нулю, тому що

$$\frac{d\vec{r}}{dt} \times m_e\vec{v} = \vec{v} \times m_e\vec{v} = 0.$$

Тоді співвідношення (5.55) можна переписати так:

$$\frac{d\vec{L}}{dt} = -e\vec{B}\vec{r}\vec{v} = -e\vec{B}\vec{r}\frac{d\vec{r}}{dt}. \qquad (5.58)$$

Проінтегруємо (5.58) за часом від 0 до t_0:

$$\int_0^{t_0} \frac{d\vec{L}}{dt} dt = -\int_0^{t_0} e\vec{B}\,(\vec{r}\frac{d\vec{r}}{dt})dt, \qquad (5.59)$$

де t_0 – час від моменту вильоту електрона з катода і до досягнення анода. Тоді (5.59) можна переписати так:

$$\int_{\vec{L}_k}^{\vec{L}_a} d\vec{L} = \vec{L}_a - \vec{L}_k = -\int_{r_k}^{r_a} e\vec{B}\, d\left(\frac{r^2}{2}\right) =$$

$$= -e\vec{B}\frac{r^2}{2}\Big|_{r_k}^{r_a} = -e\vec{B}\frac{r_a^2 - r_k^2}{2}, \qquad (5.60)$$

де $\vec{L}_a = \vec{r}_a \times m_e\vec{v}_a$ – момент кількості руху електрона біля анода; $\vec{L}_k = \vec{r}_k \times$

$\times m_e\vec{v}_k$ – момент кількості руху електрона біля катода; $\vec{v}_k -$ початкова швидкість вильоту електрона з катода. Будемо вважати, що початкова швидкість електрона v_k є значно меншою від швидкості електрона біля анода v_a. Справді, використовують електронну лампу з холодним катодом (підігрів незначний). Якщо вважати, що температура катода $T \cong 400 \ldots 500$ К, то енергія теплового руху $kT = 0{,}034 \ldots$ $\ldots 0{,}043$ еВ У нашій лампі електрони розганяються електричним полем з різницею потенціалів між анодом і катодом U_a~10 ... 30 В. Тоді електрон біля анода набуває енергії $eU_a \approx 10 \ldots 30$ еВ. Інакше кажучи, $eU_a \gg kT$, тому будемо вважати, що початкова швидкість електрона біля катода дорівнює нулю ($v_k = 0$). Тоді співвідношення (5.60) можна переписати так:

$$\vec{L}_a = -e\vec{B}\frac{r_a^2 - r_k^2}{2}, \qquad (5.61)$$

або

$$\vec{r}_a \times m_e\vec{v}_a = -e\vec{B}\frac{r_a^2 - r_k^2}{2}. \qquad (5.62)$$

Знак «мінус» у (5.62) враховує знак заряду електрона. Якщо індукція магнітного поля є критичною $B_{\text{кр}}$, то швидкість електрона біля анода $\vec{v}_a$ перпендикулярна до радіуса-вектора $\vec{r}_a$, й тоді (5.62) можна переписати так:

$$m_e r_a v_a = eB_{\text{кр}}\frac{r_a^2 - r_k^2}{2}. \qquad (6.63)$$

Швидкість електрона біля анода знайдемо із співвідношення

$$\frac{m_e v_a^2}{2} = eU_a. \qquad (5.64)$$

Тоді (5.63) можна переписати так:

$$m_e r_a\sqrt{\frac{2eU_a}{m_e}} = eB_{\text{кр}}\frac{r_a^2 - r_k^2}{2},$$

звідки знаходимо, що

$$\frac{e}{m_e} = \frac{8U_a}{B_{\text{кр}}^2 r_a^2\left(1 - \frac{r_k^2}{r_a^2}\right)^2}. \qquad (5.65)$$

За формулою (5.65) можна розрахувати питомий заряд електрона e/m_e, визначаючи $B_{\text{кр}}$ для анодної напруги U_a.

Для створення магнітного поля використовують соленоїд, діаметр перерізу якого значно менший за його довжину. Це дозволяє вважати його нескінченно довгим і використовувати формулу для індукції магнітного поля довгого соленоїда $B = \mu_0 n I_c$, де $\mu_0 = 4\pi \cdot 10^{-7}$ Гн/м – магнітна стала вакууму; $n -$ густина намотки витків соленоїда, тобто їхня кількість, що припадає на одиницю довжини соленоїда; $I_c -$ сила струму у витках соленоїда. Тоді величина критичної індукції магнітного поля становитиме

$$B_{\text{кр}} = \mu_0 n I_{\text{кр}}. \qquad (5.66)$$

Підставляючи вираз (5.66) у (5.65), отримаємо, що

$$\frac{e}{m_e} = \frac{8U_a}{\mu_0^2 n^2 I_{\text{кр}}^2 r_a^2\left(1 - \frac{r_k^2}{r_a^2}\right)^2}. \qquad (5.67)$$

Вираз (5.67) можемо дещо спростити. Для ниткоподібного катода електронної лампи $r_k \ll r_a$, тому $r_k^2/r_a^2 \ll 1$. Величини n та r_a є сталими для експериментальної установки, тому можна ввести таке позначення:

$$k = \frac{8}{\mu_0^2 n^2 r_a^2}. \quad (5.68)$$

Тоді формулу (5.67) можна переписати так:

$$\frac{e}{m_e} = \frac{kU_a}{I_{\text{кр}}^2}, \quad (5.69)$$

або

$$kU_a = \frac{e}{m_e} I_{\text{кр}}^2. \quad (5.70)$$

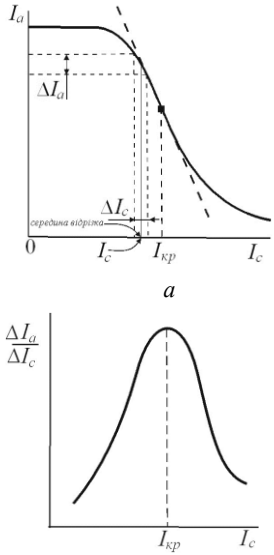
Таку залежність U_a від $I_{\text{кр}}^2$ називають ***параболою критичного режиму магнетрона***.

Визначення критичного струму в соленоїді

Як видно з (5.69), для знаходження відношення e/m_e потрібно визначити критичний струм $I_{\text{кр}}$ у соленоїді, який визначає критичну індукцію $B_{\text{кр}}$. Методика знаходження $I_{\text{кр}}$ – дискусійне питання, але більшість дослідників погоджуються, що за $B_{\text{кр}}(I_{\text{кр}})$ основна кількість електронів, які виходять з катода внаслідок явища термоелектронної емісії, не доходять до анода лампи, тому $I_{\text{кр}}$ відповідає *найбільш крутій* ділянці залежності анодного струму I_a від струму в соленоїді I_c (рис. 5.45, *б*). Ділянка, де функція $I_a = f(I_c)$, найбільш швидко змінюється, має точку, в якій похідна dI_a/dI_c має максимальне значення. Щоб знайти $I_{\text{кр}}$, потрібно побудувати графік похідної анодного струму I_a по струму I_c в соленоїді. Для цього треба розділити вісь струму I_c на інтервали ΔI_c і з рис. 5.47, *а* знайти відповідні їм величини ΔI_a. Знайти посередині кожного інтервалу ΔI_c на рис. 5.47, *а* значення струму I_c й побудувати графік залежності $\frac{\Delta I_a}{\Delta I_c} = f(I_c)$, як показано на рис. 5.47, *б*. Максимум побудованої функції $\frac{\Delta I_a}{\Delta I_c} = f(I_c)$ припадає для значення струму в соленоїді $I_c = I_{\text{кр}}$. Таку процедуру визначення критичного струму в соленоїді для певного значення анодної напруги називають графічним диференціюванням.

Таку операцію потрібно проводити для кожної залежності $I_a = f(I_c)$, отриманих для різних значень анодної напруги U_a. Це значно утруднює процес обробки результатів експеримен ту, тому можна використати простіший спосіб визначення $I_{\text{кр}}$ для кожного значення U_a. Для цього проводимо дотичну лінію до залежності $I_a = f(I_c)$ (рис. 5.47, *а*, пунктирна лінія) і знаходимо на цій залежності точку, в якій ця дотична лінія (пунктирна лінія на рис. 5.47, *а*) буде мати найбільший кут нахилу до осі

I_c. Координата цієї точки дотику на осі I_c визначає струм у соленоїді, за якого похідна dI_a/dI_c має максимальне значення (згідно з геометричним тлумаченням змісту похідної), а отже, струм у соленоїді відповідає значенню струму, що дорівнює $I_{\text{кр}}$. Якщо дотична пряма лінія має декілька точок дотику до залежності $I_a = f(I_c)$, то слід брати за точку дотику середину цього відрізка (рис. 5.47, *а*). Це дозволяє швидше і з меншими затратами часу знаходити $I_{\text{кр}}$ для кожного значення U_a.



а

б

Рис. 5.47. Метод графічного диференціювання

Методика визначення питомого заряду e/m_e. Питомий заряд електрона e/m_e можна визначати за формулою (5.69) за одними значеннями U_a і $I_{\text{кр}}$. Це буде неточно, оскільки отримано всього одне вимірювання. Для більш точного визначення відношення e/m_e за формулою (5.69) потрібно провести вимірювання для різних значень анодної напруги U_a, і для кожного значення U_a визначити відповідне значення $I_{\text{кр}}$. Згідно з формулою (5.70) залежність $kU_a = f(I_{\text{кр}}^2)$ буде лінійною в координатних осях kU_a і $I_{\text{кр}}^2$, як показано на рис. 5.48.

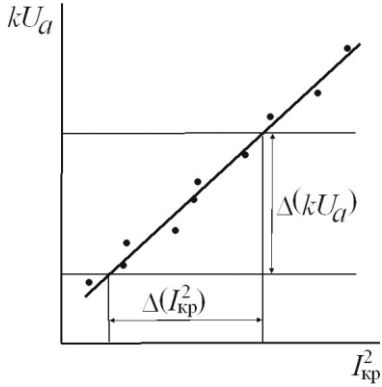


*Рис. 5.48. Лінійна залежність $kU_a = f(I_{\text{кр}}^2)$*Пряму лінію проводимо між експериментальними точками. Нахил цієї прямої лінії $\frac{\Delta(kU_a)}{\Delta(I_{\text{кр}}^2)}$ буде дорівнювати експериментально визначеній величині e/m_e. Більше того, можна використати метод найменших квадратів і математично за допомогою програми опрацювати експериментальні дані. Цей метод дозволяє визначити величину

e/m_e, а також її невизначеність (похибку).

Контрольні запитання

1. Що є джерелом існуванн магнітного поля?

2. Що є характеристикою магнітного поля у вакуумі?

3. У яких фізичних одиницях вимірюють індукцію магнітного поля?

4. Поясніть формулу для індукції магнітного поля електричного заряду, який рухається.

5. Сформулюйте та поясніть закон Біо−Савара−Лапласа.

6. Поясніть суть принципу суперпозиції магнітних полів.

7. Сформулюйте й поясніть теорему про циркуляцію вектора індукції магнітного поля для вакууму та теорему Остроградского−Гаусса для магнітного поля. На який характер магнітного поля вони вказують?

8. Яке магнітне поле прямолінійного провідника та колового витка зі струмом?

9. Яке магнітне поле соленоїда і тороїда?

10. Напишіть формули і поясніть їх для сили Ампера та сили Лоренца.

11. Де використовують поведінку контура зі струмом у магнітному полі?

12. В яких дослідах вивчають рух заряджених частинок у схрещених електричному і магнітному полях?

13. Чи змінює магнітне поле енергію зарядженої частинки, яка рухається в ньому? Відповідь поясніть.

14. Заряджена частинка влітає у взаємно перпендикулярні електричне і магнітне поля. За якої умови частинка буде рухатись рівномірно і прямолінійно?

15. Від чого залежить період обертання зарядженої частинки, яка рухається у магнітному полі?

16. Який вигляд матиме траєкторія заряджених частинок, що влітають зі сталою швидкістю v у магнітне поле з індукцією B, якщо: а) $\vec{v} \perp \vec{B}$; б) $\vec{v}$ утворює з $\vec{B}$ деякий кут; *в*) $\vec{v} \parallel \vec{B}$? Поясніть відповідь.

17. Який прилад називають магнетроном? Які фізичні явища лежать в основі його роботи? Де на практиці використовують магнетрони?

18. Які можливі траєкторії руху електронів у магнетроні за різних значень індукції магнітного поля B (покажіть траєкторії на рисунку)? Що таке критичне поле $B_{\text{кр}}$?

Розділ 6. Постійне магнітне поле в речовині

6.1. Намагнічування середовищ

Досі ми розглядали магнітне поле у вакуумі. Якщо провідники зі струмом розміщені не у вакуумі, а в іншому середовищі, то магнітне поле змінюється. Це означає, що речовини в магнітному полі намагнічуються, тобто самі стають джерелами магнітного поля. Результуюче магнітне поле в середовищі є сумою полів, які створюються провідниками зі струмом $\vec{B}_0$ і намагніченим середовищем $\vec{B}'$, і тому обидва поля в сумі дають результуюче поле:

$$\vec{B} = \vec{B}_0 + \vec{B}'. \qquad (6.1)$$

Речовини, які здатні намагнічуватись, називають магнетиками. Згідно з гіпотезою Ампера причина намагнічування полягає в тому, що в усіх речовинах існують найдрібніші електричні струми, які замикаються у межах кожного атома (молекулярні струми, рис. 6.1). Кожний такий струм має магнітний момент і створює в оточуючому просторі магнітне поле. Якщо немає зовнішнього магнітного поля, молекулярні струми орієнтовані хаотично, унаслідок чого їхнє сумарне магнітне поле дорівнює нулю. Під дією зовнішнього поля магнітні моменти молекул орієнтуються по полю (див. підрозд. 5.8), внаслідок чого магнетик намагнічується – його сумарний магнітний момент стає відмінним від нуля. Магнітні поля окремих молекулярних струмів у цьому випадку вже не компенсують один одного і виникає деяке середнє поле $\vec{B}'$, тому намагнічений магнетик можна подати як систему найдрібніших орієнтованих струмів (рис. 6.1).

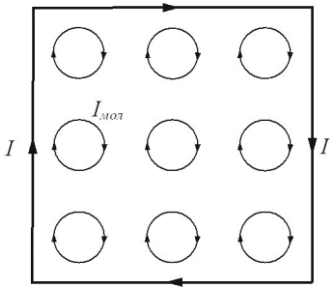


Рис. 6.1. Система найдрібніших орієнтованих струмів

У підрозд. 5.8 ми бачили, що всі магнітні дії замкнених струмів визначаються їх магнітним моментом:

$$\vec{p}_m = IS\vec{n},$$

де $I-$ сила струму; $S-$ площа, яка обігається струмом; $\vec{n}-$ одиничний вектор нормалі до площини витка зі струмом, який з напрямком струму утворює правогвинтову систему. Кожний молекулярний струм у магнетику має певний магнітний момент, а отже, і магнетик загалом під час намагнічування набуває магнітного моменту, який

дорівнює векторній сумі моментів усіх молекулярних струмів, тому магнітний стан речовини можна цілком охарактеризувати магнітним моментом його одиниці об'єму. Цю величину називають *вектором намагнічування* йпозначають літерою $\vec{J}$. Якщо магнетик намагнічений неоднорідно, то намагніченість у цій точці визначають таким виразом:

$$\vec{J} = \frac{\sum_{\Delta V} \vec{p}_m}{\Delta V}, \qquad (6.2)$$

де $\Delta V -$ фізично нескінченно малий об'єм, взятий біля точки, яку розглядаємо; $\vec{p}_m$ – магнітний момент окремої молекули. Сумування виконують по всіх молекулах, що перебувають в об'ємі ΔV (порівняйте з формулою (2.9)).

Вектор намагнічування є основною величиною, яка характеризує магнітний стан речовини. Знаючи вектор намагнічування у кожній точці якого-небудь тіла, можна визначити і магнітне поле, яке створює намагнічене тіло, що розглядається. Завдання значно спрощується, якщо вектор намагнічування однаковий в усіх точках магнетика (*однорідне* намагнічування). У цьому випадку під час додавання молекулярних струмів прилеглі їхні відрізки, що мають протилежні напрямки струмів, взаємно компенсуються, і залишаються лише відрізки струмів, які прилягають до поверхні магнетика; тому дія всіх молекулярних струмів буде така сама, як дія деякого поверхневого струму, що обтікає намагнічений магнетик (рис. 6.1). У цьому розумінні можна сказати, що під час занурення у соленоїд залізного осердя на поверхні осердя ніби виникають додаткові невидимі ампер-витки, які додаються до ампер-витків намагніченої котушки (рис. 6.2).

Величину зазначеного поверхневого струму визначають значенням намагнічування $\vec{J}$. Помістимо досить довгий циліндричний стрижень в магнітне поле досить довгого соленоїда (рис. 6.2) і позначимо через i_1 лінійну густину поверхневого струму, тобто величину струму на одиницю довжини стрижня. Тоді повна величина поверхневого струму стрижня є $i_1 l$, де l – довжина стрижня. Якщо S – площа перерізу стрижня, то величина його магнітного момента

$$i_1 lS = i_1 V,$$

де $V -$ об'єм стрижня. Утім, за визначенням намагнічування J цей самий момент дорівнює JV. Прирівнюючи обидва вирази, знаходимо:

$$J = i_1. \qquad (6.3)$$

У разі однорідного намагнічування значення J безпосередньо дає лінійну густину поверхневого струму магнетика.

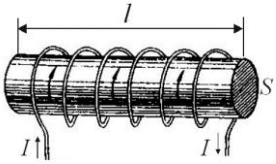


Рис. 6.2. Поверхневі струми в намагніченому циліндрі

Поле $\vec{B}'$ так само, як і поле $\vec{B}_0$, не має джерел (магнітних зарядів), тому потік результуючого поля (6.1) через замкнуту поверхню дорівнює нулю:

$$\oint_S \vec{B} d\vec{S} = \oint_S \vec{B}_0 d\vec{S} + \oint_S \vec{B}' d\vec{S} = 0. \quad (6.4)$$

Таким чином, формула (5.24) справедлива не тільки для поля у вакуумі, а й для поля у речовині.

6.2. Магнітомеханічні явища

Природа молекулярних струмів стала зрозумілою після того, як дослідами Резерфорда було встановлено, що атоми всіх речовин складаються із позитивно зарядженого ядра і негативно заряджених електронів, які рухаються навколо нього.

Рух електронів в атомах підпорядковується квантовим законам, тому поняття траєкторії до руху електронів в атомі не застосовують. Утім, магнетизм речовини вдається пояснити, користуючись простою моделлю Бора, згідно з якою електрони в атомах рухаються по стаціонарних колових орбітах. Рух електрона по одній із таки орбіт (рис. 6.3) еквівалентний коловому струму, тому він має ***орбітальний магнітний момент*** $\vec{p}_m = IS\vec{n}$, модуль якого $p_m = IS$, де S – площа орбіти електрона: $S = \pi r^2$. Вектор $\vec{p}_m$ напрямлений у той самий бік, що й індукція магнітного поля в центрі колового струму I. Швидкість обертання електрона за секунду становить $\nu = \frac{v}{2\pi r}$. Тоді сила струму $I = e\nu = \frac{ev}{2\pi r}$ і

$$p_m = \frac{ev}{2\pi r}\pi r^2 = \frac{evr}{2}. \quad (6.4)$$

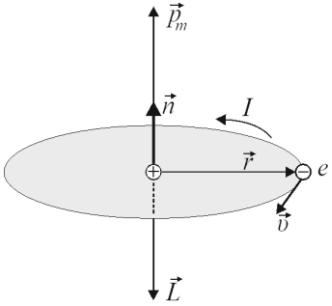


Рис. 6.3. Рух електрона по стаціонарній коловій орбіті

Заразом кожний електрон, що рівномірно рухається по орбіті, має ***орбітальний механічний момент імпульсу***, який дорівнює $\vec{L} = \vec{r} \times m_e\vec{v}$, і цей вектор направлений перпендикулярно до площини руху електрона. Оскільки $\vec{L} \perp \vec{v}$, то величина

$$L = m_e vr. \quad (6.5)$$

Ця формула нерелятивістська, але для атома вона достатньо справедлива, оскільки для електрона на борівській орбіті відношення v/c (c – швидкість

світла у вакуумі) дорівнює $e^2/(4\pi\varepsilon_0\hbar c) = 1/137$, що складає близько 1 % від швидкості світла ($\hbar$ – стала Планка).

Тоді справедливе таке відношення:

$$\frac{p_m}{L} = \frac{e}{2m_e}.$$

Вектори $\vec{p}_m$ та $\vec{L}$ напрямлені у взаємно протилежні сторони, тому

$$\vec{p}_m = -\frac{e}{2m_e}\vec{L} \text{ і } \vec{p}_m = -g\vec{L}, \quad (6.6)$$

де $g = \frac{e}{2m_e}$ – гіромагнітне відношення орбітальних моментів.

Гіромагнітне відношення для електрона виявилось у два рази більшим від гіромагнітного відношення орбітальних моментів.

Щоб пояснити результат експерименту, припустили, що електрон, крім орбітальних моментів $\vec{p}_m$ та $\vec{L}$, має ***власний механічний момент імпульсу*** $\vec{L}_s$, який називають ***спіном електрона*** (детальніше розглянуто у розд. 17). Спін є невід'ємною властивістю електрона, як і заряд e та маса m_e. Спін електрона проявляється в багатьох експериментальних дослідах. Спіну електрона $\vec{L}_s$ відповідає власний (спіновий) магнітний момент $\vec{p}_{ms}$, який становить

$$\vec{p}_{ms} = -g_s\vec{L}_s = -\frac{e}{m_e}\vec{L}_s, \quad (6.7)$$

де $g_s = \frac{e}{m_e}$ – гіромагнітне відношення спінових моментів.

Спін мають не тільки електрони, але й інші елементарні частинки. Максимальне значення проєкції спіну на виділений у просторі напрямок, наприклад на напрямок зовнішнього поля, дорівнює цілому або напівцілому значенню величини $\hbar$, яка дорівнює сталій Планка $h/(2\pi)$. Для електрона $L_s = \frac{1}{2}\hbar$, тому говорять, що спін електрона дорівнює $1/2$. Таким чином, $\hbar$ є природною одиницею моменту імпульсу, подібно до того, як елементарний заряд e є природною одиницею заряду.

Згідно з (6.7) спіновий магнітний момент електрона

$$p_{ms} = -\frac{e}{m_e}L_s = -\frac{e\hbar}{2m_e}. \quad (6.8)$$

Величину

$$\mu_Б = \frac{e\hbar}{2m_e} = 0{,}927 \cdot 10^{-23} \text{ Дж/Тл}$$

називають ***магнетоном Бора***.

Магнітний момент електрона складається з його орбітального і спінового магнітного моментів. Відповідно, магнітний момент атома $\vec{p}_{ma}$ дорівнює сумі магнітних моментів електронів, що входять до складу атома, і магнітного моменту ядра, який значно менший від магнітного моменту

електронів, тому магнітним моментом ядра зазвичай нехтують. Отже,

$$\vec{p}_{ma} = \sum_{i=1}^{Z} \vec{p}_{mi} + \sum_{i=1}^{Z} \vec{p}_{msi},$$

де Z – атомний номер хімічного елемента.

Якщо немає зовнішнього магнітного поля, магнітні моменти орієнтовані хаотично, унаслідок чого сумарний магнітний момент речовини дорівнює нулю, і результуюча магнітна індукція поля, яке створене мікроскопічними струмами, дорівнює нулю. Під дією зовнішнього магнітного поля магнітні моменти молекул орієнтуються в певному напрямку і виникає внутрішнє магнітне поле, сумарний магнітний момент молекул відмінний від нуля, а це означає, що речовина намагнічується.

Таким чином, магнетизм речовин обумовлений трьома причинами:

1) орбітальним рухом електронів навколо атомних ядер;

2) спіном електрона;

3) магнітним моментом атомних ядер.

Важкі атомні ядра рухаються значно повільніше, ніж легкі електрони, тому магнітні моменти атомних ядер у тисячі разів менші від орбітальних і спінових магнітних моментів електронів.

6.3. Закон повного струму для магнітного поля в речовині. Напруженість магнітного поля. Магнітна сприйнятливість і проникність

Закон повного струму, згідно з формулою (5.23), для результуючого магнітного поля (6.1) у речовині буде мати такий вигляд:

$$\oint_{\Gamma} \vec{B} d\vec{l} = \oint_{\Gamma} \vec{B}_0 d\vec{l} + \oint_{\Gamma} \vec{B}' d\vec{l} = \mu_0 (\sum_i I_i + \sum I_{мол}), \quad (6.9)$$

де $\sum_i I_i$ – алгебраїчна сума макрострумів (струмів провідності); $\sum I_{мол}$ – алгебраїчна сума мікрострумів (молекулярних струмів), що охоплюються заданим контуром Г. З контуром Г пов'язана деяка поверхня S (рис. 6.4), тому

$$\sum I_{мол} = \int_S \vec{j}_{мол} \cdot d\vec{S}, \quad (6.10)$$

де $\vec{j}_{мол}$ – густина молекулярних струмів, яка залежить від значень вектора $\vec{B}$, який, у свою чергу, залежить від $\vec{j}_{мол}$. Щоб обійти цю перешкоду, знайдемо таку допоміжну фізичну величину, циркуляція якої визначається лише силою макроскопічних струмів. Щоб встановити вид цієї допоміжної величини, виразимо густину молекулярних струмів $\vec{j}_{мол}$ через намагніченість магнетика. Для цього вирахуємо суму (6.10). В алгебраїчну суму (6.10) входять тільки ті молекулярні

струми, які є «нанизані» на контур Г (рис. 6.4).

Струми, які не «нанизуються» на контур Г або не перетинають цю поверхню зовсім, або перетинають цю поверхню два рази – один раз в одному напрямку, другий раз в іншому (струм $I''_{мол}$ на рис. 6.4), не роблять внеску в алгебраїчну суму струмів (6.10).

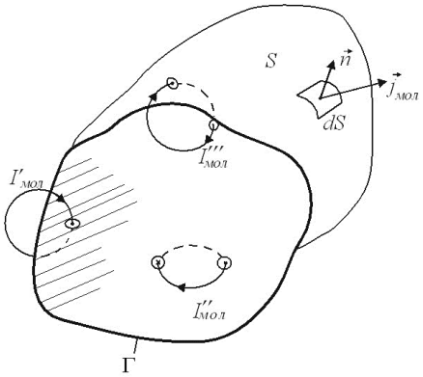


Рис. 6.4. До підрахунку суми $\sum I_{мол}$

Із рис. 6.5 видно, що елемент контура $d\vec{l}$, який утворює з напрямком намагніченості $\vec{J}$ кут α, «нанизує» ті молекулярні струми, центри яких потрапляють всередину косого циліндра з об'ємом $S_{мол} \cdot dl \cdot \cos\alpha$ ($S_{мол}$ – площа, яка охоплюється одним молекулярним струмом). Якщо n – кількість молекул в одиниці об'єму магнетика, то сумарний струм, який охоплюється елементом контура dl, дорівнює $I_{мол} \cdot nS_{мол}dl \cdot \cos\alpha$. Добуток $I_{мол} \cdot S_{мол} = p_m$, де p_m – магнітний момент окремого молекулярного струму.

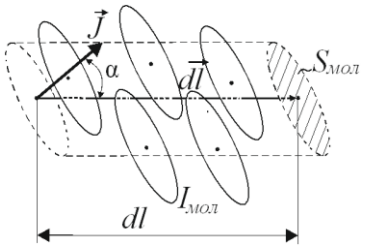


Рис. 6.5

Отже, вираз $I_{мол} \cdot S_{мол} \cdot n$ є магнітним моментом одиниці об'єму, тобто модуль вектора $\vec{J}$, а $I_{мол} \cdot S_{мол} \cdot n \cdot \cos\alpha$ – проєкція вектора $\vec{J}$ на напрямок елемента $d\vec{l}$. Таким чином, сумарний молекулярний струм, що охоплюється елементом контура $d\vec{l}$, дорівнює $\vec{J}d\vec{l}$, а сума молекулярних струмів (6.10), які охоплюються усім контуром Г, становить

$$\int_S \vec{j}_{мол} \cdot d\vec{S} = \oint_\Gamma \vec{J}d\vec{l}. \qquad (6.11)$$

З урахуванням (6.11) вираз (6.9) можна переписати так:

$$\oint_\Gamma \left(\frac{\vec{B}}{\mu_0} - \vec{J}\right) d\vec{l} = \sum_i I_i. \qquad (6.12)$$

Вираз у дужках (6.12)

$$\vec{H} = \frac{\vec{B}}{\mu_0} - \vec{J} \qquad (6.13)$$

і є шуканою допоміжною величиною, циркуляцію якої визначають лише макроскопічними струмами, тоді (6.12) можна переписати так:

$$\oint_{\Gamma} \vec{H}\, d\vec{l} = \sum_i I_i. \qquad (6.14)$$

Вектор $\vec{H}$ *називають напруженістю магнітного поля.*

Отримана формула (6.14) виражає теорему про циркуляцію вектора $\vec{H}$ (закон повного струму для магнітного поля у середовищі): *циркуляція вектора напруженості магнітного поля по деякому замкнутому контуру дорівнює алгебраїчній сумі макроскопічних струмів, що охоплює цей контур*.

Напруженість магнітного поля $\vec{H}$ є аналогом електричного зміщення $\vec{D}$ (2.18).

У вакуумі $\vec{J} = 0$, тому $\vec{H} = \vec{B}/\mu_0$ і

$$\vec{B} = \mu_0 \vec{H}. \qquad (6.15)$$

Як показують досліди, у несильних магнітних полях намагніченість $\vec{J}$ прямо пропорційна напруженості поля $\vec{H}$, що викликає намагнічення, тобто

$$\vec{J} = \chi \vec{H}, \qquad (6.16)$$

де χ – магнітна сприйнятливість речовини. Тоді (6.13) з урахуванням (6.16) можна переписати так:

$$\vec{H} = \frac{\vec{B}}{\mu_0(1+\chi)}. \qquad (6.17)$$

Безрозмірну величину

$$\mu = 1 + \chi \qquad (6.18)$$

називають *відносною магнітною проникністю* речовини.

На відміну від діелектричної сприятливості χ (формула (2.11)), яка може мати тільки позитивні значення (поляризованість $\vec{P}$ в ізотропному діелектрику завжди направлена по полю $\vec{E}$), магнітна сприйнятливість χ буває як позитивною, так і негативною, тому магнітна проникність μ може бути як більшою від одиниці, так і меншою за одиницю.

З урахуванням (6.18) формула (6.17) набуде вигляду

$$\vec{B} = \mu\mu_0 \vec{H}. \qquad (6.19)$$

Таким чином, напруженість магнітного поля $\vec{H}$ є вектором, який має той самий напрямок, що і вектор $\vec{B}$, але в $\mu\mu_0$ разів меншу за модулем величину (в анізотропних середовищах вектори $\vec{H}$ і $\vec{B}$ загалом не збігаються за напрямком).

6.4. Магнітне поле на межі двох магнетиків

Поблизу поверхні поділу двох магнетиків вектори $\vec{B}$ і $\vec{H}$ повинні задовольняти граничні умови, які отримують з таких співвідношень (див. формули (5.24) і (6.14)):

$$\oint_S \vec{B} d\vec{S} = 0, \oint_\Gamma \vec{H}\, d\vec{l} = \sum_i I_i. \quad (6.20)$$

Розглянемо стаціонарні поля. Візьмемо на межі поділу двох магнетиків з проникностями μ_1 та μ_2 уявну циліндричну поверхню, висота якої h, а площі основ $S_1 = S_2 = S$, які розміщені по різні сторони поверхні поділу (рис. 6.6). Потік вектора $\vec{B}$ через цю циліндричну поверхню

$$\Phi_B = \vec{B}_1\vec{n}_1S_1 + \vec{B}_2\vec{n}_2S_2 + < B > S_{бічна}$$

або

$$\Phi_B = -B_{1n}S_1 + B_{2n}S_2 + < B > S_{бічна},$$

де $<B>$ – деяке усереднене значення магнітної індукції на межі поділу середовищ, B_{1n} і B_{2n} – проєкції векторів $\vec{B}_1$ і $\vec{B}_2$ на відповідні нормалі $\vec{n}_1$ і $\vec{n}_2$ до поверхні циліндра (рис. 6.6). Оскільки потік вектора $\vec{B}$ через будь-яку замкнуту поверхню дорівнює нулю, то за умови $h \to 0$ $(S_{бічна} \to 0)$ отримаємо співвідношення

$$B_{1n} = B_{2n}. \qquad (6.21)$$

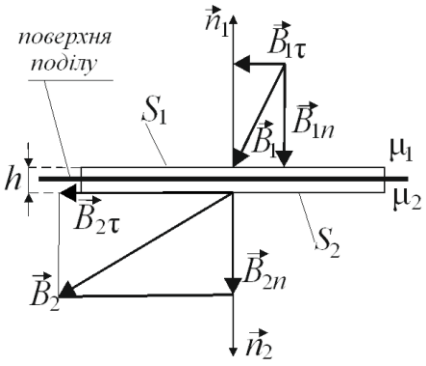


Рис. 6.6

Замінивши, згідно з (6.19), складові вектора $\vec{B}$ відповідними складовими вектора $\vec{H}$, помноженими на $\mu_0\mu$, отримаємо співвідношення $\mu_0\mu_1H_{1n} = \mu_0\mu_2H_{2n}$, з якого випливає, що

$$\frac{H_{1n}}{H_{2n}} = \frac{\mu_2}{\mu_1}. \qquad (6.22)$$

Тепер візьмемо на межі поділу двох магнетиків прямокутний контур Г (рис. 6.7 і вирахуємо для нього циркуляцію $\vec{H}$. За малих розмірів контура Г циркуляцію можна подати у такому вигляді:

$$\oint_\Gamma \vec{H}\,d\vec{l} = \int_1^2 \vec{H}_1\,d\vec{l} + \int_2^3 \langle\vec{H}\rangle\,d\vec{l} + \int_3^4 \vec{H}_2\,d\vec{l} + \int_4^1 \langle\vec{H}\rangle\,d\vec{l},$$

або

$$\oint_\Gamma \vec{H}\,d\vec{l} = -H_{1\tau}a + 2\langle H_l\rangle b + H_{2\tau}a,$$

де $\langle H_l\rangle$ – середнє значення H_l на перпендикулярних до поверхні поділу ділянках контура; $H_{1\tau} = \vec{H}_1 \cdot \vec{\tau}_1$, $H_{2\tau} = \vec{H}_2 \cdot \vec{\tau}_2$, $\vec{\tau}_1$ і $\vec{\tau}_2$ – одиничні вектори, дотичні до межі поділу магнетиків; $H_{1\tau}$ і $H_{2\tau}$ – тангенційні складові векторів $\vec{H}_1$ і $\vec{H}_2$.

Якщо на межі поділу не проходять макроскопічні струми, то згідно з (6.20) за умови, що $b \to 0$, отримаємо

$$\oint_\Gamma \vec{H}\,d\vec{l} = -H_{1\tau}a + H_{2\tau}a = 0.$$

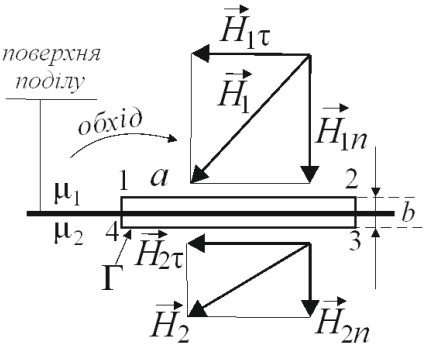


Рис. 6.7

Звідси отримаємо таке співвідношення (порівняйте з (2.25)):

$$H_{1\tau} = H_{2\tau}. \qquad (6.23)$$

Замінимо складові $\vec{H}$ у (6.23) відповідними складовими $\vec{B}$, згідно з (6.19) поділеними на $\mu_0\mu$, отримаємо співвідношення

$$\frac{B_{1\tau}}{\mu_0\mu_1} = \frac{B_{2\tau}}{\mu_0\mu_2},$$

з якого випливає, що

$$\frac{B_{1\tau}}{B_{2\tau}} = \frac{\mu_1}{\mu_2}. \qquad (6.24)$$

Із співвідношень (6.21) і (6.23) можна зробити висновок, що під час переходу через межу поділу двох магнетиків нормальна складова вектора $\vec{B}$ і тангенційна складова вектора $\vec{H}$ змінюються неперервно. Із (6.22) і (6.24) випливає, що тангенційна складова вектора $\vec{B}$ і нормальна складова вектора $\vec{H}$ під час переходу через межу поділу двох магнетиків зазнають розриву. Таким чином, під час переходу через межу поділу двох середовищ вектор $\vec{B}$ поводиться аналогічно вектору $\vec{D}$, а вектор $\vec{H}$ – аналогічно вектору $\vec{E}$.

На рис. 6.8 показано поведінку вектора $\vec{B}$ під час перетину межі поділу двох магнетиків ($\mu_2 > \mu_1$). Позначимо кути між лініями $\vec{B}$ і нормаллю до поверхні поділу, відповідно, α_1 та α_2 (рис. 6.8). Відношення тангенсів цих кутів

$$\frac{\mathrm{tg}\alpha_1}{\mathrm{tg}\alpha_2} = \frac{B_{1\tau}/B_{1n}}{B_{2\tau}/B_{1n}}.$$

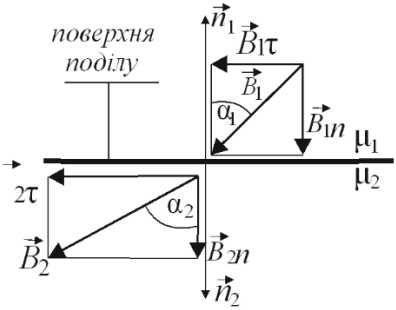


Рис. 6.8. Заломлення силової лінії магнітної індукції на межі поділу двох магнетиків

Звідси, з урахуванням (6.21) і (6.24), отримаємо закон заломлення ліній магнітного індукції:

$$\frac{\mathrm{tg}\alpha_1}{\mathrm{tg}\alpha_2} = \frac{\mu_1}{\mu_2}. \qquad (6.25)$$

Під час переходу в магнетик з більшою μ ($\mu_2 > \mu_1$) лінії магнітної

індукції відхиляються від нормалі до поверхні поділу двох середовищ. Це призводить до згущення ліній. Згущування ліній $\vec{B}$ у речовині з більшою магнітною проникністю дає можливість формувати магнітні пучки, тобто надавати їм потрібних форми і напрямку. Щоб здійснити *магнітний захист* деякого об'єму, його оточують залізним екраном. Із рис. 6.9 видно, що згущення ліній магнітної індукції в товщі екрана призводить до ослаблення поля всередині.

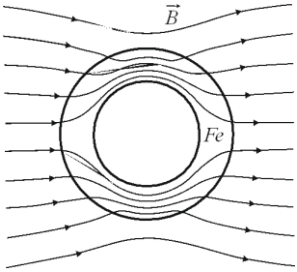


Рис. 6.9. Захист від магнітних полів

6.5. Класифікація магнетиків

Формула (6.16) визначає магнітну сприйнятливість χ одиниці об'єму речовини. Замість цієї сприйнятливості часто використовують сприйнятливість для одного кіломоля речовини, яку називають кіломолярною $\chi_{км}$ (для хімічно простих речовин кілоатомною $\chi_{кат}$). Очевидно, що $\chi_{км}(\chi_{кат})$ має розмірність $м^3/кмоль$, $(м^3/кат)$.

Залежноі від знака і величини магнітної сприйнятливості всі магнетики поділяються на три групи.

1. Діамагнетики, в яких χ від'ємна і мала за абсолютною величиною ($\chi_{км} \sim -10^{-8} \ldots -10^{-7}$ $м^3/кмоль$). До таких речовини належать: вуглекислота, вода, срібло, вісмут, мідь, цинк, золото, аргон та ін. Діамагнетики послаблюють зовнішнє магнітне поле. Від'ємне значення χ у діамагнетиках означає, що в цих речовинах вектор намагнічення направлений протилежно зовнішньому (намагнічуючому) полю.

2. Парамагнетики – це речовини, в яких χ також невелика, але позитивна величина ($\chi_{км} \sim 10^{-7} \ldots 10^{-6}$ $м^3/кмоль$). Такими речовинами є кисень, алюміній, літій, натрій, калій, магній, кальцій, титан, вольфрам, платина, уран, плутоній, хлористе залізо ($FeCl_3$) та ін.

3. Феромагнетики – це речовини, в яких χ додатна і досягає дуже великих значень ($\chi_{км} \sim 10^3$ $м^3/кмоль$). Крім того, на відміну від діа- та парамагнетиків, для яких χ не залежить від H, сприйнятливість феромагнетиків залежить від напруженості магнітного поля. Феромагнетиками є хімічні елементи Fe, Co, Ni ($3d$-метали) і рідкоземельні метали Gb, Tb, Dy, Ho, Er та ін. Феромагнетиками є також численні металічні бінарні й більш складні

(багатокомпонентні) сплави згаданих вище металів між собою та з іншими неферомагнітними системами, наприклад Fe_3Al, Ni_3Mn, $MnPt_3$, $ZrZn_2$, $ZnCMn_3$, $AlCMn_3$ та ін.

6.6. Діа- і парамагнетизм

Діамагнетизм (від грец. dia – префікс, який тут означає розходження (силових ліній) і магнетизм) – це властивість речовини намагнічуватись назустріч напрямку зовнішнього магнітного поля, яке діє на неї. Діамагнетизм властивий усім речовинам. Розглянемо вплив зовнішнього магнітною поля $\vec{B}$ на рух електронів в атомах речовини.

Нехай навколо ядра атома рівномірно рухається по коловій орбіті електрон. Якщо магнітного поля немає, то кулонівська сила є доцентровою, і згідно з другим законом Ньютона

$$\frac{Ze^2}{4\pi\varepsilon_0 r^2} = \frac{m_e \upsilon_0^2}{r} = m_e \omega_0^2 r, \quad (6.26)$$

де υ_0 – лінійна швидкість електрона на орбіті; ω_0 – його кутова швидкість; r – радіус орбіти.

В однорідному магнітному полі, лінії індукції якого перпендикулярні до площини орбіти електрона, на кожний електрон діятиме ще й сила Лоренца:

$$F_л = e\upsilon B = e\omega r B,$$

де ω – кутова швидкість за наявності магнітного поля (рис. 6.10).

Тепер рівняння Ньютона запишемо так:

$$\frac{Ze^2}{4\pi\varepsilon_0 r^2} - e\omega r B = m_e \omega^2 r,$$

або із урахуванням (6.26)

$$m_e r(\omega^2 - \omega_0^2) = -e\omega r B. \quad (6.27)$$

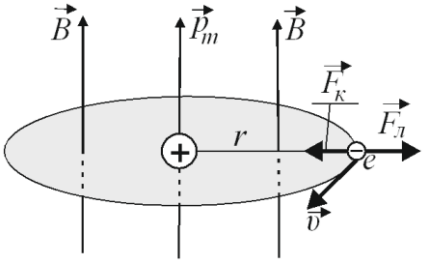


Рис. 6.10. Рух електрона в магнітному полі

Із (6.27) отримаємо, що

$$(\omega - \omega_0)(\omega + \omega_0) = -\frac{e}{m_e} B\omega.$$

Якщо позначити $\omega - \omega_0 = \omega_L$, а $\omega + \omega_0 \approx 2\omega$, то

$$\omega_L = -\frac{eB}{2m_e}. \quad (6.28)$$

Частоту (6.28) називають частотою *ларморівської прецесії* або просто *ларморовою частотою*.

Отже, магнітне поле викликає зміну кутової швидкості обертання електрона по орбіті. У випадку, коли $\vec{B}$ не перпендикулярне площині орбіти, то дія поля полягає у збудженні прецесії орбіти навколо поля (рис. 6.11).

Магнітний момент електрона $\vec{p}_m$ описує конус навколо $\vec{B}$. У цьому випадку

кутову швидкість ω_L такої прецесії також виражають формулою (6.28). Кутова швидкість ω_L є однаковою для всіх електронів атома й не залежить ні від кута нахилу орбіти φ відносно вектора $\vec{B}$, ні від радіуса орбіти або швидкості електрона.

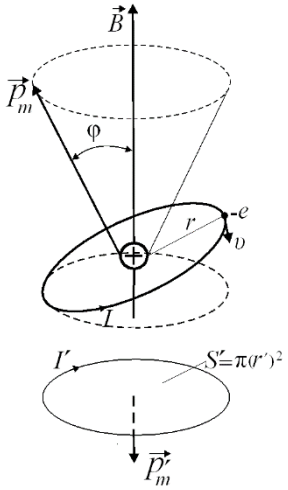


Рис. 6.11. Прецесії орбіти електрона навколо магнітного поля

Сформулюємо ***теорему Лармора:*** *єдиним наслідком впливу магнітного поля на орбіту електрона в атомі є прецесія орбіти і вектора* $\vec{p}_m$ *з кутовою швидкістю* ω_L *навколо осі, яка проходить через ядро атома і паралельна вектору* $\vec{B}$ *індукції магнітного поля.*

Прецесія орбіти обумовлює додатковий рух електрона навколо напрямку поля. Якщо б відстань r електрона від осі, що проходить через центр орбіти, не змінювалась, то додатковий рух електрона відбувався б по колу, радіус якого r.

Йому відповідав би коловий струм

$$I' = e\frac{\omega_L}{2\pi} = \frac{e^2 B}{4\pi m_e},$$

магнітний момент якого становить

$$p'_m = I'S' = \frac{e^2}{4\pi m_e} S'B$$

і напрямлений у бік, протилежний до вектора $\vec{B}$ (рис. 6.11):

$$\vec{p}'_m = -\frac{e^2}{4\pi m_e} S'\vec{B}. \qquad (6.29)$$

Магнітний момент $\vec{p}'_m$ називають індукованим магнітним моментом. Розрахунок показує, що $S' = \pi(r')^2 = \frac{2}{3}\pi r^2$.

У результаті отримуємо середнє значення індукованого магнітного моменту одного електрона:

$$\langle p'_m \rangle = -\frac{e^2 B}{6m_e} r^2. \qquad (6.30)$$

Для отримання виразу для $\langle p'_m \rangle$ передбачалось, що електрони рухаються по колових орбітах. Загалом, наприклад, для еліптичної орбіти треба замість r^2 брати $\langle r^2 \rangle$, тобто середній квадрат відстані електрона від ядра.

Індукований магнітний момент атома дорівнює сумі індукованих моментів електронів, що входять до складу атома:

$$p'_{ma} = \sum\langle p'_m\rangle = -\frac{e^2B}{6m_e}\sum_{i=1}^{Z}\langle r_i^2\rangle, \quad (6.31)$$

де Z – атомний момент хімічного елемента (кількість електронів в атомі дорівнює Z).

Таким чином, під дією зовнішнього магнітного поля відбувається прецесія електронних орбіт з однаковою для всіх електронів кутовою швидкістю (6.28). Обумовлений прецесією додатковий рух електронів приводить до виникнення індукованого магнітного моменту атома (6.31), який направлений проти зовнішнього поля. Ларморова прецесія виникає у всіх без винятку речовин, однак у тих випадках, коли атоми мають самі по собі магнітні моменти, магнітне поле не тільки індукує моменти (6.31), але й орієнтує магнітні моменти атомів за напрямком зовнішнього поля. Водночас виникає позитивний (тобто направлений по полю) магнітний момент, який буває значно більшим, ніж індукований від'ємний момент (6.31). Це означає, що результуючий момент стає позитивним і речовина поводить себе як парамагнетик.

Отже, діамагнетиками називають речовини, магнітні моменти атомів або молекул яких дорівнюють нулю, коли немає зовнішнього магнітного поля.

Коли внести діамагнетик у зовнішнє магнітне поле, у кожному його атомі індукується магнітний момент (6.31), який направлений протилежно вектору $\vec{B}$ магнітної індукції поля.

Якщо (6.31) помножити на число Авогадро N_A, то отримаємо магнітний момент одного кіломоля ізотропного діамагнетика J. Поділивши J на напруженість магнітного поля H, отримаємо кіломолярну сприйнятливість $\chi_{км}$. Магнітна проникність діамагнетиків фактично дорівнює одиниці, тому $B = \mu_0 H$:

$$\chi_{км} = -\frac{J}{H} = -\frac{\mu_0 J}{B} = -\frac{\mu_0 N_A\, e^2}{6m_e}\sum_{i=1}^{Z}\langle r_i^2\rangle. \quad (6.32)$$

Зазначимо, що строга квантово-механічна теорія дає такий самий вираз для $\chi_{км}$, як і (6.32).

Підставивши у (6.32) числові позначення для μ_0, N_A, e та m_e, отримаємо, що

$$\chi_{км} = -3{,}55\cdot 10^{12}\sum_{i=1}^{Z}\langle r_i^2\rangle. \quad (6.33)$$

Радіуси електронних орбіт мають величину порядку 10^{-10} м. Отже, кіломолярна діамагнітна сприйнятливість буде становити порядку $10^{-8} - -10^{-7}$ м^3/кмоль, що добре узгоджується з експериментальними даними. Стрижень з діамагнітного матеріалу намагнічується у напрямку, протилежному до напрямку зовнішнього магнітного поля, тому в неоднорідному

магнітному полі діамагнетик виштовхується в область слабшого поля і встановлюється так, щоб його вісь була перпендикулярною до вектора магнітної індукції $\vec{B}$ поля (рис. 6.12).

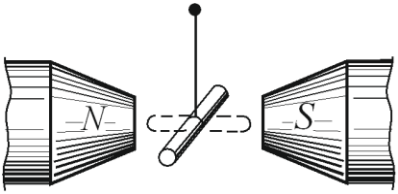


Рис. 6.12. Діамагнітна паличка вісмуту в магнітному полі

Якщо векторна сума орбітальних магнітних моментів усіх електронів атома або молекули не дорівнює нулю, то атом загалом має деякий магнітний момент $\vec{p}_{ma}$. *Такі атоми (молекули) називають парамагнітними, а речовини, що складаються з них – парамагнетиками.*

Парамагнетизм (від грец. para – біля, поруч і магнетизм) – властивість речовин (парамагнетиків), які розміщені у зовнішньому магнітному полі, намагнічуватись (набувати магнітного моменту) в напрямку, який збігається з напрямком цього поля.

Якщо немає зовнішнього магнітного поля, парамагнетик ненамагнічений, оскільки внаслідок теплового руху власні магнітні моменти атомів орієнтовані хаотично ($J = 0$).

Розглянемо, що станеться у разі внесення парамагнетика в однорідне магнітне поле, магнітна індукція якого дорівнює $\vec{B}$.

Кожний електрон атома парамагнітного тіла бере участь у двох рухах – орбітальному і прецесійному. Згідно з теоремою Лармора всі магнітні моменти електронів атома і результуючий магнітний момент атома прецесують навколо напрямку $\vec{B}$ з однаковою кутовою швидкістю ω_L.

Тепловий рух атомів парамагнетика і їх зіткнення спричиняють поступове згасання прецесії магнітних моментів атомів і, як наслідок, зменшення кутів між напрямками $\vec{p}_{ma}$ і $\vec{B}$ (рис. 6.11). Отже, незважаючи на утворюване тепловим рухом «розкидання» атомів, цей рух водночас сприяє переважній орієнтації магнітних моментів атомів у напрямку зовнішнього магнітного поля, оскільки саме по собі магнітне поле може спричиняти лише прецесію $\vec{p}_{ma}$ навколо напрямку $\vec{B}$.

Магнітний момент p_{ma} окремого атома має величину приблизно 10^{-23}Дж/Тл, але сукупна дія магнітних моментів усіх атомів, що містяться в одиниці об’єму речовини, приводить до ефекту намагнічування, що значно

перевищує діамагнітний ефект, тому парамагнетик намагнічується «за полем», тобто в напрямку магнітної індукції $\vec{B}$ зовнішнього магнітного поля.

Класичну теорію парамагнетизму розвинув П. Ланжевен, який розглянув задачу про поведінку молекулярних струмів в однорідному магнітному полі.

Орієнтувальна дія магнітного поля на атом залежить від магнітного моменту атома і від магнітної індукції $\vec{B}$ поля. «Розкидальна» дія теплового руху визначається величиною kT, пропорційною до середньої теплової енергії однієї частинки. Виявилося, що результуюча дія двох протилежних чинників залежить від відношення $\alpha = \frac{p_{ma}B}{kT}$, де $p_{ma}B$ – енергія магнітного моменту p_{ma} у магнітному полі B.

Ланжевен знайшов залежність намагніченості J парамагнетика від параметра α:

$$J = np_{ma}L(\alpha),$$

де n – концентрація атомів парамагнетика; $L(\alpha)$ – класична функція Ланжевена, яка має такий вигляд:

$$L(\alpha) = \left\{\frac{e^{\alpha}+e^{-\alpha}}{e^{\alpha}-e^{-\alpha}} - \frac{1}{\alpha}\right\}.$$

Якщо $\alpha \ll 1$, то $L(\alpha) \sim \frac{\alpha}{3}$; якщо $\alpha \gg 1$, тоді $L(\alpha) \to 1$.

За $T \approx 300$ К умова $\alpha = 1$ може справджуватись лише в дуже сильних магнітних полях.

У дуже сильних магнітних полях або за дуже низьких температур, коли більша частина векторів $\vec{p}_{ma}$ «орієнтована» вздовж напрямку індукції магнітного поля, спостерігається *явище насиченості*: намагніченість J не залежить від $\vec{B}$.

Далеко від області насиченості $L(\alpha) \sim \frac{\alpha}{3}$, тоді

$$\vec{J} = \frac{np_{ma}^2}{3kT}\vec{B} = \chi\frac{\vec{B}}{\mu_0}.$$

Магнітна сприйнятливість парамагнітних речовин

$$\chi = \frac{\mu_0 np_{ma}^2}{3kT}.$$

Замінивши n числом Авогадро N_A, отримаємо вираз для кіломолярної сприйнятливості:

$$\chi_{км} = \frac{\mu_0 N_A p_{ma}^2}{3kT}. \qquad (6.34)$$

Отже, $\chi_{км} > 0$. Якщо вважати, що $p_{ma} \approx 0{,}2 \ldots 0{,}4\ \mu_Б$, тоді значення $\chi_{км} \sim 10^{-7} \ldots 10^{-6}$ м^3/кмоль.

Макроскопічно парамагнетизм виявляється в тому, що парамагнетики втягуються в неоднорідне магнітне поле, а в однорідному полі парамагнітний стрижень орієнтується паралельно до ліній індукції магнітного поля (рис. 6.13).

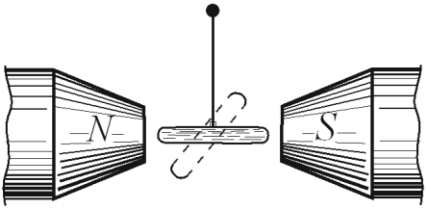


Рис. 6.13. Ампула з парамагнітним розчином хлористого заліза в магнітному полі

У разі нагрівання парамагнетика, який внесений у зовнішнє магнітне поле, тепловий рух атомів зростатиме й руйнуватиме ту орієнтацію елементарних магнітних моментів частини атомів, яка встановилася під дією зовнішнього поля. Отже, магнітна сприйнятливість парамагнетиків χ як величина, що характеризує з макроскопічного погляду магнітні властивості речовини, має залежати від температури T:

$$\chi_{км} = C/T, \qquad (6.35)$$

де $C = \frac{\mu_0 N_A P_{ma}^2}{3k}$ – стала Кюрі.

Співвідношення (6.35) виражає ***закон Кюрі***: *магнітна сприйнятливість парамагнетиків обернено пропорційна до його термодинамічної температури.*

Квантова теорія парамагнетизму враховує те, що можливі тільки дискретні орієнтації магнітного моменту атома $\vec{p}_{ma}$ відносно індукції поля $\vec{B}$. Вона дає для $\chi_{км}$ вираз, аналогічний (6.34).

6.7. Феромагнетизм

Речовини, які здатні добре намагнічуватися, називають феромагнетиками (від лат. ferrum – залізо). Їх намагніченість J велика і в 10^{10} кількість разів перевищує намагніченість діа- і парамагнетиків, які належать до категорії слабомагнітних речовин.

Магнітна проникність більшості феромагнетиків за звичайних температур вимірюється багатьма сотнями і тисячами одиниць, а в деяких спеціально виготовлених та оброблених феромагнетиків вона досягає мільйона.

Феромагнетики, крім здатності добре намагнічуватися, мають низку властивостей, які істотно відрізняють їх від діа- і парамагнетиків.

Крива намагнічування. Характерною особливістю феромагнетиків є складна нелінійна залежність між індукцією B і напруженістю поля H. Цю залежність було встановлено на прикладі м'якого (відпаленого) заліза. Залежність індукції B від напруженості магнітного поля H у феромагнетиках має вигляд, показаний на рис. 6.14. Індукція спочатку швидко збільшується, але в міру намагнічування магнетика її зростання сповільнюється.

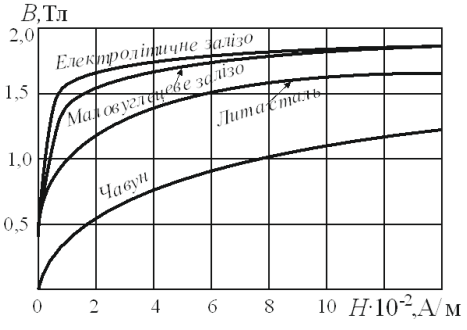


Рис. 6.14. Залежність магнітної індукції від напруженості магнітного поля

За значеннями індукції B і поля H можна знайти намагнічення магнетика $J = B/\mu_0 - H$ (магнітний момент одиниці об'єму). Характер залежності J від H для феромагнетиків зображено на рис. 6.15.

Намагнічення J, подібно до індукції, спочатку швидко зростає, але потім настає *магнітне насичення*, за якого намагнічення досягає деякого максимального значення $J_{нас}$ і фактично перестає залежати від напруженості поля.

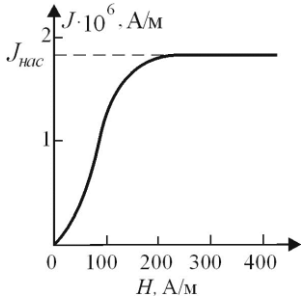


Рис. 6.15. Крива намагнічення феромагнетика

Унаслідок нелінійної залежності B від H магнітна проникність $\mu = B/(\mu_0 H)$ *залежить від напруженості магнітного поля.* Крива залежності μ від H (рис. 6.16) зростає зі збільшенням поля від початкового значення $\mu_п$ до деякої максимальної величини μ_m, але потім, після проходження через максимум, зменшується й асимптотично прямує до значення, дуже близького до одиниці.

Магнітна сприйнятливість феромагнетика $\chi = J/H$ також не є сталою і залежить від напруженості поля. Цю залежність зображено на рис. 6.16. Магнітна сприйнятливість χ, як і проникність μ, має максимум, і за великих полів асимптотично прямує до значення, близького до нуля.

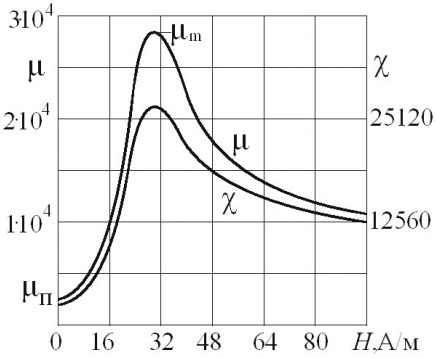


Рис. 6.16. Залежність магнітної проникності μ *і магнітної сприйнятливості* χ *від напруженості поля для заліза армко*

Особливості намагнічування розглядуваних феромагнетиків показують, що використання феромагнетиків для отримання потужних магнітних полів дуже ефективне в

областях намагнічування, які далекі від насичення. У випадку дуже потужних полів настає магнітне насичення, і магнітна проникність значно спадає.

Феромагнетики застосовують для отримання полів, напруженість яких не перевищує приблизно 10^6А/м. Якщо потрібне ще більш потужне поле, то застосовувати феромагнетики фактично неефективно. Так, наприклад, у дослідах поля з напруженістю до $3 \cdot 10^7$А/м були добуті тільки за допомогою котушок зі струмом без феромагнетиків.

Анізотропія намагнічування

Усі феромагнетики в магнітному розумінні є *анізотропними.* Утім, якщо феромагнетик має дрібнокристалічну структуру й окремі кристалики в ньому розміщені цілком безладно, то ця анізотропія не проявляється і його намагнічення не залежить від напрямку поля. Коли феромагнетик є єдиним кристалом, то вигляд кривої намагнічення буде різний і залежатиме від напрямку намагнічувального поля відносно осей кристала. На рис. 6.17 зображено елементарну комірку кристала заліза (центрований куб) і показано кристалографічні напрямки, які позначено символами [100] (ребро куба), [110] (діагональ грані) і [111] (просторова діагональ). Для кожного феромагнетика є напрямок, в якому намагнічення (при цьому полі) найбільше (*напрямок легкого намагнічування* [100]), і напрямок, в якому намагнічення найменше (*напрямок важкого намагнічування* [111] (рис. 6.17)).

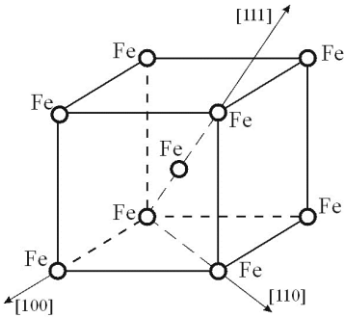


Рис. 6.17. Елементарна кристалічна комірка заліза і її основні кристалографічні напрямки: [100] – напрямок легкого намагнічування; [111] – напрямок важкого намагнічування

Гістерезис. Припустимо, що ми спочатку намагнічуємо ненамагнічений феромагнетик і, помістивши його всередину намагнічувальної котушки, збільшуємо магнітне поле всередині магнетика від нуля до деякого значення H_1 (рис. 6.18). Значення індукції в магнетику визначатимемо кривою індукції $01A$ і зобразимо відрізком ординати $0B_1$. Якщо тепер знову зменшувати магнітне поле, то зменшення індукції зображатимемо вже не відрізком кривої індукції $1O$, а кривою $1B'$, і коли поле знову дорівнюватиме нулю,

індукція не дорівнюватиме нулю, а виражатиметься відрізком OB'.

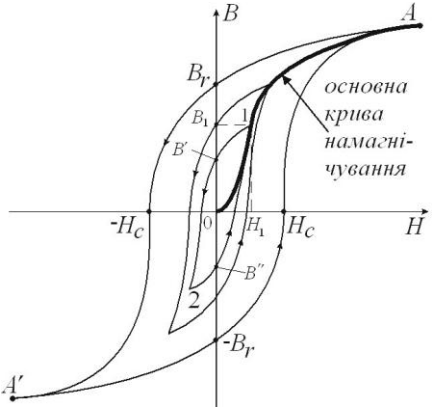


Рис. 6.18. Магнітний гістерезис

Феромагнетик у цьому стані є *постійним магнітом.* Далі, якщо змінити напрямок струму в намагнічувальній котушці й перемагнічувати зразок у зворотному напрямку, то зміна індукції описуватиметься відрізком кривої $B'2$. Під час наступної зміни поля у зворотному напрямку індукція змінюватиметься згідно з кривою $2B''1$. Під час циклічного перемагнічування феромагнетика зміну індукції в ньому зображатимуть петлеподібною замкненою кривою $1B'2B''1$.

Видно, що значення індукції у феромагнетику визначається не тільки існуючим магнітним полем, а ще й залежить від попередніх станів намагнічування, причому відбувається своєрідне *відставання* зміни індукції від зміни напруженості поля. Це явище називають магнітним *гістерезисом,* а зазначену вище петлеподібну криву залежності B від H під час циклічного перемагнічування називають *петлею гістерезису*. Магнітний гістерезис подібний до діелектричного гістерезису в сегнетоелектриках (див. підрозд. 2.6).

З кривих (рис. 6.18) видно, що під час усунення намагнічувального поля феромагнетик зберігає залишкове намагнічення, причому всередині магнетика є деяка залишкова індукція. У разі збільшення амплітуди намагнічувального поля індукція прямує до граничного значення B_r за $H = 0$ (рис. 6.18). Щоб «знищити» це залишкове намагнічення, всередині феромагнетика слід створити певне поле, напрямлене протилежно початковому намагнічувальному полю, яке зображено відрізком OH_C. Це поле називають *затримуючою*, або *коерцитивною* силою феромагнетика. Коерцитивна сила характеризує властивість феромагнетика зберігати намагніченість. Матеріали з великою коерцитивною напруженістю дають широку петлю гістерезису і їх називають «твердими» магнітними матеріалами. З них виготовляють постійні магніти. «М'які» магнітні матеріали мають малу коерцитивну силу і дають вузьку петлю гістерезису. Ці

матеріали використову-ють для виготовлення осердь трасформаторів.

З цим пов'язаний практичний спосіб, який використовують для *розмагнічування* феромагнетиків. Для цього феромагнетик поміщають всередину котушки, яка живиться змінним струмом, і величину струму поступово зменшують до нуля. Водночас феромагнетик піддається багаторазовому циклічному перемагнічуванню, що відповідає різним петлям гістерезису, які, поступово зменшуючись, стягуються до точки 0 (рис. 6.18), де намагнічування дорівнює нулю. Гістерезис значною мірою залежить від складу феромагнетика і від його обробки.

Для чистого «м'якого» заліза, тобто відпаленого і потім повільно охолодженого, гістерезис виражений дуже слабко і петля гістерезису дуже вузька. Утім, у загартованої сталі гістерезис значний (рис. 6.19).

Температура Кюрі. Здатність пара- і феромагнетиків намагнічуватися різна за різних температур, тобто їх магнітна сприйнятливість залежить від температури. Вона зменшується з підвищенням температури. Навпаки, магнітна сприйнятливість діамагнетиків фактично не залежить від температури.

Для багатьох *парамагнітних* речовин зміна χ з температурою підпорядковується закону, встановленому Кюрі (6.35). Магнітна сприйнятливість таких речовин монотонно змінюється зі зміною температури. Такі речовини називають *нормальними парамагнетиками*.

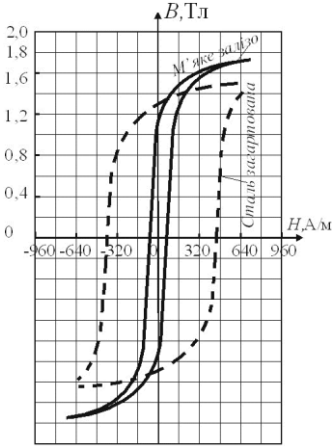


Рис. 6.19. Криві гістерезису м'якого заліза і загартованої сталі

Залежність магнітної сприйнятливості від температури для феромагнетиків має складніший характер. У разі підвищення температури здатність феромагнетиків намагнічуватися зменшується. Водночас зменшуються значення їх магнітної сприйнятливості та проникності за будь-якого значення магнітного поля, ослаблюється гістерезис і зменшується намагнічення насичення $J_{нас}$. За деякої температури T_K, яку називають *температурою Кюрі*,

феромагнітні властивості зникають зовсім. Температура Кюрі різна для різних феромагнетиків, її значення для деяких речовин наведено в табл. 6.1. За температур, вищих від температури Кюрі, феромагнетик перетворюється на парамагнетик. Залежність магнітної сприйнятливості χ від температури для таких парамагнетиків підпорядковується закону Кюрі–Вейсса, який має такий вигляд:

$$\chi = \frac{C}{T - T_K}. \qquad (6.36)$$

де C – стала, яка залежить від виду речовини, а T_K – температура Кюрі.

Таблиця 6.1

Температура Кюрі

Речовина	T_K, °C
Гадоліній	17
Залізо	770
Кобальт	1150
Нікель	360
30 % - й пермалой (сплав 70 % Fe, 30 % Ni)	70
78 % - й пермалой	550

Як видно з табл. 6.1, температура Кюрі для деяких сплавів дуже низька, наприклад для 30-процентного пермалою – близько 70 °С, тому, коли пластинку із пермалою піднести до полюсів постійного магніту за кімнатної температури, то вона притягується до магніту і втримує деякий вантаж. Утім, коли магніт із пластинкою помістити у воду, нагріту до температури 80... 90 °С, то пластинка втрачає феромагнітні властивості, її намагнічення фактично зникає і вона під дією вантажу відривається від магніту (рис. 6.20).

Робота під час намагнічування

Із розрахунку видно, що робота dw, яка потрібна для збільшення індукції на малу величину dB в одиниці об'єму магнетика,

$$dw = HdB. \qquad (6.37)$$

Оскільки робота намагнічування залежить від процесів всередині магнетика, то вираз (6.37) буде справедливий, очевидно, для магнетика будь-якої форми.

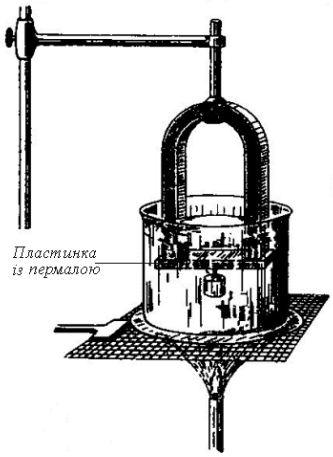


Рис. 6.20. У разі намагнічування пластинки з пермалою вище температури Кюрі вона перестає притягуватися до магніт

Розглянемо спочатку магнетик без гістерезису, для якого пряма і

зворотна гілки кривої намагнічення збігаються (рис. 6.21).

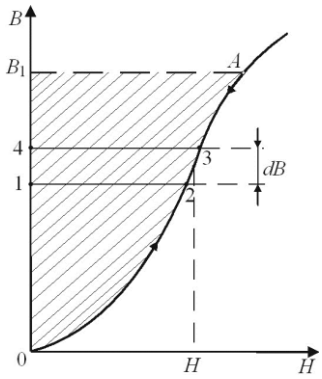


Рис. 6.21. Робота намагнічування магнетика без гістерезису

Елементарна робота, яка потрібна для нескінченно малого збільшення намагнічення, на цьому графіку виражена згідно з формулою (6.37) величиною площі *1 2 3 4 1*. Повна робота, яка затрачається у разі збільшення індукції від нуля до заданого значення B_1, дорівнює площі $0AB_10$, обмеженої кривою намагнічення і віссю B (заштрихована площа). Цю роботу виконує джерело під час намагнічування магнетика. У разі розмагнічування магнетика в коло джерела повертається енергія, яка запасена в магнітному полі. Вона, як і раніше, виражається величиною площі, обмеженої кривою намагнічення і віссю *B*. Якщо гістерезису немає, то обидві гілки кривої збігаються, а в разі розмагнічування повертається та сама робота, яка була затрачена під час намагнічування.

Інші процеси властиві для магнетика з гістерезисом. У цьому разі зі збільшенням індукції від значення B_1 до деякого іншого значення B_2 (рис. 6.22) потрібна буде робота, яка дорівнює площі, що обмежена гілкою кривої намагнічення *1*, тобто площі $B_1a1bB_2B_1$, а у разі розмагнічування до вихідного стану робота, яка повертається, дорівнюватиме площі $B_2b2aB_1B_2$, яка має меншу величину, тому під час повного циклу перемагнічування у кожну одиницю об'єму магнетика вводиться енергія

$$w = S, \qquad (6.38)$$

де S – площа петлі гістерезису, яка має розмірність одиниць роботи (енергії). Ця енергія витрачається на виконання роботи проти коерцитивних сил у магнетику і у підсумку перетворюється в тепло, тому феромагнетики під час циклічного перемагнічування нагріваються, й тим більше, чим сильніше виражений гістерезис.

В існуванні тепла гістерезису можна впевнитися за допомогою простого досліду. Якщо всередину дротяної котушки, яку живить технічний змінний струм, опустити сталевий стрижень, то він виконуватиме 50 циклів

перемагнічування за секунду і вже через 1–2 хв досліду значно нагріється.

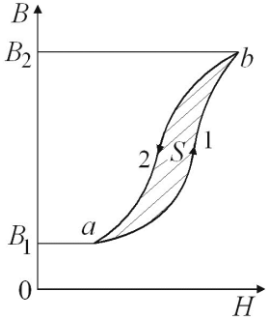


Рис. 6.22. Робота під час циклічного перемагнічування пропорційна площі петлі гістерезису

Тепло гістерезису завжди враховують під час розрахунку різних електричних пристроїв змінного струму, які містять феромагнетики, що зазнають періодичного перемагнічування. Такими є, наприклад, залізні осердя трансформаторів та обертові залізні якорі динамомашини. Наявність гістерезису в них призводить до некорисної витрати частини енергії на тепло гістерезису і зменшує коефіцієнт корисної дії установок, тому для таких пристроїв і застосовують спеціальні сорти м'якого заліза (трансформаторне залізо), в яких гістерезис виражений слабко.

6.8. Магнітні матеріали. Ферити

У сучасній електротехніці феромагнетики відіграють значну роль. Використовуючи феромагнетики, ми змушуємо елементарні струми брати участь в утворенні магнітного поля і, так би мовити, «безкоштовно» збільшуємо магнітне поле в сотні й тисячі разів порівняно з полем одних намагнічувальних котушок. Залежно від призначення феромагнетиків до них висувають різні вимоги. Так, наприклад, для застосування у трансформаторах найважливішими вимогами є висока магнітна проникність і слабкий гістерезис («магнітно м'які» матеріали, табл. 6.2), а для виготовлення постійних магнітів – залишкове намагнічення і велика коерцитивна сила (табл. 6.3).

Як феромагнітні матеріали тепер широко застосовують залізо та його сплави з іншими елементами. Підбираючи склад сплавів і варіюючи їх обробку, можна дістати різні феромагнітні матеріали, які різняться надзвичайною різноманітністю магнітних властивостей. У табл. 6.2 і 6.3 наведено магнітні характеристики деяких речовин, які застосовують в сучасній техніці. Наведені дані показують, що у виготовленні магнітних матеріалів досягнуто значних успіхів.

Деякі сплави (алніко, магніко) відзначаються дуже високими значеннями коерцитивної сили і залишкової індукції, і тому дають змогу виготовляти виняткові за якістю постійні

магніти, які широко застосовують і тепер у магнітоелектричних вимірювальних приладах й інших пристроях, у яких потрібне потужне постійне магнітне поле.

Таблиця 6.2

«Магнітно м'які» матеріали

Матеріали магніто - м'які	Склад (%), останнє залізо і домішки	Відносна початкова проникність	Відносна максим. проникність	Коерцитивна сила, А/м	Залишкова індукція,Тл
Залізо	3,3 Si	600	10000	16	2,00
Залізо чисте	0,05 (дом.)	10000	200000	4	2,15
Залізо технічне	0,20 (дом.)	150	5000	80	2,15
Сталь м'яка	0,2 C	120	2000	140	2,12
Пермалой	78,5 Ni	8000	100000	4	1,08
Нікель	99 Ni	110	600	56	0,61
Ферит	Ni-Zn	1000	2000	8	0,25

Дуже цікавими є магнітні властивості рідкоземельних елементів. Серед цієї групи елементів і їх хімічних сполук є діамагнетики, парамагнетики і феромагнетики. З них найсильнішим феромагнетиком є гадоліній. Він намагнічується до насичення в порівняно слабких полях і його намагнічення насичення більше, ніж у чистого заліза.

Дослідження магнітних сплавів і хімічних сполук привели і до важливих, і до цікавих результатів, які відкривають нові технічні можливості застосування магнітних матеріалів. Так, наприклад, було виявлено, що деякі сплави з неферомагнітних елементів за певного співвідношення між компонентами мають великий феромагнетизм. Такими є сплави «марганець – вісмут», «марганець – сурма», «хром – телур» тощо.

Таблиця 6.3

Магнніто-жорсткі матеріали

Матеріали магніто - жорсткі	Склад(%), останнє залізо	Коерцитивна сила, А/м	Залишкова індукція,Тл
Сталь вуглецева	0,9C, 1Mn	4000	1,00
Сталь вольфрамова	0,4C, 6W	5200	1,05
Сталь кобальтова	1,0C, 3Co, 4Cr, 0,4Mn	6400	1,00
Алніко	19Ni, 10Al, 18Co, 3Cu	52000	0,90
Платина – залізо	78Pt	120000	0,60
Платина - кобальт	77Pt, 23Co	320000	0,50
Магніко	13,5Ni, 9Al, 24Co, 3Cu	56000	1,30

Інше важливе досягнення полягає в добуванні й дослідженні *феритів*, які є феромагнітними хімічними сполуками типу $MeO{\cdot}Fe_2O_3$, де Me – один (або суміш) двовалентних катіонів Mn, Co, Ni, Cu, Mg, Zn, Cd, Fe^{+2}. На відміну від заліза та інших феромагнітних металів, ферити є магнітними напівпровідниками і мають великий питомий електричний опір порядку $1\ldots10^4$ Ом · м. Цим і зумовлене велике технічне значення феритів. Феромагнітні метали не можна використовувати в радіотехніці високих частот через їх велику електропровідність і великі втрати на вихрові струми, які виникають при цьому. Ферити цього недоліку не мають і дають змогу по-новому розв'язувати низку завдань радіотехніки.

6.9. Пояснення феромагнетизму

Сучасна теорія феромагнетиків у своєму розвитку ґрунтувалася на таких основних дослідних даних.

По-перше, ми бачили в підрозд. 6.7, що в деяких феромагнетиків можна змінювати намагнічення від початкового нульового значення до величезного значення насичення під дією мізерного намагнічувального поля. Це є характерним для феромагнетиків і різко відрізняє їх від парамагнетиків. Зазначимо для порівняння, що нормальна парамагнітна сіль, наприклад $FeS0_4$, за кімнатної температури під дією поля в 10 А/м збільшує своє намагнічення приблизно на 10^{-9} Тл, тобто на величину у сотні мільйонів разів меншу, ніж деякі магнітно м'які феромагнітні сплави.

Друга особливість стосується величини магнітного моменту атомів феромагнітних речовин. Прямі досліди показують, що магнітні моменти атомів феромагнітних речовин мають такий самий порядок величини, що й атоми парамагнетиків, і їх вимірюють не багатьма магнетонами Бора. Звідси випливає, що феромагнетизм не можна пояснити за допомогою теорії, подібної до теорії парамагнетизму, і що феромагнітні властивості не зумовлені наявністю магнітного моменту атома загалом.

Третій важливий дослідний випадок пов'язаний з величиною гіромагнітного відношення (див. підрозд. 6.2). У феромагнетиках воно приблизно у два рази більше, ніж очікуване теоретичне значення для електронних орбіт, і відповідає відношенню власних магнітного і механічного моментів електрона (формула (6.8)). Ця умова вказує, що намагнічування феромагнетиків зумовлюється дуже значним орієнтуванням *власних магнітних моментів електронів* (електронних спінів), але не магнітних моментів атомів загалом. За сучасними уявленнями, які надійно обґрунтовані на досліді, суть феромагнетизму полягає в тому, що значне орієнтування елементарних магнітних моментів виникає у феромагнетиках *незалежно від зовнішнього магнітного поля,* оскільки феромагнетик намагнічений до насичення, яке відповідає цій температурі, уже без будь-якого поля (рис. 6.23, *а*).

Наявність такого *самовільного*, або *спонтанного, намагнічування* є найбільш характерною властивістю феромагнетиків. Спонтанне намагнічування виникає через те, що на елементарні магнітні моменти діють сили молекулярної взаємодії немагнітного

походження. Ці уявлення висловив у працях П. Вейсс у 1907 р.

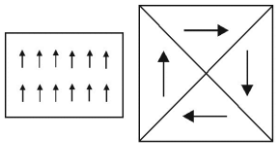
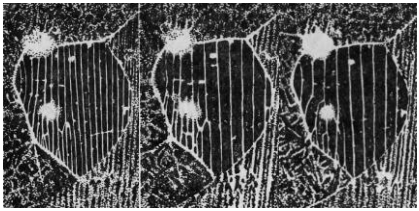


Рис. 6.23. Схематичне зображення: а – розміщення електронних спінів під час спонтанного намагнічування (низька температура); б – напрямки намагнічення в окремих доменах

Дослід показує, що феромагнетики без впливу зовнішнього поля можуть бути і ненамагнічені. Для пояснення цієї позірної суперечності Вейсс вивів другу основну гіпотезу, за якою феромагнетик розділяється на велику кількість малих (але макроскопічних) областей, або *доменів* (визначення поняття «домен» наведено в підрозд. 2.3). Кожна з цих областей за температур, нижчих за температуру Кюрі, намагнічена значною мірою, але напрямки намагнічування в окремих доменах різні, а саме такі, що повний магнітний момент феромагнетика дорівнює нулю (рис. 6.23, *б*).

Питання про фізичну причину спонтанного намагнічування принципово розв'язали в 1928 р. Френкель, а згодом Гайзенберг, які показали, що значне орієнтування електронних спінів спричиняється силами «*обмінної взаємодії*». Наявність цього нового класу сил, який не пояснюється у класичній фізиці, було з'ясовано тільки з розвитком квантової фізики атома. Існування доменів у феромагнетиках тепер доведено різними дослідами. Найбільш прямий метод полягає в добуванні «*порошкових фігур*». Якщо на добре відполіровану поверхню феромагнетика помістити шар рідини, в якій є завислі найдрібніші крупинки феромагнітного порошку (наприклад, Fe_2O_3), то ці крупинки осідатимуть переважно на ті місця, поблизу яких магнітне поле неоднорідне. Утім, саме поблизу меж доменів і виникають неоднорідності поля, і тому осілий порошок позначить межі областей спонтанного намагнічування. На рис. 6.24 подано фотографії порошкових фігур, які видно в мікроскоп під час невеликого збільшення.



а *б* *в*

Рис. 6.24. Домени у деформованому кременистому залізі: а – без поля; б – магнітне поле перпендикулярне до площини рисунка; в – магнітне поле тієї самої напруженості, але протилежного напрямку

Особливо слід зазначити, що домени справді спостерігаються і коли

немає впливу зовнішнього магнітного поля. Порошкові фігури були використані в численних дослідженнях феромагнетиків і дали змогу визначити розміри, форму й розміщення доменів, а також зміни доменів у зовнішньому магнітному полі.

Розглянемо тепер детальніше, чому у феромагнетиках виникають численні домени. Для простоти вважатимемо, що феромагнетик дуже анізотропний і що його вісь легкого намагнічування перпендикулярна до поверхні зразка. На рис. 6.25, *а* зображено феромагнетик, який складається з одного домена. У цьому випадку в зовнішньому просторі виникає магнітне поле, яке містить певну енергію. На рис. 6.25, *б* є два домени із протилежним напрямком намагнічування.

Зовнішнє магнітне поле тут зменшується зі збільшенням відстані швидше, ніж у випадку *а*, а енергія, яка є в полі, менша. У випадку, зображеному на рис. 6.25, *в*, магнітне поле фактично існує тільки безпосередньо близько від поверхні магнетика й енергія поля ще швидше зменшується. На рис. 6.25, *г* зображено такий випадок, коли в зовнішньому просторі магнітного поля немає зовсім. Тут є «замикаючі» домени у формі тригранних призм, бічні поверхні яких скрізь утворюють кут 45° з вектором намагнічення. Унаслідок цього магнітний потік проходить виключно всередині феромагнетика, він замикається граничними доменами, чим і зумовлена їх назва – «замикаючі домени». Стан, зображений на рис. 6.25, *г*, енергетично вигідніший, ніж попередні стани, і тому феромагнетик, який перебуває, наприклад, у стані, зображеному на рис. 6.25, *а*, прагнутиме перейти у стан, зображений на рис. 6.25, *г.* Нарешті, на рис. 6.25, *д* зображено сукупність доменів разом з доменами, що їх замикають, у якій також немає зовнішнього потоку.

Така форма доменів справді спостерігається під час проведення досліду. Отже, феромагнетик на домени розділяється тому, що під час утворення доменних структур енергія феромагнетика зменшується.

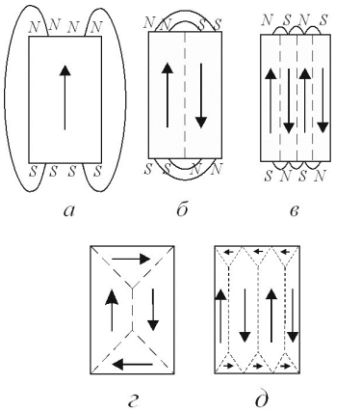


Рис. 6.25. Пояснення виникнення доменів

Вище ми розглядали феромагнетик без дії зовнішнього магнітного поля і враховували тільки його власну магнітну енергію. Якби були наявні ще й інші джерела енергії, наприклад механічні напруги або зовнішнє магнітне поле, то форма доменів визначалася б повною енергією системи. Отже, за наявності механічних напруг і зовнішнього поля доменна структура змінюється. У разі виникнення спонтанного намагнічування (орієнтування електронних спінів) магнетик деформується. Якщо за температури, вищої за температуру Кюрі, вирізати з монокристала феромагнетика кулю, то під час охолодження нижче температури Кюрі куля перетвориться в еліпсоїд. Форма і розміри доменів змінюються і під час перемагнічування, тому феромагнетик загалом не намагнічений і в разі намагнічування деформується.

Явище деформації під час намагнічування відкрив у середині минулого століття Джоуль – його називають *магнітострикцією.* Деформації, які виникають при цьому, дуже малі: відносні видовження зразка $\Delta l/l$ у полях приблизно 10^5 А/м звичайно мають порядок $10^{-5}...10^{-6}$. Магнітострикцію використовують, як і п'єзоелектричний ефект (див. підрозд. 2.6), для побудови потужних випромінювачів ультразвукових хвиль та інших цілей.

Сили обмінної взаємодії спричиняють у феромагнетиках паралельне орієнтування електронних спінів, однак обмінні сили залежать від структури тіла, і тому характер орієнтування спінів, який вони спричиняють, може бути різний. Виявляється, що є речовини, в яких також виникає значне орієнтування електронних спінів, але, на відміну від феромагнетиків, електронні спіни орієнтовані в них *попарно антипаралельно.* У найпростішому випадку електронні спіни утворюють нібито дві просторові підрешітки вставлені одна в одну і намагнічені у протилежних напрямках (рис. 6.26, *а*). Речовини, в яких намагнічення обох підрешіток однакове за величиною, називають *антиферомагнетиками.* Їх наявність теоретично передбачив Л. Д. Ландау в 1933 р. Антиферомагнетиками є деякі сполуки марганцю (MnO, MnS), хрому (NiCr, Cr_20_3), ванадію (VO_2) та ін. Такі речовини за низьких температур мають дуже малу магнітну сприйнятливість.

У разі підвищення температури точна попарна антипаралельність електронних спінів порушується і

магнітна сприйнятливість збільшується. За деякої температури (*антиферомагнітна точка Кюрі*, або *точка Неєля*) області спонтанного орієнтування електронних спінів руйнуються, а антиферомагнетик перетворюється в парамагнетик. Якщо підвищувати температуру далі, то магнітна сприйнятливість, як у будь-якого парамагнетика, зменшується, а отже, за антиферомагнітної температури Кюрі магнітна сприйнятливість має максимум.

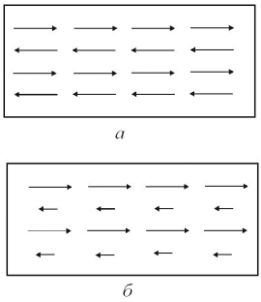


Рис. 6.26. Характер намагнічування в антиферомагнетиках (а) та у феритах (б) (схематично)

Якщо намагнічення обох підрешіток неоднакове за величиною, то виникає нескомпенсований антиферомагнетизм, і речовина може набути значного магнітного моменту. Такий характер намагнічування спостерігається у *феритах* (рис. 6.26, *б*).

6.10. Процеси намагнічування феромагнетика

Одним з основних завдань теорії феромагнетизму є пояснення технічної кривої намагнічування, тобто залежність намагнічення J від напруженості поля H, або індукції B всередині феромагнетика від H. Розглянемо полікристалічний звичайний шматок заліза. Всередині нього є велика кількість маленьких кристаликів (рис. 6.27), кристалічні осі яких направлені в різні сторони.

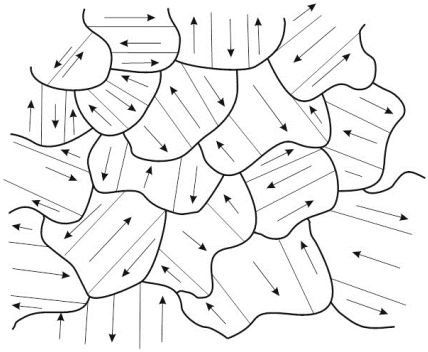


Рис. 6.27. Мікроструктура ненамагніченого полікристалічного феромагнітного матеріалу. Кожний кристалик має напрямок легкого намагнічування і розбивається на домени, які спонтанно намагнічені в цьому напрямку

Ці кристалики (це не те саме, що домени) починають зміщуватись, і домени, напрямок намагніченості яких збігається з напрямком легкого намагнічування, ростуть все більше й більше. Поки поле залишається дуже малим, цей ріст оборотний; якщо поле

виключити, намагніченість знову повернеться до нуля. Ця ділянка кривої намагнічування позначена на рис. 6.28 літерою a.

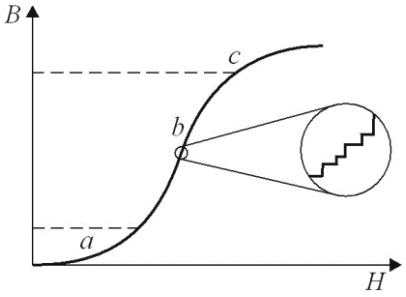


Рис. 6.28. Крива намагнічування полікристалічного заліза

Для більших полів в області, яка на рис. 6.28 позначена літерою b, все стає більш складним. У кожному маленькому кристалику матеріалу є напруги і дислокації (порушення кристалічних ґраток), домішки й дефекти. І при всіх полях, за винятком дуже слабких, стінки доменів під час руху наштовхуються на них.

Між доменною стінкою і дислокацією (або домішкою) виникає взаємодія. У результаті, коли стінка наштовхується на перешкоду, вона ніби приклеюється і тримається так, поки поле не досягне певної величини. Коли поле трохи зросте, стінка раптово зривається. Таким чином, рух доменної стінки не є плавним, як в ідеальному кристалі; вона рухається стрибкоподібно, зупиняючись на мить. Якщо розглянути криву намагнічування в мікроскопічному масштабі, то можна побачити щось подібне зображеному на вставці рис. 6.28. Найважливіше полягає в тому, що ці стрибки намагнічування викликають втрату енергії. Найперше, коли стінка домена проскакує на кінець перешкоди, вона дуже швидко рухається до наступної перешкоди. Швидкий рух спричиняє і швидку зміну магнітного поля, що, у свою чергу, створює у кристалі вихрові струми (див. наступний розділ). Останні витрачають енергію на нагрівання металу.

Інший ефект полягає в тому, що, коли домен раптово змінюється, частина кристалів через магнітострикцію змінює розміри. Кожний раптовий зсув доменної стінки створює невелику звукову хвилю, яка теж виносить енергію. Через такі ефекти ця частина кривої намагнічування є необоротною і відбувається втрата енергії. Це і є причиною гістерезисного ефекту, оскільки рух стрибками вперед – одне, а рух назад – вже інше, а під час руху в обидва кінці втрачається енергія.

Зрештою, якщо поля достатньо сильні й усі доменні стінки зсунуті, а намагніченість кожного кристалика направлена по найближчій до поля осі легкого намагнічування, залишаються ще деякі кристалики, у яких вісь легкого

намагнічування далека від напрямку зовнішнього магнітного поля. Щоб повернути ці магнітні моменти, потрібне ще й додаткове поле. Таким чином, у сильних полях саме в області, яка позначена на рис. 6.28 літерою *c*, намагніченість зростає повільно, але гладко. Намагнічення не відразу досягає свого насичення, оскільки в цій останній частині кривої відбувається до-орієнтація атомних магнітиків в сильному полі. Тепер зрозуміло, чому крива намагнічення полікристалічного матеріалу має вигляд, показаний на рис. 6.28. Такий характер залежності $B(H)$ дає можливість пояснити залежність $\mu(H)$ (див. рис. 6.16). Для цього на рис. 6.29, *а* зображено ще раз основну криву намагнічення. Від початку координат проведемо пряму лінію, яка проходить через довільну точку кривої $B(H)$. Тангенс кута нахилу цієї прямої пропорційний відношенню B/H, тобто магнітній проникності (див. формулу (6.19)) для відповідного значення напруженості поля. Зі збільшенням H від нуля кут нахилу (а отже й μ) спочатку зростає. Уточці *2* він досягає максимуму (пряма *0–2* є дотичною до кривої $B(H)$), а потім зменшується.

На рис. 6.29, *б* показано графік залежності μ від H. Із рисунка видно, що максимальне значення проникності досягається дещо раніше, ніж насичення. У разі необмеженого зростання H проникність асимптотично наближається до одиниці. Із формул (6.13) і (6.19) легко отримати вираз для μ, а саме

$$\mu = 1 + J/H. \qquad (6.37)$$

Оскільки намагнічення J не може перевищувати $J_{\text{нас}}$ (рис. 6.15), то за $H \to \infty, \mu \to 1$.

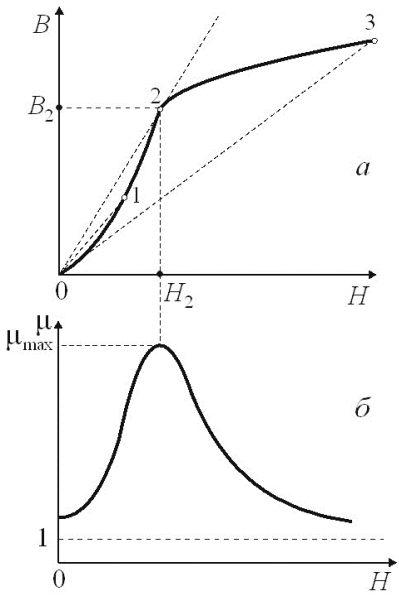


Рис. 6.29. а - основна крива намагнічення B(H); б - графік залежності відносної магнітної проникності μ *від H*

Контрольні запитання

1. Що таке процес намагнічування магнетика? Дайте визначення вектора намагнічення речовини. Який фізичний

зміст вектора намагнічування? Що таке магнітне насичення?

2. За допомогою яких фізичних величин проводять класифікацію магнетиків?

3. Провести класифікацію магнетиків. Опишіть властивості діа- та парамагнетиків. Які речовини є феромагнетиками?

5. Назвіть характеристики магнітного поля в речовині.

6. Який є зв'язок між вектором напруженості та вектором індукції магнітного поля у феромагнетиках?

7. Схематично зобразіть залежність індукції магнітного поля від напруженості магнітного поля у феромагнетиках. Поясніть цю залежність.

8. У чому суть явища магнітного гістерезису? Що таке коерцитивна сила та залишкова намагніченість?

9. Запишіть співвідношення між складовими векторів B і H на межі двох магнетиків.

10. Який фізичний зміст магнітної проникності феромагнетика? Як залежить магнітна проникність від напруженості магнітного поля і як пояснюється ця залежність?

11. Чому значення магнітної проникності для різних феромагнетиків не наводять у довідникових таблицях?

12. Які феромагнетики мають назву магніто м'яких та магніто жорстких?

13. Яка природа феромагнетизму? Що таке домени? Як вони реагують на магнітне поле?

14. Що таке точка Кюрі для феромагнетика?

15. Де в техніці зв'язку застосовують феромагнетики?

Розділ 7. Електромагнітна індукція

7.1. Явище електромагнітної індукції

У попередніх розділах було розглянуто властивості статичних, незмінних в часі електричних та магнітних полів. Переходячи в область динамічних електромагнітних явищ, ми бачимо ряд якісно нових властивостей полів та їх закономірностей. В основі електродинаміки змінних полів лежить фундаментальний закон електромагнітної індукції, який дослідно відкрив у 1831 р. видатний англійський учений М. Фарадей. Суть цього закону полягає у виникненні струму в довільному замкненому провідному контурі у разі зміни магнітного потоку, що пронизує цей контур.

Відкриття зв'язку між електрикою та магнетизмом стало визначною

подією в історії розвитку електромагнетизму, оскільки ці явища раніше вважали абсолютно взаємонезалежними. Спочатку у 1820 р. Х. Ерстед відкрив, що навколо провідників зі струмом виникає магнітне поле, а трохи пізніше в цьому ж році А. Ампер встановив, що на провідники зі струмом у магнітному полі діє сила. Результати цих досліджень стали основою припущення, що й електричне поле можна отримати за допомогою магнітного. Для перевірки цієї гіпотези було проведено різноманітні експерименти. Один із них полягає у спробі виявити струм у замкненому провіднику, розміщеному біля іншого провідника, по якому протікає постійний електричний струм. У ранніх дослідах Фарадей з подивом виявив, що постійний струм не впливає на контур, який лежав близько до нього. Він виготовляв різні провідні котушки, намотуючи два проводи так, щоб їхні витки розміщувались дуже щільно один до одного, але були розділені ізолюючим шаром тканини або паперу. Одна з котушок складала із гальванометром замкнене коло. Через іншу котушку пропускали великий струм від батареї. Фарадей був розчарований, бо стрілка гальванометра не відхилялась, але в одному з дослідів він помітив дуже слабкий рух стрілки гальванометра при вмиканні струму, й такий самий рух при вимиканні. Продовжуючи ці досліди, він швидко встановив, що струми в інших провідниках наводяться не постійним, а змінним струмом. Однією з блискучих експериментальних знахідок в дослідах Фарадея того періоду була заміна гальванометра (він зрозумів, що гальванометр був недостатньо чутливим приладом для виявлення короткочасних імпульсів струму) простою маленькою котушкою, в яку помістив ненамагнічену стальну стрілку (рис. 7.1). Фарадей виявив, що стрілка намагнічувалась імпульсом індукованого струму після вмикання первинного струму і що напрямок намагнічування був протилежним, якщо його причиною було вимикання струму у первинному контурі. Так було відкрито явище електромагнітної індукції.

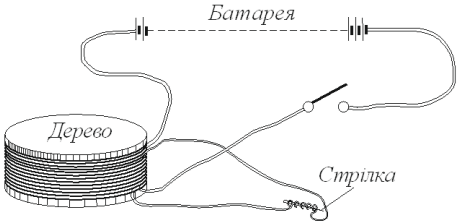


Рис. 7.1

Розглянемо деякі досліди, що ілюструють *електромагнітну* індукцію. Для цього скористаємось двома дротяними котушками *1* і *2* (рис. 7.2), одну з яких (*1*) можна надівати на іншу

(*2*). Сполучимо котушку *1* з гальванометром, а котушку *2* – із джерелом струму. Якщо котушка *1* нерухома відносно котушки *2* (тобто відносно магнітного поля), то в колі котушки *1* не буде струму, якими б сильними не були струм та магнітне поле котушки *2*. Почнемо тепер переміщувати котушку *1*. Ми побачимо, що гальванометр покаже появу струму. Цей струм існує тільки якщо котушка рухається, і він тим більший, чим швидше рухається котушка. Струм у колі котушки *1* виникає і в разі зближеня котушок, і в разі їх віддалення, однак струми в обох випадках мають протилежні напрямки. Якщо залишити котушку *1* нерухомою, а рухати котушку зі струмом *2*, то гальванометр теж показує струм під час руху котушки.

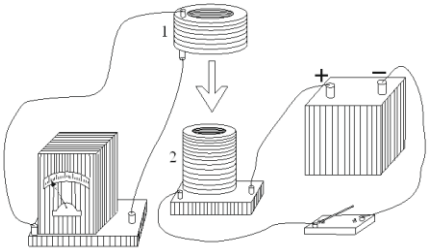


Рис. 7.2

Ми могли б залишити обидві котушки (*1* та *2*) нерухомими, але змінювати величину струму в котушці *2* за допомогою реостата. Тоді за будь-якої зміни величини струму (тобто магнітного поля котушки *2*) у колі котушки *1* виникає струм. Напрямок струму в котушці *1* в разі підсилення магнітного поля протилежний напрямку струму в разі ослаблення поля; струму в котушці *1* не виникає, якщо магнітне поле залишається постійним. Ці досліди показують, що причиною появи індукційного струму є *зміна магнітного поля.*

Те, як створюється ця зміна, значення не має. У досліді, проілюстрованому на рис. 7.3, обидві котушки (*1* та *2*) нерухомі, а в котушку *2* ми всовуємо або висуваємо з неї залізне осереддя *С*. Після всування осердя воно намагнічується, і магнітне поле посилюється; після висування осердя поле слабшає. Струм у колі котушки *1* проходить тільки за умови руху осердя. Змінне магнітне поле ми можемо створити і рухом постійного магніту. Якщо зовсім забрати котушку зі струмом *2* і всувати в котушку *1* постійний магніт (або висувати магніт з котушки), то гальванометр також показує наявність струму. Цей струм виникає й тоді, коли рухається котушка *1*, а магніт перебуває у спокої. Спостерігаючи напрямки струму у процесі зближення магніту й котушки і в разі їх віддалення, можна впевнитися, що вони протилежні, як і в попередніх дослідах.

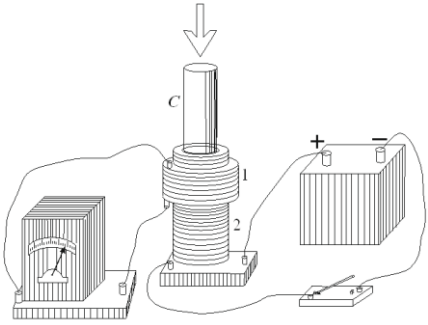


Рис. 7.3

Результати своїх численних дослідів Фарадей подав у такій наочній формі. Зображатимемо магнітне поле за допомогою ліній магнітної індукції. Тоді величина магнітної індукції поля характеризуватиметься густотою ліній індукції. Уявімо собі тепер, що замкнений провідник рухається в магнітному полі й переходить в область сильнішого поля. Тоді кількість ліній індукції, які охоплює провідник, збільшиться. Навпаки, в разі руху провідника в область слабшого поля кількість ліній індукції, які охоплює провідник, зменшиться. Утім, магнітне поле є полем *вихровим* (див. підрозд. 5.6), і його лінії індукції завжди замкнені. Унаслідок цього лінії індукції поля зчеплені з дротяним контуром як ланки ланцюга (рис. 7.4), тому будь-яка зміна кількості ліній індукції, які охоплює контур, може статися тільки внаслідок *перетину* ними дротяного контура, точно так само, коли провідник перебуває у спокої, але змінюється величина магнітної індукції. У разі посилення поля густота ліній індукції збільшуватиметься, і вони стягуватимуться одна до одної, а в разі ослабленні поля – розходитимуться одна від одної. У цьому випадку провідник перетне деяку кількість ліній індукції, тому Фарадей зробив висновок, що *індукційний струм виникає у провіднику тоді, коли провідник або яка-небудь його частина перетинає лінії магнітної індукції.* Можемо зробити такий висновок: виникнення індукційного струму завжди пов’язане зі зміною магнітного потоку крізь поверхню, яку охоплює провідник. Ці зміни можуть відбуватися з різних причин, зокрема через:

– переміщення постійного магніту відносно нерухомого провідника;

– переміщення контура відносно нерухомого магніту;

– замикання й розмикання струму в обмотці нерухомого електромагніту, розміщеного поблизу провідника;

– відносне переміщенні контура та електромагніту;

– зміну магнітної індукції поля електромагніту (виймання осердя за сталого струму в обмотці) або зміну струму реостатом;

– зміну комутатором напрямку струму в обмотці електромагніту;

– постійний рух контура в неоднорідному магнітному полі;

– обертальний рух контура в однорідному магнітному полі.

Відкриття електромагнітної індукції мало величезне наукове і технічне значення. Це явище показало, що можна не тільки створити магнітне поле за допомогою струмів, а й навпаки, мати електричні струми за допомогою магнітного поля. Цим було остаточно встановлено взаємний зв'язок між електричними і магнітними явищами. Завдяки явищу електромагнітної індукції можна перетворювати механічну енергію в електричну, а також передавати електричну енергію з одного кола в інше.

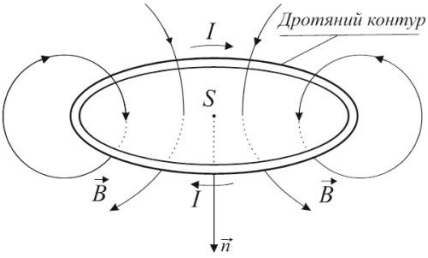


Рис. 7.4

Технічне значення цього відкриття полягає в тому, що на явищі електромагнітної індукції ґрунтується дія всіх потужних джерел струму (промислових генераторів постійного і змінного струму, трансформаторів тощо), без яких сучасний розвиток електротехніки був би неможливим.

7.2. Закон Ленца

Фарадей встановив, що напрямок індукційного струму в провіднику залежить від характеру зміни (збільшення чи зменшення) магнітного потоку ($\Delta\Phi > 0$ чи $\Delta\Phi < 0$) через його контур. Якщо після внесення постійного магніту в котушку стрілка гальванометра відхиляється в один бік, то після виймання магніту вона відхиляється в інший бік.

Загальне правило, за допомогою якого можна визначити напрямок індукційного струму в замкненому провіднику, сформулював Е. Х. Ленц: «*Індукційний струм у замкненому провіднику завжди має такий напрямок, що створений цим струмом власний магнітний потік протидіє тим змінам зовнішнього магнітного току, які збуджують індукційний струм*». Стисло закон Ленца можна подати так: *індукційний струм у всіх випадках напрямлений так, що його дія протилежна дії причини, яка зумовила виникнення цього струму.* Пояснимо закон Ленца на прикладі (рис. 7.5).

Закон Ленца справджується і в тих випадках, коли провідники нерухомі, а змінюється магнітне поле (величина

стуму). У цьому випадку індукційні струми завжди спричиняють таке поле, яке прагне протидіяти змінам зовнішнього поля, що спричинили ці струми. Так, наприклад, якщо в досліді (рис. 7.2) котушки нерухомі, то після вмикання струму в котушці *2* (його наростання) напрямок струму в котушці *1* буде протилежний (індукційний струм прагне послабити наростаюче поле котушки *2*), а після вимикання струму в котушці *2* (його зменшення) струм у котушці *1* буде напрямлений так само, як і в котушці *2*, прагнучи підтримати магнітне поле, яке послаблюється.

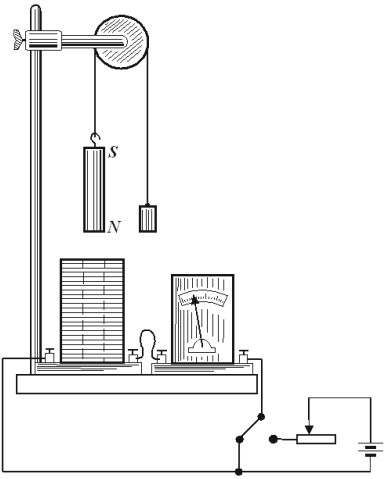


Рис. 7.5. Демонстрація закону Ленца

Закон Ленца випливає із закону збереження енергії. Справді, індукційні струми, як і будь-які електричні струми, виконують певну роботу. Це означає, що під час руху замкненого провідника в магнітному полі має бути виконана додаткова робота зовнішніх сил. Ця робота й виникає тому, що індукційні струми, взаємодіючи з магнітним полем, спричиняють сили, напрямлені протилежно рухові, тобто перешкоджають йому переміщатись.

7.3. Основний закон електромагнітної індукції

Індукційний струм I_i у замкненому провіднику, опір якого R, виникає під дією е.р.с. $\mathcal{E}_i$, яку можна виразити за законом Ома:

$$\mathcal{E}_i = I_i R.$$

Оскільки та сама е.р.с. у провідниках з різним опором створює неоднакові струми, то для кількісної характеристики явища електромагнітної індукції зручніше користуватися величиною $\mathcal{E}_i$, а не силою індукційного струму I_i.

Унаслідок численних дослідів Фарадей встановив основний кількісний закон електромагнітної індукції, однак ми розглянемо спочатку інше формулювання цього закону, яке дав згодом Максвелл.

У підрозд. 7.1 ми бачили, що індукційний струм, а отже й е.р.с. індукції, виникають тільки тоді, коли провідник *перетинає* лінії магнітної

індукції, тобто коли повна кількість ліній індукції, що проходять через площу, обмежену дротяним контуром, змінюється (рис. 7.4), але певна кількість ліній магнітної індукції, що проходять через яку-небудь поверхню, створює магнітний потік крізь цю поверхню. Звідси можна зробити висновок, що причиною виникнення е.р.с. електромагнітної індукції є зміна *магнітного потоку*. Аналізуючи результати дослідів Фарадея, Максвелл встановив, що у всіх випадках е.р.с. *електромагнітної індукції пропорційна швидкості зміни магнітного потоку через площу, обмежену контуром.* Або, інакше:

$$\mathcal{E}_i = kd\Phi/dt, \qquad (7.1)$$

де $\mathcal{E}_i$ – е.р.с. індукції; Φ – магнітний потік; t – час; k – множник пропорційності, який залежить від вибору системи одиниць. Електрорушійну силу електромагнітної індукції вважають позитивною, якщо магнітний момент $\vec{p}_m$ відповідного їй індукційного струму утворює гострий кут з лініями магнітної індукції того поля, яке наводить цей струм (рис. 7.6). Тоді у випадку, зображеному на рис. 7.6, *а*, $\mathcal{E}_i < 0$, а для зображеного на рис. 7.6, *б*, $\mathcal{E}_i > 0$. У СІ $k = 1$ та

$$\mathcal{E}_i = -d\Phi/dt. \qquad (7.2)$$

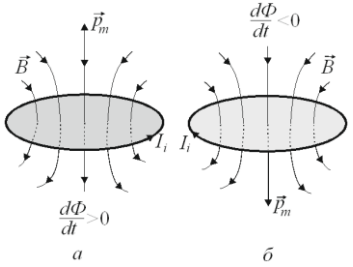


Рис. 7.6

Знак «–» є математичним виразом правила Ленца. Формула, яка об'єднує закони Фарадея і Ленца, є математичним виразом основного закону електромагнітної індукції, яку дав Максвелл: *е.р.с. електромагнітної індукції в замкненому контурі чисельно дорівнює і протилежна за знаком швидкості зміни магнітного потоку крізь поверхню, обмежену контуром*. Якщо е.р.с. індукції виникає під час зміни магнітного потоку, який пронизує котушку з N витків, то її величина буде, відповідно, в N разів більшою, ніж для одного витка, тобто

$$\mathcal{E}_i = -Nd\Phi/dt = -d\psi/dt, \qquad (7.3)$$

де $\Psi = N\Phi$ – повний магнітний потік через котушку, який називають *потокозчепленням*.

На основі закону електромагнітної індукції можна дати визначення одиниці магнітного потоку вебера: 1 Вб – це такий магнітний потік, зі зменшенням

якого до нуля протягом 1 с в колі, яке він пронизуває, виникає е.р.с. індукції в 1 В.

Формула (7.2) виражає закон електромагнітної індукції в загальній формі. Вона застосовна як до нерухомих контурів, так і до рухомих провідників у магнітному полі. Похідна від магнітного потоку за часом, що входить до неї, у загальному випадку складається із двох частин, одна з яких зумовлена зміною магнітної індукції з часом, а друга – рухом контура відносно магнітного поля (або його деформацією).

7.4. Пояснення закону електромагнітної індукції

7.4.1. Виведення закону електромагнітної індукції із закону збереження енергії

Німецький фізик Гельмгольц показав, що основний закон електромагнітної індукції (закон Фарадея) є наслідком закону збереження енергії.

Помістимо в однорідне магнітне поле з індукцією $\vec{B}$ провідник *CD* завдовжки *l* з опором *R*, який є ділянкою замкненого кола *CDKLC* (рис. 7.7). Опором з'єднувальних провідників *DK*, *KL* і *LC* нехтуємо. Провідник *DC* рухається без тертя по ділянках *DK* і *LC*. Джерело постійного струму з е.р.с. $\mathcal{E}$ і внутрішнім опором *r* створює в замкненому колі за законом Ома струм $I = \frac{\mathcal{E}}{R+r}$. При цьому на провідник діє сила Ампера (5.7) $F = BIl$, унаслідок чого провідник переміщуватиметься прискорено і за час *dt* пройде шлях *dh*. Під час проходження струму *I* в замкненому колі *CDKLC* протягом часу *dt* відбуваються такі процеси:

а) виконується робота з подолання електричного опору $R + r$ кола:

$$dA_1 = I^2(R + r)dt;$$

б) виконується механічна робота при переміщенні провідника на шляху *dh*:

$$dA_2 = Fdh = IBld = IBdS = Id\Phi.$$

За законом збереження енергії сума робіт dA_1 і dA_2 має дорівнювати повній роботі джерела струму:

$$dA = \mathcal{E}dq = \mathcal{E}Idt,$$

тобто

$$I^2(R + r)dt + Id\Phi = \mathcal{E}Idt.$$

Звідси:

$$I = \frac{\mathcal{E} - \frac{d\Phi}{dt}}{R + r}.$$

Ця формула визначає закон Ома для замкненого кола, яке поміщене в магнітне поле. Роль діючої е.р.с. у такому колі відіграє вираз

$$\mathcal{E} - \frac{d\Phi}{dt} \quad \text{або} \quad \mathcal{E} + \left(-\frac{d\Phi}{dt}\right).$$

Додатковий член $\left(-\frac{d\Phi}{dt}\right)$, який відрізняється від нуля тільки у разі зміни магнітного потоку через поверхню *dS*,

повинен мати розмірність е.р.с. Він і визначає е.р.с. індукції $\mathcal{E}_i$:

$$\mathcal{E}_i = -\frac{d\Phi}{dt}.$$

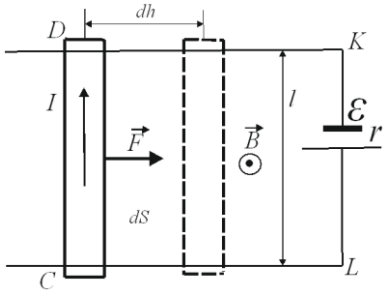


Рис. 7.7

7.4.2. Пояснення закону електромагнітної індукції за допомогою сили Лоренца

Розглянемо контур з рухомою перемичкою 1–2 (рис. 7.8). Розмістимо його в однорідне магнітне поле, яке перпендикулярне до площини контура Г і направлене за межі рисунка.

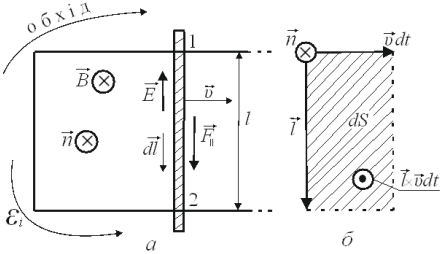


Рис. 7.8

Надамо перемичці рух із швидкістю $\vec{v}$. З тією ж швидкістю відносно поля стануть переміщуватись і носії струму в перемичці – електрони. У результаті на кожний електрон почне діяти направлена вздовж перемички магнітна сила

$$\vec{F}_\parallel = -e\vec{v} \times \vec{B} \qquad (7.4)$$

(див. підрозд. 5.10, формула (5.47); заряд електрона дорівнює $-e$). Дія цієї сили еквівалентна дії на електрон електричного поля $\vec{E}$ (рис. 7.8, *а*), напруженість якого

$$\vec{E} = \vec{v} \times \vec{B}. \qquad (7.5)$$

Це поле не електростатичного походження. Згідно з формулою (4.25) його циркуляція по контуру Г дає величину е.р.с., яка індукується в контурі:

$$\mathcal{E}_i = \oint_\Gamma \vec{E}d\vec{l} = \oint_\Gamma \vec{v}\times\vec{B}d\vec{l} = \int_1^2 \vec{v}\times\vec{B}d\vec{l} \qquad (7.6)$$

(підінтегральна функція відмінна від нуля тільки на ділянці перемички 1–2, яка рухається).

Будемо вважати $\mathcal{E}_i$ позитивною в тому випадку, коли її напрямок утворює з напрямком нормалі $\vec{n}$ до контуру Г правогвинтову систему.

Виберемо обхід по контуру Г за годинниковою стрілкою, тоді нормаль $\vec{n}$ буде мати напрямок, як показано на рис. 7.8, *а*. Під час розрахунку циркуляції слід обходити контур за годинниковою стрілкою і, відповідно, вибрати напрямок вектора $d\vec{l}$. Якщо у виразі (7.6) винести постійний вектор $\vec{v} \times \vec{B}$ за знак інтеграла, то отримаємо

$$\mathcal{E}_i = \int_1^2 \vec{v} \times \vec{B}\, d\vec{l} = \vec{v} \times \vec{B} \cdot \vec{l},$$

де $\vec{l}$ – вектор, який показаний на (рис. 7.8, *б*). В отриманому виразі здійснимо циклічну перестановку векторів (множників), після чого помножимо і розділимо його на dt:

$$\mathcal{E}_i = \vec{v} \times \vec{B} \cdot \vec{l} = \vec{B}\frac{\vec{l} \times \vec{v} dt}{dt} \cdot \qquad (7.7)$$

Із рис. 7.8, *б* видно, що $\vec{l} \times \vec{v}\, dt = -\vec{n} dS$, де dS – приріст площі контура Г за час dt. За визначенням магнітного потоку вираз

$$\vec{B} d\vec{S} = \vec{B}\vec{n} dS$$

є потоком через площину dS, тобто приріст потоку $d\Phi$ через контур Г.

Тоді

$$\vec{B} \cdot \vec{l} \times \vec{v}\, dt = -\vec{B}\vec{n} dS = -d\Phi.$$

З урахуванням цього виразу (7.7) перепишемо так:

$$\mathcal{E}_i = -\frac{d\Phi}{dt}. \qquad (7.7')$$

Таким чином, ми отримали закон (7.2) за допомогою сили Лоренца. Швидкість зміни магнітного потоку $\frac{d\Phi}{dt}$ і електрорушійної сили $\mathcal{E}_i$ мають протилежні знаки. Знак зміни потоку і знак $\mathcal{E}_i$ пов'язані з вибором напрямку обходу по контуру Г (напрямком нормалі). На рис. 7.8, *а* знак $\frac{d\Phi}{dt}$ додатний, а знак $\mathcal{E}_i$ – від'ємний. Якби ми вибрали обхід проти годинникової стрілки (нормаль направлена на нас), то знак $\frac{d\Phi}{dt}$ був би від'ємним, а знак $\mathcal{E}_i$ – додатним.

Зауважимо, що в міркуваннях, які привели нас до формули (7.7′), роль сторонніх сил, які підтримували струм у контурі, відіграють магнітні сили. Робота цих сил над одиничним позитивним зарядом, яка стандартно дорівнює е.р.с., відмінна від нуля. Ця обставина начебто суперечить твердженню про те, що магнітна сила роботи над зарядами не виконує. Це протиріччя зникає, якщо врахувати, що сила (7.4) є не повною магнітною силою, що діє на електрон, а лише її складовою, що діє паралельно проводу й обумовлена швидкістю $\vec{v}$ (сила $\vec{F}_\parallel$ на рис. 7.9).

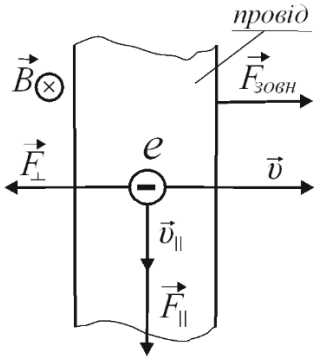


Рис.7.9. Дві складові сили Лоренца

Під дією сили $\vec{F}_\parallel$ електрон починає рухатись вздовж проводу із швидкістю $\vec{v}_\parallel$, у результаті чого виникає перпендикулярна до проводу інша

складова магнітної сили: $\vec{F}_\perp = -e\vec{v}_\parallel \times \vec{B}$ (ця складова не робить внеску в циркуляцію, оскільки перпендикулярна до $d\vec{l}$). Повна магнітна сила, яка діє на електрон,

$$\vec{F} = \vec{F}_\parallel + \vec{F}_\perp.$$

Робота цієї сили над електроном за час dt

$$dA = \vec{F}_\parallel \vec{v}_\parallel dt + \vec{F}_\perp \vec{v} dt = F_\parallel v_\parallel dt - F_\perp v dt = \\ = -evv_\parallel Bdt + evv_\parallel Bdt = 0,$$

при цьому враховано вирази для сил $\vec{F}_\parallel$ і $\vec{F}_\perp$, а також те, що напрямки векторів $\vec{F}_\parallel$ та $\vec{v}_\parallel$ однакові, а векторів $\vec{F}_\perp$ та $\vec{v}$ – протилежні.

Таким чином, робота повної магнітної сили над зарядом дорівнює нулю.

Наведене нами пояснення виникнення е.р.с. індукції належить до випадку, коли магнітне поле постійне, а змінюється геометрія контура. Утім, магнітний потік через контур може змінюватись також внаслідок зміни $\vec{B}$. У цьому випадку виникнення е.р.с. є принципово іншим і це головне у відкритті Фарадея.

7.4.3. Пояснення закону електромагнітної індукції за допомогою вихрового електричного поля

Розглянемо випадок електромагнітної індукції, коли провідний контур, в якому індукується струм, нерухомий, а зміна магнітного потоку обумовлена зміною магнітного поля. Виникнення індукційного струму свідчить про те, що зміна магнітного поля викликає появу в контурі сторонніх сил, які діють на носії заряду. Ці сторонні сили не пов'язані ні з хімічними, ні з тепловими процесами у проводі; вони не можуть бути і магнітними силами, тому що такі сили роботи над зарядами не виконують. Залишається стверджувати, що індукційний струм обумовлений електричним полем, яке виникає у проводі. Позначимо напруженість цього поля $\vec{E}_\text{в}$ (індекс «в» означає «*вихрове*»). Згідно з (4.25) циркуляція цього поля по замкненому контуру Г дорівнює е.р.с., що діє в цьому контурі, тобто

$$\mathcal{E}_i = \oint_\Gamma \vec{E}_\text{в}\, d\vec{l}. \qquad (7.8)$$

Якщо враховувати закон (7.2) і вираз $\Phi = \int_S \vec{B} d\vec{S}$, то (7.8) можна переписати так:

$$\oint_\Gamma \vec{E}_\text{в}\, d\vec{l} = -\frac{d}{dt}\int_S \vec{B} d\vec{S},$$

де інтеграл у правій частині беруть за довільною поверхнею *S*, яка опирається на контур Г (рис. 7.10).

Оскільки контур і поверхня нерухомі, операцію диференціювання за часом та інтегрування по поверхні можна поміняти місцями:

$$\oint_{\Gamma} \vec{E}_{\text{в}}\, d\vec{l} = -\int_{S} \frac{\partial \vec{B}}{\partial t} d\vec{S}. \qquad (7.9)$$

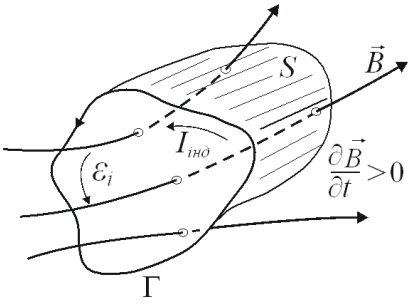


Рис. 7.10

У зв'язку з тим, що вектор $\vec{B}$ у загальному випадку залежить від часу і координат, ми записали під знаком інтеграла символ частинної похідної за часом $\frac{\partial}{\partial t}$ (інтеграл $\int_S \vec{B} d\vec{S}$ є функцією тільки від часу). Формула (7.9) виражає закон електромагнітної індукції (7.2), записаний в іншій формі.

Аналізуючи явище електромагнітної індукції, Максвелл зробив висновок, що причина виникнення е.р.с. індукції полягає у виникненні *електричного поля* (рис. 7.11). *При цьому провідники відіграють другорядну роль і є тільки своєрідним приладом, який виявляє це поле*. Під дією поля електрони провідності у дротині починають рух; коли дротина замкнена, в ній виникає індукційний струм. Істотна особливість цього явища та, що електричне поле, яке виникає, не є електростатичним. Силові лінії електростатичного поля завжди розімкнені. Вони починаються і закінчуються на електричних зарядах, відповідно до цього циркуляція вектора $\vec{E}$ в електростатичному полі завжди дорівнює нулю. Через це електростатичне поле не може підтримувати замкнений рух зарядів, а отже, не може призвести до виникнення е.р.с. Навпаки, електричне поле, яке виникає за електромагнітної індукції, має замкнені силові лінії, тобто є вихровим полем. Таке поле спричиняє у дротині рух електронів по замкнених траєкторіях і веде до виникнення е.р.с., при цьому сторонніми силами є сили вихрового електричного поля.

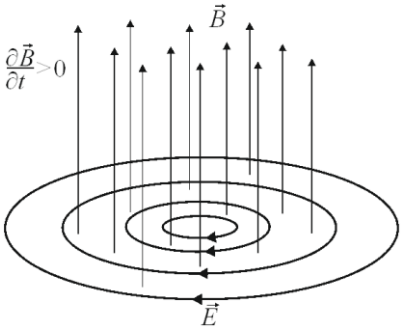


Рис. 7.11. Зі зміною магнітного поля виникає вихрове електричне поле. Зображений на рисунку напрямок $\vec{E}$ відповідає зростанню $\vec{B}$

Циркуляція вектора $\vec{E}$ по замкненому контуру в цьому полі не дорівнює нулю; її значення між двома будь-якими точками вже не визначається тільки положенням цих точок, як це було

у випадку електростатичного поля, але залежить ще й від форми контура (провідника), який сполучає ці точки. Таким чином, електричне поле може бути як потенціальним, так і вихровим. У загальному випадку електричне поле може складатись із поля $\vec{E}_q$, яке створюється зарядами, і поля $\vec{E}_в$, яке обумовлене змінним магнітним полем. Сумарне поле $\vec{E} = \vec{E}_q + \vec{E}_в$, але $\oint_Г \vec{E}_q d\vec{l} = 0$, тому співвідношення (7.9) можемо переписати так:

$$\oint_Г \vec{E} d\vec{l} = -\int_S \frac{d\vec{B}}{dt} d\vec{S}. \qquad (7.10)$$

Це рівняння є одним з основних в електромагнітній теорії Максвелла.

Існування взаємозв'язку між електричним і магнітним полями (що виражається, наприклад, рівнянням (7.10)) є причиною того, що окремий розгляд електричного і магнітного полів має тільки відносний зміст. Справді, електричне поле створюється системою нерухомих зарядів. Утім, якщо заряди нерухомі відносно деякої інерційної системи відліку, то відносно інших інерційних систем ці заряди рухаються, а отже, породжують не лише електричне, а й магнітне поле. Нерухомий провід з постійним струмом створює в кожній точці простору постійне магнітне поле. Утім, відносно інших інерційних систем цей провід перебуває в русі, тому магнітне поле в будь-якій точці з координатами *x*, *y*, *z*, яке він створює, буде змінюватись з часом, і, отже, породжувати *вихрове електричне поле*. Таким чином, поле, яке відносно деякої системи відліку є «чисто» електричним або «чисто» магнітним, відносно інших систем відліку буде являти собою сукупність електричного і магнітного полів, які утворюють єдине електромагнітне поле.

7.5. Самоіндукція. Індуктивність

Явище електромагнітної індук-ції спостерігається в усіх випадках, коли змінюється магнітний потік, який пронизує контур. Зокрема, цей потік може створюватися струмом, який проходить у розглядуваному контурі, тому за будь-якої зміни величини струму в якому-небудь контурі в ньому виникає е.р.с. індукції, яка спричиняє додатковий струм у контурі. Це явище називають *самоіндукцією*, а додаткові струми, які спричинюються е.р.с. самоіндукції – *екстраструмами* самоіндукції.

На рис. 7.12 подано схему досліду для спостереження екстраструмів. Котушка, яка має кілька тисяч витків дроту, увімкнена в коло, що містить батарею $\mathcal{E}$, реостат R і вимикач K. Паралельно до котушки приєднаний «грубий» гальванометр Г. Якщо ключ

замкнений, струм батареї розгалужуються: його частина I проходить через котушку, а частина I_1 – через гальванометр. Якщо розімкнути ключ, то магнітний потік у котушці зникатиме і в ній виникне екстраструм самоіндукції I_s (*екстраструм розмикання*). Згідно із законом Ленца він перешкоджатиме спаданню магнітного потоку, тобто буде напрямлений у котушці так само, як і спадний струм. Цей екстраструм проходить повністю через гальванометр (ключ *K* розімкнений), де його напрямок протилежний початковому струму I_1, тому гальванометр дає відкидання у зворотний бік. Після замикання ключа (установлення струму) в котушці також виникає екстраструм (*екстраструм замикання*). Його напрямок у котушці протилежний зростаючому струму батареї, однак у випадку замикання екстраструм розподіляється між батареєю і гальванометром і, крім того, його напрямок у гальванометрі такий самий, як і в наростаючого струму батареї, тому екстраструм замикання помітно значно гірший.

Якщо помістити в котушку залізне осердя, то екстраструми значно збільшуються. У цьому разі гальванометр можна замінити невеликою лампою розжарювання. Після розмикання ключа лампа дає яскравий спалах.

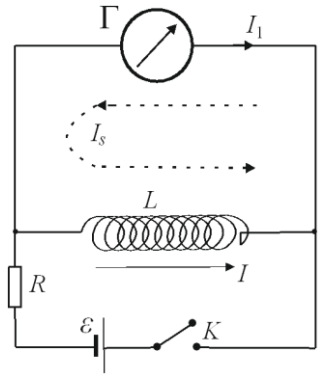


Рис. 7.12. Спостереження екстраструму самоіндукції

Розглянемо, від чого залежить е.р.с. самоіндукції. Нехай котушка L (соленоїд) має N витків, якими проходить струм I від зовнішнього джерела. Цей струм створює в соленоїді магнітне поле, яке зосереджене в його об'ємі, і це поле можна вважати однорідним. Нехай весь об'єм магнітного поля соленоїда заповнений однорідною речовиною з відносною магнітною проникністю μ. Якщо змінювати струм в соленоїді, то власний магнітний потік Ф, який пронизує його поперечний переріз, теж змінюватиметься. Магнітний потік крізь площу S, що обмежена одним витком,

$$\Phi = BS = \mu\mu_0 nSI = \mu\mu_0 \frac{NS}{l} I,$$

де l – довжина соленоїда; $B = \mu\mu_0 H$, де

$$H = nI = \frac{N}{l} I.$$

За законом Фарадея

$$\mathcal{E}_s = -N\frac{d\Phi}{dt} = -\frac{d}{dt}(\mu\mu_0 \frac{N^2 S}{l} I).$$

Припустимо, що сили, які деформують соленоїд, дуже малі й тому l т а S залишаються сталими. Якщо всередині соленоїда міститься феромагнітне осердя, то магнітний потік Φ змінюється не лише унаслідок зміни сили струму I, але і від зміни μ. Позначимо:

$$L = \mu\mu_0 \frac{N^2 S}{l}. \qquad (7.11)$$

Тоді

$$\mathcal{E}_s = -\frac{d}{dt}(LI) = -L\frac{dI}{dt} - I\frac{dL}{dt}.$$

Якщо магнітна проникність середовища μ не змінюється, то L = const і

$$\mathcal{E}_s = -L\frac{dI}{dt}. \qquad (7.12)$$

Електрорушійна сила самоіндукції прямо пропорційна швидкості зміни сили струму в контурі. Параметр L характеризує індивідуальні властивості контура, його називають *коефіцієнтом індуктивності* або просто *індуктивністю* контура. Індуктивність контура залежить від форми контура, його розмірів та відносної магнітної проникності середовища, в якому він міститься.

З виразів (7.3) і (7.12) можемо записати, що

$$LdI = d\psi_s,$$

де ψ_s – *потокозчеплення самоіндукції*.

Після інтегрування ця рівність має вигляд

$$LI = \psi_s.$$

Отже, індуктивністю замкненого контура називають скалярну величину L, яка дорівнює відношенню потокозчеплення самоіндукції контура ψ_s *до сили струму I в цьому контурі:*

$$L = \frac{\psi_s}{I}. \qquad (7.13)$$

Одиниця індуктивності – генрі (Гн); 1 Гн – це індуктивність такого контура, потокозчеплення самоіндукції якого за струму 1 А дорівнює 1 Вб.

Формулу (7.12) можна використати для провідників довільної форми, якщо знайти потокозчеплення $\psi_s = LI$, яке зв'язане з цим провідником за сили струму I. У виразі для е.р.с. самоіндукції знак «–» зумовлений правилом Ленца і показує, що наявність індуктивності контура викликає сповільнення зміни струму в ньому, тобто індуктивність контура є мірою його інертності відносно зміни струму.

Знайдемо закон зміни струму I в електричному колі, показаному на рис. 7.12, коли немає гальванометра під час вмикання і вимикання з нього джерела е.р.с. $\mathcal{E}$. Сила струму I в колі за законом Ома

$$I = \frac{\mathcal{E} + \mathcal{E}_s}{R} = \frac{\mathcal{E} - L\frac{dI}{dt}}{R},$$

якщо зовнішнє магнітне поле постійне.

Розділимо змінні у цьому рівнянні:

$$IR = \mathcal{E} - L\frac{dI}{dt}, \qquad \frac{dI}{dt} = \frac{\mathcal{E}-IR}{L},$$

$$\frac{dI}{\mathcal{E}-IR} = \frac{dt}{R}, \quad -\frac{d(\mathcal{E}-IR)}{\mathcal{E}-IR} = \frac{dt}{R}.$$

Вважаючи $\mathcal{E}$, R, L сталими, проінтегруємо останній вираз:

$$\int \frac{d(\mathcal{E}-IR)}{\mathcal{E}-IR} = -\frac{1}{R}\int dt,$$

$$\ln(\mathcal{E}-IR) = -\frac{t}{R} + \ln C,$$

де C – довільна стала інтегрування. Потенціюючи це рівняння, отримаємо $\mathcal{E} - IR = Ce^{-\frac{R}{L}t}$. У початковий момент часу $t = 0$, сила струму дорівнює I_0. Тоді: $\mathcal{E} - I_0R = C$. У результаті

$$\mathcal{E} - IR = (\mathcal{E} - I_0R)e^{-\frac{R}{L}t}.$$

Звідси знайдемо вираз для сили струму:

$$IR = I_0Re^{-\frac{R}{L}t} + \mathcal{E} - \mathcal{E}e^{-\frac{R}{L}t}$$

та

$$I = I_0e^{-\frac{R}{L}t} + \frac{\mathcal{E}}{R}(1 - e^{-\frac{R}{L}t}).$$

Для випадку вмикання джерела е.р.с. $\mathcal{E}$, коли початковий струм $I_0 = 0$, отримуємо вираз для екстраструму замикання:

$$I = \frac{\mathcal{E}}{R}(1 - e^{-\frac{R}{L}t}). \qquad (7.14)$$

Струм у колі поступово збільшується від нуля за $t = 0$ до значення $\mathcal{E}/R$, яке відповідає величині постійного струму (рис. 7.13). Зростання струму відбувається тим швидше, чим більше відношення R/L.

Для випадку вимкнення джерела ($\mathcal{E} = 0$)

$$I = I_0e^{-\frac{R}{L}t}. \qquad (7.15)$$

Струм у колі поступово зменшується від початкового значення I_0 до нуля за експоненціальним законом. Струм зменшується тим швидше, чим більше значення R/L, тобто чим більший опір кола і чим менша його індуктивність (рис. 7.14).

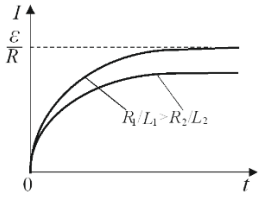


Рис. 7.13. Зростання струму в колі, що містить індуктивність для випадку вмикання джерела е.р.с. $\mathcal{E}$

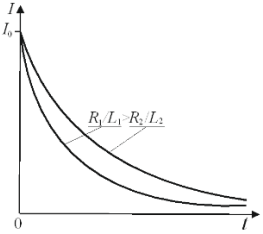


Рис. 7.14. Струм у колі для випадку вимкнення джерела е.р.с. $\mathcal{E}$

7.6. Взаємна індукція

Розглянемо тепер два контури зі струмом, наприклад два колові витки *1* і *2* (рис. 7.15). Частина ліній індукції поля,

яка створюється контуром *1*, проходитиме через контур *2*, тобто буде зчеплена із цим контуром. І навпаки, певна кількість ліній індукції поля, створюваного контуром *2*, буде зчеплена з контуром *1*. У цьому випадку ми говоримо, що між обома контурами існує *магнітний зв'язок*. Індукція поля контура *1* пропорційна величині струму I_1 в цьому контурі, тому магнітний потік $\Phi_{1,2}$ через контур *2*, який створюється контуром *1*, також пропорційний струму I_1:

$$\Phi_{12} = L_{12} I_1. \qquad (7.16)$$

Коефіцієнт $L_{1,2}$ називають *коефіцієнтом взаємної індукції* контурів *1* та *2*. Він, очевидно, дорівнює магнітному потоку через контур *2*, що створюється контуром *1* за величини струму в ньому, яка дорівнює одиниці. З порівняння (7.16) з (7.13) видно, що розмірність L_{12} – така сама, як і розмірність індуктивності, і тому коефіцієнт взаємної індукції вимірюють тими самими одиницями, що й індуктивність. Так само, коли в контурі *2* є струм деякої величини I_2, то він створює магнітний потік Φ_{21} через контур *1*, причому

$$\Phi_{21} = L_{21} I_2. \qquad (7.16\ а)$$

Тут L_{21} є коефіцієнтом взаємної індукції контурів *2* і *1*.

Можна показати, що для будь-яких двох контурів коефіцієнти взаємної індукції завжди однакові:

$$L_{12} = L_{21}. \qquad (7.17)$$

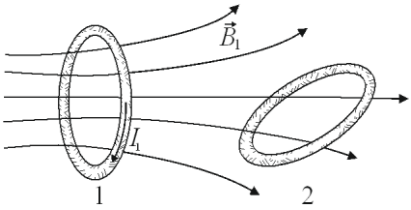


Рис. 7.15. Магнітний зв'язок двох контурів

Наявність магнітного зв'язку між контурами проявляється в тому, що за будь-якої зміни величини струму в одному з контурів у другому контурі виникає е.р.с. індукції. За основним законом електромагнітної індукції (див. підрозд. 7.3) маємо:

$$\mathcal{E}_{i2} = -\frac{d\Phi_{12}}{dt} = -L_{12}\frac{dI_1}{dt}, \qquad \mathcal{E}_{i1} = -\frac{d\Phi_{21}}{dt} = -L_{12}\frac{dI_2}{dt}, \qquad (7.18)$$

де $\mathcal{E}_{i2}$ – е.р.с. індукції, яка виникає в контурі *2*, а $\mathcal{E}_{i1}$ – е.р.с. індукції в контурі *1*.

Коефіцієнт взаємної індукції залежить від форми й розмірів контурів, від їх взаємного розміщення. Він залежить також від властивостей навколишнього середовища.

Розглянемо простий приклад визначення коефіцієнта взаємної індукції. Нехай є дві тороїдальні одношарові котушки *1* і *2*, які щільно прилягають одна до одної (рис. 7.16). У цьому випадку всі лінії індукції поля, які створюються однією котушкою,

проходять і через другу котушку. Індукція магнітного поля котушки *1*

$$B_1 = \mu_0 n_1 I_1 = \mu_0 \frac{N_1}{l} I_1,$$

де N_1 – кількість витків котушки 1; l – довжина тороїда по середній лінії. Це поле створює крізь один виток котушки *2* магнітний потік

$$B_1 S = \frac{\mu_0 N_1 I_1}{l} S,$$

де S – площа перерізу котушок. Повний потік крізь усі N_2 витків котушки *2* буде таким:

$$\Phi_{12} = \frac{\mu_0 N_1 N_2 S}{l} I_1.$$

Звідси для коефіцієнта взаємної індукції отримаємо вираз

$$L_{12} = \frac{\mu_0 N_1 N_2 S}{l} \cdot \qquad (7.19)$$

Якщо обчислити магнітний потік, який створюється котушкою *2* крізь котушку *1*, то дістанемо:

$$B_2 = \mu_0 \frac{N_2 I_2}{l}, \quad \Phi_{2,1} = \frac{\mu_0 N_2 N_1 S}{l} I_2.$$

Звідси ми для коефіцієнта взаємної індукції L_{21} знаходимо попередній вираз (7.19) згідно з формулою (7.17).

Коли всередині котушок є осердя з речовини з магнітною проникністю μ, магнітний потік збільшується в μ разів, і коефіцієнт взаємної індукції буде в μ раз більший.

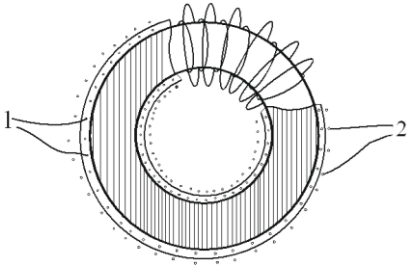


Рис. 7.16. Дві обмотки з магнітним зв'язком

7.7. Власна енергія струму

Щоб збільшити струм в електричному колі, треба виконати деяку роботу. Цю роботу виконує джерело струму, увімкнене в коло. Навпаки, під час кожного зменшення струму в колі вивільняється деяка енергія і джерело струму виконує меншу роботу, ніж за постійного струму. Розглянемо це питання докладніше.

Повернімось до кола, зображеного на рис. 7.17, що містить індуктивність L і має повний опір R, і припустимо спочатку, що в ньому є постійний струм, який встановився. Величина цього струму визначається е.р.с. джерела струму $\mathcal{E}$ та опором кола: $I = \frac{\mathcal{E}}{R}$.

За час dt в колі виділиться тепло Ленца–Джоуля $RI^2 dt$. Оскільки $RI = \mathcal{E}$, то

$$RI^2 dt = \mathcal{E} I dt = \mathcal{E} dq = dA,$$

але права частина цієї рівності виражає роботу джерела струму за той самий час dt. Ми бачимо, що робота джерела точно

дорівнює теплу Ленца–Джоуля, а отже, щоб підтримати постійне магнітне поле котушки, ніякої роботи виконувати не потрібно.

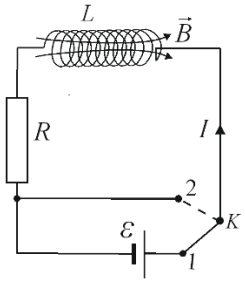


Рис. 7.17

Уявімо собі тепер, що струм у колі збільшується зі швидкістю dI/dt. У цьому разі в колі виникає ще е. р. с. самоіндукції $\mathcal{E}_s$, яка спричинить екстраструм

$$\delta I = \frac{|\mathcal{E}_s|}{R} = \frac{L}{R}\frac{dI}{dt}.$$

Він направлений проти струму I, й тому повна величина струму в колі буде $(I - \delta I)$.

Надалі вважатимемо, що процес збільшення струму відбувається дуже повільно ($\delta I \ll I$) і під час розрахунків враховуватимемо тільки члени першого порядку малості.

У процесі, який розглядається, за час dt виділиться тепло Ленца–Джоуля $dQ' = R(I - \delta I)^2 dt$. Воно менше порівняно з випадком постійного струму на величину

$$dQ' - dQ = R(I - \delta I)^2 dt - RI^2 dt = \\ = -2RI\delta I dt = -2LIdI.$$

Тут dI є збільшення струму в колі за час dt.

За цей самий час джерело струму виконає роботу $dA' = \mathcal{E}(I - \delta I)dt$, і тому батарея $\mathcal{E}$ виконає меншу роботу:

$$dA' - dA = \mathcal{E}(I\text{-}\delta I)dt - \mathcal{E}Idt = -\mathcal{E}\delta I dt = \\ = -LIdI.$$

Отже, у випадку наростаючого струму робота джерела струму більша за кількість тепла, яке виділилося (в алгебраїчному розумінні). Надлишок роботи джерела знайдемо так:

$$(dA' - dA) - (dQ' - dQ) = LIdI.$$

Вираз $dW = LIdI$ і характеризує ту роботу, яка потрібна для збільшення величини струму в колі від значення I до $I + dI$. Повна робота, яка потрібна для встановлення струму I,

$$W = L\int_0^I IdI = \frac{1}{2}LI^2. \qquad (7.20)$$

Після вимикання джерела струму робота W виділяється в колі; її виконують екстраструми розмикання, тому вираз (7.20) дає енергію, яку запасає контур зі струмом. Її називають *власною енергією струму.* Саме ця енергія і проявлялася в досліді, описаному в підрозд. 7.5 (рис.7.12) у вигляді відкидання стрілки гальванометра і спалаху лампочки від екстраструмів розмикання.

Корисно зіставити вирази для власної енергії контура зі струмом та

енергії зарядженого конденсатора (див. підрозд. 3.4): $q^2/(2C)$ ($q-$ заряд конденсатора, $C-$ його ємність). Енергія конденсатора пропорційна квадрату *заряду*, а енергія струму пропорційна квадрату *величини струму*, тобто залежить від *швидкості* руху зарядів.

У механіці є два види енергії: потенціальна і кінетична. Потенціальна енергія стисненої пружини дорівнює $kx^2/2$, де $x-$ зміщення кінця пружини, $k-$ її пружність (жорсткість), а кінетична енергія рухомого тіла є $mv^2/2$, де $m-$ маса тіла, а $v-$ його швидкість.

Розвиваючи аналогію між електричними і механічними явищами, ми бачимо, що енергія конденсатора відповідає потенціальній енергії в механіці, а власна енергія струму – кінетичній енергії. При цьому величина $1/C$, обернена ємності, аналогічна пружності (жорсткості) пружини k, а індуктивність $L-$ масі тіла m.

7.8. Енергія магнітного поля

Будь-який електричний струм завжди оточений магнітним полем, тому можна запитати, де саме зосереджена, локалізована власна енергія струму – всередині проводів, де рухаються електричні заряди, чи в магнітному полі, тобто в середовищі, яке оточують струми?

Відповідь на це запитання може дати тільки дослід. Утім, поки ми маємо справу з постійними струмами, такі досліди неможливі, бо в цьому випадку струми завжди оточені магнітним полем, і навпаки, магнітне поле завжди супроводжується струмами, які його підтримують. Це стосується і постійних магнітів, які також мають замкнені молекулярні струми (див. розд. 6).

Відповісти на це запитання можна після дослідження змінних магнітних полів або електромагнітних хвиль. В електромагнітних хвилях ми маємо магнітні поля, які змінюються у просторі й часі та здатні існувати без струмів, які їх підтримують. Оскільки електромагнітні хвилі містять у собі і переносять певну енергію, то звідси ми робимо висновок, що *енергія зосереджена в магнітному полі.*

Знайдемо величину енергії, яка міститься в одиниці об'єму магнітного поля. Розглянемо замкнену тороїдальну котушку. Її індуктивність (формула (7.11))

$$L = \mu\mu_0 \frac{N^2 S}{l},$$

де $\mu-$ магнітна проникність осердя тороїда, а μ_0- магнітна стала вакууму. Підставимо цей вираз у (7.20) й отримаємо

$$W = \frac{1}{2}\mu\mu_0 \frac{N^2 S}{l} I^2,$$

але врахувавши співвідношення (5.18) та (6.19) отримаємо, що

$$H = nI = \frac{N}{l}I$$

є напруженістю поля всередині котушки, тому

$$W = \frac{1}{2}\mu\mu_0 H^2 V,$$

де $V = Sl$ – об'єм котушки. Ми бачимо, що енергія однорідного магнітного поля пропорційна об'єму V, який займає поле, тому енергія одиниці об'єму поля, або *об'ємна густина енергії* магнітного поля,

$$\omega = \frac{\mu\mu_0 H^2}{2}.$$

Оскільки індукція магнітного поля (формула (6.19)) $B = \mu\mu_0 H$, то

$$\omega = \frac{BH}{2} = \frac{B^2}{2\mu\mu_0} = \frac{\mu\mu_0 H^2}{2} \qquad (7.21)$$

Якщо магнітне поле неоднорідне, то його можна розбити на нескінченно малі елементи об'єму dV, у кожному з яких поле можна вважати однорідним. Енергія, яка є в елементі об'єму dV, буде ωdV. Повна енергія будь-якого магнітного поля

$$W = \int_V \omega dV = \frac{1}{2}\int_V BHdV, \quad (7.22)$$

де інтегрування поширюється на весь об'єм V, зайнятий магнітним полем.

7.9. Взаємна енергія двох струмів

Обчислимо енергію магнітного поля, створеного двома контурами зі струмом. Розглянемо найпростіший випадок двох тороїдальних котушок (рис. 7.16), в яких магнітне поле однорідне, що перебувають у вакуумі ($\mu = 1$). Напруженість сумарного поля всередині котушок

$$H = H_1 \pm H_2 = \frac{N_1 I_1 \pm N_2 I_2}{l}.$$

Вибір знака «+» або «−» залежить від того, однакові чи протилежні напрямки струмів в обох котушках. В одиниці об'єму цього поля є енергія $\mu_0 H^2/2$, а повна енергія, яка міститься в усьому об'ємі поля $V = Sl$,

$$W = \frac{1}{2}\mu_0 H^2 Sl = \frac{\mu_0 S}{2l} \cdot$$
$$\cdot [(N_1 I_1)^2 + (N_2 I_2)^2 \pm 2(N_1 I_1)(N_2 I_2)].$$

Скориставшись виразами для індуктивностей L_1 та L_2 обох котушок (7.11) і їх коефіцієнтом взаємної індукції L_{12} (7.19), знайдений результат можна подати у такому вигляді:

$$W = \frac{1}{2}L_1 {I_1}^2 + \frac{1}{2}L_2 {I_2}^2 \pm L_{12} I_1 I_2. \quad (7.23)$$

Перший член цієї формули дає власну енергію струму I_1 (7.20), а другий доданок виражає власну енергію струму I_2. Формула показує, що енергія двох струмів у контурах, які мають магнітний зв'язок, відрізняється від суми власних енергій струмів на величину

$$W_{12} = \pm L_{12} I_1 I_2, \qquad (7.24)$$

яку називають *взаємною енергією* двох струмів.

З викладеного видно, що причина виникнення взаємної енергії струмів

полягає в тому, що в полі кількох контурів зі струмом додаються напруженості поля (принцип накладання, або суперпозиції полів), тоді як енергія магнітного поля пропорційна квадрату напруженості, а квадрат суми не дорівнює сумі квадратів.

Уявімо собі, що ми спочатку створюємо струм I_1 в контурі *1*, а струм $I_2 = 0$. Для цього потрібна буде робота $\frac{1}{2}L_1{I_1}^2$. Створюючи потім струм у контурі *2*, ми виконаємо роботу, яка згідно з (7.23) дорівнює $(\frac{1}{2}L_2{I_2}^2 \pm L_{12}I_1I_2)$. Якби ми спочатку створили струм I_2 (а струм I_1 дорівнював би нулю), то для цього потрібна була б робота $\frac{1}{2}L_2{I_2}^2$, але тоді наступне створення струму I_1 потребувало б роботи $(\frac{1}{2}L_1{I_1}^2 \pm L_{12}I_1I_2)$. Ми бачимо, що робота утворення струму в контурі за наявності другого контура зі струмом не дорівнює роботі утворення того самого струму у відокремленому від інших контурів. Вона відрізняється на величину взаємної енергії обох струмів W_{12}.

Ми дістали формули (7.23) і (7.24), розглядаючи окремий випадок тороїдальних котушок. Утім, легко впевнитися, що цей результат справедливий для будь-яких контурів довільної форми, які перебувають у будь-якому середовищі.

7.10. Вихрові струми

Якщо у змінному магнітному полі розміщений який-небудь масивний провідник, то вихрове електричне поле спричиняє в цьому провіднику індукційний струм. Густина цього струму в якій-небудь точці провідника за законом Ома становить $\vec{j} = \sigma\vec{E}$. Оскільки лінії напруженості $\vec{E}$ замкнені, то й лінії струму також замикаються всередині провідника, унаслідок чого такі струми називають *вихровими струмами.*

Вихрові струми спричиняють нагрівання провідників. Якщо всередині котушки зі змінним струмом помістити провідне тіло, наприклад металевий диск, орієнтований перпендикулярно до осі котушки, то диск можна розжарити до високої температури і розплавити.

Нагрівання провідників вихровими струмами застосовують в індукційних металургійних печах для плавлення металів і приготування їхніх сплавів, в лабораторній практиці для розжарювання металів, а також для приготування їжі та інших цілей. Вихрові струми виникають також під час руху масивних провідників у магнітному полі. Взаємодіючи з магнітним полем, вихрові струми спричиняють виникнення сил, що діють

на рухоме провідне тіло, які за законом Ленца завжди *протидіють* руху. Ця гальмівна дія дає змогу також легко виявити вихрові струми. На рис. 7.18 зображено маятник з масивною сочевицею з червоної міді, яка під час коливань маятника проходить між полюсами електромагніту. Якщо електромагніт вимкнений, маятник коливається завдяки великій масі з дуже малим загасанням, але після вмикання електромагніту різко зупиняється.

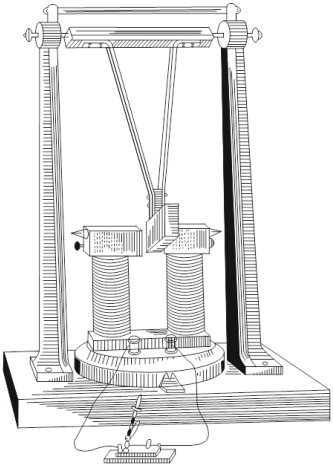


Рис. 7.18. Металевий маятник гальмується в магнітному полі внаслідок утворення вихрових струмів

Якщо мідній монеті надати можливість вільно падати без початкової швидкості між полюсами потужного електромагніту, вона рухатиметься майже рівномірно з дуже малою швидкістю приблизно 1 см/с так, як тоді, коли рух відбувається в дуже в'язкому середовищі. Якщо мідний диск розмістити поблизу постійного магніту і надати магніту обертання, то диск почне обертатися в той самий бік, що й магніт (рис. 7.19).

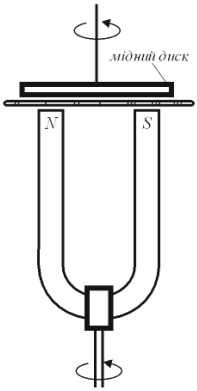


Рис. 7.19. Мідний диск, підвішений поблизу магніту, який обертається, розпочинає обертання

Взаємодія вихрових струмів з магнітним полем підлягає третьому закону Ньютона, тому, якщо в останньому досліді замість магніту обертати диск, то магніт також розпочинає обертання. Сили, які виникають від вихрових струмів і діють на рухомі провідники в магнітному полі, використовують у багатьох вимірювальних приладах (наприклад, електричні лічильники, електромагнітне заспокоювання вимірювальних приладів, тахометри тощо).

У деяких технічних пристроях вихрові струми відіграють шкідливу роль. Так, наприклад, у залізних осердях трансформаторів та обертових частинах електричних генераторів вихрові струми спричиняють некорисне нагрівання і зменшують к.к.д. цих пристроїв. Для ослаблення вихрових струмів такі деталі виготовляють з тонких листів, розділених найтоншими шарами ізолятора, щоб ізолюючі шари перетинали можливі лінії вихрових струмів. Суть цього прийому пояснюють дослідом, який зображено на рис. 7.20.

Підвісимо кубик, набраний з тонких металевих пластинок, на нитці, помістимо його між полюсами електромагніту і, закрутивши попередньо нитку, спостерігатимемо обертання кубика під час розкручування нитки. Якщо підвісити кубик у положення *а,* то ізолюючі шари перетинатимуть лінії вихрових струмів (одну з них зображено пунктиром). У цьому випадку вихрові струми виникають тільки в межах кожної із пластинок, їх дія дуже слабка і кубик вільно обертається з великою кутовою швидкістю. А якщо кубик підвісити в положення, як показано на рис. 7.20, *б*, то ізолюючі шари будуть паралельні лініям вихрових струмів і не перешкоджатимуть їх утворенню. Кубик поводитиме себе так, ніби ізолюючих шарів немає зовсім, і повертатиметься дуже повільно внаслідок гальмівної дії вихрових струмів.

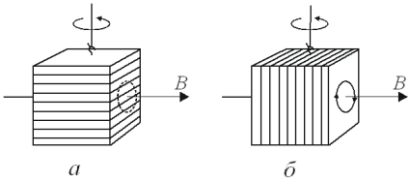


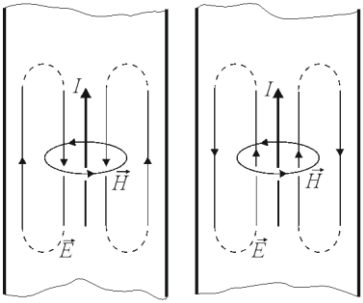
Рис. 7.20. Ослаблення вихрових струмів
У випадку а вихрові струми значно слабші, ніж у випадку б

7.11. Витіснення змінного струму (скін-ефект)

Якщо в однорідному провіднику зі сталим поперечним перерізом є постійний струм, то густина струму однакова в різних точках перерізу провідника. Інше спостерігається за змінного струму. У цьому випадку густина струму неоднакова за перерізом – вона найбільша на поверхні й найменша на осі провідника. Ця нерівномірність тим більша, чим товстіший провідник і чим більша частота змінного струму, і за дуже великих частот струм фактично існує тільки в тонкому поверхневому шарі, унаслідок чого явище і називають ***скін-ефектом*** (від англ. skin – шкіра, поверхневий шар).

Це цікаве явище також пояснюється виникненням вихрового електричного поля електромагнітної індукції. Розглянемо провідник зі

змінним струмом, і нехай в якийсь момент часу струм I має напрямок, показаний на рис. 7.21. Цей струм створює всередині провідника магнітне поле H, силові лінії якого лежать у площині, перпендикулярній до осі провідника.



а *б*

Рис. 7.21. Пояснення витіснення змінного струму на поверхню провідника: H – магнітне поле струму; E – вихрове електричне поле, яке ним спричиняється; а – струм зростає; б – струм спадає

Припустимо, що струм I підсилюється. Тоді зростаюче магнітне поле H спричинить виникнення вихрового електричного поля E (рис. 7.21, *а*), яке біля поверхні провідника напрямлене однаково зі струмом I, а на осі провідника – протилежно струму. Отже, це поле підсилюватиме струм на поверхні й ослаблюватиме його на осі.

Припустимо тепер, що струм I зменшується. У цьому випадку послаблююче магнітне поле H спричинить електричне поле E, яке буде напрямлене протилежно порівняно з першим випадком (рис. 7.21, *б*), тобто буде біля поверхні протилежне струму, а на осі – збігатиметься зі струмом. В обох випадках – і в разі зростання, і в разі послаблення струму I – вихрове електричне поле на осі провідника перешкоджатиме, а на поверхні сприятиме змінам струму I, а отже, на осі провідника змінний струм буде змінюватись слабо, а на поверхні сильно.

Закон розподілу густини струму по перерізу провідника особливо простий, коли провідник має форму плоского шару, товщина якого $2D$ значно менша за ширину (пластина). Розрахунок показує, що цей розподіл залежить від величини

$$d = \frac{1}{\sqrt{\pi \nu \mu \mu_0 \sigma}}, \qquad (7.25)$$

яку називають глибиною проникнення струму. Тут μ – магнітна проникність речовини провідника; μ_0 – магнітна стала вакууму; σ – питома електропровідність; ν – частота змінного струму. Виражаючи ці величини в одиницях СІ, d дістанемо в метрах.

Якщо $d \gg D$, то густина струму фактично стала по перерізу провідника. Якщо $d \ll D$ (значний скін-ефект), то закон розподілу струму має приблизно такий вигляд:

$$j(y) = j_{\text{пов}} e^{-\frac{y}{d}}, \qquad (7.26)$$

де $j(y)$ – густина струму на відстані y від поверхні провідника; $j_{\text{пов}}$ – густина струму біля поверхні. З (7.26) видно, що глибиною проникнення d є така відстань від поверхні провідника, на якій густина струму зменшується в $e = 2{,}7182...$ разів.

Завдяки скін-ефекту електричний струм за високих частот протікає переважно через поверхневий шар провідника. Це призводить до зменшення наявного перерізу провідника і, як наслідок, до *збільшення опору* провідника. За високих частот або товстих проводів це збільшення опору може бути значним.

Унаслідок нерівномірного розподілу струму в провіднику змінюється не тільки його опір, а й індуктивність. Справді, за постійного струму магнітне поле виникає як у зовнішньому просторі, так і всередині провідника. За наявності значного скін-ефекту струм існує фактично тільки в поверхневому шарі й магнітного поля всередині провідника немає. Це означає, що магнітна енергія стає меншою на величину енергії поля всередині провідника, а отже, індуктивність провідника зменшується.

Існування скін-ефекту завжди враховують у техніці швидкозмінних струмів. Оскільки такі струми фактично не проходять у глибину провідника, то провідні лінії для них складають з порожнистих труб. У сучасній радіотехніці надвисоких частот багато деталей (хвилеводи, коаксіальні лінії) покривають тонким шаром срібла (який є добрим провідником), бо їхній опір фактично зумовлений тільки поверхневим шаром.

Скін-ефект у поєднанні з нагріванням вихровими струмами використовують тепер з великим успіхом для *поверхневого гартування сталі.* Якщо використовувати цей метод гартування, на деталі діють магнітним полем змінних струмів, які дістають від спеціальних машин, і потім швидко охолоджують. Унаслідок скін-ефекту розігрівання і наступне гартування відбуваються тільки у поверхневому шарі, а вся основна маса залишається незагартованою. Це дає змогу виготовити вироби з високою стійкістю поверхні проти стирання, але не крихкі, як за звичайного гартування.

7.12. Метод вимірювання магнітної індукції

Нехай повний магнітний потік Ψ, який пронизує деякий провідний контур, змінюється від значення Ψ_1 до Ψ_2. Знайдемо заряд q, який протікає через кожний переріз контура. Миттєве значення сили струму в контурі згідно з законом Ома для замкненого кола запишемо так:

$$I = \frac{\mathcal{E}_i}{R} = -\frac{1}{R}\frac{d\Psi}{dt}.$$

Звідси знайдемо заряд, який протікає за час dt через поперечний переріз контура:

$$dq = Idt = -\frac{1}{R}\frac{d\Psi}{dt}dt = -\frac{1}{R}d\Psi.$$

Проінтегруємо цей вираз за зміною повного магнітного потоку від Ψ_1 до Ψ_2. Знайдемо повний заряд, який при цьому протікає через поперечний переріз контура, а саме:

$$q = \int dq = -\frac{1}{R}\int_{\Psi_1}^{\Psi_2} d\Psi = \frac{1}{R}(\Psi_1 - \Psi_2). \quad (7.26)$$

Таким чином, кількість заряду, що протікає через поперечний переріз контура, визначають зміною повного магнітного потоку, який пронизує цей контур і не залежить від швидкості його зміни.

Співвідношення (7.26) лежить в основі балістичного методу визначення магнітної індукції, суть якого в такому. В магнітне поле поміщають котушку, яка має N витків. Максимальний повний магнітний потік, що пронизує котушку, буде становити

$$\Psi_1 = NBS,$$

де $S-$ площа одного витка (має бути малою, щоб магнітне поле в її межах можна було вважати однорідним). У цьому випадку нормаль до площини котушки $\vec{n}$ і вектор індукції магнітного поля $\vec{B}$ спрямовані однаково. У разі повороту котушки на $180°$ повний магнітний потік

$$\Psi_2 = -NBS.$$

У цьому випадку нормаль $\vec{n}$ і вектор індукції поля $\vec{B}$ спрямовані протилежно. Тоді зміна магнітного потоку

$$\Psi_1 - \Psi_2 = 2NBS.$$

Якщо поворот котушки здійснити достатньо швидко, то в колі котушки виникне короткочасний імпульс струму, за якого протікає заряд

$$q = \frac{1}{R}2NBS. \quad (7.27)$$

З формули (7.27) можна визначити величину індукції B. Величини N і S відомі, а заряд q визначають балістичним гальванометром чи будь-яким іншим способом.

Контрольні запитання

1. У чому полягає суть явища електромагнітної індукції?

2. Сформулюйте закон Фарадея для електромагнітної індукції.

3. Сформулюйте й поясніть правило Ленца.

4. Знаменитий англійський фізик М. Фарадей у 1831 р. відкрив явище електромагнітної індукції. Фарадей довго ретельно шукав це явище, керуючись загальною ідеєю про зв'язок явищ

електрики і магнетизму. Одночасно з Фарадеєм і незалежно від нього в цьому ж напрямі працював швейцарський фізик Колладон, керуючись тією ж ідеєю. Суть досліду Колладона була в такому: кінці соленоїда з'єднувались з гальванометром, який для усунення впливу магніту був винесений в сусідню кімнату. Колладон всовував магніт у соленоїд і йшов в сусідню кімнату дивитись, що показує гальванометр. В чому була помилка Колладона? Чому йому не вдалось відкрити явище електромагнітної індукції? В часи Колладона застосовували гальванометри, в яких легка магнітна стрілка підвішувалась всередині котушки. За відхиленням стрілки судили про наявність струму в котушці. За такої системи гальванометра був неминучий вплив магніту, який переміщувався біля гальванометра.

5. У чому суть явища самоіндукції?

6. Запишіть вираз для е.р.с. самоіндукції.

7. Як визначають кількість електрики, що переноситься індукційним струмом? Самостійно виведіть потрібну формулу із закону Фарадея для електромагнітної індукції.

8. Пояснення явища електромагнітної індукції за допомогою сили Лоренца.

9. Поясніть явища електромагнітної індукції за допомогою вихрового електричного поля.

10. Поясніть поняття «самоіндукція», «індуктивність», «е.р.с. самоіндукції».

11. Що таке взаємна індукція? Поясніть залежність сили струму від часу в колах, які містять індуктивність.

12. Де зосереджена енергія магнітного поля?

13. Виведіть формулу для енергії магнітного поля соленоїда.

14. Виведіть формулу для густини енергії магнітного поля.

15. Наведіть приклади використання явища електромагнітної індукції в техніці зв'язку.

16. Які струми називають вихровими? Наведіть приклади використання вихрових струмів.

17. Чому швидкозмінні струми «витісняються» на поверхню провідника?

Розділ 8. Рівняння Максвелла

8.1. Перше рівняння Максвелла

У 60-х роках XIX століття Д. Максвелл, взявши за основу ідеї Фарадея про електричне і магнітне поля, узагальнив закони, виведені експериментально, і розробив завершену теорію

єдиного електромагнітного поля, яке утворюється довільною системою зарядів і струмів.

Теорія Максвелла є феноменологічною теорією електромагнітного поля. Це означає, що у цій теорії не розглядається внутрішній механізм явищ, які відбуваються в середовищі й викликають появу електричних і магнітних полів.

Електричні й магнітні властивості середовища характеризуються у теорії Максвелла трьома величинами: відносною діелектричною проникністю ε, відносною магнітною проникністю μ і питомою електропровідністю σ. В теорії Максвелла досліджується залежність ε, μ, σ від властивостей середовища, фізичний зміст цих явищ, які відбуваються у ньому у процесі його поляризації і намагнічування.

Теорія Максвелла є макроскопічною теорією електромагнітного поля. У ній розглядаються електричні й магнітні поля, які утворюються в об'ємах набагато більших, ніж об'єми окремих атомів і молекул. Крім того, припускається, що відстані від джерел полів до розглядуваних точок у багато разів більші від розмірів молекул.

У теорії Максвелла розглядаються усереднені електричні й магнітні поля, причому усереднення відповідних мікрополів виконується для інтервалів часу, значно більших від періодів обертання або коливання електричних зарядів і для ділянок поля, об'єми яких у багато разів більші від об'ємів атомів і молекул.

Теорія Максвелла є теорією близькодії, згідно з якою електричні й магнітні взаємодії здійснюються за допомогою електричних і магнітних полів і у яких вони поширюються зі скінченною швидкістю.

Максвелл узагальнив закон електромагнітної індукції для замкненого нерухомого провідного контура, що перебуває у змінному магнітному полі, і рівняння (7.10) є *першим рівнянням Максвелла*:

$$\oint_\Gamma \vec{E} d\vec{l} = -\int_S \frac{\partial \vec{B}}{\partial t} d\vec{S}. \qquad (8.1)$$

Фактично рівняння (8.1) є математичним формулюванням закону електромагнітної індукції Фарадея.

8.2. Струм зміщення. Друге рівняння Максвелла

У попередніх підрозділах ми бачили, що будь-яке змінне магнітне поле

спричиняє вихрове електричне поле. Аналізуючи різні електромагнітні процеси, Максвелл дійшов висновку, що має існувати й обернене явище: *будь-яка зміна електричного поля спричиняє виникнення вихрового магнітного поля.* Це твердження визначає найважливіші властивості електромагнітного поля і є другим основним положенням теорії Максвелла. Оскільки магнітне поле є основною, обов'язковою ознакою будь-якого струму, то Максвелл назвав змінне електричне поле *струмом зміщення*, на відміну від струму провідності, зумовленого рухом заряджених частинок (електронів та іонів). Слід зауважити, що цей термін (*струм зміщення*) не є вдалим. Він має деякий сенс у випадку діелектриків, бо в них насправді зміщуються заряди в атомах і молекулах. Утім, поняття струму зміщення ми застосовуємо і для вакууму, де ніяких зарядів, а отже, і їх «зміщення», немає. Утім, цей термін зберігся внаслідок історичних традицій.

Поняття струму «зміщення» можна пояснити за допомогою досліду, зображеного на рис. 8.1. Конденсатор з позитивно зарядженою правою пластиною розряджається через опір *R*. Навколо проводу створюється магнітне поле $\vec{H}$. Згідно із законом повного струму (6.14) циркуляція вектора напруженості магнітного поля по замненому контуру Г буде

$$\oint_{\Gamma} \vec{H} d\vec{l} = \sum I_i = \int_S \vec{j} d\vec{S}, \quad (8.2)$$

де $\vec{j}$ – густина струму, який протікає через поверхню *S*, яка пов'язана з контуром Г. Через поверхню *S*, яка показана на рис. 8.1 і 8.2, протікає струм *I*. Для цієї поверхні формула (8.2) може бути записана так: $\oint_{\Gamma} \vec{H} d\vec{l} = I$. Поверхня S' на рис. 8.2 теж пов'язана з контуром Г і має однакові права на участь у законі (теоремі) (8.2). Утім, через цю поверхню не протікає жодного струму. Тим не менше циркуляція $\vec{H}$ по контуру Г не може дорівнювати нулю, а це означає, що і $\int_{S'} \vec{j} d\vec{S} \neq 0$.

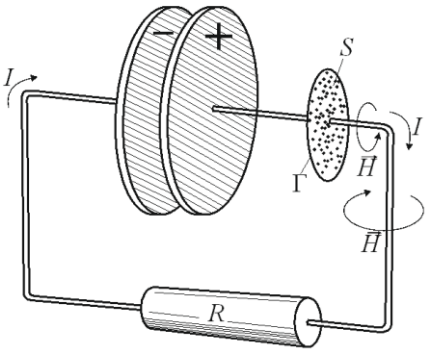


Рис. 8.1

Отже, рівняння (8.2) має бути замінене якимсь іншим, яке підходить для випадку, коли відбувається зміна розподілу зарядів (сила струму I змінюється з часом).

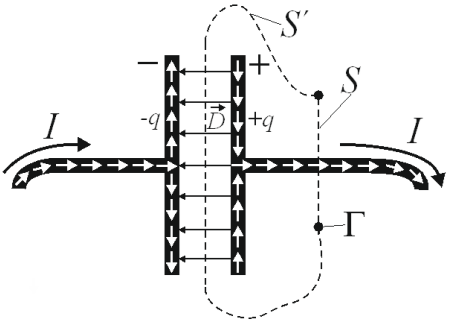


Рис. 8.2. Білими стрілками показано струм у провіднику

Перепишемо (8.2) у такому вигляді:

$$\oint_{\Gamma} \vec{H}d\vec{l} = \int_{S'} \vec{j}'d\vec{S}\,, \qquad (8.3)$$

і спробуємо з'ясувати, що являє собою $\vec{j}'$.

За теоремою Остроградського–Гаусса (2.19)

$$\Phi_D = \oint_{S'+S} \vec{D}d\vec{S} = q, \qquad (8.4)$$

де q – заряд правої пластини конденсатора, яка міститься всередині замкненої поверхні $S' + S$. Продиференціюємо вираз (8.4) за часом:

$$\frac{dq}{dt} = \frac{d\Phi_D}{dt} = \frac{d}{dt}\oint_{S+S'} \vec{D}d\vec{S}.$$

Якщо поверхня $S'+S$ нерухома і не деформується, то зміни з часом потоку зміщення через поверхню $S' + S$ зумовлюються лише зміною електричного зміщення $\vec{D}$ із плином часу, тому повну похідну d/dt можна замінити частинною похідною за часом $\partial/\partial t$ і диференціювання внести під знак інтеграла:

$$\frac{dq}{dt} = I' = \oint_{S'+S} \frac{\partial \vec{D}}{\partial t}d\vec{S} = \oint_{S'+S} \vec{j}'d\vec{S}.$$

Із цього виразу ми бачимо, що $\frac{d\vec{D}}{dt}$ має розмірність густини струму. Максвелл запропонував назвати вираз $\frac{d\vec{D}}{dt}$ густиною струму зміщення:

$$\vec{j}_{зм} = \frac{\partial \vec{D}}{\partial t}\cdot \qquad (8.5)$$

Густина струму зміщення у цій точці простору дорівнює швидкості зміни вектора електричного зміщення у цій точці.

Струмом зміщення через довільну поверхню S називають фізичну величину, яка дорівнює потоку вектора густини струму зміщення крізь цю поверхню:

$$I_{зм} = \int_S \vec{j}_{зм}d\vec{S} = \int_S \frac{\partial \vec{D}}{\partial t}d\vec{S} = \frac{\partial \Phi_D}{\partial t}. \qquad (8.6)$$

Таким чином, закон (8.2) треба записати так:

$$\oint_L \vec{H}d\vec{l}=\int_S \vec{j}d\vec{S}+ \\ +\int_S \vec{j}_{зм}d\vec{S}=\int_S \vec{J}_{повн}d\vec{S}, \quad (8.7)$$

де $\vec{J}_{повн}=\vec{j}+\vec{j}_{зм}-$ *повний струм*, що дорівнює сумі струмів провідності та конвекційних, а також струму зміщення.

Тепер співвідношення (8.7) можна назвати *законом повного струму* в загальному випадку. Поняття струму зміщення усуває складність, що виникла в ситуації, яка показана на рис. 8.2. Оскільки заряд конденсатора зменшується, то й величина індукції електричного поля, конфігурація якого для будь-якого моменту часу показана на рис. 8.3, зменшується. У цьому випадку вектор $\frac{d\vec{D}}{dt}$ направлений у бік, протилежний до напрямку вектора $\vec{D}$. Векторна функція $\frac{d\vec{D}}{dt}$ (струм зміщення) на рис. 8.4 зображена чорними стрілками. У співвідношенні (8.3) густина струму $\vec{j}'$ є густиною струму зміщення і

$$\int_{S'} \vec{j}'\,d\vec{s}=\int_{S'} \frac{\partial \vec{D}}{\partial t}\,d\vec{S},$$

якщо поверхня S перебуває на значній відстані від конденсатора.

Увівши уявлення про струм зміщення, Максвелл зробив висновок про те, що кола будь-яких непостійних струмів замкнені так само, як і кола постійного струму. Замкненість таких кіл забезпечується струмами зміщення, які «проходять» у тих ділянках, де немає провідників, наприклад, між обкладинками конденсатора під час його заряджання або розрядження.

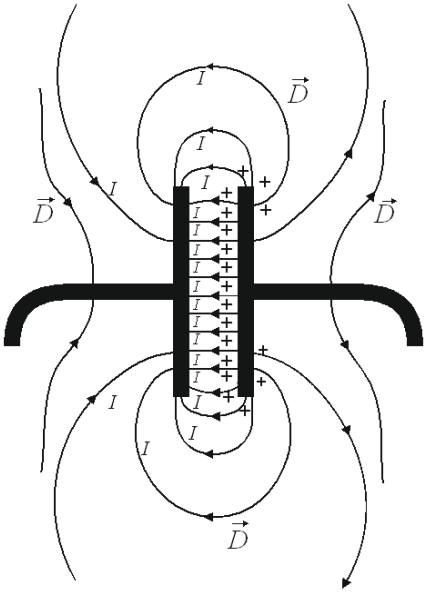


Рис. 8.3. Конфігурація індукції $\vec{D}$ електричного поля

За Максвеллом, струм зміщення, як і звичайні струми провідності, є джерелом виникнення вихрового магнітного поля, циркуляція напруженості $\vec{H}$ якого по замкненому контуру не дорівнює нулю.

В діелектрику вектор зміщення $\vec{D}$ згідно з (2.18) складається із двох доданків: $\vec{D}=\varepsilon_0\vec{E}+\vec{P}$, де $\vec{E}$ –

напруженість електричного поля, $\vec{P}$ – вектор поляризації, який характеризує дійсне зміщення електричних зарядів у неполярних молекулах і поворот полярних молекул, що містяться в одиниці об'єму діелектрика. Густина струму зміщення в діелектрику

$$\vec{j}_{зм} = \varepsilon_0 \frac{\partial \vec{E}}{\partial t} + \frac{\partial \vec{P}}{\partial t}, \qquad (8.8)$$

де $\varepsilon_0 \frac{\partial \vec{E}}{\partial t}$ – густина струму зміщення у вакуумі; $\frac{\partial \vec{P}}{\partial t}$ – густина струму поляризації, що обумовлений зміщенням або поворотом зарядів в діелектрику.

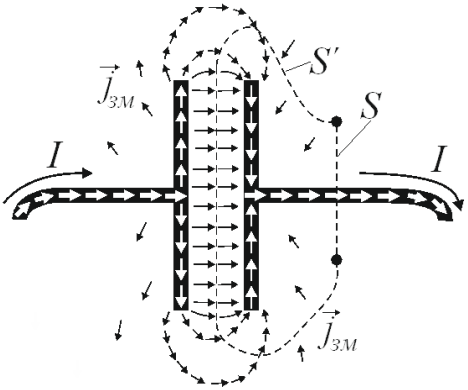


Рис. 8.4. Струм провідності (білі стрілки), струм зміщення (чорні стрілки)

Струм зміщення у вакуумі не виділяє тепла. Густина цього струму тим більша, чим більша швидкість зміни напруженості електричного поля. Густину повного струму запишемо так:

$$j_{повн} = \vec{j} + \frac{\partial \vec{D}}{\partial t}. \qquad (8.9)$$

Повний струм у колах змінного струму завжди замкнений. Тепер співвідношення (8.7) можемо записати так:

$$\oint_\Gamma \vec{H} d\vec{l} = \int_S \left(\vec{j} + \frac{\partial \vec{D}}{\partial t}\right) d\vec{S}. \qquad (8.10)$$

Якщо струми провідності відсутні ($\vec{j} = 0$), то (8.10) перепишеться як:

$$\oint_\Gamma \vec{H} d\vec{l} = \int_S \frac{\partial \vec{D}}{\partial t} d\vec{S}. \qquad (8.11)$$

Порівнявши (8.11) з (7.10) можемо сказати, що (8.11) визначає наявність *нового явища індукції*, в якому змінне електричне поле супроводжується магнітним полем. Якщо це явище індукції існує, то чому Фарадей його не відкрив? Насамперед він його не шукав, але є й більш фундаментальна причина, чому досліди, подібні до фарадеївських, не могли виявити будь-яких нових явищ, що пов'язані зі струмом зміщення (8.5).

Щоб побачити нове явище індукції (8.11), потрібні швидкозмінні електричні поля. Ось чому від теоретичного відкриття струмів зміщення Максвеллом до експериментального доказу їх існування в дослідах Герца пройшло багато років.

Введення струму зміщення, який визначається виразом (8.5), «зрівняло у правах» електричне і магнітне поля. Із

явища електромагнітної індукції (8.1) витікає, що змінне магнітне поле породжує електричне поле. З рівняння (8.11) випливає, що змінне електричне поле породжує магнітне поле. Співвідношення (8.10) є другим рівнянням Максвелла.

8.3. Рівняння Максвелла

Відкриття струму зміщення дозволило Максвеллу створити єдину теорію електричних і магнітних явищ. Ця теорія пояснила всі відомі на той час експериментальні факти і передбачила ряд нових явищ, існування яких підтвердилось пізніше.

Рівняння Максвелла пов'язують зміни основних характеристик електромагнітного поля – векторів $\vec{E}$, $\vec{D}$, $\vec{H}$, $\vec{B}$ у певному матеріальному середовищі з розподілом у ньому електричних зарядів і струмів. Середовище в рівняннях Максвелла враховується феноменологічно, тобто без розкриття внутрішнього механізму взаємодії речовин з полем.

В основі теорії Максвелла лежать чотири рівняння. Першим рівнянням Максвелла є співвідношення (8.1), яке пов'язує значення $\vec{E}$ зі змінною $\vec{B}$ з часом і є виразом закону електромагнітної індукції.

Перше рівняння Максвелла вказує на те, що джерелами електричного поля можуть бути не тільки електричні заряди, але й змінні з часом магнітні поля.

Другим рівнянням Максвелла є закон повного струму (8.10), який показує, що магнітні поля можуть збуджуватись або рухомими зарядами (струмами), або змінними електричними полями.

Третім рівнянням Максвелла є теорема Остроградського–Гаусса (2.19) для потоку вектора електричного зміщення $\vec{D}$ крізь довільну замкнену поверхню S, що охоплює сумарний заряд q:

$$\oint_S \vec{D} d\vec{S} = q.$$

Якщо заряд розподілений всередині замкненої поверхні з об'ємною густиною ρ, то

$$\oint_S \vec{D} d\vec{S} = \int_V \rho dV. \qquad (8.12)$$

Четвертим рівнянням Максвелла є теорема Остроградського–Гауса (5.24) для магнітного потоку крізь довільну замкнену поверхню:

$$\oint_S \vec{B} d\vec{S} = 0. \qquad (8.13)$$

Отже, повна система рівнянь Максвелла в інтегральній формі має такий вигляд:

$$\begin{aligned} &1.\ \oint_\Gamma \vec{E}d\vec{l} = -\int_S \frac{\partial \vec{B}}{\partial t}d\vec{S}, \\ &2.\ \oint_\Gamma \vec{H}d\vec{l} = \int_S \left(\vec{j}+\frac{\partial \vec{D}}{\partial t}\right)d\vec{S}, \end{aligned} \qquad (8.14)$$

$$\begin{aligned} &3.\ \oint_S \vec{D}d\vec{S} = \int_V \rho dV, \\ &4.\ \oint_S \vec{B}d\vec{S} = 0. \end{aligned} \qquad (8.14)$$

Величини, що входять у рівняння Максвелла (8.14), не є незалежними. Між ними є зв'язок (2.21), (6.19), (4.14):

$$\vec{D} = \varepsilon\varepsilon_0\vec{E},\ \ \vec{B} = \mu\mu_0\vec{H},\ \ \vec{j} = \sigma\vec{E}.$$

Зазначимо, що до першого і четвертого рівнянь Максвелла входять лише основні характеристики поля $\vec{E}$ і $\vec{B}$, а до другого і третього – лише допоміжні величини $\vec{D}$ і $\vec{H}$. Рівняння Максвелла несиметричні відносно полів.

Це пов'язано з тим, що у природі існують лише електричні заряди, а магнітних зарядів немає.

Для стаціонарних полів ($\vec{E} = = \text{const}$ і $\vec{B} = \text{const}$) рівняння Максвелла мають такий вигляд:

$$\oint_\Gamma \vec{E}d\vec{l} = 0, \quad \oint_S \vec{D}d\vec{S} = q,$$

$$\oint_\Gamma \vec{H}d\vec{l} = I, \quad \oint_S \vec{B}d\vec{S} = 0. \quad (8.15)$$

У цьому випадку електричні й магнітні поля існують незалежно одне від одного.

8.4. Рівняння Максвелла у диференціальній формі

Рівняння Максвелла (8.14) застосовні до поверхні будь-якої величини; тому величини, які входять до них, стосуються різних точок поля. Так, наприклад, у лівій частині рівняння (8.10) $\vec{H}$ є напруженість магнітного поля в точках контура Г, що обмежує поверхню S, яка розглядається, тоді як потік у правій частині залежить від значень $\vec{D}$ у точках розглядуваної поверхні. Утім, можна перетворити ці рівняння в таку форму, щоб усі величини стосувалися однієї точки поля. Для цього рівняння Максвелла треба застосувати до контурів і поверхонь, які пов'язані між собою, нескінченно малих величин.

Для цього розглянемо деякі математичні поняття. Оператор набла, або оператор Гамільтона позначають символом $\vec{\nabla}$, і під ним розуміють вектор з компонентами $\frac{\partial}{\partial x}$, $\frac{\partial}{\partial y}$, $\frac{\partial}{\partial z}$, тобто

$$\vec{\nabla} = \vec{\imath}\frac{\partial}{\partial x} + \vec{\jmath}\frac{\partial}{\partial y} + \vec{k}\frac{\partial}{\partial z}\cdot \qquad (8.16)$$

Сам по собі цей вектор сенсу не має. Він набуває змісту в поєднанні із скалярною або векторною функцією, на яку його символічно множать. Так, якщо

помножити вектор $\vec{\nabla}$ на скаляр φ, то отримуємо вектор

$$\vec{\nabla}\varphi = \vec{i}\frac{\partial\varphi}{\partial x} + \vec{j}\frac{\partial\varphi}{\partial y} + \vec{k}\frac{\partial\varphi}{\partial z} = \text{grad}\varphi, \quad (8.17)$$

який є градієнтом функції φ (див. співвідношення (1.33)).

Якщо вектор $\vec{\nabla}$ помножити скалярно на вектор $\vec{E}$, то отримаємо скаляр

$$\vec{\nabla}\vec{E} = \nabla_x E_x + \nabla_y E_y + \nabla_z E_z = \\ = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = \text{div}\vec{E}, \quad (8.18)$$

який є не чим іншим, як дивергенцією вектора $\vec{E}$ (див. (1.36)). Якщо помножити $\vec{\nabla}$ на вектор $\vec{H}$ ***векторно***, то отримаємо вектор, компоненти якого знаходять за допомогою визначника

$$\vec{\nabla} \times \vec{H} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ H_x & H_y & H_z \end{vmatrix} = \text{rot}\vec{H}. \quad (8.19)$$

Такий векторний добуток називають ротором вектора (у нашому випадку як приклад вектора $\vec{H}$). Із (8.19) знаходимо, що

$$\begin{cases} \text{rot}_x\vec{H} = \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z}, \\ \text{rot}_y\vec{H} = \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x}, \\ \text{rot}_z\vec{H} = \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y}. \end{cases} \quad (8.20)$$

Теорему Остроградського–Гауса в математиці записують так:

$$\oint_S \vec{a}d\vec{S} = \int_V \text{div}\vec{a}dV, \quad (8.21)$$

де V – об'єм, обмежений замкненою поверхнею S; $\vec{a}$ – деяка векторна функція. Є теорема Стокса, яку математично записують так:

$$\oint_\Gamma \vec{a}d\vec{l} = \int_S \text{rot}\vec{a} \cdot d\vec{S}, \quad (8.22)$$

де контур Г і поверхня S взаємопов'язані.

Тоді перше і друге рівняння Максвелла з урахування (8.22) можуть бути записані так:

$$\oint_\Gamma \vec{E}d\vec{l} = \int_S \text{rot}\vec{E}d\vec{S} = -\int_S \frac{\partial\vec{B}}{\partial t}d\vec{S},$$

$$\oint_\Gamma \vec{H}d\vec{l} = \int_S \text{rot}\vec{H}d\vec{S} = \int_s \left(\vec{j} + \frac{\partial\vec{D}}{\partial t}\right)d\vec{S}.$$

Оскільки S є довільною поверхнею, яка пов'язана з контуром Г, то рівність інтегралів можна замінити на рівність підінтегральних виразів, тоді отримаємо:

$$\text{rot}\vec{E} = -\frac{\partial\vec{B}}{\partial t}, \quad (8.23)$$

$$\text{rot}\vec{H} = \vec{j} + \frac{\partial\vec{D}}{\partial t}\cdot \quad (8.24)$$

Скориставшись теоремою Остроградського–Гаусса (8.21), третє (8.12) і четверте (8.13) рівняння Максвелла перепишемо так:

$$\oint_S \vec{D}d\vec{S} = \int_V \mathrm{div}\vec{D}dV = \int_V \rho dV,$$

$$\oint_S \vec{B}d\vec{S} = \oint_V \mathrm{div}\vec{B}d\vec{S} = 0.$$

Оскільки замкнена поверхня S і пов'язаний з нею об'єм V є довільними, то рівність інтегралів замінимо на рівність підінтегральних виразів, тому:

$$\mathrm{div}\vec{D} = \rho, \qquad (8.25)$$

$$\mathrm{div}\vec{B} = 0. \qquad (8.26)$$

Співвідношення (8.23)–(8.26) є рівняннями Максвелла в диференціальній формі, в яких характеристики поля $\vec{E}$, $\vec{B}$, $\vec{D}$, $\vec{H}$ належать до однієї точки простору. Кожне з векторних рівнянь (8.23) і (8.24) еквівалентне трьом скалярним рівнянням, які пов'язують компоненти векторів, що стоять у лівій і правій частинах рівності. Скориставшись формулами (8.18), (8.19) та (8.20), подамо рівняння Максвелла в скалярній формі:

$$\begin{aligned} \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} &= -\frac{\partial B_x}{\partial t} \\ \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} &= -\frac{\partial B_y}{\partial t}, \end{aligned} \qquad (8.27)$$

$$\begin{aligned} \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} &= -\frac{\partial B_z}{\partial t} \\ \frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} &= 0 \end{aligned} \qquad (8.28)$$

(перша пара рівнянь),

$$\begin{aligned} \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} &= j_x + \frac{\partial D_x}{\partial t} \\ \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} &= j_y + \frac{\partial D_y}{\partial t}, \end{aligned} \qquad (8.29)$$

$$\begin{aligned} \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} &= j_z + \frac{\partial D_z}{\partial t} \\ \frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} &= \rho \end{aligned} \qquad (8.30)$$

(друга пара рівнянь).

Всього отримано вісім рівнянь, в які входять 12 функцій (по три компоненти векторів $\vec{E}$, $\vec{B}$, $\vec{D}$, $\vec{H}$).

Оскільки кількість рівнянь менша від кількості невідомих функцій, то рівнянь (8.27)–(8.30) недостатньо, щоб знайти характеристики полів за заданим розподілом зарядів і струмів. Щоб виконати розрахунок полів, слід доповнити рівняння Максвелла співвідношеннями (2.21), (6.19) і (4.14), а саме:

$$\vec{D} = \varepsilon\varepsilon_0\vec{E}, \qquad (8.31)$$

$$\vec{B} = \mu\mu_0\vec{H}, \qquad (8.32)$$

$$\vec{j} = \sigma\vec{E}. \qquad (8.33)$$

Сукупність рівнянь (8.23)–(8.26) і (8.31)–(8.33) становить основу електродинаміки середовищ, які перебувають у стані спокою.

8.5. Значення теорії Максвелла

Теорія Максвелла відіграла значну роль у розвитку знань людства про електрику. Щоб краще зрозуміти значення цієї теорії, потрібно згадати

історичну послідовність основних відкриттів у галузі електрики до праць Максвелла.

Кількісне вивчення електричних явищ почалося із праць Кулона (1785 р.), який встановив спочатку закон взаємодії електричних зарядів і поширив його пізніше на взаємодію «магнітних зарядів». Утім, аж до 1820 р. електричні й магнітні явища розглядали як різні явища, не пов'язані між собою.

Відкриття Ерстедом у 1820 р. дії струму на магнітну стрілку показало, що між магнітними та електричними явищами існує зв'язок і що магнітні дії можна дістати за допомогою електричних струмів. Магнітну дію струмів докладно вивчив Ампер, який дійшов висновку, що всі магнітні явища у природі, зокрема й пов'язані з постійними магнітами, спричинені електричними струмами (теорія молекулярних струмів Ампера).

Особливе значення мало відкриття електромагнітної індукції Фарадеєм. Фарадей виходив з основної ідеї про взаємозв'язок явищ природи. Він вважав, що коли струм здатний викликати магнітні явища, то і навпаки, за допомогою магнітів або інших струмів можна дістати електричні струми. Унаслідок наполегливих і численних спроб Фарадей справді у 1831 р. відкрив це явище, яке ще більше зміцнювало уявлення про зв'язок між електрикою і магнетизмом.

Іншою найважливішою ідеєю у працях Фарадея було визнання основної визначальної ролі проміжного середовища в електричних явищах. Фарадей не допускав дії на відстані, яка, як тепер добре відомо, фізично не має сенсу, і вважав, що електричні й магнітні взаємодії передаються проміжним середовищем і що саме в цьому середовищі відбуваються основні електричні й магнітні процеси.

У працях Максвелла ідеї Фарадея було надалі поглиблено, розвинуто і перетворено в точну математичну теорію. У теорії Максвелла думка про тісний зв'язок електричних і магнітних явищ дістала остаточне оформлення у вигляді двох основних положень теорії, розглянутих у цьому підручнику у підрозд. 7.4 і 8.2, і була у строгій формі виражена у вигляді рівнянь Максвелла (підрозд. 8.3 та 8.4), тому теорія Максвелла була завершенням важливого етапу в розвитку вчення про електрику і привела до класичного уявлення про *електромагнітне поле*, яке містить у загальному випадку і електричне, і

магнітне поля, пов'язані між собою і здатні взаємно перетворюватись одне в одне.

Рівняння Максвелла містять усі основні закони електричного і магнітного полів, включаючи електромагнітну індукцію, і тому є *загальними* рівняннями електромагнітного поля в середовищах, які перебувають у спокої.

Теорія Максвелла не тільки пояснила вже відомі факти, а й завбачила нові важливі явища. Зовсім новим у цій теорії було припущення Максвелла про магнітне поле струмів зміщення (підрозд. 8.2). На основі цього припущення Максвелл теоретично завбачив існування *електромагнітних хвиль*, тобто змінного електромагнітного поля, яке поширюється у просторі зі скінченною швидкістю. Теоретичне дослідження властивостей електромагнітних хвиль привело потім Максвелла до створення *електромагнітної теорії світла*, за якою світло є теж лектромагнітними хвилями. Надалі електромагнітні хвилі справді були виявлені під час досліду, а ще пізніше електромагнітна теорія світла, а з нею і вся теорія Максвелла, дістала повне і беззаперечне підтвердження.

Відкладемо розгляд електромагнітних хвиль до вивчення електричних коливань і пружних хвиль.

Контрольні запитання

1. Скільки є рівнянь Максвелла?

2. Чим відрізняється струм зміщення від струму провідності?

3. Чи переносить електричний заряд струм зміщення?

4. Який напрямок має струм зміщення?

5. Чи створює магнітне поле струм зміщення?

6. Чи створює електричне поле струм зміщення?

7. Чи створює теплову дію струм зміщення подібно струму провідності?

8. Поясніть, чому конденсатор проводить змінний струм і не проводить постійний.

9. Які фізичні закони відображені у рівняннях Максвелла?

10. Яку роль відіграла теорія Максвелла про електромагнітне поле в розвитку науки в цілому?

Розділ 9. Коливання

9.1. Види коливань. Загальні поняття

Коливання – це рухи або процеси, які мають тією чи іншою мірою повторюваність із часом. Коливання властиві всім явищам природи: пульсуюче випромінювання зірок, всередині яких відбуваються циклічні ядерні реакції; з високим ступенем періодичності обертаються планети Сонячної системи; рух Місяця викликає припливи і відпливи на Землі; в земній іоносфері й атмосфері циркулюють потоки заряджених і нейтральних частинок; вітри збуджують коливання і хвилі на поверхнях водойм і т. д. Всередині будь-якого живого організму неперервно відбуваються різноманітні процеси, які циклічно повторюються, наприклад, з дивною надійністю б'ється людське серце, навіть психіка людей піддається коливанням. У вигляді складної сукупності коливань частинок і полів (електронів, протонів, фотонів та ін.) можна подати «будову» мікросвіту.

У техніці коливання або виконують певні функціональні обов'язки (маятник, коливальний контур, генератор коливань і под.), або виникають як неминучий прояв фізичних властивостей (вібрації машин і споруд, нестійкість вихрових потоків під час руху тіл у газах та ін.).

У фізиці виділяють коливання механічні, електромагнітні та їх комбінації. Це обумовлено тією винятковою роллю, яку відіграють гравітаційна й електромагнітна взаємодія у масштабах, характерних для життєдіяльності людини. За допомогою поширення механічних коливань густини і тиску повітря, які сприймаються нами як звук, а також дуже швидких коливань електричного і магнітного полів, які сприймаються нами як світло, ми отримуємо більшість прямої інформації про навколишній світ.

З вивченням коливань різної фізичної природи виникла впевненість про можливість загального, «позапредметного» підходу до них, який ґрунтується на властивостях і закономірностях коливальних процесів взагалі. У результаті з'явилась теорія коливань і хвиль, основним математичним апаратом якої спочатку були диференціальні рівняння із звичайними похідними.

Кінематика коливань дозволяє виділити декілька прикладів (рис. 9.1). Будемо говорити про коливання, які описуються функцією часу $u(t)$. На рис. 9.1, *а – г*, показано періодичні

коливання різної форми, в яких будь-яке значення $u(t)$ повторюється через однакові проміжки часу T, які називають **періодом** коливань, тобто $u(t) = u(t + T)$. Величина, яка обернена періоду T і дорівнює кількості коливань за одиницю часу, називають **частотою коливань** $\left(\nu = \frac{1}{T}\right)$; користуються також або циклічною (коловою) частотою $\omega = 2\pi\nu$. У випадку просторових коливань вводять аналогічні поняття просторового періоду (довжини хвилі λ) і хвильового числа $k = 2\pi/\lambda$.

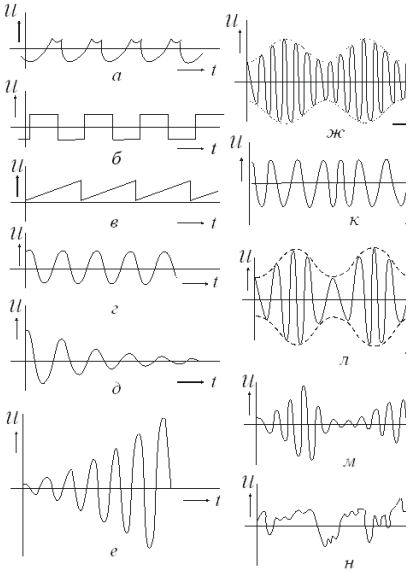


Рис.9.1. Різні види коливань: а – періодичні коливання складної форми; б – прямокутні коливання; в – пилкоподібні коливання; г – синусоїдальні; д – згасаючі; е – наростаючі; ж – амплітудно-модульовані; к – частотно-модульовані; л – коливання модульовані за амплітудою і фазою; м – коливання, амплітуда і фаза яких випадкові функції; н – випадкові коливання; и – величина, яка коливається; е - час

Різновидами періодичних коливань є прямокутні (рис. 9.1, *б*), пилкоподібні (рис. 9.1, *в*) і найбільш важливі – синусоїдальні, або **гармонічні** коливання (рис. 9.1, *г*). Останні можуть бути записані у вигляді

$$\begin{aligned} u(t)=A\cos\varphi=A\cos(\omega_0 t+\varphi_0), \\ u(t)=A\sin\varphi=A\sin(\omega_0 t+\varphi_0'), \end{aligned} \quad (9.1)$$

де A – амплітуда, максимальне значення за модулем величини u; $\omega t + \varphi_0$ – фаза; φ_0 або φ_0' – початкове значення фази; ω_0 – циклічна частота.

Для показаного на рис. 9.1, *д* згасаючого коливання

$$u(t) = Ae^{-\beta t}\cos(\omega t + \varphi_0), \quad (9.2)$$

де β – коефіцієнт згасання; ω – частота згасаючих коливань. На рис. 9.1, *е* показано наростаюче коливання.

Для передавання інформації застосовують модульовані коливання (рис. 9.1, *ж–л*), амплітуда, фаза або частота яких змінюються за певним законом відповідно до сигналу, який передається. Наприклад, у радіомовленні високочастотні (ВЧ) коливання модулюються коливаннями звукової частоти, які передають мовлення, музику.

На рис. 9.1, *м* показано приклад синусоїдального коливання, яке

модульоване за амплітудою і фазою випадковими функціями, а на рис. 9.1, *н* показано одну з реалізацій зовсім неупорядкованого процесу («білого шуму»), який тільки умовно можна віднести до коливань.

9.2. Власні коливання

Найпростішою коливальною системою в механіці є тягарець, який підвішений на пружині та рухається без тертя (рис. 9.2). При цьому ми припускаємо, що масою пружини можна знехтувати порівняно з масою тягарця і що вся пружність належить тільки пружині (система із зосередженими масою і пружністю). З механіки відомо, що тягарець, виведений із стану рівноваги, здійснює гармонічні коливання, за яких його відхилення від положення рівноваги змінюється з часом за законом синуса (косинуса). Коли тягарець перебуває у крайніх положеннях (рис. 9.2, *а* та *в*), кінетична енергія тягарця дорівнює нулю, але потенціальна енергія пружини сягає максимуму. Навпаки, коли тягарець проходить положення рівноваги (рис. 9.2, *б* та *г*), кінетична енергія тягарця має найбільше значення, а потенціальна енергія пружини, яка у цьому положенні і не стискається, і не розтягується, дорівнює нулю, тому механічне коливання, яке розглядається, є **періодичним перетворенням енергії** системи з кінетичної в потенціальну і навпаки.

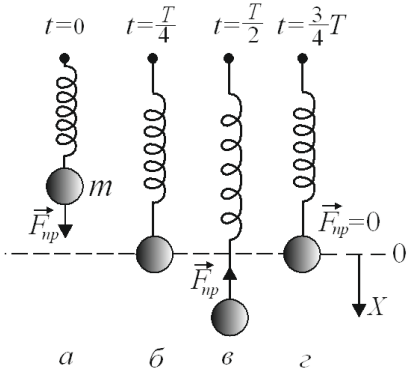


Рис. 9.2. Механічні коливання із зосередженими масою і пружністю: $F_{\text{пр}}$ *– сила пружності пружини*

Аналогічні процеси відбуваються при електричних коливаннях. Найпростіший електричний коливальний контур складається з конденсатора та індуктивності, сполучених між собою (рис. 9.3). Вважатимемо, що ємність між витками котушки дуже мала порівняно з ємністю конденсатора, а індуктивність конденсатора і з'єднувальних проводів мала порівняно з індуктивністю котушки (контур із зосередженими ємністю та індуктивністю, або закритий коливальний контур). Припустимо, що ми, розімкнувши контур, зарядили конденсатор. Між пластинами конденсатора з'явиться електричне поле *E*, яке матиме певну енергію (рис. 9.3, *а*).

Замкнемо тепер конденсатор на індуктивність. Конденсатор почне розряджатись, а його електричне поле зменшуватиметься. При цьому в контурі виникне електричний струм розряджання конденсатора, внаслідок чого в котушці індуктивності виникне магнітне поле B. Через деякий час, що дорівнює чверті періоду коливання, конденсатор розрядиться повністю й електричне поле зникне зовсім, але магнітне поле при цьому досягне максимуму, а отже, енергія електричного поля перетвориться в енергію магнітного поля.

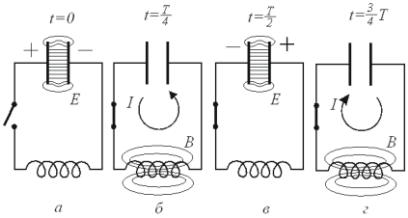


Рис. 9.3. Електричні коливання в контурі із зосередженими індуктивністю та ємністю

У наступні моменти часу магнітне поле зникатиме, бо немає струмів, які б його підтримували. Це зникаюче поле приведе до появи екстраструму самоіндукції, який за законом Ленца прагнутиме підтримати струм заряджання конденсатора, й отже, матиме такий самий напрямок, як і цей останній, тому конденсатор знову перезаряджатиметься і між його пластинами виникне електричне поле протилежного напрямку. Протягом часу, який дорівнює половині періоду коливання, магнітне поле зникне зовсім, а електричне поле досягне максимуму й енергія магнітного поля знову перетвориться в енергію електричного поля. Далі конденсатор знову розряджатиметься, і в контурі виникне струм, направлений протилежно струму в попередній стадії процесу. Після $3T/4$ конденсатор знову буде розряджений, а енергія електричного поля знову перетвориться в енергію магнітного поля (рис. 9.3, *г*) і т. д. Через проміжки часу, які дорівнюють повному періоду коливання T, електричний стан контура буде такий самий, як і на початку коливань (рис. 9.3, *а*).

Якщо опір контура дорівнює нулю, то такий процес періодичного перетворення електричної енергії в магнітну і навпаки триватиме нескінченно довго й ми дістанемо незгасаючі електричні коливання.

Механічні коливання, що виникають під дією сил, які розвиваються у коливальній системі, називають **власними** коливаннями. Вони виникають за всякого порушення рівноваги коливальної системи. Подібно до цього електричні коливання, що відбуваються під дією процесів у коливальному контурі, називають **власними**

електричними коливаннями. Розглянуті вище коливання є, очевидно, власними.

9.3. Рівняння власних коливань. Коливання у разі відсутності загасання

Для тягарця, що підвішений на пружині й рухається вертикально (по осі x) без тертя, рівняння руху згідно з другим законом Ньютона можна записати так:

$$ma_x = F_{пр}, \qquad (9.3)$$

де $a_x = \frac{d^2x}{dt^2} = \ddot{x}$ – прискорення тягарця (друга похідна за часом від його координати зміщення x відносно положення рівноваги); $F_{пр} = -kx$ – сила пружності пружини, яка завжди направлена проти зміщення x; k – коефіцієнт пружності пружини; m – маса тягарця.

Тоді рівняння (9.3) можна переписати у такому вигляді:

$$\ddot{x} + \omega_0^2 x = 0\,, \qquad (9.4)$$

де

$$\omega_0 = \sqrt{\frac{k}{m}} \qquad (9.5)$$

є циклічною частотою коливань. Рівняння (9.4) називають диференціальним рівнянням власних (гармонічних) коливань другого порядку, однорідним з постійними коефіцієнтами. Розв'язком цього рівняння є співвідношення (9.1), які називають рівняннями власних коливань в інтегральному вигляді. Величини A та φ в (9.1) знаходять із початкових умов.

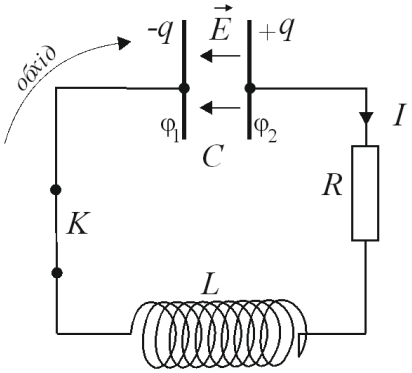


Рис. 9.4.

Тепер кількісно розглянемо власні коливання в електричному контурі на рис. 9.3. Для цього розглянемо коливальний контур (рис. 9.4). Будемо вважати, що електричні процеси в контурі **квазістаціонарні**. Це означає, що миттєве значення величини струму I – одне й те саме в будь-кому місці контуру, і що до миттєвих значень електричних величин можна застосовувати правила Кірхгофа. Нехай заряд конденсатора q_0, а струм у колі $I = \frac{dq}{dt} = \dot{q}$. Тоді згідно з другим правилом Кірхгофа (4.40) можна записати, що

$$IR = \varphi_1 - \varphi_2 + \mathcal{E}_s, \qquad (9.6)$$

де $\varphi_1 - \varphi_2 = -(\varphi_2 - \varphi_1) = -U_c = -\frac{q}{C}$ – напруга на конденсаторі;

$$\mathcal{E}_s = -L\frac{dI}{dt} = -L\frac{d\dot{q}}{dt} = -L\ddot{q} \; -$$

е.р.с. самоіндукції, яка виникає в котушці індуктивності L.

Для випадку ідеалізованого контура $(R = 0)$ рівняння (9.6) запишемо так:

$$\ddot{q} + \frac{1}{LC}q = 0,$$

або

$$\ddot{q} + \omega_0^2 q = 0, \qquad (9.7)$$

де

$$\omega_0 = \frac{1}{\sqrt{LC}}. \qquad (9.8)$$

Рівняння (9.7) математично еквівалентне рівнянню (9.4) і воно є лінійним однорідним диференціальним рівнянням другого порядку із звичайними похідними та зі сталими коефіцієнтами, що описує коливання заряду на конденсаторі в контурі на рис. 9.4. Розв'яжемо рівняння (9.7). Для цього помножимо його на $I = \dot{q}$ й отримаємо:

$$\ddot{q}\dot{q} + \omega^2 q\dot{q} = 0,$$

або

$$\frac{d}{dt}(\dot{q}^2 + \omega_0^2 q^2) = 0.$$

Звідси отримаємо, що $\dot{q}^2 + +\omega_0^2 q^2 = \text{const} = q_0^2\omega_0^2$. Із цього співвідношення знаходимо, що

$$\frac{dq}{dt} = \dot{q} = \omega_0\sqrt{q_0^2 - q^2}$$

і

$$\frac{dq}{\sqrt{q_0^2 - q^2}} = \omega_0 dt.$$

Інтегрування останнього виразу дає, що

$$\arccos\frac{q}{q_0} = \omega_0 t + \varphi_0,$$

або

$$q = q_0\cos(\omega_0 t + \varphi_0). \qquad (9.9)$$

Вираз (9.9) є інтегральним рівнянням власних (гармонічних) коливань заряду на конденсаторі.

Таким чином, заряд на обкладках конденсатора змінюється за гармонічним законом з частотою, яку визначають виразом (9.8). Цю частоту називають **власною частотою контура**. Для періоду коливань отримують **формулу Томсона**:

$$T = 2\pi/\omega_0 = 2\pi\sqrt{LC}. \qquad (9.10)$$

Напруга на конденсаторі відрізняється від заряду на множник $1/C$:

$$U = \frac{q}{C} = \frac{q_0}{C}\cos(\omega_0 t + \varphi_0) = \\ = U_0\cos(\omega_0 t + \varphi_0). \qquad (9.11)$$

Продиференціювавши функцію (9.9) за часом, отримуємо вираз для сили струму:

$$I = -q_0\omega_0\sin(\omega_0 t + \varphi_0) = \\ = I_0\cos(\omega_0 t + \varphi_0 + \pi/2), \qquad (9.12)$$

де $I_0 = q_0\omega_0$ – амплітуда сили струму. Таким чином, сила струму випереджає за

фазою напругу на конденсаторі на $\frac{\pi}{2}$. Врахувавши, що $U_0 = \frac{q_0}{C}$, а $I_0 = q_0\omega_0$, отримуємо, що

$$U_0 = \sqrt{\frac{L}{C}} I_0. \qquad (9.13)$$

Порівнявши це співвідношен-ня із законом Ома (4.8), величину $\sqrt{\frac{L}{C}}$ називають **хвильовим опором** контура.

Енергія електричного поля конденсатора

$$W_c = \frac{q^2}{2C} = \frac{q_0^2}{2C}\cos^2(\omega_0 t + \varphi_0). \quad (9.14)$$

Енергія магнітного поля котушки становитиме

$$W_L = \frac{LI^2}{2} = \frac{Lq_0^2\omega_0^2}{2}\sin^2(\omega_0 t + \varphi_0)$$

або

$$W_L = \frac{q_0^2}{2C}\sin^2\left(\omega_0 t + \varphi_0\right) =$$

$$= \frac{q_0^2}{2C}\cos^2\left(\omega_0 t + \varphi_0 + \pi/2\right). \quad (9.15)$$

Сумарна енергія електричного і магнітного полів

$$W = W_C + W_L = q_0^2/(2C), \quad (9.16)$$

що відповідає енергії, яка була надана контуру через заряджання конденсатора зарядом q_0 (до напруги U_0). Співвідношення (9.16) відображає закон збереження енергії в ідеалізованому контурі. Зауважимо, що формули і співвідношення для механічних коливань можна отримати із формул і співвідношень для електричних коливань, і навпаки, зберігаючи таку заміну:

– заряд конденсатора q – зміщення тягарця від положення рівноваги x;

– сила струму $I = \frac{dq}{dt} = \dot{q}$ – швидкість коливань $\upsilon = \frac{dx}{dt} = \dot{x}$;

– індуктивність контура L – маса тягарця m;

– величина $1/C$ – коефіцієнт пружності пружини k.

9.4. Геометричне зображення гармонічних коливань. Додавання гармонічних коливань однакового напрямку та однакової частоти

Перш ніж розглядати додавання коливальних рухів, проаналізуємо способи зображення коливань за допомогою обертального вектора амплітуди. Розглянемо рівняння (9.1), вважаючи, що зміна величини U еквівалентна зміщенню від положення рівноваги коливальної системи x. Із довільної точки O, яка вибрана на осі X, під кутом φ_0, що дорівнює початковій фазі коливання, відкладемо вектор $\vec{A}$, модуль якого дорівнює амплітуді A коливання (рис. 9.5).

Проєкція вектора $\vec{A}$ на вісь OX дорівнює зміщенню x_0 в момент початку відліку часу ($t = 0$):

$$x_0 = A\cos\varphi_0.$$

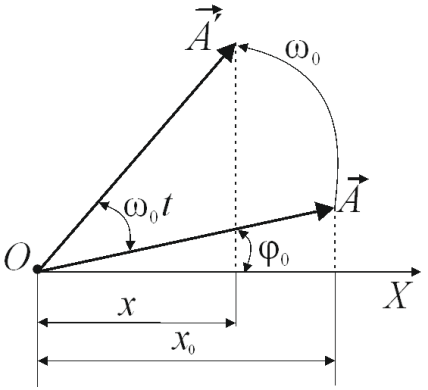


Рис. 9.5. Геометричне зображення гармонічних коливань

Обертатимемо у площині рисунка вектор амплітуди навколо осі OO, яка перпендикулярна до площини рисунка, з постійною кутовою швидкістю ω_0. За проміжок часу t вектор амплітуди повертається на кут $\omega_0 t$. Проєкція вектора $\vec{A}$ в цьому положенні на вісь OX

$$x = A\cos(\omega_0 t + \varphi_0).$$

За час T, що дорівнює періоду коливань, вектор амплітуди повертається на кут 2π, а проєкція його кінця зробить одне повне коливання навколо положення рівноваги O. Отже, обертовий вектор амплітуди повністю характеризує гармонічне коливання.

Нехай точка бере участь у двох гармонічних коливаннях однакової частоти, які напрямлені вздовж однієї прямої:

$$x_1 = A_1\cos(\omega_0 t + \varphi_{01}), \quad (9.17)$$

$$x_2 = A_2\cos(\omega_0 t + \varphi_{02}). \quad (9.18)$$

Ці коливання зручно додати, користуючись методом вектора амплітуди, що обертається. Для цього відкладемо з точки O під кутом φ_{01} вектор амплітуди $\vec{A}_1$, а під кутом φ_{02} – вектор амплітуди $\vec{A}_2$ (рис. 9.6). Оскільки вектори $\vec{A}_1$ та $\vec{A}_2$ обертаються з однаковою кутовою швидкістю, то різниця фаз $\varphi_{02} - \varphi_{01}$ між ними постійна. Оскільки сума проєкцій двох векторів на одну вісь дорівнює проєкції на ту саму вісь вектора, який є їх сумою, то результуюче коливання можна подати вектором амплітуди $\vec{A}$, що дорівнює сумі векторів $\vec{A}_1$ і $\vec{A}_2$:

$$\vec{A} = \vec{A}_1 + \vec{A}_2,$$

який обертається навколо точки O з тією самою кутовою швидкістю ω_0, що й вектори $\vec{A}_1$ і $\vec{A}_2$. Результуюче коливання описують рівнянням

$$x = x_1 + x_2 = A\cos(\omega_0 t + \varphi), \quad (9.19)$$

де A – амплітуда результуючого коливання, а φ – його початкова фаза.

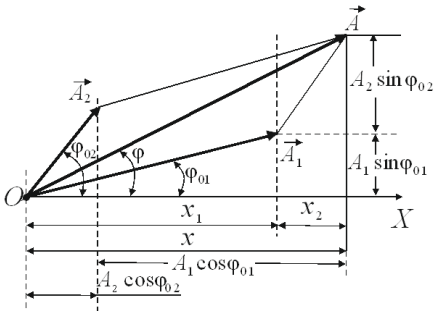


Рис. 9.6. Додавання гармонічних коливань однакового напрямку та однакової частоти

Застосовуючи теорему косинусів до одного із трикутників, на які паралелограм розбивається діагонал-лю, з рис. 9.6 видно, що

$$A^2 = A_1^2 + A_2^2 + 2A_1A_2\cos(\varphi_{02} - \varphi_{01}), \quad (9.20)$$

$$\mathrm{tg}\varphi = \frac{A_1\sin\varphi_{01} + A_2\sin\varphi_{02}}{A_1\cos\varphi_{01} + A_2\cos\varphi_{02}}. \quad (9.21)$$

Амплітуда A результуючого коливання залежить від різниці фаз початкових фаз $\varphi_{02} - \varphi_{01}$ коливань, що додаються. Можливі значення A лежать у межах

$$A_1 + A_2 \geq A \geq |A_2 - A_1|. \quad (9.22)$$

9.5. Додавання гармонічних коливань однакового напрямку з різними частотами. Биття

Якщо частоти коливань (9.17) і (9.18) неоднакові, то вектори $\vec{A}_1$ і $\vec{A}_2$ будуть обертатися з різною швидкістю. У цьому випадку результуючий вектор $\vec{A}$ пульсує за величиною та обертається зі змінною швидкістю. Результуючим рухом у цьому випадку буде не гармонічне коливання, а деякий складний коливальний процес.

Особливий інтерес становить випадок, коли два гармонічні коливання однакового напрямку, що додаються, майже не різняться за частотою.

Періодичні зміни амплітуди коливання, які виникають при додаванні двох гармонічних коливань одного напрямку з близькими частотами, називають **биттям**.

Нехай амплітуди коливань $A_1 = A_2 = A$, а частоти дорівнюють $\omega_1 = \omega_0$, $\omega_2 = \omega_0 + \Delta\omega$ та $\Delta\omega \ll \omega_0$.

Тоді рівняння коливань матимуть такий вигляд:

$$x_1 = A\cos\omega_0 t; \quad x_2 = A\cos(\omega_0 t + \Delta\omega t). \quad (9.23)$$

Порівнюючи (9.23) з (9.17) та (9.18), можемо вважати, що $\varphi_{01} = 0$, а $\varphi_{02} = \Delta\omega t$.

Використаємо формулу (9.20) і знайдемо амплітуду результуючого коливання (биття):

$$A_б^2 = 2A^2(1 + \cos\Delta\omega t) =$$

$$= 4A^2\cos^2\frac{\Delta\omega}{2}t.$$

Звідси знайдемо, що

$$A_б = 2A\left|\cos\frac{\Delta\omega}{2}t\right|. \qquad (9.24)$$

Фазу результуючого коливання знайдемо за допомогою формули (9.21), а саме:

$$\mathrm{tg}\varphi = \frac{\sin\Delta\omega t}{1+\cos\Delta\omega t} = \mathrm{tg}\frac{\Delta\omega}{2}t,$$

звідки

$$\varphi = \frac{\Delta\omega}{2}t.$$

Тоді результуюче коливання можна записати у такому вигляді:

$$x = 2A\left|\cos\frac{\Delta\omega}{2}t\right|\cos\left(\omega_0 t + \frac{\Delta\omega t}{2}\right). \qquad (9.25)$$

Для випадку, коли $\omega_0 \gg \Delta\omega$, формулу (9.25) записують так:

$$x = 2A\left|\cos\frac{\triangle\omega t}{2}\right|\cos\omega_0 t. \qquad (9.26)$$

Отриманий вираз (9.26) є добутком двох коливань. Оскільки $\Delta\omega \ll \omega_0$, то множник $2A\left|\cos\frac{\triangle\omega}{2}t\right|$ майже не зміниться, коли множник $\cos\omega_0 t$ здійснює кілька повних коливань, тому результуюче коливання x можна розглядати як гармонічне з частотою ω_0 й амплітудою (9.24). Частота зміни $A_б$ вдвоє більша за частоту зміни косинуса (оскільки її беруть за модулем). Частота биття дорівнює різниці частот коливань, що додаються, тобто $\omega_б = \Delta\omega$. Період биття

$$T_б = 2\pi / \Delta\omega. \qquad (9.27)$$

Суцільні лінії на рис. 9.7 дають графік результуючого коливання (9.26) у випадку $\omega_0/\Delta\omega = 10$ і графік амплітуди $A_б$ (пунктирна лінія).

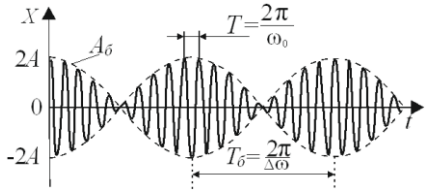


Рис. 9.7 Биття

9.6. Додавання взаємно перпендикулярних коливань

Нехай матеріальна точка M одночасно бере участь у двох гармонічних коливаннях з однаковою частотою у двох взаємно перпендикулярних напрямках, як уздовж осі X, так і вздовж осі Y (рис. 9.8). Якщо збудити обидва коливання, матеріальна точка буде рухатись уздовж деякої криволінійної траєкторії, форма якої залежить від різниці фаз обох коливань.

Виберемо початок відліку часу так, щоб початкова фаза першого коливання дорівнювала нулю. Тоді рівняння коливань матимуть такий вигляд:

$$\begin{aligned} x &= A\cos\omega_0 t, \\ y &= B\cos(\omega_0 t + \varphi), \end{aligned} \qquad (9.28)$$

де φ – різниця фаз обох коливань; A і B – їх амплітуди коливань.

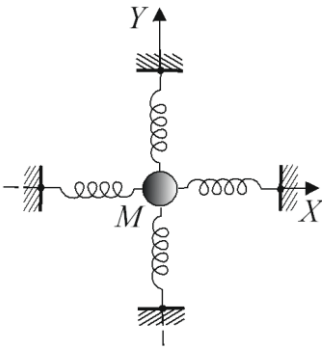


Рис. 9.8. Модель двох взаємно перпендикулярних гармонічних коливань

Ці вирази – параметрична форма рівняння траєкторії, вздовж якої рухається точка, що бере участь в обох коливаннях. Щоб отримати рівняння траєкторії у звичайному вигляді, треба виключити з цих рівнянь параметр t. Проведемо такі перетворення:

$$\frac{x}{A} = \cos\omega_0 t,\ \sin\omega_0 t = \sqrt{1-\frac{x^2}{A^2}},$$

$$\frac{y}{B} = \cos(\omega_0 t + \varphi) = \cos\omega_0 t \cdot \cos\varphi -$$

$$-\sin\omega_0 t \cdot \sin\varphi = \frac{x}{A}\cos\varphi - \sqrt{1-\frac{x^2}{A^2}}\sin\varphi;$$

$$\left(\frac{y}{B} - \frac{x}{A}\cos\varphi\right)^2 = \left(-\sqrt{1-\frac{x^2}{A^2}}\sin\varphi\right)^2;$$

$$\frac{x^2}{A^2}\cos^2\varphi - 2\frac{x}{A}\frac{y}{B}\cos\varphi + \frac{y^2}{B^2} =$$

$$= -\frac{x^2}{A^2}\sin^2\varphi + \sin^2\varphi.$$

У результаті отримаємо:

$$\frac{x^2}{A^2} - 2\frac{x}{A}\frac{y}{B}\cos\varphi + \frac{y^2}{B^2} = \sin^2\varphi. \quad (9.29)$$

Це рівняння еліпса, осі якого повернуті відносно координатних осей OX і OY. Орієнтація еліпса і величини його півосей залежать від амплітуд A і B та різниці фаз φ.

Розглянемо окремі випадки:

1) $\varphi = \pm 2m\pi,\ m = 0, 1, 2, 3, \ldots$.

Тоді

$$\left(\frac{x}{A} - \frac{y}{B}\right)^2 = 0.$$

Звідси $y = \frac{B}{A}x$. Результуюче коливання є гармонічним уздовж прямої з частотою ω та амплітудою $\sqrt{A^2 + B^2}$ (рис. 9.9). Пряма утворює з віссю OX кут $\alpha = \operatorname{arctg}\frac{B}{A}$;

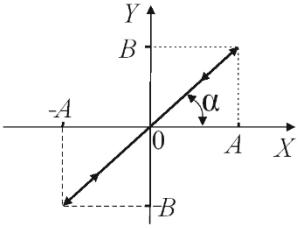


Рис. 9.9

2) $\varphi = \pm(2m+1)\pi,\ m = 0,\ 1,\ 2,\ 3, \ldots$.

У цьому випадку $\left(\frac{x}{A} + \frac{y}{B}\right)^2 = 0$ й $y = -\frac{B}{A}x$. Результуючий рух є гармонічним коливанням вздовж прямої $y = -\frac{B}{A}x$ (рис. 9.10);

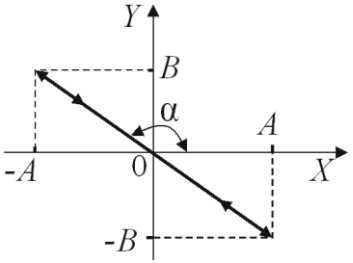


Рис. 9.10

3) $\varphi = \pm(2m+1)\frac{\pi}{2}$,

де m = 0, 1, 2, 3, …. . Тоді рівняння траєкторії буде:

$$\frac{x^2}{A^2} + \frac{y^2}{B^2} = 1.$$

Це рівняння еліпса, осі якого збігаються з осями координат, а його пів осі дорівнюють відповідним амплітудам (рис. 9.11).

Якщо $A = B$, то еліпс вироджується в коло. Випадки

$\varphi = (2m+1)\frac{\pi}{2}$ та $\varphi = -(2m+1)\frac{\pi}{2}$

відрізняються напрямком руху по еліпсу (колу) (за годинниковою стрілкою чи проти неї).

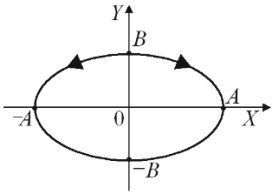


Рис. 9.11

Якщо частоти взаємно перпендикулярних коливань, що додаються, різні, то замкнена траєкторія результуючого коливання досить складна.

Замкнені траєкторії, що кресляться точкою, яка здійснює одночасно два взаємно перпендикулярні коливання з різними частотами, називають фігурами **Ліссажу**. Форма цих кривих залежить від співвідношення амплітуд, частот і різниці фаз коливань, що додаються (рис. 9.12).

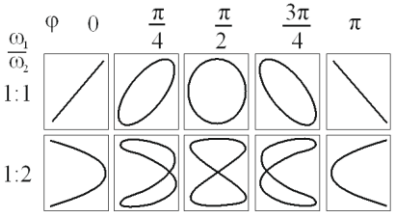


Рис. 9.12. Фігури Ліссажу

Фігури Ліссажу можна спостерігати, наприклад, на екрані електронно-променевого осцилографа, якщо до двох пар пластин, які відхиляють електронний промінь, підведені змінні напруги з рівними або кратними частотами. Спостереження фігур Ліссажу – зручний метод дослідити співвідношення між частотами і фазами коливань, а також і форми коливань. За допомогою фігур Ліссажу знаходять невідомі частоти коливань.

9.7. Диференціальне рівняння загасаючих коливань та його розв'язання

Розглянемо цей випадок на прикладі електричних коливань, що виникають у контурі, показаному на рис. 9.4, що має активний опір R. Енергія, що запасається у контурі, у процесі коливань в контурі поступово витрачається **на нагрівання** опору R, унаслідок чого вільні коливання згасають.

Диференціальне рівняння загасаючих коливань отримаємо із (9.6), розділивши його на L:

$$\ddot{q} + \frac{R}{C}\dot{q} + \frac{1}{LC}q = 0.$$

Взявши до уваги формулу (9.8) і ввівши позначення

$$\beta = \frac{R}{2L}, \qquad (9.30)$$

який називають *коефіцієнтом загасання*, отримаємо рівняння

$$\ddot{q} + 2\beta\dot{q} + \omega_0^2 q = 0. \qquad (9.31)$$

Ми дістали лінійне диференціальне рівняння другого порядку із звичайними похідними й зі сталими коефіцієнтами. Таке саме рівняння можна отримати для напруги U і для величини струму I. Зауважимо, що коливання, які описують лінійними диференціальними рівняннями, називають лінійними коливаннями, а відповідні коливальні системи – лінійними системами.

Розв'язок рівняння (9.31) будемо шукати у вигляді

$$q = u(t)e^{-\beta t}. \qquad (9.32)$$

Підставимо (9.32) в (9.31). Врахуємо, що $\dot{q} = (\dot{u} - \beta u)\, e^{-\beta t}$ та

$$\ddot{q} = (\ddot{u} - 2\beta\dot{u} + \beta^2 u)e^{-\beta t}.$$

Тоді для змінної $u(t)$ отримаємо рівняння

$$\ddot{u} + (\omega_0^2 - \beta^2)u = 0. \qquad (9.33)$$

Розв'язок цього рівняння має різний вигляд залежно від співвідношення між коефіцієнтами.

9.1. Припустимо спочатку, що

$$\omega_0^2 > \beta^2, \qquad (9.34)$$

й тоді

$$\omega = \sqrt{\omega_0^2 - \beta^2}. \qquad (9.35)$$

За такої умови рівняння (9.33) перепишемо так:

$$\ddot{u} + \omega^2 u = 0,$$

розв'язок якого за аналогією з (9.7) має вигляд

$$u = q_0 \cos(\omega t + \varphi),$$

де q_0 і φ – сталі, які визначаються з початкових умов. Отже, розв'язком рівняння (9.31) буде

$$q = q_0 e^{-\beta t} \cos(\omega t + \varphi), \qquad (9.36)$$

де $q_0 e^{-\beta t}$ – «амплітуда» загасаючих коливань; q_0 – початкова амплітуда (максимальний заряд конденсатора).

Амплітуда загасаючих коливань зменшується із плином часу і тим

швидше, чим більший коефіцієнт β. Величину (9.35) називають *власною циклічною частотою загасаючих коливань*. Графік залежності q від часу зображено на рис. 9.13.

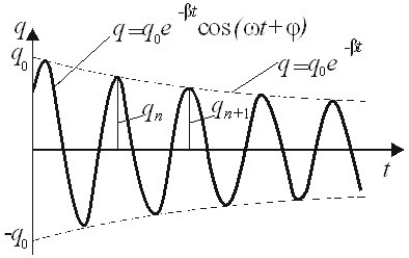


Рис. 9.13. Загасаючі коливання

Щоб знайти силу струму в контурі, продиференціюємо (9.36) за часом:

$$I = \frac{dq}{dt} = \dot{q} = q_0 e^{-\beta t} \times$$
$$\times (-\beta\cos(\omega t + \varphi) - \omega \sin(\omega t + \varphi)). \quad (9.37)$$

Помножимо і розділимо цей вираз на $\sqrt{\omega^2 + \beta^2} = \sqrt{\omega_0^2} = \omega_0$.

Тоді

$$I = \omega_0 q_0 e^{-\beta t}\left(-\frac{\beta}{\sqrt{\omega^2+\beta^2}}\cos(\omega t+\varphi) - \right.$$
$$\left. -\frac{\omega}{\sqrt{\omega^2+\beta^2}}\sin(\omega t+\varphi)\right).$$

Якщо ввести кут ψ, який визначають умовами:

$$\cos\psi = -\frac{\beta}{\sqrt{\omega^2+\beta^2}} = -\frac{\beta}{\omega_0},$$

$$\sin\psi = \frac{\omega}{\sqrt{\omega^2+\beta^2}} = \frac{\omega}{\omega_0},$$

то можна написати, що

$$I = \omega_0 q_0 e^{-\beta t}\cos(\omega t + \varphi + \psi). \quad (9.38)$$

Оскільки $\cos\psi < 0$, а $\sin\psi > 0$, то значення ψ знаходять у межах від $\pi/2$ до π ($\pi/2 < \psi < \pi$). Таким чином, за наявності в контурі активного опору сила струму за фазою випереджає напругу (заряд) на конденсаторі більше, ніж на $\pi/2$ (за $R = 0$ випередження становить $\pi/2$).

Дослідимо докладніше розв'язок (9.36) і знайдемо насамперед ті моменти часу, в які заряд q досягає максимумів і мінімумів. Для цього за правилом знаходження екстремумів прирівняємо (9.37) до нуля. Це дає

$$\operatorname{tg}(\omega t + \varphi) = -\frac{\beta}{\omega}.$$

Нехай $t = t_1$ є яким-небудь розв'язком цього рівняння. Тоді його розв'язками також будуть:

$$\ldots, t_1 - 2T, t_1 - T, t_1 + T, t_1 + 2T, \ldots;$$
$$\ldots, t_1 - \frac{3}{2}T, t_1 - \frac{T}{2}, t_1 + \frac{T}{2}, t_1 + \frac{3}{2}T, \ldots$$

де

$$T = 2\pi/\omega = 2\pi/\sqrt{\omega_0^2 - \beta^2}. \quad (9.39)$$

При цьому легко впевнитись, що коли за $t = t_1$ маємо максимум q, то всі значення t, що стоять у першому рядку,

відповідають також максимуму ($\frac{d^2q}{dt^2} < 0$), а всі значення другого рядка – мінімуму q. Отже, хоч згасаючі коливання не є періодичним процесом у строгому розумінні цього слова, однак цей процес має певну ***повторюваність*** у тому розумінні, що максимальних і мінімальних значень заряд (а також струм і напруга) набувають через однакові проміжки часу T. Цей проміжок часу T (формула (9.39)) ми і називаємо **періодом** загасаючих коливань.

Нехай q_n і q_{n+1} – максимальні значення заряду конденсатора (рис. 9.13) у двох послідовних максимумах з номерами n і $(n+1)$ у моменти часу t_n і t_{n+1}, причому $t_{n+1} = t_n + T$. Згідно з (9.36) і (9.39) маємо:

$$q_n = q_0 e^{-\beta t_n} \cos(\omega t_n + \varphi),$$

$$q_{n+1} = q_0 e^{-\beta(t_n+T)} \times$$

$$\times \cos\left[\omega\left(t_n + \frac{2\pi}{\omega}\right) + \varphi\right] =$$

$$= q_0 e^{-\beta(t_n+T)} \cos(\omega t_n + \varphi).$$

Поділивши почленно обидві ці рівності, знаходимо:

$$D = \frac{q_n}{q_{n+1}} = e^{\beta T}. \qquad (9.40)$$

Ми бачимо, що відношення двох послідовних максимальних значень заряду не залежить від номера максимумів, або, інакше кажучи, ці максимальні значення зменшуються в геометричній прогресії. Відношення (9.40) називають **декрементом загасання**, а його натуральний логарифм

$$\lambda = \ln D = \beta T \qquad (9.41)$$

називають **логарифмічним декрементом загасання**.

Логарифмічний декремент загасання λ можна визначити й інакше. Позначимо через τ час, протягом якого амплітуда коливань зменшується в $e = 2{,}718\,28\ldots$ разів. Тоді згідно з (9.36)

$$\frac{q_0 e^{-\beta(t+\tau)}}{q_0 e^{-\beta t}} = e^{-\beta\tau} = \frac{1}{e},$$

а отже,

$$\beta\tau = 1.$$

Поділивши почленно (9.41) на останнє співвідношення, отримаємо:

$$\frac{T}{\tau} = \frac{1}{N_e} = \lambda,$$

де N_e – кількість повних коливань за час τ, який називають **часом релаксації**. Отже, логарифмічний декремент є величиною, оберненою до кількості коливань, після якого амплітуда зменшується в e разів.

Щоб охарактеризувати загасання коливальних контурів, часто користуються, особливо в радіотехніці, ще й іншою величиною, яку називають **добротністю контура** і позначають літерою Q. З логарифмічним декрементом добротність контура пов'язана співвідношенням

$$Q = \pi/\lambda. \qquad (9.42)$$

Через те, що $\lambda = 1/N_e$, то

$$Q = \pi N_e. \qquad (9.43)$$

Добротність контура є помноженою на π кількість повних коливань, після закінчення яких амплітуда зменшується в e разів. Отже, добротність контура тим вища, чим меншим є загасання коливань у ньому.

Формула (9.35) показує, що частота електричних коливань ω залежить від коефіцієнта загасання β і не дорівнює частоті коливань ω_0 того самого контура за опору $R = 0$ ($\beta = 0$). Зі збільшенням опору контура частота ω зменшується, а період коливань T збільшується.

Покажемо, що за слабкого загасання (коли $\beta^2 \ll \omega_0^2$) добротність коливальної системи з точністю до множника 2π дорівнює відношенню енергії, яка запасена в системі у цей момент, до зменшення цієї енергії за один період коливань. Згідно з (9.36) амплітуда заряду конденсатора зменшується за законом $e^{-\beta t}$. Енергія, яка запасається в конденсаторі, пропорційна квадрату амплітуди заряду, значить енергія конденсатора W зменшується за законом $e^{-2\beta t}$. Відносне зменшення енергії за період коливань

$$\frac{\Delta W}{W} = \frac{W(t) - W(t+T)}{W(t)} = \\ = \frac{1 - e^{-2\beta T}}{1} = 1 - e^{-2\lambda}.$$

Якщо загасання слабке (тобто за умови, що $\lambda \ll 1$), за формулами приблизного підрахунку можна вважати, що $e^{-2\lambda} \approx 1 - 2\lambda$. Тоді

$$\frac{\Delta W}{W} = 1 - (1 - 2\lambda) = 2\lambda.$$

Скориставшись формулою (9.42) отримаємо, що

$$Q = 2\pi \frac{W}{\Delta W}. \qquad (9.44)$$

9.2. Розглянемо випадок, коли $\omega_0^2 = \beta^2$. Це відбуватиметься, коли опір коливального контура

$$R = 2\sqrt{L/C} = R_{\text{кр}},$$

де $R_{\text{кр}}$ називають критичним опором. Тоді (9.33) набуде вигляду $\ddot{u} = 0$. Розв'язком цього рівняння є: $u = a + bt$. Це означає, що розв'язок (9.32) буде мати вигляд

$$q = (a + bt)e^{-\beta t},$$

де a і b визначають з початкових умов, наприклад, $t = 0$, $q = 0$, $I = I_0$. Тоді $a = 0$, а $b = I_0$ й

$$q = I_0 t \cdot e^{-\beta t}.$$

Цей розв'язок описує «аперіодичний процес», графік якого показаний на рис. 9.14, *а*. Коливальна система, що виведена з положення

рівноваги, повертається (експоненцій-но) у рівноважний стан. Такий процес використовують у балістичних гальванометрах, приладах для вимірювання кількості заряду, який протік через поперечний переріз провідника в електричному колі.

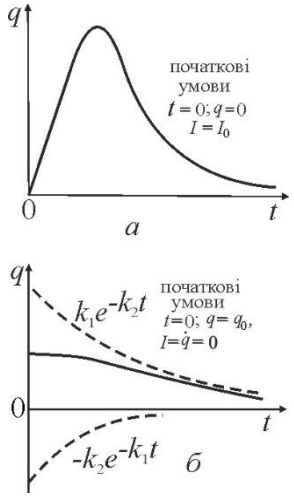


Рис. 9.14. Аперіодичний процес у коливальному контурі (а) та аперіодичний розряд конденсатора (б)

9.3. Припустимо тепер, що опір контура великий наскільки, що

$$\omega_0^2 < \beta^2.$$

У цьому випадку частота ω, яку виражають формулою (9.35), буде уявною. Це означає, що розв'язку (9.36) вже не існує, а отже, електричних коливань у контурі не буде. В цьому випадку розв'язок основного рівняння (9.31) має вигляд

$$q = A_1 e^{-k_1 t} + A_2 e^{-k_2 t}, \qquad (9.45)$$

де

$$k_1 = \beta + \sqrt{\beta^2 - \omega_0^2},$$

$$k_2 = \beta - \sqrt{\beta^2 - \omega_0^2},$$

а A_1 і A_2 – довільні сталі. Підставляючи (9.45) у (9.31) можна впевнитись, що рівняння при цьому задовольняється тотожно, а отже, (9.45) є справді шуканим розв'язком. Через те, що $\omega_0^2 < \beta^2$, то обидва k_1 і k_2 є дійсними й додатними.

Значення сталих A_1 і A_2 визначають із початкових умов задачі. Якщо такими є умови

$$t = 0:\ q = q_0,\ \ I = \dot{q} = 0,$$

то

$$q|_{t=0} = A_1 + A_2 = q_0,$$

$$\dot{q}|_{t=0} = -A_1 k_1 - A_2 k_2 = 0.$$

Це дає, що

$$A_1 = -q_0 \frac{k_2}{k_1 - k_2},$$

$$A_2 = q_0 \frac{k_1}{k_1 - k_2}.$$

Після цього розв'язок (9.45) набуває вигляду

$$q = \frac{q_0}{k_1 - k_2}(k_1 e^{-k_2 t} - k_2 e^{-k_1 t}).$$

На рис. 9.14, *б* графічно зображено обидва доданки цієї формули (пунктир) і їх суму (суцільна крива), яка відображає аперіодичний розряд конденсатора.

Якщо опір контура дуже великий, так що $\beta^2 \gg \omega_0^2$, то $k_1 \gg k_2$, і в останньому виразі можна знехтувати другим доданком порівняно з першим, а у знаменнику $-k_2$ порівняно з k_1. Тоді

$$q = q_0 e^{-k_2 t}.$$

Цей випадок відповідає розряду конденсатора.

Із зазначеного вище видно, що для можливості електричних коливань потрібно, щоб виконувалась умова (9.34). Підставляючи замість β і ω_0 їх значення (9.30) і (9.8) знаходимо умову можливості коливань у вигляді

$$\frac{1}{LC} > \frac{R^2}{4L^2},$$

або

$$R < 2\sqrt{L/C} = R_{\text{кр}}. \qquad (9.46)$$

Формула (9.46) визначає *критичний опір,* за якого коливальний процес переходить в аперіодичний.

9.8. Вимушені електричні коливання. Резонанс

Для техніки зв'язку дуже важливе місце займають вимушені коливання в коливальній системі.

9.8.1. Вимушені коливання виникають тоді, коли на коливальну систему впливає зовнішня періодично змінна дія. У випадку електричних коливань це можна здійснити, якщо в електричне коло (рис. 9.4) включити послідовно з елементами контура змінну е.р.с. або, розірвавши контур, подати на контакти, які утворились, періодичну напругу (рис. 9.15):

$$U = U_m \cos \omega t, \qquad (9.47)$$

де U_m – амплітуда, а ω – частота вимушуючої сили. Цю напругу слід додати до е.р.с. самоіндукції в рівнянні (9.6). У результаті формула (9.6) набуває вигляду

$$IR = -\frac{q}{c} - L\frac{dI}{dt} + U_m \cos \omega t. \qquad (9.48)$$

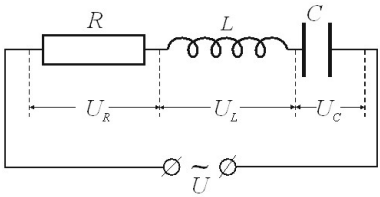


Рис. 9.15. Послідовний коливальний контур

Після перетворення цього рівняння отримаємо:

$$\ddot{q} + 2p\dot{q} + \omega_0^2 q = \frac{U_m}{L}\cos \omega t, \qquad (9.49)$$

де ω_0 і β визначають формулами (9.8) і (9.30). Рівняння (9.49) називають диференціальним, неоднорідним, другого порядку рівнянням вимушених коливань.

Загальним розв'язком рівняння (9.49) є сума загального розв'язку лінійного однорідного рівняння (9.31), тобто розв'язок (9.36), і частинного розв'язку рівняння (9.49). Розв'язок (9.36) відіграє помітну роль лише на початковій

стадії виникнення вимушених коливань. З часом унаслідок експоненціального множника $e^{-\beta t}$ роль розв'язку (9.36) зменшується, амплітуда вимушених коливань зростає, доки не досягне певного значення.

Серед частинних розв'язків рівняння (9.49) є такий, який змінюється з часом синусоїдально з частотою ω зовнішньої сили, що періодично змінюється (9.47). Будемо шукати його в комплексній формі, що можна робити, оскільки всі математичні операції, з якими доведеться мати справу, є лінійними й дійсними. Замінимо праву частину рівняння (9.49) на комплексну величину $U_m e^{i\omega t}$, тобто напишемо, що

$$\ddot{q} + 2\beta\dot{q} + \omega_0^2 q = \frac{U_m e^{i\omega t}}{L}. \quad (9.50)$$

Будемо шукати частинний розв'язок рівняння (9.50) у комплексному вигляді. Знайшовши його, відкинемо його уявну частину й отримаємо розв'язок рівняння (9.49). Частинний розв'язок рівняння (9.50) шукаємо у вигляді

$$q = q_1(\omega)e^{i\omega t}, \quad (9.51)$$

звідки отримаємо, що

$$\dot{q} = \frac{dq}{dt} = i\omega q_1 e^{i\omega t} = i\omega q,$$

$$\ddot{q} = \frac{d^2q}{dt^2} = i^2\omega^2 q_1 e^{i\omega t} = -\omega^2 q.$$

Підставимо (9.51) в (9.50) й отримаємо

$$q_1 = \frac{U_m/L}{\omega_0^2 - \omega^2 + i2\beta\omega}, \quad (9.52)$$

тоді частинний розв'язок (9.51) запишемо так:

$$q = \frac{U_m/L}{\omega_0^2 - \omega^2 + i2\beta\omega} e^{i\omega t}. \quad (9.53)$$

Розв'язок (9.53) тільки символічно описує вимушене коливання. В ньому має бути залишена лише ***дійсна частина***. Щоб знайти цю частину, введемо позначення

$$\omega_0^2 - \omega^2 + i2\beta\omega = \rho e^{i\psi}, \quad (9.54)$$

де ρ і ψ – дійсні величини, які визначають згідно з теорією комплексних чисел за формулами:

$$\rho = \sqrt{(\omega_0^2 - \omega^2)^2 + 4\beta^2\omega^2}, \quad (9.55)$$

$$\mathrm{tg}\psi = \frac{2\beta\omega}{\omega_0^2 - \omega^2} = \frac{R}{\frac{1}{\omega C} - \omega L}. \quad (9.56)$$

Таким чином, розв'язок (9.53) може бути записаний, як

$$q = q_m e^{i(\omega t - \psi)}, \quad (9.57)$$

де

$$q_m = \frac{U_m/L}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4\beta^2\omega^2}} \quad (9.58)$$

є амплітудою вимушених коливань. Виділивши у (9.57) дійсну частину, отримаємо частинний розв'язок рівняння (9.49) у вигляді

$$q = q_m \cos(\omega t - \psi), \quad (9.59)$$

який для часу $t \gg \tau = 1/\beta$ стане *загальним розв'язком рівняння* (9.49), оскільки вільні коливання, описувані рівнянням (9.36), фактично зовсім згаснуть.

Таким чином, вимушені коливання будуть гармонічними з частотою ω зовнішньої сили, що періодично змінюється (9.47). Амплітуда q_m цих коливань залежить від ω, її визначають за формулою (9.58). Коливання (9.59) відстають за фазою від прикладеної напруги (9.47) на величину ψ, яка визначається формулою (9.56) і теж залежить від ω.

9.8.2. Дослідимо отриманий результат. Продиференціюємо розв'язок (9.59) за t і знайдемо силу струму в контурі для коливань, які встановились:

$$I = \dot{q} = -\omega q_m \sin(\omega t - \psi) = \\ = I_m \cos(\omega t - \psi + \pi/2),$$

де $I_m = \omega q_m$.

Запишемо цей вираз у вигляді

$$I = I_m \cos(\omega t - \varphi), \quad (9.60)$$

де $\varphi = \psi - \pi/2$ – зсув за фазою між струмом і прикладеною напругою (9.47). Відповідно до (9.56)

$$\mathrm{tg}\varphi = \mathrm{tg}\left(\psi - \frac{\pi}{2}\right) = -\frac{1}{\mathrm{tg}\psi} = \frac{\omega L - \frac{1}{\omega C}}{R} \quad (9.61)$$

З цієї формули видно, що струм за фазою відстає від напруги (9.47) ($\varphi > 0$) у тому випадку, коли $\omega L > 1/(\omega C)$ й випереджає напругу (9.47) ($\varphi < 0$) за умови, що $\omega L < 1/(\omega C)$. Згідно з (9.58)

$$I_m = \omega q_m = \frac{U_m}{\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}}. \quad (9.62)$$

Порівнюючи (9.62) із законом Ома (4.8), бачимо, що $\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}$ визначає опір кола, де $R-$ активний опір; ωL і $1/(\omega C)-$ реактивні опори (буде розглянуто далі).

Подамо співвідношення (9.48) у вигляді

$$IR + \frac{q}{C} + L\frac{dI}{dt} = U_m \cos \omega t. \quad (9.63)$$

Добуток IR дорівнює напрузі U_R на активному опорі, q/C є напругою на конденсаторі U_c, вираз $L(dI/dt)$ визначає напругу U_L на індуктивності. Тоді можна записати, що

$$U_R + U_C + U_L = U_m \cos \omega t. \quad (9.64)$$

Таким чином, ця сума напруг (це векторна сума коливань) на окремих елементах контура в кожний момент часу дорівнює напрузі, яка прикладена зовні (рис. 9.15).

Згідно з (9.60)

$$U_R = RI_m \cos(\omega t - \varphi). \quad (9.65)$$

Розділивши вираз (9.59) на ємність, отримаємо напругу на конденсаторі:

$$U_C = \frac{q_m}{C}\cos(\omega t - \psi) = \\ = U_{cm} \cos(\omega t - \varphi - \pi/2\,). \quad (9.66)$$

Тут $U_{cm} = q_m/C = I_m/(\omega C)$, де враховано, що $I_m = \omega q_m$.

Помноживши похідну за часом від функції (9.59) на L, отримаємо напругу на індуктивності:

$$U_L = L(dI/dt) = -\omega L I_m \sin(\omega t - \varphi) = \\ = U_{Lm}\cos(\omega t - \varphi + \pi/2). \quad (9.67)$$

Тут

$$U_{Lm} = \omega L I_m. \quad (9.68)$$

Зіставлення формул (9.60), (9.65), (9.66), (9.67) показує, що напруга на ємності за фазою відстає від сили струму на $\pi/2$, а напруга на індуктивності випереджає струм на $\pi/2$. Напруга на активному опорі змінюється у фазі зі струмом. Згідно з геометричним поданням гармонічних коливань фазові співвідношення можна наочно подати за допомогою векторної діаграми. Візьмемо пряму, від якої відраховують початкову фазу, за вісь струмів. Тоді вийде діаграма, яку показано на рис. 9.16. Згідно з (9.63) три функції (U_R, U_C і U_L) в сумі мають дорівнювати прикладеній напрузі U. Відповідно до цього напруга U зображено на діаграмі вектором, який дорівнює сумі векторів U_R, U_C і U_L. Зауважимо, що із прямокутного трикутника, який утворений на діаграмі векторами U, U_R і різницею $U_L - U_C$, легко отримати формулу (9.62).

9.8.3. Резонанс (франц. resonance, від лат. resono – звучу у відповідь, відзиваюсь) відносно великий селективний (вибірковий) відгук коливальної системи на періодичну дію з частотою, близькою до частоти її власних коливань. Під час резонансу відбувається різке зростання амплітуди вимушених коливань. Резонанс як механічне та акустичне явище вперше описав італійський учений Г. Галілей, а в електромагнітних системах – на прикладі коливального контура – англійський учений Дж. Максвелл (1868 р.).

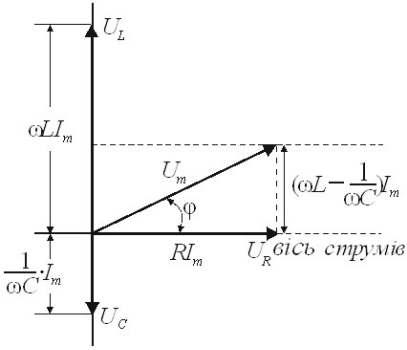


Рис. 9.16. Векторна діаграма напруг в коливальному контурі

Резонансну частоту для заряду q і напруги на конденсаторі U_C знайдемо з умови мінімуму підкореневого виразу у формулі (9.58), тобто похідна за ω від цього виразу має дорівнювати нулю:

$$\frac{d}{d\omega}((\omega_0^2 - \omega^2)^2 + 4\beta^2\omega^2) = 0.$$

Звідси знаходимо, що

$$\omega_{q_{\text{рез}}} = \omega_{U_{\text{рез}}} = \sqrt{\omega_0^2 - 2\beta^2} = \\ = \sqrt{\frac{1}{LC} - \frac{R^2}{2L^2}} \leq \omega_0. \quad (9.69)$$

Резонансні криві для U_C зображені на рис. 9.17 (резонансні криві для q мають

такий самий вигляд). За $\omega \to 0$ резонансні криві сходяться в одній точці з ординатою $U_{C_m} = U_m$. Це означає, що конденсатор підключений до джерела постійної напруги U_m. Максимум при резонансі тим вищий і гостріший, чим менший коефіцієнт згасання $\beta = R/(2L)$, тобто чим менший активний опір і більша індуктивність контура.

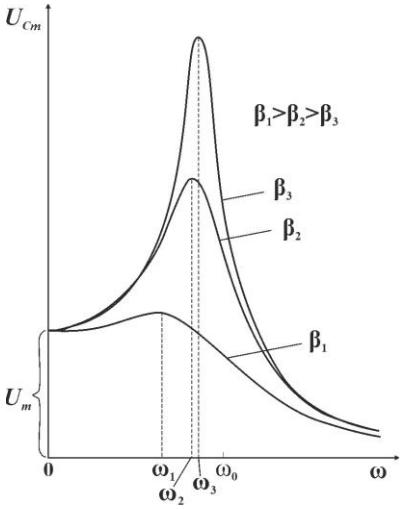


Рис. 9.17. Резонансні криві для напруги на конденсаторі

Резонансні криві для сили струму зображені на рис. 9.18 для різних коефіцієнтів згасання β. Із формули (9.62) видно, що амплітуда сили струму має максимальне значення за $\omega L - \frac{1}{(\omega C)} = 0$. Отже, резонансна частота для сили струму збігається із власною частотою контура ω_0:

$$\omega_{I_{\text{рез}}} = \omega_0 = 1/\sqrt{LC}. \qquad (9.70)$$

Якщо $\omega \to 0$, $I_m = 0$, за постійної напруги постійний струм в колі з конденсатором протікати не може. Згідно з (9.62) за $\omega \to \omega_0$ маємо $I_{m\,\text{рез}} = U_m/R$.

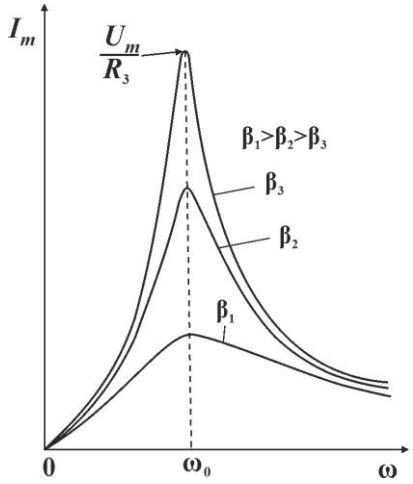


Рис. 9.18. Резонансні криві для сили струму в коливальному контурі

Резонансне значення напруги на конденсаторі знайдемо з формули (9.58), поділивши її на ємність C і підставивши замість частоти ω її резонансне значення з формули (9.69). Тоді отримаємо, що

$$U_{cm\,\text{рез}} = \frac{U_m\omega_0^2}{2\beta\sqrt{\omega_0^2 - \beta^2}}. \qquad (9.71)$$

Особливий інтерес становить випадок слабкого згасання (коли $\beta^2 \ll \omega_0^2$). Тоді

$$\frac{U_{cm\,\text{рез}}}{U_m} = \frac{1}{\omega_0 CR} = \frac{\sqrt{LC}}{CR} = \frac{1}{R}\sqrt{\frac{L}{C}} = Q. \qquad (9.72)$$

При цьому враховано, що для випадку слабкого згасання

$$Q = \frac{\pi}{\lambda} = \frac{\pi}{\beta T} \cong \frac{\pi\omega_0}{2\pi\beta} = \frac{2L}{2\sqrt{LC}\cdot R} = \frac{1}{R}\sqrt{\frac{L}{C}}.$$

Таким чином, добротність контура (для слабкого згасання) показує, у скільки разів напруга на конденсаторі може перевищувати прикладену напругу.

Добротність контура визначає також гостроту резонансних кривих. На рис. 9.19 показано одну з резонансних кривих для сили струму в контурі. По вертикальній осі відкладено не значення I_m (формула (9.62)), що відповідає цій частоті, а відношення $I_m/I_{m\,\text{рез}}$. Взагалі кажучи, чим більша добротність контура Q, тим вужчий і вищий максимум на кривій залежності амплітуди струму від частоти зовнішньої сили, що періодично змінюється. Щоб з'ясувати це питання з більшою точністю, розглянемо частоти, близькі до ω_0, наприклад $\omega = \omega_0 \pm \pm \Delta\omega/2$. Потім вирахуємо величину $\omega L - 1/(\omega C)$, яка входить у знаменник рівняння (9.62), з точністю до першого степеня відношення $\Delta\omega/(2\omega_0)$. Тоді отримаємо:

$$\omega L - \frac{1}{\omega C} = \omega_0 L\left(1 \pm \frac{\Delta\omega}{2\omega_0}\right) - $$

$$- \frac{1}{\omega_0 C\left(1 \pm \frac{\Delta\omega}{2\omega_0}\right)} =$$

$$= \omega_0 L\left(1 \pm \frac{\Delta\omega}{2\omega_0} - \frac{1}{1 \pm \frac{\Delta\omega}{2\omega_0}}\right) =$$

$$= \pm\omega_0 L \cdot \frac{\Delta\omega}{\omega_0}.$$

При цьому враховано, що $\omega_0 L = \frac{1}{\omega_0 C}$ і

$$\frac{1}{1 \pm \frac{\Delta\omega}{2\omega_0}} = \left(1 \pm \frac{\Delta\omega}{2\omega_0}\right)^{-1} \approx 1 \mp \frac{\Delta\omega}{2\omega_0}.$$

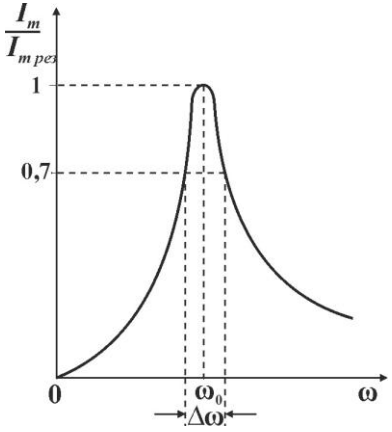


Рис. 9.19. Резонансна крива для сили струму в контурі

Розглянемо ширину $\Delta\omega$ кривої $\frac{I_m}{I_{m\,\text{рез}}}(\omega)$ на рис. 9.19, яку взято на висоті 0,7 (відношення амплітуд струмів, яке дорівнює 0,7, відповідає відношенню потужностей за частот $\omega_{1,2} = \omega_0 \pm \Delta\omega/2$ до потужності за частоти ω_0, як $0{,}7^2 \approx$ $\approx 0{,}5$, оскільки потужність пропорційна I^2). Тоді

$$\left(\frac{I_m}{I_{m\,\text{рез}}}\right)^2 = \frac{R^2}{R^2 + \left(\omega_0 L \frac{\Delta\omega}{\omega_0}\right)^2} =$$

$$= \frac{1}{1 + \left(\frac{\omega_0 L}{R}\frac{\Delta\omega}{\omega_0}\right)^2} = \frac{1}{1 + \left(Q\frac{\Delta\omega}{\omega_0}\right)^2} = \frac{1}{2}.$$

Звідси отримаємо, що $Q \cdot \frac{\Delta\omega}{\omega_0} = 1$, тоді

$$Q = \omega_0/\Delta\omega. \qquad (9.73)$$

Точки на осі ω, яким відповідають значення $\omega_0 \pm \Delta\omega/2$, називають точками «половинної енергії». За ширину резонансної кривої часто беруть відстань $\Delta\omega$ між точками половинної енергії. Радіоприймач налаштовують на певну станцію і розрізняють її від інших за допомогою резонансного контура з добротністю Q, яка дорівнює декільком сотням. Неважко виготовити мікрохвильовий резонансний контур з добротністю Q порядку 10^4 або навіть 10^5.

Розглянемо тепер різницю фаз φ між струмом (9.60) і прикладеною напругою (9.47). З (9.61) видно, що за дуже малих частот, коли $\omega L \ll \frac{1}{(\omega C)}$, $\text{tg}\varphi$ дуже великий і від'ємний, а отже, $\varphi \cong\cong -\frac{\pi}{2}$. У цьому разі струм випереджає напругу і коло має ємнісний характер. Зі зростанням частоти ω реактивний опір $\left(\omega L - \frac{1}{(\omega C)}\right)$, залишаючись від'ємним, зменшується за абсолютним значенням, і різниця фаз φ зменшується. Коли $\omega = \omega_0$, формула (9.61) дає $\text{tg}\varphi = 0$, а отже, $\varphi = 0$. Надалі зі зростанням ω реактивний опір $\left(\omega L - 1/(\omega C)\right) > 0$ і збільшується зі зростанням ω. Тому $0 < \text{tg}\varphi < +\infty$ і $0 < \varphi < +\pi/2$. Отже, якщо $\omega > \omega_0$, струм відстає від напруги, і коло набуває індуктивного характеру, причому кут φ зі зростанням частоти ω асимптотично прямує до граничного значення $+\pi/2$. Залежність різниці фаз φ від частоти коливань зображено графічно на рис. 9.20. Так само, як і I_m, φ залежить ще й від активного опору контура R. Чим менше R, тим швидше змінюється φ поблизу $\omega = \omega_0$, й у граничному випадку за $R = 0$ зміна фази набуває стрибкоподібного характеру (рис. 9.20).

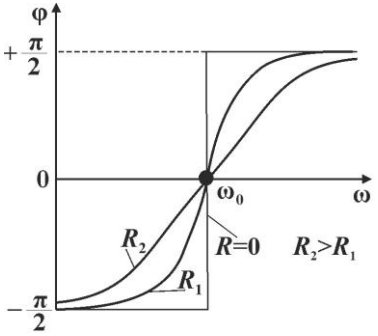


Рис. 9.20. Залежність різниці фаз φ від частоти коливань

Зазначимо, що частота ω_0, за якої настає резонанс, не дорівнює частоті вільних коливань контура $\sqrt{\omega_0^2 - \beta^2}$

(формула (9.35)). Утім, у переважній більшості практичних випадків $\beta^2 \ll \omega_0^2$, і тому з великим наближенням цією різницею можна знехтувати. Вище ми припускали, що змінюється частота прикладеної напру-ги ω, а параметри контура залиша-ються незмінними. Водночас зрозуміло, що для отримання резонансу можна робити й інакше: змінювати в контурі індуктивність або ємність (тобто змінювати ω_0), залишаючи частоту ω сталою.

9.9. Параметричний резонанс

Під час вивчення електричного резонансу (напруги і струмів) ми розглядали дію періодичної е.р.с. генератора змінної напруги (9.47) на коливальний контур. Утім виявляється, що аналогічні явища спостерігаються і в разі зовнішнього впливу інших типів, причому коливання, які виникають, як і за дії е.р.с., істотно залежать від частоти дій на контур, тому поняття резонансу можна узагальнити й поширити на більшу кількість явищ. Припустимо, що дія зовнішніх факторів змінює один із параметрів коливальної системи, і розглянемо спочатку простий механічний приклад (рис. 9.21).

Нехай є маятник, довжину якого можна змінювати (рис. 9.21), підтягуючи кінець нитки, перекинутої через блок, або навпаки, опускаючи його. Періодично змінюватимемо довжину маятника, підтягуючи нитку (зменшуючи довжину) кожного разу, коли маятник перебуватиме поблизу положення рівноваги *0*, і відпускати нитку (збільшувати довжину) за крайніх положень маятника *1* і *2*, тобто з частотою, яка дорівнює подвоєній частоті власних коливань маятника. Ми знайдемо, що маятник почне здійснювати коливання із *зростаючою* амплітудою, яка збільшуватиметься доти, поки нитка не зіскоче з блока. Це спостерігатиметься й тоді, коли частота зміни довжини не дорівнює подвоєній власній частоті, але близька до неї.

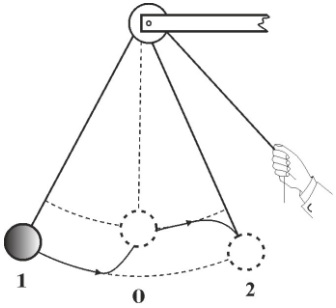


Рис. 9.21. Параметричний резонанс у механіці

Отже, у цьому досліді ми маємо наростаючі коливання, як і у випадку резонансу під дією періодичної зовнішньої сили, однак вони виникають унаслідок періодичної зміни одного з *параметрів* системи (довжини), тому

описане явище називають *параметричним резонансом.*

Причину зростання коливань можна пояснити, зважаючи на енергетичні міркування. Коли ми вкорочуємо нитку в положенні *0* (рис. 9.21), то зовнішня сила (сила руки) виконує роботу не тільки проти сили тяжіння, а й проти відцентрової сили, бо, проходячи через положення рівноваги *0*, маятник має найбільшу швидкість. Якщо нитку видовжити, роботу виконує маятник, проте ця робота виконується тільки за рахунок сили тяжіння, бо в положеннях *1* і *2* відцентрова сила дорівнює нулю (швидкість дорівнює нулю) і тому вона менша від роботи за вкороченої нитки. Отже, у коливальну систему (маятник) безперервно вводиться енергія за рахунок роботи зовнішньої сили, що й веде до зростання коливань. Такі явища параметричного резонансу спостерігаються і в електричних коливальних контурах, якщо параметри контура (ємність чи індуктивність) періодично змінюються.

Розглянемо, наприклад, коливальний контур (рис. 9.22), що має конденсатор з рухомою пластиною, яку можна періодично наближати до другої пластини або віддаляти від неї. Далі припустимо, що в контурі унаслідок яких-небудь випадкових причин виникли коливання і що в момент часу, коли заряд конденсатора дорівнює нулю, ми зближуємо пластини. Це явище не супроводжується жодною роботою, бо заряд конденсатора дорівнює нулю, а отже, й сила притягання між пластинами також дорівнює нулю. Через час, який дорівнює чверті періоду власних коливань $T/4$, заряд конденсатора буде найбільшим. Якщо у цей час розсунути пластини, то зовнішні сили виконають роботу, яка витрачається на подолання взаємного притягання пластин. При цьому ємність конденсатора зменшиться, а напруга між його обкладками збільшиться, сприяючи коливанням у контурі. Якщо потім знову через час $T/4$ зблизити пластини, то енергія контура не зміниться, бо заряд конденсатора у цей момент знову дорівнює нулю. З подальшим розсуванням пластин у контур знову буде введена певна енергія і т. д., тому істотно змінюючи ємність конденсатора з частотою, яка дорівнює подвоєній частоті власних коливань (чи близька до неї), ми дістанемо в контурі електричні коливання із зростаючою амплітудою, яка збільшуватиметься доти, поки не буде пробитий конденсатор.

Зауважимо, що початкові малі коливання і в механічних, і в електричних

системах завжди виникають під дією випадкових зовнішніх факторів, чи флуктуацій, тому за досить значної періодичної зміни параметрів системи спостерігається *самозбудження* коливань. Правильне співвідношення між фазою коливань і фазою зміни параметра здійснюється автоматично, бо підсилюються тільки ті з коливань, які мають потрібну початкову фазу.

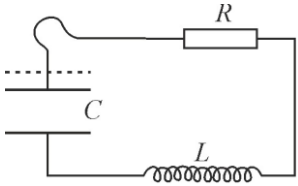


Рис. 9.22. Коливальний контур із ємністю, яка змінюється

Описаний дослід з електричним параметричним резонансом вперше провели Л. І. Мандельштам і М. Д. Папалексі в 1933 р. Сконструйована ними ємнісна «параметрична машина» мала обертовий конденсатор, який складався із двох систем пластин, що мали радіальні вирізи. Одна із систем була нерухома, а друга оберталася за допомогою електродвигуна. Така машина розвивала напруги до багатьох тисяч вольт.

Явище параметричного резонансу можна використати для технічного добування змінних струмів.

9.10. Змінний струм

Описані у підрозд. 9.8 вимушені коливання, які встановились у контурі, можна розглядати як протікання у колі, яке має ємність, індуктивність та активний опір змінного струму, обумовленого змінною напругою (9.47).

Силу струму визначають за формулами (9.60), (9.61) і (9.62). У випадку, коли $\varphi < 0$, струм фактично випереджає напругу.

Вираз

$$Z = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}, \quad (9.73)$$

який є знаменником у формулі (9.61), називають *повним електричним опором*, або *імпедансом*.

Якщо коло складається тільки з активного опору R, то рівняння закону Ома має вигляд

$$IR = U_m \cos \omega t.$$

Звідси випливає, що струм у цьому випадку змінюється у фазі з напругою, а амплітуда сили струму

$$I_m = U_m / R.$$

Порівняння цього виразу з (9.62) показує, що заміна конденсатора закороченою ділянкою кола означає перехід не до $C = 0$, а до $C = \infty$.

Будь-яке реальне коло має кінцеві R, L та C. В окремих випадках деякі з цих параметрів бувають такі, що їхнім

впливом на струм можна знехтувати. Припустимо, що R кола дорівнює нулю, а C дорівнює нескінченності. Тоді з формул (9.61) і (9.62) отримаємо, що

$$I_m = \frac{U_m}{\omega L}, \qquad (9.74)$$

а $\mathrm{tg}\varphi = \infty$ (відповідно, $\varphi = \pi/2$). Величину

$$X_L = \omega L \qquad (9.75)$$

називають *реактивним індуктивним опором* або просто *індуктивним опором*. Якщо L виразити в Генрі, а ω – у рад/с, то X_L буде виражено в Омах. Із (9.75) видно, що індуктивний опір зростає зі зростанням частоти ω. Індуктивність не чинить опору постійному струму $(\omega = 0)$. Напруга на індуктивності випереджає струм на $\pi/2$ (рис. 9.16).

Тепер припустимо, що $R = 0$ й $L = 0$. Тоді згідно з формулою (9.61) і (9.62)

$$I_m = \frac{U_m}{1/(\omega C)}, \qquad (9.76)$$

$\mathrm{tg}\varphi = -\infty$ (тобто $\varphi = -\pi/2$). Величину

$$X_c = \frac{1}{\omega C} \qquad (9.77)$$

називають *реактивним ємнісним опором.* Якщо C виразити у фарадах, а ω – у рад/с, то X_c буде виражено в омах. Із (9.77) видно, що ємнісний опір зменшується із зростанням частоти ω. Для постійного струму $(\omega = 0\,, X_c = \infty)$ постійний струм через конденсатор протікати не може. Оскільки $\varphi = -\pi/2$, то струм, який протікає через конденсатор, випереджає напругу на $\pi/2$. Відповідно, напруга на конденсаторі відстає від струму на $\pi/2$ (рис. 9.16).

Припустимо, що R = 0. У цьому випадку формула (9.62) набуває вигляду

$$I_m = \frac{U_m}{\left|\omega L - \frac{1}{\omega C}\right|}. \qquad (9.78)$$

Величину

$$X = \omega L - \frac{1}{\omega C} = X_L - X_c \qquad (9.79)$$

називають *реактивним опором*, або *реактансом*.

Формули (9.60) і (9.61) можна подати у вигляді

$$\mathrm{tg}\varphi = \frac{X}{R};\ I_m = \frac{U_m}{\sqrt{R^2 + X^2}} = \frac{U_m}{Z}. \qquad (9.80)$$

Знайдемо потужність, яка виділяється в колі змінного струму. Миттєве значення потужності дорівнює добутку миттєвих значень напруги і сили струму:

$$P(t) = U(t) \cdot I(t) = \\ = U_m I_m \cos\omega t \cdot \cos(\omega t - \varphi). \qquad (9.81)$$

Скориставшись формулою

$$\cos\alpha \cdot \cos\beta = \frac{1}{2}\cos(\alpha - \beta) + \\ + \frac{1}{2}\cos(\alpha + \beta),$$

вираз (9.81) можна подати у вигляді

$$P(t) = \frac{1}{2} U_m I_m \cos\varphi + \\ + \frac{1}{2} U_m I_m \cos(2\omega t - \varphi). \qquad (9.82)$$

Практичний інтерес становить середнє за часом значення $P(t)$, яке

позначимо просто P. Оскільки середнє значення $\cos(2\omega t - \varphi)$ дорівнює нулю, то

$$P = \frac{U_m I_m}{2}\cos\varphi. \qquad (9.83)$$

Із (9.82) видно, що миттєва потужність коливається поблизу середнього значення з частотою, яка у два рази перевищує частоту струму (рис. 9.23). Оскільки $\cos\varphi = (1 + \mathrm{tg}\varphi)^{-1/2}$, то, використавши формулу (9.61), отримаємо

$$\cos\varphi = \frac{R}{\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}} = \frac{R}{Z}. \qquad (9.84)$$

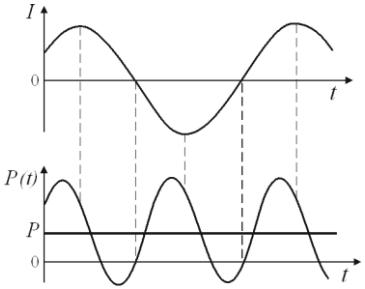


Рис. 9.23. Миттєва потужність коливається поблизу середнього значення з частотою, яка у два рази перевищує частоту струму

Підставивши це значення $\cos\varphi$ у формулу (9.83) і врахувавши (9.80), отримаємо

$$P = \frac{RI_m^2}{2}. \qquad (9.85)$$

Таку ж потужність розвиває постійний струм, сила якого

$$I = I_m/\sqrt{2}. \qquad (9.86)$$

Величину (9.86) називають *діючим* (або *ефективним*) *значенням сили струму*. Аналогічно величину

$$U = U_m/\sqrt{2} \qquad (9.87)$$

називають *діючим значенням напруги*.

Формула (9.83) через діючі значення сили струму і напруги набуде вигляду

$$P = IU\cos\varphi. \qquad (9.88)$$

Множник $\cos\varphi$, який входить у формулу (9.88), називають *коефіцієнтом потужності*. У техніці $\cos\varphi$ намагаються зробити якомога більшим. Якщо $\cos\varphi$ малий, для виділення в колі необхідної потужності потрібно пропускати струм великої сили, що є причиною зростання втрат у підвідних проводах.

9.11. Автоколивання

Автоколивання – це незгасаючі коливання, які підтримуються зовнішніми джерелами енергії в нелінійних коливальних системах, де є розсіювання енергії, вид і властивості яких визначаються цими системами.

Автоколивання принципово різняться від інших коливальних процесів у коливальних системах, де є розсіювання енергії тим, що для їхньої підтримки не потрібно періодичної дії ззовні. Коливання струни скрипки при рівномірному русі смичка, коливання

струму в радіотехнічному генераторі, повітря в органній трубі, маятника в годиннику, спів соловейка – приклади автоколивань. У найпростіших автоколивальних системах можна виділити коливальну систему із згасанням (коливальний контур), підсилювач коливань, нелінійний обмежувач і ланку *оберненого зв'язку.*

Одну з найпростіших класичних схем лампового генератора незгасаючих коливань подано на рис. 9.24. Коливальний контур, що має ємність C та індуктивність L, увімкнений в коло сітки електронної лампи. В колі анода, крім батареї живлення, є ще котушка L', розміщена безпосередньо близько до котушки L так, що між обома котушками існує індуктивний зв'зок.

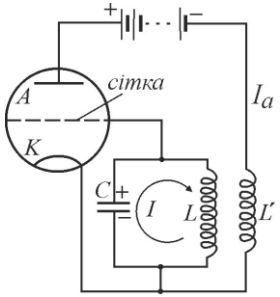


Рис. 9.24. Класичний генератор з коливальним контуром у колі сітки лампи

Розглянемо принцип дії генератора. Коли в коливальному контурі виникають коливання (при вмиканні батареї або під впливом яких-небудь випадкових причин), то між обкладками конденсатора виникає змінна напруга. Така сама напруга виникає між сіткою й катодом K, бо вони приєднані до обкладок конденсатора. Унаслідок цього в колі анода виникає змінний струм I_a, але котушки L' та L індуктивно пов'язані між собою, й тому змінний струм I_a спричиняє в котушці L змінну е.р.с. взаємної індукції:

$$\mathcal{E} = -L_{12}\frac{dI_a}{dt}, \qquad (9.89)$$

де L_{12} – коефіцієнт взаємної індукції. *Ця е.р.с. залежно від спеціального взаємного напрямку намотування витків котушок L' та L може або перешкоджати коливанням струму в контурі, або сприяти їм.*

Очевидно, що перемикаючи за потреби кінці котушки L, можна завжди домогтися, щоб е.р.с. $\mathcal{E}$, яка виникає, мала такий самий знак, як і струм I в контурі. Тоді робота цієї е.р.с. буде додатною, й коливальний контур отримає енергію за рахунок джерела струму так, що згасання коливань в контурі зменшиться. Якщо зв'язок між котушками досить сильний, енергія, яку отримує контур, може бути більша за ту, яка витрачається в контурі. У цьому разі амплітуда коливань зростатиме доти, поки отримувана енергія не дорівнюватиме втратам, після чого встановляться незгасаючі коливання.

Розглянемо тепер, яку кількісну умову мають задовольняти параметри лампового генератора (опір контура *R,* коефіцієнт взаємоіндукції L_{12} і т. д.),щоб автоколивання були можливими.

Знову розглянемо схему на рис. 9.24. Коливальний контур цього генератора можна подати еквівалентною схемою з рис. 9.25, в якій деякий генератор здійснює позитивний зворотний зв'язок (подібно до зовнішньої сили, що періодично змінюється (9.47) у рівнянні (9.48)) з величиною змінної е.р.с. (9.89).

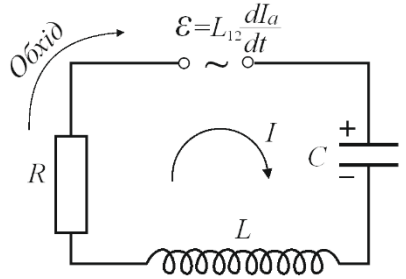


Рис. 9.25. Еквівалентна схема генератора, зображеного на рис. 9.24

Застосовуючи до цього контура друге правило Кірхгофа, отримаємо

$$IR + U_C = -L\frac{dI}{dt} + L_{12}\frac{dI_a}{dt}.$$

Враховуючи, що $U_C = q/C$ й $I = \frac{dq}{dt} > 0$ (зарядка конденсатора), це рівняння перепишем так:

$$\frac{d^2q}{dt^2} + \frac{R}{L}\frac{dq}{dt} + \frac{q}{LC} - \frac{L_{12}}{L}\cdot\frac{dI_a}{dt} = 0. \quad (9.90)$$

У цих рівняннях анодний струм I_a є функцією сіткової напруги u_c (що виражається сітковою характеристикою лампи $I_a(u_c)$), яка, у свою чергу, залежить від заряду конденсатора. Утім, сіткова характеристика лампи нелінійна, тому й написане рівняння є нелінійним рівнянням, а отже, автоколивання є *нелінійними коливаннями.* Водночас, щоб розв'язати поставлене питання, задачу можна спростити. Припустимо, що ми маємо малі коливання і що робочу точку вибрано в середній частині сіткової характеристики. Тоді малий відрізок сіткової характеристики можна вважати відрізком прямої лінії й рівняння коливань буде *лінійним*. Крім того, для простоти вважатимемо, що проникність сітки лампи дуже мала, так, що I_a фактично залежить тільки від потенціалу сітки u_c (але не від потенціалу анода U_a). Тоді потенціал сітки визначають напругою на конденсаторі U_C й

$$I_a = a + SU_C = a + \frac{S}{C}q,$$

де $S-$ крутість сіткової характеристики лампи; $a-$ постійна. Тому

$$\frac{dI_a}{dt} = \frac{S}{C}\frac{dq}{dt}.$$

Підставивши цей вираз у рівняння коливань (9.90), отримаємо:

$$\ddot{q} + \left(\frac{R}{L} - \frac{SL_{12}}{LC}\right)\dot{q} + \frac{1}{LC}q = 0. \quad (9.91)$$

Це рівняння того самого вигляду, що й (9.31), і його розв'язком є (9.36):

$$q = q_0 e^{-\beta' t}\cos(\omega t + \varphi), \quad (9.92)$$

але в цьому випадку коефіцієнт згасання

$$\beta' = \frac{1}{2}\left(\frac{R}{L} - \frac{SL_{12}}{LC}\right). \quad (9.93)$$

Він менший за коефіцієнт $\beta = R/(2L)$, коли немає зворотного зв'язку, і тому можна сказати, що дія зворотного зв'язку еквівалентна введенню в контур *від'ємного опору.*

З (9.93) видно, що за певних умов коефіцієнт згасання β' може дорівнювати нулю або навіть бути від'ємним, що відповідає виникненню автоколивань. Це буде тоді, коли

$$SL_{12}/C > R. \quad (9.94)$$

Записана умова є *умовою самозбудження* лампового генератора.

Якщо виконується умова (9.94), то $\beta' < 0$ і розв'язок (9.92) відображає *наростаючі* коливання, амплітуда яких збільшується з часом до нескінченності. Причина цього дивного результату полягає в тому, що нелінійне рівняння коливань (9.90) ми замінили наближеним лінеаризованим рівнянням (9.91), яке справджується тільки для малих коливань (початкової стадії процесу), але не відповідає всім властивостям генератора. Насправді ж після самозбудження амплітуда наростаючих коливань прагнутиме до деякої *скінченної* границі, що визначається властивостями генератора і не залежить від початкових умов.

Контрольні запитання

1. Дайте визначення коливального руху.
2. Які коливання називають гармонічними? Запишіть рівняння гармонічних коливань.
3. Опишіть характеристики гармонічних коливань: амплітуди, фази, циклічної частоти, лінійної частоти, початкова фази.
4. Як проявляється закон збереження енергії при гармонічних коливаннях?
5. Як геометрично можна зобразити гармонічне коливання?
6. За якої умови утворюються коливання, які називають «биттям»?
7. Що таке фігури Ліссажу?
8. Яка відповідність між гармонічними механічними коливаннями і коливан-нями, які виникають в ідеалізованому електричному контурі?
9. Запишіть рівняння згасаючих коливань.
10. Дайте визначення коефіцієнта і логарифмічного декремента згасання.
11. Розкрийте фізичний зміст добротності коливального контура.
12. Як відрізняється циклічна частота згасаючих коливань від «власної» циклічної частоти коливальної системи?
13. Дайте визначення аперіодичного процесу. Запишіть формулу для

визначення критичного опору коливального контура.

14. Запишіть рівняння вимушених коливань.

15. З якою частотою відбуваються вимушені коливання?

16. Запишіть і поясніть формулу для амплітуди вимушеного коливання.

17. Що таке резонанс?

18. Чим пояснюється зсув за фазою між зовнішньою силою, що періодично змінюється (напругою) і зміщенням (зарядом) при вимушених коливаннях?

19. Який зв'язок між вимушеними коливаннями і змінним струмом?

20. Які є способи використання параметричних коливань і автоколивань у техніці звязку?

Розділ 10. Пружні хвилі

Поняття хвилі

Хвилі – це зміни стану середовища (збурення), які поширюються в цьому середовищі й несуть із собою енергію. Найбільш важливі хвилі, які трапляються часто, це такі: пружні хвилі, хвилі на поверхні рідини та електромагнітні хвилі. Окремим випадком пружних хвиль є звук і сейсмічні хвилі, а з електромагнітних – радіохвилі, світло, рентгенівські промені та ін.

Основна властивість усіх хвиль незалежно від їхньої природи полягає в тому, що у хвилях здійснюється перенесення енергії без перенесення речовини (останнє може відбуватися тільки як побічне явище). Хвильові процеси трапляються майже у всіх областях фізичних явищ, тому їх вивчення має велике значення.

10.1. Утворення хвиль у пружному середовищі. Поздовжні й поперечні хвилі

Розглянемо пружне середовище, між частинками якого існують сили взаємодії, що перешкоджають тому або іншому виду його деформації. Тіло, яке коливається у пружному середовищі, періодично діє на прилеглі до нього частинки середовища, виводячи їх з положення рівноваги і змушуючи здійснювати вимушені коливання. При цьому середовище поблизу тіла деформується і в ньому виникають пружні сили. Ці сили діють як на прилеглі до тіла частинки, намагаючись повернути їх у положення рівноваги, так і на найбільш віддалені від тіла частинки, виводячи їх з положення рівноваги. Віддалені від тіла області середовища поступово втягуються в коливальний рух. *Процес поширення коливань в суцільному середовищі, яке неперервно розподілене у просторі й має пружні властивості,*

називають механічним хвильовим процесом, або механічною хвилею.

Під час поширенні хвилі частинки середовища не рухаються разом із хвилею, а коливаються поблизу своїх положень рівноваги.

Пружними (або механічними) хвилями називають механічні збурення, що поширюються у пружному середовищі. Пружні хвилі бувають поперечними й поздовжніми.

У поперечних хвилях частинки середовища коливаються у площинах, які перпендикулярні до напрямку поширення хвилі (рис. 10.1). Поперечні хвилі можуть поширюватись у середовищі, в якому виникають пружні сили під час деформації зсуву, тобто лише у твердих тілах.

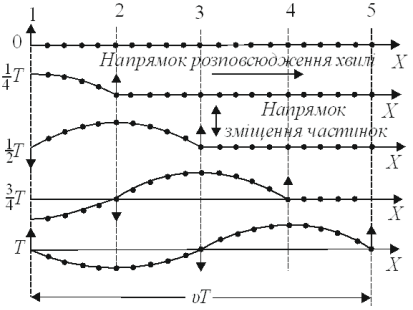


Рис. 10.1. Поперечна хвиля

У поздовжніх хвилях частинки коливаються у напрямку поширення хвилі (рис. 10.2). Ці хвилі можуть поширюватись у середовищах, в яких виникають пружні сили під час деформації стиску і розтягу, тобто у твердих, рідких і газоподібних тілах. У твердих тілах з обмеженими розмірами (наприклад, у стрижнях, пластинках) картина поширення хвиль більш складна: тут виникають ще й інші типи хвиль, які є комбінацією перших двох основних типів.



Рис. 10.2. Поздовжня хвиля

Загальні характеристики і властивості хвиль

Хвилі можуть мати різну форму. Поодинокою хвилею, або імпульсом, називають порівняно коротке збурення, яке не має регулярного характеру (рис. 10.3, *а*). Обмежений ряд збурень, які повторюються, називають цугом хвиль. Найчастіше поняття цуга відносять до відрізка синусоїди (рис. 10.3, *б*). Особливе значення в теорії хвиль має уявлення про гармонічні хвилі, тобто нескінченні синусоїдальні хвилі, в яких усі зміни стану середовища відбуваються за законом синуса або косинуса (рис. 10.3, *в*), оскільки такі хвилі могли б

поширюватись в однорідному середовищі без спотворення форми (якщо амплітуда їх невелика).

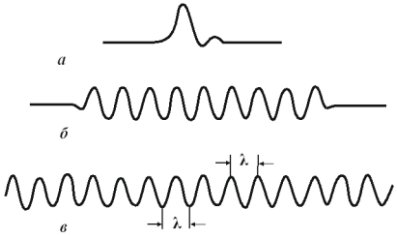


Рис. 10.3. Види хвиль: а – поодинока хвиля; б – цуг хвиль; в – нескінченна синусоїдальна хвиля

Пружну хвилю називають *гармонічною*, якщо відповідні їй коливання частинок середовища є гармонічними.

Нехай поперечна гармонічна хвиля поширюється вздовж осі *OX*. На рис. 10.4 показано гармонічну хвилю (поперечну або поздовжню), яка поширюється вздовж осі *OX* зі швидкістю v, тобто наведена залежність зміщень *S* усіх частинок середовища, що беруть участь у хвильовому процесі, від відстані *x* цих частинок від джерела коливань *O* для якогось фіксованого моменту часу. Основними характеристиками гармонічної хвилі є *довжина хвилі* λ – відстань між двома максимумами або мінімумами (рис. 10.3, *в*, та рис. 10.4) і *період T* – час, за який здійснюється один повний цикл коливань. Довжина хвилі λ пов'язана з періодом коливань *T* співвідношенням

$$\lambda = vT, \qquad (10.1)$$

де v – швидкість поширення хвилі (рис. 10.1, 10.2 та 10.4). Співвідношення (10.1) справедливе для гармонічної хвилі будь-якої природи. Довжину хвилі можна визначити ще й як відстань, на яку поширюється хвиля за час, що дорівнює періоду коливань частинок середовища, або як відстань між найближчими точками середовища, які коливаються з різницею фаз, що дорівнює 2π (рис. 10.3, *в*). Замість періоду *T* нерідко користуються *частотою* ν, яка дорівнює кількості періодів (коливань) за одиницю часу:

$$\nu = 1/T, \qquad (10.2)$$

при цьому

$$\lambda\nu = v. \qquad (10.3)$$

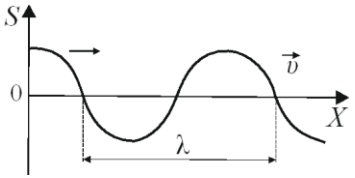


Рис. 10.4. Гармонічна хвиля, яка поширюється вздовж осі OX зі швидкістю v

В теорії хвиль користуються також поняттям *хвильового вектора* $\vec{k}$. За абсолютною величиною (*хвильове число*)

$$k = \frac{2\pi}{\lambda} = \frac{2\pi\nu}{v} = \frac{\omega}{v}, \qquad (10.4)$$

тобто він дорівнює кількості довжин

хвилі, що вкладаються на відрізку 2π й орієнтовані у напрямку поширення хвилі.

Нехай точка, від якої йдуть коливання, коливається в суцільному середовищі. Коливання поширюються від центра в усі боки.

Поверхню, до якої доходять коливання в деякий момент часу, називають *фронтом хвилі* (це поверхня, яка відокремлює частину простору, уже залучену у хвильовий процес, від області, в якій коливання ще не виникли). *Принцип Гюйгенса* встановлює спосіб побудови фронту хвилі в момент часу $t + \Delta t$, якщо відоме положення фронту в момент часу t. Згідно з принципом Гюйгенса кожна точка, до якої доходить хвильовий рух, стає центром вторинних хвиль; огинальна цих хвиль дає положення фронту хвилі в наступний момент часу (див. рис. 10.4′).

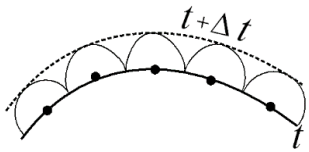


Рис. 10.4′. Поширення хвилі за принципом Гюйгенса

Поверхню, всі частинки якої коливаються з однаковими фазами, називають *хвильовою поверхнею*.

Хвильову поверхню можна провести через довільну точку простору, який охоплений хвильовим процесом. Отже, хвильових поверхонь існує нескінченна множина, а хвильовий фронт у кожний момент часу лише один. Хвильові поверхні залишаються нерухомими, а хвильовий фронт весь час переміщується.

Хвильові поверхні можуть бути довільної форми.

Хвилю називають *плоскою*, якщо її хвильові поверхні мають вигляд площин, які паралельні до площини, що проходить через джерело хвиль.

Хвилю називають *сферичною*, якщо її хвильові поверхні мають вигляд концентричних сфер. Центр цих сфер називають *центром хвилі*.

Напрямки, в яких поширю-ються коливання, називають *променями*. В ізотропному середовищі промені перпендикулярні до фронту хвилі.

Поширення у пружному середовищі механічних збурень, збуджених джерелом хвиль, пов’язане з перенесенням хвилями енергії, тому такі хвилі називають *біжучими хвилями*.

10.2. Рівняння плоскої і сферичної хвиль

Рівнянням хвилі називають вираз, який описує зміщення частинки, яка коливається, як функцію її рівноважних координат x, y, z й часу t.

Розглянемо плоску хвилю, яка поширюється вздовж осі OX і збуджується у площині $x = 0$ (рис. 10.5). Нехай коливання у площині $x = 0$ описують рівнянням

$$S(0,t) = A\cos(\omega t + \varphi_0).$$

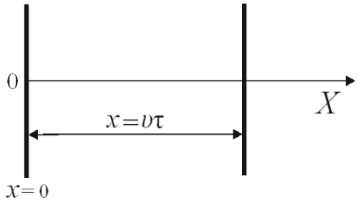


Рис. 10.5. Плоска хвиля

Знайдемо вигляд рівняння коливань частинок у площині, що відповідає довільному значенню x. Щоб пройти шлях від площини $x = 0$ до площини на відстані x, хвилі потрібний час $\tau = x/v$, де v – швидкість поширення хвилі. Отже, коливання частинок, що лежать у площині x, будуть запізнюватись на час τ від коливань частинок у площині x=0, тобто матимуть вигляд

$$S(x,t) = A\cos[\omega(t-\tau)+\varphi_0] =$$
$$= A\cos\left(\omega t - \frac{\omega x}{v} + \varphi_0\right) =$$
$$= A\cos\left(\omega t - \frac{2\pi}{vT}x + \varphi_0\right) =$$
$$= A\cos\left(\omega t - \frac{2\pi}{\lambda}x + \varphi_0\right).$$

Тоді рівняння біжучої плоскої хвилі, що поширюється вздовж осі OX, із врахуванням (10.4) має такий вигляд:

$$S(x,t) = A\cos(\omega t - kx + \varphi_0), \quad (10.5)$$

де A – амплітуда коливань, яку називають амплітудою хвилі; $\omega = \frac{2\pi}{T}$ – циклічна частота хвилі; k – хвильове число хвилі; φ_0 – початкова фаза коливань у площині за $x = 0$.

Величина $\omega t - kx + \varphi_0$ визначає фазу коливань у довільній площині з координатою x, її називають фазою плоскої хвилі. Зафіксуємо певне значення фази:

$$\omega t - kx + \varphi_0 = \text{const}. \quad (10.6)$$

Цей вираз визначає зв'язок між часом t і тим місцем x, в якому фаза має зафіксоване значення. Величина $\frac{dx}{dt}$ дає швидкість, з якою переміщується розглядуване значення фази. Продиференціюємо вираз (10.6):

$$\omega dt - k dx = 0.$$

Звідси

$$\frac{dx}{dt} = \frac{\omega}{k} = \frac{\omega\lambda}{2\pi} = \frac{\lambda}{T} = v. \quad (10.7)$$

Отже, швидкість v поширення хвилі є нічим іншим, як швидкістю переміщення фази хвилі, її називають ***фазовою швидкістю***.

Якщо плоска хвиля поширюється в довільному напрямку, то рівняння хвилі запишемо так:

$$S(\vec{r},t) = A\cos\left(\omega t - \vec{k}\vec{r} + \varphi_0\right), \quad (10.8)$$

де $\vec{k} = k\vec{n}$ – вектор, який дорівнює за

модулем хвильовому числу і має напрямок нормалі $\vec{n}$ до хвильової поверхні.

У випадку сферичної хвилі рівняння хвилі буде

$$S(r,t) = \frac{A}{r}A\cos(\omega t - kr + \varphi_0), \quad (10.9)$$

де r – відстань від центра хвилі до розглядуваної точки середовища. У випадку сферичної хвилі навіть у середовищі, яке не поглинає енергію, амплітуда коливань не залишається постійною, а зменшується з відстанню за законом A/r. Це співвідношення справедливе лише для r, що значно більші за розміри джерела.

Згідно з (10.7) $dx/dt > 0$. Це означає, що рівняння (10.5) описує хвилю, яка поширюється у бік зростання x. Хвилю, яка поширюється у протилежному напрямку, описують рівнянням

$$S(x,t) = A\cos(\omega t + kx + \varphi_0). \quad (10.10)$$

Справді, взявши фазу хвилі за константу і продиференціювавши за часом, отримаємо, що $dx/dt = -v$. Це означає, що хвиля, яку описують рівнянням (10.10), поширюється у бік зменшення x.

10.3. Хвильове рівняння

Рівняння довільної хвилі є розв'язком ***рівняння***, яке називають ***хвильовим***.

Щоб встановити вид цього рівняння, використаємо рівняння плоскої хвилі (10.8), що поширюється у довільному напрямку:

$$S(\vec{r},t) = A\cos\left(\omega t - \vec{k}\vec{r} + \varphi_0\right) = A\cos\left(\omega t - k_x x - k_y y - k_z z + \varphi_0\right).$$

Продиференціюємо цю функцію двічі за кожною змінною:

$$\frac{\partial^2 S}{\partial t^2} = -\omega^2 A\cos\left(\omega t - \vec{k}\vec{r} + \varphi_0\right) = -\omega^2 S;$$

$$\frac{\partial^2 S}{\partial x^2} = -k_x^2 A\cos\left(\omega t - \vec{k}\vec{r} + \varphi_0\right) = -k_x^2 S;$$

$$\frac{\partial^2 S}{\partial y^2} = -k_y^2 A\cos\left(\omega t - \vec{k}\vec{r} + \varphi_0\right) = -k_y^2 S;$$

$$\frac{\partial^2 S}{\partial z^2} = -k_z^2 A\cos\left(\omega t - \vec{k}\vec{r} + \varphi_0\right) = -k_z^2 S.$$

Додамо ліві та праві частини похідних за такими координатами:

$$\frac{\partial^2 S}{\partial x^2} + \frac{\partial^2 S}{\partial y^2} + \frac{\partial^2 S}{\partial z^2} = -\left(k_x^2 + k_y^2 + k_z^2\right)S = -k^2 S.$$

Зіставимо цю суму з похідною за часом і врахуємо згідно з (10.7), що $\omega^2 = v^2 k^2$. Тоді поширення хвиль в однорідному ізотропному середовищі описують ***хвильовим рівнянням*** – диференціальним рівнянням в частинних похідних:

$$\frac{\partial^2 S}{\partial x^2} + \frac{\partial^2 S}{\partial y^2} + \frac{\partial^2 S}{\partial z^2} = \frac{1}{v^2}\frac{\partial^2 S}{\partial t^2}, \quad (10.11)$$

або

$$\Delta S = \frac{1}{v^2}\frac{\partial^2 S}{\partial t^2}, \quad (10.12)$$

де v – фазова швидкість, а

$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ – оператор Лапласа.

Легко переконатись у тому, що хвильове рівняння (10.11) задовольняє не тільки функція (10.8), але і будь-яка функція вигляду

$$S(x, y, z, t) = f(\omega t - k_x x - k_y y - k_z z + \varphi_0) \quad (10.13)$$

Будь-яка функція, яка задовольняє рівняння вигляду (10.11), описує деяку хвилю, причому корінь квадратний з величини, оберненої коефіцієнту за $\frac{\partial^2 S}{\partial t^2}$, дає фазову швидкість цієї хвилі.

Надалі будемо розглядати плоскі хвилі, які поширюються вздовж осі *X*, і рівняння (10.11) буде мати вигляд

$$\frac{\partial^2 S}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 S}{\partial t^2}. \quad (10.13)$$

Остроградський Михайло Васильович (1801–1862) розв'язав у 1826 р. задачу поширення хвиль на поверхні води.

10.4. Енергія пружної хвилі

Нехай у деякому середовищі в додатному напрямку осі *OX* поширюється плоска хвиля (10.5).

Визначимо зміну енергії малого об'єму dV пружного середовища, пов'язану з поширенням у середовищі плоскої хвилі. Оскільки об'єм dV дуже малий, то можна вважати, що всі частинки середовища, які містяться в цьому об'ємі, коливаються з однаковою фазою так, що їх швидкості однакові й $v = \frac{dS}{dt} = -A\omega\sin(\omega t - kx + +\varphi_0)$.

Тому кінетичну енергію об'єму dV, яка пов'язана з коливальним рухом, знайдемо за формулою

$$dW_k = \frac{v^2}{2}dm = \frac{v^2}{2}\rho dV =$$
$$= \frac{1}{2}\rho A^2\omega^2 dV\sin^2(\omega t - kx + \varphi_0),$$

де ρ – густина середовища.

Визначаючи роботу деформації об'єму dV середовища під час хвильового руху, можна показати, що потенціальна енергія dW_p об'єму dV середовища дорівнює його кінетичній енергії:

$$dW_{\text{п}} = dW_{\text{к}}.$$

Повна механічна енергія коливального руху об'єму dV

$$dW = dW_{\text{к}} + dW_{\text{п}} =$$
$$= \rho A^2\omega^2 dV\sin^2(\omega t - kx + \varphi_0).$$

Об'ємна густина енергії хвиль у пружному середовищі

$$w = \frac{dW}{dV} = \rho A^2\omega^2\sin^2(\omega t - kx + \varphi_0). \quad (10.14)$$

Густина енергії в кожний момент часу в різних точках простору різна. В одній і тій самій точці густина енергії змінюється з часом за законом квадрата синуса. Середнє значення квадрата синуса дорівнює $1/2$. Відповідно, середнє за часом значення об'ємної густини енергії в кожній точці середовища

$$\langle w\rangle = (1/2)\rho A^2\omega^2. \quad (10.15)$$

Поширення хвиль у пружному середовищі нерозривно пов'язане з процесом передавання енергії від одних ділянок середовища до інших – саме тому при хвильовому русі об'ємна густина енергії w коливань у кожній точці середовища змінюється з часом.

Об'ємну густину енергії (10.14) можна записати так:

$$w = (1/2)\,\rho A^2\omega^2 \times$$
$$\times\left(1-\cos 2\left(\omega t - kx + \varphi_0\right)\right). \quad (10.16)$$

Швидкість u поширення енергії хвилі дорівнює швидкості переміщення у просторі поверхні, яка відповідає максимальному значенню об'ємної густини енергії w. Рівняння поверхні $w = w_{max}$ має вигляд

$$2(\omega t - kx\ + \varphi_0) = \pi.$$

Продиференціюємо цей вираз і прирівняємо до нуля:

$$\omega dt - kdx = 0.$$

Звідси швидкість переміщення поверхні, що відповідає максимальному значенню об'ємної густини енергії w_{max}, буде

$$u = \frac{dx}{dt} = \frac{\omega}{k} = v.$$

Отже, швидкість поширення енергії хвилі збігається з фазовою швидкістю хвилі.

Для характеристики процесу перенесення енергії хвилями введемо поняття про потік енергії.

Потоком енергії Φ крізь яку-небудь поверхню площею S називають фізичну величину, яка чисельно дорівнює кількості енергії dW, яка переноситься через цю поверхню за одиницю часу:

$$\Phi = \frac{dW}{dt}. \quad (10.17)$$

Потік енергії – скалярна величина, розмірність якої дорівнює розмірності енергії, поділеній на розмірність часу, тобто збігається з розмірністю потужності, тому Φ вимірюють у ватах.

Знайдемо потік енергії хвилі, що рухається з фазовою швидкістю $\vec{v}$, через площину dS (рис. 10.6). За час dt хвиля перенесе енергію, що міститься всередині косого циліндра, об'єм якого

$$dV = vdt\ dS\cos\alpha = vdt\ dS_n.$$

Тоді $dW = wdV = wv\ dt\ dS_n$ й потік енергії $d\Phi = wvdS_n = w\vec{v}\cdot d\vec{S}$, де w – об'ємна густина енергії хвилі; $d\vec{S} = \vec{n}dS$ – вектор площини dS; $\vec{n}$ – одинична нормаль до dS.

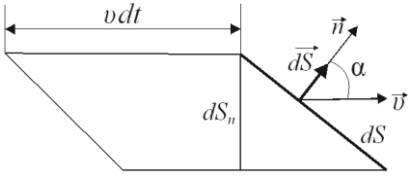


Рис. 10.6. Потік енергії хвилі, що рухається з фазовою швидкістю $\vec{v}$, через площину dS

Для різних точок простору вводять векторну величину $\vec{j}$, яку називають *густиною потоку енергії*.

Густина потоку енергії – векторна величина, яка напрямлена у бік поширення хвилі й чисельно дорівнює потоку енергії $d\Phi$ через одиницю площі dS поверхні, яка розміщена перпендикулярно до напрямку поширення хвилі:

$$j = \frac{d\Phi}{dS} = \frac{wvdS_n}{dS_n} = wv.$$

Оскільки швидкість $\vec{v}$ – це вектор, модуль якого дорівнює фазовій швидкості хвилі, а напрямок збігається з напрямком поширення хвилі (і перенесення енергії), то

$$\vec{j} = w\vec{v}. \qquad (10.18)$$

Вектор густини потоку енергії хвилі, який називають вектором Умова, дорівнює добутку вектора швидкості поширення енергії хвилі на величину її об'ємної густини.

Вектор $\vec{j}$ в різних точках простору має неоднакові значення, а у цій точці простору змінюється з часом за законом квадрата синуса. Середнє значення вектора Умова

$$\langle\vec{j}\rangle = \langle w\rangle\vec{v} = (1/2)\,\rho A^2\omega^2\vec{v}. \qquad (10.19)$$

Знаючи $\vec{j}$ у всіх точках довільної поверхні S, можна обчислити потік енергії через цю поверхню:

$$\Phi = \int_S \vec{j}d\vec{S}.$$

Скалярну величину I, яка дорівнює модулю середнього значення вектора Умова, називають **інтенсивністю хвилі**:

$$I = |\langle\vec{j}\rangle|.$$

Інтенсивність хвилі чисельно дорівнює енергії, яка переноситься хвилею за одиницю часу через одиницю площі поверхні, яка перпендикулярна до напрямку поширення хвилі:

$$I = \langle w\rangle v = (1/2)\rho\omega^2 A^2 v. \qquad (10.20)$$

Інтенсивність синусоїдальної хвилі пропорційна квадрату її амплітуди.

10.5. Групова швидкість поширення хвилі

Швидкість поширення гармонічної хвилі, яку описують рівняннями (10.5) і (10.9), *дорівнює фазовій швидкості* v (10.7).

Постає запитання, чи для всіх видів хвиль справедлива така залежність? Усі реальні хвилі не є гармонічними, проте можна довести, що будь-яку несинусоїдальну хвилю можна розкласти на еквівалентну їй систему синусоїдальних хвиль. Таке розкладання називають гармонійним, або спектральним аналізом несинусоїдальної хвилі. Сукупність значень амплітуд, початкових фаз і частот еквівалентної системи синусоїдальних хвиль називають *спектром*, відповідно, *амплітуди*,

початкових фаз і *частот* несинусоїдальної хвилі, що розглядається. Спектр може бути як дискретним, або лінійчастим, тобто множиною окремих різних значень, так і неперервним (суцільним).

Встановлено, що всі синусоїдальні (гармонічні) хвилі поширюються в середовищі незалежно одна від одної, а результуюче зміщення будь-якої частинки середовища дорівнює векторній сумі її зміщень, зумовлених кожною із хвиль. Цей результат справедливий для хвиль будь-якої природи, *його називають принципом суперпозиції хвиль* (див. підрозд. 10.6 та 10.7).

Суперпозицію гармонічних хвиль, які майже не відрізняються одна від одної за частотою (чи довжиною хвиль), називають групою хвиль, або хвильовим пакетом.

Якщо швидкість поширення синусоїдальних хвиль у середовищі не залежить від їх частоти (довжини хвилі), то швидкість поширення пакета дорівнюватиме фазовій швидкості. Утім, є й такі гармонічні хвилі, фазова швидкість яких залежить від їх частоти (довжини хвилі). Залежність фазової швидкості від частоти (довжини хвилі) називають ***дисперсією хвиль***, а середовище, де спостерігається це явище, – ***дисперсійним***.

Якщо хвильовий пакет поширюється в дисперсійному середовищі, то швидкість окремих його синусоїдальних хвиль різна. Унаслідок цього неперервно змінюватимуться різниці фаз між хвилями, імпульс складного коливання дещо змінюватиме свою форму, а амплітудне значення хвилі зміщуватиметься відносно складових хвиль. При цьому швидкість поширення хвильового пакета не збігається із швидкістю жодної складової хвилі.

Для прикладу розглянемо простий хвильовий пакет, утворений двома плоскими синусоїдальними хвилями, що поширюються вздовж осі *X*. Не порушуючи загальності результату, вважатимемо, що амплітуди цих хвиль однакові, початкові фази $\varphi_{01} = \varphi_{02} = 0$, а циклічні частоти і хвильові числа їх різні, але майже не різняться між собою:

$$\begin{cases} S_1(x,t) = A_0 \cos(\omega_1 t - k_1 x) \\ S_2(x,t) = A_0 \cos(\omega_2 t - k_2 x) \end{cases}. \quad (10.21)$$

Згідно з принципом суперпозиції хвиль рівняння результуючої хвилі буде

$$S(x,t) = S_1(x,t) + S_2(x,t) = \\ = 2A_0 \cos\left(\frac{\Delta\omega}{2} t - \frac{\Delta k}{2} x\right) \times \\ \times \cos(\omega_{\text{сер}}\, t - k_{\text{сер}} x), \quad (10.22)$$

де

$$\Delta\omega=\omega_1-\omega_2,\ \omega_{\text{сер}}=(\omega_1+\omega_2)/2,$$
$$\Delta k=k_1-k_2,\ \ k_{\text{сер}}=(k_1+k_2)/2. \quad (10.23)$$

При отриманні (10.22) враховано, що

$$\cos\alpha+\cos\beta=$$
$$=2\cos\frac{1}{2}(\alpha+\beta)\cdot\cos\frac{1}{2}(\alpha-\beta).$$

З виразу (10.22) випливає, що результуюча хвиля є плоскою, її циклічна частота і хвильове число дорівнюють півсумі циклічних частот і хвильових чисел синусоїдальних хвиль, що утворюють хвильовий пакет. Утім, амплітуда цієї хвилі не стала, а залежить від координати *x* і часу *t*:

$$A(x,\,t)=2A_0\cos(\tfrac{\Delta\omega}{2}t-\tfrac{\Delta k}{2}x). \quad (10.24)$$

Цей вираз для амплітуди хвильового пакета сам по собі є виразом рівняння плоскої синусоїдаль-ної хвилі − хвилі амплітуди коливань, фаза якої

$$\varphi_A=\tfrac{\Delta\omega}{2}t-\tfrac{\Delta k}{2}x. \quad (10.25)$$

Таке коливання, як зазначалось у підрозд. 9.5, називають *биттям*. Графік його для коливань, що поширюються у недисперсійному середовищі, зображено на рис. 9.7.

Знайдемо, чому дорівнює швидкість результуючої хвилі у дисперсійному середовищі. Хвиля, поширюючись у середовищі, переносить енергію.

Швидкість v поширення енергії хвилі дорівнює швидкості переміщення у просторі поверхні, через яку переноситься максимальна кількість енергії за одиницю часу. Оскільки енергія в разі коливального процесу пропорційна квадрату амплітуди, то швидкість поширення енергії хвильового пакета збігається з фазовою швидкістю хвилі амплітуди. Вочевидь, що для того, щоб амплітуда $A(x,t)$ (див. (10.24)) мала стале значення, наприклад максимальне, її фаза φ_A мусить залишатися сталою, тобто

$$\tfrac{\Delta\omega}{2}t-\tfrac{\Delta k}{2}x=\text{const}. \quad (10.26)$$

Взявши повний диференціал від цього виразу, отримаємо

$$\tfrac{\Delta\omega}{2}dt-\tfrac{\Delta k}{2}dx=0.$$

Звідси

$$v_{\text{гр}}=\frac{dx}{dt}=\frac{\Delta\omega}{\Delta k}=\frac{\omega_1-\omega_2}{k_1-k_2}\cdot \quad (10.27)$$

За умови, що $\Delta\omega$, а отже, і Δk, прямують до нуля,

$$v_{\text{гр}}=\frac{d\omega}{dk}. \quad (10.28)$$

Цю швидкість називають ***груповою швидкістю пакета хвиль***.

Швидкість гармонічних хвиль (див. (10.7)) дорівнює $v=\omega/k$. Отже, згідно із співвідношеннями (10.7) і (10.28) групова швидкість пакета хвиль не дорівнює жодній із швидкостей синусоїдальних складових цього пакета.

Вираз для групової швидкості

(10.28) можна подати ще й у такій формі. Згідно з (10.7) $\omega = vk$. Підставивши це значення у (10.28), отримаємо

$$v_{\text{гр}} = \frac{d(vk)}{dk} = v + k\frac{dv}{dk}. \quad (10.29)$$

Якщо врахувати, що $k = 2\pi/\lambda$ й $dk = -\frac{2\pi}{\lambda^2}d\lambda$, то отримаємо такий вираз для групової швидкості:

$$v_{\text{гр}} = v - \lambda\frac{dv}{d\lambda}. \quad (10.30)$$

Якщо немає дисперсії хвиль у середовищі ($\frac{dv}{d\lambda} = 0$), фазові швидкості хвиль однакові й не залежать від довжини хвилі, тому в таких середовищах групова швидкість хвиль збігається з фазовою швидкістю.

10.6. Інтерференція хвиль

Інтерференція хвиль (від лат. inter – взаємно, між собою та ferio – вдаряю, уражаю) – *складання у просторі двох (або декількох) хвиль, за якого в різних його точках виникає підсилення або послаблення амплітуди результуючої хвилі*. Інтерференція характерна для хвиль будь-якої природи; хвиль на поверхні рідини, пружних (наприклад, звукових), електромагнітних (наприклад, радіохвиль або світла). Інтерференційна картина – регулярне чергування областей підвищеної і зниженої інтенсивності хвиль. Як приклад наведемо інтерференційну картину хвиль на поверхні води, які збуджуються у двох різних точках (рис. 10.7). ***В основі явища*** інтерференції лежить *принцип суперпозиції хвиль*, справедливість якого підтверджується на дослідах. Суть його полягає у такому. Якщо в середовищі поширюється одночасно декілька хвиль, то коливання частинок середовища є геометричною сумою коливань, які здійснювали б частинки під час поширення кожної хвилі окремо. Отже, хвилі просто накладаються одна на одну, не спотворюючись.

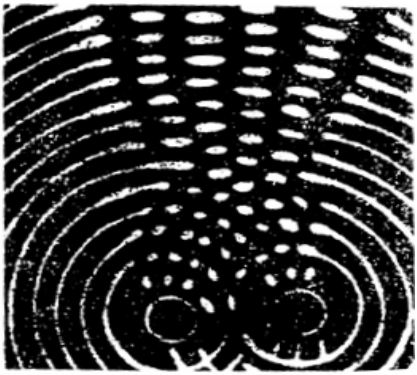


Рис. 10.7. *Інтерференційна картина хвиль на поверхні води*

У місцях зустрічі хвиль коливання середовища, які викликані кожною із хвиль, складаються одне з одним. Результат додавання (результуюча хвиля) залежить від співвідношення фаз,

періодів, напрямків і амплітуд хвиль, що накладаються.

Розглянемо накладання двох косинусоїдальних сферичних хвиль, які збуджуються точковими джерелами S_1 і S_2 (рис. 10.8). Ці хвилі в точці накладання P викликають коливання частинок середовища. Ці коливання згідно з (10.9) можна описати такими рівняннями:

$$S_1 = \frac{A_1}{r_1}\cos(\omega_1 t - k_1 r_1 + \varphi_{01}) = \\ = \frac{A_1}{r_1}\cos\varphi_1,$$

$$S_2 = \frac{A_2}{r_2}\cos(\omega_2 t - k_2 r_2 + \varphi_{02}) = \\ = \frac{A_2}{r_2}\cos\varphi_2.$$

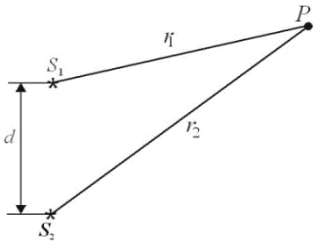


Рис. 10.8. *накладання двох косинусоїдальних сферичних хвиль, які збуджуються точковими джерелами* S_1 *і* S_2

Результат додавання цих коливань залежить від спрямованості. Будемо вважати, що відстань між джерелами коливань S_1 та S_2, $d << r_1$ та $d << r_2$. Тоді можна вважати, що в точці P додаються коливання однаково спрямовані і, згідно з формулою (9.20), амплітуда результуючого коливання (хвилі) в точці P

$$A_P^2 = \left(\frac{A_1}{r_1}\right)^2 + \left(\frac{A_2}{r_2}\right)^2 + 2\frac{A_1}{r_1}\frac{A_2}{r_2}\cos\varphi, \quad (10.31)$$

де

$$\varphi = \varphi_1 - \varphi_2 = (\omega_1 - \omega_2)t + \\ + (k_2 r_2 - k_1 r_1) + \varphi_{01} - \varphi_{02} \quad (10.32)$$

є різницею фаз хвиль у точці P.

Якщо частоти хвиль різні, тобто $\omega_1 \neq \omega_2$, $k_1 \neq k_2$, та ще й ω_1, ω_2, k_1, k_2 φ_{01}, φ_{02} хаотично залежать від часу, то φ є функцією часу й $\cos\varphi$ може змінюватись хаотично від -1 до $+1$. Тоді середнє значення $\langle\cos\varphi\rangle = 0$ і середнє значення результуючої амплітуди хвилі в точці P буде

$$\langle A_p^2\rangle = \langle\left(\frac{A_1}{r_1}\right)^2\rangle + \langle\left(\frac{A_2}{r_2}\right)^2\rangle. \quad (10.33)$$

Оскільки згідно з (10.20) квадрат амплітуди хвилі пропорційний її інтенсивності, то в цьому випадку в точці P будемо мати додавання інтенсивності двох хвиль. Такі хвилі називають *некогерентними.*

Дамо визначення, що ж називають когерентністю.

Когерентність (від лат. cohaerens – що перебуває у зв'язку) – узгоджене протікання з часом й у просторі декількох коливальних або хвильових процесів, яке проявляється при

їхньому додаванні. Коливання (хвилі) називають когерентними, якщо різниця фаз залишається сталою або закономірно змінюється з часом і при додаванні коливань (хвиль) визначає амплітуду сумарного коливання.

Таким чином, хвилі, а значить і коливання, в точці P будуть когерентними, якщо різниця фаз φ (формула (10.32)) не буде залежати від часу. Це буде у випадку, якщо $\omega_1 = \omega_2$ (монохроматичні хвилі) і $\varphi_{01} - \varphi_{02} = =$ const. За умови, що швидкість поширення хвиль однакова $(v_1 = v_2)$, $k_1 = k_2 = \frac{\omega_1}{v_1} = \frac{\omega_2}{v_2} = k$. Хвильове число не залежатиме від часу, якщо частота ω і швидкість v не залежатимуть від часу. За цих умов

$$\varphi = k(r_2 - r_1) + \varphi_{01} - \varphi_{02} \quad (10.34)$$

не буде залежати від часу і коливання в точці P, а значить і хвилі, які збудили ці коливання, будуть когерентними.

Тоді результат інтерференції двох хвиль, що визначається формулою (10.31), у різних точках залежить від величини $\Delta = r_2 - r_1$, яку називають *геометричною різницею ходу хвиль*.

У точках, для яких $\cos\varphi = +1$, буде максимум інтерференції. Амплітуда результуючого коливання в таких точках згідно з (10.31)

$$A_P = \frac{A_1}{r_1} + \frac{A_2}{r_2}. \quad (10.35)$$

Умову максимуму згідно з (10.34) можна записати так:

$$k(r_2 - r_1) + \varphi_{01} - \varphi_{02} = \pm 2m\pi, \quad (10.36)$$

де $m = 0, 1, 2, 3, \ldots$.

У точках, для яких $\cos\varphi = -1$, буде мінімум інтерференції: амплітуда результуючого коливання в таких точках згідно з (10.31)

$$A_P = \left|\frac{A_1}{r_1} - \frac{A_2}{r_2}\right|. \quad (10.37)$$

Умову мінімуму згідно з (10.34) можна записати так:

$$k(r_2 - r_1) + \varphi_{01} - \varphi_{02} = \pm(2m + 1)\pi,$$

де $m = 0, 1, 2, 3, \ldots$. (10.38)

Дуже часто реалізується випадок, коли $\varphi_{01} - \varphi_{02} = 0$. Тоді умова (10.36) набуває такого вигляду (з урахуванням, що $k = 2\pi/\lambda$, де λ – довжина хвилі в цьому середовищі):

$$\Delta = r_2 - r_1 = \pm 2m\frac{\lambda}{2}. \quad (10.39)$$

Умова мінімуму (10.38) тоді набуде вигляду

$$\Delta = (r_2 - r_1) = \pm(2m + 1)\frac{\lambda}{2}. \quad (10.40)$$

Під час інтерференції хвиль їхня інтенсивність (енергія) механічно не підсумовується. Інтерференція хвиль спричиняє перерозподіл інтенсивності

(енергії) коливань між сусідніми областями середовища.

На рис. 10.7 показано дві хвилі, які інтерферують: гребні хвиль – світлі лінії, западини – темні місця.

У місцях перетину двох гребнів або двох западин розміщено максимуми коливань (світлі плями), в місцях перетину гребнів і западин розміщено мінімуми (темні місця).

Особливим випадком інтерференції є стоячі хвилі.

Стоячі хвилі *– це хвилі, які утворюються при накладанні двох біжучих хвиль, що поширюються назустріч одна одній з однаковими частотами та амплітудами.*

Нехай дві плоскі хвилі поширюються назустріч одна одній уздовж осі *OX* в середовищі без згасання. Рівняння цих хвиль

$$S_1 = A_0 \cos(\omega t - kx),$$

$$S_2 = A_0 \cos(\omega t + kx + \varphi),$$

де φ – різниця фаз хвиль у точці $x = 0$; $k = 2\pi/\lambda$ (рис. 10.9).

Додавши ці рівняння й перетворивши результат за формулою для суми косинусів з урахуванням, що $k = 2\pi/\lambda$, отримаємо

$$S = S_1 + S_2 = 2A_0 \cos\left(\frac{2\pi x}{\lambda} + \frac{\varphi}{2}\right) \cos\left(\omega t + \frac{\varphi}{2}\right). \quad (10.41)$$

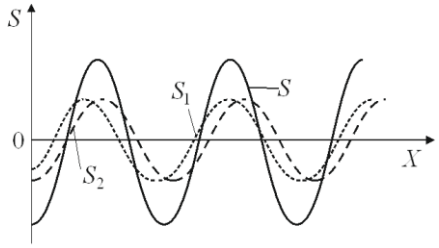


Рис. 10.9. Дві плоскі хвилі поширюються назустріч одна одній уздовж осі OX

Множник $\cos\left(\omega t + \frac{\varphi}{2}\right)$ показує, що в точках середовища виникає коливання з тією ж самою частотою ω, що і коливання зустрічних хвиль.

Множник $2A_0 \cos\left(\frac{2\pi x}{\lambda} + \frac{\varphi}{2}\right)$, який не залежить від часу, виражає амплітуду $A_{ст}$ результуючих хвиль, точніше – амплітуда як величина позитивна дорівнює абсолютному значенню цього множника:

$$A_{ст} = \left|2A_0 \cos\left(\frac{2\pi x}{\lambda} + \frac{\varphi}{2}\right)\right|.$$

Амплітуда результуючого коливання залежить від координати *x*, що визначає положення точок середовища.

У точках середовища, де

$$\frac{2\pi x}{\lambda} + \frac{\varphi}{2} = \pm m\pi \quad (m = 0, 1, 2, 3, ...),$$

амплітуда $A_{ст}$ досягає максимального значення $2A_0$. Точки, в яких $A_{ст}$ максимальна, називають *пучностями стоячої хвилі.*

У точках середовища, де

$$\frac{2\pi x}{\lambda}+\frac{\varphi}{2}=\pm(2m+1)\frac{\pi}{2}$$

за $m = 0, 1, 2, 3, \ldots$ маємо $A_{ст} = 0$. Ці точки називають *вузлами стоячої хвилі.* Точки середовища, що перебувають у вузлах, не коливаються. Виберемо початок відліку x так, щоб φ дорівнювало нулю. Тоді координати пучностей

$$x_{п} = \pm m\frac{\lambda}{2}, \qquad (10.42)$$

а вузлів

$$x_{в} = \pm(2m+1)\frac{\lambda}{4}\cdot \qquad (10.43)$$

Відстань між двома сусідніми пучностями отримаємо, якщо знайдемо різницю двох значень $x_{п}$ для двох послідовних значень m:

$$x_{п}(m+1) - x_{п}(m = (m+1)\frac{\lambda}{2} - m\frac{\lambda}{2} = \frac{\lambda}{2},$$

тобто відстань між сусідніми пучнос-тями дорівнює половині довжини хвилі тих хвиль, у результаті інтерференції яких утворюється ця стояча хвиля.

Відстань від вузла до найближчої пучності

$$x_{в} - x_{п} = (2m+1)\frac{\lambda}{4} - m\frac{\lambda}{2} = \frac{\lambda}{4}.$$

Множник $(2A_0 \cos 2\pi\frac{x}{\lambda})$ у виразі для амплітуди під час переходу через нульове значення змінює знак. Відповідно до цього фаза коливань по різні сторони від вузла відрізняється на π. Це означає, що точки, які лежать по різні сторони від вузла, коливаються у протифазі. Усі точки, які містяться між сусідніми вузлами, коливаються синфазно (тобто в однаковій фазі). На рис. 10.10 зображено ряд «моментальних фотографій» відхилення точок від положення рівноваги. Перша «фотографія» відповідає моменту, коли відхилення досягає найбільшого абсолютного значення. Наступні «фотографії» зроблені з інтервалом у четверть періоду. Стрілками показано напрямки швидкостей частинок.

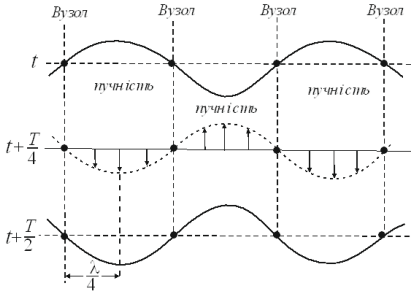


Рис. 10.10. «моментальних фотографій» відхилення точок від положення рівноваги

Отже, у стоячій хвилі є ряд нерухомих вузлових точок, які розміщені на відстані півхвилі одна від одної. Частинки між вузлами коливаються з різними амплітудами, від нуля у вузлі до подвійної амплітуди у пучності. Усі частинки одночасно проходять через положення рівноваги й одночасно досягають максимальних відхилень, а отже, коливаються в однакових фазах. В суміжному інтервалі між вузлами характер коливань такий самий, але фази

протилежні.

У стоячій хвилі енергія не переноситься – повна енергія коливань кожного елемента об'єму середовища, обмеженого сусіднім вузлом і пучністю, не залежить від часу. Вона лише переходить з кінетичної енергії в потенціальну енергію пружно деформованого середовища, і навпаки. Відсутність перенесення енергії стоячою хвилею є результатом того, що падаюча й відбита хвилі, які утворюють цю стоячу хвилю, переносять енергію в рівних кількостях у протилежних напрямках.

Стоячі хвилі утворюються, наприклад, на струні, повітряному стовпі в духових музичних інструментах, у камертоні. Вони є результатом інтерференції біжучої хвилі в одному напрямку й відбитої – у протилежному. Як засвідчують досвід і теорія, у разі відбивання хвиль від густішого середовища на межі поділу двох середовищ стояча хвиля матиме вузол, у разі її відбивання від середовища з меншою густиною – пучність.

Утворення вузла стоячої хвилі на межі поділу менш густого і більш густого середовищ можна пояснити так. На межі двох середовищ коливання, спричинені біжучою хвилею, додаються. Оскільки на межі таких середовищ утворюється вузол (тобто сума зміщень точок дорівнює нулю), можна дійти висновку, що в разі відбивання хвилі від густішого середовища її фаза змінюється на протилежну. Як відомо, фаза змінюється на протилежну на відстані, що дорівнює половині довжини хвилі. Тому можна стверджувати, що під час відбивання хвилі від густішого середовища «втрачається півхвилі».

Відбившись від поверхні менш густого середовища, хвиля не змінює фазу у місці відбивання, тому фази падаючої й відбитої хвиль на межі поділу таких двох середовищ однакові. У цьому разі зміщення точок буде вдвічі більшим, ніж зміщення, спричинене падаючою хвилею, тобто виникне пучність.

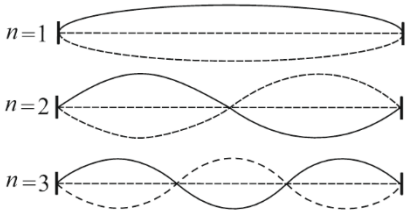


Рис. 10.11

У закріпленої з обох кінців натягнутої струни у разі збудження поперечних коливань встановлюються стоячі хвилі, причому в місцях закріплення струни мають утворюватись вузли, тому у струні збуджуються із значною інтенсивністю тільки такі коливання, половина довжини хвилі яких

вкладається на довжині струни цілу кількість разів (рис. 10.11).

Звідси випливає така умова:

$$l = n\lambda/2 \text{ або } \lambda_n = \frac{2l}{n}, \quad (10.44)$$

де l – довжина струни. Довжинам хвиль (10.44) відповідають частоти

$$\nu_n = \frac{v}{\lambda_n} = \frac{v}{2l}n, \quad (10.45)$$

за $(n = 1,\ 2,\ 3,\ \ldots)$, де v – фазова швидкість хвилі, яку визначають силою натягу струни і масою одиниці довжини, тобто лінійною густиною струни.

Частоти ν_n називають власними частотами струни. Власні частоти є кратними частоті $\nu_1 = \frac{v}{2l}$, яку називають основною частотою.

Гармонічні коливання з частотами (10.45) називають власними, або нормальними коливаннями. Їх називають також гармоніками. У загальному випадку коливання струни є накладанням різних гармонік.

Коливання струни дивні в тому сенсі, що для них, за класичними уявленнями, отримують дискретні значення однієї з величин, яка характеризує коливання (частоти). Для класичної фізики така дискретність є винятком. Для квантових процесів дискретність є швидше правилом, ніж винятком.

10.7. Звук

Звук у широкому сенсі – коливальний рух частинок пружного середовища, що поширюється у вигляді хвилі в газоподібному, рідкому чи твердому середовищі – це те ж саме, що і пружні хвилі; у вузькому сенсі – явище, що суб'єктивно сприймається органами слуху людини і тварин.

Людина чує звук в діапазоні частот від 16 Гц до 20 кГц. Людина не чує звуки з частотою, нижчою за 16 Гц (інфразвук) і вищою за 20 кГц (ультразвук). Високочастотні пружні хвилі в діапазоні від 10^9 до 10^{12}–10^{13} Гц називають гіперзвуком.

Важливою характеристикою звуку є його спектр, який отримують у результаті розкладання звуку на прості гармонічні коливання (частотний аналіз звуку). Основна частота, яку сприймають на слух, визначає при цьому висоту звуку, а набір гармонічних складових – тембр звуку. Набір частот коливань, які наявні у певному звуці, називають його акустичним спектром. Якщо у звуці є коливання всіх частот у деякому інтервалі від ν' до ν'', то спектр називають *суцільним*. Якщо звук складається з коливань дискретних частот ν_1, ν_2, ν_3 і т.д., то спектр називають *лінійчастим*. Суцільний акустичний спектр мають

шуми. Коливання з лінійчастим спектром викликають відчуття звуку з більш-менш визначеною частотою. Такий звук називають *тональним.*

Висоту тонального звуку визначають основною (найменшою) частотою ν_1. Відносна інтенсивність *обертонів* (тобто коливань з частотами ν_2, ν_3 і т.д.) визначає забарвлення, або тембр звуку. Різний спектральний склад звуків, які збуджуються різними музичними інструментами, дозволяє відрізнити на слух, наприклад, флейту від скрипки.

Енергетичною характеристикою звукових коливань є *інтенсивність звуку* I, тобто середня за часом енергія, яка переноситься хвилею за одиницю часу через одиничну площадку, що перпендикулярна до напрямку поширення хвилі, й виражається через амплітуду звукового тиску:

$$I = \frac{\Delta p_0^2}{2 v_{\parallel} \rho}, \qquad (10.46)$$

де Δp_0 – амплітуда звукового тиску; $v_{\parallel}$ – швидкість звуку в середовищі; ρ – густина середовища.

Звуковий тиск – це змінна частина тиску, яка виникає під час проходження звукової хвилі в середовищі. Поширюючись в середовищі, звукова хвиля утворює його згущення й розрідження, які створюють додаткові зміни тиску щодо його середнього значення в середовищі. Звуковий тиск здійснюється з частотою, яка дорівнює частоті звукової хвилі. Звуковий тиск – основна кількісна характеристика звуку. Амплітуда звукового тиску Δp_0 й амплітуда швидкості коливань v_m частинок середовища у звуковій хвилі пов'язані співвідношенням

$$\Delta p_0 = v_{\parallel} \rho v_m. \qquad (10.47)$$

Щоб викликати звукове відчуття, хвиля повинна мати деяку мінімальну інтенсивність, яку називають *порогом чутності* (рис. 10.12, нижня крива). Поріг чутності дещо різний для різних осіб і сильно залежить від частоти звуку. Найбільш чутливе людське вухо до частот 1000 … 5000 Гц (рис. 10.12). У цій області частот поріг чутності дорівнює середній інтенсивності звуку близько $5 \cdot 10^{-13}$ Вт/м^2. За інших частот поріг чутності має більшу інтенсивність звуку (рис. 10.12, нижня крива).

Якщо інтенсивність звуку становить 0,5... 5 Вт/м^2, хвиля перестає сприйматись як звук і викликає тільки відчуття болю та тиску. Значення інтенсивності, за якого це відбувається, називають *порогом больового відчуття.* Поріг больового відчуття, як і поріг чутності, залежить від частоти (верхня крива на рис. 10.12; дані, подані на цьому рисунку, належать до середнього

нормального слуху).

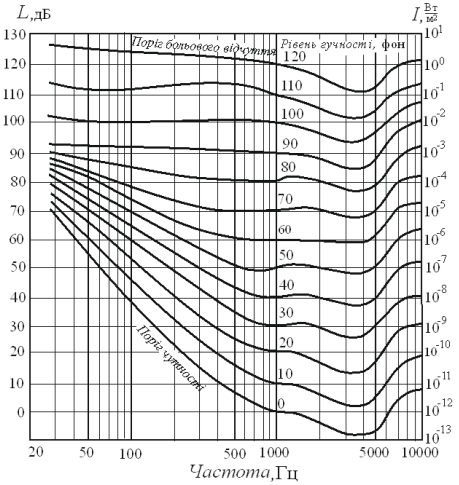


Рис. 10.12. Криві однакової гучності, які виражають залежність рівня звукового тиску (в децибелах) від частоти за заданої гучності (у фонах). Нижня крива – поріг чутності, верхня крива – поріг больового відчуття

Гучність звуку – величина, яка характеризує слухове відчуття для певного звуку. Гучність звуку складним чином залежить від інтенсивності (тиску) звуку, частоти й форми коливань (рис. 10.12). За незмінних частоти й форми коливань гучність звуку зростає зі збільшенням інтенсивності (тиску) звуку (рис. 10.12). За однакової інтенсивності (тиску) звуку гучність звуку чистих тонів (гармонічних коливань) різних частот різна, тобто на різних частотах однакову гучність можуть мати звуки різної інтенсивності (тиску).

Гучність звуку, яку оцінюють суб'єктивно, зростає значно повільніше, ніж інтенсивність (тиск) звукових хвиль. Зі зростанням інтенсивності (тиску) в геометричній прогресії гучність зростає приблизно в арифметичній прогресії, тобто лінійно.

Виходячи з цього *рівень гучності* L визначають як логарифм відношення інтенсивності (тиску) дапевного звуку I (або Δp) до інтенсивності (тиску) I_0 (або Δp_0), взятих за вихідні:

$$L = \lg \frac{I}{I_0} \cdot \qquad (10.48)$$

Вихідну інтенсивність I_0 беруть за величину 10^{-12} Вт/м^2, що відповідає пороговому звуковому тиску $\Delta p_п = 2 \times 10^{-5}$ Па. Як, видно, з рис. 10.12, поріг чутності за частоти чистого тону 1000 Гц лежить на нульовому рівні ($L = 0$).

Одиницю рівня гучності L, яку визначають формулою (10.48), називають белом (Б). Зазвичай користуються у 10 разів меншими одиницями – децибелами (дБ).

Значення L у децибелах знаходять за формулою

$$L = 10 \lg \frac{I}{I_0} . \cdot \qquad (10.49)$$

Децибел (дБ) – рівень звукового тиску Δp, для якого виконується співвідно-шення $201g(\Delta p/\Delta p_п) = 1$, де $\Delta p_п = 2 \cdot 10^{-5}$ Па – пороговий звуковий тиск (враховано, що $I \sim \Delta p^2$). Одному

децибелу відповідає рівень звукового тиску $\Delta p = 2{,}244 \cdot 10^{-5}$ Па.

Таблиця 10.1

Класифікація звуку за рівнем гучності

Звук	Рівень гучності L, дБ
Найбільш тихий звук, який здатне вловити людське вухо	10-15
Тікання годинника	20
Шепіт на відстані 1м	30
Негучна розмова	35-40
Крик	60
Шум працюючого поряд двигуна грузового автомобіля	80
Допустима гучність для людського вуха	Не більше 85
Шум в метро і шум від грузового залізнодорожного потяга	90
Навушники *mp3* плеера (*max*)	115-120
Шум літака, що злітає, на відстані 30м від нього	130
Можливість контузії	130-155
Шок	155-160
Розрив барабанних перетинок	160-175
Розрив легенів	175-200
Вибух атомної бомби (причина смерті)	>200

За формулою (10.49) один децибел відповідає зміні інтенсивності звуку в $10^{0,1} = 1{,}26$ рази. Різні рівні гучності наведено в табл. 10.1. Рівень гучності звуку L_N (у фонах) залежить від інтенсивності звуку, його частоти, його розраховують за формулою

$$L_N = 10 \lg\left(\frac{I_N}{I_0}\right), \qquad (10.50)$$

де I_N – інтенсивність звуку з частотою $\nu = 1000$ Гц, який має однакову гучність із досліджуваним звуком. Залежність L_N від інтенсивності звуку та його частоти складна і простим аналітичним співвідношенням її не виражають. Рівень гучності звуку L_N визначають за кривими (рис. 10.12). Фон – одиниця рівня гучності певного звуку. Рівень гучності цього звуку у фонах дорівнює рівню звукового тиску в децибелах для чистого тону з частотою 1000 Гц, гучність якого, якщо порівнювати на слух, дорівнює гучності цього звуку. Для звуку еталонної частоти (ν = 1000 Гц) рівень гучності звуку, який виражений у фонах, чисельно дорівнює рівню інтенсивності, що виражається у децибелах, $L = L_N$ (числова рівність справедлива тільки для частоти ν = 1000 Гц).

Діапазон інтенсивності звуку, за якого хвиля викликає в людському вусі звукове відчуття, становить від $2 \cdot 10^{-13}$ Вт/м^2 за 3000 ... 4000 Гц до майже 10 Вт/м^2 близько частоти 30 Гц. Відповідні значення рівня гучності від -8 до +130 дБ показано на рис. 10.12.

Джерелами звуку можуть бути будь-які явища, які викликають місцеву зміну тиску або механічних напружень. Широко поширені джерела звуку у вигляді твердих тіл, які коливаються (наприклад, дифузори гучномовців і мембрани телефонів, струни й деки музичних інструментів); в ультразвуковому діапазоні частот це пластинки і

стрижні із п'єзодіелектричних або магнітострикційних матеріалів. Широкий клас джерел звуку – електроакустичні перетворювачі.

До приймачів звуку належать, зокрема, слухові апарати людини і тварин. У техніці для приймання звуку застосовують, головним чином, електроакустичні перетворювачі: у повітрі – мікрофони, у воді – гідрофони, в земній корі – геофони.

Поширення звукової хвилі характеризується насамперед *швидкістю звуку.* У деяких випадках спостерігається дисперсія швидкості звуку, тобто залежність швидкості його поширення від частоти. Під час поширення звукової хвилі відбувається поступове згасання звуку, тобто зменшення його інтенсивності й амплітуди, яке значною мірою обумовлюється поглинанням звуку, що пов'язане з необоротним переходом звукової енергії в інші форми (головним чином у теплоту).

Швидкість звуку становить:

– у газах

$$v = \sqrt{\gamma \frac{p}{\rho}}, \qquad (10.51)$$

де γ – відношення теплоємності газу за постійного тиску до теплоємності за постійного об'єму ($\gamma \approx 1{,}5$); p і ρ – тиск і густина незбуреного хвилею газу відповідно;

– в ізотропних твердих тілах для поздовжньої хвилі

$$v_{\parallel} = \sqrt{\frac{E}{\rho}}, \qquad (10.52)$$

а для поперечної хвилі

$$v_{\perp} = \sqrt{\frac{G}{\rho}}, \qquad (10.53)$$

де E – модуль Юнга; G – модуль зсуву; ρ – густина речовини твердого тіла.

10.8. Ефект Доплера для звукової хвилі

У 1842 р. в журналі Празького наукового товариства з'явилась стаття під назвою «Про кольорове світло подвійних зірок і деяких інших сузір'їв неба». Автор статті, австрійський професор математики Християн Доплер, запропонував нове й неочікуване пояснення забарвлення небесних об'єктів. Свою теорію Доплер будував на *твердженні про те, що частота хвильових коливань має залежати від швидкості руху джерела відносно спостерігача* (тобто приладу, який сприймає ці хвилі). Інакше кажучи, один і той самий звук буде мати різну висоту, а світло – різне забарвлення (колір) для джерела, яке рухається або перебуває у стані спокою.

Щоб краще зрозуміти ефект Доплера для звуку, розглянемо такий наочний випадок. В озері плавають два човни. На одному з них у воду занурений

прилад, який може випромінювати звукові коливання з частотою ν_0. Із другого човна у воду занурений приймач, який може реєструвати частоту сигналу, який надходить.

Слід встановити, яку частоту звуку буде реєструвати приймач у таких випадках:

1. Обидва човни перебувають у стані спокою на воді.

2. Човен із джерелом перебуває у стані спокою, човен із приймачем рухається.

3. Човен із джерелом рухається, човен із приймачем перебуває у стані спокою.

Розглянемо кожний випадок окремо.

1. Відповідь на це запитання очевидне: приймач реєструє звук, частота якого ν_0. Справді, якщо швидкість звуку у воді v_0, то довжина звукової хвилі, яку випромінює джерело, $\lambda_0 = v_0/\nu_0$. Спостерігач у другому човні, який вимірює швидкість звуку та довжину хвилі, записує саме ці величини: $v_1 = v_0$ та $\lambda_1 = \lambda_0$. Так що частота звуку, яку зареєструє спостерігач, буде становити

$$\nu_1 = \frac{v_1}{\lambda_1} = \frac{v_0}{\lambda_0} = \nu_0. \qquad (10.54)$$

2. Нехай тепер човен із спостерігачем почав рухатись вздовж прямої, яка з'єднує обидва човни. Позначимо швидкість човна із спостерігачем через $v_{\text{сп}}$ й будемо вважати, що вона додатна, якщо відстань між човнами скорочується. Швидкість звуку, яку тепер буде вимірювати спостерігач, буде дорівнювати не v_0, а

$$v_2 = v_0 + v_{\text{сп}} = v_0\left(1 + \frac{v_{\text{сп}}}{v_0}\right).$$

Довжина хвилі звуку у воді не залежить від того, рухається спостерігач чи ні, тому $\lambda_2 = \lambda_0 = v_0/\nu_0$. Водночас $\lambda_2 = v_2/\nu_2$, отже, $v_0/\nu_0 = v_2/\nu_2$, тобто

$$\nu_2 = \frac{v_2}{v_0}\nu_0 = \frac{v_0 + v_{\text{сп}}}{v_0}\nu_0 = \left(1 + \frac{v_{\text{сп}}}{v_0}\right)\nu_0. \qquad (10.55)$$

Це означає, що якщо відстань між човнами скорочується, то приймач зареєструє більш високу частоту, ніж частота джерела звуку. Навпаки, якщо човен із спостерігачем віддаляється від човна із джерелом, то частота звуку, яку реєструє приймач, зменшується:

$$\nu_2' = \left(1 - \frac{v_{\text{сп}}}{v_0}\right)\nu_0. \qquad (10.55')$$

3. Розглянемо випадок, коли човен із джерелом рухається у напрямку до спостерігача із швидкістю $v_{\text{дж}}$. Швидкість звуку у воді, як і раніше, дорівнює v_0. Довжина хвилі, яку буде вимірювати спостерігач, тепер буде меншою, ніж у тому випадку, коли човни перебували у стані спокою. Справді, довжина хвилі – це відстань між двома

сусідніми точками, які коливаються у фазі, наприклад, між сусідніми гребнями хвилі. За час $T = \lambda_0/v_0$, поки один гребінь хвилі віддалиться від точки, в якій він був «висланий», на відстань λ_0, джерело зміститься в тому ж напрямку на відстань $v_{дж} \cdot T = v_{дж}\lambda_0/v_0$, тому відстань між «першим» гребнем і наступним буде дорівнювати не λ_0, а

$$\lambda_3 = \lambda_0 - \frac{v_{дж}\lambda_0}{v_0} = \lambda_0(1 - v_{дж}/v_0).$$

Отже, частота звуку, яку зареєструє спостерігач, буде дорівню-вати

$$\nu_3 = \frac{v_0}{\lambda_3} = \frac{v_0}{\lambda_0(1 - v_{дж}/v_0)},$$

або

$$\nu_3 = \frac{\nu_0}{(1 - v_{дж}/v_0)}. \qquad (10.56)$$

У випадку, коли човен із джерелом віддаляється від човна з приймачем, частота звуку, яку буде реєструвати приймач,

$$\nu_3' = \frac{\nu_0}{1 + v_{дж}/v_0} \cdot \qquad (10.56')$$

Ми бачимо, що формули (10.56) і (10.56′) відрізняються від формул (10.55), хоч основні закономірності зберігаються: частота звуку підвищується, коли човни зближаються, і знижується, коли човни розходяться.

Утім, якщо швидкість човна мала порівняно із швидкістю звуку у воді, то ми можемо скористатись наближеними рівностями:

$$\frac{1}{1-\alpha} \approx 1 + \alpha, \quad \frac{1}{1+\alpha} \approx 1 - \alpha,$$

де $\alpha \ll 1$.

У цьому випадку формули (10.56) та (10.56′) приблизно збігаються з формулами (10.55):

$$\nu_3 \approx \nu_2 = (1 \pm \frac{v_{дж}}{v_0})\, \nu_0.$$

Можна написати і загальну формулу для випадку, коли рухаються обидва човни – і джерело, та приймач. Таку формулу отримують послідовним застосуванням формул (10.55) і (10.56), отриманих вище:

$$\nu = \frac{1 \pm v_{сп}/v_0}{1 \mp v_{дж}/v_0}\nu_0, \qquad (10.57)$$

або наближено, враховуючи, що

$(1 + \alpha)(1 + \beta) \approx 1 + \alpha + \beta$ за $\alpha, \beta \ll 1$, отримаємо

$$\nu \approx \left(1 \pm \frac{v_{сп} \pm v_{дж}}{v_0}\right)\nu_0. \qquad (10.58)$$

Неважко бачити, що $\nu_{сп} \pm v_{дж}$ є відносною швидкістю джерела і приймача. Знак «+» відповідає їх руху в протилежних напрямках, знак «−» – в одному напрямку, тому формулу для ефекту Доплера в наближенні малих швидкостей можна записати так:

$$\nu \approx \left(1 \pm \frac{v}{v_0}\right)\nu_0, \qquad (10.59)$$

де ν – відносна швидкість джерела і приймача звуку.

Формулу (10.59) можна ще раз узагальнити на той випадок, коли також рухається середовище. (Це узагальнення

означає, що в розглянутому нами прикладі човни перебувають не в озері, а у річці). У цьому випадку загальний вигляд формули (10.59) не змінюється, але під величинами $v_{дж}$ і $v_{сп}$ тепер слід розуміти швидкості джерела і спостерігача відносно річки; відповідно до цього слід підставляти й значення v у формулу (10.59).

Ми розглядали випадок, коли джерело і приймач рухаються вздовж однієї прямої й уздовж цієї ж прямої поширюється фронт звукової хвилі. Якщо ж напрямок поширення звуку утворює кут θ з напрямком відносної швидкості v, то частота, яку реєструє приймач,

$$\nu \approx \left(1 + \frac{v}{v_0}\cos\theta\right)\nu_0. \quad (10.60)$$

Зміна частоти ν_0 визначається проєкцією $v\cos\theta$ відносної швидкості на напрямок поширення звуку. Якщо $\theta = 90^\circ$, то $\nu = \nu_0$, тобто поперечного ефекту Доплера у звичайній акустиці немає.

Ефект Допплера для звукових хвиль вперше експериментально перевірили у 1845 р. Музиканти на слух оцінювали звучання музичного інструмента, встановленого на платформі потяга, що проходив повз них.

Контрольні запитання

1. Дайте визначення рівняння хвилі.

2. Запишіть рівняння біжучої плоскої і сферичної хвиль. У чому відмінність між цими рівняннями?

3. Запишіть і поясніть структуру і зміст хвильового рівняння для одновимірного випадку.

4. Поясніть зміст характеристик хвилі: амплітуда, фаза, частота, довжина хвилі (хвильове число), фазова швидкість хвиль у пружному середовищі.

5. Розкрийте суть густини потоку енергії хвиль – вектора Умова.

6. Що таке групова швидкість хвиль? Чим вона відрізняється від фазової швидкості?

7. Що таке звукові хвилі? Дайте визначення характеристик звукових хвиль: амплітуда звукового тиску, інтенсивність звуку, гучність звуку.

8. Які хвилі є когерентними? Дайте визначення явища інтерференції хвиль. Які умови максимуму та мінімуму інтерференції?

11. Як проявляється ефект Доплера для звукових хвиль?

Розділ 11. Електромагнітні хвилі

11.1. Загальні відомості про електромагнітні хвилі

Електромагнітні хвилі – це електромагнітні коливання, які поширюються у просторі з кінцевою швидкістю. Існування електромагніт-них хвиль передбачив англійський фізик М. Фарадей у 1832 р. Англійський фізик Дж. Максвелл у 1865 р. теоретично показав, що електромагнітні коливання поширюються в вакуумі із швидкістю світла. У 1888 р. максвеллівська теорія електромагнітних хвиль отримала підтвердження в дослідах німецького фізика Г. Герца, що відіграло вирішальну роль в її утвердженні.

Теорія Максвелла дозволила встановити, що радіохвилі, світло, рентгенівське і гамма-випромінювання є електромагнітмими хвилями з різною довжиною хвилі (див. табл. 11.1), причому між сусідніми діапазонами шкали електромагнітних хвиль не має різких границь (рис. 11.1). Назви різним діапазонам спектрів було дано за способом їх збудження і за методами спостереження. Властивості різних діапазонів хвиль також різняться за їх дією, за проникною здатністю. Кожний із цих діапазонів має широке практичне

Таблиця 11.1

Спектр електромагнітних хвиль

Частота ν, Гц	Довжина хвилі λ, м	Назва діапазону	Джерела. Основні методи збудження
$0 - 10^3$	$\infty - 3 \cdot 10^5$	Низькочастотні хвилі	Змінні струми. Генератори
$10^3 - 10^{12}$	$3 \cdot 10^5 - 3 \cdot 10^{-4}$	Радіохвилі	Генератори радіочастот. Генератори НВЧ
$8{,}6 \cdot 10^{11} - 4 \cdot 10^{14}$	$3{,}5 \cdot 10^{-4} - 7{,}5 \cdot 10^{-7}$	Інфрачервоні промені	Теплове випромінювання атомів і молекул
$4 \cdot 10^{14} - 7{,}5 \cdot 10^{14}$	$7{,}5 \cdot 10^{-7} - 4 \cdot 10^{-7}$	Видиме світло	Випромінювання атомів і молекул при нагріванні і при електричних розрядах
$7{,}5 \cdot 10^{14} - 6 \cdot 10^{16}$	$0{,}4 \cdot 10^{-6} - 5 \cdot 10^{-9}$	Ультрафіолетові промені	Випромінювання атомів під час дії на них швидких електронів
$3 \cdot 10^{16} - 3 \cdot 10^{20}$	$10^{-8} - 10^{-12}$	Рентгенівські промені	Випромінювання атомів при збудженні зовнішніх і внутрішніх електронів
$3 \cdot 10^{19}$ і більше	10^{-11} і менше	γ– промені	Ядерні процеси, радіоактивний розпад, космічні процеси

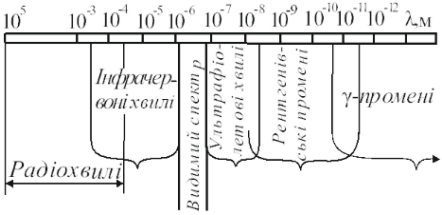


Рис. 11.1. Шкала електромагнітних хвиль

застосування. В електродинаміці розглядають переважно низькочастотні радіохвилі. Інші частотні діапазони вивчають в оптиці та в атомній і ядерній

фізиці. Особливості електромагнітних хвиль, закони їх збудження і поширення описують рівняннями Максвелла.

11.2. Доведення необхідності існування електромагнітних хвиль

У розд. 8 ми з'ясували, що змінне електричне поле породжує магнітне, яке, взагалі кажучи, теж є змінним. Це змінне магнітне поле породжує змінне електричне і т.д. Таким чином, якщо в якому-небудь місці простору виникає змінне електричне або змінне магнітне поле, то можливе збудження в послідовних точках простору змінного електричного поля змінним магнітним полем і навпаки, тобто поширення цих полів у вигляді єдиного електромагнітного поля у просторі, що і є утворенням електромагнітної хвилі. Покажемо, що існування електромагнітних хвиль випливає з рівнянь Максвелла. Розглянемо випадок однорідного, нейтрального ($\rho = 0$) непровідного ($\vec{j} = 0$) середовища з постійними проникностями ε та μ. Будемо розглядати одновимірну задачу: вектори $\vec{E}$, $\vec{D}$, $\vec{B}$ і $\vec{H}$ залежать тільки від однієї координати (від x) і від часу. Тоді, оскільки всі похідні за y та z у формулах (8.29)–(8.27) дорівнюють нулю, то їх можна переписати так:

$$\frac{\partial B_x}{\partial t} = 0, \tag{11.1}$$

$$\frac{\partial E_z}{\partial x} = \frac{\partial B_y}{\partial t}, \tag{11.2}$$

$$\frac{\partial E_y}{\partial x} = -\frac{\partial B_z}{\partial t}, \tag{11.3}$$

$$\frac{\partial B_x}{\partial x} = 0, \tag{11.4}$$

$$\frac{\partial D_x}{\partial t} = 0, \tag{11.5}$$

$$\frac{\partial H_z}{\partial x} = -\frac{\partial D_y}{\partial t}, \tag{11.6}$$

$$\frac{\partial H_y}{\partial x} = \frac{\partial D_z}{\partial t}, \tag{11.7}$$

$$\frac{\partial D_x}{\partial x} = 0. \tag{11.8}$$

Із рівнянь (11.1) і (11.5) випливає, що складові векторів $\vec{B}$ і $\vec{D}$ по осі OX не залежать від часу, а з (11.4) і (11.8) випливає, що B_x і D_x не залежать також і від x. Утім, вище припускалось, що вектори $\vec{E}$, $\vec{D}$, $\vec{B}$ і $\vec{H}$ не залежать від y і z. Таким чином, B_x і D_x не залежать ані від часу, ані від однієї з координат, тобто постійні з часом та у просторі:

$$B_x = \text{const}, \; D_x = \text{const},$$

де B_x і D_x – постійні однорідні статичні поля, які накладаються на змінне електромагнітне поле. Статичні поля не впливають на переміщення електромагнітного поля, тобто на утворення електромагнітної хвилі, і їх можна не враховувати. Будемо вважати, що $B_x = 0$ та $D_x = 0$.

Для випадку, який ми розглядаємо, диференціальні рівняння Максвелла можуть бути записані так:

$$\frac{\partial D_y}{\partial t} = -\frac{\partial H_z}{\partial x}, \tag{11.9}$$

$$\frac{\partial B_z}{\partial t} = -\frac{\partial E_y}{\partial x}, \qquad (11.10)$$

$$\frac{\partial D_z}{\partial t} = \frac{\partial H_y}{\partial x}, \qquad (11.11)$$

$$\frac{\partial B_y}{\partial t} = \frac{\partial E_z}{\partial x}. \qquad (11.12)$$

Із (11.9) випливає, що змінне з часом електричне поле, напрямлене вздовж осі *OY*, викликає появу магнітного поля, яке напрямлене вздовж осі *OZ*. Зміна останнього з часом, як випливає з (11.10), призводить до появи електричного поля, яке напрямлене вздовж осі *OY*. Із (11.11) випливає, що змінне з часом електричне поле, яке направлене вздовж осі *OZ*, викликає появу магнітного поля, напрямленого вздовж осі *OY*, а змінне з часом магнітне поле, яке направлене вздовж осі *OY* (співвідношення (11.12)), призводить до появи направленого вздовж осі *OZ* електричного поля.

Із сказаного можна зробити такий висновок: змінні з часом електрична й магнітна складові електромагнітного поля взаємно перпендикулярні. Оскільки $B_x = 0$ й $D_x = 0$, то складові електромагнітного поля обидві перпендикулярні до осі *OX*. Поля E_y і H_z не викликають появу полів E_z і H_y, тому для опису плоскої електромагнітної хвилі достатньо взяти одну пару рівнянь (11.9) і (11.10), а компоненти, що є у рівняннях (11.11) і (11.12), прирівняємо до нуля, тобто $D_z = 0$ і $H_y = 0$. Врахуємо співвідношення (8.28) і (8.29) та (11.9) і (11.10) перепишемо так:

$$\varepsilon\varepsilon_0 \frac{\partial E_y}{\partial t} = -\frac{\partial H_z}{\partial x}, \qquad (11.13)$$

$$\mu\mu_0 \frac{\partial H_z}{\partial t} = -\frac{\partial E_y}{\partial x}. \qquad (11.14)$$

Продиференціюємо перше рівняння за *t* й виконаємо перестановку: $\frac{\partial}{\partial t}\frac{\partial H_z}{\partial x} = \frac{\partial}{\partial x}\frac{\partial H_z}{\partial t}$. У це співвідношення замість $\frac{\partial H_z}{\partial t}$ з рівняння (11.14) підставимо $\frac{\partial H_z}{\partial t} = -\frac{1}{\mu\mu_0}\frac{\partial E_y}{\partial x}$. Тоді із (11.13) отримаємо рівняння

$$\frac{\partial^2 E_y}{\partial x^2} = \varepsilon\mu\varepsilon_0\mu_0 \frac{\partial^2 E_y}{\partial t^2}. \qquad (11.15)$$

Якщо продиференціювати рівняння (11.14) за *t* й виконати аналогічні перетворення, отримаємо таке рівняння для H_z:

$$\frac{\partial^2 H_z}{\partial x^2} = \varepsilon\mu\varepsilon_0\mu_0 \frac{\partial^2 H_z}{\partial t^2}. \qquad (11.16)$$

Порівняння (11.15), (11.16) з (10.13) дає підстави стверджувати, що (11.15) і (11.16) є хвильовими рівняннями, які подають хвилі E_y і H_z, тобто процеси поширення у просторі одночасно існуючих і взаємно обумовлених змінних з часом електричного і магнітних полів, тобто електромагнітного поля, що є електромагнітною хвилею. У цьому випадку хвильові рівняння (11.15) і (11.16) описують поширення електромагнітної хвилі вздовж осі *OX*.

Нагадаємо, що $E_x = E_z = 0$ і $H_x = H_y = 0$, так що $E_y = E$ і $H_z = H$. У рівняннях (11.15) і (11.16) збережені індекси y та z за E та H, щоб підкреслити ту обставину, що вектори $\vec{E}$ і $\vec{H}$ направленні вздовж взаємно перпендикулярних осей OY і OZ. Таким чином, $\vec{E} \perp \vec{H} \perp OX$, і це значить, що *електромагнітна хвиля є поперечною хвилею*.

В електродинаміці розглядають більш загальну задачу, без викладених нами обмежень, і досягають аналогічного результату: з рівнянь Максвелла випливає необхідність існування електромагнітних хвиль.

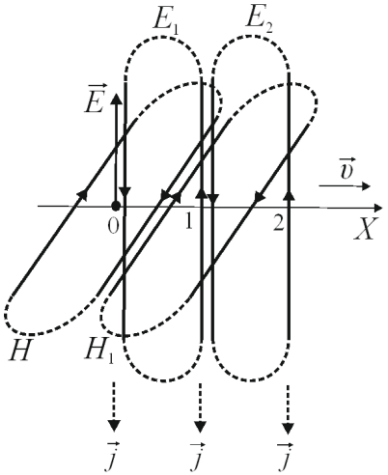


Рис. 11.2. Вільні електромагнітні хвилі

Розглянемо якісну картину утворення електромагнітних хвиль. Уявімо, що в якійсь точці O (рис. 11.2) всередині нескінченного непровідного середовища створено яким-небудь способом електричне поле $\vec{E}$. Якщо електричних зарядів, які підтримували б це поле, немає, то воно зникатиме, але спадне поле $\vec{E}$ за Максвеллом створює магнітне поле $\vec{H}$.

Через те що поле $\vec{E}$ спадає, то густина струму зміщення $\vec{j}_{\text{зм}} = \varepsilon\varepsilon_0 \frac{\partial \vec{E}}{\partial t}$ напрямлена протилежно $\vec{E}$ і лінії сил магнітного поля напрямлені за рухом стрілки годинника (коли дивитись згори за напрямком струму, рис. 11.2). Оскільки в середовищі немає постійних струмів, які підтримували б поле $\vec{H}$, то це останнє, у свою чергу, зникатиме і спричинюватиме вихрове електричне поле $\vec{E}_1$. Силові лінії цього поля будуть напрямлені проти руху стрілки годинника, як показано на рис. 11.2. Поле $\vec{E}_1$ знищить початкове поле $\vec{E}$ у точці *0*, але зате виникне в сусідній точці *1*. Зникаючи в точці *1*, електричне поле $\vec{E}_1$ спричинить виникнення магнітного поля $\vec{H}_1$, яке буде напрямлене, як і поле $\vec{H}$, за рухом стрілки годинника. Поле $\vec{H}_1$ знищить поле $\vec{H}$ і виникне в більш віддаленій точці. Зникаючи, воно спричинить вихрове електричне поле $\vec{E}_2$, яке знищить поле $\vec{E}_1$ у точці *1*, але виникне в точці *2*, і т. д.

Таким чином, замість початкового поля $\vec{E}$ ми дістанемо й електричне поле, і магнітне поле, які взаємно пов'язані одне з одним і які поширюються у просторі, тобто електромагнітну хвилю.

З рис. 11.2 також видно, що $\vec{E}$ перпендикулярний до $\vec{H}$, причому обидва ці вектори перпендикулярні до швидкості поширення хвилі $\vec{v}$. Усі три вектори пов'язані між собою правилом правого свердлика: якщо обертати свердлик із правою нарізкою так, щоб його рукоятка переміщувалась від вектора $\vec{E}$ до вектора $\vec{H}$, то напрямок поступального руху свердлика збігатиметься з напрямком $\vec{v}$.

Вище ми розглядали електромагнітні хвилі якісно, проте теорія Максвелла не тільки завбачила існування електромагнітних хвиль, а й дала змогу точно, у кількісній формі встановити всі основні їх властивості.

11.3. Властивості електромагнітних хвиль

Розглянемо основні властивості електромагнітних хвиль.

1. Порівняння (11.15) й (11.16) з (10.13) дає можливість стверджувати, що фазова швидкість електромагнітної хвилі

$$v = \frac{1}{\sqrt{\varepsilon\mu\varepsilon_0\mu_0}}. \qquad (11.17)$$

Для вакууму ($\varepsilon = 1,\ \mu = 1, \varepsilon_0 = 8{,}85 \cdot 10^{-12}$ Ф/м, $\mu_0 = 4\pi \cdot 10^{-7}$ Гн/м, і тому фазова швидкість (позначимо її літерою c) становить

$$c = \frac{1}{\sqrt{\varepsilon_0\mu_0}} = 3 \cdot 10^8 \text{ м/с}, \qquad (11.18)$$

що дорівнює швидкості світла у вакуумі. Цей важливий результат Максвелл отримав до експериментального відкриття електромагнітних хвиль. Це привело Максвелла до думки про електромагнітну природу світла. Для середовища з ε та μ фазова швидкість електромагнітних хвиль згідно з (11.17)

$$v = \frac{c}{\sqrt{\varepsilon\mu}} = \frac{c}{n}, \qquad (11.19)$$

де $n = \sqrt{\varepsilon\mu}$ – показник заломлення середовища. Це співвідношення дає зв'язок між оптичними n, електричними ε і магнітними μ властивостями середовища. Слід мати на увазі, що ε та μ залежать від частоти коливань $\vec{E}$ та $\vec{H}$.

2. Розв'язками рівнянь (11.15) і (11.16) є функції

$$E = E_m \cos(\omega t - kx + \alpha_1) \qquad (11.20)$$

та

$$H = H_m \cos(\omega t - kx + \alpha_2). \qquad (11.21)$$

У цих формулах $\omega -$ циклічна частота хвилі, k – хвильове число, яке дорівнює ω/v, α_1 та α_2 – початкові фази коливань у точках з координатами $x = 0$, E_m та H_m – амплітуди хвилі. Підставимо (11.20) і (11.21) у рівняння (11.13) і (11.14):

$$\varepsilon\varepsilon_0\omega E_m \sin(\omega t - kx + \alpha_1) = \\ = kH_m \sin(\omega t - kx + \alpha_2),$$
$$\mu\mu_0\omega H_m \sin(\omega t - kx + \alpha_2) = \\ = kE_m \sin(\omega t - kx + \alpha_1).$$

Щоб рівняння (11.13) і (11.14) задовольнялись розв'язками (11.20) і (11.21), потрібна рівність початкових фаз α_1 й α_2. Крім того, мають виконуватись співвідношення

$$kE_m = \mu\mu_0\omega H_m,$$
$$\varepsilon\varepsilon_0\omega E_m = kH_m.$$

Перемноживши ці дві рівності, знайдемо:

$$\varepsilon\varepsilon_0 E_m^2 = \mu\mu_0 H_m^2. \quad (11.22)$$

Таким чином, коливання електричного і магнітного векторів в електромагнітній хвилі відбувається з однаковою фазою ($\alpha_1 = \alpha_2$), а амплітуди цих векторів пов'язані співвідношенням

$$E_m\sqrt{\varepsilon\varepsilon_0} = H_m\sqrt{\mu\mu_0}. \quad (11.23)$$

Оскільки вектори $\vec{E}$ і $\vec{H}$ коливаються у фазі, то співвідношення (11.23) справедливе не тільки для амплітудних значень E_m і H_m, а й для миттєвих значень E і H в якийсь момент часу, тобто

$$E\sqrt{\varepsilon\varepsilon_0} = H\sqrt{\mu\mu_0}. \quad (11.24)$$

Помноживши рівняння (11.20) на орт осі OY ($E\vec{j} = \vec{E}$), а рівняння (11.21) на орт осі OZ ($H\vec{k} = \vec{H}$), отримаємо рівняння плоскої електромагнітної хвилі у векторному вигляді:

$$\vec{E} = \vec{E}_m \cos(\omega t - kx), \\ \vec{H} = \vec{H}_m \cos(\omega t - kx) \quad (11.25)$$

(ми прирівняли $\alpha_1 = \alpha_2 = 0$).

Косинусоїдальну або синусоїдальну електромагнітну хвилю називають монохроматичною хвилею. У кожній точці електромагнітного поля монохроматичної хвилі проєкції векторів $\vec{E}$ та $\vec{H}$ на осі координат інерційної системи відліку здійснюють гармонічні коливання однакової лінійної частоти $\nu = \omega/(2\pi)$, яку називають частотою хвилі.

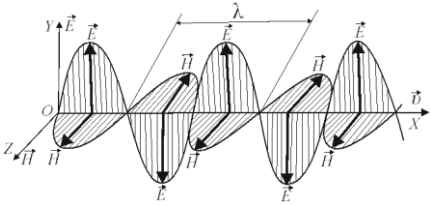


Рис. 11.3. Плоска електромагнітна хвиля в якмісь фіксований момент часу

На рис. 11.3 зображено вектори $\vec{E}$ і $\vec{H}$ поля плоскої (вектор $\vec{E}$ коливається у площині *YOX*) монохроматичної хвилі в один і той самий момент часу. У фіксованій точці простору вектори $\vec{E}$ та $\vec{H}$ змінюються з часом за гармонічним законом. Вони одночасно збільшуються від нуля, потім через ¼*T* досягають найбільшого значення. Ще через ¼*T* обидва вектори одночасно дорівнюють нулю. Потім знову обидва вектори досягають максимального значення, але протилежного напрямку, ніж півперіоду

тому. Через час, який дорівнює періоду коливання векторів, вектори $\vec{E}$ та $\vec{H}$ знову стають нульовими. Така зміна векторів $\vec{E}$ та $\vec{H}$ відбувається у всіх точках простору, але зі зсувом за фазою, що визначається відстанню між точками, яка відраховується уздовж осі *OX*.

Площину, яка проходить через вектор $\vec{E}$ і вектор швидкості $\vec{v}$, називають *площиною поляризації хвилі.* Довжина хвилі λ, період T, частота ν і швидкість поширення електромагнітної хвилі пов'язані між собою співвідношеннями

$$\lambda = vT = v/\nu. \qquad (11.26)$$

Якщо плоска електромагнітна біжуча хвиля поширюється в додатному напрямку осі *OX*, то її можна подати парою рівнянь (11.25). Тоді відбитій біжучій хвилі, яка поширюється в оберненому напрямку, від'ємному напрямку осі *OX*, відповідає одна пара із сукупності рівнянь:

$$\begin{aligned} \vec{E} &= \pm\vec{E}_m \cos(\omega t + kx), \\ \vec{H} &= \mp\vec{H}_m \cos(\omega t + kx). \end{aligned} \qquad (11.27)$$

У результаті попарного додавання (11.25) і (11.27) та відповідного математичного перетворення отримуємо рівняння стоячої електромагнітної хвилі:

$$E = 2E_m \cos kx \cos \omega t, \qquad (11.28)$$

$$H = 2H_m \sin kx \sin \omega t. \qquad (11.29)$$

Отже, (11.28) – рівняння електричної, (11.29) – рівняння магнітної стоячої хвилі. Тут $|2E_m \cos kx|$ і $|2H_m \sin kx|$ – амплітуди електричних і магнітних коливань відповідно. Оскільки максимальному значенню синуса відповідає нульове значення косинуса, то максимальному (нульовому) значенню амплітуди електричних коливань відповідає нульове (максимальне) значення амплітуди магнітних коливань, інакше кажучи, пучності (вузлу) електричної стоячої хвилі відповідає вузол (пучність) магнітної стоячої хвилі. З рівнянь (11.28) і (11.29) випливає, що коливання електричної компоненти хвилі зсунуті на $\pi/2$ за фазою відносно коливань магнітної компоненти хвилі ($\cos \omega t$ відносно $\sin \omega t$).

11.4. Енергія електромагнітних хвиль. Потік енергії. Вектор Пойнтинга

Електромагнітне поле має енергію, тому поширення електромагнітних хвиль пов'язане з перенесенням енергії у полі, подібно до того, як поширення пружних хвиль у речовині пов'язане з перенесенням механічної енергії.

Об'ємна густина енергії електромагнітної хвилі складається з об'ємних густин електричного і магнітного полів:

$$w = w_E + w_H = \frac{\varepsilon\varepsilon_0 E^2}{2} + \frac{\mu\mu_0 H^2}{2}.$$

Врахувавши (11.24), отримаємо, що густини енергії електричного і магнітного полів у кожний момент часу однакові, тобто $w_E = w_H$, тому

$$w = 2w_E = \varepsilon\varepsilon_0 E^2 = \\ = \sqrt{\varepsilon\varepsilon_0 \mu\mu_0} EH = \frac{1}{v} EH. \quad (11.30)$$

У випадку плоско поляризованої монохроматичної хвилі, що поширюється (рис. 11.3) вздовж додатного напрямку осі *OX*, напруженість поля

$$E = E_m \cos(\omega t - kx).$$

Відповідно, об'ємна густина енергії цієї хвилі $w = \varepsilon\varepsilon_0 E_m^2 \cos^2(\omega t - kx)$. Значення w у кожній точці поля періодично змінюється з частотою 2ω у межах від 0 до $w_{max} = \varepsilon\varepsilon_0 E_m^2$. Середнє значення w за період π/ω пропорційне квадрату амплітуди напруженості поля:

$$\langle w \rangle = \frac{\omega}{\pi} \int_0^{\pi/\omega} w dt = \\ = \varepsilon\varepsilon_0 E_m^2 \langle \cos^2(\omega t - kx) \rangle = \frac{1}{2} \varepsilon\varepsilon_0 E_m^2. \quad (11.31)$$

Помноживши густину енергії w на швидкість v поширення хвилі в середовищі, отримуємо модуль густини потоку енергії.

Модуль густини потоку енергії чисельно дорівнює енергії, яку переносить хвиля за одиницю часу через одиницю площі поверхні, що розміщена перпендикулярно до напрямку поширення хвилі:

$$\Pi = wv = EH.$$

Оскільки вектори $\vec{E}$ та $\vec{H}$ взаємно перпендикулярні й утворюють з напрямком поширення хвилі правогвинтову систему, то напрямок вектора $\vec{E} \times \vec{H}$ збігаються з напрямком перенесення енергії, а модуль цього вектора дорівнює *EH*. Отже, вектор густини потоку енергії електромагнітної хвилі, який називають **вектором Пойнтинга**,

$$\vec{\Pi} = \vec{E} \times \vec{H}. \quad (11.32)$$

Потік Φ електромагнітної енергії через деяку поверхню *S* можна знайти за допомогою інтегрування:

$$\Phi = \int_S \vec{\Pi} d\vec{S}. \quad (11.33)$$

Інтенсивність електромагнітної хвилі *I* дорівнює модулю середнього значення вектора Пойнтинга за проміжок часу, який дорівнює періоду *T* повного коливання:

$$I = \left|\langle \vec{\Pi} \rangle\right| = \left|\langle \vec{E} \times \vec{H} \rangle\right| = \frac{1}{T} \left| \int_0^T \vec{E} \times \vec{H} dt \right|,$$

або

$$I = \langle w \rangle v. \quad (11.34)$$

Таким чином, інтенсивність плоско поляризованої монохроматичної

біжучої хвилі прямо пропорційна квадрату амплітуди E_m коливань вектора $\vec{E}$ поля хвилі:

$$I = \frac{1}{2}\sqrt{\frac{\varepsilon\varepsilon_0}{\mu\mu_0}}E_m^2. \qquad (11.35)$$

11.5. Генерування та експериментальне дослідження електромагнітних хвиль (радіохвиль)

Із попередніх параграфів відомо, що для отримання електромагнітної хвилі треба створити у просторі досить швидко змінне електричне поле. Очевидно, що для цієї мети електричні коливальні контури із зосередженими ємністю й індуктивністю (закриті контури), які були розглянуті у розд. 9, непридатні. У таких випадках усе електричне поле зосереджене у вузькому зазорі конденсатора, усе магнітне поле – всередині індуктивності, а в навколишньому просторі електричне поле фактично дорівнює нулю (рис. 11.4, *а*).

Делокалізацію змінних електричного й магнітного полів, перенесення їх з частин простору, які обмежені обкладками конденсатора і котушкою, можна здійснити, розсуваючи обкладки конденсатора і зменшуючи їх площу, а також розсуваючи витки котушки і зменшуючи їх кількість (рис. 11.4, *б*, *в*).

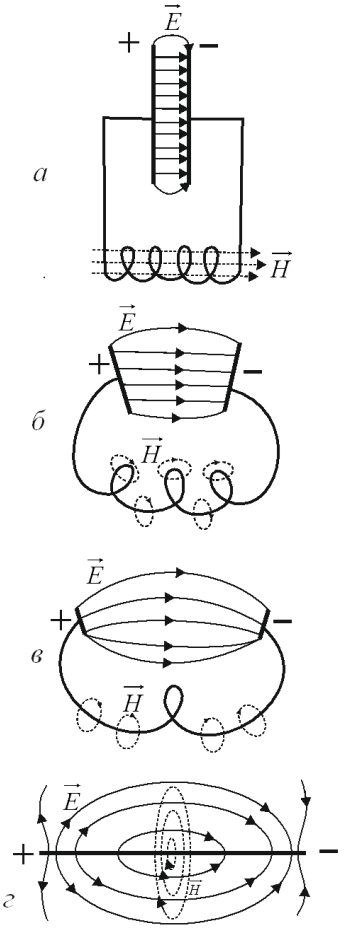


Рис. 11.4. Генерація електромагнітних хвиль

При цьому одночасно зменшуються ємність та індуктивність контура, а отже, збільшується власна частота коливального контура. Зрештою, отримуємо прямолінійний відрізок проводу (металічного стрижня), як показано на рис. 11.4, *г* – це відкритий

коливальний контур (відкритий вібратор). Як видно з рисунка, змінним електричному і магнітному полям (електромагнітному полю) відкритого вібратора повністю забезпечений вихід в навколишній простір. Такий вібратор винайшов і вперше застосував для генерування електромагнітних хвиль (для перевірки теорії Максвелла) у 1887–1888 рр. німецький фізик Г. Герц.

Для вивчення електромагнітних хвиль Герц використав власні електричні коливання відкритого вібратора (диполя), який складався із двох однакових металевих стрижнів *вв* (рис. 11.5), розділених іскровим проміжком.

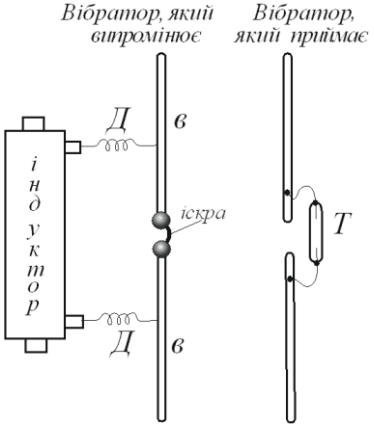


Рис. 11.5. Відкритий вібратор Герца

Обидві половини вібратора заряджались від джерела високої напруги індуктора. Коли різниця потенціалів досягала пробійного значення, в розряднику проскакувала іскра, яка замикала обидві половини вібратора, й у вібраторі виникали згасаючі електричні коливання високої частоти. Щоб швидкозмінні струми не переходили з вібратора в котушку, між вібратором і котушкою ввімкнено дроселі *Д*, реактивний опір яких дорівнює ωL.

Оскільки зміна контура від вигляду, зображеного на рис. 11.4, *а*, до вигляду, зображеного на рис. 11.4, *г*, спричиняє дуже сильне зменшення ємності відкритого контура, то з метою деякого збільшення ємності кінці половинок стрижнів іскрового проміжку мали сферичні потовщення (рис. 11.5).

Індуктор – котушка із залізним осердям і двома обмотками. За допомогою механічного переривача в первинній обмотці створювався пульсуючий струм з частотою його розриву 10^{-3} ... 10^{-4} $с^{-1}$. У вторинній обмотці індукувалась змінна е.р.с. (до 10^4 В) тієї ж частоти, яка прикладалась до двох половинок вібратора. Коли напруга між половинками вібратора досягала пробивної величини, між ними проскакувала іскра, яка закорочувала обидві половинки вібратора і в ньому виникали високочастотні власні згасаючі електромагнітні коливання, частота яких визначалась величинами його ємності та

індуктивності. Герц отримував з різними вібраторами коливання, частота яких була порядку $10^7 \dots 10^8$ Гц та електромагнітні хвилі з $\lambda = (3 \dots 30)$ м. Половинки вібратора зазнавали багатократної перезарядки, після якої електромагнітні коливання повністю згасали.

Згасання коливань обумовлене тим, що енергія, яку отримав вібратор під час його зарядки, витрачалась на виділення тепла і на випромінювання хвиль. Унаслідок згасання коливань і погасання іскри контур розмикається, вібратор заряджається від індуктора до наступної іскри, процес багаторазової перезарядки половинок вібратора повторюється і т. д. Такі періодичні порції багатократних перезарядок половин вібратора (з періодом роботи переривача індуктора) супроводжується випромінюванням порції (***цуга***) електромагнітних хвиль у навколишній простір.

Період роботи переривача індуктора набагато більший, ніж період електромагнітних коливань вібратора, тому й використовують дроселі *Д* (рис. 11.5) для захисту індуктора від високочастотних струмів. Для виявлення електромагнітних хвиль Герц застосовував вібратори різної форми. Найпростішим і найзручнішим є прямий відкритий вібратор, який за формою і розмірами тотожний випромінювальному вібратору (рис. 11.5). Під дією змінного електричного поля прохідної електромагнітної хвилі електрони всередині вібратора починають здійснювати вимушені коливання, унаслідок чого у вібраторі виникає швидкозмінний струм, а між обома половинами вібратора – змінна напруга. Якщо довжина приймального вібратора дорівнює довжині випромінювального вібратора і дорівнює $\lambda/2$, то власні частоти обох вібраторів збігаються й електричні коливання у приймальному вібраторі внаслідок резонансу підсилюються. Появу змінної напруги Герц виявляв за виникненням електричної іскорки у мікроскопічному зазорі всередині вібратора або за світінням мініатюрної газорозрядної трубки *Т*, увімкненої між обома половинками вібратора (рис. 11.5).

Тепер для випромінювання електромагнітних хвиль майже завжди використовують генератори, які дають змогу діставати електричні коливання фактично будь-якої потужності й до того ж правильної синусоїдальної форми. Щоб збудити коливання у вібраторі, між обома його половинками *вв* можна ввімкнути один або кілька витків зв'язку (рис. 11.6) і помістити їх поблизу котушки індуктивності *К* генератора Г (магнітний зв'язок). Є й інші способи зв'язку між

вібраторами і генераторами. Щоб підсилити коливання у вібраторі, використовують явище резонансу, для чого частоту генератора роблять такою, щоб вона дорівнювала одній із власних частот вібратора, зазвичай частоті основного його коливання.

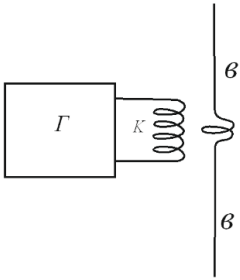


Рис. 11.6. Схема сучасного способу генерування електромагнітних хвиль

Герц продовжив також дослідження властивостей електромагнітних хвиль. За допомогою великих металічних дзеркал та асфальтової призми (розміром більшої за 1 м і масою 1200 кг) він здійснював відбивання і заломлення електромагнітних хвиль і виявив, що обидва ці явища підпорядковуються законам, які встановлені в оптиці для світлових хвиль. Відбиваючи біжучу плоску хвилю за допомогою металічного дзеркала в оберненому напрямку, Герц отримав стоячу хвилю. Відстань між вузлами і пучностями хвилі дозволяла визначити довжину хвилі λ. Помноживши λ на частоту коливань вібратора ν, можна було знайти швидкість електромагнітних хвиль, яка виявилась близькою до швидкості світла *c*. Розміщуючи на шляху хвиль решітку із паралельних одна одній мідних дротин, Герц виявив, що під час обертання решітки навколо променя інтенсивність хвиль, які пройшли через решітку, сильно змінюється. Якщо дротини, які утворюють решітку, були перпендикулярні до вектора $\vec{E}$, хвиля проходила через решітку без завад. У разі розміщення дротин паралельно $\vec{E}$ хвиля через решітку не проходила. Так було доведено поперечність електромагнітних хвиль.

Досліди Герца відіграли вирішальну роль для підтвердження і визнання теорії Максвелла. Експериментальне відкриття Герцом електромагнітних хвиль лягло в основу великого винаходу бездротової телефонії і телеграфії (радіо) у 1895–1896 рр.

11.6. Випромінювання елементарного диполя

З-поміж різних електричних систем, які випромінюють електромагнітні хвилі, особливо важливе значення має електричний диполь. У підрозд. 11.5 ми вже розглядали застосування диполя (вібратора) для випромінювання

електромагнітних хвиль. Утім, там йшлося про *півхвильовий* диполь, довжина якого дорівнює половині довжини хвилі. Тепер розглянемо диполь, довжина якого мала порівняно з довжиною хвилі (елементарний диполь). Найпростішим прикладом елементарного диполя є дві металеві кулі, які заряджаються від якого-небудь генератора електричних коливань, якщо відстань між кулями $l \ll \lambda$ (рис. 11.7). Момент такого диполя $p = ql$. Якщо генератор видає гармонічні коливання, то $q = q_0 \cos \omega t$, і момент диполя змінюється з часом також за гармонічним законом:

$$p = p_0 \cos \omega t, \qquad (11.36)$$

де $p_0 = q_0 l$ є амплітудою електричного момента диполя. Її можна подати також через амплітуду величини струму I_0. Справді, величина струму в диполі $I = \frac{dq}{dt} = q_0 \omega \sin \omega t$, а амплітуда струму $I_0 = q_0 \omega$, тому $q_0 = I_0/\omega$, а отже,

$$p_0 = \frac{l}{\omega} I_0. \qquad (11.37)$$

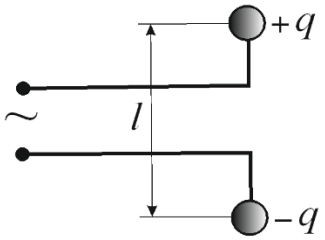


Рис. 11.7. Елементарний диполь

З елементарними диполями нам доводиться зустрічатися досить часто. Антени радіостанцій разом з поверхнею землі, на якій індукуються заряди протилежного знака, в багатьох випадках можна розглядати як елементарні диполі. Утім, найважливішим прикладом елементарних диполів є електрони всередині атомів. Коловий та еліптичний рухи електронів, які обертаються навколо позитивного ядра, можна розкласти на два прямолінійні гармонічні коливання, але електрон, який здійснює прямолінійні гармонічні коливання, разом з позитивним ядром (яке не бере участі у випромінюванні) можна розглядати як диполь, момент якого змінюється згідно з формулою (11.36). Через те, що довжини хвиль, які випромінює атом (для видимого світла округлено $5 \cdot 10^{-7}$ м), набагато більші за розміри атомів ($\sim 10^{-10}$ м), то диполі, які розглядаються, можна з великою точністю вважати елементарними.

Розглянемо тепер, який характер мають електромагнітні хвилі, випромінювані диполем. При цьому ми не подаватимемо строгого виведення виразів для електромагнітного поля з рівнянь Максвелла, а обмежимося якісними результатами.

Характер електромагнітного поля диполя істотно залежить від того, близько чи далеко розміщено точку, яка розглядається. Якщо відстань r від центра диполя до цієї точки мала порівняно з довжиною хвилі ($r \ll \lambda$), то справедливі ті самі формули, що і для постійних електричного і магнітного полів. Електричне поле диполя виражають формулами, воно спадає з відстанню пропорційно $1/r^3$ (див. підрозд. 2.2). Магнітне поле диполя виражають тією самою формулою (5.8) (чи скалярною формулою (5.9)), що й поле елемента струму, і пропорціональне $1/r^2$. Такі поля можна вважати квазістаціонарними. З ними не може бути пов'язане випромінювання енергії. Потік енергії через одиничну площину за одиницю часу визначається складовою вектора Пойтинга $\vec{\Pi} = \vec{E} \times \vec{H}$ (формула (11.32)), яка перпендикулярна до цієї площини. У квазістаціонарних полях $\vec{E}$ та $\vec{H}$ зсунуті за фазою на $\pi/2$ (як у стоячих хвилях), тому вектор Пойтинга, що змінюється з частотою 2ω, в середньому за період коливань точно дорівнює нулю. Відмінність $\vec{\Pi}$ від нуля може бути зумовлена тільки полями $\vec{E}$ та $\vec{H}$, які коливаються з однаковою фазою (як у біжучих хвилях, формула (11.25)) і зменшуються пропорційно $1/r$. ($\Pi \sim 1/r^2$) Це безпосередньо випливає із закону збереження енергії, оскільки, якщо немає втрат у середовищі, повний потік енергії у просторі не має змінюватись із відстанню. Оскільки площа, яка охоплює диполь (площа сфери) збільшується як r^2, то потрібно, щоб Π був пропорційним $1/r^2$.

Таким чином, поле у ближній зоні ($r \ll \lambda$) диполя (зоні індукції) служить для формування біжучих складових полів, які відповідають за випромінювання. На великих відстанях від диполя ($r \gg \lambda$) закон зміни полів зовсім інший. Ця *хвильова* область становить основний інтерес, і тому ми розглянемо її докладніше. З'ясуємо, який вигляд має хвильовий фронт електромагнітної хвилі диполя. Оскільки електромагнітне збурення поширюється на всі боки від диполя з однаковою швидкістю *c* (ми припускаємо, що диполь міститься у вакуумі), то час проходження хвилі до всіх точок, віддалених від диполя на одну й ту саму відстань *r*, однаковий, тому в усіх точках сфери, центр якої збігається з диполем, фаза коливань однакова, тобто ми маємо *сферичний* хвильовий фронт, а отже хвиля, яка випромінюється диполем, є сферичною чи кульовою хвилею. Через те, що електричне поле E у хвилі перпендику-

лярне до напрямку поширення, то вектор $\vec{E}$ в різних точках перпендикулярний до радіусів-векторів (рис. 11.8, *а*). Це поле періодично змінюється, і тому, переміщуючись уздовж радіуса-вектора, ми знаходитимемо поля взаємно протилежного напрямку (рис. 11.8, *б*). З'єднавши на рис. 11.8, *б* стрілки суцільною лінією, ми дістанемо одну з електричних силових ліній (рис. 11.8, *в*).

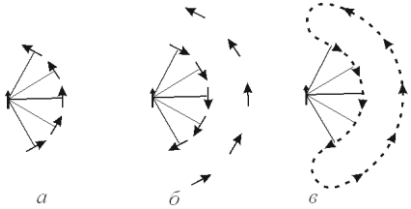


Рис. 11.8. Форма електричних силових ліній поля диполя, який випромінює електромагнітні хвилі

На рис. 11.9 зображено картину послідовного відокремлення силових ліній створюваного електричного поля $\vec{E}$. У першій чверті періоду T коливань ($t = T/4$) виникає квазістатична частина поля (рис. 11.9, *а*), яка за $t = T/2$ перетворюється на нуль, але від поля «відриваються» замкнуті самі на себе силові лінії поля $\vec{E}$ і «зчеплені» з ними кільцеві ортогональні магнітні лінії (на рис. 11.9, *б* не показані). Разом вони утворюють автономну напівхвильову тороїдальну (завдяки аксіальній симетрії) комірку сферично розбіжної хвилі, яка несе електромагнітну енергію (рис. 11.9, *в*, *г*).

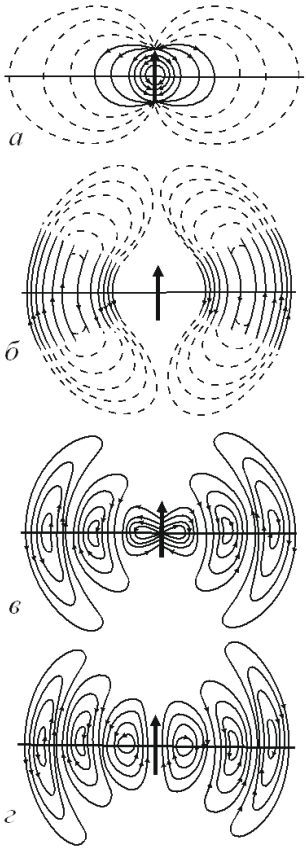


Рис. 11.9. Послідовно відокремлені силові лінії: а – електричні силові лінії біля електричного диполя (за постійного заряду); б–г – силові лінії, які відділились від диполя: б – через ½ періоду (T/2) після підключенння генератора (заряду на диполі немає); в – через (3/2)T (масштаб змінений); г – через (7/4)T (масштаб змінений)

Напрямок магнітного поля $\vec{H}$ у кожній точці перпендикулярний до $\vec{E}$ і до

напрямку поширення, тому магнітними силовими лініями будуть концентричні кола, які лежать у площинах, перпендикулярних до диполя, і мають центр на осі диполя.

Розглянемо докладніше, який вигляд має електромагнітне поле диполя у хвильовій області.

Відомо, що заряд, який рухається з постійною швидкістю, спричиняє тільки постійне магнітне поле *H*. Водночас згідно з основним положенням теорії Максвелла для виникнення вихрового електричного поля потрібно, щоб $\frac{dH}{dt}$ було відмінне від нуля, тому для випромінювання потрібно, щоб заряд рухався з певним *прискоренням* $\dot{v} = \ddot{l}$ (крапка означає диференціювання за часом) або, інакше, щоб друга похідна за часом від моменту диполя $\ddot{p}$ = $(\ddot{q}\vec{l})$ не дорівнювала нулю. Відповідно до цього точний розрахунок показує, що електричне поле *E* хвилі (а отже, і магнітне поле *H*) пропорційне $\ddot{p}(t - r/c)$. Аргумент $(t - r/c)$ показує, що коливання поля у точці, яка лежить на відстані *r* від диполя, запізнюється відносно коливань $\ddot{p}$ у часі на r/c, так що *E*(*t*) і *H*(*t*) у точці, яка розглядається, в момент *t* визначаються значенням $\ddot{p}$ у більш ранній момент часу $(t - r/c)$. Раніше вже було показано, що виходячи із закону збереження енергії *E* та *H* пропорційні $1/r$.

Величина полів *E* та *H* залежить ще від напрямку випромінювання, тобто від кута θ, утвореного радіусом-вектором $\vec{r}$ з віссю диполя (рис. 11.10). Уже згадувалось, що поблизу диполя його магнітне поле виражається тією самою формулою (5.9), що й для елемента струму. Це магнітне поле пропорційне $\sin\theta$ і дорівнює нулю в будь-якій точці, яка лежить на продовжені осі диполя. Так само і напруженість полів *E* та *H* у хвильовій зоні пропорційна $\sin\theta$.

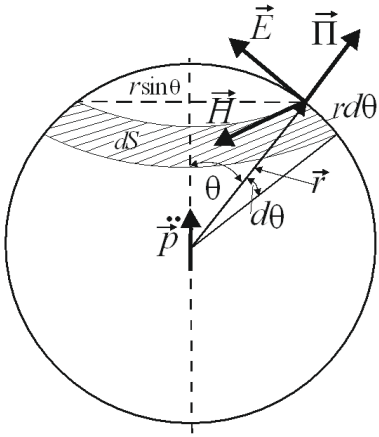


Рис 11.10. Електромагнітне поле у сферичній хвилі, яку випромінює диполь

Точний розрахунок дає такий результат:

$$E = \frac{1}{4\pi\varepsilon_0}\frac{\ddot{p}(t-r/c)\sin\theta}{c^2 r},\ H = \sqrt{\frac{\varepsilon_0}{\mu_0}}E. \tag{11.38}$$

Тут $c = 3 \cdot 10^8$ м/с – швидкість поширення електромагнітних хвиль у вакуумі. Усі інші величини в (11.38) мають розмірність в СІ.

Обидва поля $\vec{E}$ та $\vec{H}$ напрямлені перпендикулярно до напрямку поширення, тобто до напрямку $\vec{r}$, і пов'язані з ним правилом правого свердлика (рис. 11.10). Зауважимо, що формула (11.38) справедлива тільки для таких рухів зарядів, швидкість яких v мала порівняно зі швидкістю світла c (нерелятивістський випадок, $\beta = v/c \ll$ $\ll 1$). В окремому випадку, коли p змінюється за гармонічним законом (11.36),

$$\ddot{p}(t - r/c) = -p_0\,\omega^2 \cos\omega(t - r/c) = \\ = -p_0\,\omega^2 \cos(\omega t - kr),$$

і вираз (11.38) набуває вигляду

$$E = -\frac{1}{4\pi\varepsilon_0}\frac{p_0\,\omega^2 \sin\theta}{c^2 r}\cos(\omega t - kr). \quad (11.39)$$

Енергія, випромінювана диполем. Значенням вектора потоку енергії в якій-небудь точці з координатами r, θ є:

$$\Pi = EH = \frac{{p_0}^2\omega^4 \sin^2\theta}{16\pi^2\varepsilon_0 c^3 r^2}\cos^2(\omega t - kr).$$

Інтерес становить не миттєве значення П, а його середнє значення $\langle\Pi\rangle$ за досить великий проміжок часу, тобто інтенсивність випромінювання. Оскільки $\langle\cos^2(\omega t - kr)\rangle = 1/2$, то

$$\langle\Pi\rangle = I = \frac{1}{32\pi^2\varepsilon_0 c^3}\frac{{p_0}^2\omega^4 \sin^2\theta}{r^2}. \quad (11.40)$$

Інтенсивність випромінювання пропорційна частоті коливань ω у четвертому степені. Вона залежить також від напрямку випромінювання (пропорційна $\sin^2\theta$). Залежність інтенсивності випромінювання диполя від напрямку подано на рис. 11.11 («діаграма направленості» випромінювання). Таким чином, у напрямку своєї осі диполь нічого не випромінює. Навпаки, у напрямках, перпендикулярних до осі диполя, інтенсивність випромінювання найбільша. Цю діаграму будують так, щоб довжина відрізка, який відсікається нею на промені, що проведений із центра диполя, давала інтенсивність випромінювання під кутом θ.

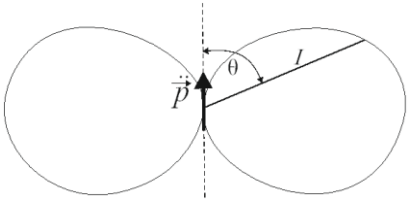


Рис. 11.11. Діаграма направленості випромінювання елементарного диполя

Нарешті обчислимо повну енергію (потужність) P, яку випромінює диполь в усіх напрямках за одиницю часу. Через те, що $\langle\Pi\rangle$ є енергією, яка проходить через одиницю поверхні за 1 с, то

$$P = \int_S \langle \vec{\Pi} \rangle d\vec{S} = \int_S \langle \vec{\Pi}\vec{n} \rangle dS =$$

$$= \int_S \langle \Pi_n \rangle dS,$$

де інтегрування виконують по довільній поверхні S, яка охоплює диполь. Вибираючи сферу, радіус якої r, із центром у точці розміщення диполя і користуючись сферичними координатами r, θ (рис. 11.10), отримаємо:

$$dS = 2\pi r^2 \sin\theta d\theta,$$

$$\langle \Pi_n \rangle dS = \langle \Pi \rangle dS = \frac{{p_0}^2 \omega^4}{16\pi^2 \varepsilon_0 c^3} \sin^3 \theta \mathrm{d}\theta,$$

$$P = \frac{{p_0}^2 \omega^4}{16\pi^2 \varepsilon_0 c^3} \int_0^{\pi} \sin^3 \theta d\theta = \frac{1}{12\pi} \frac{{p_0}^2 \omega^4}{\varepsilon_0 c^3}. \quad (11.41)$$

Написаний вираз можна подати і в іншому вигляді. Для цього виразимо амплітуду коливань моменту p_0 через амплітуду струму в диполі I_0 згідно з (11.37). Тоді отримаємо:

$$P = \frac{1}{12\pi} \frac{l^2 \omega^2}{\varepsilon_0 c^3} {I_0}^2, \quad (11.42)$$

де l – довжина диполя. Цей вираз подібний до виразу для потужності, яку виділяє змінний струм з опором R:

$$P = \frac{1}{2} R {I_0}^2,$$

тому величину

$$R_{\text{випром}} = \frac{1}{6\pi} \frac{l^2 \omega^2}{\varepsilon_0 c^3} \quad (11.43)$$

називають *опором випромінювання диполя*. Це поняття застосовують в радіотехніці до випромінювання антен передавальних станцій.

Контрольні запитання

1. Напишіть та поясніть рівняння плоскої біжучої електромагнітної хвилі.

2. Як визначають фазову швидкість електромагнітної хвилі у вакуумі та середовищі?

3. Яке співвідношення між напруженістю електричного та напруженістю магнітного поля в електромагнітної хвилі?

4. Як визначають густину потоку електромагнітної енергії – вектор Пойтинга? Чому вектор $\vec{E}$ електромагнітної хвилі називають світловим?

5. Якою характеристикою електромагнітної хвилі визначається її інтенсивність?

6. Поясніть, чому стоячі хвилі не переносять енергію?

7. Нарисуйте й поясніть полярну діаграму випромінювання електричного диполя у хвильовій зоні. Які можна зробити висновки?

8. Яка потрібна умова, щоб електричний диполь та електричний заряд випромінювали електромагнітні хвилі?

Розділ 12. Інтерференція електромагнітних хвиль

12.1. Основні відомості про хвилі оптичного діапазону. Інтенсивність електромагнітних хвиль

У *вузькому розумінні* світло – те саме, що й видиме випромінювання, тобто електромагнітні хвилі в інтервалі частот, які сприймаються людським оком ($7{,}5\cdot10^{14}$ … $4{,}0\cdot10^{14}$ Гц), що відповідає довжинам хвиль у вакуумі від ~ 0,4 до ~ 0,75 мкм. Світло високої інтенсивності око сприймає в дещо більш широкому діапазоні. Світлові хвилі різних частот людина сприймає як різні кольори.

Світло в *широкому розумінні* – синонім оптичного випромінювання, що включає, крім видимого випромінювання, ультрафіолетову (УФ) та інфрачервону (ІЧ) області спектра.

Світло – складне явище: в одних випадках воно поводить себе як електромагнітна хвиля, а в інших – як потік особливих частинок (фотонів). У цьому розділі розглянемо хвильову оптику, тобто коло явищ, в основі яких лежить хвильова природа світла.

В електромагнітній хвилі коливаються вектори $\vec{E}$ та $\vec{H}$. Як показують дослідження, фізіологічну, фотохімічну, фотоелектричну та інші види дії світла обумовлені коливаннями електричного вектора $\vec{E}$. Відповідно до цього надалі будемо говорити про світловий вектор, розуміючи під ним вектор напруженості електричного поля $\vec{E}$. Про магнітний вектор світлової хвилі ми згадувати майже не будемо. Модуль амплітуди світлового вектора будемо позначати літерою E_m. Відповідно, зміни з часом й у просторі проєкції світлового вектора на напрямок, вздовж якого він коливається, описують рівнянням (11.20):

$$\vec{E} = \vec{E}_m \cos(\omega t - kr + \alpha). \quad (12.1)$$

Тут $k = 2\pi/\lambda$ – *хвильове* число; r – відстань, яка відраховується вздовж напрямку поширення світлової хвилі. Для плоскої хвилі, що поширюється в середовищі, яке не поглинає світло, $E_m = \text{const}$, для сферичної хвилі E_m зменшується з відстанню пропорційно $1/r$.

Згідно з (11.19) відношення швидкості світлової хвилі у вакуумі до фазової швидкості v в деякому середовищі називають абсолютним показником заломлення цього середовища. Таким чином,

$$n = c/v. \quad (12.2)$$

Із формули (11.19) видно, що $n = \sqrt{\varepsilon\mu}$. Для більшості прозорих речовин μ фактично не відрізняється від одиниці, тому можна вважати, що

$$n = \sqrt{\varepsilon}. \qquad (12.3)$$

Формула (12.3) пов'язує оптичні властивості речовин з їх електричними властивостями. На перший погляд може видатись, що ця формула неправильна. Наприклад, для води $\varepsilon = 81$, а n = 1,33. Утім, треба мати на увазі, що значення $\varepsilon = 81$ отримано з електростатичних вимірювань. У швидкозмінних електричних полях значення ε отримують іншим, причому воно залежить від частоти коливань поля. Цим пояснюється дисперсія світла, тобто залежність показника заломлення (або швидкості світла) від частоти (або довжини хвилі). Підстановка у формулу (12.3) значення ε, яке отримане для відповідної частоти, дає правильне значення n.

Значення показника заломлення характеризує *оптичну густину* середовища. Середовище з більшим n називають оптично більш густим, ніж середовище з меншим n. Відповідно, середовище з меншим n називають оптично менш густим, ніж середовище з більшим n.

Довжини хвиль видимого світла $\lambda_0 = 0{,}40 \ldots 0{,}76$ мкм належать до світлових хвиль у вакуумі. У речовині довжини світлових хвиль будуть іншими. У випадку коливань з частотою ν довжина хвилі у вакуумі буде $\lambda_0 = \frac{c}{\nu}$.

В середовищі, в якому фазова швидкість світлової хвилі $v = \frac{c}{n}$, довжина хвилі буде $\lambda = \frac{v}{\nu} = \frac{c}{\nu n} = \lambda_0/n$. Таким чином, довжина світлової хвилі в середовищі з показником заломлення n зв'язана з довжиною хвилі у вакуумі співвідношенням

$$\lambda = \lambda_0/n\,. \qquad (12.4)$$

Інтенсивність світла I згідно з (11.35) у розглядуваній точці простору буде:

$$I = \frac{1}{2}\sqrt{\frac{\varepsilon_0}{\mu_0}}\, nE_m^2, \qquad (12.5)$$

тобто інтенсивність світла пропорційна показнику заломлення середовища і квадрату амплітуди світлової хвилі.

Зауважимо, що під час проходження світла в однорідному середовищі можна вважати, що інтенсивність пропорційна квадрату амплітуди світлової хвилі:

$$I \sim E_m^2. \qquad (12.6)$$

Утім, у випадку проходження світла через межу поділу середовищ вираз для інтенсивності, який не враховує множник n, приводить до незбереження світлового потоку.

Лінії, вздовж яких поширюється світлова енергія, називають променями. Усереднений вектор Пойтинга $\langle \vec{\Pi} \rangle$ направлений в кожній точці по дотичній до променя. В ізотропних середовищах

(властивості в усіх напрямках однакові) напрямок $\langle\vec{\Pi}\rangle$ збігається з нормаллю до хвильової поверхні, тобто з напрямком хвильового вектора $\vec{k}$, значить промені перпендикулярні до хвильових поверхонь. В анізотропних середовищах (властивості в різних напрямках є різними) нормаль до хвильової поверхні в загальному випадку не збігається з напрямком вектора Пойтинга, тобто промені не перпендикулярні хвильовим поверхням.

12.2. Інтерференція світла

Хвильові властивості світла найбільш виразно проявляють себе в інтерференції. Ці явища характерні для хвиль будь-якої природи і порівняно просто спостерігаються на досліді для хвиль на поверхні води (див. підрозд. 10.7) або для звукових хвиль. Спостерігати ж інтерференцію світлових хвиль можна тільки за певних умов.

Визначення інтерференції хвиль надано у підрозд. 10.7. *Під інтерференцією світла зазвичай розуміють широке коло явищ, в яких у разі накладання пучків світла результуюча інтенсивність не дорівнює сумі інтенсивності окремих пучків: в одних місцях вона більша, в інших менша, тобто виникають світлі й темні ділянки – інтерференційні смуги, які чергуються.*

Умови, за яких виникає інтерференція, було розглянуто у підрозд. 10.7. Передусім хвилі мають бути монохроматичними ($\omega_1 = \omega_2$). Світло, яке випромінюється звичайними (не лазерними) джерелами, не буває строго монохроматичним, тому для спостереження інтерференції світло від одного джерела слід розділити на два пучки, а потім накласти їх один на одного. Відомі експериментальні методи отримання когерентних пучків з одного світлового пучка можна поділити на два класи. У методі ділення хвильового фронту пучок пропускають, наприклад, через два близько розташованих отвори в непрозорому екрані. Такий метод придатний тільки за досить малих розмірів джерела. В іншому методі пучок розділяють однією або декількома поверхнями, які частково відбивають і пропускають світло. Цей метод поділу амплітуди можна застосовувати і для протяжних джерел. Він забезпечує більшу інтенсивність і лежить в основі дії різних інтерферометрів. Залежно від кількості пучків, які інтерферують, розрізняють двопроменеві й багатопроменеві інтерферометри, які мають важливе практичне застосування у техніці, метрології і спектроскопії.

Розглянемо інтерференцію монохроматичного світла. Світлові коливання в деякій точці, через яку проходить строго монохроматична хвиля, мають відбуватись нескінченно довго й мати незмінні частоту і амплітуду. Просте світло, яке випромінюється будь-яким реальним джерелом, не має цієї властивості. Тим не менше, монохроматична ідеалізація є достатньою для розв'язання багатьох задач. Зокрема, під час вивчення явищ інтерференції вона придатна для визначення положення максимумів і мінімумів інтерференційної картини.

Нехай розділення на дві когерентні хвилі відбувається в точці O (рис. 12.1). До точки P, в якій спостерігається інтерференційна картина, одна хвиля в середовищі з показником n_1 проходить шлях r_1, а друга – в середовищі з показником заломлення n_2 – шлях r_2. Згідно з (11.20) і (12.1) перша хвиля викличе в точці P коливання:

$$\vec{E}_1 = \vec{E}_{m1}\cos\left(\omega t - k_1 r_1 + \varphi_{01}\right), \tag{12.7}$$

а друга –

$$\vec{E}_2 = \vec{E}_{m2}\cos\left(\omega t - k_2 r_2 + \varphi_{02}\right), \tag{12.8}$$

де

$$\begin{aligned} k_1 &= \omega/v_1 = \omega n_1/c = 2\pi n_1/\lambda_0, \\ k_2 &= \omega/v_2 = \omega n_2/c = 2\pi n_2/\lambda_0 \end{aligned} \tag{12.9}$$

є хвильовими числами першої та другої хвиль; c – швидкість світла у вакуумі; $\lambda_0 = c/\nu$ – довжина хвилі у вакуумі; φ_{01} та φ_{02} – початкові фази.

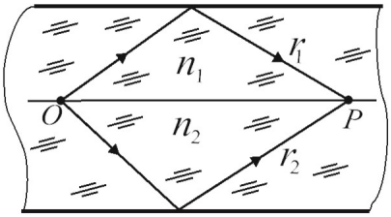


Рис. 12.1. Загальна схема інтерференції

Згідно з принципом суперпозиції напруженість результуючого поля в точці P дорівнює їх векторній сумі:

$$\vec{E}_P = \vec{E}_1 + \vec{E}_2. \tag{12.10}$$

У результаті додавання двох гармонічних коливань однакової частоти отримують коливання тієї ж частоти, а незмінна з часом його амплітуда залежить від співвідношення фаз коливань, які додаються (див. підрозд. 9.4), і тому в різних точках спостереження амплітуда має, взагалі кажучи, різні значення.

Зважаючи на дуже велику частоту оптичних коливань ($\nu \sim 10^{14}$ Гц) напруженість $\vec{E}_P$ неможливо виміряти безпосередньо. Усі приймачі випромінювання вимірюють енергетичні величини (інтенсивність світла або освітленість поверхні), усереднені за проміжок часу, який набагато більший за період оптичних коливань, тому

величини, які експериментально вимірюють, пропорційні середньому значенню квадрату напруженості електричного поля $I_p \sim \langle E_P^2 \rangle$ за час, який визначається інерційністю приймача випромінювання:

$$\langle E_P^2 \rangle = \langle (\vec{E}_1 + \vec{E}_2)^2 \rangle = \\ = \langle E_1^2 \rangle + \langle E_2^2 \rangle + 2\langle \vec{E}_1 \cdot \vec{E}_2 \rangle. \quad (12.11)$$

Вираз для результуючої інтенсивності, крім суми інтенсивностей $I_1 \sim \langle E_1^2 \rangle$ та $I_2 \sim \langle E_2^2 \rangle$ кожної із хвиль, містить ще один доданок, пропорційний $2\langle \vec{E}_1 \cdot \vec{E}_2 \rangle$, який називають *інтерференційним членом*. У тих випадках, коли він дорівнює нулю, результуюча інтенсивність $I_P = I_1 + I_2$ й інтерференції немає.

Скалярний добуток $\vec{E}_1 \cdot \vec{E}_2$ дорівнює нулю, коли $\vec{E}_1 \perp \vec{E}_2$, тобто хвилі, що додаються, поляризовані у взаємно перпендикулярних напрямках (поляризацію електромагнітних хвиль розглянемо далі). *Відсутність інтерференції променів, які поляризовані у взаємно перпендикулярних напрямках, виявили Френель і Араго в 1816 р., її у 1817 р. інтерпретував Юнг як доказ поперечності світлових хвиль. Електромагнітна теорія світла це підтвердила.*

Будемо вважати, що вектори $\vec{E}_1$ та $\vec{E}_2$ в точці спостереження P здійснюють коливання вздовж однієї прямої. Тоді інтерференційний член можна записати так:

$$2\langle \vec{E}_1 \cdot \vec{E}_2 \rangle = 2\langle E_1 E_2 \rangle.$$

Додавання однаково спрямо-ваних з однаковими частотами коливань розглянуто у підрозд. 9.4. Згідно з формулою (9.20) амплітуда результуючого коливання E_{mP} в точці P із врахуванням (12.7) і (12.8) становитиме:

$$E_{mP}^2 = E_{m1}^2 + E_{m2}^2 + \\ + 2E_{m1}E_{m2}\cos\left(k_2 r_2 - k_1 r_1 + \varphi_{01} - \varphi_{02}\right). \quad (12.12)$$

Такий самий результат можна отримати із співвідношення (12.11), якщо врахувати, що

$$\begin{cases} \langle E_P^2 \rangle = \frac{1}{2}E_{mP}^2;\ \langle E_1^2 \rangle = \frac{1}{2}E_{m1}^2,;\langle E_2^2 \rangle = \frac{1}{2}E_{m2}^2; \\ \langle \vec{E}_1 \vec{E}_2 \rangle = \frac{1}{2}E_{m1}E_{m2} \times \\ \times \cos\left(k_2 r_2 - k_1 r_1 + \varphi_{01} - \varphi_{02}\right). \end{cases}$$

Враховуючи, що інтенсивність хвиль пропорційна квадратам їхніх амплітуд, для інтенсивності результуючого коливання в точці P отримаємо

$$I_P = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\varphi, \quad (12.13)$$

де

$$\varphi = k_2 r_2 - k_1 r_1 + \varphi_{01} - \varphi_{02} \quad (12.14)$$

є різницею фаз двох когерентних хвиль від одного джерела. З урахуванням (12.9) запишемо (12.14) у такому вигляді:

$$\varphi = \frac{2\pi}{\lambda_0}(r_2 n_2 - r_1 n_{1)} + \varphi_{01} - \varphi_{02}). \quad (12.15)$$

Добуток *геометричної довжини шляху* світлової хвилі на показник заломлення цього середовища називають *оптичною довжиною шляху*, а різницю оптичної довжини шляхів, що пройшли хвилі, $\Delta = r_2 n_2 - r_1 n_1 -$ *оптичною різницею ходу*.

У випадку когерентних хвиль $\varphi_{01} - \varphi_{02} = \text{const}$, і для зручності будемо вважати, що $\varphi_{01} = \varphi_{02}$.

У результаті (12.13) перепишемо так:

$$I_P = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\frac{2\pi}{\lambda_0}\Delta. \quad (12.16)$$

Якщо оптична різниця ходу дорівнює парній кількості півхвиль у вакуумі,

$$\Delta = 2m\frac{\lambda_0}{2} \quad (m = 0,\ \pm1,\ \pm2,\ \pm3, \ldots), \quad (12.17)$$

то $\varphi = \frac{2\pi}{\lambda_0}\Delta = 2m\pi$, і коливання, що збуджуються в точці P обома хвилями, відбуваються в однакових фазах. *Число m називають порядком інтерференції.* Отже, $\cos\frac{2\pi}{\lambda_0}\Delta = 1$, і результуюча інтенсивність максимальна:

$$I_{Pmax} = I_1 + I_2 + 2\sqrt{I_1 I_2} \quad (E_{m1} \neq E_{m2}); \qquad I_{Pmax} = 4I_1 \quad (E_{m1} = E_{m2}). \quad (12.18)$$

Таким чином, максимум інтенсивності у разі накладання двох когерентних хвиль буде у точках, для яких виконується умова (12.17), тобто на оптичній різниці ходу укладається парна кількість півхвиль ($\lambda_0/2$). Сукупність таких точок утворює інтерференційні лінії (смуги), порядок яких визначають числом m.

Якщо оптична різниця ходу

$$\Delta = (2m+1)\frac{\lambda_0}{2}, \quad (12.19)$$

де $m = 0, \pm1, \pm2, \pm3, \ldots$, то $\varphi = (2m + +1)\pi$ і коливання, що збуджуються в точці P обома хвилями, перебувають у протифазі. Це означає, що $\cos\frac{2\pi}{\lambda_0}\Delta = -1$, і результуюча інтенсивність у цій точці згідно з (2.16) буде мінімальною:

$$I_{Pmin} = I_1 + I_2 - 2\sqrt{I_1 I_2} \quad (E_{m1} \neq E_{m2}); \qquad I_{Pmin} = 0 \quad (E_{m1} = E_{m2}). \quad (12.20)$$

Мінімум інтенсивності отримаємо в точках, для яких оптична різниця ходу променів вміщує непарну кількість півхвиль ($\lambda_0/2$).

Таким чином, (12.17) – *умова інтерференційного максимуму,* а (12.19) *– умова інтерференційного мінімуму*.

12.3. Розрахунок інтерференційної картини від двох когерентних джерел

Розглянемо випадок інтерференції хвиль від двох однакових синфазних ($\alpha_{01} = \alpha_{02}$) монохроматичних точкових джерел S_1 і S_2 (рис. 12.2), які розміщені на відстані d одне від одного в середовищі з показником заломлення

$n = 1$. Якщо відстань l до екрана, де спостерігаються інтерференційні смуги, набагато більша за відстань між джерелами ($l \gg d$), то коливання в точці P, що збуджуються хвилями від S_1 і S_2, мають однаковий напрямок та однакові амплітуди ($I_1 = I_2 = I_0$). Інтенсивність результуючого коливання в точці P згідно з (12.16)

$$I_P = 2I_0\left(1+\cos\frac{2\pi}{\lambda_0}\Delta\right) = \\ = 4I_0\cos^2\left(\frac{\pi}{\lambda_0}\Delta\right), \qquad (12.21)$$

де I_0 – інтенсивність коливань від одного джерела; $\Delta = (r_2 - r_1)$ – різниця ходу хвиль, що інтерферують. Освітленість екрана в мінімумах дорівнює нулю, а в максимумах – $4I_0$. Положення максимумів визначають умовою (12.17).

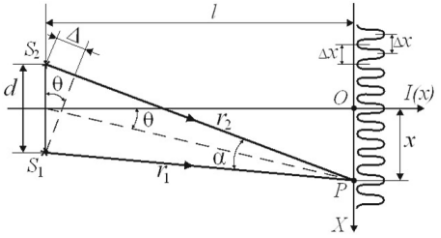


Рис. 12.2. інтерференції хвиль від двох однакових синфазних ($\alpha_{01} = \alpha_{02}$) *монохроматичних точкових джерел* S_1 *і* S_2

Якщо хвилі від джерел поширюються не у вакуумі, а в середовищі з показником заломлення n, формула (12.21) залишається справедливою, але в ній під Δ слід розуміти не геометричну, а оптичну різницю ходу хвиль, що інтерферують: $\Delta = n(r_2 - r_1)$.

Щоб знайти залежність освітленості екрана від координати x (рис. 12.2), треба різницю ходу Δ виразити через координату x точки спостереження P. Для зручності введемо кут θ, що утворюється напрямком на точку P з перпендикуляром до лінії, яка з'єднує джерела (тобто з оптичною віссю розглядуваної схеми). Для практично важливого випадку малих значень θ ($\theta \ll \ll 1$), для різниці ходу можна записати $\Delta \approx d\sin\theta \approx d\cdot\theta$. Оскільки $\mathrm{tg}\theta \approx \theta \approx \approx x/l$, то $\Delta \approx xd/l$. Підставляючи цей вираз для Δ в (12.21), отримаємо

$$I_P(x) = 2I_0\left(1+\cos\frac{2\pi dx}{\lambda_0 l}\right). \qquad (12.22)$$

У точці, для якої $x = 0$, розміщений максимум, який відповідає нульовій різниці ходу. Для цього максимуму порядок інтерференції $m = 0$. Це центр інтерференційної картини. Відстань між *сусідніми максимумами* або *мінімумами* (просторовий період інтерференційної картини) визначають з умови

$$2\pi d\cdot\Delta x/(\lambda_0 l) = 2\pi,$$

звідки

$$\boldsymbol{\Delta x = \frac{\lambda_0 l}{d}}. \qquad (12.23)$$

Формула (12.23) визначає *ширину інтерференційної смуги.*

Якщо ввести кут сходження променів $\alpha \approx d/l$, тобто кут, під яким видно джерела S_1 і S_2 із точки спостереження, то вираз для Δx можна записати так:

$$\Delta x = \lambda_0/\alpha. \qquad (12.24)$$

Формулу (12.24) використовують для випадку інтерференції плоских хвиль (паралельних променів), які поширюються під кутом α одна відносно одної. Справді, для великих відстаней від джерел сферичні хвилі на невеликих ділянках наближено можна вважати плоскими, кут між напрямками яких за $\theta \ll 1$ приблизно дорівнює d/l.

Зазначимо, що середнє значення освітленості по екрану в інтерференційній картині згідно з формулою (12.22) буде:

$$\langle I(x)\rangle = \langle 2I_0\rangle + \langle 2I_0 \cos\frac{2\pi dx}{\lambda_0 l}\rangle = 2I_0,$$

що дорівнює подвоєній освітленості від одного джерела. Це означає, що під час інтерференції відбувається тільки перерозподіл енергії у просторі, а повний потік енергії залишається незмінним.

До цього часу ми розглядали тільки точки екрана, які лежать у площині рис. 12.2. У просторі поверхнями максимальної та мінімальної інтенсивності є гіперболоїди обертання з фокусами в точках S_1 і S_2, оскільки відповідають множині точок, для яких різниця відстаней від двох заданих точок S_1 і S_2 має одне й те саме значення, тобто $r_2 - r_1 = \text{const}$. Для максимальної інтенсивності $r_2 - r_1 = 2m\frac{\lambda_0}{2}$, для мінімальної $r_2 - r_1 = (2m+1)\frac{\lambda_0}{2}$.

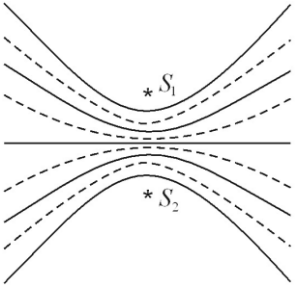


Рис. 12.3. Переріз інтерференційної картини в просторі

Переріз таких гіперболічних поверхонь зображено на рис. 12.3, де суцільними лініями показано розміщення максимумів інтенсивності, штрихованими – мінімуми інтенсивності.

Контрастність інтерференційної картини характеризують за допомогою параметра

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}, \qquad (12.25)$$

де I_{max} та I_{min} – інтенсивність світлих і темних смуг. Параметр V вперше ввів Майкельсон, його називають *функцією видності* (або просто *видністю*). Якщо інтенсивність темної смуги дорівнює нулю ($I_{min} = 0$), то $V = 1$, тобто когерентність променів максимальна. Для рівномірно освітленого екрана

($I_{max} = I_{min}$) видність $V = 0$ (промені не когерентні).

12.4. Способи спостереження інтерференції світла

1. Метод поділу хвильового фронту

Вперше експериментальну установку для демонстрації інтерференції світла сконструював Томас Юнг на початку XIX ст. Яскравий пучок сонячних променів освітлював екран A з малим отвором S (рис. 12.4). Світло, яке пройшло через отвір S, має вигляд розбіжного пучка, який падає на другий екран B із двома малими отворами S_1 і S_2, які розміщені близько один від одного на однакових відстанях від S. Ці отвори діють як вторинні точкові когерентні (синфазні) джерела і хвилі, які виходять із них, перекриваються і створюють інтерференційну картину, яку спостерігають на віддаленому екрані E.

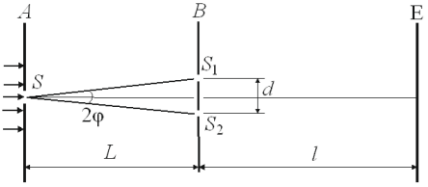


Рис. 12.4. Схема досліду Юнга

Розміщення світлих і темних смуг на екрані знаходять, користуючись монохроматичною ідеалізацією, яку розглядали раніше. Відстань Δx між сусідніми смугами знаходять за формулою (12.23). Вимірюючи ширину інтерференційних смуг, Юнг у 1802 р. вперше визначив довжини світлових хвиль для різних кольорів, хоча ці виміри й не були точними.

Інший інтерференційний дослід провів Френель у 1816 р. Дві когерентні світлові хвилі були отримані у результаті відбивання від двох дзеркал, площини яких нахилені під невеликим кутом α одна до одної (рис. 12.5). Джерелом світла є вузька яскраво освітлена щілина S, яка паралельна ребру між дзеркалами. Відбиті від дзеркал пучки падають на екран, і в тій області, де вони перекриваються, виникає інтерференційна картина. Від прямого потрапляння променів від джерела S екран захищений ширмою. Для розрахунку освітленості екрана можна вважати, що хвилі, які інтерферують, випромінюють вторинні джерела S_1 і S_2, які є уявними зображеннями щілини S у дзеркалах, тому $I(x)$ визначають формулою (12.22), в якій відстань l від джерел до екрана слід замінити на $r + b$, де $r -$ відстань від S до ребра дзеркал; $b -$ відстань від ребра до екрана (рис. 12.5). Відстань між джерелами, як видно з рис. 12.5, $d \approx 2r\alpha$, тому ширина інтерференційної смуги на екрані

$$\Delta x \approx \frac{\lambda l}{d} = \frac{\lambda(r+b)}{2r\alpha}. \qquad (12.26)$$

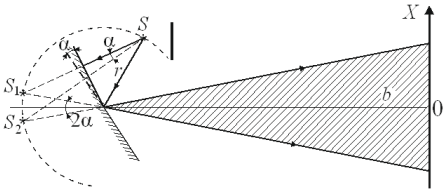


Рис. 12.5. Дзеркала Френеля

В іншому інтерференційному досліді Френеля для розділення початкової світлової хвилі на дві використовують біпризму із кутом при вершині α, який дуже малий (біпризма Френеля). Джерелом світла служить яскраво освітлена вузька щілина *S*, яка паралельна заломному ребру біпризми (рис. 12.6). Можна вважати, що тут утворюються два уявні зображення S_1 і S_2 джерела *S*, які розміщені близько, оскільки кожна половинка біпризми відхиляє промені на невеликий кут $(n-1)\alpha$.

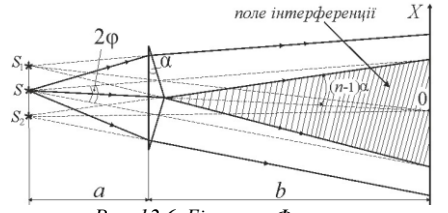


Рис. 12.6. Біпризма Френеля

Аналогічний пристрій, в якому роль когерентних джерел відіграють дійсні зображення яскраво освітленої щілини, отримують, якщо збиральну лінзу розрізати по діаметру, а половинки трохи розсунути (білінза Бі'є, рис. 12.7). Проріз закривають непрозорим екраном, а промені, які падають на лінзу, проходять через дійсні зображення щілини *S*, S_1 і S_2 й надалі перекриваються, утворюючи інтерференційне поле. У всіх описаних пристроях інтерференційні смуги можна спостерігати у будь-якому місці в області перекриття розбіжних променів від джерел S_1 і S_2 (заштриховані області на рис. 12.5 – 12.7).

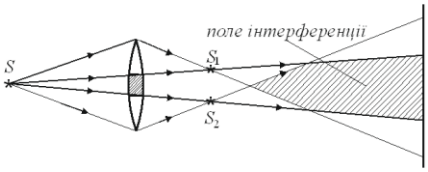


Рис. 12.7. Білінза Бі'є

2. Метод поділу амплітуд

Інтерференцію світла за методом поділу амплітуди насправді спостерігати простіше, ніж у дослідах з поділом хвильового фронту. У дослідах Поля світло від джерела *S* відбивається двома поверхнями тонкої прозорої плоско-паралельної пластинки (рис. 12.8).

У будь-яку точку *P*, яка міститься з тієї ж сторони від пластинки, що і джерело, приходять два промені з різницею ходу $\Delta = 2h\cos\theta$ (рис. 12.8). Якщо коефіцієнт відбивання малий, повторне відбивання від внутрішніх поверхонь пластинки можна не враховувати через малу енергію променів, які зазнають два відбивання і більше.

Перші два промені утворюють інтерференційну картину. Щоб визначити вигляд смуг, можна уявити, що промені йдуть з уявних зображень S_1 S_2 джерела *S*, які утворюються поверхнями пластинки. На віддаленому екрані, який розміщений паралельно до пластинки, інтерференційні смуги мають вигляд концентричних кілець із центрами на перпендикулярі до пластинки, що проходить через джерело *S*. За допомогою листочка слюди (завтовшки 0,03 … 0,05 мм) можна отримати яскраву інтерференційну картину прямо на стелі та стінах аудиторії. Чим тонша пластина, тим більша відстань між смугами.

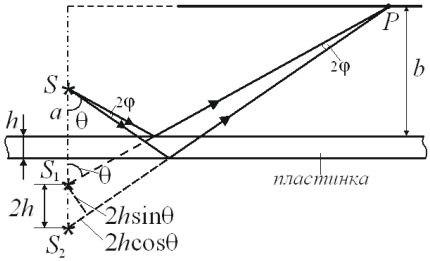


Рис. 12.8. Інтерференцію світла за методом поділу амплітуди

Особливо важливим окремим випадком інтерференції світла, яке відбивається двома поверхнями плоскопаралельної пластини, є такий, коли точка спостереження *P* міститься на нескінченності, тобто спостереження ведеться або оком, яке акомодоване на нескінченність, або на екрані, що розміщений у фокальній площині збиральної лінзи (рис. 12.9). У цьому випадку обидва промені, які йдуть від *S* до *P*, породжені одним падаючим променем і після відбивання від передньої і задньої поверхонь пластинки паралельні між собою (рис. 12.9). Оптична різниця ходу між ними в точці *P* така сама, як на лінії *DC*:

$$\Delta = n(|AB| + |BC|) - |AD| \pm \frac{\lambda}{2}.$$

Тут у фориулі n – показник заломлення матеріалу пластини. Вважають, що над пластиною є повітря, тобто $n_1 \approx 1$. Оскільки $|AB| = |BC| = \frac{h}{\cos\beta}$, $|AD| = 2h \cdot \mathrm{tg}\beta \cdot \sin\alpha$ (h – товщина пластинки; α і β – кути падіння і заломлення на верхній грані; $\sin\alpha = n\sin\beta$), то для різниці ходу отримаємо

$$\Delta = 2nh\cos\beta \pm \frac{\lambda_0}{2} = 2h\sqrt{n^2 - \sin^2\alpha} \pm \frac{\lambda_0}{2}. \quad (12.27)$$

Тут враховано, що в разі відбивання хвилі в точці *A* вона «втрачає півхвилі», тобто фаза хвилі змінюється на π, тому різниця фаз φ хвиль, які складаються в точці *P*, становить

$$\varphi = \frac{4\pi nh\cos\beta}{\lambda_0} \pm \pi = \frac{4\pi h\sqrt{n^2 - \sin^2\alpha}}{\lambda_0} \pm \pi, \quad (12.28)$$

де λ_0 – довжина хвилі у вакуумі. Різницю фаз визначають кутом α, який однозначно

пов'язаний з положенням точки P у фокальній площині лінзи (рис. 12.9).

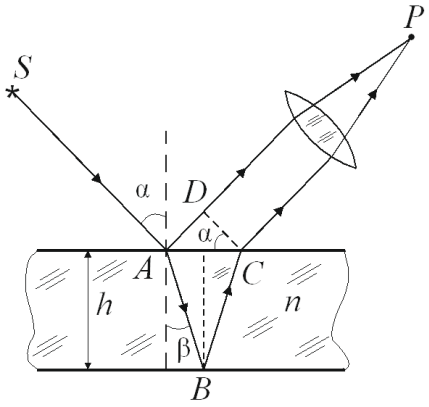


Рис. 12.9. інтерференції світла, яке відбивається двома поверхнями плоско-паралельної пластини

Згідно з формулою (12.17) світлі смуги розміщені в місцях, для яких

$$2h\sqrt{n^2-\sin^2\alpha}\pm\frac{\lambda_0}{2}=2m\frac{\lambda_0}{2}, \quad (12.29)$$

де m – ціле число, яке називають порядком інтерференції. Смуга, що відповідає такому порядку інтерференції, обумовлена світлом, яке падає на пластину під певним кутом α, тому такі смуги називають *інтерференційними смугами однакового нахилу*. Смуги однакового нахилу можна отримати не тільки у відбитому світлі, але й у світлі, яке проходить крізь пластину. В цьому випадку один із променів проходить прямо, а другий – після двох відбивань на внутрішній стороні пластини.

3. Інтерференція при протяжних джерелах.

Ми розглянули інтерференційні досліди, в яких поділ амплітуди світлової хвилі від джерела відбувався в результаті часткового відбивання на поверхнях плоскопаралельної пластини. У випадку точкового джерела смуги спостерігаються скрізь, тобто вони нелокалізовані. Утім, на нескінченності або у фокальній площині збиральної лінзи смуги спостерігаються і при протяжному джерелі. Локалізовані смуги при протяжному джерелі можна спостерігати і в інших умовах. Виявляється, що для достатньо тонкої пластинки або плівки (поверхні якої не обов'язково мають бути паралельними і взагалі плоскими) можна спостерігати інтерференційну картину, яка локалізована біля поверхні, що відбиває світло. У білому світлі інтерференційні смуги забарвлені, тому таке явище називають *кольорами тонких плівок.* Його легко спостерігати на мильних бульбашках, на тонких плівках мастила або бензину, що плавають на поверхні води, на плівках окислів на поверхні металів і плівках лазерних дисків тощо.

Локалізовані біля поверхні плівки або тонкої пластинки інтерференційні смуги можна спостерігати неозброєним оком або за допомогою лупи чи

мікроскопа. За допомогою збиральної лінзи інтерференційну картину з поверхні плівки можна відобразити на екрані (рис. 12.10, *а*).

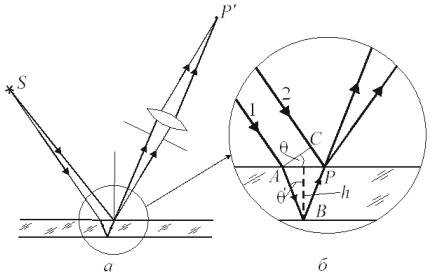


Рис. 12.10. Інтерференція на тонких плівках для точкового джерела світла

Справді, промені, які виходять з точки *P* (рис. 12.10, *б*), знову збираються в точці *P′*, яка є зображенням точки *P*, що створюється лінзою. Промені, які інтерферують, прийдуть у точку *P′* з тією ж різницею фаз, яку вони мали в точці *P*. Щоб зрозуміти причину виникнення інтерференційних смуг, знайдемо різницю ходу двох променів, які приходять у точку *P* від джерела *S*.

Із рис.12.10, *б* видно, що

$$\Delta = n(|AB| + |BP|) - |CP| \pm \frac{\lambda_0}{2}.$$

Якщо плівка мало відрізняється від плоско-паралельної, трикутники можна вважати прямокутними і вважати, що $|AB| \approx |BP| \approx \frac{h}{\cos\theta'}$, $|CP| \approx 2h\mathrm{tg}\theta' \times \sin\theta$. Тут θ – кут падіння променів на плівку; θ' – кут заломлення променів у точці *A*. За законом заломлення $\sin\theta = n\sin\theta'$. Тоді

$$\Delta = \frac{2nh}{\cos\theta'} - 2h\mathrm{tg}\theta'\sin\theta \pm \frac{\lambda_0}{2} \approx$$
$$\approx 2hn\cos\theta' \pm \frac{\lambda_0}{2},$$

де враховано «втрату півхвилі» у разі відбивання променя *2* в точці *P*. Різницею фаз φ двох променів у точці *P* буде

$$\varphi = \frac{4\pi nh\cos\theta'}{\lambda_0} \pm \pi. \qquad (12.30)$$

Формула (12.30) відповідає іншим умовам інтерференції, ніж формула (12.28). Для точки *P* товщина h має певне значення і, якщо використовувати протяжне джерело світла, то різні величини φ (формула (12.30)) для променів, що надходять від різних точок джерела, пов'язані з різним для них значенням $\cos\theta'$. У точці *P* (і, відповідно, у точці *P′*) буде максимум інтенсивності, якщо різниця ходу Δ згідно з (12.17) становитиме

$$2nh\langle\cos\theta'\rangle \pm \frac{\lambda_0}{2} = m\lambda_0 \quad (m = 0,1,2,3,\ldots), \qquad (12.31)$$

де $\langle\cos\theta'\rangle$ означає усереднення за тими точками джерела, світло від яких потрапляє в *P′*. Співвідношення (12.31) є справедливим і при неплоских поверхнях плівки за умови, що кут між ними залишається малим, тому інтерференційні смуги відповідають сукупності місць плівки, де її оптична товщина має одне й те саме значення (за умови, звичайно, що

$\langle\cos\theta'\rangle$ є достатньо однаковий для всієї області спостереження). З цієї причини такі смуги називають *смугами однакової товщини.*

Кільцеві смуги однакової товщини, які спостерігаються у повітряному зазорі між випуклою сферичною поверхнею лінзи малої кривизни і плоскою поверхнею скла, які дотикаються (рис. 12.11), називають *кільцями Ньютона*. Їх спільний центр розміщений у точці дотику. Якщо світло на лінзу падає нормально, для відбитих променів *1* і *2* (рис. 12.11) різниця ходу

$$\Delta = 2h + \frac{\lambda_0}{2},$$

де враховано «втрату півхвилі» променя *2* у разі відбивання від поверхні скла. Товщина h повітряного зазору зв'язана з відстанню r до точки дотику (рис. 12.11):

$$h = R - \sqrt{R^2 - r^2} =$$
$$= R\left(1 - \sqrt{1 - \frac{r^2}{R^2}}\right) \approx r^2/(2R).$$

Тут використано умову $(r^2/R^2) \ll 1$. Тоді різниця ходу

$$\Delta = \frac{r^2}{R} + \frac{\lambda_0}{2}.$$

Згідно з умовою (12.17) світлі кільця будуть спостерігатись за умови, що

$$\Delta = \frac{r_{\text{св}}^2}{R} + \frac{\lambda_0}{2} = 2m\frac{\lambda_0}{2}, (m = 1, 2, 3, \dots)$$

а радіуси світлих кілець

$$r_{\text{св}} = \sqrt{\frac{(2m-1)\lambda_0 R}{2}}. \qquad (12.32)$$

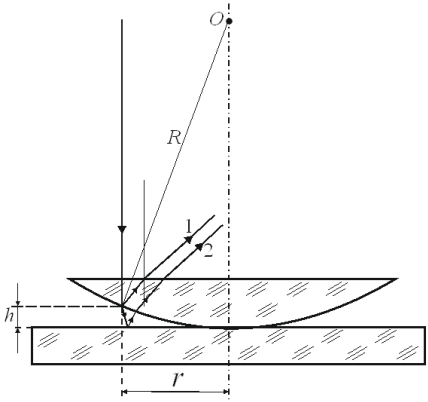


Рис. 12.11. Схема для отримання кілець Ньютона

Темні кільця відповідно до умови (12.19) становитимуть

$$\Delta = \frac{r_{\text{тем}}^2}{R} + \frac{\lambda_0}{2} = (2m+1)\frac{\lambda_0}{2}$$
$$(m = 0,\ 1,\ 2,\ 3, \dots),$$

а радіуси темних кілець знаходять за формулою

$$r_{\text{тем}} = \sqrt{mR\lambda_0}. \qquad (12.33)$$

У відбитому світлі центр темний ($r_{\text{тем}} = 0$), оскільки за товщини повітряного зазору $h \ll \lambda_0$ різниця ходу $\Delta \approx \lambda_0/2$, а різниця фаз хвиль, які інтерферують, близька до π. Якщо лінзу поступово відсувати від поверхні скла, то інтерференційні кільця будуть стягуватись до центру. Для світла, яке проходить крізь лінзу й пластину, розміщення кілець Ньютона обернене до розміщення кілець Ньютона у відбитому світлі, тобто за формулою (12.32) визначають радіуси темних кілець, а за

формулою (12.33) – радіуси світлих кілець.

За допомогою кілець Ньютона, як і в досліді Юнга, можна порівняно простими засобами приблизно визначити довжину хвилі світла.

12.5. Інтерференція квазімонохроматичного світла. Часова когерентність

Щодо всіх описаних вище інтерференційних дослідів під час їх інтерпретації вважали, що джерело випромінює монохроматичне світло (світло однієї частоти $\nu\,(\Delta\nu = 0)$ або однієї довжини хвилі $\lambda\,(\Delta\lambda = 0)$). Природно, що результати, отримані для монохроматичного світла, мають обмежене застосування. У цьому підрозділі ми з'ясуємо, до яких змін в інтерференційних явищах приводить врахування спектрального складу ($\Delta\lambda \neq \neq 0$) реальних джерел світла.

Розглянемо найпростіший випадок, коли джерело випромінює дві дуже вузькі близькі одна до одної спектральні лінії з частотами ω_1 і ω_2. Якби випромінювання на кожній із частот являло собою нескінченну синусоїдальну хвилю, то результуюче випромінювання було б хвилею середньої частоти з амплітудою, що періодично змінювалась би. Утім, насправді випромінювання кожної із спектральних компонентів є хаотична послідовність більш-менш довгих хвильових цугів. Зазвичай за час спостереження проходить багато цугів, коливання в яких ніяк не пов'язані за фазою, тому можна вважати, *що замість одного є два розміщених в одному місці джерела і незалежно одне від одного випромінюють хвилі з частотами* ω_1 і ω_2. Під час виконання інтерференційних дослідів з такими джерелами кожна із хвиль створює свою інтерференційну картину й ці картини просто накладаються одна на одну.

Якщо частоти ω_1 і ω_2 (λ_1 і λ_2) майже не різняться (*таке світло називають квазімонохроматичним, тобто майже монохроматичним*), то інтерференційні смуги в кожній картині мають майже однакову ширину (формула (12.23)). У тих випадках, де світлі смуги однієї картини накладаються на світлі смуги іншої, чіткість смуг в сумарній картині найбільша. Навпаки, там, де світлі смуги однієї картини збігаються з темними смугами іншої картини, чіткість смуг зменшується майже до повного їх зникнення.

Дослідимо це питання кількісно. Інтенсивність результуючого коливання в точці P (рис. 12.1) екрана знаходять за формулою (12.13), де різниця фаз двох

коливань (див. підрозд. 10.7, формула (10.32)) становитиме

$$\varphi=(\omega_1-\omega_2)t+ \\ +(k_2r_2-k_1r_1)+\varphi_{01}-\varphi_{02}, \quad (12.34)$$

або

$$\varphi = \Delta\omega t + \varphi_{\text{р}}, \quad (12.35)$$

де $\varphi_{\text{р}} = (k_2r_2 - k_1r_1) + \varphi_{01} - \varphi_{02}$ – різниця фаз двох коливань у точці P, яку визначають за формулою (12.14). Якщо інтенсивності хвиль з частотами ω_1 і ω_2 однакові ($I_1 = I_2 = I_0$), то формулу (12.13) можна переписати так:

$$I_P = 2I_0\big(1 + \cos\big(\Delta\omega t + \varphi_{\text{р}}\big)\big),$$

або

$$I_P = 4I_0\cos^2\left(\frac{\Delta\omega}{2}t + \frac{\varphi_{\text{р}}}{2}\right). \quad (12.36)$$

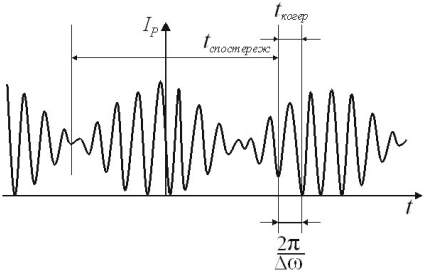


Рис. 12.12. Залежність інтенсивності світла в точці спостереження від часу

Приблизний вигляд показаний на рис. 12.12 згідно з формулою (12.36). Якщо час спостереження в точці значний, то ані максимуму, ані мінімуму не зафіксуємо (рис. 12.12). Для часу $t_{\text{спостереж}} > t_{\text{когер}} = \frac{2\pi}{\Delta\omega}$ в точці P виникне пульсація інтенсивності. Якщо час спостереження $t_{\text{когер}} = 2\pi/\Delta\omega$, то в точці буде або максимум, або мінімум інтерференційної картини. Час $t_{\text{когер}}$ називають *часом когерентності*. Це такий проміжок часу, протягом якого випадкова зміна фази хвилі досягає значення π. Тоді

$$\tfrac{\Delta\omega}{2}(t+t_{\text{когер}}) + \tfrac{\varphi_{\text{р}}}{2} - \left(\tfrac{\Delta\omega}{2}t + \tfrac{\varphi_{\text{р}}}{2}\right) = \pi\,.$$

Звідси знаходимо, що

$$t_{\text{когер}} = 2\pi/\Delta\omega = 1/\Delta\nu. \quad (12.37)$$

Якщо врахувати, що $\nu = c/\lambda$, $|\Delta\nu| = c\Delta\lambda/\lambda^2$, то

$$t_{\text{когер}} = \lambda^2/(c\Delta\lambda). \quad (12.38)$$

Відстань $l_{\text{когер}} = ct_{\text{когер}}$, на яку переміщається хвиля за час $t_{\text{когер}}$, називають *довжиною когерентності* (або довжиною цуга). Довжиною когерентності є та відстань, на якій випадкова зміна фази хвилі досягає значення приблизно π. Щоб отримати інтерференційну картину методом поділу природної хвилі на дві частини потрібно, щоб оптична різниця ходу Δ була меншою, ніж довжина когерентності $l_{\text{когер}}$. Ця вимога обмежує кількість видимих інтерференційних смуг, що спостерігаються за схемою, яка показана на рис. 12.2. Зі збільшенням номера смуги m різниця ходу зростає, унаслідок чого чіткість смуг (їх видність) стає дедалі меншою.

Довжині когерентності відповідає максимально можливий порядок інтерференції:

$$m_{max}\lambda \approx l_{\text{когер}} = \lambda^2/\Delta\lambda\,.$$

Звідси

$$\boldsymbol{m_{max} = \lambda/\Delta\lambda.} \qquad (12.39)$$

Для білого світла (сонце, лампа розжарювання, дуга з вугільними електродами) ефективним діапазоном довжин хвиль для візуального спостереження є довжини хвиль приблизно від 0,4 до 0,7 мкм, тобто $\Delta\lambda \sim \lambda$. У цьому випадку $m_{max}\sim 1$ й інтерференційні смуги, здавалось би, спостерігатись не повинні. Справді, приймач випромінювання, який має однакову чутливість на різних ділянках спектру, наприклад термоелемент, покаже при переміщенні в полі зору впоперек смуг майже рівномірний розподіл освітленості. Врахуємо, що око є селективним приймачем із сильно змінною чутливістю залежно від довжини хвилі, що дає деяким хвилям перевагу над іншими. Візуальне спостереження смуг у білому світлі полегшується, дякуючи здатності нашого зору розрізняти кольори, а не тільки інтенсивність світла, тому в білому світлі око розрізняє близько десятка забарвлень інтерференційних смуг. Коли різниця ходу дорівнює нулю для деяких місць, куди обидві інтерферуючі хвилі надходять в однаковій фазі, умова максимуму виконується для всіх довжин хвиль. У цьому місці виникає ахроматична (тобто незабарвлена) світла смуга. По обидва боки від неї містяться забарвлені максимуми й мінімуми, а за ними поле зору здається оку рівномірно освітленим білим світлом. Таке походження мають чудові інтерфереційні кольори в тонких плівках мастила або бензину на поверхні води. Легко оцінити максимальну товщину плівки, за якої можливе візуальне спостереження інтерференції в білому світлі. Вважаючи, що $m_{\max} \approx 10$, для максимальної можливої різниці ходу отримуємо $\Delta_{\max}\sim 10\langle\lambda\rangle$, (де $\langle\lambda\rangle = 0{,}5$ мкм). При цьому товщина плівки удвічі менша: $h_{\max} \sim 5\langle\lambda\rangle\sim$ $\sim$2,5 мкм.

12.6. Роль кінцевих розмірів джерела світла. Просторова когерентність

Під час аналізу інтерференційних дослідів первинне джерело хвиль вважаюьб точковим. Утім, всі реальні джерела світла мають кінцеві розміри. Збільшення розмірів джерела, як і розширення спектру випромінюваного світла, призводить до погіршення контрастності (видності) інтерференційних смуг і навіть до їх

повного зникнення. Щоб з'ясувати роль розмірів джерела, будемо вважати тут випромінюванням монохроматичним ($\Delta\lambda = 0$).

Протяжне джерело, яке світиться, складається з великої кількості точкових взаємно некогерентних елементів, тому інтенсивність у будь-якому місці дорівнює сумі інтенсивностей в інтерференційних картинах, що створюються окремими точковими елементами джерела.

В інтерференційних дослідах за методом поділу хвильового фронту (підрозд. 12.4) смуги на екрані перпендикулярні площині, в якій містяться первинне точкове джерело S і вторинні джерела S_1 та S_2. Використання замість S лінійного джерела (досить вузької щілини, витягнутої перпендикулярно до цієї площини), збільшує інтенсивність і не викликає погіршення чіткості інтерференційних смуг. Утім, збільшення ширини щілини приводить до того, що смуги стають менш чіткими або пропадають зовсім.

Якщо розміри джерела (тобто ширина щілини S) набагато менша за довжину світлової хвилі, то інтерференційна картина буде різкою, оскільки різниця ходу променів, що інтерферують, від будь-якої точки джерела (за його шириною) до деякої точки спостереження P буде фактично тією самою. Утім, як правило, розміри джерела значно більші за довжину хвилі, тому однакові інтерференційні картини від різних його елементів зсунуті одна відносно одної. У результаті накладання цих картин інтерференційні смуги стають більш-менш розмитими. Їх можна спостерігати тільки в разі використання певних умов, які накладаються на геометрію експерименту. Ці умови розглянемо далі.

Розглянемо протяжне джерело світла, розміри якого D (це ширина щілини). Кожна точка перерізу такого джерела є точковим джерелом світла. Розглянемо дві такі точки S_1 і S_2 (рис. 12.13, *а*), які є некогерентними світними точками. В інтерференційних дослідах світло від кожного джерела потрапляє в деяку точку спостереження на екрані двома різними шляхами. Розглянемо випадок, коли промені, що інтерферують, виходять з кожної точки джерела симетрично відносно нормалі до лінії S_1S_2. Нехай промінь, який іде від S_1 по одному з цих шляхів, утворює кут β_1 (промінь 1) з лінією S_1S_2, а іншим шляхом – кут β_2 (промінь 2). Джерелу S_2 відповідають промені $1'$ і $2'$. Зсув двох інтерференційних картин на екрані, що

створюються джерелами S_1 і S_2, визначається сумою різниць ходу Δ_1 променів 1 і $1'$ та Δ_2 променів 2 і $2'$:

$$\Delta_1 = D\sin\left(\frac{\pi}{2} - \beta_1\right) = D\cos\beta_1,$$

$$\Delta_2 = D\sin\left(\beta_2 - \frac{\pi}{2}\right) = -D\cos\beta_2.$$

Тоді

$$\Delta_1 + \Delta_2 = D(\cos\beta_1 - \cos\beta_2).$$

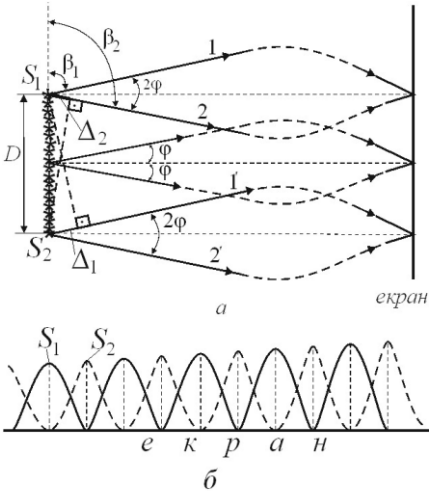


Рис. 12.13. Роль кінцевих розмірів джерела світла на інтерфенційну картину

Для симетричного випадку (рис. 12.13, *а*) кут $\beta_1 = \frac{\pi}{2} - \varphi$, кут $\beta_2 = \frac{\pi}{2} + \varphi$. Тоді $\Delta_1 + \Delta_2 = 2D\sin\varphi$. Якщо $\Delta_1 + \Delta_2 = 0$, то максимуми однієї картини збігаються з максимумами іншої. Якщо $\Delta_1 + \Delta_2 = \frac{\lambda}{2}$, світлі смуги інтерференційної картини від джерела S_1 суміщаються з темними смугами інтерференційної картини від джерела S_2 (рис. 12.13, *б*), тому як умову отримання хорошої інтерференційної картини від протяжного джерела можна взяти нерівність

$$\Delta_1 + \Delta_2 = 2D\sin\varphi \le \frac{\lambda}{2}. \qquad (12.40)$$

Для випадку малого кута $\sin\varphi \approx \varphi$ і (12.40) можна переписати так:

$$\boldsymbol{D \le \lambda/(2 \cdot 2\varphi)}, \qquad (12.41)$$

де 2φ – кут між інтерферуючими променями, що виходять із джерела, який називають *апертурою інтерференції. Апертура інтерференції — це кут, під яким з однієї точки джерела світла виходять два промені, що приходять в одну точку екрана. Цей параметр визначає видимість інтерференційної картини: зі збільшенням кута (через збільшення розмірів джерела) чіткість смуг падає, і вони зникають.*

Якщо апертура велика, спостерігати інтерференцію можна тільки від джерел, розміри яких менші від довжини світлової хвилі. Якщо $\varphi \cong \pi/2$, тобто промені, що інтерферують, виходять із джерела майже у протилежних напрямках, то із (12.41) отримуємо, що його протяжність D має бути меншою за $\lambda/4$. Для спостереження інтерференції з використанням джерела, розміри якого набагато більші за довжину хвилі світла, геометрія експерименту має бути такою, щоб промені, які

інтерферують, виходили із джерела під малим кутом один до одного.

Застосуємо критерій (12.41) до інтерференційних дослідів, які було розглянуто вище. У досліді Юнга (рис. 12.4) $2\varphi \approx \frac{d}{L}$, тому ширина додаткової щілини S має задовільняти умову $D \leq \frac{\lambda L}{2d}$. За $\lambda = 5 \cdot 10^{-7}$ м, $L = 1$ м, d = 0,5 мм знаходимо, що ширина щілини має бути меншою за 0,5 мм.

Розглянемо дослід із дзеркалами Френеля (рис. 12.5). Для точки спостереження P, яка лежить у центрі інтерференційного поля, кут $2\,\varphi$ між інтерферуючими променями, які виходять із джерела S, легко знайти з побудови, зображеної на рис. 12.14. Кут α (кут між дзеркалами) є зовнішнім для трикутника S_1OP, й тому $\alpha = \varphi + \frac{\beta}{2}$. Для половини кута сходження променів можемо записати $\frac{\beta}{2} = \frac{r\alpha}{r+b}$. Виключаючи з цих рівнянь $\frac{\beta}{2}$, знаходимо, що $\varphi = \frac{\alpha b}{(r+b)}$.

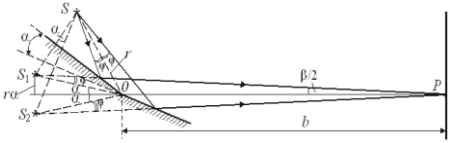


Рис. 12.14. Апертура інтерференції для бідзеркал Френеля

Підставимо φ в (12.40) й отримаємо таке обмеження на ширину D щілини джерела S:

$$D \leq \lambda(r+b)/(4b\alpha).$$

Якщо $b \gg r$, то ця умова набуває вигляду $D \leq \lambda/(4\alpha)$. Щоб можна було спостерігати смуги інтерференції із джерелом, для якого $D \gg \lambda$, кут між дзеркалами має бути дуже малим ($\alpha \ll 1$). Можна показати, що в дослідах із дзеркалами Френеля апертура інтерференції 2φ має фактично одне й те саме значення за будь-якого положення точки спостереження P на екрані в області, де перекриваються інтерферуючі промені, тому чіткість смуг однакова по всьому інтерференційному полю.

У досліді з біпризмою Френеля (рис. 12.6.) апертура інтерференції також фактично однакова по всьому полю:

$2\,\varphi = \alpha(n-1)b/(r+b) \approx 2\alpha(n-1)$ (останнє справедливе за $r \ll b$). Підставляючі 2φ у (12.40), знаходимо, що $D \leq \lambda/(4\alpha(n-1))$.

Під час аналізу інтерференційного досліду Поля (рис. 12.8) для спрощення будемо нехтувати заломленням у слюді, тобто замінимо пластинку двома паралельними площинами, які відбивають світло, відстань між якими дорівнює товщині h пластини. Тоді відстань між вторинними джерелами S_1 і S_2 (уявними зображеннями джерела S) дорівнює $2h$. Кут сходження променів у точку

спостереження в цьому досліді дорівнює апертурі інтерференції 2φ. Відстань $|S_1P|$ дорівнює $\frac{a+b}{\cos\theta}$, тому $2\varphi = \frac{2h\sin\theta}{|S_1P|} = \frac{h\sin 2\theta}{a+b}$. Товщина h листочка слюди дуже мала ($\approx 0{,}05$ мм) порівняно з $a+b \approx 5$ м, тому *мала* й апертура інтерференції 2φ (за будь-якого положення точки спостереження P, включаючи $\theta = 45°$). Значить, розмір джерела S може бути достатньо великим: із (12.40) знаходимо $D \leq \frac{\lambda(a+b)}{2h\sin 2\theta}$. Якщо $\lambda = 5 \cdot 10^{-7}$ м, $h = 0{,}05$ мм, $a+b = 5$ м і $\theta = 45°$ отримаємо, що $D \ll 25$ мм. Для демонстрації досліду Поля можна використати велику лампу без всяких щілин. За допомогою листочка слюди, площа якого декілька квадратних сантиметрів, можна отримати яскраву інтерференційну картину великих розмірів, що покриває стелю та стіни аудиторії. Розмір джерела (≈ 10 мм) набагато більший за відстань $|S_1S_2|$ ($2h \approx 0{,}1$ мм), так що уявні зображення джерела майже повністю перекриваються.

Під час спостереження локалізованих на нескінченності смуг однакового нахилу (рис. 12.9) видно, що обидва промені, які інтерферують, виходять із джерела в одному напрямку, тобто апертура інтерференції дорівнює нулю, тому тут немає ніяких обмежень на розміри джерела.

Залежність видності (12.25) смуг від апертури інтерференції можна наочно продемонструвати на досліді із дзеркалом Ллойда (рис. 12.15). Тут прямий пучок світла від джерела S інтерферує з пучком, який відбитий від дзеркала під час майже ковзаючого падіння. На відміну від дослідів Юнга або Френеля, в досліді Ллойда апертура інтерференції сильно залежить від положення точки спостереження P на екрані, який встановлений перпендикулярно до площини дзеркала. Із (рис. 12.15) видно, що $2\varphi \approx 2x/l$. Апертура тим менша, чим ближче точка P до площини дзеркала. Якщо використано протяжне джерело S, видність смуг помітно спадає в разі віддалення від точки $x = 0$.

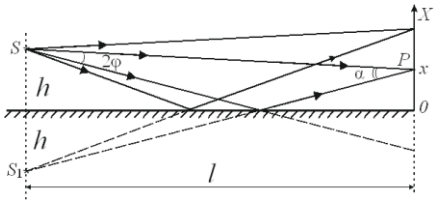


Рис. 12.15. Дослід із дзеркалом Ллойда

Критерій (12.40) дозволяє визначити розмір області, у межах якої видність смуг (12.25) більша від 2/3 за ширини D щілини джерела S: $x \leq \frac{\lambda l}{4D}$. На вдвічі більшій відстані $\lambda l/(2D)$ смуги пропадають. Кут сходження променів, як видно з рис. 12.15, $\alpha \approx 2h/l$. Ширина

інтерференційної смуги $\Delta x \approx \frac{\lambda}{\alpha} \approx \frac{\lambda l}{4h'}$, тому повна кількість смуг на цій відстані буде $N = 2x/\Delta x = h/D$. Це дає оцінку найвищого порядку інтерференції в монохроматичному світлі від протяжного джерела у досліді Ллойда.

12.7. Практичне застосування інтерференції світла

Явище інтерференції світла використовують для створення різних оптичних приладів, за допомогою яких можна виконувати такі операції:

– вимірювати довжину світлової хвилі з великою точністю;

– вимірювати лінійні розміри тіл і коефіцієнти лінійного розширення тіл;

– визначати якість і точність шліфування різних поверхонь;

– вимірювати показники заломлення газоподібних, рідких і твердих тіл;

– вимірювати кутові розміри небесних тіл;

– вивчати й контролювати однорідність речовин;

– вивчати структури спектрів різних речовин;

– досліджувати ударні хвилі в газах.

За допомогою явища інтерференції можна виготовляти інтерференційні світлофільтри і вдається значно зменшити відбивання світла від оптичних систем.

Розглянемо більш детально деякі практичні застосування явища інтерференції.

О. Смакула розробив спосіб просвітлення оптики для зменшення втрат світла, зумовлених його відбиванням на межі поділу двох середовищ, оскільки проходження світла крізь кожну заломну поверхню лінзи супроводжується відбиттям $\approx 4\,\%$ падаючого потоку світла. У складних об'єктивах кількість відбивань велике і сумарна втрата світлового потоку досить значна, що приводить до зменшення світлосили оптичних приладів. Крім того, відбивання від поверхні лінз спричиняє виникнення бліків.

Щоб елементи оптичних систем зробити просвітленими, їх поверхні покривають прозорими плівками речовин, показник заломлення яких менший, ніж скла. Під час відбивання світла від межі поділу «повітря – плівка» й «плівка – скло» виникає інтерференція когерентних хвиль $1'$ і $2'$ (рис. 12.16.). Товщину плівки d і показники заломлення скла n_c та плівки n можна підібрати так, щоб хвилі, які інтерферують, гасили одна одну. Для цього їх амплітуди мають бути рівними, а оптична різниця ходу $\Delta = (2m+1)\frac{\lambda_0}{2}$.

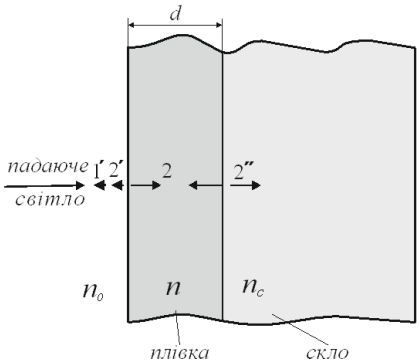


Рис. 12.16. Просвітлення оптики

Розрахунки показують, що амплітуди відбитих хвиль однакові, якщо $n = \sqrt{n_c}$. Речовину плівки підбирають так, щоб виконувалась умова $n_c > n > n_0$, де n_0 – показник заломлення повітря. Втрата півхвилі в цьому випадку відбувається на обох поверхнях. Умова мінімуму за кута падіння $\alpha = 0$ має такий вигляд:

$$2dn = (2m + 1)\frac{\lambda_0}{2}\cdot$$

Для мінімальної товщини плівки $m = 0$ й тоді

$$dn = \frac{\lambda_0}{4}\cdot$$

Ця формула показує, що неможливо добитися одночасного гасіння для всіх довжин хвиль видимого спектра, тому ця умова повинна виконуватись для довжини хвиль $\lambda_0 =$ 0,555 мкм, яку найкраще сприймає людське око.

Отже, якщо $n = \sqrt{n_c}$ й оптична товщина плівки дорівнює $\lambda_0/4$, то в результаті інтерференції спостерігається гасіння відбитих хвиль.

Явище інтерференції використовують у деяких дуже точних вимірювальних приладах, які називають **інтерферометрами** (рис. 12.17).

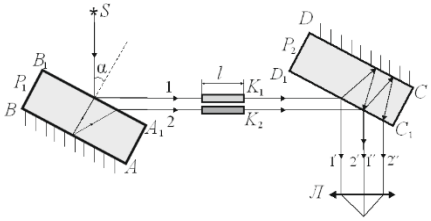


Рис. 12.17. Схема інтерферометра Жамена

Розглянемо схему інтерферометра Жамена, який використовують для точних вимірювань показників заломлення газів і їх залежності від температури, тиску та вологості (рис. 12.17). Дві однакові скляні товсті плоскопаралельні пластини P_1 та P_2 установлені майже паралельно одна до одної. Поверхні *AB* й *CD* – дзеркальні. Промені світла довжиною λ_0 від монохроматичного джерела S падають на поверхню пластини P_1 під різними кутами α, близькими до 45°. На рис. 12.17 показано лише один падаючий промінь. Унаслідок його відбивання від поверхонь *AB* і A_1B_1 пластини P_1 з неї виходять дві когерентні паралельні хвилі *1* і *2*. Пройшовши крізь дві цілком однакові закриті скляні кювети K_1 і K_2 довжиною

l, ці хвилі відбиваються поверхнями CD і C_1D_1 пластини P_2. У результаті виникають чотири промені $1'$ і $1''$, $2'$ і $2''$, з яких промені $1''$ і $2'$ накладаються один на одного й інтерферують.

Якщо одну з кювет K_1 заповнено газом з відомим показником заломлення n, а другу – газом з показником заломлення n_x, який потрібно знайти, то оптична різниця ходу між інтерферуючими хвилями $\Delta = (n_x - n)l$. Якщо ця різниця ходу $\Delta = 2m\frac{\lambda_0}{2}$, то вся інтерференційна картина зміститься на m смуг. Визначивши m, можна знайти показник заломлення n_x:

$$n_x = n + \frac{m\lambda_0}{l}.$$

Унаслідок великої чутливості інтерферометра його використовують для вимірювання показників заломлення газів, що майже не відрізняються від одиниці.

Розглянемо спрощену схему інтерферометра Майкельсона (рис. 12.18). Монохроматичне світло від джерела S падає під кутом 45° на плоскопаралельну пластину P_1. Сторона пластини AB, яка посріблена і напівпрозора, розділяє промінь на дві частини: промінь *1*, який відбивається від посрібленого шару, і промінь *2*, що проходить крізь нього. Промінь *1* відбивається від дзеркала M_1 й повертається назад, знову проходить крізь пластину $P_1(1')$. Промінь *2* поширюється до дзеркала M_2, відбивається від нього, повертається назад і відбивається від пластини $P_1(2')$. Оскільки промінь *1* проходить пластину P_1 двічі, то для компенсації різниці ходу, яка виникає на шляху другого променя, ставлять пластину P_2, яку називають компенсатором.

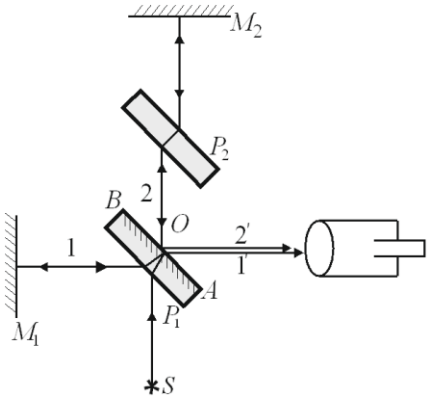


Рис. 12.18. Інтерферометр Майкельсона

Дзеркала M_1 і M_2 розміщують перпендикулярно одне до одного, одне з них можна переміщати за допомогою мікрометричних гвинтів паралельно саме до себе.

Хвилі $1'$ і $2'$ – когерентні: отже, за допомогою зорової труби буде спостерігатися інтерференційна картина. Оптична різниця хвиль $1'$ і $2'$ становить $\Delta = 2n_0(l_1 - l_2)$, де n_0 – абсолютний

показник заломлення повітря, а l_1 і l_2 – відстані від точки O до дзеркал M_1 і M_2.

Якщо $l_1 = l_2$, то спостерігається інтерференційний максимум. Переміщення одного із дзеркал на відстані $\lambda_0/4$ приводить до виникнення інтерференційного мінімуму. Отже, за зміною інтерференційної картини можна зробити висновки про малі переміщення одного із дзеркал і тим самим використати інтерферометр Майкельсона для точних вимірювань довжини світлової хвилі, довжини тіла в разі зміни температури і т. д. Похибки під час вимірювання довжини становлять до 10^{-8} мм.

Інтерферометр Майкельсона можна використати для знаходження незначних змін показника заломлення прозорих тіл залежно від тиску, температури і домішок. Такий інтерферометр називають *інтерференційним рефрактометром.*

В. П. Лінник використав принцип дії інтерферометра Майкельсона для створення *мікроінтерферометра* – високочутливого приладу, який використовують для контролю чистоти обробки поверхонь металевих виробів. Світло від джерела S падає на скляний кубик A (рис. 12.19), який складається із двох половин, склеєних по діагональній площині. Одну з діагональних поверхонь напівпосріблено так, що вона частково пропускає, а частково відбиває промені, які падають на неї. Промінь *1*, який пройшов крізь кубик, потрапляє на дзеркало M_1, відбивається і, пройшовши знову крізь кубик, відбивається від діагональної площини і потрапляє в мікроскоп M. Промінь *2* відбивається від напівпрозорої діагональної пощини, падає на досліджувану поверхню BC, відбивається від неї і проходить крізь кубик у мікроскоп.

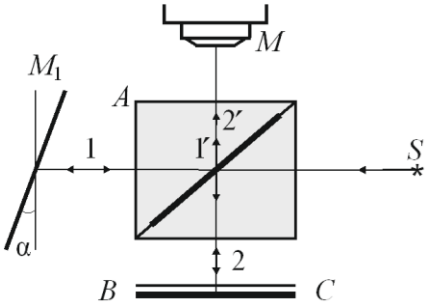


Рис. 12.19. Мікроінтерферометр Лінника

Дзеркало M_1 нахилене під невеликим кутом α. Завдяки цьому, якщо поверхня BC ідеально гладка, у полі зору мікроскопа видно інтерференційні смуги однакової товщини повітряного проміжку між кубиком і дзеркалом. У тих місцях поверхні BC, де є виступи або заглибини, інтерференційні смуги викривлятимуться. За допомогою цього приладу можна

виявити штрихи на поверхні деталі, глибина яких дорівнює $(2...3)\cdot10^{-7}$ мм.

Контрольні запитання

1. Що являє собою світло у вузькому і широкому розумінні?

2. Як визначають інтенсивність світла?

3. Як змінюється довжина світлової хвилі під час переходу з одного середовища в інше з різними показниками заломлення. Відповідь поясніть та аргументуйте.

4. Дайте визначення оптичної довжини шляху. Що називають оптичною і геометричної різницею ходу променів (хвиль)? У чому полягає суть явища інтерференції?

5. За яких умов можливе явище інтерференції?

6. Який зв'язок між явищем інтерференції і додаванням коливань однакового напрямку з однаковими частотами?

7. Дайте визначення когерентності хвиль. Які властивості когерентних хвиль? Чому світлові хвилі, що випускаються звичайними джерелами, некогерентні?

8. Які джерела світла називають монохроматичними? Чи завжди монохроматичні джерела світла є когерентними?

9. Чому відстань між когерентними джерелами коливань S_1 і S_2 має бути значно меншою за відстань від цих джерел до екрана, на якому спостерігається інтерференція?

10. На що вказує відсутність інтерференції променів, які поляризовані у взаємно перпендикулярних напрямках?

11. Запишіть і поясніть умову максимального підсилення й послаблення електромагнітних хвиль унаслідок інтерференції.

12. Дайте визначення ширини інтерференційної смуги. Від яких величин залежить ширина інтерференційної смуги?

13. Які відомі способи отримання когерентних хвиль оптичного діапазону? Поясніть суть цих способів. Побудуйте хід променів від лінійного джерела світла через біпризму Френеля й покажіть зону інтерференції. Чому заломлюючі кути біпризми мають бути дуже малими?

14. Поясніть явище «втрати півхвилі» під час відбивання хвиль від перешкоди.

15. Як впливають немонохроматичність світла та розміри джерела світла на інтерференційну картину?

16. Як визначають максимально можливий порядок інтерференції? Від чого він залежить? Поясніть, чому для видимого світла, для якого $\Delta\lambda\sim\lambda$, ми бачимо ряд різнокольорових інтерференційних смуг?

17. Наведіть приклади інтерференції електромагнітних хвиль у природі й техніці. Розкажіть про застосування явища інтерференції.

Розділ 13. Дифракція електромагнітних хвиль

13.1. Дифракція хвиль

Дифракція хвиль *(від лат.* diffractus *– розламаний, заломлений) у початковому вузькому смислі – це огинання хвилями перешкод, у сучасному більш широкому – будь-яке відхилення від законів геометричної оптики.* За такого загального тлумачення дифракція хвиль переплітається з явищами поширення і розсіяння хвиль в неоднорідних середовищах. Унаслідок дифракції хвилі можуть потрапляти в область геометричної тіні: огинати перешкоди, стелитись уздовж поверхонь, проникати через невеликі отвори в екранах тощо. Наприклад, звук можна почути за рогом будинку чи радіохвиля може проникати за горизонт навіть без відбивання від іоносфери. В явищах дифракції світла, як і в інтерференції, на перший план виступають хвильові властивості світла. Для спостереження дифракції світлових хвиль треба створити спеціальні умови. Це зумовлено малою довжиною світлових хвиль.

Між інтерференцією та дифракцією немає суттєвих фізичних відмінностей. Суть обох явищ полягає у перерозподілі світлового потоку в результаті суперпозиції хвиль. Через історичні причини перерозподіл інтенсивності світлового потоку, який виникає у результаті суперпозиції хвиль, збуджуваних кінцевою кількістю дискретних когерентних хвиль, називають інтерференцією хвиль. ***Перерозподіл інтенсивності, що виникає внаслідок суперпозиції хвиль, які збуджуються когерентними джерелами, розміщеними неперервно, називають дифракцією хвиль.*** Тому й говорять, наприклад, про інтерференційну картину від двох вузьких щілин і про дифракційну картину від однієї щілини.

Зазвичай дифракцію спостерігають за схемою, зображеною на рис. 13.1. На шляху світлової хвилі, що поширюється від деякого джерела S_0, розміщують непрозорий екран з малим отвором, розміри якого D (наприклад, порівняльні з довжиною хвилі λ світлової хвилі або й менші λ). За перешкодою розміщений екран, на якому виникає дифракційна картина. Структура дифракційного поля на екрані суттєво залежить від відстані L між джерелом S_0 і точкою спостереження P. Розрізняють ***дифракцію Френеля***, якщо $L \sim D^2/\lambda$, і ***дифракцію Фраунгофера*** за $L \gg D^2/\lambda$. Дифракцію Фраунгофера можна спостерігати, якщо між S_0 й отвором, а також між отвором та екраном

розмістити по збиральній лінзі так, щоб точки S_0 і P опинились у фокальній площині відповідної лінзи.

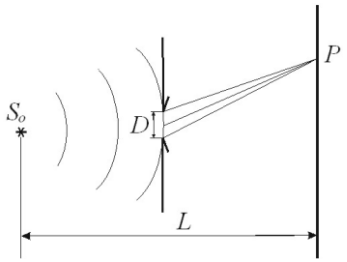


Рис. 13.1. Схема спостереження дифракції

13.2. Принцип Гюйгенса–Френеля

Перше пояснення дифракції світла дав Френель у 1818 р. У мемуарах він зазначив, що кількісне описання дифракційних явищ можливе на основі принципу Гюйгенса, якщо його доповнити принципом інтерференції вторинних хвиль. Кірхгоф у 1882 р. дав строге математичне обґрунтування принципу Гюйгенса–Френеля.

У межах електромагнітної теорії світла для описання дифракційних явищ не потрібно вводити які-небудь нові принципи. Точний розв'язок задачі про поширення світла на основі рівнянь Максвелла з відповідними граничними умовами досить складний математично. У більшості випадків, цікавих практично, наближений метод розв'язання задачі про розподіл світла біля межі між світлом і тінню, який оснований на принципі Гюйгенса–Френеля, є достатнім.

Згідно з ***принципом Гюйгенса*** кожну точку, до якої надійшла хвиля від джерела, можна вважати центром вторинних хвиль, що поширюються у всі сторони. Результуючу хвилю розглядають як накладання вторинних хвиль. Гюйгенс вважав, що окремі вторинні хвилі не мають періодичності, що вони дуже слабкі й помітну дію виконує тільки їх обвідна. За такого довільного припущення принцип Гюйгенса дає тільки деякий рецепт побудови хвильових фронтів, тобто поверхонь, до яких дійшло світлове збурення. Побудови Гюйгенса наочно пояснюють закони прямолінійного поширення, відбивання і заломлення світла, але в цих побудовах не використано поняття довжини хвилі, тому вони не дозволяють визначити умови застосування згаданих законів.

Френель вклав у принцип Гюйгенса зрозумілий фізичний зміст, відмовившись від штучного припущення про обвідну вторинних хвиль, і розглядав повне світлове поле як результат ***інтерференції вторинних хвиль***. При цьому не тільки отримано фізичне пояснення рецепту Гюйгенса (до точок на обвідній усі вторинні хвилі надходять в однаковій фазі), також з'явилась можливість розрахунку розподілу світлового поля у просторі. Вивчаючи

розподіл світла біля межі між світлом і тінню на основі принципу Гюйгенса–Френеля, можна отримати кількісний опис дифракційних явищ.

Принцип Гюйгенса–Френеля можна виразити такими положеннями:

1) під час розрахунку амплітуди світлових коливань, що збуджуються джерелом S_0 в довільній точці P, джерело S_0 можна замінити еквівалентною йому системою вторинних джерел – малих ділянок dS будь-якої замкненої допоміжної поверхні S, проведеної так, щоб вона охоплювало джерело S_0 і не охоплювала розглядувану точку P (рис. 13.2);

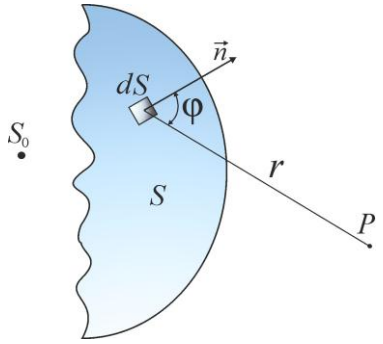


Рис. 13.2

2) вторинні джерела, які еквівалентні джерелу S_0, когерентні між собою, тому вторинні хвилі, збуджені ними, інтерферують, і розрахунок інтерференції найбільш простий, якщо S – хвильова поверхня для світла джерела S_0, оскільки при цьому фази коливань всіх вторинних джерел однакові;

3) амплітуда dE_p коливань, що збуджуються в точці P вторинним джерелом, пропорційна площі відповідної ділянки хвильової поверхні, обернено пропорційна відстані r від неї до точки P і залежить від кута φ між зовнішньою нормаллю $\vec{n}$ до хвильової поверхні та напрямком від елемента поверхні dS до точки P (рис. 13.2):

$$dE_p = K(\varphi)\frac{a_0}{r}dS\cos(\omega t - kr + \alpha_0), \quad (13.1)$$

де $(\omega t + \alpha_0)$ – фаза коливань вторинних джерел dS хвильової поверхні; a_0 – величина, пропорційна амплітуді первинних хвиль одиниці хвильової поверхні в точках елемента поверхні dS; $K(\varphi)$ – коефіцієнт, який монотонно спадає від 1 за $\varphi = 0$ до 0 за $\varphi \geq \pi/2$ (вторинні джерела не випромінюють назад); кут φ називають кутом дифракції. Багато практично важливих задач дифракції можна розв'язати, враховуючи ці досить загальні припущення відносно $K(\varphi)$, не уточнюючи конкретної залежності K від φ;

4) якщо частина поверхні S закрита непрозорими екранами, то вона не випромінює енергію, а інші випромінюють так само, як і за відсутності екранів. Врахування амплітуд і фаз вторинних хвиль дозволяє в

кожному конкретному випадку знайти амплітуду результуючої хвилі в довільній точці простору, тобто визначити закономірності поширення світла. У загальному випадку розрахунок інтерференції вторинних хвиль досить складний і громіздкий, однак для ряду випадків знаходження амплітуди результуючого коливання здійснюється алгебраїчним або геометричним додаванням. Результуюче коливання в точці P є суперпозицією коливань dE_p, взятих для всієї хвильової поверхні S:

$$E_p = \int_S K(\varphi)\frac{a_0}{r}\cos(\omega t - kr + \alpha_0)dS. \quad (13.2)$$

Формула (13.2) є аналітичним виразом принципу Гюйгенса–Френеля.

13.3. Метод зон Френеля. Прямолінійне поширення світла

За допомогою принципу Гюйгенса–Френеля із хвильових властивостей світла можна обґрунтувати *закон прямолінійного поширення світла в однорідному середовищі.* Френель розв'язав цю задачу, розглянувши взаємну інтерференцію вторинних хвиль, і застосував прийом, який отримав назву ***методу зон Френеля***. Щоб зрозуміти суть методу Френеля, визначимо амплітуду світлового коливання, яке збуджується в точці P сферичною хвилею, що поширюється в однорідному середовищі від точкового джерела S_0 (рис. 13.3).

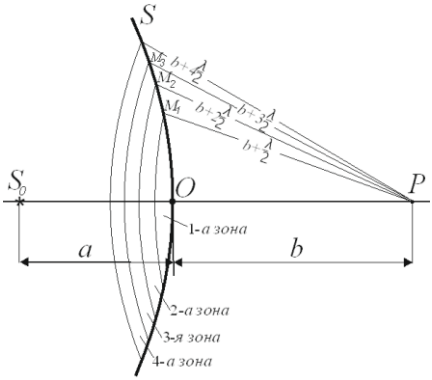


Рис. 13.3. Розбиття хвильової поверхні S на кільцеві зони

Згідно з принципом Гюйгенса–Френеля замінимо дію джерело S_0 дією уявних джерел, які розміщені на допоміжній поверхні S, що є однією із хвильових поверхонь хвиль, які поширюються від джерела S_0.

Ця допоміжна поверхня є поверхнею сфери із центром в S_0. Френель розбив хвильову поверхню S на кільцеві зони такого розміру, щоб відстані від країв зони до точки P відрізнялись на $\lambda/2$, тобто

$$PM_1 - PO = PM_2 - PM_1 = PM_3 - PM_2 = \cdots = \lambda/2.$$

Подібне розбивання хвильової поверхні S на зони можна виконати, провівши з точки P концентричні сфери з радіусами $b + \lambda/2$; $b + 2\lambda/2$; $b + 3\lambda/2$; … $b + m\lambda/2$, де $b = PO$. Точки сфери S, що лежать від точки P на відстанях

$b + \lambda/2\,;\ b + 2\lambda/2\,;\ b + 3\lambda/2$ і т.д. утворюють межі 1-ї, 2-ї, 3-ї і т.д. зон Френеля. Оскільки коливання від сусідніх зон приходять до точки P відстані, які відрізняються на $\lambda/2$, то в точку P вони надходять із протилежними фазами і в разі накладання ці коливання будуть взаємно ослаблюватися; тому амплітуда результуючого коливання в точці P

$$E_P = E_1 - E_2 + E_3 - E_4 + \cdots \pm E_m, \quad (13.3)$$

де $E_1, E_2, \ldots, E_m$ – амплітуди коливань у точці P, що збуджуються 1-ю, 2-ю, …, m-ю зонами. У цей вираз всі амплітуди коливань від непарних зон входять зі знаком «+», а від парних зон – зі знаком «−».

Величина E_m залежить від площі σ_m m-ї зони, відстані $b + \frac{m\lambda}{2}$ і кута φ_m між зовнішньою нормаллю до поверхні зони в якій-небудь її точці і прямою, яка напрямлена з цієї точки в точку P. На рис. 13.4 точки B і B_1 відповідають зовнішній межі m-ї зони: $BC = r_m$ – радіус зовнішньої межі m-ї зони. $CO = h_m$ – висота кульового сегмента BOB_1. Із трикутників S_0BC і PBC видно, що

$$r_m^2 = a^2 - (a - h_m)^2 = \\ = (b + m\,\lambda/2)^2 - (b + h_m)^2. \quad (13.4)$$

Звідси

$$a^2 - a^2 + 2ah_m - {h_m}^2 = b^2 + \\ + 2bm\,\lambda/2 + (m\,\lambda/2)^2 - b^2 - 2bh_m - h_m^2.$$

Тоді

$$2(a + b)h_m = m\lambda b + (m\,\lambda/2)^2\,.$$

Оскільки $\lambda << b$, то за не дуже великих m доданком $(m\,\lambda/2)^2$ можна знехтувати і

$$h_m = \frac{m\lambda b}{2(a+b)}\,. \quad (13.5)$$

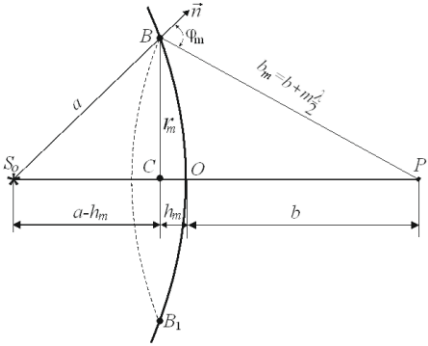


Рис. 13.4

Бічна поверхня кульового сегмента BOB_1, яка є сумою площ усіх m зон, починаючи з першої, $S_m = \sigma_1 + \sigma_2 + \ldots + \sigma_m = 2\pi a h_m = \frac{\pi ab\lambda m}{a+b}$, а площа m-ї зони Френеля

$$\sigma_m = S_m - S_{m-1} = \\ = \frac{\pi ab\lambda}{a+b}m - \frac{\pi ab\lambda}{a+b}(m-1).$$

Звідси отримуємо:

$$\sigma_m = \frac{\pi ab\lambda}{a+b}\,. \quad (13.6)$$

Цей вираз не залежить від m, отже, за не дуже великих m площі зон Френеля однакові. У такий спосіб побудова зон Френеля розбиває поверхню сферичної хвилі на однакові за площею зони. Відстань $b_m = b + m\lambda/2$ від зони до точки P повільно зростає, а це

значить, що згідно з (13.1) амплітуда коливань, збуджених у точці P зонами Френеля, зменшується із зростанням номера зони m. Зростання m супроводжується зростанням кута φ_m (рис. 13.4) і зменшенням коефіцієнта $K(\varphi)$ у формулі (13.1). Таким чином, амплітуди коливань, які збуджуються у точці P зонами Френеля, утворюють монотонну спадну послідовність:

$$E_1 > E_2 > E_3 > \cdots > E_{m-1} > E_m > \\ > E_{m+1} > \cdots.$$

Загальне число N зон Френеля, які вміщуються на частині сфери, яка повернена до точки P (рис. 13.5), дуже велике.

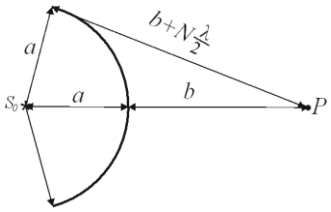


Рис. 13.5

З рис. 13.5 видно, що

$$b + N\lambda/2 = \sqrt{(b+a)^2 - a^2}.$$

Звідси:

$$N = \frac{2}{\lambda}\left(\sqrt{b^2 + 2ba}\ - b\right).$$

Якщо $a = b = 1$ м і $\lambda = 5\times \times 10^{-7}$м, то $N \sim 3 \cdot 10^6$, тому можна вважати, що в межах не дуже великих змін m залежність величини E_m від m є лінійною й амплітуда коливань, яка викликана якою-небудь m-ю зоною, дорівнює півсумі амплітуд коливань, що викликані m − 1-ю і m + 1-ю зонами, тобто

$$E_m = (E_{m-1} + E_{m+1})/2. \qquad (13.7)$$

Тоді амплітуда результуючого коливання в точці P буде

$$E_p = E_1/2 + (E_1/2 - E_2 + E_3/2) + \\ +(E_3/2 - E_4 + E_5/2\ ) + \cdots = E_1/2,$$

оскільки усі вирази, що стоять у дужках, дорівнюють нулю. Отже, амплітуда коливань, що створюється в точці P сферичною хвильовою поверхнею, дорівнює половині амплітуди коливань, що створюється однією центральною зоною. Дія всієї хвильової поверхні в точці P зводиться до дії її малої ділянки, меншої, ніж центральна зона (1-ша зона на рис. 13.3).

Якщо у виразі (13.4) покладемо, що висота сегмента $h_m \ll a$ (за не дуже великих m), тоді $r^2{}_m = 2ah_m$ і з урахуванням (13.5) радіус зовнішньої границі m-ї зони Френеля

$$r_m = \sqrt{2ah_m} = \sqrt{\frac{m\lambda ab}{a+b}}\ . \qquad (13.8)$$

Якщо $a = b = 1$ м і $\lambda = 0{,}5$ мкм, $r_1 = 0{,}5$ мм, то згідно з (13.6) $\sigma_1 = 0{,}8$ мм^2. Отже, поширення світла від S_0 до P відбувається так, немовби світловий потік поширюється всередині дуже вузького каналу вздовж прямої S_0P, тобто ***прямолінійно***. У такий спосіб хвильовий принцип Гюйгенса–Френеля дозволяє

пояснити прямолінійне поширення світла в однорідному середовищі.

Виразимо кількість зон Френеля m через радіус зовнішньої границі (13.8):

$$m = \frac{r^2{}_m}{\lambda}\frac{(a+b)}{ab} = \frac{r^2{}_m}{\lambda}\left(\frac{1}{b}+\frac{1}{a}\right).$$

Кількість зон m симетрична відносно a і b. Це означає, що точкове джерело викликає в точці P таку дію, яку викликало б воно в точці S_0, якщо його розмістити у точці P.

Наведемо наочну геометричну інтерпретацію розрахунку напруженості поля в точці P на основі принципу Гюйгенса–Френеля. Розіб'ємо хвильову поверхню на кільцеві зони, аналогічні зонам Френеля, але значно менші за шириною (різниця ходу від країв зони до точки P складає однакову для всіх зон малу частку λ). Коливання, що створюються в точці P кожною із цих зон, зобразимо у вигляді вектора $d\vec{E}$, довжина якого дорівнює амплітуді коливань, а кут, який утворює вектор $d\vec{E}$ з напрямком, який беруть за початок відліку, дає початкову фазу коливань. Амплітуди коливань, що створюються такими зонами у точці P, повільно зменшуються з переходом від зони до зони. Кожне наступне коливання відстає за фазою від попереднього на одну й ту ж величину. Наприклад, коливання в точці P, що збуджується хвилею, що надходить від 1-ї зони біля точки O (рис. 13.3), відповідає вектор $d\vec{E}_1$ на рис. 13.6, *а*. Коливання в точці P, яке збуджується хвилею від наступної (такої самої за площею) елементарної зони, зображують таким самим за модулем вектором $d\vec{E}_2$, але повернутим відносно $d\vec{E}_1$ на невеликий кут, оскільки друге коливання дещо відстає за фазою від попереднього коливання (рис. 13.6, *а*).

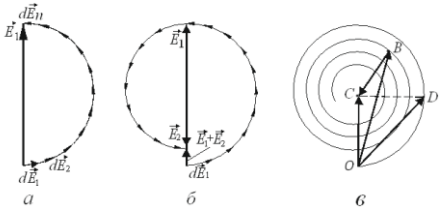


Рис. 13.6. Вектори для результуючого коливання в точці P

Коливанню в точці P, що збуджується хвилею, яка приходить від ділянки, що прилягає до межі першої зони Френеля (точка M_1 на рис. 13.3), буде відповідати вектор $d\vec{E}_n$, що повернутий відносно $d\vec{E}_1$ на π, оскільки згідно з визначенням зон Френеля різниця ходу $M_1P - OP = \lambda/2$.

Результуюче коливання в точці P, що збуджується хвилями від всієї 1-ї зони Френеля, зображено на діаграмі (рис. 13.6, *а*) вектором $\vec{E}_1$, який замикає ломану лінію, утворену векторами $d\vec{E}_1$, $d\vec{E}_2$, ... , $d\vec{E}_n$. Йому відповідає перший доданок у (13.3).

В окремому випадку, коли площі кільцевих зон дуже малі ($d\sigma \to 0$), вектор $\vec{E}_1$ проходить по діаметру напівкола. Продовжимо побудову. Векторну діаграму результуючого коливання в точці P від двох перших зон Френеля показано на рис. 13.6, *б*, і йому відповідає амплітуда $\left|\vec{E}_1 + \vec{E}_2\right| = E_1 - E_2$. Якщо б усі елементарні амплітуди dE_i, що додаються у точці P, були строго однаковими за величиною, то результуюча амплітуда коливань у точці P, що збуджуються хвилями від перших двох зон Френеля, дорівнювала б нулю (хвиля від першої і другої зон Френеля в точці P унаслідок інтерференції повністю гасила б одна одну). Утім, коефіцієнт $K(\varphi)$, що зменшується зі збільшенням кута φ, характеризує поступове зменшення амплітуди коливань dE_i у точці P, тому амплітуда $E_1 - E_2$ коливань від двох зон має кінцеве, хоч і мале значення. Цьому відповідає два перших доданки у (13.3).

Амплітуду коливань у точці P, що збуджується всією хвильовою поверхнею, зображено на рис. 13,6, *в* вектором $\overrightarrow{OC}$. Із рисунка видно, що амплітуда в цьому випадку дорівнює половині амплітуди коливання, що збуджується хвилею від першої зони Френеля, **тобто** $\boldsymbol{\left|\overrightarrow{OC}\right| = E_1/2}$. Цей результат ми отримали раніше алгебраїчно. Зауважимо, що коливання, яке збуджується внутрішньою половиною першої зони Френеля, на рис. 13.6, *в* зображено вектором $\overrightarrow{OD}$. Таким чином, дія внутрішньої половини першої зони Френеля не еквівалентна половині дії першої зони. Вектор $\overrightarrow{\mathrm{OD}}$ в $\sqrt{2}$ рази більший за вектор $\overrightarrow{OC}$, тобто інтенсивність світла, яка створюється внутрішньою половиною першої зони Френеля, у два рази перевищує інтенсивність $I_0 = (E_1/2)^2$, що створюється всією хвильовою поверхнею:

$$|\overrightarrow{OD}|^2 = 2|\overrightarrow{OC}|^2 = 2(E_1/2)^2.$$

Інтенсивність у точці P можна набагато підсилити, якщо виготовити екран, який буде відкривати, наприклад, першу, третю, п'яту і т.д., тобто всі непарні зони Френеля. Вторинні хвилі від усіх відкритих зон Френеля будуть надходити в точку спостереження P з однаковою фазою і в результаті інтерференції підсилять одна одну. Амплітуда результуючого коливання в точці P буде визначена сумою

$$E_p = E_1 + E_3 + E_5 + \cdots.$$

Такий екран називають ***зонною пластинкою*** (рис. 13.7). Можна виготовити аналогічний екран, який відкриває всі парні зони Френеля. Зонна пластинка, яка містить n відкритих зон Френеля, створює в точці P освітленість

приблизно в $4n^2$ разів більшу від освітленості, коли пластинки немає.

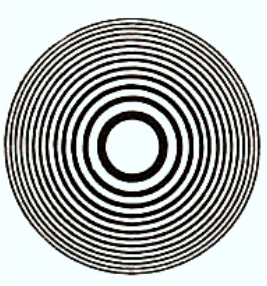


Рис. 13.7. Зонна пластинка

Підсилення інтенсивності світла зонною пластинкою аналогічне дії збиральної лінзи. Відстані від пластинки до джерела S_0 і від джерела S_0 до його «зображення» (точка P) пов'язані таким же співвідношенням, як і відповідні величини для лінзи. У цьому легко переконатись, переписавши формулу (13.8) так:

$$\frac{1}{a}+\frac{1}{b}=\frac{1}{F}, \qquad (13.9)$$

де «фокусна відстань» $F = r_m^2/(m\lambda) \approx$ $\approx r_1^2/\lambda$ для розглядуваної пластинки є постійною величиною (оскільки $r_m \sim \sqrt{m}$, то F не залежить від m). На відміну від лінзи, зонна пластинка має декілька фокусів. Справді, якщо наблизити точку P до пластинки, можна знайти таке її положення, коли в центральному світлому колі пластинки вміститься три перші зони Френеля (а не одна). Наступні три зони будуть перекриті темним кільцем, ще наступні три зони будуть суміщатися зі світлим кільцем і т. д. Дії сусідніх зон Френеля майже знищують одна одну, тому результат буде такий, як у випадку, коли відкриті 1-ша, 7-ма, 13-та і т. д. зони. У цьому випадку в точці P буде максимум освітленості, хоча і слабший, ніж в основному фокусі, а відповідна йому фокусна відстань $F_1 = F/3$. Аналогічно для фокусів вищих порядків можна записати $F_m = F/(2m+1)$, де m – цілі числа.

Інтенсивність світла в голов-ному фокусі можна збільшити ще в чотири рази (порівняно із зонною пластинкою), якщо змінити на π фази вторинних хвиль, що виходять від всіх зон з парними (або навпаки, з непарними) номерами. Таку пластинку виготовив Вуд. Хімічним травленням у певних місцях товщину скляної пластини зменшували на $(m-1)\lambda/2$ – у цьому випадку вторинні хвилі від усіх точок хвильової поверхні надходять у точку P в однаковій фазі.

13.4. Дифракція Френеля на круглому отворі та диску

Дифракція на круглому отворі

Дифракція Френеля – це дифракція сферичних світлових хвиль, а дифракційна картина спостерігається на скінченній відстані від перешкоди, яка викликає дифракцію.

Нехай сферична хвиля, що поширюється із точкового джерела S_0, зустрічає на своєму шляху екран із круглим отвором, радіус якого r_0 (рис. 13.8, *а*). Дифракційну картину спостерігаємо на екрані **Е** в точці *P*, що лежить на лінії, яка з'єднує S_0 із центром отвору *O*. Екран **Е** паралельний до площини отвору і розташований від нього на відстані *b*. Побудуємо на відкритій частині фронту хвилі зони Френеля, що відповідають точці *P*. Вигляд дифракційної картини залежить від кількості зон Френеля, які укладаються в отворі. Якщо відстані *a* та *b* задовольняють співвідношення (13.8) так, що $r_o = r_m$, де m – ціле число, то в отворі укладається m зон Френеля.

Тоді згідно з (13.3)

$$E_p = E_1 - E_2 + E_3 - E_4 + \cdots \pm E_m.$$

Із врахуванням (13.7) амплітуда E_p результуючих коливань у точці *P* залежатиме від парності або непарності *m*.

Якщо кількість зон Френеля *m* непарна, то

$$E_p = \frac{E_1}{2} + \left(\frac{E_1}{2} - E_2 + \frac{E_3}{2}\right) + \left(\frac{E_3}{2} - E_4 + \frac{E_5}{2}\right) + \\ + \cdots + \left(\frac{E_{m-2}}{2} - E_{m-1} + \frac{E_m}{2}\right) + \frac{E_m}{2} = \frac{E_1}{2} + \frac{E_m}{2}. \quad (13.10)$$

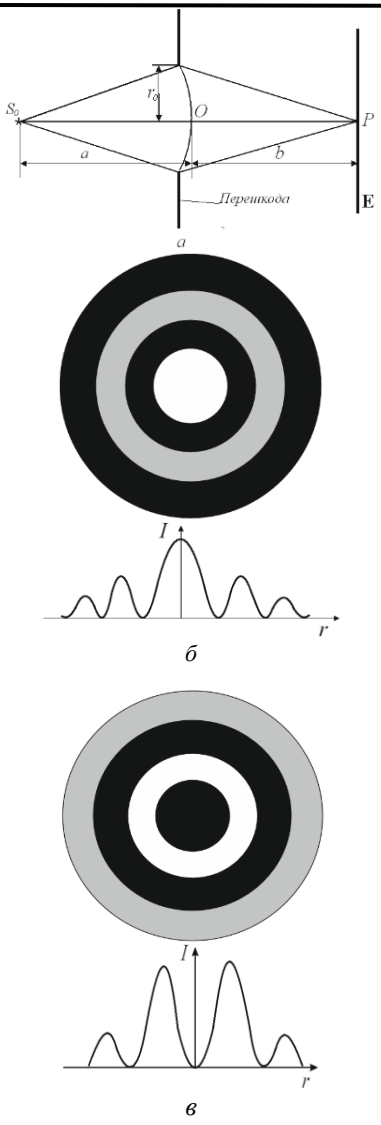


Рис. 13.8. Схема дифракції Френеля на круглому отворі (а) і теоретичний розподіл інтенсивності світла на екрані ***Е*** *(б, в), якщо отвір відкриває непарну (б) і парну (в) кількість зон Френеля*

Якщо в отворі укладається парна кількість m зон, то

$$E_p = \frac{E_1}{2} + \left(\frac{E_1}{2} - E_2 + \frac{E_3}{2}\right) + \\ + \left(\frac{E_3}{2} - E_4 + \frac{E_5}{2}\right) + \cdots + \\ + \left(\frac{E_{m-3}}{2} - E_{m-2} + \frac{E_{m-1}}{2}\right) + \\ + \frac{E_{m-1}}{2} - E_m = \frac{E_1}{2} + \frac{E_{m-1}}{2} - E_m. \quad (13.11)$$

Отже, коли отвір відкриває непарну кількість зон Френеля, то амплітуда в точці P буде більшою, ніж у разі вільного поширення хвилі, й у точці P буде інтерференційний максимум. Якщо m – парне, то у точці P буде інтерференційний мінімум. Амплітуди коливань від двох сусідніх зон фактично однакові. Тому

$$\frac{E_{m-1}}{2} - E_m = -\frac{E_m}{2}.$$

У результаті

$$E_p = \frac{E_1}{2} \pm \frac{E_m}{2}, \qquad (13.12)$$

де знак «+» стосується непарної, а знак «−» парної кількості зон Френеля. Якщо в отворі вкладається одна зона Френеля, то в точці P амплітуда $E_p = E_1$, тобто вдвоє більша, ніж коли немає непрозорого екрана з отвором. Якщо в отворі вкладаються дві зони Френеля, то їх дія в точці P фактично взаємно компенсується через інтерференцію. Таким чином, дифракційна картина від круглого отвору поблизу точки P матиме вигляд темних і світлих кілець із центрами в точці P (якщо m – непарне, то в центрі буде світлий круг (рис. 13.8, *б*), коли m – парне, круг буде темний (рис. 13.8, *в*), причому інтенсивність ($I = E^2$) максимумів зменшується з відстанню від центра картини.

Якщо положення джерела світла незмінне, кількість зон m залежить від радіуса отвору r_0 і відстані b. Зі зміною радіуса отвору в отворі змінюватиметься і кількість зон Френеля, що вміщуються на відкритій ділянці хвильового фронту, – з парної на непарну і навпаки. Унаслідок цього інтенсивність світла в точці P періодично то зменшуватиметься, то збільшуватиметься. Така зміна інтенсивності також спостерігається, якщо екран **Е** переміщувати вздовж лінії S_0P.

Коли діаметр отвору великий, так що $E_m \ll E_1$, то ніякої дифракційної картини на екрані не буде спостерігатись − світло в цьому випадку поширюється прямолінійно. Якщо отвір освітлюється не монохроматичним, а білим світлом, то кільця будуть мати кольорове забарвлення, бо кількість зон Френеля, які вкладаються в отворі, залежить від довжини хвилі світла.

Дифракція на диску

Нехай сферична хвиля, яка поширюється від точкового джерела S_0, зустрічає на своєму шляху непрозорий диск. У такому разі закриту диском

ділянку фронту хвилі треба виключити з розгляду, і будувати зони Френеля потрібно, починаючи з країв диска. На рис. 13.9 показано побудову цих зон для точки P екрана **E**, яка лежить навпроти центра диска. Амплітуду E_m в точці P визначають спільною дією всіх відкритих зон, починаючи з першої:

$$\begin{cases} E_P = E_1 - E_2 + E_3 - E_4 + \cdots = \\ = \dfrac{E_1}{2} + \left(\dfrac{E_1}{2} - E_2 + \dfrac{E_3}{2}\right) + \\ + \left(\dfrac{E_3}{2} - E_4 + \dfrac{E_5}{2}\right) + \cdots = \dfrac{E_1}{2}. \end{cases}$$

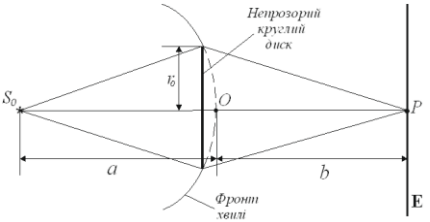


Рис. 13.9. Дифракція від круглого диска

Отже, в точці P завжди буде інтерференційний максимум, який відповідає половині дії першої відкритої зони Френеля. Центральний максимум оточений концентричними темними і світлими інтерференційними кільцями, що чергуються. Зміну інтенсивності світла I з відстанню від точки P на екрані показано на рис. 13.10. Із збільшенням радіуса диска перша відкрита зона віддаляється від точки P з одночасним збільшенням кута φ між нормаллю до поверхні цієї зони в якій-небудь її точці і напрямком випромінювання у бік точки P (див. підрозд. 13.2), тому інтенсивність центрального максимуму послаблюється зі збільшенням розмірів диска. Якщо радіус диска набагато більший за радіус закритої ним центральної зони Френеля, то за диском буде звичайна тінь з дуже слабкою інтерференційною картиною в її межах.

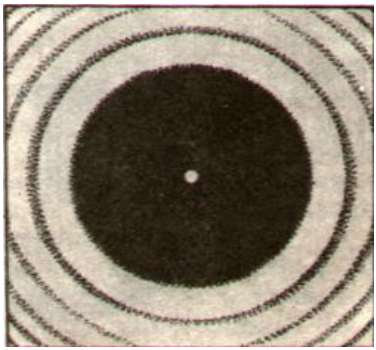


Рис. 13.10. Дифракційна картина від круглого диска

Якщо диск закриває лише невелику частину центральної зони Френеля, то він зовсім не дає тіні – освітленість екрана скрізь залишається такою ж, як і без перешкоди. Отримані вище результати можна легко отримати за допомогою векторної діаграми (рис. 13.6, *в*), якщо вважати, що вектор $\overrightarrow{OC}$, який відображає коливання напруженості поля в точці P за повністю відкритої хвильової поверхні як суму двох векторів, один з яких $\overrightarrow{OB}$ відображає коливання від тих зон Френеля, які перекриті диском, а інший $\overrightarrow{BC}$ – коливання від відкритої ділянки хвильової поверхні, яка лежить за

межами диску. Вектор $\overrightarrow{BC}$ тільки трохи менший за модулем від модуля вектора $\overrightarrow{OC}$.

Ці висновки теорії Френеля викликали сильні враження у його сучасників. У 1818 р. член конкурсного комітету французької академії наук Пуассон, який розглядав представлений на премію мемуар про дифракцію світла на круглому диску Френеля, дійшов висновку, що в центрі тіні маленького диска має міститись світла пляма, і вважав цей висновок настільки абсурдним, що висунув його як заперечення проти хвильової природи світла, яку розвивав Френель. Утім, інший член того ж комітету Араго виконав експеримент, який показав, що це дивне передбачення Френеля правильне. Довга суперечка між прихильниками корпускулярної і хвильової теорій світла завершилась на користь хвильової теорії. Світлу точку в центрі дифракційної картини від круглого диска назвали плямою Пуассона.

Вивчення дифракції Френеля в цьому підрозділі було проведено із припущенням, що джерело світла точкове, а світло, випромінюване цим джерелом, монохроматичне. У випадку протяжного джерела світло від кожного його елемента дає свою дифракційну картину. Внаслідок повної незалежності (некогерентності) світла окремих елементів просто додають інтенсивність у кожній точці, а результат дифракції визначають накладанням таких дещо зміщених одна відносно одної дифракційних картин. Щоб спостерігати дифракцію під час проведення дослідів, розміри джерела мають бути достатньо малими, щоб темні та світлі смуги картин від його окремих елементів не перекривались. Аналогічно для випадку немонохроматичного джерела різні спектральні компоненти його випромінювання створюють дифракційні картини, які не збігаються, оскільки розміри зон Френеля залежать від довжини хвилі. Розподіл інтенсивності світла, який спостерігається, відповідає накладанню цих дифракційних картин.

13.5. Дифракція Френеля на прямолінійному краю напівплощини

Принцип Гюйгенса–Френеля можна застосовувати для знаходження розподілу інтенсивності поблизу межі тіні, яка відкидається краєм великого екрана.

Обмежимось випадком плоскої хвильової поверхні падаючої хвилі, яка збігається з напівплощиною (рис. 13.11). На відстані b за напівплощиною поставимо паралельний їй екран, на якому візьмемо точку P (рис. 13.11). Розіб’ємо відкриту частину хвильової

поверхні на зони Френеля, які мають вигляд дуже вузьких прямолінійних смуг, паралельних краю напівплощини. Ширину зон виберемо так, щоб виміряні у площині рисунка відстані від точки *P* до країв будь-якої зони відрізнялись на однакову величину Δ (Δ може дорівнювати і λ/2). За такої умови коливання, які створюються в точці *P* сусідніми зонами, будуть відрізнятися за фазою на постійну величину ($\Delta\varphi = 2\pi\Delta/\lambda$).

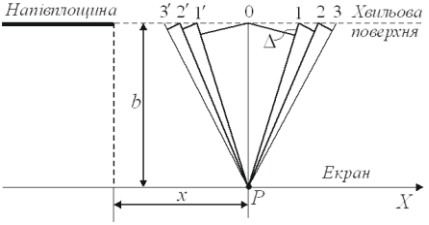


Рис. 13.11. Зони Френеля, які мають вигляд дуже вузьких прямолінійних смуг, паралельних краю напівплощини

Зонам, що розміщені справа від точки *P*, надамо номери 1, 2, 3 і т.д., а зонам, розміщеним зліва – номери $1'$, $2'$, $3'$ і т.д. Зони з номерами m і m' мають однакову ширину і розміщені відносно точки *P* симетрично, тому коливання в точці *P*, які вони створюють, збігаються за фазою і амплітудою. Щоб встановити залежність амплітуди від номера зони *m*, оцінимо площі зон. Із рис. 13.12 видно, що сумарна ширина перших *m* зон

$$d_1 + d_2 + \cdots + d_m = \sqrt{(b + m\Delta)^2 - b^2} = \sqrt{2bm\Delta + m^2\Delta^2},$$

де $d_1, d_2, \ldots, d_m$ – ширина 1-ї, 2-ї і т.д. зон. Унаслідок того, що зони вузькі, $\Delta \ll b$, тому за не дуже великих m матимемо $d_1 + d_2 + \cdots + d_m = \sqrt{2bm\Delta}$. Якщо $m = 1$, за цією формулою отримаємо, що $d_1 = \sqrt{2b\Delta}$. Тоді сумарна ширина перших m зон буде $d_1 + d_2 + \cdots + d_m = d_1\sqrt{m}$. Звідси

$$d_m = d_1\left(\sqrt{m} - \sqrt{m-1}\right). \quad (13.13)$$

Розрахунок за формулою (13.13) дає, що

$$d_1{:}d_2{:}d_3{:}d_4{:}\ldots = 1{:}0{,}41{:}0{,}32{:}0{,}27{:}\ldots. \quad (13.14)$$

У таких самих співвідношеннях знаходять і площі зон.

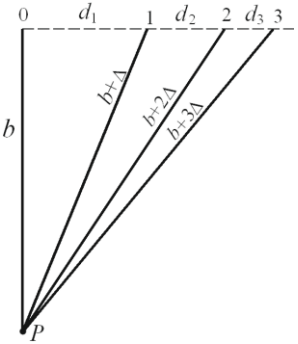


Рис. 13.12

Із (13.14) випливає, що амплітуди коливань, які створюються в точці *P* окремими зонами, спочатку (для перших зон) спадають дуже швидко, потім це спадання стає повільнішим. Через це ламана лінія, яку отримують під час графічного додавання коливань, які збуджуються прямолінійними зонами, спочатку йде більш полого, ніж у випадку

кільцевих зон (площі яких за аналогічної побудови приблизно однакові) (див. підрозд. 13.3, рис. 13.6). На рис. 13.13 зіставленно обидві векторні діаграми. В обох випадках відставання за фазою кожного наступного коливання є одним і тим же. Значення амплітуди для кільцевих зон (рис. 13.13, *а*) беруть постійним, а для прямолінійних зон (рис. 13.13, *б*) – спадним відповідно до пропорцій (13.14). Графіки на рис. 13.13 є наближеними. Щоб точно побудувати графіки, треба враховувати залежність амплітуди від r і φ (див. підрозд. 13.3). Утім, на загальному характері діаграм це не відбивається.

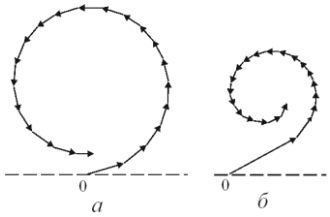


Рис. 13.13. Метод графічного додавання амплітуд

На рис. 13.13, *б* показано тільки коливання в точці P, обумовлені зонами, що лежать справа від неї. Зони з номерами m і m' розміщені симетрично відносно P, тому природно, що при побудові діаграми вектори, які відображають відповідні цим зонам коливання, розміщені симетрично відносно початку координат O (рис. 13.14). Якщо ширину зон спрямувати до нуля, то ламана лінія, яка зображена на рис. 13.14, перетворюється на плавну криву (рис. 13.15), яку називають **спіраллю Корню**.

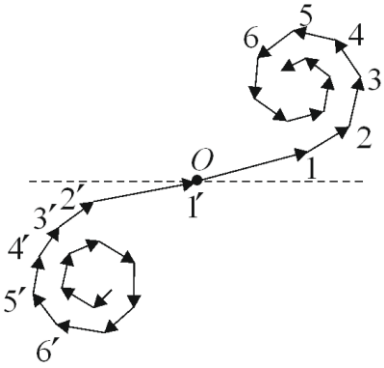


Рис. 13.14. Спіраль Корню

Рівняння спіралі Корню в параметричній формі має такий вигляд:

$$\xi = \int_0^v \cos\frac{\pi u^2}{2}\,du,$$

$$\eta = \int_0^v \sin\frac{\pi u^2}{2}\,du.$$

Ці інтеграли називають інтегралами Френеля, значення яких наводять у таблицях залежно від параметра v. Зміст параметра v полягає в тому, що $|v|$ дає довжину дуги кривої Корню, яку вимірюють від початку координат. Числа, які відмічені вздовж кривої на рис. 13.15, дають значення параметра v. Точки F_1 і F_2, до яких асимптотично наближається крива за $v \to \to \pm\infty$, називають фокусами спіралі Корню. Їхні координати дорівнюють:

– для точки F_1 $\xi = +1/2, \eta = +1/2$;

– для точки F_2 $\xi = -1/2, \eta = -1/2$.

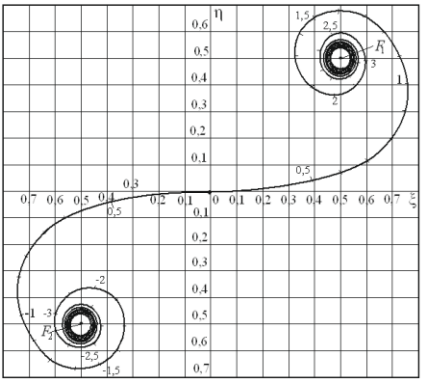


Рис. 13.15. Плавна спіраль Корню

Правий завиток спіралі (ділянка OF_1) відповідає зонам, які розміщені справа від точки *P,* лівий завиток (ділянка OF_2) – зонам, які розміщені зліва від точки *P*. Спіраль Корню дає можливість знайти амплітуду світлового коливання в будь-якій точці екрана. Положення точки будемо характеризувати координатою x, яку відраховують від межі геометричної тіні (рис. 13.11). Для точки *P*, що лежить на межі геометричної тіні ($x = 0$), усі штриховані зони будуть закриті. Коливання від нештрихованих зон відкриває правий завиток спіралі. Отже, результуюче коливання буде зображено вектором, початок якого міститься в точці *O*, а кінець – у точці F_1 (рис. 13.16, *а*). У разі зміщення точки *P* в область геометричної тіні напівплощина закриває все більше ділянок нештрихованих зон, тому початок результуючого вектора переміщується по правому завитку у напрямку полюса F_1 (рис. 13.16, *б*). У результаті амплітуда коливання монотонно прямує до нуля. Якщо точка *P* зміщується від межі геометричної тіні вправо, крім нештрихованих зон відкривається більше штрихованих зон, тому початок результуючого вектора ковзає по лівому завитку спіралі в напрямку полюса F_2. При цьому амплітуди проходять через ряд максимумів (перший із них дорівнює довжині відрізка MF_1 на рис. 13.16, *в*) і мінімумів (перший із них дорівнює довжині відрізка NF_1 на рис. 13.16, *г*). Якщо хвильова поверхня повністю відкрита, амплітуда дорівнює довжині відрізка F_2F_1 (рис. 13.16, *д*), тобто у два рази перевищує амплітуду на межі геометричної тіні (див. рис. 13.16, *а*). Відповідно, інтенсивність на межі геометричної тіні становить 1/4 інтенсивності I_0, яка отримується на екрані, коли немає перешкод.

Залежність інтенсивності світла I від координати x точки *P* зображено на рис. 13.17. Із рисунка видно, що після переходу в область геометричної тіні інтенсивність змінюється не стрибком, а поступово прямує до нуля. За допомогою спіралі Корню можна також знайти відносну величину інтенсивності у

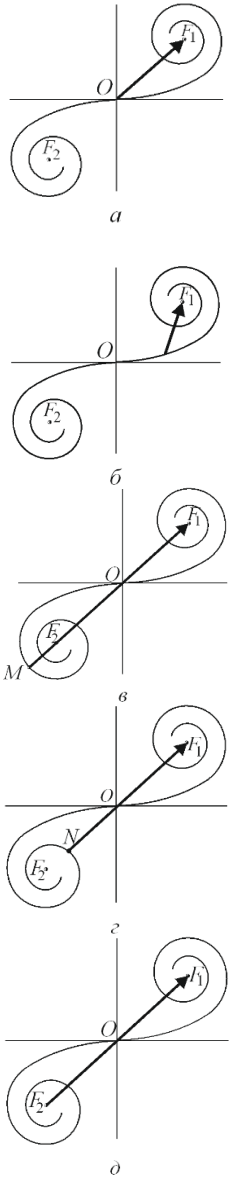


Рис. 13.16

максимумах і мінімумах. Для першого максимуму отримано значення $1{,}37I_0$, для першого мінімуму – $0{,}78I_0$.

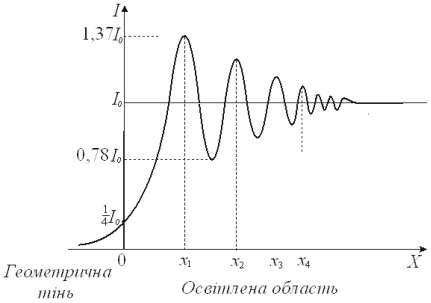


Рис. 13.17. Залежність інтенсивності світла I в точці P від координати x відносно краю великого екрана ($x = 0$)

Зробимо декілька зауважень з приводу використаного вище наближеного методу розв'язку дифракційних задач. Детальний аналіз показує, що в основі методу Френеля лежать припущення, які справедливі, коли розміри перешкод (або отворів в екранах) великі порівняно з довжиною світлової хвилі. У цьому випадку відхилення від геометричної оптики незначні, тобто помітна інтенсивність спостерігається тільки за малих кутів дифракції. Різниця в істинному і розрахованому напрямках $\vec{E}$ при цьому не суттєва. Вплив краю перешкоди (екран з отвором) на поле в отворі поширюється тільки на відстані, співмірні з довжиною хвилі, і за великих розмірів отвору заміна істинних значень dE в підінтегральному виразі формули (13.2) на напруженість E хвилі в сумі

(13.3) не приводить до суттєвих помилок, оскільки на більшій частині поверхні S ці значення збігаються. За великих порівняно з довжиною хвилі розмірів отворів розрахунки дифракційної картини методом Френеля добре підтверджуються під час дослідів та узгоджуються з результатами точних розрахунків. За цих умов підтверджується і припущення про незалежність дифракційної картини від матеріалу перешкоди. Для більшості задач метод Френеля є єдиний шлях розв'язку, який приводить до практично задовільних результатів.

13.6. Дифракція Фраунгофера

Найбільший практичний інтерес становлять дифракційні явища, які спостерігаються при падінні на екран (або на отвір в екрані) паралельного пучка світла. У результаті дифракції пучок втрачає паралельність, тобто з'являється світло, яке поширюється в напрямках, що відрізняються від початкового. Розподіл його інтенсивності на дуже великій (у граничному випадку нескінченно великій) відстані від екрана відповідає *дифракції Фраунгофера.* Хвилі, які виникають у результаті обмеження фронту падаючої «плоскої» хвилі під час проходження через отвір в екрані, називають *дифракційними*, а нормалі до їх хвильових поверхонь – дифракційними променями. Їх не існує в межах геометричної оптики.

Дифракція на одній щілині

Нехай паралельний пучок монохроматичного світла падає нормально на непрозору плоску поверхню, в якій прорізано вузьку щілину, що має сталу ширину b і довжину $l >> b$ (рис. 13.18).

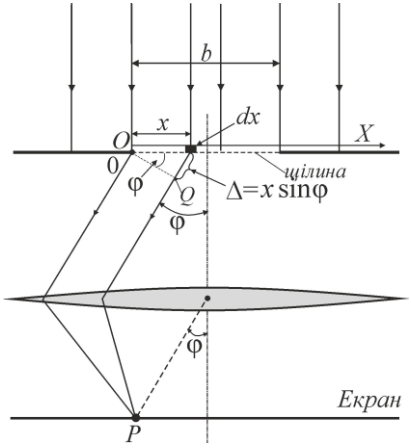


Рис. 13.18. Дифракція на одній щілині

За щілиною розмістимо збиральну лінзу, а у фокальній площині лінзи – екран. Хвильова поверхня падаючої хвилі, площина щілини та екран паралельні одні одним. Оскільки щілина вузька, картина, яка спостерігається в будь-якій площині, що перпендикулярна щілині, буде однаковою. Отже, достатньо дослідити характер картини в одній такій площині, наприклад у площині на рис. 13.18. Усі величини, які будуть вводити

надалі, наприклад кут φ між променем та оптичною віссю лінзи, належать до цієї площини. Розіб'ємо відкриту частину хвильової поверхні на паралельні краю щілини елементарні зони, ширина яких dx.

Вторинні хвилі, які посилають цими зонами у напрямку, який визначають кутом φ, зберуться в точці екрана P (рис. 13.18). Лінза збирає у фокальній площині плоскі (а не сферичні) хвилі, тому множника $1/r$ у виразі (13.1) для dE_P для випадку дифракції Фраунгофера не буде. Якщо обмежитись розглядом невеликих кутів φ, коефіцієнт $K(\varphi)$ у формулі (13.1) можна вважати постійним і взяти за одиницю. Тоді амплітуди коливань, які збуджуються зоною в будь-якій точці екрана, будуть залежати від площі зон. Площа пропорційна ширині зони dx. Отже, амплітуда коливання dE_P, що збуджується зоною, ширина якої dx, в будь-якій точці екрана має вигляд $a_0 dx/b$. Тут a_0 є амплітудою хвилі, яка падає на щілину.

Тепер визначимо фазові співвідношення між коливанням dE_P. Зіставимо фази коливань, які збуджуються в точці P елементарними зонами з координатами 0 і x (рис. 13.18). Оптичні шляхи OP і QP однакові, тому різниця фаз між коливаннями, які розглядаються, утворюється на шляху Δ, який дорівнює $x\sin\varphi$ (рис. 13.18). Якщо за фазу коливань, що збуджується елементарною зоною, яка примикає до лівого краю щілини ($x = 0$), взяти ωt, то фаза коливань, яка збуджується зоною з координатою x, буде

$$\omega t - 2\pi\frac{\Delta}{\lambda} = \omega t - \frac{2\pi}{\lambda}x\sin\varphi,$$

де λ – довжина хвилі в розглядуваному середовищі.

Таким чином, коливання, що збуджуються елементарною зоною з координатою x, положення якої на екрані у точці P визначається кутом φ, може бути подане у вигляді

$$dE_P = \frac{a_0}{b}\cos\left(\omega t - \frac{2\pi}{\lambda}x\sin\varphi\right)dx.$$

Згідно з принципом Гюйгенса–Френеля (13.2) результуюче коливання, яке збуджується в точці P відкритою ділянкою хвильової поверхні, знайдемо, проінтегрувавши dE_P по всій ширині щілини:

$$\begin{cases} E_P = \int_0^b \frac{a_0}{b}\cos\left(\omega t - \frac{2\pi}{\lambda}x\sin\varphi\right)dx = \\ = \frac{a_0}{b}\left(-\frac{\lambda}{2\pi\sin\varphi}\right)\left(\sin\left(\omega t - \frac{2\pi}{\lambda}b\sin\varphi\right) - \sin\omega t\right) = \\ = a_0\frac{\sin\left((\pi b\sin\varphi)/\lambda\right)}{(\pi b\sin\varphi)/\lambda}\cos\left(\omega t - \frac{\pi b\sin\varphi}{\lambda}\right). \end{cases}$$

При цьому було використано формулу

$$\sin\alpha - \sin\beta = 2\sin\frac{\alpha-\beta}{2}\cdot\cos\frac{\alpha+\beta}{2}.$$

Амплітуда A_φ результуючого коливання в точці P, положення якої визначають кутом φ, буде:

$$A_\varphi = \left|a_0\frac{\sin\left((\pi b\sin\varphi)/\lambda\right)}{(\pi b\sin\varphi)/\lambda}\right|. \qquad (13.15)$$

Для точки, яка лежить навпроти центра лінзи, $\varphi = 0$. Підстановка цього значення у формулу (13.15) дає значення амплітуди a_0. Цей результат можна отримати більш простим шляхом. За $\varphi = 0$ коливання від усіх елементарних зон надходять у точку P в однаковій фазі, тому амплітуда результуючого коливання дорівнює алгебраїчній сумі амплітуд коливань, які додаються.

За значень φ, які задовольняють умову $\pi b\sin\varphi/\lambda = \pm k\pi$, тобто у випадку, якщо

$$b\sin\varphi = k\lambda \ \ (k = \pm 1, \pm 2, \pm 3 \ldots), \qquad (13.16)$$

амплітуда A_φ дорівнює нулю. Таким чином, умова (13.16) визначає положення мінімумів амплітуди (інтенсивності). Зазначимо, що $b\sin\varphi$ є різниця ходу Δ променів, які йдуть у точку P від країв щілини (рис. 13.18).

Умову (13.16) легко отримати із таких міркувань. Якщо різниця ходу Δ від країв щілини дорівнює $k\lambda$, відкриту частину хвильової поверхні можна розбити на $2k$ однакових за шириною зон, причому різниця ходу від країв кожної зони буде дорівнювати $\lambda/2$ (рис. 13.19, який виконаний для $k = 2$). Коливання від кожної пари сусідніх зон взаємно компенсують одне одного, так що результуюча амплітуда дорівнює нулю. Якщо для точки P різниця ходу Δ дорівнює $(k + 1/2)\lambda$, кількість зон буде непарною, дія однієї з них буде нескомпенсованою і спостерігатиметься максимум інтенсивності.

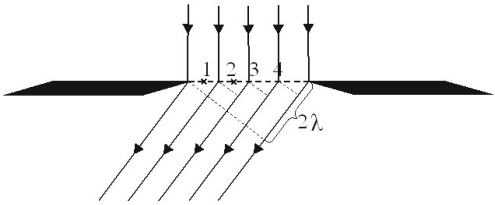


Рис. 13.19

Інтенсивність світла пропорційна квадрату амплітуди. Отже, відповідно до (13.15)

$$I_\varphi = I_0 \frac{\sin^2((\pi b\sin\varphi)/\lambda)}{((\pi b\sin\varphi)/\lambda)^2}, \qquad (13.17)$$

де $I_0 = a_0^2$ – інтенсивність у центрі дифракційної картини (навпроти центра лінзи); I_φ – інтенсивність у точці, положення якої визначається значенням φ.

Із формули (13.17) отримуємо, що $I_{-\varphi} = I_\varphi$. Це означає, що дифракційна картина симетрична відносно центра лінзи. Графік функції (13.17) зображено на рис. 13.20. По осі абсцис відкладено значення $\sin\varphi$, по осі ординат – інтенсивність I_φ. Між мінімумами інтенсивності, які визначаються умовою (13.16), мвстяться максимуми різних порядків. Їх положення знаходять з умови

$$\frac{dI_\varphi}{du} = 0, \ \text{де} \ \ u = \frac{\pi b\sin\varphi}{\lambda}, \ \text{а}$$

$$I_\varphi = I_0((\sin u)/u)^2.$$

Диференціювання I_φ за u і прирівнювання похідної до нуля дає рівняння $\mathrm{tg}u = u$, яке має корені $u_0 = 0$; $u_1 = 1{,}43\pi$; $u_2 = 2{,}46\pi$; $u_3 = 3{,}47\pi$ і т.д. Це дає, що для максимумів $\sin\varphi_0 = 0$, $\sin\varphi_1 = \pm 1{,}43\lambda$; $\sin\varphi_2 = 2{,}46\lambda$, $\sin\varphi_3 = \pm 3{,}47\lambda$ і т.д.

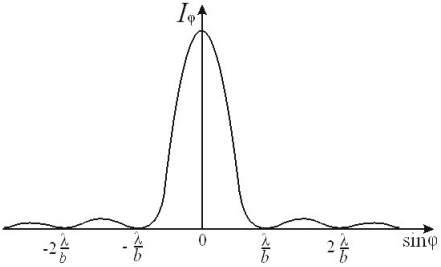


Рис. 13.20

Фактично можна вважати, що максимуми містяться посередині між сусідніми мінімумами. Значення інтенсивності в максимумах швидко спадає зі збільшенням порядку. Їх відношення приблизно можна виразити у вигляді

$$1 : 0{,}0472 : 0{,}0165 : \ldots = 1 : (2/(3\pi))^2 : (2/(5\pi))^2 : \ldots, \quad (13.18)$$

що отримують із (13.17).

Таким чином, основна частина світлового потоку зосереджена у центральній смузі між мінімумами порядків $k = \pm 1$, тобто у межах кутів $-\varphi_1 < \varphi < \varphi_1$, де $\sin\varphi_1 = \pm\lambda/b$. Кутова ширина максимума зменшується зі збільшенням ширини b щілини, якщо $\varphi \ll 1$, то $\varphi_1 \approx \lambda/b$. Центральний максимум стає різким, а перші мінімуми присуваються ближче до центра картини. Висота максимуму інтенсивності пропорційна квадрату ширини щілини, оскільки зростаючий пропорційно b світловий потік поширюється в межах спадного кута ($\varphi \sim 1/b$). Відносна інтенсивність залишається незмінною: розподіл світла по максимумах різних порядків (13.18) не залежить від ширини щілини. У разі звуження щілини картина розширюється, а її яскравість зменшується. Кількість мінімумів інтенсивності визначають відношенням ширини щілини b до довжини хвилі λ. Із умови (13.16) випливає, що $\sin\varphi_1 = \pm\frac{\lambda}{b}$.

Модуль $\sin\varphi$ не може перевищувати одиницю, тому $k\lambda/b \le 1$, звідки

$$k \le b/\lambda. \quad (13.19)$$

Якщо ширина щілини менша за довжину хвилі, мінімуми взагалі не виникають. У цьому випадку інтенсивність світла монотонно спадає від середини картини до її країв. Якщо $b = 2\lambda$, то k може набувати значення 1 і 2. Це означає, що будуть проявлятись чотири мінімуми, і за щілиною будуть поширюватись *три когерентні пучки світла* (центральний і два бічних максимуми).

Отримані результати можна використати для оцінювання дифракційної розбіжності пучків світла, ширина яких обмежена, наприклад, у результаті проходження крізь діафрагму. Основна частина світлого потоку припадає на центральний дифракційний максимум, тому його ширину можна прийняти як оцінку кутової розбіжності $\Delta\varphi$ пучка з поперечним перерізом a:

$$\Delta\varphi \sim \lambda/a. \qquad (13.20)$$

Це кутове розширення пучка обумовлене хвильовою природою світла і його у принципі не можна усунути за заданої ширини перерізу пучка. Строго паралельних світлових пучків не існує. На шляху, довжина якого l, пучок зазнає дифракційного розширення, яке визначають виразом $l \cdot \Delta\varphi = \lambda l/a$. Цим розширенням можна знехтувати тільки тоді, коли воно мале порівняно з початковою шириною пучка, тобто за $l \cdot \Delta\varphi \ll a$ або $\lambda l \ll a^2$. У таких умовах пучок приблизно можна вважати паралельним і використовувати для його описання геометричну оптику.

Пучок світла з мінімально можливою за діаметра a кутовою розбіжністю $\Delta\varphi \sim \lambda/a$ формується у результаті інтерференції вторинних хвиль від всього поперечного перерізу. Така інтерференція можлива тільки тоді, коли світлові коливання когерентні по всьому поперечному перерізу пучка. *Висока просторова когерентність лазерного пучка* обумовлена самою природою процесу випромінювання світла (вимушене випромінювання) і пояснює його малу розбіжність.

Якщо плоска хвиля падає на щілину похило під кутом φ' до нормалі, то різниця ходу між вторинними хвилями, які поширюються від країв щілини в напрямку φ, буде $b(\sin\varphi - \sin\varphi')$. Умова дифракційних мінімумів замість (13.16) набуває вигляду $b(\sin\varphi - \sin\varphi') = k\lambda$. Кутовий розподіл інтенсивності, як і раніше, визначають формулою (13.17), в якій тепер $\sin\varphi$ треба замінити на $\sin\varphi - \sin\varphi'$. Центральний максимум дифракційної картини розміщений у напрямку $\varphi = \varphi'$, тобто в напрямку падаючої хвилі.

13.7. Дифракційна ґратка

Важливе практичне застосування має явище дифракції Фраунгофера на системі з великої кількості паралельних щілин, які містяться на однакових відстанях одна від одної. Так побудована найпростіша *дифракційна ґратка*, яку виготовив у 1786 р. астроном Ріттенгауз, була у вигляді натягнутих на рамку паралельних тонких дротин. Починаючи із Фраунгофера, який виконав у 1821 р.

перші дослідження за допомогою дифракційного спектроскопа, ґратки виготовляють методом нанесення штрихів на поверхню скляної або металічної пластинки. Дифракційною ґраткою можна вважати будь-який пристрій, який забезпечує просторову періодичну модуляцію падаючої світлової хвилі за амплітудою або фазою. Відстань d між серединами сусідніх щілин називають *періодом* ґратки (рис. 13.21).

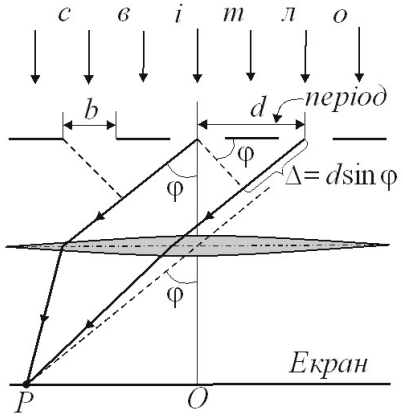


Рис. 13.21. Схема спостереження дифракції на дифракційній ґратці

Розмістимо паралельно до ґратки збиральну лінзу, у фокальній площині якої поставимо екран. З'ясуємо характер дифракційної картини, яку отримують на екрані в разі падіння на ґратку плоскої світлової хвилі (для простоти будемо вважати, що хвиля падає на ґратку нормально). Кожна із щілин дасть на екрані картину, яка описується кривою, зображеною на рис. 13.20. Картини від усіх щілин надходитимуть на одне й те саме місце екрана (незалежно від положення щілини центральний максимум лежить навпроти центра лінзи).

Якщо коливання, які надходять у точку P від різних щілин, були б некогерентними, то результуюча картина від N щілин відрізнялась би від картини, яка створюється однією щілиною, тільки тим, що всі значення інтенсивності зросли би в N разів. Утім, коливання від різних щілин є більшою чи меншою мірою когерентними, тому результуюча інтенсивність буде відмінна від NI_φ (I_φ – інтенсивність, що створюється однією щілиною, див. формулу (13.17)).

Надалі ми будемо вважати, що світлові коливання падаючої хвилі по всьому її поперечному перерізу є когерентними й переріз набагато перевищує довжину ґратки, так що коливання від усіх щілин можна вважати когерентними одне відносно одного. У цьому випадку результуюче коливання в точці P, положення якої визначається кутом φ, буде сумою N коливань з однаковими амплітудами A_φ (13.15), зсунутими одне відносно одного за фазою на одну й ту ж величину δ.

Результатом такої багатопроменевої інтерференції є такий розподіл інтенсивності на екрані

$$I_{грат} = I_{\varphi}\frac{\sin^2(N\delta/2)}{\sin^2(\delta/2)}. \quad (13.21)$$

Із рис. 13.21 видно, що різниця ходу від сусідніх щілин $\Delta = d\sin\varphi$. Отже, різниця фаз

$$\delta = \frac{2\pi}{\lambda}\Delta = \frac{2\pi}{\lambda}d\sin\varphi, \quad (13.22)$$

де λ – довжина хвилі в досліджуваному середовищі.

Підставивши у формулу (13.21) вираз (13.17) для I_{φ} та (13.22) для δ, отримаємо, що

$$I_{грат} = I_0\frac{\sin^2((\pi b\sin\varphi)/\lambda)}{((\pi b\sin\varphi)/\lambda)^2}\times \times\frac{\sin^2((N\pi d\sin\varphi)/\lambda)}{\sin^2((\pi d\sin\varphi)/\lambda)}, \quad (13.23)$$

де I_0 – інтенсивність, що створюється однією щілиною навпроти центра лінзи. Перший множник у (13.23) перетворюється в нуль у точках, для яких виконується умова (13.16). У цих точках інтенсивність, яка створюється кожною із щілин окремо, дорівнює нулю.

Другий множник в (13.23) має значення N^2 у точках, які задовільнять умови

$$\boldsymbol{d\sin\varphi = m\lambda} \ (m = 0, \pm1, \pm2, \dots). \quad (13.24)$$

Умова (13.24) визначає положення максимумів інтенсивності, які називають ***головними***. Число m визначає ***порядок головного максимуму***. Максимум нульового порядку тільки один, максимумів 1-го, 2-го і т.д. порядків є по два. Із (13.23) випливає, що інтенсивність головних максимумів в N^2 раз більша за інтенсивність I_{φ}, яку дає одна щілина у напрямку, що визначається кутом φ.

Крім мінімумів, які визначаються умовою (13.16), у проміжку між сусідніми головними максимумами є по $(N-1)$-му додатковому мінімуму. Ці мінімуми виникають у тих напрямках, для яких коливання від окремих щілин взаємно гасять одне одного. Умовою додаткових мінімумів є

$$d\sin\varphi = \pm\frac{k'}{N}\lambda, \quad (13.25)$$

де $k' = 1,2,\dots,N-1,N+1,\dots,2N-1, 2N+1,\dots$.

У формулі (13.25) k' набуває всіх цілочисельних значень, крім 0, N, 2 N, … , тобто тих, за яких умова (13.25) переходить у (13.24). Отже, між двома сусідніми головними максимумами міститься $N-1$ додаткових мінімумів і $N-2$ вторинних максимумів. На них накладатимуться мінімуми, що виникають під час дифракції від однієї щілини. Із формул (13.16) і (13.24) видно, що головний максимум m-го порядку збігається з k-м мінімумом від однієї щілини, якщо виконується рівність

$$\frac{k}{b}=\frac{m}{d},\ \text{або}\ \ \frac{m}{k}=\frac{d}{b}.$$

На рис. 13.22 показано розподіл інтенсивності I_φ від $\sin\varphi$ для $N=4$ і $\frac{d}{b}=3$. Пунктирна крива, що проходить через вершини головних максимумів, зображує інтенсивність, яка зумовлена дифракцією на одній щілині. Як видно з рис. 13.22, за відношення $d/b=3$ головні максимуми 3-го, 6-го та ін. порядків збігаються з мінімумами інтенсивності від однієї щілини, тому ці максимуми зникають. Чим більше щілин N, тим більша кількість світлової енергії пройде через ґратку, тим більше мінімумів утворюється між сусідніми головними максимумами, тим інтенсивнішими і гострішими будуть максимуми. Оскільки $\sin\varphi$ не може бути більшим від одиниці, то кількість головних максимумів $m \le d/\lambda$.

Якщо дифракційну ґратку освітлюють білим світлом, то для різних значень λ положення всіх головних максимумів, крім центрального, не збігаються один з одним, тому центральний максимум має вигляд білої смужки, а всі інші – райдужних смужок, які називають дифракційними спектрами першого, другого та ін. порядків. У межах кожної смужки забарвлення змінюється від фіолетового біля внутрішнього краю, який найближчий до максимуму нульового порядку, до червоного біля зовнішнього краю. На рис. 13.23 зображено центральний і два головні максимуми й зазначено їхнє забарвлення ($ф$ – фіолетовий, $ч$ – червоний).

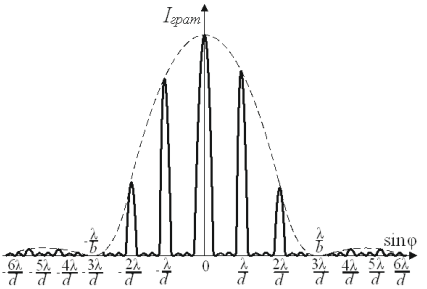


Рис. 13.22. Розподіл інтенсивності I_φ від $\sin\varphi$ для дифракційної гратки з $N=4$ і $d/b=3$

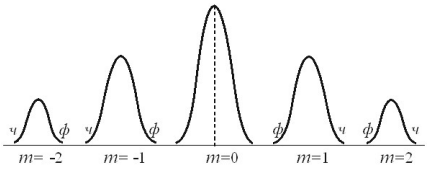


Рис. 13.23. Розклад білого світла по спектру за допомогою дифракційної гратки

Кутовою дисперсією називають величину

$$D_\varphi=\frac{\delta\varphi}{\delta\lambda},$$

де $\delta\varphi$ – кутова відстань між двома спектральними лініями, яким відповідають довжини хвиль λ та $\lambda+\delta\lambda$ (рис. 13.24).

Продиференціювавши формулу (13.24) за λ за сталого m, отримаємо $d\cos\varphi\cdot\delta\varphi=m\delta\lambda$. Звідси

$$D_\varphi = \frac{\delta\varphi}{\delta\lambda} = \frac{m}{d\cos\varphi}. \quad (13.26)$$

Якщо кути дифракції φ малі, то $\cos\varphi \approx \approx 1$ і

$$D_\varphi = \frac{m}{d} = m\frac{N}{l}, \quad (13.27)$$

де l – довжина робочої ділянки ґратки, яка містить *N* штрихів.

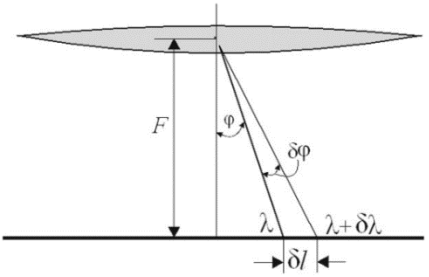


Рис.13.24

Із виразу (13.27) видно, що кутова дисперсія обернено пропорційна періоду ґратки *d* і чим вищий порядок спектру m, тим більша дисперсія.

Лінійною дисперсією називають величину

$$D_l = \frac{\delta l}{\delta\lambda}, \quad (13.28)$$

де δl – лінійна відстань на екрані між двома максимумами одного й того самого порядку m для хвиль λ і $\lambda + \delta\lambda$ (рис.13.24). Якщо фокусна відстань лінзи, у фокальній площині якої спостерігається дифракційна картина дорівнює *F*, то $\delta l \approx F\delta\varphi$ і

$$D_l = F\frac{\delta\varphi}{\delta\lambda} = FD = F\frac{m}{d}. \quad (13.29)$$

Дисперсійною областю спектрального приладу, зокрема дифракційної ґратки, називається ширина спектрального інтервалу $\Delta\lambda$, в якому спектри не перекриваються. Нехай довжини світлових хвиль, що падають на решітку знаходиться в інтервалі від λ до $\lambda_1 = \lambda + \Delta\lambda$. Інтервал $\Delta\lambda$ буде дисперсійною областю ґратки тоді, коли лівий край спектра $(m+1)$-го порядку для довжини хвилі λ збігатиметься із правим краєм спектра *m*-го порядку для довжини хвилі λ_1.

Цю умову згідно з (13.24) можна записати так: $d\sin\varphi = m\lambda_1$, $d\sin\varphi = (m+1)\lambda$. Звідси $m\lambda_1 = (m+1)\lambda$, $\lambda_1 - \lambda = \Delta\lambda = \lambda/m$.

Найменша різниця довжин хвиль двох спектральних ліній $\delta\lambda$, за якої спектральний прилад розділяє їх окремо, називають *спектральною розділяючою відстанню*, а величину

$$R = \frac{\lambda}{\delta\lambda}$$

роздільною здатністю приладу.

Для дифракційної ґратки Релей запропонував такий критерій спектрального розділення: спектральні лінії з довжинами хвиль λ та $\lambda_1 = \lambda + \delta\lambda$ вважають розділеними, якщо головний максимум дифракційної картини для хвилі з довжиною λ збігається за розміщенням з першим дифракційним

мінімумом того самого порядку для хвилі з довжиною λ_1, а інтенсивність у проміжку між максимумами становить не більше, ніж 80 % від інтенсивності максимуму (рис. 13.25).

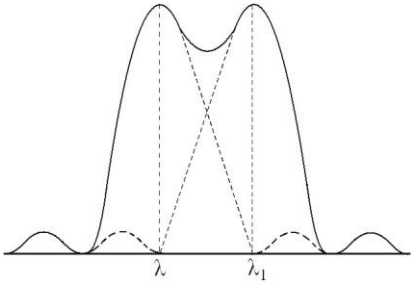


Рис. 13.25. Критерій Релей для спектрального розділення: спектральних ліній

Нехай головний максимум m-го порядку для хвилі з $\lambda_1 = \lambda + \delta\lambda$ (формула (13.24)) перебуває на місці першого мінімуму (формула (13.25), $k' = 1$) спектра того самого порядку для хвилі з λ:

$$d\sin\varphi = m\lambda_1, \quad d\sin\varphi = \left(m + \frac{1}{N}\right)\lambda.$$

Звідси

$$m(\lambda + \delta\lambda) = \left(m + \tfrac{1}{N}\right)\lambda,$$

тобто

$$\delta\lambda = \frac{\lambda}{mN}.$$

Тоді роздільна здатність ґратки

$$R = \frac{\lambda}{\delta\lambda} = mN. \qquad (13.30)$$

Отже, роздільна здатність ґратки пропорційна порядку спектру m і числу N щілин (штрихів). Оскільки $N = l/d$, а $m = (d\sin\varphi)/\lambda$, то

$$R = mN = \frac{l}{d} \cdot \frac{d\sin\varphi}{\lambda},$$

$$R_{max} = \frac{l\sin\varphi_{\max}}{\lambda},$$

де $\varphi_{\max}$ – максимальний кут дифракції ($\varphi_{\max} \le 90^\circ$), тому максимальна роздільна здатність ґратки буде

$$R_{max} = (\lambda/\delta\lambda)_{\max} \le l/\lambda.$$

Таким чином, роздільна здатність дифракційної ґратки, як видно із (13.30), пропорційна порядку спектра m і кількості щілин ґратки N.

На рис. 13.26 зіставлені дифракційні картини, отримані для двох спектральних ліній за допомогою граток, які відрізняютьсязначенням дисперсії D_φ і роздільної здатності R. Ґратки I та II мають однакову роздільну здатність R (у них однакова кількість щілин N), але різну дисперсію (у ґратки I період d у два рази більший, відповідно, дисперсія D_φ у два рази менша, ніж у ґратки II). Ґратки II та III мають однакову дисперсію (у них однакові d), але різну R (у ґратки II кількість щілин N і роздільна здатність R у два рази більша, ніж у ґратки III).

Дифракційну ґратку використовують як для пропускання світла, так і для його відбивання. Прозорі ґратки виготовляють із скляних або кварцових пластин, на поверхню яких за допомогою спеціальної машини алмазним різцем

наносять ряд паралельних штрихів. Щілинами є проміжки між штрихами.

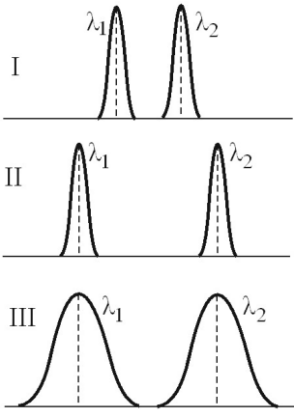


Рис. 13.26

Ґратки, що відбивають світло, отримують нанесенням алмазним різцем штрихів на поверхню металічного дзеркала. Світло падає на таку ґратку похило. При цьому ґратка з періодом d діє так, як у разі паралельного падіння діяла би прозора ґратка з періодом $d\cos\theta$, де θ – кут падіння. Кращі ґратки мають 1200 штрихів на 1 мм ($d \approx 0{,}8$ мкм). Оскільки кількість головних максимумів $m \le d/\lambda$, то спектрів другого порядку у видимому світлі за такого періоду не спостерігають. Загальна кількість штрихів у таких ґратках сягає 200 тисяч (довжина близько 200 мм). Для фокусної відстані приладу $F = 2$ м довжина видимого спектра 1-го порядку становить більше 0,7 м.

13.8. Дифракція рентгенівського випромінювання

Рентгенівським випромінюванням є електромагнітні хвилі, довжина яких лежить в інтервалі 10^{-8} ... 10^{-12} м (див. підрозд. 11.1). Залежно від умов отримання рентгенівське випромінювання утворює суцільний або лінійчастий (характеристичний) спектр. Для спостереження дифракції на ґратці потрібно, щоб її період був того самого порядку, що і довжина хвилі падаючого випромінювання.

Щоб зрозуміти, як відбувається дифракція рентгенівських променів, поставимо дві дифракційні ґратки одна за одною так, щоб штрихи були взаємно перпендикулярними. Перша ґратка, штрихи якої будемо вважати вертикальними, дасть у горизонтальному напрямку ряд максимумів, положення яких визначають умовою (13.24) (тут кут дифракції φ зазвичай позначають літерою θ):

$$d_1 \sin\theta_1 = m_1\lambda \ (m_1 = 0,\ \pm1,\ \pm2,\ \ldots). \quad (13.31)$$

Друга ґратка (з горизонтальними штрихами) розіб'є кожний із пучків, що йдуть від першої ґратки, на розміщені по вертикалі максимуми, положення яких теж визначають умовою (13.24):

$$d_2 \sin\theta_2 = m_2\lambda \ (m_2 = 0, \pm1, \pm2, \ldots). \quad (13.32)$$

У результаті дифракційна картина буде мати вигляд правильно розміщених плям, кожній з яких відповідають два цілочисленних індекси m_1 і m_2 (рис. 13.27).

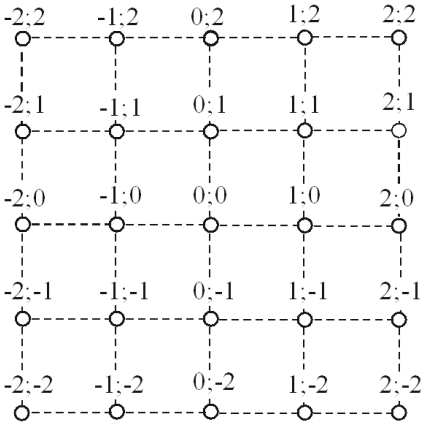


Рис. 13.27. Вигляд дифракційної картини рентгенівських променів

Таку саму дифракційну картину отримують, якщо замість двох окремих ґраток взяти одну прозору пластину, на яку нанесено дві системи взаємно перпендикулярних штрихів. Подібна пластина є *двовимірною періодичною структурою* (звичайна ґратка – одновимірна структура). Вимірявши кути θ_1 і θ_2, які визначають положення максимумів, і знаючи довжину хвилі λ можна знайти за формулами (13.31) і (13.32) періоди структури d_1 і d_2.

Дифракційну картину, аналогічну тій, що зображена на рис. 13.27, дають будь-які двовимірні періодичні структури, наприклад система невеликих отворів або система непрозорих маленьких кульок.

Для виникнення дифракційних максимумів потрібно, щоб період структури d був більший за λ. Інакше умови (13.31) і (13.32) можуть задовольнятись тільки за значень m_1 і m_2, які дорівнюють нулю (модуль $\sin\theta$ не може перевищувати одиницю).

Дифракція спостерігається також на тривимірних структурах, тобто просторових утвореннях, які мають періодичність за трьома напрямками, що не лежать в одній площині. Подібними структурами є всі кристалічні тіла. Для кристалів умова $d > \lambda$ виконується для рентгенівських променів. Вперше дифракцію рентгенівських променів спостерігали у 1913 р. у дослідах Лауе, Фрідріха і Кніппінга.

Для рентгенівських променів ідеальними природними дифракційними ґратками є монокристали, в яких атоми й іони розмішені на відстані 10^{-10} м. Кристалографічні дослідження показали, що у будь-якому кристалі можна виявити певні площини, де атоми або іони, які утворюють його кристалічну ґратку, розміщені найбільш густо. Такі площини

відбиватимуть монохроматичне рентгенівське випромінювання, яке може інтерферувати, відбившись від різних площин.

На рис. 13.28 зображено сусідні площини кристала AA_1 і BB_1. Абсолютний показник заломлення всіх середовищ для рентгенівських променів близький до одиниці, тому оптична різниця ходу Δ між двома променями *1′* і *2′*, які відбиваються від площин AA_1 і BB_1, дорівнює

$$\Delta = MO + ON = 2d\sin\theta,$$

де d – відстань між площинами, а θ – кут між площиною AA_1 та падаючими й відбитими променями, або кут ковзання. Якщо довжина хвилі рентгенівських променів λ, то інтерференційні максимуми у відбитих променях спостерігатимуться, коли

$$2d\sin\theta = m\lambda, \quad (m = \pm1, \pm2, \ldots). \quad (13.33)$$

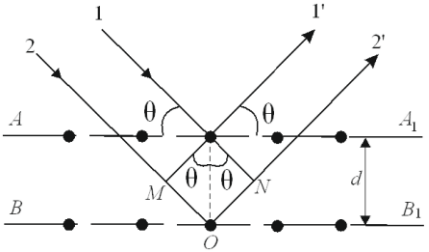


Рис. 13.28

Співвідношення (13.33) є ***формулою Вульфа–Брегга***, яке означає, що дифракція виникає не в довільному напрямку падіння монохроматичного випромінювання. Для її спостереження потрібно повертати кристал так, щоб кут ковзання задовольняв умову (13.33). Якщо обертати кристал або проводити експеримент з полікристалічною системою, в якій окремі кристалики орієнтовані довільно, то можна отримати певну систему інтерференційних картин від усіх можливих типів атомних площин певного кристала.

Дифракція рентгенівських променів є основою ***рентгеноструктурного аналізу***, який дає можливість вивчати структуру кристалів, аморфних твердих тіл, рідин, а також *рентгеноспектрального аналізу*, який використовують для вивчення рентгенівських спектрів, а також для визначення хімічного складу речовин.

Контрольні запитання

1. У чому суть явища дифракції? Що таке дифракція світла? За яких умов її спостерігають?

2. Сформулюйте й поясніть принцип Гюйгенса–Френеля.

3. Поясніть суть методу зон Френеля.

4. Поясніть утворення дифракційної картини від круглого отвору і круглого диска. Як пояснити розбіжність променя лазера?

5. Поясніть метод графічного складання амплітуд для якісного пояснення дифракційної картини.

6. Поясніть явище дифракції Фраунгофера на одній щілині. Що визначає кількість дифракційних пучків під час дифракції світла на одній щілині?

7. Що являє собою дифракційна ґратка? Якими параметрами вона характеризується?

8. Опишіть дифракційну картину від ґратки. Чим вона відрізняється від такої для однієї щілини? Як впливає дифракція від однієї щілини на дифракційну картину дифракційної ґратки?

9. Поясніть умову головних фраунгоферових максимумів під час дифракції світла на дифракційній ґратці.

10. Як визначити максимальний порядок спектра, який може дати ґратка?

11. Дайте визначення кутової та лінійної дисперсії дифракційної ґратки. Який зв'язок між лінійною і кутовою дисперсією ґратки за великих кутів дифракції?

12. Як визначають роздільну здатність дифракційної ґратки? Чим визначають її величину?

13. Якщо ґратка складається з N щілин, то у скільки разів інтенсивність у головному максимумі дифракції на екрані відрізняється від інтенсивності, що створюється в цій точці однією щілиною?

14. Від чого і як залежить ширина дифракційних максимумів?

15. Яке найменше значення періоду ґратки дозволяє з її допомогою спостерігати дифракцію?

16. Як згідно з теорією мають співвідноситися величини кутової дисперсії ґратки у довгохвильовій і короткохвильовій областях спектра?

17. Поясніть, чому дифракційна ґратка розкладає біле світло за спектром (за кольорами)?

18. Поясніть, чому кутова дисперсія ґратки майже не залежить від довжини хвилі світла?

19. Як співвідносяться величини кутової дисперсії ґратки для першого і другого порядків спектра? Відповідь поясніть.

20. Дайте визначення роздільної здатності дифракційної ґратки. Що треба знати, щоб її визначити?

21. У чому особливість дифракції рентгенівського випромінювання?

Розділ 14. Поляризація світла

14.1. Природне і поляризоване світло. Закон Малюса

Наслідком теорії Максвелла є поперечність світлових хвиль: вектори напруженості електричного $\vec{E}$ і

магнітного $\vec{H}$ полів електромагнітної хвилі взаємно перпендикулярні й коливаються перпендикулярно до вектора швидкості $\vec{v}$ поширення хвилі. Оскільки основне значення під час дії світла на речовину має вектор напруженості $\vec{E}$ електричного поля – *світловий вектор*, – що діє на електрони в атомах речовини, тому надалі будемо розглядати напрямок коливання лише цього вектора.

Розглянемо випромінювання реального джерела світла, яке складається з великої кількості атомів – елементарних випромінювачів (диполів). В основу моделі випромінювання звичайного (нелазерного) джерела світла покладено *статистичну гіпотезу* про те, що у випадку спонтанного випромінювання різні атоми джерела випромінюють окремі цуги хвиль (див. підроздію 11.6) незалежно один від одного у випадкові моменти часу. Фази коливань електромагнітного поля у випромінюванні різних атомів не скорельовані одна з одною, тому виявляється, що розподіл інтенсивності випромінювання всіх атомів джерела в такій некогерентній суперпозиції визначають підсумовуванням розподілів інтенсивності для індивідуальних атомів. Процес випромінювання окремого атома триває близько 10^{-8} с. За цей час утворюється послідовність цугів хвиль, протяжність яких становить приблизно 3 м. «Погаснувши», атом через деякий час «спалахує» знову. Одночасно «спалахує» багато атомів. Збуджені ними цуги хвиль, накладаючись один з одним, утворюють світлову хвилю, яку випромінює тіло. Площина коливань вектора $\vec{E}$ для кожного цуга орієнтована випадковим чином, тому в результуючій хвилі коливання вектора $\vec{E}$ в різних напрямках подані з однаковою ймовірністю (рис. 14.1). Такий рівномірний розподіл векторів $\vec{E}$ пояснюють великою кількістю атомарних випромінювачів, а рівність амплітудних значень векторів $\vec{E}$ – однаковою в середньому інтенсивністю випромінювання.

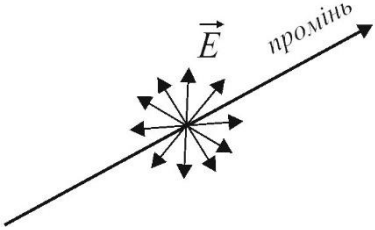


Рис. 14.1. Природне світло

Природним називають світло з усіма можливими рівноймовірними орієнтаціями вектора $\vec{E}$ (й отже, $\vec{H}$).

Поляризованим називають світло, в якому напрямки коливань вектора $\vec{E}$ якимось чином упорядковані.

Якщо коливання світлового вектора $\vec{E}$ відбуваються тільки в одній площині, яка проходить через промінь, то світло називають *плоскополяризованим* (або *лінійно поляризованим*). Упорядкованість може бути в тому, що вектор $\vec{E}$ повертається навколо променя, одночасно пульсуючи за величиною. У результаті кінець вектора $\vec{E}$ описує еліпс. Таке світло називають *еліптично поляризованим*. Якщо кінець вектора $\vec{E}$ описує коло, світло називають *поляризованим по колу.*

Будь-яке світло можна подати як суперпозицію двох плоскополяризованих у взаємно перпендикулярних площинах хвиль. Розглянемо дві монохроматичні взаємно перпендикулярні хвилі, що поширюються вздовж додатного напрямку осі *OX*:

$$E_y = E_{m1}\cos(\omega t - kx), \quad (14.1)$$

$$E_z = E_{m2}\cos(\omega t - kx + \psi), \quad (14.2)$$

де $\omega = 2\pi\nu$ – циклічна частота; $k = \omega/v$ – хвильове число; E_{m1} і E_{m2} – амплітуди хвиль; ψ – різниця фаз коливань E_y і E_z.

Щоб знайти траєкторію результуючого коливання світлового вектора під час додавання двох взаємно перпендикулярних коливань (див. підрозд. 9.10), визначимо $\cos(\omega t - kx)$ із (14.1): $\cos(\omega t - kx) = \frac{E_y}{E_{m1}}$. Тоді $\sin(\omega t - kx) = \sqrt{1 - \left(\frac{E_y}{E_{m1}}\right)^2}$. Оскільки $\cos(\omega t - kx + \psi) = \cos(\omega t - kx)\cos\psi - \sin(\omega t - kx)\sin\psi$, то

$$\frac{E_z}{E_{m2}} - \frac{E_y}{E_{m1}}\cos\psi = \sqrt{1 - \left(\frac{E_y}{E_{m1}}\right)^2}\sin\psi.$$

Піднесемо до квадрата це рівняння і в результаті отримаємо

$$\frac{E_y^2}{E_{m1}^2} + \frac{E_z^2}{E_{m2}^2} - 2\frac{E_y E_z}{E_{m1}E_{m2}}\cos\psi = \sin^2\psi. \quad (14.3)$$

Отримане співвідношення означає, що результуючий вектор $\vec{E} = \vec{E}_y + \vec{E}_z$ (точніше, кінець вектора $\vec{E}$) в кожній точці поля описує еліпс, який лежить у площині, що перпендикулярна до осі *OX* і довільно орієнтований відносно осей *OY* і *OZ*. Якщо різниця фаз ψ зазнає випадкових хаотичних змін (тобто хвилі (14.1) і (14.2) некогерентні) і якщо $E_{m1} = E_{m2}$, то напрямок світлового вектора $\vec{E}$ буде зазнавати стрибкоподібних неупорядкованих змін, а величина його буде однаковою у всіх напрямках, перпендикулярних до напрямку променя. Таке світло буде природним. Подібне уявлення природного світла значно спрощує розгляд його проходження через поляризаційні пристрої.

Якщо хвилі (14.1) і (14.2) некогерентні та різної інтенсивності ($E_{m1} \neq E_{m2}$), то в результаті суперпозиції таких хвиль отримаємо *частково поляризоване* світло. Таке

світло можна розглядати як суміш природного і плоскополяризованого.

Якщо хвилі (14.1) і (14.2) когерентні, це означає, що різниця фаз ψ не залежить від часу і $\psi = \text{const}$.

Якщо $\psi = \pm (2m+1)\frac{\pi}{2}$, ($m = 0, 1, 2\ldots$), то із (14.3) отримаємо рівняння еліпса, орієнтованого відносно осей *OY* і *OZ*:

$$\frac{E_y^2}{E_{m1}^2} + \frac{E_z^2}{E_{m2}^2} = 1. \qquad (14.4)$$

Таку хвилю (тобто світло) називають еліптично поляризованою.

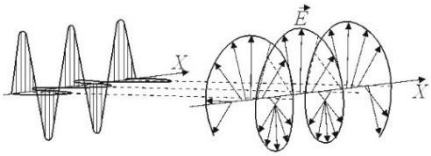


Рис. 14.2. Циркулярна поляризація світла

Якщо $E_{m1} = E_{m2} = E_m$, еліпс перетворюється на коло. Таку хвилю (світло) називають *циркулярно поляризованою* (*поляризованою по колу*).

Схематично на рис. 14.2 зображено циркулярно поляризовану хвилю. Якщо щодо напрямку променя вектор $\vec{E}$ обертається проти годинникової стрілки, поляризацію називають *лівою*, якщо у протилежному напрямку – *правою*. Якщо $\delta = \pm m\pi$ ($m = 0, 1, 2, \ldots$), то згідно з (14.3) еліпс вироджується у пряму:

$$\left(\frac{E_y}{E_{m1}} \mp \frac{E_z}{E_{m2}}\right)^2 = 0,$$

$$E_y = \pm \frac{E_{m1}}{E_{m2}} E_z. \qquad (14.5)$$

Таку хвилю називають *лінійно поляризованою* (*плоскополяризованою*) (рис. 14.3).

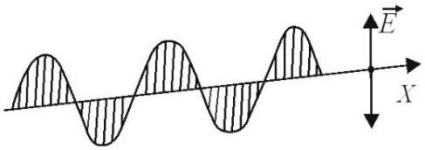


Рис. 14.3. Плоско поляризована хвиля

Площина, в якій відбувається коливання вектора $\vec{E}$, називається ***площиною поляризації***, а перпендикулярна до неї площина – ***площиною коливань***.

Кількісною характеристикою поляризації служить *ступінь поляризації*. Щоб його визначити, треба провести у площині, яка перпендикулярна променю світла і в якій лежать електричні вектори, декартову систему координат, спроєктувати амплітудні значення векторів $\vec{E}$ з однієї чверті на ці осі. Потім знайти суми проєкцій $\sum E_x$ та $\sum E_y$. Оскільки інтенсивність світла пропорційна квадрату амплітуди, то $I_x = (\sum E_x)^2$ та $I_y = \left(\sum E_y\right)^2$. Тоді ступінь поляризації визначають виразом

$$P = \frac{|I_x - I_y|}{I_x + I_y}. \qquad (14.6)$$

Для лінійно поляризованого світла $I_x = 0$. Тоді $P = 1$. Для природного світла $I_x = I_y$, оскільки

$\sum E_x = \sum E_y$, тому $P = 0$. Для частково поляризованого світла $0 < P < 1$.

Поляризацією світла називають виділення лінійно поляризованого або частково поляризованого світла із природного світла.

Плоскополяризоване світло можна отримати із природного за допомогою приладів, які називають *поляризаторами*. Ці прилади вільно пропускають коливання вектора $\vec{E}$, паралельні до площини поляризації, яку називають *головною площиною*, і повністю або частково затримують коливання вектора $\vec{E}$, перпендикулярні цій площині. У ролі поляризаторів можуть бути середовища, анізотропні відносно коливань вектора $\vec{E}$, наприклад кристали. Одним із природніх кристалів, які використовують як поляризатори, може бути турмалін – природний синтетичний монокристал алюмосилікату. Турмалін оптично анізотропний (подвійне променезаломлення розглянемо далі), є *дихроїчним* (грецьке *дихроїзм* означає *двокольорність*).

Прилади, за допомогою яких аналізують ступінь поляризації світла, називають *аналізаторами*.

Нехай на поляризатор падає природне світло (рис. 14.4). Виберемо природну хвилю, вектор напруженості електричного поля якої $\vec{E}_{\text{пр}}$ коливається у площині, що утворює з головною площиною поляризатора $P - P$ кут φ. При вході в поляризатор падаючу природну хвилю можна зобразити у вигляді суми двох коливань у взаємно перпендикулярних площинах:

$$E_{\text{пр}\parallel} = E_{\text{пр}} \cos\varphi, \quad E_{\text{пр}\perp} = E_{\text{пр}} \sin\varphi.$$

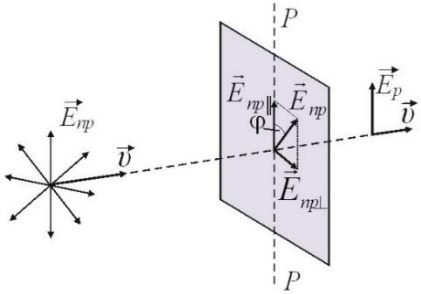


Рис. 14.4. Поляризація світла за допомогою поляризатора

Перше коливання пройде через поляризатор, друге буде затримане. Інтенсивність хвилі, що пройшла, пропорційна $E^2_{\text{пр}\parallel} = E^2_{\text{пр}} \cos^2 \varphi$, тобто

$$I_P = I_{пр} \cos^2\varphi, \qquad (14.7)$$

де $I_{пр}$ – інтенсивність хвилі з амплітудою природного світла $E_{\text{пр}}$.

У природному світлі значення φ рівноймовірні, тому частина світла, що пройшла через поляризатор, буде дорівнювати середньому значенню $\cos^2\varphi$, тобто $1/2$ і

$$I_P = I_{пр}/2. \qquad (14.8)$$

Під час обертання поляризатора навколо напрямку поширення природного світла інтенсивність світла,

що проходить через поляризатор, залишається однаковою згідно з (14.8), а змінюється лише орієнтація його площини поляризації.

Нехай на аналізатор падає лінійно поляризоване світло, отримане за допомогою поляризатора, головна площина якого $P-P$ утворює кут φ з головною площиною аналізатора $A-A$ (рис. 14.5).

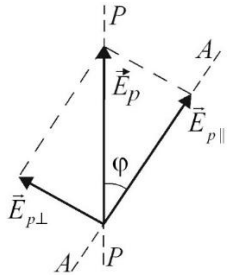


Рис. 14.5

Під час входу в аналізатор лінійно поляризованого променя амплітуду його електричного вектора $\vec{E}_p$ можна подати як суму двох лінійно поляризованих променів, площини поляризацій яких взаємно перпендикулярні. Амплітуди вектора напруженості електричного поля цих променів будуть

$$E_{P\perp} = E_P \sin\varphi \quad \text{та} \quad E_{P\parallel} = E_P \cos\varphi.$$

Оскільки аналізатор пропускає коливання електричного вектора, що відбуваються у площині $A-A$, а інтенсивність пропорційна квадрату амплітуди E_P, то інтенсивність світла після аналізатора буде

$$I_A = I_P \cos^2\varphi. \qquad (14.9)$$

Якщо аналізатор не абсолютно прозорий, то

$$I_A = k_A I_P \cos^2\varphi, \qquad (14.10)$$

де k_A – коефіцієнт прозорості аналізатора.

Отримані співвідношення (14.9) і (14.10) виражають ***закон Малюса***.

Отже, інтенсивність світла, яке пройшло через аналізатор, змінюється від мінімуму за $\varphi = \pi/2$ до максимуму за $\varphi = 0$.

Якщо пропустити природне світло через поляризатор та аналізатор, площини яких утворюють кут φ, то з першого вийде плоскополяризоване світло, інтенсивність якого $I_P = \frac{1}{2} k_P I_{np}$, де k_P – коефіцієнт прозорості поляризатора, а з другого вийде світло, інтенсивність якого $I_A = k_A I_P \cos^2\varphi$. Отже, інтенсивність світла, яке пройшло через два поляризатори, буде:

$$I_A = \frac{1}{2} k_P k_A I_{np} \cos^2\varphi.$$

Звідси $I_{max} = \frac{1}{2} I_{np} k_P k_A$ (якщо $\varphi = 0$ – поляризатори паралельні) і $I_{min} = 0$ (якщо $\varphi = \pi/2$ – поляризатори схрещені). Нехай еліптично поляризоване світло падає на аналізатор, який пропускає складову $E_{\text{пр}\parallel}$ вектора $\vec{E}_P$ за напрямком площини $A-A$ (рис. 14.6).

Максимального значення $E_{\text{пр}\parallel}$ досягають у точках 1 і 2. Отже, максимальна амплітуда $\vec{E}_A$ плоскополяризованого світла, що вийде з приладу, буде дорівнювати довжині відрізка $01'$ або $02'$. У разі повертання аналізатора навколо напрямку променя спостерігається зміна інтенсивності у межах від I_{Amax}, яку отримують за збігу осі $A-A$ з великою віссю еліпса, до I_{Amin}, яка спостерігається в разі збігу осі $A-A$ з малою віссю еліпса.

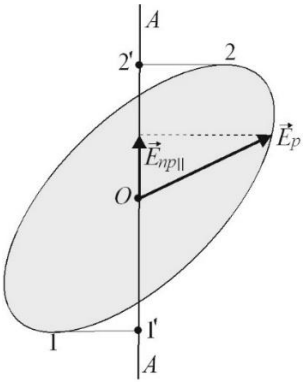


Рис. 14.6

У випадку світла, поляризованого по колу, обертання аналізатора не супроводжується зміною інтенсивності світла, яке пройшло через нього.

14.2. Поляризація під час відбивання та заломлення

Під час проходження електромагнітної хвилі з одного середовища в інше відбуваються її відбивання і заломлення. Виходячи з теорії електромагнітних хвиль Максвелла і враховуючи, що на межі середовищ тангенціальні складові векторів $\vec{E}$ і $\vec{H}$ хвилі задовільняють умови $E_{\tau_1} = E_{\tau_2}$, $H_{\tau_1} = H_{\tau_2}$, можна отримати формули для інтенсивності відбитих і заломлених хвиль. Вперше такі формули вивів Френель у 1823 р., коли розглядав проходження пружної хвилі через межу двох середовищ. Розглянемо детальніше питання відбивання і заломлення електромагнітних хвиль на межі двох діелектриків на основі аналізу формул Френеля. Нехай промінь природного світла падає під кутом α до межі двох діелектриків (наприклад «скло – повітря»), які характеризуються показниками заломлення n_1 і n_2 (чи, відповідно, діелектричними проникностями, $\varepsilon_1 = n_1^2$, $\varepsilon_2 = n_2^2$ (рис. 14.7).

Природне світло зручно подати як суперпозицію двох лінійно поляризованих у взаємно перпендикулярних площинах, некогерентних, однакової інтенсивності хвиль. Світлові вектори (вектори напруженості електричного поля) цих хвиль $\vec{E}_{i\parallel}$ і $\vec{E}_{i\perp}$. Індекс «*i*» (від слова incident – падаючий) означає промінь, що падає, індекс «$\parallel$» означає, що вектор $\vec{E}_{i\parallel}$ лежить у площині рисунка, індекс «$\perp$» означає, що вектор $\vec{E}_{i\perp}$ пендикулярний до площини рисунка і

позначений точкою. Для природного світла $|\vec{E}_{i\parallel}| = |\vec{E}_{i\perp}|$.

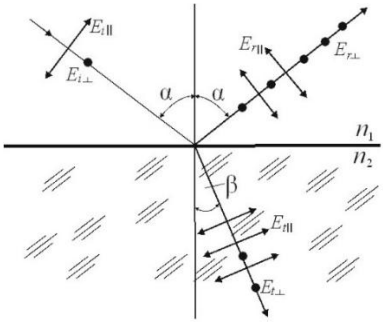


Рис. 14.7. Поляризація під час відбивання та заломлення

На рис. 14.7 індекс «*r*» (від слова reflection – відбивання) означає відбитий промінь, індекс «*t*» (від слова tranzit – проходження) означає промінь, що пройшов (заломлений промінь).

Формули Френеля встановлюють зв'язок між складовими амплітудами падаючої хвилі $E_{i\parallel}$ і $E_{i\perp}$, амплітудами відбитої хвилі $E_{r\parallel}$ і $E_{r\perp}$ та амплітудами $E_{t\parallel}$ і $E_{t\perp}$ хвилі, яка пройшла межу поділу двох середовищ. Ці формули такі:

$$E_{r\perp} = -E_{i\perp}\frac{\sin(\alpha-\beta)}{\sin(\alpha+\beta)}, \quad (14.11)$$

$$E_{r\parallel} = E_{i\parallel}\frac{\operatorname{tg}(\alpha-\beta)}{\operatorname{tg}(\alpha+\beta)}, \quad (14.12)$$

$$E_{t\perp} = E_{i\perp}\frac{2\sin\beta\cos\alpha}{\sin(\alpha+\beta)}, \quad (14.13)$$

$$E_{t\parallel} = E_{i\parallel}\frac{2\sin\beta\cos\alpha}{\sin(\alpha+\beta)\cos(\alpha-\beta)}, \quad (14.14)$$

де α – кут падіння; β – кут заломлення променя. Співвідношення між кутами α та β задають законом заломлення:

$$n_1\sin\alpha = n_2\sin\beta. \quad (14.15)$$

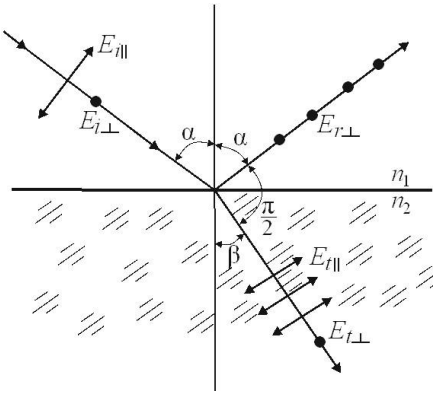


Рис. 14.8. Поляризація при падінні променя світла під кутом Брюстера

Особливу цікавість має випадок, коли $\alpha+\beta = \pi/2$, тобто відбитий і заломлений промені взаємно перпендикулярні. Оскільки $\operatorname{tg}(\alpha+\beta) = \operatorname{tg}(\pi/2) = \infty$, то із формули (14.12) видно, що $E_{r\parallel} = 0$, тобто відбитий промінь є лінійно поляризованим і містить тільки перпендикулярні до площини падіння складові напруженості електричного поля $E_{r\perp}$ (рис. 14.8).

Кут падіння α, за якого $\alpha+\beta = \pi/2$, називають кутом Брюстера $\alpha_Б$, і закон (14.15) при цьому можна переписати так:

$$\operatorname{tg}\alpha_Б = n_2/n_1, \quad (14.16)$$

його називають ***законом Брюстера***.

Таким чином, у разі падіння природного світла на межу поділу двох середовищ під кутом Брюстера $\alpha_Б$ відбите світло теоретично має бути повністю поляризованим, проте на практиці це не так. Навіть у разі падіння променя світла під кутом $\alpha_Б$ у відбитому світлі $E_{r\parallel} \neq 0$, і це означає, що воно є частково поляризованим, оскільки $E_{r\perp} > > E_{r\parallel}$. Формули (14.13) і (14.14) показують, що після проникнення у друге середовище фаза хвилі не зазнає стрибків (знаки $\vec{E}_{i\perp}$ і $\vec{E}_{t\perp}$, $\vec{E}_{i\parallel}$ і $\vec{E}_{t\parallel}$ однакові).

Фазові співвідношення між відбитою і падаючою хвилями залежать від співвідношення між показниками заломлення n_1 і n_2 першого і другого середовищ, а також від співвідношення між кутом падіння α і кутом Брюстера $\alpha_Б$. Якщо $\alpha < \alpha_Б$, то відбивання від оптично більш густого середовища $(n_2 > n_1)$ супроводжується стрибком фази на π («втрата півхвилі» – знак мінус у формулі (14.11)); відбивання від оптично менш густого середовища $(n_2 < n_1)$ відбувається без зміни фази. У випадку, коли $\alpha > \alpha_Б$, фаза для обох компонентів відбитої хвилі змінюється на π.

Фізичний зміст закону Брюстера. Хвиля, яка падає, збуджує в середовищі II (рис. 14.9) коливання електронів, які стають джерелом вторинних хвиль; ці хвилі й дають відбите світло. Напрямок коливань збігається з напрямком електричного вектора світлової хвилі, тобто для середовища II він перпендикулярний до *OC*.

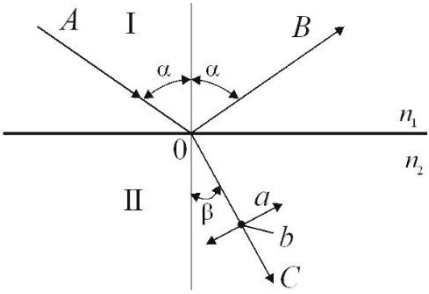


Рис. 14.9

Ми можемо подати це коливання як суму двох коливань, одне з яких (*a* на рис. 14.9) лежить у площині *AOC*, а інше (*b* на рис. 14.9) – до нього перпендикулярне. Інакше кажучи, ми зображуємо коливання електронів у молекулі як суперпозицію коливань двох елементарних випромінювачів, осі яких напрямлені, відповідно, за *a* та *b*.

Якщо світло падає під кутом Брюстера, тобто $\alpha + \beta = \pi/2$, то очевидно, що $OB \perp OC$. Отже, $OB \parallel a$. Відомо, однак, що електричний заряд не випромінює електромагнітних хвиль уздовж напрямку свого руху, тому випромінювач типу a вздовж OB не випромінює. Таким чином, за напрямком OB поширюється світло, яке посилає

випромінювач типу b, напрямок коливань яких перпендикулярний до OB, тобто перпендикулярний до площини рис. 14.9.

Інакше кажучи, відбите світло поляризоване, і коливання вектора напруженості електричного поля в ньому перпендикулярні до площини падіння (закон Брюстера).

Якщо кут падіння відрізняється від кута Брюстера, то вздовж OB може поширюватись хвиля, яка крім компоненти b містить і компоненту a, частка якої буде тим більша, чим більший кут між напрямком a і напрямком відбитої хвилі. Таким чином, відбите світло буде частково поляризованим і ступінь поляризації зростає з наближенням до кута Брюстера.

Проаналізуємо коефіцієнти відбивання для $E_\perp$ і $E_\parallel$. Згідно з формулами (14.11) і (4.12) коефіцієнт відбивання для $E_\perp$:

$$r_\perp^2 = \left(\frac{E_{r\perp}}{E_{i\perp}}\right)^2 = \left(\frac{\sin(\alpha-\beta)}{\sin(\alpha+\beta)}\right)^2, \qquad (14.17)$$

для $E_\parallel$:

$$r_\parallel^2 = \left(\frac{E_{r||}}{E_{i||}}\right)^2 = \left(\frac{\mathrm{tg}(\alpha-\beta)}{\mathrm{tg}(\alpha+\beta)}\right)^2. \qquad (14.18)$$

Для випадку нормального падіння світла на межу поділу двох середовищ ($\alpha = \beta = 0$) із формул (14.11) і (14.12), розкриваючи невизначеність, знаходимо амплітудні коефіцієнти відбивання:

$$r_\perp = \frac{E_{r\perp}}{E_{i\perp}} = -\frac{n_2-n_1}{n_2+n_1}, \quad r_\parallel = \frac{E_{r||}}{E_{i||}} = \frac{n_2-n_1}{n_2+n_1}. \qquad (14.19)$$

Для $n_2 = 1{,}5$, $n_1 = 1$ («скло – повітря») знаходимо, що

$$r_\perp^2 = r_\parallel^2 = 1/25 = 4\ \%.$$

Відбивання світла від багатьох поверхонь, навіть якщо падіння близьке до нормального, може значно послабити інтенсивність світла, що слід враховувати під час побудови складних оптичних систем.

Якщо кут падіння $\alpha = \alpha_Б$, тоді $\mathrm{tg}(\alpha+\beta) \to \infty$, амплітудний коефіцієнт відбивання $r_\parallel$ для $E_{r\parallel}$ дорівнює нулю, а $r_\perp$ визначають за формулою (14.11):

$$r_\perp = \frac{n_2^2-n_1^2}{n_2^2+n_1^2}. \qquad (14.20)$$

Таким чином, за $\alpha = \alpha_Б$ відбите світло лінійно поляризоване у площині, яка перпендикулярна площині падіння, а коефіцієнт відбивання («скло – повітря»)

$$r_\perp^2 = \left(\frac{n_2^2-n_1^2}{n_2^2+n_1^2}\right)^2 = 15\ \%.$$

Якщо $\alpha \to \infty$ (ковзаюче падіння), то $r_\parallel^2 = r_\perp^2 = 1$, тобто відбувається повне відбивання світла. Цим пояснюють яскраві зображення предметів у спокійній воді (береги річок, ліхтарі, місячна доріжка, сонце, яке заходить і под.).

На рис. 14.10 зображено графіки залежності $r_\perp^2$ і $r_\parallel^2$ (криві I та III) від кута падіння α для $n_2 = 1{,}52$ й $n_1 = 1$. Кут

Брюстера при цьому дорівнює 56°40′. Крива II відповідає коефіцієнту відбивання для неполяризованого світла. У цьому випадку $E_{i\parallel}^2 = E_{i\perp}^2$, і коефіцієнт відбивання

$$R = \frac{I_r}{I_i} = \frac{1}{2}(r_\parallel^2 + r_\perp^2), \qquad (14.21)$$

де I_i – інтенсивність падаючого світла; I_r – інтенсивність відбитого світла. Із (14.21) видно, що коефіцієнт відбивання дорівнює середньому арифметичному із $r_\perp^2$ і $r_\parallel^2$.

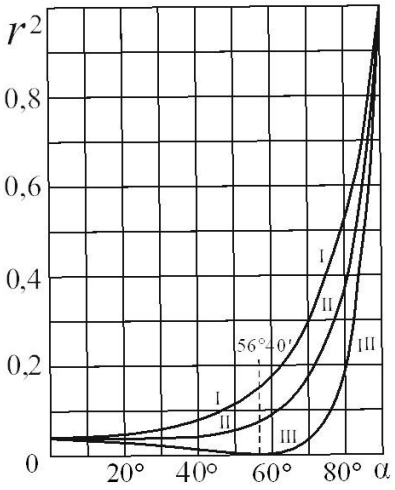


Рис. 14.10. Відбивання природного та поляризованого світла від скляної поверхні

Для природного світла, яке падає на пластинку, $E_{i\parallel}^2 = E_{i\perp}^2$. Для відбитого світла інтенсивність $I_{r\perp} = \langle E_{i\perp}^2 \rangle \neq I_{r\parallel} = \langle E_{i\parallel}^2 \rangle$, тому відбите світло буде частково поляризованим. За міру його ступеня поляризації природно взяти співвідношення

$$P_r = \frac{I_{r\perp} - I_{r\parallel}}{I_{r\perp} + I_{r\parallel}} \cdot 100\ \%,$$

де $I_{r\perp}$ та $I_{r\parallel}$ – інтенсивність відбитого світла, що відповідає компонентам $E_{r\perp}$ та $E_{r\parallel}$ відповідно.

Відбивання під кутом Брюстера дає змогу отримати лінійно поляризоване світло, однак його інтенсивність невелика і для скла (n = 1,5) дорівнює близько 15 %, тобто основна його частина поширюється у напрямку заломленої хвилі, яка поляризована частково.

Справді, із формул (14.13) і (14.14) отримаємо, що для природного світла

$$\frac{E_{t||}}{E_{t\perp}} = \frac{1}{\cos(\alpha - \beta)} = \frac{1}{\sin 2\alpha_Б} > 1. \qquad (14.22)$$

Щоб отримати (14.22) слід враховувати, що $E_{i\perp} = E_{i||}$ й $\alpha + \beta = \pi/2$. Формула (14.22) показує, що $E_{t\parallel} > E_{t\perp}$. Для збільшення ступеня поляризації заломлених хвиль їх треба пропустити крізь стопу скляних пластинок (рис. 14.11).

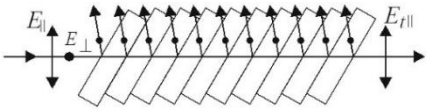


Рис. 14.11. Стопа скляних пластинок

Так, для $\alpha = \alpha_Б$ стопа з десяти скляних пластинок дає змогу отримати

майже стовідсоткову поляризацію заломлених хвиль.

14.3. Подвійне променезаломлення

Проходження світла крізь анізотропну речовину, оптичні властивості якої в різних напрямках неоднакові, супроводжуються рядом своєрідних явищ, які мають принципове практичне значення. Особливості оптичних явищ в анізотропних середовищах пов'язані з тим, що індукований електромагнітною хвилею дипольний момент фізично нескінченно малого об'єму середовища, взагалі кажучи, не співпадає за напрямком з електричним полем хвилі. Так відбувається тому, що в анізотропній речовині під дією зовнішньої сили елементарні заряди зміщуються в одних напрямках легше, ніж в інших.

Фізична природа анізотропії речовини пов'язана з особливостями будови її молекул або особливостями кристалічної ґратки, у вузлах якої містяться атоми або іони.

Після падіння світлової хвилі на межу *ізотропного середовища* в цьому середовищі від межі поширюється одна хвиля, яка лежить у площині падіння і яка складає з нормаллю до межі середовищ кут β, який визначається відомим законом заломлення: $n_1 \sin\alpha = n_2 \sin\beta$, де α – кут падіння; n_1 і n_2 – показники заломлення середовищ. Якщо середовище анізотропне, то в ньому в загальному випадку виникають дві хвилі, які поширюються від межі в різних напрямках і з різними швидкостями. Це явище називають *подвійним променезаломленням.* Його відкрив Бартолін у 1670 році у кристалах *ісландського шпату*, згодом його детально дослідив Гюйгенс – він дав пояснення подвійного променезаломлення на основі гіпотези про те, що хвиля, яка падає на межу середовищ, породжує у кристалі елементарні вторинні хвилі двох видів: сферичні (*звичайні*) й еліптичні (*незвичайні*), швидкість поширення яких залежить від напрямку поширення. Ісландський шпат – це різновид вуглекислого кальцію $CaCO_3$ (кальцит), який належить до гексагональної системи. У природі він трапляється у вигляді досить великих оптично чистих кристалів. Розколюючи його по певних площинах, кристалу можна надати форму ромбоедра – фігури, яку найлегше уявити собі як куб, дещо стиснутий вздовж просторової діагоналі AB (рис. 14.12). Грані його мають вигляд ромбів з кутами 78° і 102°. У двох протилежних вершинах A і B сходяться сторони трьох тупих кутів.

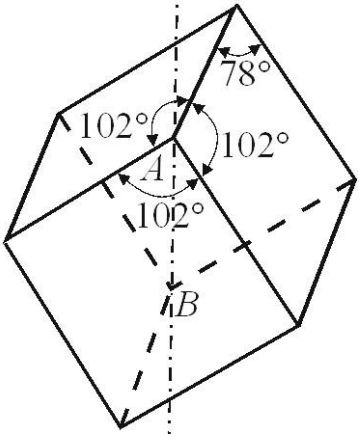


Рис. 14.12 Кристал ісландського шпату

Для демонстрації подвійного променезаломлення вузький паралельний пучок світла направляють перпендикулярно до грані природного ромбоедра. Із протилежної грані виходять два пучки, які мають напрямки, паралельні початковому. Один із них є продовженням первинного, а другий зміщений у сторону, тобто для нього кут заломлення відмінний від нуля попри те, що кут падіння дорівнює нулю. Ця обставина дала привід назвати другий пучок *незвичайним* (e – extraordinary), а перший, який підпорядковується закону заломлення – *звичайним* (o – ordinary). Якщо пучок, який падає, досить вузький, а кристал має достатню товщину, то пучки, що виходять із кристала, просторово розділені (рис. 14.13) й утворюють дві плями на екрані. Якщо падаюче світло природне, то обидві плями мають однакову освітленість. Під час повороту кристала навколо напрямку падаючого променя одна пляма залишається на місці, а друга рухається навколо неї. За допомогою аналізатора легко переконатись, що пучки, які виходять із кристала, *лінійно поляризовані* у взаємно перпендикулярних напрямках (рис. 14.13).

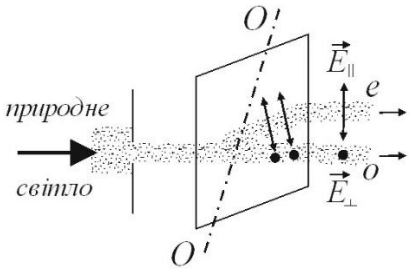


Рис. 14.13. Подвійне променезаломлення

Збільшивши переріз падаючого пучка, можна отримати часткове перекриття плям на екрані. Під час повороту аналізатора частини плям, які не перекриваються, періодично стають світлими і темними, коли в однієї частинки освітленість максимальна, а у другої освітленість відсутня. Освітленість частини плям, що перекриваються, залишається незмінною.

У кристалі ісландського шпату просторова діагональ AB ромбоедра (рис. 14.12) є *віссю симетрії*: під час обертання на 120° навколо цієї осі кристал суміщається сам із собою. Якщо спилити

тупі кути у вершинах ромбоедра по площинах, які перпендикулярні осі, й відшліфувати ці площини, можна дослідити поширення світла у напрямку осі *OO*, яка паралельна *AB* (рис. 14.14).

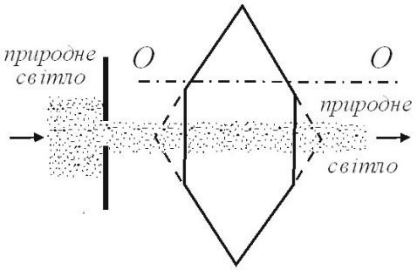


Рис. 14.14. OO – оптична вісь кристала

Виявляється, що в цьому випадку подвійного променезаломлення немає: падаючий пучок світла не роздвоюється і стан його поляризації не змінюється. Такий напрямок у кристалі, що має описану властивість, називають його *оптичною віссю*. Зазначимо, що мова йде саме про *напрямок*, а не про окрему пряму, оскільки пучок не буде зазнавати подвійного заломлення при поширенні вздовж будь-якої прямої, яка паралельна оптичній осі.

Кристали, які мають лише один напрямок, вздовж якого не відбувається подвійного променезаломлення, називають одновісними (ісландський шпат, кварц, турмалін та ін.). Є кристали, які мають два такі напрямки, їх називають двовісними (слюда, гіпс). У двовісних кристалах обидва промені незвичайні. Надалі будемо розглядати одновісні кристали, які широко використовують в оптичних експериментах.

Площину, яка проходить через промінь та оптичну вісь кристала, що перетинає промінь, називають *головною площиною*, або *головним* перерізом кристала, що відповідає цьому променю. Через кристал можна провести нескінченну множину паралельних оптичних осей і нескінченну множину паралельних головних перерізів. Лінія перетину двох довільних головних перерізів завжди є оптичною віссю.

Дослідження звичайного і незвичайного променів показує, що електричний вектор у звичайному промені перпендикулярний до площини головного перерізу, його позначають (•), а в незвичайному промені лежить у площині головного перерізу, його позначають (↕) (рис. 14.13), тобто площина поляризації звичайного променя перпендикулярна до площини головного перерізу, а незвичайного збігається із площиною головного перерізу.

Після виходу з кристала, якщо не брати до уваги поляризацію у двох взаємно перпендикулярних напрямках (площинах), ці два промені нічим не відрізняються один від одного.

Пояснення подвійного променезаломлення

Подвійне променезаломлення пояснюють анізотропією кристалів. У кристалах некубічної системи діалектрична проникність ε є залежною від напрямку дії електричного поля $\vec{E}$. В одновісних кристалах ε у напрямку оптичної осі та в напрямку, який перпендикулярний до неї, має різні значення $\varepsilon_{||}$ і $\varepsilon_{\perp}$, які називають поздовжньою і поперечною діелектричною проникністю кристала. В інших напрямках ε має проміжне значення. Оскільки $n = \sqrt{\varepsilon}$ і $v = \frac{c}{n} = \frac{c}{\sqrt{\varepsilon}}$, то з анізотропії ε випливає, що світловим хвилям з різними напрямками коливання вектора $\vec{E}$ відповідають різні значення показника заломлення n і швидкості світлових хвиль v.

Виділимо у кристалі площину головного перерізу і розглянемо промені, що виходять з деякої точки C в різних напрямках у цій площині. На рис. 14.15 напрямки, які паралельні до оптичної осі OO, зображені пунктирними лініями, а напрямки коливань електричного вектора $\vec{E}$ показані точками. Якщо напрямок звичайного променя довільний, наприклад C_{a_1}, C_{a_2}, C_{a_3}, вектор $\vec{E}$ утворює прямий кут з оптичною віссю кристала, і швидкість буде однаковою в усіх напрямках та дорівнюватиме $v_o = \frac{c}{\sqrt{\varepsilon_{\perp}}}$. Геометричним місцем точок, до яких промені, що виходять із центра C, будуть доходити за однаковий проміжок часу, буде коло (рис. 14.15). Якщо розглянути сукупність усіх можливих головних перерізів, що проходять через точку C, то геометричним місцем кінців усіх променів буде сфера.

Коливання вектора $\vec{E}$ в незвичайному промені здійснюються у площині головного перерізу (рис. 14.16), тому для різних напрямків коливання вектора $\vec{E}$ утворюють з оптичною віссю OO різні кути α. Для променя C_{a_1} кут $\alpha = \pi / 2$ і швидкість поширення коливань $v_o = c/\sqrt{\varepsilon_{\perp}}$, для променя C_{a_2} кут α = 0 і швидкість $v_e = c/\sqrt{\varepsilon_{\|}}$ Для променя C_{a_3} швидкість має проміжне значення між v_o і v_e, у такому разі хвильова поверхня має вигляд еліпсоїда обертання навколо осі, переріз якого є однією з головних площин, зображених на рис. 14.16.

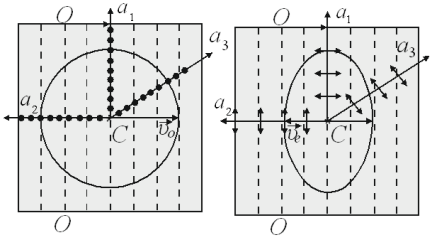


Рис. 14.15 Рис. 14.16

Одновісні кристали характеризуються показником заломлення звичай-

ного променя, який становить $n_o = c/v_o$, і показником заломлення незвичайного променя $n_e = c/v_e$. Значення n_e найбільше відрізняється від n_o для напрямку, який перпендикулярний до оптичної осі.

Залежно від того, яка із швидкостей v_o чи v_e більша, розрізняють додатні й від'ємні одновісні кристали.

Якщо $v_o > v_e$ і, відповідно, $n_o < n_e$, то кристал називають оптично додатним, а еліпсоїд вписаний у сферу (рис. 14.17). Для від'ємного кристала $v_o < v_e,\ n_o > n_e$. У такому разі еліпсоїд описаний навколо сфери (рис. 14.18).

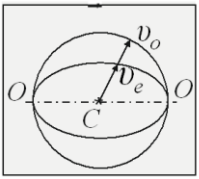


Рис. 14.17.
Додатний кристал

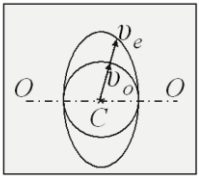


Рис. 14.18.
Від'ємний кристал

Хід звичайних і незвичайних променів у кристалі можна визначити за допомогою принципу Гюйгенса (рис. 14.19). Нехай на плоску поверхню *ab* одновісного оптично додатного кристала під кутом α падає плоска неполяризована світлова хвиля. Оптична вісь *OO* з поверхнею кристала утворює гострий кут і лежить у площині рисунка. У момент часу *t* фронт *AB* падаючої хвилі досяг точки *A* поверхні кристала. За час, поки промінь з точки *B* досягне точки *D* у кристалі, з точки *A* виникнуть сферична та еліпсоїдальна хвильові поверхні. Із проміжних точок, які лежать на поверхні кристала між *A* і *D*, також утворяться подібні поверхні, але менших розмірів.

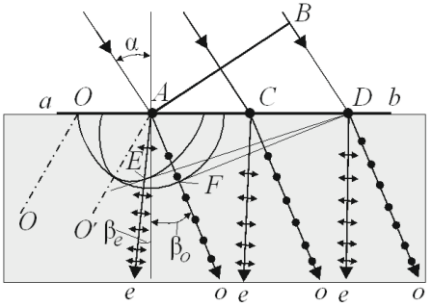


Рис. 14.19. Хід звичайних і незвичайних променів у кристалі

Згідно із принципом Гюйгенса прямі *DE* і *DF*, які дотикаються відповідно до еліпсоїдальної та сферичної поверхонь, будуть фронтами незвичайної і звичайної хвиль. Промені *Ao* і *Do* – звичайні, вони нормальні до хвильового фронту *DF*; промені *Ae* і *De* – незвичайні, вони не перпендикулярні до фронту *DE*.

На рис. 14.20 зображено три випадки нормального падіння світла на поверхню кристала, які відрізняються напрямком оптичної осі *OO*. У випадку *а* промені *o* та *e* поширюються вздовж оптичної осі, і тому поширюються, не розділяючись. Із рис. 14.20, *б* видно, що навіть за нормального падіння світла на заломну поверхню незвичайний промінь

може відхилитись від нормалі до цієї поверхні. На рис. 14.20 оптична вісь кристала паралельна заломній поверхні. У цьому випадку за нормального падіння світла звичайний і незвичайний промені поширюються в одному і тому ж напрямку, але поширюються з різною швидкістю, унаслідок чого між ними виникає зростаюча різниця ходу, а відповідно, і різниця фаз. Якщо товщина кристала d, то оптична різниця ходу

$$\Delta = (n_e - n_o)d, \qquad (14.23)$$

а різниця фаз

$$\delta = 2\pi\frac{(n_e - n_o)d}{\lambda_0}, \qquad (14.24)$$

де λ_0 – довжина хвилі у вакуумі.

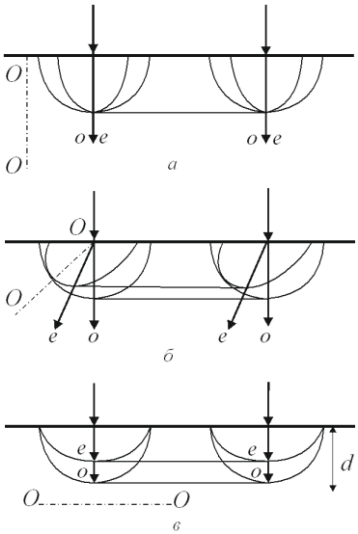


Рис. 14.20

Таким чином, якщо пропустити природне світло крізь вирізану паралельно оптичній осі кристалічну пластину, із пластини вийдуть два поляризовані у взаємно перпендикулярних площинах промені o та e, між якими буде існувати різниця фаз (14.24).

Якщо світло, яке падає на пластину є природним, то промені o та e некогерентні, тому що вони містять хвилі, які належать різним цугам, що випромінюють атоми джерела світла.

Якщо на пластину падає плоскополяризоване світло, то в цьому випадку коливання кожного цуга розділяється між звичайним і незвичайним променями в однаковій пропорції, яка залежить від орієнтації оптичної осі пластини відносно площини коливань у падаючому промені, тому промені o та e будуть когерентними і можуть інтерферувати.

Залежно від товщини пластини різниця фаз (14.24) може набувати різних значень, і світло на виході із пластини може мати різну поляризацію (у загальному випадку еліптичну) через додавання двох взаємно перпендикулярно поляризованих хвиль.

14.4. Поляризаційні призми і поляроїди

Явище подвійного променезаломлення лежить в основі роботи поляризаційних пристроїв, які служать для отримання поляризованого світла. Замість окремого кристала для цієї цілі

більш зручними є їх комбінації, які називають *поляризаційними призмами.*

Призми поділяють на два класи:

1) призми, що дають лише плоскополяризований промінь (поляризаційні призми);

2) призми, що дають два поляризовані у взаємно перпендикулярних площинах промені (двозаломні призми).

Найбільш придатний матеріал для виготовлення призм – ісландський шпат, в якого порівняно велика різниця між звичайними і незвичайним показниками заломлення (n_o = 1,658; n_e = 1,658−1,486) рідше застосовують кварц (n_o = 1,544; n_e = 1,553).

Першу поляризаційну призму винайшов Нікол у 1828 р. Вона була виготовлена з отриманого розколюванням по площинах спайності шматка ісландського шпату, довжина якого приблизно у 3,5 рази більша за товщину. Торцеві основи *AC* і *BD* (рис. 14.21) були відшліфовані під кутом 68° до його довгих ребер (замість 71° у природного кристала). Потім кристал розрізали по площині, перпендикулярній до торцевих поверхонь (розріз *AB* перпендикулярний *AC*), і відшліфовані половинки склеювали в попередньому положенні канадським бальзамом, який утворював між ними тонкий прозорий шар з показником заломлення n = 1,549, що має проміжне значення між n_o і n_e. Для звичайного променя канадський бальзам є оптично менш густим середовищем, для незвичайного – більш густим, тому звичайний промінь може пройти шар бальзаму тільки тоді, коли кут його падіння на площину розрізу менший від граничного кута повного відбивання $\alpha_{гр}$, який визначають з умови $n_0 \sin\alpha_{гр} = n \sin(\pi/2)$. У такому випадку $\sin\alpha_{гр} = n/n_o$ = 1,549/1,658 та $\alpha_{гр} = 69°$.

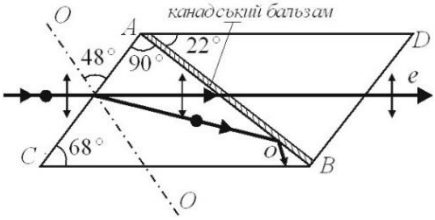


Рис. 14.21. Призма Ніколя

Оптична вісь *OO* спрямована під кутом 48° до вхідної грані *AC*. Промінь при падінні на грань *AC* унаслідок подвійного заломлення розділяється на звичайний промінь *o* і незвичайний промінь *e*. За певного вибора кутів призми звичайний промінь падає на шар бальзаму під кутом 76°, який більший за граничний (69°), зазнає повного внутрішнього відбивання, падає на зачорнену грань *CB* і нею поглинається. Незвичайний промінь виходить із призми паралельно до грані *CB*. Площина його поляризації збігається із площиною

головного перерізу. Для незвичайного променя повне внутрішнє відбивання від шару канадського бальзаму виключене ($n_e < n$).

Повне відбиття звичайного променя відбувається тільки тоді, коли кути падіння на вхідну грань призми лежать у певних межах. Кут між крайніми променями падаючого пучка (29°), який задовольняє цю умову, визначає *апертуру повної поляризації призми.*

Є багато модифікацій призми Ніколя. Для роботи в ультрафіолетовій області спектра канадський бальзам не підходить через сильне поглинання світла, тому використовують *призму Фуко* з тонким повітряним прошарком. Повне відбиття звичайного променя у площині розрізу відбувається за менших кутів, і тому призма Фуко виходить значно коротшою за призму Ніколя. Апертура повної поляризації призми Фуко становить всього 8°.

Деколи із падаючого світла треба отримати відразу два промені, які поляризовані у взаємно перпендикулярних напрямках. Призми, які використовують для цієї цілі, називають двозаломними.

Прикладом двозаломної призми є *призма Волластона*, яка складається із двох прямокутних призм з ісландського шпату (рис. 14.22).

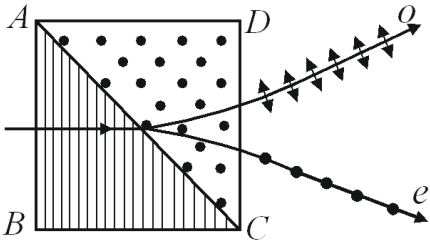


Рис. 14.22. Призма Волластона

У призмі *ABC* оптична вісь паралельна до катета *AB*, а у призмі *ACD* – до ребра *C*, яке перпендикулярне до площини рисунка. Природне світло падає нормально на грань *AB*; обидва промені, звичайний і незвичайний, що виникають у призмі *ABC*, йдуть за одним напрямком, відповідно, зі швидкістю v_o і v_e. В іншій призмі вони також поширюються в напрямку, який перпендикулярний до оптичної осі, але оскільки оптичні осі в обох призмах взаємно перпендикулярні, то звичайний промінь першої призми перетворюється на незвичайний промінь в іншій, і навпаки. Отже, звичайний промінь у першій призмі заломиться на межі обох призм з відносним показником заломлення n_e/n_o, а промінь незвичайний – з n_o/n_e. Для ісландського шпату $n_o > n_e$, й тому $n_e/n_o < 1$, а $n_o/n_e > 1$. Перший промінь заломиться у бік ребра *C* призми *ACD*, а другий – у бік її основи *AD.* Цим досягається значне розходження променів. Обидва промені плоскополяризовані.

Усі двозаломні кристали тією чи іншою мірою поглинають світло. Коефіцієнт поглинання неоднаковий для звичайного й незвичайного променів і залежить від напрямку поширення світла у кристалі. Це явище називають *дихроїзмом*. Значний дихроїзм у видимій області спектра мають кристали турмаліну, в якому коефіцієнт поглинання для звичайних променів набагато більший, ніж для незвичайних. Пластинка турмаліну завтовшки 1 мм майже повністю поглинає звичайні промені, і світло, яке проходить крізь неї, буде лінійно поляризованим. Водночас для деяких ділянок спектра незвичайний промінь теж зазнає відчутного поглинання, що обмежує застосування турмаліну як поляризатора.

Дуже зручні у використанні поляризаційні пристрої із синтетичних дихроїчних матеріалів, які відомі як *полярoїди*. Полярoїд – це плівка целулоїда або іншого прозорого матеріалу, в яку вкраплені певним чином орієнтовані мікроскопічні кристалики сильно дихроїчної речовини, найчастіше герапатиту (сульфат йодистого хініну) або інших споріднених йому сполук. Плівка завтовшки 0,1 мм повністю поглинає звичайний промінь видимої області спектра. Таким способом отримують листи великої площі, які забезпечують високий ступінь поляризації світла, яке проходить, за великих апертур і мають порівняно невелику вартість. Недоліки полярoїдів пов'язані із спектральною селективністю поглинання герапатиту, через що фіолетова частина спектра є поляризованою тільки частково, а плівка виявляється неоднаково прозорою для променів різних кольорів.

14.5. Штучна оптична анізотропія

Подвійне променезаломлення спостерігається у природних анізотропних середовищах. Є різні способи отримання штучної оптичної анізотропії, тобто надання оптичної анізотропії природно ізотропним речовинам.

Оптично ізотропні речовини стають оптично анізотропними під дією:

1) *одностороннього стиску або розтягу* (кристали кубічної системи, скла та ін.) (Т. Зеєбек, Д. Брюстер (1815 р));

2) *електричного поля* (рідини, аморфні тіла, гази) (Керр);

3) *магнітного поля* (рідини, скло, колоїди) (Коттон, Мутон).

1. Оптичну анізотропію, яка виникає під впливом деформації, можна виявити, якщо помістити досліджуване тіло T між поляризатором P і аналізатором A, які схрещені (рис. 14.23). Доки тіло не деформоване, така система світла не пропускає. Під час односторон-

нього стиску або розтягу тіла вздовж напрямку *OO* в ньому виникає оптична анізотропія, яка еквівалентна анізотропії одновісного кристала з оптичною віссю *OO*. Звичайний і незвичайний промені будуть поширюватися в напрямку, який перпендикулярний до *OO*, з різною швидкістю v_e і v_o. Якщо головний переріз поляризатора не паралельний і не перпендикулярний до *OO*, то світло, яке пройшло крізь деформоване тіло, стане еліптично поляризованим і його не можна погасити аналізатором.

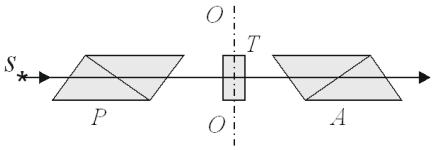


Рис. 14.23

Різниця коефіцієнтів заломлення $n_e - n_o = c/v_e - c/v_o$ може служити мірою анізотропії, що виникла. Досвід показує, що різниця $n_e - n_o$ пропорційна до напруги $\sigma = F/S$ у певній точці тіла (F – сила, S – площа, до якої перпендикулярно прикладена сила F):

$$n_e - n_o = k_1\sigma,$$

де k_1 – коефіцієнт пропорційності, що залежить від властивостей речовини. Різниця фаз, яку матимуть звичайний і незвичайний промені, пройшовши тіло завтовшки l,

$$\delta = \frac{2\pi l}{\lambda_0}(n_e - n_o) = 2\pi C\sigma l, \quad (14.25)$$

де λ_0 – довжина світлової хвилі у вакуумі, а $C = k_1/\lambda_0$ – новий коефіцієнт. Залежно від роду речовини коефіцієнт C може бути позитивним або негативним.

Явище штучної оптичної анізотропії у разі деформацій використовують для виявлення внутрішніх залишкових напруг, які можуть виникати у виробах зі скла та інших прозорих ізотропних матеріалів унаслідок порушення технології їх виготовлення. Оптичний метод вивчення на прозорих моделях розподілу внутрішніх напруг у різних напружених частинах машин і споруд широко застосовують у сучасній техніці. Для цього використовують моделі, виготовлені з целулоїду або іншої ізотропної речовини.

Оскільки величина оптичної анізотропії пропорційна до напруги σ, то за виглядом смуг однакового кольору (ізохром), що виникають під час спостереження моделі між схрещеними ніколями, можна зробити висновок про величину напруг.

Оптична анізотропія може виникнути також і в рідині під дією зовнішніх динамічних впливів. Виникнення оптичної анізотропії в потоці може бути використане для вивчення властивостей полімерів, що

йдуть на виготовлення штучних каучуків і пластмас.

2. У 1875 р. Д. Керр виявив, що рідкий або твердий ізотропний діелектрик, уміщений у дуже сильне однорідне електричне поле, стає оптично анізотропним. Це явище називають *ефектом Керра.* Принципову схему спостереження цього явища в рідинах зображено на рис. 14.24, де P і A – поляризатор і схрещений з ним аналізатор. Між ними розміщено кювету з конденсатором («комірка Керра»), між пластинами якого міститься досліджувана рідина. Якщо немає електричного поля, світло через систему не проходить. Досліди показали, що під дією однорідного електричного поля у плоскому конденсаторі рідина набуває властивостей одновісного двозаломлюючого кристала, оптична вісь якого збігається з напрямком вектора $\vec{E}$ напруженості поля конденсатора. Різниця показників заломлення рідини для звичайного і незвичайного променів монохроматичного світла в напрямку, який перпендикулярний до вектора $\vec{E}$, пропорційна E^2:

$$n_e - n_o = k_2 E^2, \qquad (14.26)$$

де k_2 – коефіцієнт пропорційності.

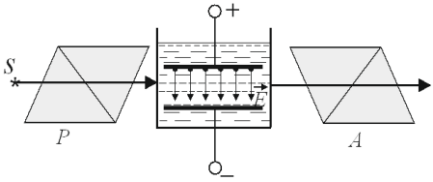


Рис. 14.24. Оптична схема ефекта Керра

Такий характер залежності (14.26) пояснюється тим, що для ізотропного середовища (коли немає поля) величина $n_e - n_o$ не може залежати від напрямку $\vec{E}$, тому розкладання $n_e - n_o$ у ряд за степенями напруженості поля має починатись із квадратичного члена. У слабких полях, які використовують на практиці, у розкладанні досить обмежитись тільки цим членом.

Якщо довжина шляху променів між обкладками конденсатора становить l, то різниця фаз між звичайними і незвичайними променями

$$\delta = \frac{2\pi}{\lambda_0}(n_e - n_o)l = \\ = \frac{2\pi}{\lambda_0} l k_2 E^2 = 2\pi l B E^2, \qquad (14.27)$$

де $B = k_2/\lambda_0$ – стала Керра, яка залежить від природи речовини, довжини хвилі λ_0, температури і швидко зменшується з її збільшенням. Часто користуються іншою константою Керра K, яка пов'язана з B співвідношенням $K = B\lambda_0/n$, n – абсолютний показник заломлення речовини за відсутності електричного поля.

У 1930 р. було виявлено існування ефекту Керра і в газах. Складність спостереження цього явища пов'язана з тим, що значення B для газів на кілька порядків менше, ніж для рідин. Для рідин стала Керра $B = (10^{-14}... 10^{-12})$ м/В^2, для газів $B = (10^{-19}... 10^{-16})$ м/В^2. Велике значення ($B = 2{,}2 \cdot 10^{-12}$ м/В^2) має нітробензол ($C_6H_5NO_2$), тому його часто використовують у технічних застосуваннях ефекту Керра.

Для більшості речовин стала Керра додатна, тобто $n_e > n_o$, що відповідає анізотропії додатного одновісного кристала. Рідше трапляються речовини, в яких $B < 0$ (етиловий ефір, спирт).

Ефект Керра пояснюють різною поляризацією молекул за різними напрямками. Коли немає впливу поля, молекули орієнтовані довільно, тому рідина загалом не виявляє анізотропії. Під дією поля молекули повертаються так, щоб у напрямку поля були орієнтовані або дипольні електричні моменти (у полярних молекул), або напрямок найбільшої поляризації (у неполярних молекул). У результаті речовина стає оптично анізотропною.

Ефект Керра фактично безінерційний, тобто перехід речовини з ізотропного стану в анізотропний (і назад) у разі вмикання поля становить приблизно 10^{-10} с, тому цей ефект може бути ідеальним світловим затвором і його застосовують у швидкоплинних процесах (звукозапис, відтворення звуку, швидкісне фото і кінознімання), в оптичній локації.

3. Штучну анізотропію можна створити також дією магнітного поля, яка спостерігається в речовинах, молекули яких анізотропні, тобто в парамагнетиках. Коли немає зовнішнього магнітного поля, молекули розміщуються хаотично, результатом чого є статистична анізотропія. Якщо таку речовину помістити у досить сильне магнітне поле, то відбудеться напрямлена орієнтація власних магнітних моментів молекул. Це зумовлює анізотропію речовини, що приводить до подвійного заломлення променів. Таке середовище поводить себе як одновісній кристал, оптична вісь якого паралельна до вектора індукції поля $\vec{B}$. Це явище, відкрите у 1905 р., називають ефектом Коттона–Мутона. Різниця показників заломлення середовища при цьому

$$(n_e - n_o) = k_3 B^2, \qquad (14.28)$$

де k_3 – коефіцієнт пропорційності, а різниця фаз між звичайним і незвичайним променями становитиме

$$\delta = \frac{2\pi}{\lambda_0}(n_e - n_o)l = \frac{2\pi}{\lambda_0}k_3 lB2 = 2\pi l D B^2, \qquad (14.29)$$

де $D = k_3/\lambda_0$ – стала, яка залежить від природи речовини, довжини хвилі світла λ_0 і температури.

4. Зміну оптичних характеристик кристала під дією зовнішнього електричного поля називають *електро-оптичним ефектом Поккельса*. В одновісних кристалах поширення світла вздовж оптичної осі відбувається з однією і тією ж фазовою швидкістю $v_o = c/n_o$ незалежно від напрямку поляризації. Якщо кристал не має центра симетрії, то при прикладанні зовнішнього електричного поля фазові швидкості хвиль з ортогональними напрямками поляризації вздовж осі стають різними. На відмінну від ефекта Керра, *квадратичного* за напруженістю зовнішнього електричного поля, в електро-оптичному ефекті різниця фазових швидкостей таких хвиль пропорційна напруженості поля (лінійний ефект Поккельса). Безінерційність ефекту Поккельса дозволяє широко використовувати його для створення швидкодійних оптичних затворів і високочастотних модуляторів світла.

14.6. Обертання площини поляризації

Явище повертання площини поляризації світлової хвилі на деякий кут під час проходження світла крізь кристалічні тіла і деякі ізотропні рідини називають обертанням площини поляризації або оптичною активністю.

Якщо речовина не перебуває у зовнішньому магнітному полі, то оптична активність буде природною. Природну оптичну активність відкрив у 1811 р. Д. Араго на пластинах *кварцу*, які вирізані перпендикулярно до оптичної осі. Пізніше Ж. Біо виявив обертання напрямку поляризації світла в розчинах цукру. Тепер відомо багато природно активних речовин, хоч у більшості з них це явище виражене дуже слабо.

Для спостереження оптичної активності шар досліджуваної речовини розміщують між встановленими на темноту (схрещеними) поляризатором і аналізатором. Світло, яке при цьому проходить, можна знову погасити поворотом аналізатором. Це означає, що після проходження крізь шар активної речовини світло залишається лінійно поляризованим, але напрямок його поляризації повернутий на деякий кут відносно початкового напрямку.

Нехай погляд спостерігача спрямований назустріч падаючому променю. Обертання називають правим (додатним), якщо площина поляризації повертається вправо (за годинниковою стрілкою) для спостерігача, і лівим (від'ємним), якщо вона повертається вліво. Кристалічний кварц трапляється у

двох модифікаціях: для однієї обертання напрямку поляризації під час поширення вздовж оптичної осі відбувається вправо, для другої – вліво (на той же кут). Для кристалічної ґратки кварцу характерна відсутність центра і площин симетрії. Дві модифікації за симетрією і зовнішньою формою кристалів (рис. 14.25) різняться одна від одної так само, як права рука відрізняється від лівої або як гвинт із правою нарізкою відрізняється від такого ж гвинта з лівою нарізкою. Жодним переміщенням їх не можна сумістити одна з одною. Одна модифікація є дзеркальним зображенням іншої. Обертальна властивість кварцу пов'язана з його кристалічною структурою, оскільки плавлений кварц не має оптичної активності.

Кут обертання площини поляризації пропорційний товщині шару оптично активної речовини:

$$\varphi = \alpha l, \qquad (14.30)$$

де l – довжина шляху променя в оптично активному середовищі; α – коефіцієнт пропорційності, який називають обертальною здатністю або питомим обертанням. Він залежить від природи речовини, від температури і довжини хвилі.

Питоме обертання α дорівнює величині кута, на який повертається площина поляризації монохроматичного світла під час проходження шару завтовшки $l = 1$ м.

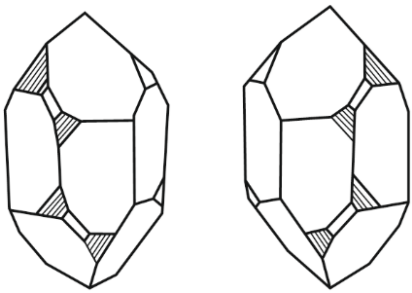


Рис. 14.25. Дві модифікації кристалічного кварцу – ліва і права

На віддалі від смуг поглинання світла речовиною залежність α від λ описують *законом Біо*: $\alpha \sim 1/\lambda^2$.

Для оптично активних рідин і розчинів Ж. Біо встановив, що кут повороту площини поляризації прямо пропорційний товщині шару і концентрації C оптично активної речовини, тобто

$$\varphi = [\alpha] Cl, \qquad (14.31)$$

де [α] – коефіцієнт пропорційності, який називають *питомим обертанням розчину*. Коефіцієнт [α] залежить від природи оптично активної речовини і розчинника, температури і довжини хвилі світла.

Властивості оптичної активності розчинів дають змогу визначити їх концентрацію. Прилади, за допомогою яких проводять такі вимірювання, називають *поляриметрами*. Оскільки для

розчину цукру питоме обертання [α] значне, то поляриметри набули широкого застосування в цукрометрії.

Теорію обертання площини поляризації оптично активними речовинами розробив О. Френель. Він вважав, що це явище зумовлене особливим видом подвійного заломлення променів, за якого швидкість поширення світла в активному середовищі різна для променів, що мають праву і ліву колову поляризацію. Знак кута повертання площини поляризації визначають співвідношенням між швидкістю поширення променів правої циркулярної поляризації $v_{\text{пр}}$ і лівої циркулярної поляризації $v_{\text{л}}$. Для $v_{\text{пр}} > v_{\text{л}}$ оптично активне середовище буде додатним, а для $v_{\text{пр}} < v_{\text{л}}$ – від'ємним.

На вході в оптично активну речовину лінійно поляризоване монохроматичне світло розкладається на дві хвилі тієї ж частоти, але поляризовані по колу у взаємно протилежних напрямках (рис. 14.26).

Вектори $\vec{E}_{\text{пр}}$ і $\vec{E}_{\text{л}}$ цих хвиль симетричні відносно площини *P–P* коливань падаючого світла. Якщо $v_{\text{пр}} > v_{\text{л}}$, то при виході з оптично активного середовища з товщиною шару l електричний вектор правоциркулярної хвилі буде повернутий на більший кут $\varphi_{\text{пр}}$, ніж для лівоциркулярної хвилі $\varphi_{\text{л}}$. Унаслідок цього площина $P'–P'$, відносно якої електричні вектори цих хвиль розміщенні симетрично, буде повернута вправо на кут φ відносно площини поляризації падаючої хвилі (рис. 14.27).

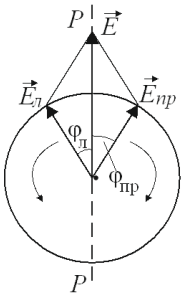


Рис. 14.26

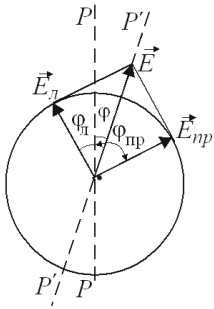


Рис. 14.27

Кут повороту знаходять з умови $\varphi_{\text{пр}} - \varphi = \varphi_{\text{л}} + \varphi$. Звідси отримаємо, що $\varphi = (\varphi_{\text{пр}} - \varphi_{\text{л}})/2$.

Кути повороту електричного вектора правої і лівої хвиль залежать від часу поширення хвилі *t* і довжини їхнього шляху в оптично активному середовищі, тобто

$$\varphi_{\text{пр}} = \omega\left(t - \frac{l}{v_{\text{пр}}}\right), \text{ а } \varphi_{\text{л}} = \omega\left(t - \frac{l}{v_{\text{л}}}\right).$$

Тоді

$$\varphi = \frac{1}{2}\left[\omega\left(t - \frac{l}{v_{\text{пр}}}\right) - \omega\left(t - \frac{l}{v_{\text{л}}}\right)\right] = \frac{\omega l}{2}\left(\frac{1}{v_{\text{л}}} - \frac{1}{v_{\text{пр}}}\right).$$

Фазову швидкість поширення лівої і правої хвиль виразимо, відповідно, через коефіцієнти заломлення $n_{\text{л}}$ і $n_{\text{пр}}$, $v_{\text{пр}} =$

$= c/n_{\text{пр}}$ і $\upsilon_{\text{л}} = c/n_{\text{л}}$. Враховуючи, що $\omega = 2\pi\nu = 2\pi/T$, з попередньої формули отримаємо

$$\varphi = \frac{2\pi l}{T}\left(\frac{n_{\text{л}}}{c} - \frac{n_{\text{пр}}}{c}\right) = \frac{\pi l}{cT}(n_{\text{л}} - n_{\text{пр}}) = \frac{\pi l}{\lambda_0}(n_{\text{л}} - n_{\text{пр}}), \quad (14.32)$$

де $\lambda_0 = cT$ – довжина світлової хвилі у вакуумі. Отже, за $n_{\text{л}} > n_{\text{пр}}$ $(\upsilon_{\text{л}} < \upsilon_{\text{пр}})$ площина поляризації повертається вправо, а за $n_{\text{л}} < n_{\text{пр}}$ $(\upsilon_{\text{л}} > \upsilon_{\text{пр}})$ – вліво.

Повертання площини поляризації під час проходження світла крізь речовину, яка перебуває в магнітному полі, вперше спостерігав Фарадей у 1846 р. Це був перший експеримент, який продемонстрував зв'язок оптичних і магнітних явищ. Фарадей так охарактеризував своє відкриття: «*Мені вдалося намагнітити і наелектризувати промінь світла та освітити магнітну силову лінію*». Для спостереження ефекту Фарадея досліджувана речовина розміщувалась між полюсами електромагніту (рис. 14.28).

Лінійно поляризоване світло пропускають вздовж магнітного поля через канал, просвердлений в осерді. Аналізатор A встановлюють так, щоб в разі відсутності магнітного поля світло не проходило (схрещенні поляризатор P і аналізатор A). Після включеня електромагніту світло проходило через аналізатор. Поворотом аналізатора його можна погасити. Це означає, що світло залишається лінійно поляризованим, і тут спостерігається саме поворот напрямку лінійної поляризації.

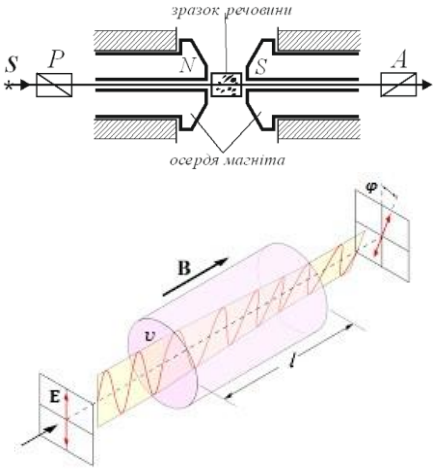


Рис. 14.28. Схема дослідження ефекта Фарадея

Якщо спостерігач дивиться у напрямку магнітного поля, то повертання праворуч вважають додатним, а повертання ліворуч – від'ємним. Досліди М. Фарадея та М. Верде показали, що кут повертання площини поляризації пропорційний довжині шляху l променя у речовині та напруженості магнітного поля H, тобто

$$\varphi = RlH, \quad (14.33)$$

де R – стала Верде, яка залежить від природи речовини і довжини хвилі світла. Більшість речовин характеризується додатним обертанням.

Ефект Фарадея широко застосовують в наукових дослідженнях. Мала інерційність ефекту (час встановлення менше 10^{-9} с) дозволяє використовувати його для модуляції світла, для створення оптичного затвору і под.

Класичну теорію ефекту Фарадея розглянуто у підрозд. 15.3.

Контрольні запитання

1. Яке світло називають поляризованим? Які є види поляризації світла?
2. Що таке ступінь поляризації?
3. Чому природне світло неполяризоване?
4. Як поляризуються відбиті від діелектрика хвилі?
5. Як поляризуються заломлені діелектриком хвилі?
6. Що встановлюють формули Френеля?
7. Що таке кут Брюстера?
8. Як визначити кут Брюстера?
9. Які поляризаційні прилади ви знаєте?
10. Які фізичні явища лежать в основі роботи поляризаційних приладів?
11. Опишіть будову і принципи дії призми Ніколя.
12. Опишіть будову і принцип дії поляроїдів.
13. Сформулюйте і поясніть закон Малюса.
14. У чому суть явища обертання площини поляризації? Як пояснюють це явище?
15. Як і в яких галузях науки і техніки враховують явище повертання площини поляризації?

Розділ 15. Взаємодія електромагнітних хвиль з речовиною

15.1. Дисперсія світла. Області нормальної та аномальної дисперсії. Класична електронна теорія дисперсії

Дисперсією світла називають залежність показника заломлення n речовини від частоти ν (*довжина хвилі* λ) *світла або залежність фазової швидкості v світла в середовищі від його частоти* ν.

Дисперсію світла подають у вигляді залежності $n = f(\lambda)$.

Наслідком дисперсії є розклад у спектр пучка білого світла під час проходження його крізь призму. Такий дослід вперше виконав у 1672 р. Ньютон. Схему досліду зображено на рис. 15.1.

Промінь світла від Сонця проходить через малий круглий отвір S у віконниці вікна, потім заломлюється у скляній призмі ABC і падає на аркуш білого паперу. При цьому кругле зображення отвору S розтягується в кольорову смугу DE, яку Ньютон назвав спектром.

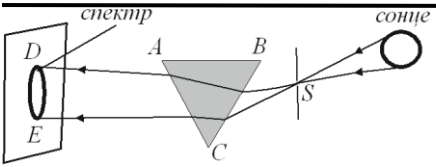


Рис. 15.1

За допомогою призми, як і за допомогою дифракційної ґратки, можна визначити спектральний склад світла.

Розглянемо відмінності в дифракційному і призматичному спектрах:

а) дифракційна ґратка розкладає світло, яке падає, безпосередньо за довжинами хвиль, тому за виміряними кутами можна обчислити довжину хвилі;

б) розклад білого світла у призмі відбувається за значеннями показника заломлення n, тому для визначення довжини хвилі λ світла потрібно знати залежність $n = f(\lambda)$.

У дифракційному спектрі кольори розміщуються у порядку зростання довжини хвилі, а в дисперсійному – навпаки. Дифракційною ґраткою червоні промені, що мають більшу довжину хвилі, ніж фіолетові, відхиляються сильніше. Дифракційний спектр рівномірно розтягнутий на всій ділянці спектру.

На рис. 15.2 показано залежності показника заломлення n від довжини хвилі λ для скла (1), кварцу (2) і флюориту (3). заломлення n монотонно зростає зі зменшенням довжини хвилі λ.

Як видно з рис.15.2, показник заломлення n монотонно зростає із зменшенням довжини хвилі λ. Отже, червоні промені, що мають менший показник заломлення порівняно з фіолетовими променями, відхиляються призмою на менший кут.

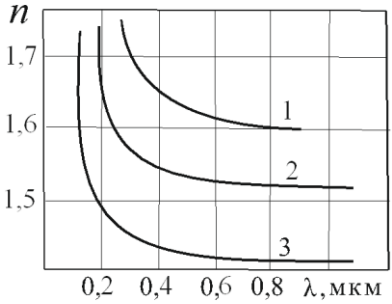


Рис. 15.2. Залежність показника заломлення від довжини хвилі

Дисперсійний спектр білого світла стиснутий у червоній області і розтягнутий у фіолетовій, тому що показник заломлення скла в області коротких хвиль зі зміною довжини хвилі випромінювання змінюється швидко, а в області довгих хвиль – повільно.

Залежність $n = f(\lambda)$ в оптичній області спектра має складний характер; для тих ділянок спектра, які слабо поглинаються певною речовиною, залежність показника заломлення від довжини хвилі може бути подана формулою ***Коші***:

$$n = n_0 + \frac{a}{\lambda^2} + \frac{b}{\lambda^4}$$

або

$$n \approx n_0 + \frac{a}{\lambda^2}. \qquad (15.1)$$

Величину, яка показує, як швидко змінюється показник заломлення речовини з довжиною хвилі λ, називають **дисперсією речовини *D***:

$$D = \frac{dn}{d\lambda}. \qquad (15.2)$$

Дисперсію світла в середовищі називають **нормальною**, якщо зі зростання частоти ν світла абсолютний показник заломлення *n* середовища також зростає:

$$\frac{dn}{d\nu} > 0, \quad \left(\frac{dn}{d\lambda} < 0\right).$$

Така залежність показника заломлення n від ν буде в тих областях частот, для яких середовище прозоре. Наприклад, звичайне скло прозоре для видимого світла і в цьому інтервалі частот має нормальну дисперсію.

Дисперсію світла в середовищі називають **аномальною**, якщо зі зростанням частоти світла абсолютний показник заломлення середовища n зменшується:

$$\frac{dn}{d\nu} < 0, \quad \left(\frac{dn}{d\lambda} > 0\right).$$

Якщо речовина поглинає частину променів, то в області поглинання і поблизу неї хід дисперсії проявляє аномалію (рис. 15.3). На деяких ділянках дисперсія речовини $dn/d\lambda$ є позитивною. Такий хід залежності n від λ називають *аномальною дисперсією.* Пунктирна крива на рис. 15.3 зображує хід коефіцієнта поглинання світла речовиною. Ділянка 1-2 і 3-4 відповідають нормальній дисперсії $dn/d\lambda < 0$. На ділянці 2-3 дисперсія є аномальною $dn/d\lambda > 0$. В області 1-2 показник заломлення менший за одиницю.

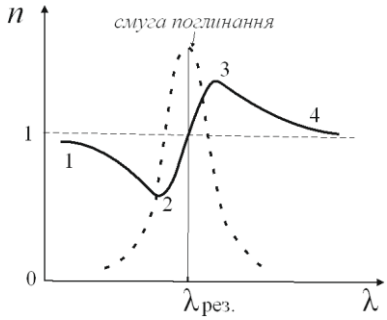


Рис. 15.3

В різних ділянках спектра дисперсія характеризується тією зміною показника заломлення, яка припадає на одиничний інтервал довжин хвиль. Цю величину $\Delta n/\Delta\lambda$ називають ***середньою дисперсією*** для ділянки спектра λ, $\lambda + \Delta\lambda$.

У довідникових таблицях для жовтої лінії натрію показник заломлення речовини становить $\lambda_D = 0{,}5893$ мкм і позначений n_D. Середню дисперсію

визначають за синьою $\lambda_F = 0{,}4861$ мкм і червоною $\lambda_C = 0{,}6563$ мкм лініями водню і позначено $n_F - n_C$. Величину $(n_F - n_C)/(n_D - 1)$ називають ***відносною дисперсією***, обернену їй величину – ***коефіцієнтом дисперсії***.

Оскільки показник заломлення визначають відношенням швидкості світла у вакуумі та обраному середовищі $n = c/v$, то очевидно, в останньому швидкість червоного світла, яке заломлюється найменше, буде найбільшою, а фіолетового – найменшою. У вакуумі швидкість світла будь-якого кольору однакова.

Зауважимо, що відсутність дисперсії світла у міжзоряному просторі засвідчує, що цей простір можна вважати вакуумом. Коли б це було не так, дисперсія світла виявлялася б, наприклад, у спостереженнях затемнення подвійних зірок. На початку затемнення колір зірки змінювався б від нормального до синього, оскільки відповідне йому світло поширюється у прозорому середовищі з найменшою швидкістю, наприкінці затемнення – від червоного до нормального, оскільки швидкість червоного світла найбільша. Ще Д. Араго (1786–1853), спостерігаючи затемнення подвійних зірок Алголь, не виявив цих ефектів.

Із кількісного боку дисперсію світла можна задовільно описати класичною теорією. Суть її зводиться до врахування впливу на середовище вимушених коливань електронів, що збуджуються електричним полем світлової хвилі. Йдеться про зовнішні (оптичні) електрони атомів; електрони внутрішніх оболонок через великі власні частоти фактично не збуджуються світлом.

Класична електронна теорія дисперсії

Щоб знайти залежність показника заломлення середовища *n* від частоти електромагнітної хвилі ω (або те саме, що діелектричної проникності ε(ω)), потрібна мікроскопічна теорія взаємодії електромагнітного поля хвилі з речовиною.

Мікроскопічна теорія ґрунтується на ідеалізованій моделі будови речовини. Найбільш простою є модель газоподібного середовища, в якому в першому наближенні можна не враховувати взаємодію атомів або молекул. Крім цього можна вважати, що поле, яке діє на окремий атом, збігається із середнім полем електромагнітної хвилі. За такої умови досить розглянути вплив поля на окремий атом, щоб отримати макроскопічне матеріальне рівняння.

Електрони, що входять до складу атома, можна поділити на зовнішні, або оптичні, і внутрішні, або електрони внутрішніх заповнених електронних оболонок. Така термінологія пов'язана з тим, що з випромінюванням оптичного діапазону взаємодіють фактично тільки зовнішні електрони. Власні частоти внутрішніх електронів настільки великі, що поле світлової хвилі оптичного діапазону на них фактично не впливає.

У класичній теорії дисперсії оптичний електрон в атомі розглядають як згасаючий дипольний випромінювач (осцилятор), що має власну частоту ω_0 і коефіцієнт згасання β.

Тоді рух оптичного електрона під впливом електромагнітної хвилі можна описати рівнянням вимушеного коливання осцилятора згідно з теорією вимушених коливань як:

$$m\vec{a} = -k\vec{r} - b\vec{v} - e\vec{E}(t), \qquad (15.3)$$

де m – маса електрона; k – коефіцієнт квазіпружної сили; $\vec{r}$ – радіус-вектор, що характеризує положення електрона; $-b\vec{v}$ – сила опору; $\vec{v} = \dot{\vec{r}}$ – швидкість електрона; $\vec{a} = \ddot{\vec{r}}$ – прискорення електрона.

Рівняння (15.3) перепишемо так:

$$\ddot{\vec{r}} + 2\beta\dot{\vec{r}} + \omega_0^2\vec{r} = -\frac{e}{m}\vec{E}(t), \qquad (15.4)$$

де $\beta = \frac{b}{2m}$ – коефіцієнт згасання; $\omega_0 = \sqrt{k/m}$ – власна частота; $-\frac{e}{m}\vec{E}(t)$ – вимушуюча сила, що припадає на одиницю маси електрона.

Рівняння (15.4) записано на основі того, що рух електрона описаний класичною механікою. З позиції сучасної фізики таке припущення зовсім невиправдане. Для атомних масштабів рух електрона описують квантовою механікою, тому може скластися враження, що класична теорія не має сенсу. Утім, це не так. Справа у тому, що квантова теорія дисперсії для описаної вище моделі гармонічного осцилятора дає ті самі результати, що і класична. Зрозуміло, що жодних квазіпружних сил та сил опору, пропорційних швидкості, в атомах немає, однак вплив решти електронів, ядра та електромагнітної хвилі на оптичний електрон можна описати цими двома силами. Насправді будова атомів і молекул визначається кулонівськими силами взаємодії електрона та ядер. Правильну теорію атома дає квантова механіка, тому і послідовна теорія дисперсії, яка використовує реалістичну модель середовища, має бути квантовою, але її розгляд виходить за межі цієї роботи.

У рівнянні (15.4) *власна частота* ω_0 *електрона в атомі може бути розрахована тільки на основі квантової теорії атома. У класичній теорії*

дисперсії її розглядають як формально введену константу. В експерименті значення ω_0 визначає частоту лінії поглинання у спектрі досліджуваної речовини.

Постійна згасання β у рівнянні (15.4), яка характеризує силу «опору», що пропорційна швидкості електрона, містить внесок, обумовлений радіаційним згасанням: у класичній теорії електрон, який осцилює (рухається з прискоренням) обов'язково випромінює. Є й інші причини згасання. Наприклад, перетворення енергії електромагнітного поля в інші форми (на теплоту) внаслідок взаємодії з іншими атомами і зіткненнями. Роль різних складових у рівнянні (15.4) залежить від області частот ω. *Для частот, далеких від власної частоти ω_0 осцилятора, згасанням можна знехтувати.* Ще зауважимо, що в рівнянні (15.3) квазіпружна сила $-k\vec{r} = -m\omega_0^2\vec{r}$ однакова в разі зміщень електрона в будь-якому напрямку, тобто є ізотропною.

В монохроматичній хвилі поле $\vec{E}(t)$, яке діє на осцилятор, у правій частині рівняння (15.4) є гармонічним, тобто

$$\vec{E}(t) = \vec{E}_0 \cos \omega t. \qquad (15.5)$$

Так само, як і в теорії вимушених коливань, нас буде цікавити частинний розв'язок рівняння (15.4), який буде описувати вимушені коливання осцилятора, що вже встановилися. Ці коливання виникають під дією зовнішньої синусоїдальної сили, тому будуть також синусоїдальними і їх частота буде збігатися з частотою вимушуючої сили.

Запишемо вимушуючу силу в комплексному вигляді:

$$\vec{E} = \vec{E}_0 e^{-i\omega t}, \qquad (15.6)$$

і тоді розв'язок рівняння (15.4) будемо шукати у такому вигляді:

$$\vec{r} = \vec{r}_0 e^{-i\omega t}. \qquad (15.7)$$

Якщо цей розв'язок (15.7) підставити у рівняння (15.4) з урахуванням (15.6), то можна визначити амплітуду коливань r_0 в (15.7). У результаті підстановки (15.7) в (15.4) отримаємо

$$\vec{r}(t) = \frac{-e/m}{\omega_0^2 - \omega^2 - 2i\beta\omega}\vec{E}(t). \qquad (15.8)$$

Індукований полем $\vec{E}(t)$ дипольний момент атома $\vec{p}(t) = -e\vec{r}(t)$ пропорційний напруженості поля:

$$\vec{p}(t) = \frac{e^2/m}{\omega_0^2 - \omega^2 - 2i\beta\omega}\vec{E}(t). \qquad (15.9)$$

Коефіцієнт пропорційності між $\vec{p}$ і $\vec{E}$ називаються поляризованістю атома $\varepsilon_0\alpha(\omega)$, тому

$$\vec{p}(t) = \varepsilon_0\alpha(\omega)\vec{E}. \qquad (15.10)$$

Якщо N є концентрацією атомів речовини, то вектор поляризації

$\vec{P} = N\vec{p}(t)$. Врахуємо зв'язок між вектором поляризації $\vec{P}$ і світловим вектором електромагнітної хвилі $\vec{E}$ як $\vec{P}(t) = \varepsilon_0\chi(\omega)\vec{E}(t)$. Тоді діелектрична сприйнятливість

$$\chi(\omega) = N\alpha(\omega). \qquad (15.11)$$

Із формул (15.9), (15.10) і (15.11) отримаємо, що

$$\chi(\omega) = \frac{Ne^2/(\varepsilon_0 m)}{\omega_0^2 - \omega^2 - 2i\beta\omega}\cdot \qquad (15.12)$$

Врахуємо, що діелектрична проникність $\varepsilon(\omega)$ і сприйнятливість пов'язані співвідношенням $\varepsilon(\omega) = 1 + + \chi(\omega)$, тоді

$$\varepsilon(\omega) = 1 + \frac{Ne^2/(\varepsilon_0 m)}{\omega_0^2 - \omega^2 - 2i\beta\omega}\cdot \qquad (15.13)$$

Для спрощення запису введемо константу

$$\omega_\text{p}^2 = N\,e^2/(\varepsilon_0 m), \qquad (15.14)$$

яка має розмірність частоти і характеризує модель середовища. Тоді (15.13) можна переписати так:

$$\varepsilon(\omega) = 1 + \frac{\omega_\text{p}^2}{\omega_0^2 - \omega^2 - 2i\beta\omega}\cdot \qquad (15.15)$$

Виділяючи з (15.15) дійсну ε' і комплексну ε'' складові, можемо отримати залежність показника заломлення $n(\omega) = \sqrt{\varepsilon'(\omega)}$ і коефіцієнта згасання $\varkappa(\omega)$ від частоти.

Нормальна дисперсія далеко від лінії поглинання

Для частот, далеких від власної частоти ω_0 атомних осциляторів, де виконується умова $2\beta\omega \ll |\omega_0^2 - \omega^2|$, уявною частиною в (15.15) можна знехтувати. Тоді залежність показника заломлення від частоти можна виразити наближеною формулою:

$$n^2(\omega) = \varepsilon(\omega) = 1 + \frac{\omega_\text{p}^2}{\omega_0^2 - \omega^2}. \qquad (15.16)$$

За досить малої концентрації N (гази або пара) далеко від власної частоти ω_0 (коли $\omega_\text{p}^2 \ll |\omega_0^2 - \omega^2|$) показник заломлення n близький до одиниці, тобто другий доданок у (15.16) малий порівняно з першим. Якщо застосувати формулу наближеного підрахунку $\sqrt{1+x} = 1 + \frac{1}{2}x$ (за $x \ll 1$), то із (15.16) отримаємо:

$$n(\omega) = 1 + \frac{1}{2}\cdot\frac{\omega_\text{p}^2}{\omega_0^2 - \omega^2}. \qquad (15.17)$$

Залежність показника заломлення від частоти показана на рис. 15.4. За частот ω, для яких показник заломлення $n(\omega)$ зростає зі зростанням частоти ω, відбувається нормальна дисперсія.

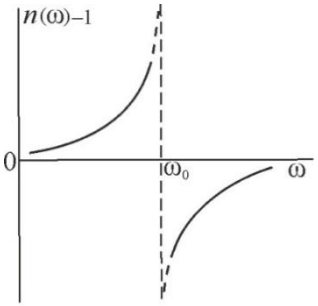


Рис. 15.4

Для частот, менших за ω_0, показник заломлення більший від

одиниці (див. формулу (15.17)), а отже фазова швидкість електромагнітної хвилі c/n в середовищі менша за швидкість світла у вакуумі. Це означає, що хвиля, яка змінена середовищем, за фазою відстає від падаючої хвилі.

Якщо частота світла ω більша від власної частоти ω_0 осциляторів, то показник заломлення $n < 1$ (рис. 15.3) і фазова швидкість хвилі в середовищі буде більшою від швидкості світла у вакуумі, тобто хвиля, змінена середовищем, за фазою випереджає падаючу хвилю.

Протиріччя з теорією відносності про граничну швидкість світла, коли $c = 3 \cdot 10^8$ м/с, тут немає. Згідно з теорією відносності швидкість матеріального тіла і швидкість сигналу не можуть перевищувати c. Поняття показника заломлення застосовують до монохроматичної хвилі, яка має нескінченну протяжність у просторі та часі, тобто до вимушених коливань осциляторів середовища, котрі вже встановились. Монохроматичну хвилю не можна використовувати для передавання сигналів (використовується пакет хвиль), тому сигнал рівносильний процесу, який ще не встановився, і із факту, що показник заломлення, менший від одиниці, ще не можна нічого сказати про швидкість сигналу. За частоти, більшій за ω_0, хвиля, яка встановилась у середовищі, за фазою випереджає падаючу хвилю, аналогічно за вимушених коливань, що встановилися, маятник рухається протилежно до сили, що діє на нього, якщо частота сили більша від власної частоти ω_0 маятника.

Багато речовин мають власні частоти ω_0 (і пов'язані з ними смуги поглинання) в далекій ультрафіолетовій області спектра, тому для частот ω у всій видимій області спектра справедлива нерівність $\omega \ll \omega_0$ (початкова ділянка зліва на рис. 15.4). Тоді, вважаючи $(\omega/\omega_0)^2$ малим параметром, вираз

$$\frac{1}{\omega_0^2-\omega^2} = \frac{1}{\omega_0^2} \cdot \frac{1}{1-\frac{\omega^2}{\omega_0^2}} \approx$$

$$\approx \frac{\left(1-\frac{\omega^2}{\omega_0^2}\right)^{-1}}{\omega_0^2} \approx \frac{1+\frac{\omega^2}{\omega_0^2}}{\omega_0^2}.$$

Тоді формула (15.16) набуде вигляду

$$n^2(\omega) = 1 + \frac{\omega_p^2}{\omega_0^2}\left(1+\frac{\omega^2}{\omega_0^2}\right). \quad (15.18)$$

Цю формулу можна переписати так:

$$n^2(\lambda) = 1 + K\left(1+\frac{L}{\lambda^2}\right), \quad (15.19)$$

якщо врахувати, що $\lambda = 2\pi c/\omega$, то

$$K = \frac{\omega_p^2}{\omega_0^2}, \quad (15.20)$$

$$L = \frac{4\pi^2 c^2}{\omega_0^2} \cdot \quad (15.21)$$

Формула (15.16) справедлива для розрідженого середовища, оскільки під час її отримання допускалося, що поле,

яке діє на окремий атом, збігається із середнім макроскопічним полем $\vec{E}$. У щільній речовині (рідини, тверді тіла) це припущення несправедливе, оскільки напруженість полів, що створюються дипольними моментами сусідніх атомів, що індукуються зовнішнім полем, можуть бути одного порядку з $\vec{E}$.

Для такого випадку замість формули (15.16) отримують формулу, яку отримали Г. Лоренц і Я. Лоренц,

$$\frac{n^2-1}{n^2+2} = \frac{\omega_p^2}{3(\omega_0^2-\omega^2)}. \qquad (15.22)$$

Для розрідженого середовища, показник заломлення якого близький до одиниці, можна вважати, що $n^2 + 2 \approx 3$ і формула (15.22) переходить у формулу (15.16).

Із формули (15.14) випливає, що ω_p^2 пропорційна концентрації атомів, а отже, і густині речовини ρ. Тому величину

$$\frac{1}{\rho} \cdot \frac{n^2-1}{n^2+2}$$

називають *питомою рефракцією*, і для речовини ρ не має змінюватись зі зміною густини. Так, при конденсації водяної пари питома рефракція змінюється всього на 3 %, хоча густина змінюється при цьому на три порядки.

15.2. Групова швидкість світла. Хвильовий пакет (група хвиль)

Оскільки показник заломлення n середовища залежить від частоти коливання ν світла, то швидкість поширення світла теж залежить від частоти коливань електромагнітної хвилі. З'ясуємо, яку швидкість поширення світлового сигналу вимірюють дослідним шляхом.

Усі приймачі світла реагують на енергію, тому в дослідах вимірюють *швидкість перенесення енергії світловим сигналом, і цю швидкість називають* ***груповою швидкістю***. Вона відрізняється від швидкості v поширення фази, яку називають ***фазовою швидкістю*** (див. підрозд. 10.3, формула (10.7)).

У реальних умовах дістати строго монохроматичну світлову хвилю неможливо. Виділяючи пучок світла одного кольору або модулюючи хвилю для передавання сигналу, фактично дістаємо пучок хвиль близьких, але все-таки різних частот, тому у раніше розглянутих методах вимірювання швидкості світла ми змушені були оцінювати швидкість поширення групи хвиль різних частот (довжин). Винятком можуть бути лише непрямі вимірювання швидкості світла. Це трапляється, наприклад, коли визначають показник заломлення середовища, а потім за

формулою $n = c/v$ знаходять фазову швидкість світла.

Як відомо, у разі накладання хвиль близьких частот виникає биття, тобто пульсація максимумів результуючої амплітуди, й отже, значень енергії. Це наочно можна бачити, якщо зобразити накладання двох хвиль з близькими частотами (довжинами хвиль) (рис. 15.5). У точках A і C хвилі перебувають у протифазах, тому результуюча амплітуда та енергія мінімальні; у точці B хвилі за фазою збігаються, тому виникає максимум амплітуди й енергії хвильового процесу.

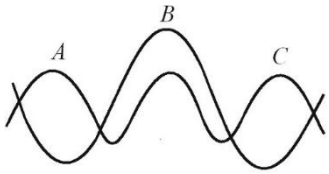


Рис. 15.5

Зауважимо, що у вакуумі внаслідок однакової швидкості світла різних частот взаємне розміщення хвиль залишається незмінним, тому швидкості поширення фаз і максимуму амплітуди хвильового процесу однакові; фазова і групова швидкості світла теж однакові. У разі переходу світлового пучка у дисперсійне середовище монохроматичні хвилі різних частот поширюватимуться з різною фазовою швидкістю, тому максимум амплітуди хвильового процесу весь час змінюватиме своє положення у смузі складових хвиль. Його швидкість у просторі не збігатиметься зі швидкістю жодної зі складових хвиль. У цьому випадку питання про групову швидкість світла ускладнюється.

Щоб оцінити групову швидкість світла в дисперсійному середовищі, розглянемо накладання двох світлових хвиль із близькими частотами та однаковими амплітудами:

$$E_1 = E_m \cos(\omega_1 t - k_1 x),$$
$$E_2 = E_m \cos(\omega_2 t - k_2 x).$$

Їх результуюче зміщення E в деякий момент часу t у точці з координатою x можна виразити сумою

$$E(x,t) = E_1(x,t) + E_2(x,t) = 2E_m \cos\left(\frac{\Delta\omega}{2}t - \frac{\Delta k}{2}x\right) \times \times \cos(\omega_{\text{сер}}\, t - k_{\text{сер}} x), \qquad (15.23)$$

де

$$\Delta\omega = \omega_1 - \omega_2,\ \omega_{\text{сер}} = (\omega_1 + \omega_2)/2,$$
$$\Delta k = k_1 - k_2,\ k_{\text{сер}} = (k_1 + k_2)/2. \qquad (15.24)$$

З виразу (15.23) випливає, що результуюча хвиля є плоскою, її циклічна частота і хвильове число дорівнюють півсумі циклічних частот і хвильових чисел синусоїдальних хвиль, що утворюють хвильовий пакет. *Утім, амплітуда цієї хвилі не стала, а залежить від координати x та часу t*:

$$A(x,\, t) = 2E_m \cos\left(\frac{\Delta\omega}{2}t - \frac{\Delta k}{2}x\right). \qquad (15.25)$$

Вираз (15.25) для амплітуди

хвильового пакета сам по собі є виразом рівняння плоскої синусоїдальної хвилі – хвилі амплітуди коливань, фаза якої

$$\varphi_A = \frac{\Delta\omega}{2}t - \frac{\Delta k}{2}x. \qquad (15.26)$$

Знайдемо, чому дорівнює швидкість результуючої хвилі у дисперсійному середовищі. Хвиля, поширюючись у середовищі, переносить енергію.

Швидкість v поширення енергії хвилі дорівнює швидкості переміщення у просторі поверхні, через яку переноситься максимальна кількість енергії за одиницю часу. Оскільки енергія в разі коливального процесу пропорційна квадрату амплітуди, то швидкість поширення енергії хвильового пакета збігається з фазовою швидкістю хвилі амплітуди. Вочевидь, що для того, щоб амплітуда $A(x,\ t)$ (формула (15.25)) мала стале значення, наприклад максимальне, її фаза φ_A мусить залишатися сталою, тобто

$$\frac{\Delta\omega}{2}t - \frac{\Delta k}{2}x = \text{const}. \qquad (15.27)$$

Взявши повний диференціал від цього виразу, отримаємо

$$\frac{\Delta\omega}{2}dt - \frac{\Delta k}{2}dx = 0.$$

Звідси

$$v_{\text{гр}} = \frac{dx}{dt} = \frac{\Delta\omega}{\Delta k} = \frac{\omega_1 - \omega_2}{k_1 - k_2}. \qquad (15.28)$$

За умови, що $\Delta\omega$, а отже і Δk, прямують до нуля,

$$v_{\text{гр}} = \frac{d\omega}{dk}. \qquad (15.29)$$

Цю швидкість називають груповою швидкістю пакета хвиль.

Зауваження. Формула (15.29) справедлива для лінійної залежності ω від k, яка характерна для квазімонохроматичної хвилі (хвильового пакета або групи хвиль, див. далі).

Швидкість гармонічних хвиль (формула (10.7), підрозд. 10.2) $v = \omega/k$. Отже, згідно зі співвідношеннями (10.7) і (15.29) групова швидкість пакета хвиль не дорівнює жодній із швидкостей синусоїдальних складових цього пакета.

Вираз для групової швидкості (15.29) можна подати ще й у такій формі. Згідно з формулою (10.7) (див. підрозд. 10.2) $\omega = vk$.

Підставивши це значення у (15.29), отримаємо

$$v_{\text{гр}} = \frac{d(vk)}{dk} = v + k\frac{dv}{dk}. \qquad (15.30)$$

Якщо врахувати, що $k = 2\pi/\lambda$ та $dk = -\frac{2\pi}{\lambda^2}d\lambda$, то отримаємо такий вираз для групової швидкості:

$$v_{\text{гр}} = v - \lambda\frac{dv}{d\lambda}. \qquad (15.31)$$

Якщо врахувати, що

$$\frac{dv}{d\lambda} = \frac{d}{d\lambda}\left(\frac{c}{n}\right) = -\frac{c}{n^2}\frac{dn}{d\lambda},$$

то формулу (15.31) можна переписати так:

$$v_{\text{гр}} = v\left(1 + \frac{\lambda}{n}\frac{dn}{d\lambda}\right). \qquad (15.32)$$

Якщо $dn/d\lambda = 0$ (немає дисперсії), то $v_{\text{гр}} = v$, тобто фазова і групова швидкість

збігаються. Цей випадок справджується для вакууму.

За нормальної дисперсії

$$\frac{dn}{d\lambda} < 0 \ \left(\frac{dn}{d\nu} > 0\right),$$

тому $v_{\text{гр}} < v$.

За аномальної дисперсії

$$\frac{dn}{d\lambda} > 0 \ \left(\frac{dn}{d\nu} < 0\right) \ \text{і} \ v_{\text{гр}} > v.$$

Поняття групової швидкості застосовне лише за умови, що поглинання енергії хвилі в певному середовищі невелике. За значного згасання хвилі поняття групової швидкості втрачає сенс. Саме в області аномальної дисперсії поглинання дуже велике і поняття групової швидкості використовувати не можна.

Наочно картину поширення групи хвиль за нормальної дисперсії можна описати так. Врахуємо, що $\frac{dn}{d\lambda} < 0$. Це означає, що для $\lambda_2 > \lambda_1$, $n_2 < n_1$ і $v_2 > v_1$. Тоді якби у початковий момент часу максимум амплітуди містився посередині смуги хвиль, то надалі у їхньому русі спостерігалося б поступове ослаблення, а потім і зникнення передніх хвиль; у голову смуги хвиль пік амплітуди хвильового процесу повільно переміщувався б, слабшав і згодом зникав би, одночасно позаду смуги хвиль формувався, повільно переміщувався і виростав би новий пік амплітуди і т. д. Оскільки передні хвилі весь час зникають, а наступний пік амплітуди формується позаду смуги хвиль, то це є свідченням того, що групова швидкість менша за фазову швидкість складових хвиль.

Групова швидкість світла в усіх випадках не може перевищувати швидкості світла у вакуумі.

Хвильовий пакет (група хвиль)

Розглянемо поняття групової швидкості для більш загального випадку. Нагадаємо, що згідно з теорією Максвелла для електромагнітних хвиль (ЕМХ) фазова швидкість їх поширення

$$v = \frac{1}{\sqrt{\varepsilon\mu\varepsilon_0\mu_0}} = \frac{c}{\sqrt{\varepsilon\mu}} = \frac{c}{n}, \qquad (15.33)$$

де $c = \frac{1}{\sqrt{\varepsilon_0\mu_0}} = 3\cdot 10^8$ м/с – швидкість поширення ЕМХ у вакуумі, для якого відносні електрична і магнітна проникність ε та μ дорівнюють одиниці; $n = \sqrt{\varepsilon\mu}$ – показник заломлення середовища (для немагнітних середовищ $\mu = 1$, тому $n = \sqrt{\varepsilon}$). Діелектрична проникність ε залежить від поляризації діелектрика ЕМХ. Ця поляризація залежить від частоти ω (довжини хвилі λ) ЕМХ, тому згідно з (15.26) n та v теж залежать від частоти ω (довжини хвилі λ) ЕМХ. Згідно з формулою (15.26) у загальному випадку швидкість v залежить від частоти ω, а отже, вона

залежить і від хвильового числа k. Тому згідно з (15.26)

$$\omega(k) = kv(k) \qquad (15.34)$$

є складною функцією, яку називають ***дисперсійним співвідношенням*** для певного середовища.

Для загального випадку запишемо пакет хвиль як суму монохро-матичних хвиль для хвильових чисел від $k_0 - \Delta k$ до $k_0 + \Delta k$ або частот від $\omega_0 - \Delta\omega$ до $\omega_0 + +\Delta\omega$:

$$E(x,t) = \\ = \int_{k_0-\Delta k}^{k_0+\Delta k} E_{01}(k)\cos(\omega t - kx)dk, \qquad (15.35)$$

$$E(x,t) = \\ = \int_{\omega_0-\Delta\omega}^{\omega_0+\Delta\omega} E_{02}(\omega)\cos(\omega t - kx)d\omega, \qquad (15.36)$$

де $E_{01}(k)$ та $E_{02}(\omega)$ – амплітуди складових монохроматичних хвиль; ω_0 та k_0 – центральні (основні) частота і хвильове число групи хвиль.

Отже, реальна ЕМХ не є монохроматичною, а містить хвилі певного інтервалу $2\Delta\omega$ частот, або хвильових чисел $2\Delta k$. У загальному випадку інтервали $2\Delta\omega$ і $(2\Delta k)$ є нескінченно великими, але для нас особливий інтерес становлять квазімонохроматичні (майже монохро-матичні) хвилі, для яких

$$\frac{\Delta\omega}{\omega_0} \ll 1, \quad \frac{\Delta k}{k_0} \ll 1. \qquad (15.37)$$

Для цього випадку основний внесок у суми (15.35) і (15.36) вносять хвилі поблизу центральної частоти ω_0 (хвильового числа k_0). Таку суму хвиль називають групою хвиль або хвильовим пакетом.

Для такого хвильового пакета, коли $\frac{\Delta k}{k_0} \ll 1$, можна залежність (15.34) $\omega(k)$ розкласти у ряд за Δk і використати перше наближення розкладу, тобто

$$\omega(k) \approx \omega_0 + \left(\frac{d\omega}{dk}\right)_{k=k_0} \cdot \Delta k = \\ = \omega_0 + b \cdot \Delta k, \qquad (15.38)$$

де

$$b = \left(\frac{d\omega}{dk}\right)\Big|_{k=k_0}. \qquad (15.39)$$

У розкладі (15.38) знехтували членами з Δk другого і т.д. порядків. Для спрощення розгляду будемо вважати, що у сумі (15.35) $E_{01}(k) = E_0 = \text{const}$. Тоді (15.35) з урахуванням (15.38) запишемо так:

$$E(x,t) = \\ = E_0\left(\int_{k_0-\Delta k}^{k_0+\Delta k} \cos(\omega_0 t + bt(k-k_0) - \right. \\ \left. - kx)dk\right) = A(x,t)\cos(\omega_0 t - k_0 x), \qquad (15.40)$$

де

$$A(x,t) = 2E_0\Delta k\frac{\sin((bt-x)\Delta k)}{(bt-x)\Delta k}. \qquad (15.41)$$

Вираз (15.40) є рівнянням хвилі з частотою ω_0 і хвильовим числом k_0, але амплітуда такої хвилі (вираз (15.41)) залежить від координати x і часу t значно слабіше, ніж $\cos(\omega_0 t - k_0 x)$, тому що $\frac{\Delta k}{k_0} \ll 1$.

Рівняння (15.41) можна розглядати як таку собі хвилю амплітуди пакета хвиль (15.40). Ця хвиля амплітуди поширюється в середовищі із швидкістю

$$b = \left(\frac{d\omega}{dk}\right)\Big|_{k=k_0} = \text{const},$$

яку ми будемо називати груповою швидкістю і позначатимемо як

$$v_{\text{гр}} = \frac{d\omega}{dk}. \qquad (15.42)$$

Для нашого квазімонохроматичного наближення $v_{\text{гр}} = \text{const}$. Це означає, що в такому наближенні пакет хвиль не буде розпливатись під час поширення в середовищі з часом, а $v_{\text{гр}}$ характеризує рух центра групи хвиль, де амплітуда максимальна. Оскільки енергія хвилі пропорційна квадрату амплітуди, то $v_{\text{гр}}$ визначає швидкість поширення енергії та імпульсу пакета хвиль.

15.3. Вплив магнітного поля на рух електрона в атомі.

Пояснення ефекту Фарадея

Вплив зовнішнього магнітного поля $\vec{B}$ на рух електрона в атомі, який випромінює електромагнітні хвилі, розглянемо для випадку, коли можна знехтувати згасанням його коливань. У рівняння (15.4) включимо силу Лоренца $-e\vec{v} \times \vec{B} = -e\dot{\vec{r}} \times \vec{B}$, вважаючи, що $\beta = 0$ і $\vec{E}(t) = 0$. Тоді (15.4) перепишемо так:

$$\ddot{\vec{r}} + \omega_0^2 \vec{r} = -\frac{e}{m}\dot{\vec{r}} \times \vec{B} \qquad (15.43)$$

або

$$\ddot{\vec{r}} + \omega_0^2 \vec{r} = \dot{\vec{r}} \times \left(-\frac{e}{m}\vec{B}\right) = 2\dot{\vec{r}} \times \vec{\omega}_L, \qquad (15.44)$$

де

$$\vec{\omega}_L = -\frac{e}{2m}\vec{B} \qquad (15.45)$$

є вектором кутової швидкості (Ларморова частота, див. підрозд. 6.6).

Тоді (15.44) перепишемо так:

$$\ddot{\vec{r}} - 2\dot{\vec{r}} \times \vec{\omega}_L + \omega_0^2 \vec{r} = 0. \qquad (15.46)$$

У декартових координатах рівняння (15.46), якщо вісь Z направити вздовж магнітного поля $\vec{B}$, запишемо так:

$$\begin{cases} \ddot{x} + 2\omega_L \dot{y} + \omega_0^2 x = 0; \\ \ddot{y} - 2\omega_L \dot{x} + \omega_0^2 y = 0; \\ \ddot{z} + \omega_0^2 z = 0. \end{cases} \qquad (15.47)$$

При цьому враховано, що

$$\dot{\vec{r}} \times \vec{\omega}_L = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ \dot{x} & \dot{y} & \dot{z} \\ 0 & 0 & -\omega_L \end{vmatrix} =$$

$$= -\omega_L \dot{y}\vec{i} + \omega_L \dot{x}\vec{j} + 0 \cdot \vec{k}.$$

Таким чином отримали систему однорідних диференціальних рівнянь, з яких випливає, що можливе гармонічне коливання електрона вздовж осі Z. Воно відбувається з тією ж частотою ω_0, що і за відсутності магнітного поля. Перші два рівняння системи (15.47) взаємозв'язані між собою й утворюють характеристичне рівняння

$$\omega^2 + 2\omega_L \omega - \omega_0^2 = 0,$$

коренями якого є частоти

$$\omega_{1,2} = -\omega_L \pm \sqrt{\omega_0^2 + \omega_L{}^2}. \qquad (15.48)$$

Ларморова частота ω_L набагато менша від власної частоти електрона ω_0

навіть у сильних магнітних полях. Наприклад, за $B = 1$ Тл формула (15.45) дає $\omega_L \approx 10^{11}\text{с}^{-1}$, а ω_0 для видимого світла ($\lambda = 0{,}5$ мкм) дорівнює близько $4 \cdot 10^{15}$ с$^{-1}$, тому $(\omega_L/\omega_0)^2 \approx 5 \cdot 10^{-10}$ і другим доданком у підкореневому виразі (15.48) можна знехтувати. Тоді

$$\omega_{1,2} = \pm\omega_0 - \omega_L. \qquad (15.49)$$

Першому кореню $\omega_1 = \omega_0 - \omega_L$ відповідає розв'язок

$$\begin{cases} x(t) = a_0 \cos(\omega_0 - \omega_L)t, \\ y(t) = -a_0 \sin(\omega_0 - \omega_L)t. \end{cases} \qquad (15.50)$$

Цей розв'язок описує рух електрона по колу у площині XY в напрямку *за годинниковою стрілкою* з кутовою швидкістю $\omega_0 - \omega_L$. Таку саму частоту має хвиля, яка випромінюється електроном. Світло з частотою $\omega_0 - \omega_L$, яке поширюється вздовж магнітного поля, має *праву колову поляризацію*.

Другому кореню характеристичного рівняння $\omega_2 = -\omega_0 - \omega_L$ відповідає розв'язок

$$\begin{cases} x(t) = a_0 \cos(\omega_0 + \omega_L)t, \\ y(t) = a_0 \sin(\omega_0 + \omega_L)t. \end{cases} \qquad (15.51)$$

Тут електрон рухається у площині XY по колу *проти годинникової стрілки* з кутовою швидкістю $\omega_0 + \omega_L$, тому світло такої самої частоти, яке поширюється вздовж магнітного поля, має *ліву колову циркуляцію*.

Знайдемо показник заломлення для хвиль, частоти яких перебувають далеко від $\omega_1 = \omega_0 - \omega_L$ та $\omega_2 = \omega_0 + \omega_L$, де згасанням осцилятора можна знехтувати. Для цього випадку рівняння (15.43) у полі світлової хвилі за наявності зовнішнього магнітного поля має вигляд

$$\ddot{\vec{r}} + \omega_0^2 \vec{r} = -\frac{e}{m}\left(\vec{E}(t) + \dot{\vec{r}} \times \vec{B}\right), \qquad (15.52)$$

або врахувавши (15.45), рівняння (15.52) перепишемо так:

$$\ddot{\vec{r}} - 2\dot{\vec{r}} \times \vec{\omega}_L + \omega_0^2 \vec{r} = -\frac{e}{m}\vec{E}(t). \qquad (15.53)$$

Направимо вісь Z уздовж магнітного поля $\vec{B}$. Тоді вектор $\vec{\omega}_L$ буде мати проєкції $(0, 0, -\omega_L)$. Вектор напруженості $\vec{E}(t)$ електричного поля для циркулярно поляризованої монохроматичної хвилі з частотою ω, яка поширюється паралельно $\vec{B}$, має проєкції

$$\begin{cases} E_x = E_0 \cos \omega t\,, \\ E_y = \pm E_0 \sin \omega t, \\ E_z = 0, \end{cases}$$

де знак «+» в E_y відповідає хвилі з вектором $\vec{E}_+$, який повертається проти годинникової стрілки, а знак «−» − хвилі з вектором $\vec{E}_-$, що повертається у протилежному напрямку. При цьому напрямок поширення хвилі (уздовж поля $\vec{B}$ чи назустріч) не має значення.

Запишемо рівняння (15.53) у проєкціях на осі X та Y:

$$\begin{cases} \ddot{x} + 2\omega_L \dot{y} + \omega_0^2 x = -\frac{e}{m} E_0 \cos \omega t\,, \\ \ddot{y} - 2\omega_L \dot{x} + \omega_0^2 y = \pm\frac{e}{m} E_0 \sin \omega t, \end{cases} \qquad (15.54)$$

Розв'язками цих рівнянь будуть:

$$\begin{cases} x(t) = \frac{(-e/m)E_x(t)}{\omega_0^2-\omega^2\pm 2\omega_L\omega}, \\ y(t) = \frac{(-e/m)E_y(t)}{\omega_0^2-\omega^2\pm 2\omega_L\omega}. \end{cases} \quad (15.55)$$

Звідси знаходимо індукований дипольний момент $\vec{p}(t) = -e\vec{r}(t)$ і атомну поляризованість $\alpha_\pm(\omega)$ у полі циркулярно поляризованої монохроматичної хвилі:

$$\alpha_\pm(\omega) = \frac{e^2/(m\varepsilon_0)}{\omega_0^2-\omega^2\pm 2\omega_L\omega}. \quad (15.56)$$

Знаменник у (15.56) можна переписати так:

$$\omega_0^2 - \omega^2 \pm 2\omega_L\omega \approx \omega_0^2 - (\omega \mp \omega_L)^2,$$

врахувавши, що $\omega_L \ll \omega$. Тоді

$$\alpha_\pm(\omega) = \frac{e^2/(m\varepsilon_0)}{\omega_0^2-(\omega\mp\omega_L)^2}, \quad (15.57)$$

а вектор поляризації

$$\vec{P}(t) = N\varepsilon_0\alpha_\pm(\omega)\vec{E}(t) = \varepsilon_0\chi(\omega)\vec{E}(t).$$

Звідси $\chi_\pm(\omega) = N\alpha_\pm(\omega)$, а діелектрична проникність

$$\varepsilon_\pm(\omega) = 1 + \chi_\pm(\omega) = \\ = 1 + \frac{Ne^2/(m\varepsilon_0)}{\omega_0^2-(\omega\mp\omega_L)^2}, \quad (15.58)$$

або

$$\varepsilon_\pm(\omega) = 1 + \frac{\omega_\text{p}^2}{\omega_0^2-(\omega\mp\omega_L)^2}, \quad (15.59)$$

де ω_p визначають за формулою (15.14). Звідси знаходимо, що

$$n_\pm(\omega) = \sqrt{\varepsilon_\pm(\omega)} = \\ = \sqrt{1 + \frac{\omega_\text{p}^2}{\omega_0^2-(\omega\mp\omega_L)^2}}. \quad (15.60)$$

Оскільки $\omega_L \ll \omega$, $\omega_L \ll \omega_0$, $\omega_\text{p}^2 \ll \ll |\omega_0^2 - \omega^2|$, то під коренем $\frac{\omega_\text{p}^2}{\omega_0^2-(\omega\mp\omega_L)^2} \ll 1$ і вираз (15.60) перепишемо за допомогою формули для наближеного підрахунку так:

$$n_\pm(\omega) = 1 + \frac{1}{2}\cdot\frac{\omega_\text{p}^2}{\omega_0^2-(\omega\mp\omega_L)^2}. \quad (15.61)$$

Важливо зауважити, що формулу (15.61) можна отримати із формули (15.17), замінивши в ній ω на $\omega \mp \omega_L$.

Для пояснення повертання площини поляризації в ефекті Фарадея важливо знайти різницю $n_\text{л} - n_\text{пр} = \\ = n_-(\omega + \omega_L) - n_+(\omega - \omega_L)$. Із формули (15.61) знаходимо:

$$n_\text{л} - n_\text{пр} = \\ = 2\omega_\text{p}^2\frac{\omega\omega_L}{(\omega_0^2-(\omega+\omega_L)^2)(\omega_0^2-(\omega-\omega_L)^2)},$$

або врахувавши, що $\omega_L \ll \omega$, отримаємо:

$$n_\text{л} - n_\text{пр} = 2\omega_\text{p}^2\frac{\omega}{(\omega_0^2-\omega^2)^2}\omega_L. \quad (15.62)$$

Перепишемо (15.62) по-іншому. Для цього продифенціюємо (15.17) за ω:

$$\frac{dn(\omega)}{d\omega} = \frac{\omega_\text{p}^2}{2}\cdot\left(\frac{2\omega}{(\omega_0^2-\omega^2)^2}\right).$$

Тоді (15.62) перепишемо так:

$$n_\text{л} - n_\text{пр} = 2\frac{dn(\omega)}{d\omega}\,\omega_L. \quad (15.63)$$

Підставимо (15.63) у формулу (14.32) й отримаємо вираз для кута повороту напрямку поляризації:

$$\varphi = \frac{\pi l}{\lambda}\left(n_\text{л} - n_\text{пр}\right) = \\ = \frac{\omega}{2c}\left(n_\text{л} - n_\text{пр}\right)l = \frac{\omega}{c}\omega_L\frac{dn}{d\omega}l. \quad (15.64)$$

Для області нормальної дисперсії показник заломлення n зростає зі

збільшенням частоти ω, тобто $\frac{dn}{d\omega} > 0$, і поворот напрямку поляризації відбувається в той бік, в який протікає струм в обмотці електромагніту, який створює магнітне поле. Це справедливо як для світла, яке поширюється вздовж поля, так і для світла, яке поширюється назустріч полю.

Врахувавши формулу (15.45), (15.64) перепишемо так:

$$\varphi = \frac{e}{2mc} \cdot \omega \frac{dn}{d\omega} Bl. \quad (15.65)$$

Для немагнітних речовин $B = \mu_0 H$ (H – напруженість магнітного поля), тому (15.65) перепишемо таким чином:

$$\varphi = \frac{\mu_0 e}{2mc} \cdot \omega \frac{dn}{d\omega} Hl. \quad (15.66)$$

Як видно із (15.66), кут повороту φ пропорційний напруженості магнітного поля H і пройденому світлом шляху l в середовищі.

Величину

$$R = \frac{\mu_0 e}{2mc} \cdot \omega \frac{dn}{d\omega} = -\frac{\mu_0 e}{2mc} \cdot \lambda \frac{dn}{d\lambda} \quad (15.67)$$

називають сталою Верде.

Для закону дисперсії, який описують формулою (15.19) (формула Коші),

$$\frac{dn}{d\lambda} = -\frac{2KL}{\lambda^3}. \quad (15.68)$$

Якщо (15.68) підставити у (15.67), то для сталої Верде отримаємо вираз

$$R = \frac{\mu_0 |e|}{mc} \cdot \frac{KL}{\lambda^2}. \quad (15.69)$$

Залежність $R \sim 1/\lambda^2$ (закон Біо) підтверджується експериментально.

Якщо експериментально визначити сталу Верде, то можна знайти таку величину, як

$$D = \omega \frac{dn}{d\omega} = -\lambda \frac{dn}{d\lambda}, \quad (15.70)$$

яка характеризує дисперсію досліджуваної речовини для частоти ω (довжини хвилі λ). *Величину D називають* ***фактором дисперсії***, її визначають за формулою

$$D = \frac{2mc}{\mu_0 |e|} \cdot R. \quad (15.71)$$

Формулу для власної частоти ω_0 електрона в атомі отримаємо із (15.22), яка має такий вигляд:

$$\frac{n^2-1}{n^2+2} = \frac{\omega_p^2}{3(\omega_0^2-\omega^2)},$$

продиференціювавши її зліва і справа за частотою ω. Тоді отримаємо:

$$\frac{2n\frac{dn}{d\omega}(n^2+2) - 2n\frac{dn}{d\omega}(n^2-1)}{(n^2+2)^2} = \frac{\omega_p^2 \cdot 2\omega}{3(\omega_0^2-\omega^2)^2}, \quad (15.72)$$

або

$$\frac{3n\frac{dn}{d\omega}\omega}{(n^2+2)^2} = \frac{\omega_p^2}{3} \cdot \frac{\omega^2}{(\omega_0^2-\omega^2)^2}. \quad (15.73)$$

Якщо врахувати (15.59), то цей вираз перепишемо так:

$$\frac{9nD}{(n^2+2)^2} = \frac{\omega_p^2 \omega^2}{(\omega_0^2-\omega^2)^2}. \quad (15.74)$$

Із формули (15.22) визначимо ω_p^2 й підставимо цей вираз для ω_p^2 у (15.74). Тоді отримаємо

$$\frac{3nD}{(n^2+2)\cdot(n^2-1)} = \frac{\omega^2}{\omega_0^2-\omega^2}, \qquad (15.75)$$

або

$$\frac{(n^2+2)\cdot(n^2-1)}{3nD} = \frac{\omega_0^2}{\omega^2} - 1. \qquad (15.76)$$

З виразу (15.76) отримаємо, що

$$\omega_0 = \omega\sqrt{\frac{(n^2+2)\cdot(n^2-1)}{3nD} + 1}. \qquad (15.77)$$

Для довжини хвилі формулу (15.77) запишемо так:

$$\lambda_0 = \lambda\left(\sqrt{\frac{(n^2+2)\cdot(n^2-1)}{3nD} + 1}\right)^{-1},$$

або

$$\lambda_0 = \lambda\sqrt{\frac{3nD}{(n^2+2)\cdot(n^2-1)+3nD}}. \qquad (15.78)$$

Якщо

$$A = \frac{nD}{n^2-1} \qquad (15.79)$$

то

$$\omega_0 = \omega\sqrt{\frac{(n^2+2)}{3A} + 1}, \qquad (15.80)$$

$$\lambda_0 = \lambda\sqrt{\frac{3A}{n^2+2+3A}}. \qquad (15.81)$$

Вкрай важливо зауважити, що зі зміною довжини хвилі світла λ резонансна довжина хвилі λ_0, що відповідає власній частоті ω_0, фактично не повинна змінюватись. Це важливий фактор, який має експериментально підтверджуватись під час дослідження дисперсії середовища $n(\lambda)$.

15.4. Поглинання світла

Поглинанням світла *назива-ють явище втрати енергії світловою хвилею, яка проходить крізь речовину, внаслідок перетворення енергії хвилі в інші форми енергії.*

Під час проходження паралельного пучка світла крізь шар прозорого середовища його інтенсивність зменшується. Поглинання світла може приводити до нагрівання, іонізації або збудження атомів і молекул речовин, до деформації, поглинання може супроводжуватись розсіянням світла та індуктивним (вимушеним) випромінюванням.

Щоб отримати співвідношення, яке виражає закон поглинання світла, розглянемо шар прозорого середовища завтовшки l, на який падає паралельний пучок променів, інтенсивність якого I_0. Виділимо в середовищі нескінченно тонкий шар dl, який обмежений паралельними поверхнями, що перпендикулярні до напрямку поширення світла (рис. 15.6).

Дослід показує, що зменшення інтенсивності світла шаром середовища dl пропорційне величині інтенсивності, що входить у цей шар, і товщині шару, тобто

$$dI = -\alpha I dl,$$

де α – коефіцієнт пропорційності, який не залежить від інтенсивності світла, його називають *коефіцієнтом поглинання.* Знак «мінус» вказує на те, що зі збільшенням товщини шару середовища, яке поглинає світло, інтенсивність світла, що проходить крізь нього, зменшується.

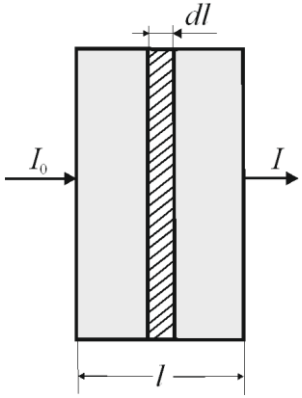


Рис. 15.6

Після розділення змінних у рівнянні дістаємо $dI/I = -\alpha dl$. Проінтегруємо це рівняння:

$$\int_{I_0}^{I} \frac{dI}{I} = -\alpha \int_{0}^{l} dl; \ \ln\frac{I}{I_0} = -\alpha l.$$

У результаті отримаємо:

$$I = I_0 e^{-\alpha l}, \qquad (15.82)$$

де I – інтенсивність світла, що виходить із шару поглинаючого середовища завтовшки l; I_0 – інтенсивність світла, що входить у поглинаюче середовище. За $l = 1/\alpha$ інтенсивність $I = I_0/e$. Отже, шар, товщина якого дорівнює $1/\alpha$, зменшує інтенсивність світла в $e = 2{,}72$ ... рази. Отримане співвідношення встановив у 1729 р. П. Бугер, його називають законом Бугера, або ***законом Бугера–Ламберта***. А. Бер встановив, що поглинання світла розчинами пропорційне молекулярній концентрації C_0 розчиненої речовини, тобто

$$\alpha = \alpha_0 C_0, \qquad (15.83)$$

де α_0 – коефіцієнт пропорційності, який залежить від природи розчиненої речовини і не залежить від її концентрації.

Тоді ***закону Бугера–Ламберта–Бера***, який справедливий для газів і розчинів малих концентрацій, можна надати вигляду

$$I = I_0 e^{-\alpha_0 C_0 l}. \qquad (15.84)$$

Коефіцієнт поглинання α залежить від довжини хвилі λ (або частоти $\omega = 2\pi\nu$) і від хімічної природи речовини.

В одноатомних газах і парах металів, в яких атоми розміщені на значних відстанях один від одного і їх можна вважати ізольованими, коефіцієнт поглинання для більшості довжин хвиль близький до нуля і лише для дуже вузьких спектральних областей спостерігаються різкі максимуми. Такий спектр поглинання називають ***лінійчастим*** (рис. 15.7). Ці лінії відповідають частотам власних коливань електронів в атомах.

У газів з багатоатомними молекулами спостерігаються системи тісно розміщених ліній, які утворюють ***смуги поглинання***. Структура цих смуг визначається складом і будовою молекул.

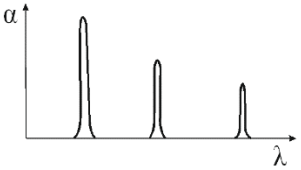


Рис. 15.7

Рідкі й тверді діелектрики мають ***суцільні спектри поглинання***, що складаються з порівняно широких смуг поглинання, у межах яких коефіцієнт поглинання змінюється плавно. За межами цих смуг $\alpha \approx 0$, тобто діелектрики прозорі.

Метали фактично непрозорі для світла. Коефіцієнт α для них має значення порядку 10^2 м$^{-1}$, тоді як для скла $\alpha \approx$ ≈ 1 м$^{-1}$. Це обумовлено наявністю в металах вільних електронів. Під дією електричного поля світлової хвилі вільні електрони починають рухатись, і в металі виникають швидкозмінні струми, що супроводжуються виділенням теплоти. У результаті енергія світлової хвилі швидко зменшується і перетворюється у внутрішню енергію металу.

Структура спектрів поглинання визначається складом і будовою молекул, тому вивчення спектрів поглинання є одним з основних методів кількісного і якісного дослідження речовин.

15.5. Розсіяння світла

З класичної точки зору процес розсіяння світла полягає в тому, що світло, яке проходить крізь середовище, викликає коливання електронів в атомах. Ці електрони, що коливаються, збуджують вторинні хвилі, які поширюються у всіх напрямках. Це явище, здавалось би, має за всіх умов приводити до розсіяння світла. Утім, вторинні хвилі є когерентними, так що треба врахувати їх взаємну інтерференцію.

З відповідного розрахунку видно, що у випадку однорідного середовища вторинні хвилі повністю гасять одна одну у всіх напрямках, крім напрямку поширення первинної хвилі, тому перерозподіл світла за напрямками, тобто розсіяння світла, не відбувається.

Вторинні хвилі не гасять одна одну в бокових напрямках тільки у разі поширення світла в неоднорідному середовищі. Світлові хвилі, зазнаючи дифракції на неоднорідностях середовища, дають дифракційну картину, яка характеризує досить рівномірний розподіл інтенсивності у всіх напрямках. Таку дифракцію на дрібних

неоднорідностях називають ***розсіюванням світла***.

Середовища з явно вираженою оптичною неоднорідністю мають назву ***каламутних*** середовищ. До них належать:

1) дим, тобто наявність у газах завислих дрібних твердих частинок;

2) тумани – гази, в яких завислі дрібні крапельки рідини;

3) суспензії, утворені твердими частинками, які плавають у рідині;

4) емульсії, тобто наявність дрібних краплинок однієї рідини в іншій, які завислі в ній і не розчиняються (прикладом емульсії може бути молоко, яке можна подати як воду, в якій завислі крапельки жиру);

5) тверді тіла, як то перламутри, опали, молочне скло і под.

Світло, розсіяне на частинках, розміри яких значно менші від довжини світлової хвилі, є *частково поляризованим.* Це пояснюють тим, що коливання електронів, викликане розсіяним світловим пучком, відбуваються у площині, яка перпендикулярна до пучка (рис. 15.8).

У результаті розсіяння світла в бокових напрямках інтенсивність у напрямку поширення зменшується швидше, ніж у випадку одного тільки поглинання, тому для каламутної речовини у виразі (15.82), поряд з коефіцієнтом поглинання α, має перебувати додатковий коефіцієнт α', обумовлений розсіянням:

$$I = I_0 e^{-(\alpha+\alpha')l}. \quad (15.85)$$

Сталу α' називають ***коефіцієнтом екстинції***.

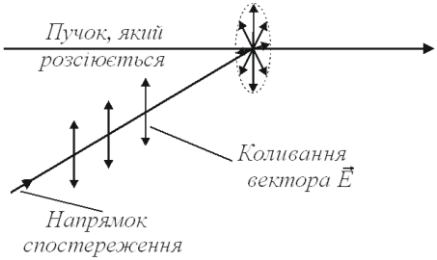


Рис. 15.8

Якщо розміри неоднорідностей малі порівняно з довжиною хвилі (не більші за $\sim 0{,}1\ \lambda$), інтенсивність розсіяного світла I пропорційна четвертому степеню частоти або обернено пропорційна четвертому степеню довжини хвилі:

$$I \sim \omega^4 \sim \frac{1}{\lambda^4}. \quad (15.86)$$

Залежність (15.86) називають ***законом Релея***. Її походження легко зрозуміти, якщо врахувати, що потужність випромінювання заряду, який коливається, пропорційна четвертому степеню частоти (див. підрозд. 11.5, формула (11.41)) і, відповідно, обернено пропорційна четвертому степеню довжини хвилі.

Якщо розміри неоднорідностей порівняльні з довжиною хвилі, електрони, що перебувають у різних місцях неоднорідності, коливаються із суттєвим зсувом за фазою. Ця обставина ускладнює явище і призводить до інших закономірностей – інтенсивність розсіяного світла стає пропорційною тільки квадрату частоти (обернено пропорційною тільки квадрату довжини хвилі).

Прояв закономірностей (15.86) легко спостерігати, якщо пропускати пучок білого світла крізь посудину з каламутною рідиною (рис. 15.9).

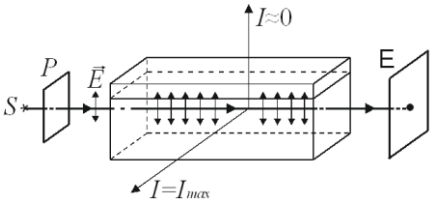


Рис. 15.9

Унаслідок розсіяння слід пучка в рідині добре видно збоку, причому, оскільки короткі світлові хвилі розсіюються значно сильніше за довгі, то цей слід стає блакитнуватим. Пучок, який проходить крізь рідину, стає збагаченим довгохвильовим випромінюванням і утворює на екрані E не білу, а червоно-жовту пляму. Якщо поставити на вході пучка в посудину поляризатор *P*, ми побачимо, що інтенсивність розсіяного світла в різних напрямках, які перпендикулярні до первинного пучка, неоднакова. Напрямок випромінювання диполя (див. рис. 11.11) приводить до того, що в напрямках, які збігаються із площиною коливань первинного пучка, інтенсивність розсіяного світла фактично дорівнює нулю, у напрямках, що перпендикулярні до площини коливань, інтенсивність розсіяного світла максимальна. Повертаючи поляризатор навколо напрямку первинного пучка, ми будемо спостерігати поперемінне підсилення та ослаблення світла, яке розсіюється у цьому напрямку.

Навіть старанно очищені від сторонніх домішок і забруднень рідини й гази у деякій мірі розсіюють світло. Встановлено, що причиною появи оптичних неоднорідностей є в цьому випадку флуктуації густини (тобто відхилення густини від її середнього значення, яке спостерігається в межах малих об'ємів). Ці флуктуації викликані хаотичним рухом молекул речовини, тому розсіяння світла, яке обумовлене ним, називають ***молекулярним***.

Молекулярним розсіянням пояснюється блакитний колір неба. Місця згущення і розрідження повітря, які безперервно виникають в атмосфері внаслідок хаотичного молекулярного руху, розсіюють сонячне світло. При цьому згідно із законом (15.86) блакитні

та сині промені розсіюються сильніше, ніж жовті й червоні, чим і обумовлюють блакитний колір неба. Коли Сонце перебуває низько над горизонтом, промені, які поширюються безпосередньо від нього, проходять більшу товщину середовища, яке розсіює, у результаті чого вони стають збагаченими більшими довжинами хвиль. З цієї причини небо забарвлене у червоні тони.

15.6. Ефект Доплера в оптичних явищах

Ефект Доплера для звукових хвиль, існування яких можливе у пружних середовищах, розглянуто в підрозд. 10.8. Перш ніж перейти до розгляду ефекту Доплера в оптичних явищах зазначимо таке: у класичній фізиці з гіпотези про існування ефіру автоматично випливало, що для світлових хвиль має виявлятися ефект Доплера так само, як і для звукових хвиль. Утім, згодом було з'ясовано, що ніякого ефіру, який міг би бути абсолютною системою відліку, не існує, що зміст має лише відносний рух джерела і приймача один відносно одного, що системою відліку може бути лише приймач або джерело світла.

З викладеного випливає, що формули, які відображають ефект Доплера в оптичних явищах, мають бути єдиними, незалежними від того, що саме переміщується – джерело чи приймач. Щоб знайти такі формули, керуватимемось принципом відносності, згідно з яким закони природи мають однаковий вигляд в усіх інерційних системах відліку.

Візьмемо для системи відліку S приймач, а для системи S' – джерело світла. Розмістимо приймач і джерело у початках відповідних систем координат, осі яких X та X' збігаються зі швидкістю переміщення джерела v (рис. 15.10).

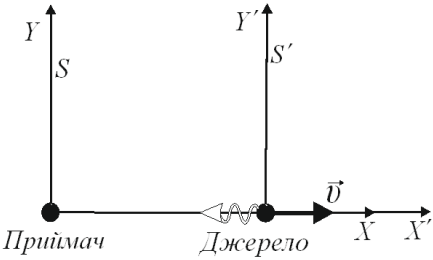


Рис. 15.10

Запишемо в системі S' рівняння плоскої світлової хвилі, що поширюється від джерела до приймача:

$$E' = E'_m \cos\omega'(t' + \frac{x'}{c}), \quad (15.87)$$

де ω' – циклічна частота хвилі в системі S'; c – швидкість світла, однакова у всіх системах відліку.

Згідно із принципом відноснос-ті в системі S рівняння хвилі, що розглядається, матиме вигляд, аналогічний формулі (15.87):

$$E = E_m \cos\omega\left(t + \frac{x}{c}\right), \quad (15.88)$$

де ω – циклічна частота хвилі, яку фіксує приймач у системі S.

Водночас від рівняння (15.87) можна перейти до рівняння (15.88), якщо за формулами перетворень Лоренца виразити координати x' та t' через x та t:

$$x' = \frac{x - vt}{\sqrt{1 - \frac{v^2}{c^2}}}; \quad t' = \frac{t - \frac{v}{c^2}x}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

Тоді отримаємо:

$$E = E'_m \cos\omega' \left(\frac{t - \frac{v}{c^2}x}{\sqrt{1 - \frac{v^2}{c^2}}} + \frac{x - vt}{c\sqrt{1 - \frac{v^2}{c^2}}} \right),$$

або в іншому вигляді:

$$E = E'_m \cos\omega' \frac{1 - \frac{v}{c}}{\sqrt{1 - \frac{v^2}{c^2}}} \left(t + \frac{x}{c} \right). \quad (15.89)$$

Зіставивши рівняння (15.88) і (15.89), знайдемо:

$$\omega = \omega' \frac{1 - \frac{v}{c}}{\sqrt{1 - \frac{v^2}{c^2}}} = \omega' \sqrt{\frac{1 - \frac{v}{c}}{1 + \frac{v}{c}}}.$$

Переходячи від циклічних до звичайних частот і позначивши частоту ν' у системі S' джерела через ν_0, дістанемо формулу, що описує поздовжній ефект Доплера, який спостерігається під час руху джерела вздовж лінії, що з'єднує його із приймачем:

$$\nu = \nu_0 \sqrt{\frac{1 - \frac{v}{c}}{1 + \frac{v}{c}}}. \quad (15.90)$$

Якщо $v \ll c$, то

$$\nu \approx \nu_0 \frac{1 - \frac{1}{2}\frac{v}{c}}{1 + \frac{1}{2}\frac{v}{c}} \approx \nu_0 \left(1 - \frac{1}{2}\frac{v}{c}\right)\left(1 - \frac{1}{2}\frac{v}{c}\right).$$

Обмежившись членами v/c, отримаємо

$$\nu = \nu_0 \left(1 - \frac{v}{c}\right).$$

Звідси відносна зміна частоти

$$\frac{\nu - \nu_0}{\nu_0} = \frac{\Delta\nu}{\nu_0} = -\frac{v}{c}. \quad (15.91)$$

Зазначимо, що з віддаленням джерела і приймача один від одного спостерігається зсув в область довших хвиль ($\nu < \nu_0$, $\lambda > \lambda_0$) – червоне зміщення, а з наближенням відбувається зсув в область коротших хвиль ($\nu > \nu_0$, $\lambda < \lambda_0$). Крім поздовжнього ефекту, для світлових хвиль існує також ***поперечний ефект Доплера***. Він полягає у зменшенні частоти, яку сприймає приймач, і спостерігається в тому випадку, коли вектор відносної швидкості напрямлений перпендикулярно вздовж прямої, що проходить через приймач і джерело (наприклад, джерело рухається по колу, в центрі якого уміщено приймач). У такому випадку

$$\nu = \nu_0 \sqrt{1 - \frac{v^2}{c^2}} \approx \nu_0 \left(1 - \frac{1}{2}\frac{v^2}{c^2}\right) < \nu_0.$$

Відносна зміна частоти за поперечного ефекту Доплера

$$\frac{\Delta\nu}{\nu_0} = -\frac{1}{2}\frac{v^2}{c^2}. \quad (15.16)$$

Вона значно менша, ніж за поздовжнього ефекту. Для звукових виль поперечного ефекту Доплера не існує.

Поздовжній світловий ефект Доплера було експериментально підтверджено дослідами А. А. Бєлопольського у 1905 р. і Б. Б. Голіцина у 1907 р.; поперечний – дослідами Айвса і Стілуела у 1938 р. В останніх дослідах була помічена зміна частоти випромінювання атомів водню в анодних променях, де їх швидкість досягла $2 \cdot 10^6$м/с. Ці досліди водночас підтверджували правильність перетворень Лоренца в теорії відносності.

Поздовжній ефект Доплера використовують в астрономії для визначення радіальної швидкості зірок. Вимірявши спектроскопом відносну зміну частоти для будь-якої лінії спектра випромінювання зірки, за формулою (15.91) можна визначити її швидкість v.

Ефект Доплера впливає на спектр випромінювання газів. Унаслідок хаотичного теплового руху молекул частоти випромінювання їх фіксуватимуться у спектрографі у межах від $\nu_0\left(1-\frac{v}{c}\right)$ до $\nu_0\left(1+\frac{v}{c}\right)$. Це зумовлюватиме розширення спектральної лінії світного газу. Отже, за розширенням спектральних ліній можна судити про температуру світного газу.

Контрольні запитання

1. Що є дисперсією світла? Яка її природа?
2. Які є види дисперсії?
3. Що означає класичне пояснення явища дисперсії?
4. Як впливає магнітне поле на рух електрона в атомі?
5. Яка природа появи оптичної активності в речовині, що міститься в магнітному полі?
6. Які властивості речовини можна визначити, досліджуючи ефект Фарадея?
7. Що характеризує групова швидкість світла? Який її зв'язок з дисперсією?
8. Поясніть суть закону Бугера–Ламперта поглинання світла в середовищах.

Розділ 16. Квантова природа випромінювання

16.1. Теплове випромінювання. Його характеристики. Абсолютно чорне тіло. Закон Кірхгофа

Світло, випромінюване джерелом, відносить із собою енергію. Залежно від того, звідки черпається ця енергія, розрізняють і види світіння.

Випромінювання, яке супроводжується хімічними перетвореннями, називають *хемілюмінесценцією*. Прикладом є світіння фосфора, який повільно окислюється на повітрі. Випромінювання променистої енергії в цьому випадку

відбувається внаслідок зменшення внутрішньої енергії тіла у разі зміни його хімічного складу.

Світіння під дією електричного поля називають *електролюмінесценцією*. У напівпровідниках випромінювання може відбуватись унаслідок рекомбінації носіїв заряду в області *p*—*n*-переходу (*інжекційна електролюмінесценція*), яка відбувається у фотодіодах. Інший приклад – світіння газів або пари в електричному розряді. У цьому випадку потрібна для випромінювання енергія надається атомам або молекулам газу електронами, які прискорюються електричним полем розряду.

Бомбардування електронним пучком може викликати також світіння твердих тіл, наприклад сірчаного цинку, який нанесений на екран електронно-променевої трубки (*катодолюмінесценція*).

Процеси випромінювання, які викликаються попереднім або одночасним освітленням тіла, об'єднують під назвою *фотолюмінесценція*. При цьому потрібна для випромінювання енергія доставляється світлом від зовнішнього джерела.

У тих випадках, коли потрібну енергію надають нагріванням, тобто підведенням теплоти, випромінювання називають ***тепловим*** або ***температурним***. З-поміж різних видів світіння воно посідає особливе місце. На противагу всім видам люмінесценції *це єдиний вид випромінювання, який може перебувати у стані термодинамічної рівноваги з тілами*.

Отже, фізика *теплового випромінювання є тією ланкою, яка пов'язує термодинаміку й оптику.* Це не тільки розширило межі можливостей термодинаміки і статистичної механіки, але й привело до таких наслідків, котрі виявились революційними в історії фізики взагалі. Велике відкриття Планка — квантова гіпотеза — і є таким наслідком.

У проблемах теплового випромінювання особливо важливе значення має поняття *рівноважного випромінювання*. Щоб встановити це поняття, розглянемо оболонку з нерухомими і непрозорими стінками, температура яких підтримується сталою (рис. 16.1). Атоми й молекули оболонки переходять у збуджені стани внаслідок енергії теплового руху і під час обернених переходів у незбуджені стани дають випромінювання, яке заповнює оболонку. Падаючи на стінки оболонки, променистa енергія частково відбивається, частково поглинається. Відбувається зміна напрямку поширення, спектрального складу, поляризації,

інтенсивності випромінювання. У результаті всіх цих процесів, як випливає із загальних законів термодинаміки, в оболонці врешті-решт встановиться макроскопічний цілком визначений стан випромінювання, за якого за кожний проміжок часу кількість променистої енергії певного кольору (частоти), напрямку поширення і поляризації, яка випромінюється, в середньому дорівнює кількості енергії того ж кольору, напрямку поширення і поляризації, що поглинається. *Як і будь-який рівноважний стан, він характеризується тим, що кожному мікропроцесу, що відбувається в системі, з тією ж ймовірністю відповідає мікропроцес, що йде в оберненому напрямку (**принцип детальної рівноваги**).*

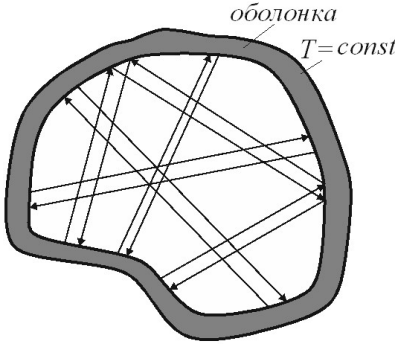


Рис. 16.1

Завдяки цьому стан випромінювання в оболонці залишається *макроскопічно незмінним* з часом. Перехід у рівноважний стан, як і будь-який статистичний процес, керується *ймовірністними законами*. В оболонці встановлюється хаотичний стан випромінювання, якому відповідає *найбільша ймовірність*. Такий стан випромінювання називають *рівноважним випромінюванням.*

Властивостями рівноважного випромінювання є такі:

1. Густина променистої енергії, її розподіл за спектром частот (довжин хвиль) і напрямком поширення, а також поляризація випромінювання *зовсім не залежать від форми і матеріалу стінок оболонки*. Ці властивості, як і стан газу в посудині, визначаються *тільки температурою стінок оболонки*.

2. Рівноважне випромінювання *однорідне*, тобто його густина однакова у всіх точках всередині оболонки. Воно *ізотропне і неполяризоване*: усі можливі напрямки поширення випромінювання є рівноймовірними, а напрямки векторів $\vec{E}$ та $\vec{H}$ у кожній точці простору хаотично змінюються з часом.

3. Оскільки випромінювання перебуває у тепловій рівновазі зі стінками, то можна говорити про температуру не тільки стінок оболонки, а і про *температуру самого випромінювання*, вважаючи обидві температури однаковими. Утім, слід підкреслити, що температура рівноважного випроміню-

вання є *властивістю самого випромінювання*, а не стінок оболонки, з якою воно перебуває у тепловій рівновазі. Про температуру випромінювання має сенс говорити й тоді, коли взагалі немає ніяких стінок. Зокрема, наприклад, густина енергії рівноважного випромінювання однозначно визначає і його температуру.

У порожній оболонці зі стінками, що є ідеально дзеркальними, поглинання і випромінювання світла не відбувається. Якщо якимось чином ввести в оболонку випромінювання, то напрямок його поширення буде змінюватись при відбиванні від стінок, але спектральний склад зберігатиметься попередній. Таке випромінювання нерівноважне й нестійке. Скільки завгодно малого відхилення властивостей стінок від ідеальності вже достатньо для того, щоб випромінювання стало рівноважним. Це відбудеться і коли стінки ідеальні, якщо внести в оболонку скільки завгодно мале тіло («порошинку»), яке здатне поглинати і випромінювати. Така «порошинка», не впливаючи на енергетичний баланс, переведе випромінювання в оболонці з будь-якого нерівноважного стану в рівноважний. Від властивостей «порошинки» залежить тільки час встановлення рівноважності.

Для експериментального вивчення спектрального складу рівноважного випромінювання можна зробити невеликий отвір у стінках оболонки, які підтримуються за певної температури. Випромінювання, яке виходить назовні через отвір, має такий самий спектральний склад, як і всередині оболонки. Від рівноважного воно відрізняється лише тим, що поширюється в межах деякого тілесного кута в одному напрямку, тобто воно не ізотропне.

Введемо поняття деяких фізичних величин, що характеризують стан випромінювання у просторі. Ці величини можна застосовувати до довільного випромінювання, а не тільки до рівноважного.

Позначимо через $U(T)$ *об'ємну густину енергії* випромінювання, тобто сумарну енергію одиниці об'єму електромагнітних хвиль усіх можливих частот ($0 \le \omega \le \infty$).

Для характеристики розподілу енергії за частотами або довжинами хвиль введемо *спектральну густину* випромінювання $u(\omega, T)$ або $u(\lambda, T)$ так, що величина $u(\omega, T)\, d\omega$ визначає енергію одиниці об'єму випромінювання з частотами в інтервалі від ω до $\omega + d\omega$, а $u(\lambda, T)\, d\lambda$ визначає енергію одиниці об'єму випромінювання з довжинами хвиль в інтервалі від λ до $\lambda + d\lambda$. Якщо

це один і той самий спектральний інтервал, то $u(\omega,T)d\omega = u(\lambda,T)d\lambda$. Вираховуючи, що $\omega = 2\pi c/\lambda$, легко виразити $u(\omega,T)$ через $u(\lambda,T)$ або навпаки:

$$u(\lambda,T) = (2\pi c/\lambda^2)u(\omega,T). \quad (16.1)$$

Очевидно, що

$$U(T) = \int_0^\infty u(\omega,T)d\omega = \int_0^\infty u(\lambda,T)d\lambda. \quad (16.2)$$

В теоретичній фізиці зазвичай користуються величиною $u(\omega,T)$, а в експериментальній фізиці віддають перевагу $u(\lambda,T)$. Для випадку рівноважного випромінювання функції $u(\omega,T)$ і $u(\lambda,T)$ залежать тільки від частоти ω (або довжини хвилі λ) і від температури випромінювання T, але не залежать від форми і матеріалу стінок оболонки. У цьому випадку $u(\omega,T)$ буде *універсальною функцією тільки* ω та T, а $U(T)$ – універсальною функцією тільки T. Знаходження функції $u(\omega,T)$ є *основним завданням* теорії теплового випромінювання.

Водночас як спектральний розподіл енергії випромінювання, що виходить з отвору в оболонці, має універсальний характер, для теплового випромінювання з відкритої поверхні тіла це не так: його спектральний розподіл залежить не тільки від температури, але і від матеріалу поверхні. Для кількісної характеристики цього спектрального розподілу вводять поняття ***випромінювальної здатності*** тіла $r(\omega,T)$ або $r(\lambda,T)$, тобто спектральної густини потоку енергії випромінювання, що випромінюється одиничною площиною поверхні у всіх напрямках (у межах тілесного кута 2π) для одиничного інтервалу частот (довжин хвиль) так, що $r(\omega,T)d\omega$ або $r(\lambda,T)d\lambda$ є випромінювання у відповідному спектральному інтервалі від ω до $\omega + d\omega$ або від λ до $\lambda + d\lambda$.

Аналогічно формулі (16.1)

$$r(\lambda,T) = \left(\frac{2\pi c}{\lambda^2}\right) r(\omega,T). \quad (16.3)$$

Інтегральна випромінювальна здатність (повний потік випромінювання для всіх частот) називають *енергетичною світністю* $R(T)$ поверхні. Очевидно, що

$$R(T) = \int_0^\infty r(\omega,T)d\omega = \int_0^\infty r(\lambda,T)d\lambda. \quad (16.4)$$

Підкреслимо, що випромінювальна здатність $r(\omega,T)$ характеризує тільки теплове випромінювання тіла (усі види люмінесценції виключаються). Для певного тіла вигляд функції $r(\omega,T)$ залежить тільки від його температури. Випромінювальна здатність тіла не залежить від навколишнього середовища, зокрема від того, чи перебувало тіло в рівновазі з випромінюванням, чи ні.

Випромінювання, яке падає на поверхню тіла, частково відбивається або розсіюється поверхнею, частково

проходить крізь поверхню. Обмежимось випадком, коли тіло непрозоре, тобто товщини його достатньо для того, щоб випромінювання, яке зайшло в нього, встигало повністю поглинутись, не досягнувши протилежної сторони. Тоді можна умовно говорити про поглинання самою поверхнею.

Поглинання, як правило, має селективний (вибірковий) характер і залежить від температури поверхні. Так, наприклад, тонкий шар сажі майже повністю поглинає видиме світло, але у значно меншій мірі інфрачервоне випромінювання. Плавлений кварцпрозорий у широкому інтервалі довжин хвиль, але починає помітно поглинати світло за температури близько 1500 °C. Для характеристики таких властивостей тіл вводять поняття *поглинальної здатності* тіла $\alpha(\omega, T)$ або $\alpha(\lambda, T)$, під якою розуміють безрозмірну величину, що показує, яка частка енергії випромінювання певної частоти (або довжини хвилі), яке падає на поверхню, поглинається нею. Стандартно $\alpha(\omega, T) \leq\leq 1$. Тіло, яке за будь-якої температури повністю поглинає випромінювання всіх частот, яке падає на нього, називають ***абсолютно чорним*** (АЧТ) або ***просто чорним***. Для нього $\alpha(\omega, T) \equiv 1$. Тіло, для якого $\alpha(\omega, T) = \alpha(T) = \text{const} < 1$, називають ***сірим***.

Між випромінювальною і поглинальною здатністю будь-якого тіла є зв'язок, який випливає із загальних принципів термодинаміки. Оскільки випромінювальна і поглинальна здатність характеризують саме тіло і не залежать від оточення, то для знаходження зв'язку між ними можна розглянути окремий випадок, коли тіло оточене рівноважним випромінюванням з тією ж температурою, наприклад, уміщене в оболонку, стінки якої є ідеально дзеркальними, або підтримуються за тієї самої температури, що й тіло. На рис. 16.2 в оболонці розміщено декілька тіл, серед яких і АЧТ. Всередині оболонки є вакуум, так що тіла можуть обмінюватись енергією між собою та оболонкою тільки шляхом випромінювання і поглинання ЕМХ.

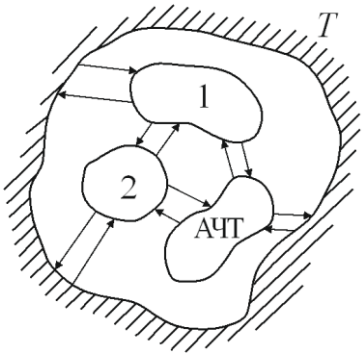


Рис. 16.2

Дослід показує, що така система через деякий час повернеться у стан рівноваги – усі тіла приймуть одну й ту

саму температуру, яка дорівнюватиме температурі оболонки T. У такому стані тіло, яке має більшу випромінювальну здатність $r(\omega,T)$, втрачає за одиницю часу з одиниці поверхні більше енергії, ніж тіло, що має меншу $r(\omega,T)$. Оскільки температура (а значить і енергія) тіл не змінюється, то тіло, яке випромінює більше енергії, має й більше поглинати, тобто мати більший $\alpha(\omega,T)$. Таким чином, чим більша випромінювальна здатність тіла $r(\omega,T)$, тим більша і його поглинальна здатність $\alpha(\omega,T)$.

Звідси випливає співвідношення

$$\left(\frac{r(\omega,T)}{\alpha\,(\omega,T}\right)_1=\left(\frac{r(\omega,T)}{\alpha\,(\omega,T}\right)_2=\left(\frac{r(\omega,T)}{\alpha\,(\omega,T}\right)_{\text{АЧТ}}=\ldots, \qquad (16.5)$$

де індекси 1, 2, АЧТ і т.д. належать до різних тіл. Співвідношення (16.5) є встановленим Кірхгофом в 1859 р. законом, який формулюють так: *відношення випромінювальної здатності тіла до його поглинальної здатності однакове для всіх тіл і є універсальною функцією частоти (довжини хвилі) та температури*,

$$\frac{r(\omega,T)}{\alpha(\omega,T)}=f(\omega,T), \qquad (16.6)$$

де $f(\omega,T)-$ *універсальна функція Кірхгофа.*

Величини $r(\omega,T)$ і $\alpha(\omega,T)$ можуть змінюватись дуже сильно під час переходу від одного тіла до іншого. Відношення ж їх є однаковим для всіх тіл. Закон Кірхгофа є точним кількісним узагальненням правила, яке емпірично встановив Прево в 1809 р. Згідно з цим правилом, якщо *поглинальна здатність тіл різна, то буде різною і їх випромінювальна здатність.* Оскільки величина $\alpha\,(\omega,T)$ не може бути більшою за одиницю, то із (16.6) випливає, що зі всіх тіл *за тієї самої температури абсолютно чорне тіло має найбільшу випромінювальну здатність.* Із закону Кірхгофа випливає також, що будь-яке тіло за певної температури випромінює переважно промені таких довжин хвиль, які воно за тієї самої температури сильніше всього поглинає.

Для АЧТ стандартно $\alpha\,(\omega,T)\equiv 1$. Тоді із формули (16.6) випливає, що $r(\omega,T)$ для АЧТ дорівнює $f(\omega,T)$. Таким чином, універсальна функція Кірхгофа $f(\omega,T)$ *є не чим іншим, як випромінювальною здатністю* АЧТ.

Між спектральною густиною енергії рівноважного випромінювання і випромінювальною здатністю АЧТ існує такий зв'язок:

$$u\,(\omega,T)=\frac{4}{c}r(\omega,T), \qquad (16.7)$$

де c – швидкість світла у вакуумі. Отже, спектральний розподіл теплового випромінювання АЧТ буде таким самим, як у рівноважного випромінювання за тієї самої температури, тому рівноважне випромінювання ще називають *чорним*

випромінюванням. Сажа і платинова чернь у видимій області спектра мають поглинальну здатність $\alpha(\omega, T)$, близьку до одиниці, але в далекій інфрачервоній області спектра $\alpha(\omega, T)$ для них помітно менша за одиницю. Абсолютно чорних тіл, як і інших ідеалізованих об'єктів, у природі немає, але можна створити пристрій, який за властивостями скільки завгодно близький до абсолютно чорного тіла. Цю оболонку показано на рис. 16.1, у її стінці зроблений невеликий отвір (рис. 16.3).

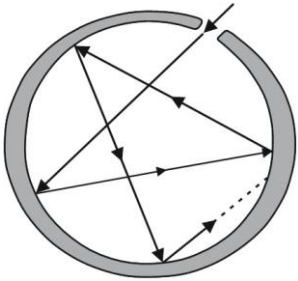


Рис. 16.3

Випромінювання, яке падає ззовні оболонки, проникає через отвір всередину, потрапляє на стінки оболонки, частково поглинається ними, частково відбивається або розсіюється і знову потрапляє на стінки. Через малі розміри отвору це відбувається багатократно, перш ніж деяка частина випромінювання знову потрапить на отвір, тому фактично все світло будь-якої частоти, яке падає на отвір, повністю поглинається. Матеріал стінок оболонки значення не має. Отвір оболонки щодо випромінювання, яке падає на нього, а також щодо теплового випромінювання, яке виходить із нього, поводить себе як поверхня АЧТ з температурою, що дорівнює температурі стінок оболонки.

Модель АЧТ можна проілюструвати простими демонстраціями. Якщо взяти коробку, яка всередині пофарбована білою фарбою, то малий отвір у ній буде здаватись зовсім чорним навіть на фоні пофарбованої чорною фарбою зовнішньої стінки. Чорними виглядають ззовні відкриті вікна будинків попри те, що у приміщенні стіни і стеля світлі. Якщо розжарити стінки оболонки з матеріалу з малою поглинальною здатністю, наприклад із фарфору, то її отвір буде яскраво світитись на тьмяному фоні прямого випромінювання стінок.

Джерело у вигляді оболонки з малим отвором, на стінках якої підтримують постійну температуру, використовують для кількісного вимірювання випромінювання чорного тіла. Аналогічним чином побудовано стандартний випромінювач, який є еталоном одиниці сили світла (кандели).

16.2. Закон Стефана–Больцмана. Закон Віна. Формула Релея–Джінса

Після встановлення закону Кірхгофа стало очевидно, що першочергове завдання теорії теплового випромінювання полягає у визначенні вигляду універсальної функції Кірхгофа $f(\omega,T)$, тобто $(r(\lambda,T)-$ випромінювальної здатності АЧТ. Утім, спочатку було знайдено залежність інтегральної (повної) випромінювальної здатності (енергетичної світності) $R(T)$ АЧТ від його температури. На основі аналізу експериментальних даних Стефан у 1879 р. дійшов висновку, що повна енергія, яку випромінює тіло у процесі теплового випромінювання, пропорційна четвертому степеню термодинамічної температури. В 1884 р. Больцман теоретично отримав формулу для енергетичної світності чорного тіла:

$$R(T)=\int_0^\infty r(\omega,T)d\omega=\sigma T^4. \quad (16.8)$$

В основу доведення формули (16.8) покладено аналогію між термодинамічними властивостями рівноважного випромінювання й ідеального газу. Суть цієї аналогії полягає у тому, що як енергія ідеального газу, так і енергія рівноважного випромінювання залежать тільки від температури; як газ, так і випромінювання здатні створювати тиск під час взаємодії з поверхнею тіла. На основі цих уявлень було з'ясовано, що енергетична світність АЧТ пропорційна четвертому степеню температури. Формула (16.8) виражає закон Стефана–Больцмана. За сучасними вимірюваннями числове значення постійної Стефана–Больцмана $\sigma = 5{,}67\cdot 10^{-8}$ Вт·м$^{-2}$·К$^{-4}$. Варто зазначити, що закон (16.8) справедливий тільки для чорного тіла. Для нечорних тіл немає простої аналітичної залежності енергетичної світності від температури. Для реальних тіл енергетична світність $R^*(T)$ може бути визначена на основі закону Кірхгофа:

$$R^*(T)=\alpha(T)\sigma T^4, \quad (16.9)$$

де $\alpha(T)$ – поглинальна здатність реального тіла.

Закон Стефана–Больцмана встановлює залежність інтегральної випромінювальної здатності АЧТ від температури, але не розкриває характеру розподілу енергії випромінювання за довжинами хвиль або частотами, тобто конкретний вигляд функції Кірхгофа $f(\omega,T)$ залишається невідомим. Основним завданням теорії теплового випромінювання є встановлення конкретного вигляду цієї функції.

Криві розподілу енергії у спектрі випромінювання АЧТ вивчав експериментально В. Ленглей у 1866 р. Як АЧТ було використано порожнину з малим отвором, а також чорну сажу. Його

досліди показали, що для заданих температур максимуми енергії випромінювання припадають на певні довжини хвиль, причому величина максимуму збільшується з підвищенням температури і зміщується у бік коротких довжин хвиль (більших частот).

Важливий крок в теоретичному вивченні рівноважного випромінювання зробив Він у 1893 р. Розглядаючи адіабатичне (без теплообміну) стискування рівноважного випромінювання в оболонці із дзеркальними стінками і враховуючи зміну частоти світла під час відбивання від дзеркала, яке рухається (ефект Доплера), Він показав, що спектральна густина $u(\omega, T)$ повинна мати вигляд

$$u\,(\omega, T) = \omega^3 \varphi\left(\frac{\omega}{T}\right), \quad (16.10)$$

де φ – деяка функція відношення частоти до температури, конкретний вид якої в межах електромагнітної теорії і феноменологічної термодинаміки встановити неможливо. Співвідношення (16.10) зводить задачу знаходження універсальної функції $u(\omega, T)$, яка залежить від двох аргументів, до задачі знаходження функції $\varphi(\omega/T)$, яка залежить тільки від одного аргумента (ω/T). Це означає, що якщо відомий спектральний розподіл енергії рівноважного випромінювання за якої-небудь довільної температури, то з допомогою формули (16.10) можна знайти його за будь-якої іншої температури.

Закон Віна (16.10) як наслідок включає і закон Стефана–Больцмана. Виражаючи випромінювальну здатність чорного тіла $r(\omega,T)$ через $u\,(\omega, T)$ за допомогою співвідношення (16.7) і переходячи від інтегрування за частотами до нової змінної $x = \omega/T$, отримаємо:

$$R(T) = \frac{c}{4}\int_0^\infty u\,(\omega, T)d\omega =$$

$$= \frac{c}{4}\int_0^\infty \omega^3\,\varphi\left(\frac{\omega}{T}\right)d\omega =$$

$$= \frac{c}{4}T^4\int_0^\infty x^3\,\varphi(x)dx = \sigma T^4.$$

Отже, інтегральна світність пропорційна T^4, оскільки інтеграл за x не залежить від T.

Вимірювання (Люммер і Прінгсгейм, 1899 р.) розподілу інтенсивності випромінювання з отвору оболонки за довжинами хвиль, на стінках якої підтримується постійна температура, показали, що за деякого значення довжини хвилі λ_m крива $r(\lambda, T)$ має чітко виражений максимум (рис. 16.4).

Ці досліди підтвердили результати В. Ленглея (1866 р.), що з підвищенням температури максимум зміщується в область більш коротких довжин хвиль. Це зміщення відбувається так, що добуток температури на довжину

хвилі, яка відповідає максимуму випромінювання, залишається незмінним:

$$\lambda_m \cdot T = b. \qquad (16.11)$$

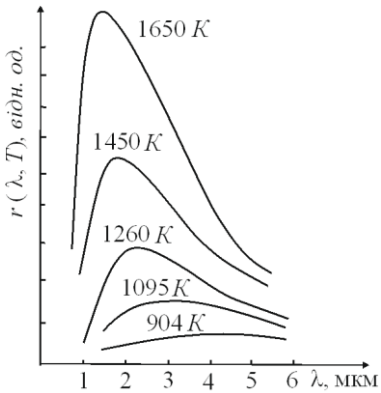


Рис. 16.4

Значення константи $b = 2{,}9 \times$ $\times 10^{-3}$ м $\cdot K$ визначено з дослідних даних. Співвідношення (6.11) називають ***законом зміщення Віна***. Воно безпосередньо випливає з термодинамічної формули Віна (16.10). Переходячи в ній від частот до довжин хвилі за схемою (16.1), отримуємо:

$$u(\lambda, T) = \left((2\pi c)^4/\lambda^5\right)) \cdot \varphi\left(\frac{2\pi c}{\lambda T}\right) = T^5 \varphi_1(\lambda T), \qquad (16.12)$$

де $\varphi_1(\lambda T) = (2\pi c)^4/(\lambda T)^5) \cdot \varphi(2\pi c/(\lambda T)$ є новою невизначеною функцією від аргумента λT. Позначимо $x = \lambda T$. За фіксованої температури для визначення положення максимуму залежності $u(\lambda, T)$ від λ отримаємо рівняння $\frac{d\varphi_1(\lambda T)}{d\lambda} = T\frac{d\varphi_1(x)}{dx} = 0$. Інакше $\frac{d\varphi_1(x)}{dx} = 0$, яке має деякий корінь $x = x_m$. Це означає, що за будь-якої температури максимум залежності $u(\lambda, T)$ від λ отримують за одного й того самого значення аргумента $x = x_m$, тобто $\lambda_m T = x_m = \text{const}$, що й доказує закон зміщення (16.11). Із (16.12) випливає, що значення спектральної густини $u(\lambda, T)$ у максимумі заи $(\lambda = \lambda_m)$ пропорційне п'ятому степеню термодинамічної температури.

За температур, яких досягають практично, максимуми випромінювальної здатності чорного тіла лежать в інфрачервоній області спектра. Так, якщо $T = 3000$ K, то $\lambda_m \approx 0{,}97$мкм. Тільки за $T > 5000$ K максимум потрапляє в зелену область спектра, до якої око найбільш чутливе. Світність чорного тіла швидко зростає з підвищенням температури $(\sim T^4)$ і за високих температур досягає дуже великих значень. Якщо $T = 6000$ K (температура поверхні Сонця), з кожного квадратного сантиметра чорного тіла випромінюється потік потужністю 7,4 кВт, але тільки невелика частина цієї потужності припадає на видиму область спектра.

Розглянуті вище закономірності теплового випромінювання використовують в оптичних методах вимірювання температури (див. підрозд. 16.4).

У 1896 р. Він опублікував роботу «Про розподіл енергії у спектрі випромінювання чорних тіл», в якій він навів доведення формули (16.10) у такому вигляді:

$$u\,(\omega,T)=\frac{h\omega^3}{2\pi^3c^3}\exp\left\{-\frac{a\omega}{2\pi T}\right\},\quad(16.13)$$

де h та a – деякі константи. Під час отримання цієї формули Він застосував до випромінювання статистику. Він вважав, що густина випромінювання має бути пропорційна кількості молекул, які випромінюють і розподілені за швидкостями згідно із законом Максвелла, а частота їх коливань пропорційна квадрату швидкості молекул. Формула (16.13) узгоджувалась із вимірюваннями Пашена в короткохвильовій області спектра, які були відомі на той час. Фактично Пашен отримав формулу (16.13) як емпіричну. В 1897 р. О. Луммер і Е. Прінсгейм показали, що формула (16.13) добре узгоджується з експериментом тільки в області коротких хвиль (високих частот).

У 1892–1905 рр. англійські вчені Д. Релей і Д. Джінс спробували теоретично отримати залежність $u(\omega,T)$ для рівноважного випромінювання всередині оболонки. Вони вважали, що таке випромінювання – це система стоячих хвиль. Їхні розрахунки показали, що кількість стоячих хвиль в одиниці об'єму, яка припадає на одиничний інтервал частот,

$$n(\omega)=\frac{\omega^2}{\pi^2c^3}.\quad(16.14)$$

Тоді

$$u\,(\omega,T)=n(\omega)\,\langle\varepsilon\rangle=\frac{\omega^2}{\pi^2c^3}\langle\varepsilon\rangle,\quad(16.15)$$

де $\langle\varepsilon\rangle$ – середня енергія, яка припадає на одну стоячу хвилю. Формула (16.15) отримана строго, без будь-яких припущень.

Релей і Джінс, виходячи з теореми класичної статистики про рівномірний розподіл енергії за степенями вільності, припустили, що на кожне електромагнітне коливання припадає в середньому енергія, яка дорівнює двом половинкам kT, одна половинка – на електричну, друга – на магнітну енергію хвилі (за класичними уявленнями на кожний степінь вільності припадає в середньому енергія, що дорівнює $\frac{1}{2}kT$). Взявши

$$\langle\varepsilon\rangle=kT,\quad(16.16)$$

Релей та Джінс із (16.15) отримали

$$u\,(\omega,T)=\frac{\omega^2}{\pi^2c^3}kT.\quad(16.17)$$

За допомогою формули (16.7) перейдемо від $u\,(\omega,T)$ до $r(\omega,T)$.

Таким чином, для випромінювальної здатності чорного тіла отримуємо вираз

$$r(\omega,T)=\frac{\omega^2}{4\pi^2c^2}kT,\quad(16.18)$$

або

$$r(\lambda, T) = \frac{2\pi c}{\lambda^4} kT. \quad (16.18')$$

Вирази (16.17), (16.18) та (16.18′) називають формулою Релея–Джінса. Ці формули дуже добре описують розподіл енергії у довгохвильовій частині спектра випромінювання, тобто для малих частот (великих довжин хвиль). Для великих частот вони непридатні.

Розбіжність між експериментальною кривою (суцільна лінія) і кривою, отриманою за допомогою формули Релея–Джінса (16.18′) (штрихова крива), видно на рис. 16.5. Крім того виявилось, що спроба вивести закон Стефана–Больцмана за допомогою формули Релея–Джінса призвела до абсурдного результату, а саме:

$$U(T) = \int_0^\infty \frac{\omega^2}{4\pi^2 c^2} kT \, d\omega = \infty.$$

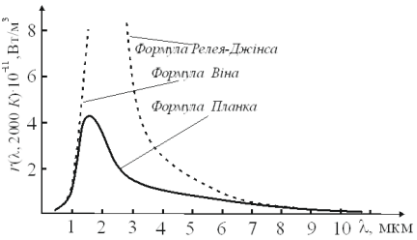


Рис. 16.5

Досліди заперечують такий результат. Ця ситуація, що створилась у теорії теплового випромінювання, П. Еренфест образно назвав «ультрафіолетовою катастрофою». Назва пов'язана з тим, що зменшення довжини хвилі (збільшення частоти) згідно з (16.17) обумовлює необмежене зростання $u(\omega, T)$. Проблема отримання функції розподілу енергії у спектрі випромінювання АЧТ, за висловленням У. Кельвіна, це «*одна з темних хмар на горизонті класичної фізики, що затьмарила її чисте небо*». Як висловився з цього приводу Лоренц, рівняння класичної фізики виявилися нездатними пояснити, чому піч, що згасла, не випромінює синіх променів разом з випромінюванням променів великих довжин хвиль.

16.3. Квантова гіпотеза і формула Планка

Планк почав публікувати роботи з теорії теплового випромінювання у 1895 р. Проблема особливо притягувала його тим, що тут мова йшла не про часткову закономірність, а про «абсолютну». Планк розглядав оболонку (рис. 16.3) як набір лінійних, пружних диполів (осциляторів, див. підрозд. 17.4), які перебувають в термодинамічній рівновазі з випромінюванням. До 1900 р. у нього були два основні результати. Суть першого в тому, що, застосувавши рівняння Максвелла до лінійного пружного осцилятора, який перебуває в оболонці у стані рівноваги з тепловим випромінюванням, він отримав формулу

$$u(\omega, T) = \frac{\omega^2}{\pi^2 c^3} \langle \varepsilon \rangle, \quad (16.19)$$

яка збігалася з формулою (16.14). Порівнявши (16.19) з формулою Віна (16.14), яку Планк розглядав як емпіричну, він отримав

$$\langle\varepsilon\rangle = \frac{\hbar\omega}{2\pi}\exp\left\{-\frac{a\omega}{2\pi T}\right\}. \quad (16.20)$$

Із досліду для коефіцієнта h було знайдено значення $h = 6{,}885 \cdot 10^{-34}$ Дж· с. Так **18 травня 1899 р.** народилась стала Планка, якій судилась така велика доля: $a = h/k$, де $k -$ стала Больцмана. Отримані Планком значення сталих h і k дуже близькі до сучасних:

$$h = 6{,}626 \cdot 10^{-34}\ \text{Дж} \cdot \text{с};$$

$$\hbar = \frac{h}{2\pi} = 1{,}054 \cdot 10^{-34}\ \text{Дж} \cdot \text{с};$$

$$k = 1{,}38 \cdot 10^{-23}\ \text{Дж/К}.$$

У нових позначеннях (16.20) можна переписати так:

$$\langle\varepsilon\rangle = \hbar\omega \cdot \exp\left\{-\frac{\hbar\omega}{kT}\right\}. \quad (16.21)$$

Іншим результатом було те, що на початку 1900 р. Планк, як він думав, довів те, на що раніше сподівався, формула Віна (16.13) є наслідком термодинаміки.

Планк закінчував доведення в момент, коли вимірювання за малих ω поставили під сумнів формулу Віна. Незабаром він відмовився і від доведення, і від самої формули. Тим не меншу 19 жовтня 1900 р. Планк знайшов свою знамениту формулу, яка визначає спектральну густину енергії випромінювання для всього спектра частот:

$$u\,(\omega, T) = \frac{\omega^2}{\pi^2 c^3} \cdot \frac{\hbar\omega}{\exp\left\{\frac{\hbar\omega}{kT}\right\} - 1}. \quad (16.22)$$

Залишилось знайти її теоретичне обґрунтування. Знайшовши помилку у своєму старому доведенні Планк виявив, *що з однієї термодинаміки функцію Кірхгофа $f(\omega, T)$ визначити неможливо.* Для теоретичного виведення формули (16.22) потрібна гіпотеза, яка докорінно протирічить уявленням класичної фізики. Рівномірно розподіляти енергію між осциляторами, аналогічно Релею і Джінсу, безперспективно, тому її треба розподіляти нерівномірно. Якщо $\langle\varepsilon\rangle$ є середньою енергією осцилятора, то сумарна енергія *N* осциляторів буде $N\langle\varepsilon\rangle$. Щоб цю енергію розподілити нерівномірно між *N* осциляторами, слід ввести порцію енергії ε. Таких порцій може бути $n = N\langle\varepsilon\rangle/\varepsilon$, і їх потрібно розподілити між осциляторами так, щоб система осциляторів перебувала в рівноважному, найбільш вірогідному стані. Таку задачу розв'язав Больцман, розв'язок цієї задачі давав формулу (16.26).

Наведемо отримання формули Планка (16.22), коли не енергію розподіляють між осциляторами, а осцилятори за енергіями. Нехай енергія осцилятора може набувати не будь-яких, а тільки певних визначених дискретних значень енергії ε_n, які відділені одне від

одного кінцевими інтервалами. Перехід осцилятора з одного стану в інший супроводжується поглинанням або випромінюванням кінцевої порції (кванта) енергії випромінювання. У такій системі з дискретним енергетичним спектром середню енергію $\langle\varepsilon\rangle$ у стані теплової рівноваги за температури T вже не можна знаходити за формулою (16.16). Ймовірність p_n того, що осцилятор перебуває у стані з енергією ε_n, відповідно до розподілу Больцмана пропорційна $\exp\{-\varepsilon_n/(kT)\}$, але під час розрахунків середніх значень інтеграли замінюють сумами

$$\langle\varepsilon\rangle = \sum_n \varepsilon_n p_n = \frac{\sum_n \varepsilon_n \exp\{-\varepsilon_n/(kT)\}}{\sum_n \exp\{-\varepsilon_n/(kT)\}}, \quad (16.23)$$

де $p_n = \frac{\exp\{-\varepsilon_n/(kT)\}}{\sum_n \exp\{-\varepsilon_n/(kT)\}}$.

Ще одна гіпотеза потрібна для встановлення значень енергії ε_n дозволених станів осцилятора. Згідно із припущенням Планка гармонічний осцилятор має еквідистанційний енергетичний спектр так, що енергія ε_n у стані з номером n становить ціле кратне найменшій порції енергії ε; $\varepsilon_n = n\varepsilon$, де $n = 0,\ 1,\ 2,\ 3,\ \ldots$ (сучасна квантова механіка для рівнів енергії гармонічного осцилятора дає значення $\varepsilon_n = (n + \frac{1}{2})\,\hbar\omega$). Тоді сума у знаменнику формули (16.23), є нескінченною геометричною прогресією зі знаменником $\exp\{-\beta\varepsilon\}$, де $\beta = \frac{1}{kT}$. Ця сума

$$S(\beta) = \sum_{n=0}^{\infty} \exp\{-\beta n\varepsilon\} = \\ = (1 - \exp\{-\beta\varepsilon\})^{-1}. \quad (16.24)$$

Легко побачити, що сума в численнику формули 16.23) дорівнює похідній $\frac{dS(\beta)}{d\beta}$, взятій із протилежним знаком:

$$\begin{cases} \sum_{n=0}^{\infty} n\varepsilon \exp\{-\beta n\varepsilon\} = \\ = -\frac{d}{d\beta}\sum_{n=0}^{\infty} \exp\{-\beta n\varepsilon\} = \\ = (1 - \exp\{-\beta\varepsilon\})^{-2} \cdot \exp\{-\beta\varepsilon\} \cdot \varepsilon\,. \end{cases} \quad (16.25)$$

Підставимо (16.25), (16.24) в (16.23), і знайдемо, що

$$\langle\varepsilon\rangle = \frac{\varepsilon}{\exp\left\{\frac{\varepsilon}{k\,T}\right\} - 1}. \quad (16.26)$$

Цей результат збігається з класичним виразом (16.16) тільки у граничному випадку за $\varepsilon/(kT) \to 0$. В оригінальній роботі Планка 1900 р. середню енергію осцилятора знаходили розрахунком середньої енергії на основі співвідношення Больцмана $S = \ln W$ між ентропією S і термодинамічною ймовірністю стану системи W. Підставляючи отримане значення середньої енергії квантового осцилятора (16.26) в (16.19), отримаємо формулу

$$u(\omega,T)=\frac{\omega^2}{\pi^2c^3}\cdot\frac{\varepsilon}{\exp\left\{\frac{\varepsilon}{kT}\right\}-1}. \quad (16.27)$$

Щоб вона не протирічила термодинамічній формулі Віна (16.10), згідно з якою температура може входити в $u(\omega,T)$ тільки в комбінації $\frac{\omega}{T}$, треба взяти, що

$$\varepsilon=\hbar\omega. \quad 6.28)$$

У результаті з формули (16.27) ми отримуємо формулу (16.22).

14 грудня 1900 р. Планк зробив доповідь, в якій навів повний розв'язок поставленої задачі й отримання формули (16.22). Цю дату можна вважати днем народження квантової теорії. Планк довів, що формулу (16.22) можна вивести тільки в тому випадку, якщо допустити квантування енергії, яке протирічить класичним уявленням. Формула (16.22) у змінних λ, T має вигляд

$$u(\lambda,\mathrm{T})=(\frac{8\pi hc}{\lambda^5})/(\exp\left\{\frac{hc}{\lambda kT}\right\}-1), \quad (16.29)$$

а для випромінювальної здатності АЧТ

$$r(\lambda,T)=\frac{c}{4}\,u(\lambda,T)=$$
$$=(\frac{2\pi hc^2}{\lambda^5})/(\exp\left\{\frac{hc}{\lambda kT}\right\}-1). \quad (16.30)$$

Формула Планка (16.22) (те саме, що і (16.30)) чудово узгоджується з дослідом за всіх частот і температур. Для малих частот (довгих хвиль) і за високих температур, коли $\hbar\omega/(kT)\le 1$, експоненту $\exp\{\hbar\omega/(kT)\}$ можна розкласти в ряд за степенями змінної $\hbar\omega/(kT)$. Якщо обмежитись лінійним членом цього розкладу, то формула Планка (16.22) переходить у формулу Релея–Джінса (16.17). Утім, в області високих частот і низьких температур формула Планка передбачає зовсім іншу поведінку $u(\omega,T)$, ніж класична теорія. За $\hbar\omega/(kT)\gg 1$ експонента $\exp\{\hbar\omega/(kT)\}$ у знаменнику формули (16.22) значно більша від одиниці, тому поведінку $u(\omega,T)$ в області коротких хвиль приблизно описують формулою

$$u\,(\omega,T)\approx\frac{\hbar\omega^3}{\pi^2c^3}\exp\left\{-\frac{\hbar\omega}{kT}\right\}, \quad (16.31)$$

яку запропонував Він у 1896 р. (див формулу (16.13)). Експоненціальний спад спектральної густини з ростом частоти усуває «ультрафіолетову катастрофу» класичної теорії.

Формула Планка (16.30) дає змогу не тільки встановити закон Стефана–Больцмана (16.8), але і знайти значення сталої σ, а саме:

$$\sigma=\frac{2\pi^5k^4}{15c^2h^3}=(5{,}67032\pm 0{,}00071)\times$$
$$\times 10^{-8}\ \text{Вт}/(\text{м}^2\cdot\text{К}^4).$$

З формули (16.30) можна отримати закон зміщення Віна і значення сталої b. Для цього потрібно розв'язати рівняння

$$\frac{dr(\lambda,T)}{d\lambda}=0,$$

з якого отримаємо

$$T\lambda_m=\frac{hc}{4{,}965k}=b=$$

$$= (2{,}897790 \pm 0{,}000090) \times$$

$$\times 10^{-3}\ \text{м} \cdot \text{К} \cong 2{,}9 \cdot 10^{-3}\ \text{м} \cdot \text{К}$$

Під час вирішення проблеми теплового випромінювання виникла стала Планка $\hbar$ (або $h = 2\pi\hbar$), яка у квантовій фізиці відіграє таку саму роль, як швидкість світла c в релятивістській фізиці. Ці світові фундаментальні константи визначають межі застосування класичного описання. В масштабах макросвіту числове значення сталої Планка надзвичайно мале. Цим пояснюється широке застосування класичної фізики, в основі якої лежить концепція неперервності під час описання макроскопічних явищ. Значення $\hbar$, отримані на основі різних фізичних явищ (теплове випромінювання, фотоефект, короткохвильова межа суцільного рентгенівського спектра, ефект Джозефсона у надпровідності та ін.), добре узгоджуються одні з одними.

Згідно із сучасним уявленням рівноважне випромінювання в оболонці розглядають як фотонний газ, до якого застосовують статистику Бозе–Ейнштейна. Середня кількість фотонів у квантовому стані

$$< n > = \frac{1}{\exp\left\{\frac{\hbar\omega}{kT}\right\} - 1},$$

де $\hbar\omega$ – енергія фотона. Кількість таких станів $n(\omega) = \frac{\omega^2}{\pi^2 c^3}$, тому спектральна густина енергії випромінювання

$$u(\omega, T) = < n > \cdot\, \hbar\omega \cdot n(\omega) =$$

$$= \frac{\omega^2}{\pi^2 c^3} \frac{\hbar\omega}{\exp\left\{\frac{\hbar\omega}{kT}\right\} - 1},$$

що збігається з формулою Планка (16.22).

16.4. Оптична пірометрія

Для вимірювання температур нагрітих тіл як термометричний параметр можна використати теплове випромінювання. Сукупність оптичних методів визначення температур, в основу яких покладено закони рівноважного випромінювання, називають оптичною термометрією, а методи вимірюванн високих температур називають *оптичною пірометрією.* Оптичні методи вимірювання температур не потребують безпосереднього контакту вимірювальних приладів із досліджуваним тілом. Завдяки цьому можна вимірювати високі температури, а також температури досить віддалених тіл. Застосування цих методів не змінює температуру досліджуваних тіл.

Залежно від того, який із законів теплового випромінювання покладено в основу методу вимірювання температур, розрізняють три умовні температури: *енергетичну* або *радіаційну*; *яскравісну*; *кольорову*. Вони функціонально пов'язані із справжньою температурою тіла і його випромінювальною здатністю. Під *радіаційною* розуміють температуру АЧТ $T_{\text{Р}}$, за якої його випромінювальна

здатність $R(T)$ однакова з випромінювальною здатністю $R^*(T)$ тіла, температуру якого визначають. Відповідно до визначення радіаційної температури та закону Стефана–Больцмана для сірих тіл (16.9) маємо:

$$R(T) = R^*(T) = \sigma T_{\text{р}}^4 = \alpha(T)\sigma T^4. \quad (16.32)$$

Звідси

$$T_{\text{р}} = T\sqrt[4]{\alpha(T)}. \quad (16.33)$$

Оскільки $\alpha(T) < 1$, то радіаційна температура менша від справжньої температури тіла. Радіаційну температуру вимірюють радіаційним пірометром (рис. 16.6), оптична система якого дає змогу сфокусувати чітке зображення джерела світла S на приймач $П$ так, щоб зображення повністю покрило поверхню приймача. За такої умови енергія випромінювання джерела, що переноситься за одиницю часу на приймач, не залежатиме від відстані між джерелом і приймачем. Приймачами у пірометрах часто є термобатареї або болометри. Пірометри попередньо градуюють за допомогою АЧТ, температура якого відома.

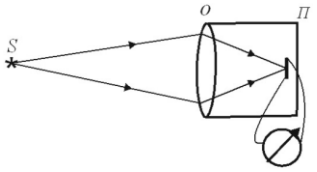


Рис. 16.6. Схема радіаційного пірометра

Під яскравісною розуміють температуру АЧТ, за якої його спектральна випромінювальна здатність $r(\lambda, T_{\text{я}})$ дорівнює спектральній випромінювальній здатності $r^*(\lambda, T)$ досліджуваного тіла для тієї самої довжини хвилі, яка згідно з (16.6) дорівнює $\alpha(\lambda, T) \cdot r(\lambda, T)$, тобто

$$r(\lambda, T_{\text{я}}) = r^*(\lambda, T) = \alpha(\lambda, T) \cdot r(\lambda, T), \quad (**)$$

де T – справжня температура тіла; $T_{\text{я}}$ – яскравісна температура; $\alpha(\lambda, T)$ – поглинальна здатність тіла.

Оскільки для світлових хвиль видимої частини спектра, що випромінюються нагрітими тілами, температура яких $T \sim 3000$ К, величина $\frac{hc}{kT\lambda} \approx 7$, і $e^7 \gg 1$, то, нехтуючи одиницею у формулі Планка (16.30), вираз $(**)$ можна записати так:

$$\exp\left\{\left(\frac{1}{T} - \frac{1}{T_{\text{я}}}\right)\frac{hc}{k\lambda}\right\} = \alpha(\lambda, T). \quad (16.34)$$

Із співвідношення (16.34) отримаємо вираз

$$\frac{1}{T} - \frac{1}{T_{\text{я}}} = \frac{k\lambda}{hc}\ln\alpha(\lambda, T), \quad (16.35)$$

який пов'язує яскравісну і справжню температури тіла. Яскравісну температуру вимірюють пірометрами із зникаючою ниткою (рис. 16.7 та 16.8).

Принцип дії такого пірометра полягає у тому, що за допомогою об'єктива O зображення поверхні нагрітого тіла, температуру якого визначають, суміщають із площиною

нитки розжарення лампи Л (рис. 16.7), увімкненої в електричне коло. Поверхня тіла і нитка розжарення спостерігаються через окуляр O_1. Перед окуляром міститься світлофільтр $Ф$, який пропускає досить вузьку частину спектра (λ_0 = 0,66 мкм). Під час вимірювань за допомогою реостата змінюють струм через лампу Л доти, поки яскравість нитки розжарення лампи і поверхні тіла не стане однаковою, при цьому зображення нитки лампи Л ніби зникає на фоні поверхні тіла (рис. 16.9), тому $T_я$ називають яскравісною температурою тіла.

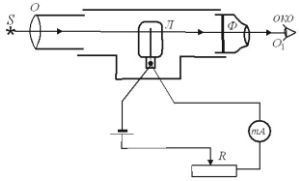


Рис. 16.7. Принципова схема роботи пірометра із зникаючою ниткою

Силу струму, що живить лампу Л, у цей момент фіксує амперметр *3* (рис. 16.8 та 16.10), який попередньо градуюють у градусах Цельсія за допомогою випромінювання АЧТ і має дві шкали: верхня шкала – 800 … 1400 ℃, нижня – 1200 … 2000 ℃ (рис. 16.10). Вимірювання за нижньою шкалою проводять з уведенням послаблюючого світлофільтра *Ф* (*4* на рис. 16.8).

Еталонна лампа Л (рис. 16.7) призначена для роботи за температур не вище 1400 ℃. Перегрівання пірометричної лампи спричиняє зміни її характеристик і порушення градуювання пірометра, тому у разі вимірювання температур вище за 1200 ℃ треба послаблювати яскравість випромінювання досліджуваного тіла. Для цього вводять послаблюючий світлофільтр, повертаючи гвинт *4* (рис. 16.8) на чверть повного оберта (при цьому біла крапка на гвинті *4* має збігатись з індексом «20» на корпусі пірометра).

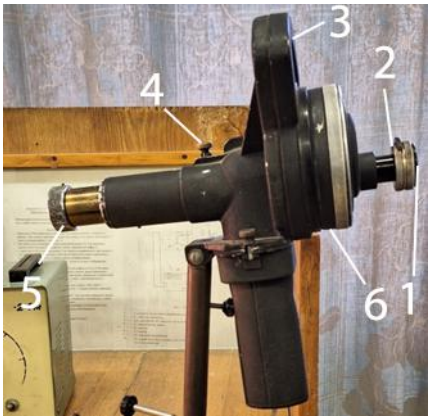
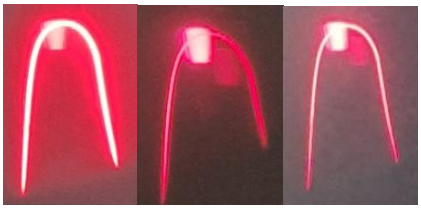


Рис. 16.8. Зовнішній вигляд пірометра із зникаючою ниткою: 1 – окуляр; 2 – червоний світлофільтр; 3 – шкала амперметра; 4 – гвинт послаблюючого світлофільтра; 5 – об'єктив; 6 – реостат

Якщо величина $\alpha(\lambda_0, T)$ тіла, яке випромінює, відома, то за яскравісною температурою можна визначити справжню температуру тіла.



а *б* *в*

Рис. 16.9. Вигляд в окулярі пірометра: а – дужка лампи пірометра перегріта; б – дужка лампи пірометра не догріта; в – яскравість дужки лампи пірометра збігається з яскравістю спіралі лампи розжарення



Рис. 16.10. Дві шкали пірометра

Визначення термодинамічної (справжньої) температури вольфрамової спіралі лампи розжарювання

Будемо вважати, що за допомогою пірометра визначили яскравісну температуру спіралі лампи Л (рис. 16.7) для довжини хвилі $\lambda_0 = 0{,}66$ мкм. Тоді із формули (16.35) можемо знайти різницю між термодинамічною (справжньою) температурою T і яскравісною $T_я$. Запишемо:

$$T - T_я = \Delta T.$$

Тоді

$$\frac{T_я - T}{T \cdot T_я} = \frac{k\lambda_0}{hc} \ln \alpha(\lambda_0, T). \quad (16.36)$$

Будемо вважати, що різниця $\Delta T \ll T$ і $T_я$. Тоді вираз (16.36) перепишемо так:

$$\Delta T = -\frac{k\lambda_0}{hc} T_я^{\,2} \ln \alpha(\lambda_0, T), \quad (16.37)$$

Або, враховуючи сталі величини,

$$\Delta T = -4{,}5873 \cdot 10^{-5} T_я^{\,2} \ln \alpha(\lambda_0, T). \quad (16.38)$$

Для нечорних тіл, яким є вольфрам, $\alpha(\lambda, T) < 1$, $\ln \alpha(\lambda, T) < 0$, $\Delta T > 0$, і тому яскравісна температура завжди менша від справжньої (термодинамічної) температури. Якщо відомі $T_я$ та $\alpha(\lambda, T)$, то можна визначити ΔT, і отже, термодинамічну температуру спіралі лампи розжарення.

Під кольоровою розуміють температуру АЧТ, за якої спектральний склад його випромінювання однаковий із спектральним складом випромінювання досліджуваного тіла, тобто відношення спектральних випромінювальних здатностей для двох заданих довжин хвиль однакові. За визначенням *поняття кольорової температури застосовують до сірих тіл*. Для тіл, в яких значною

мірою проявляється селективність випромінювання, поняття кольорової температури не має змісту. Якщо для сірого тіла $r^*(\lambda_1, T)/r^*(\lambda_2, T) = \eta$, то така ж рівність справедлива і для АЧТ:

$$r(\lambda_1, T_к)/r(\lambda_2, T_к) = \eta.$$

На основі формули Планка (16.30) це співвідношення запишемо так:

$$\eta = \left(\frac{\lambda_2}{\lambda_1}\right)^5 \cdot \frac{\exp\left\{\frac{hc}{kT_к\lambda_2}\right\} - 1}{\exp\left\{\frac{hc}{kT_к\lambda_1}\right\} - 1}. \qquad (16.39)$$

Звідси, враховуючи, що $\exp\left\{\frac{hc}{kT_к\lambda}\right\} \gg 1$, отримаємо вираз для кольорової температури:

$$T_к = \frac{\frac{hc}{k}\left(\frac{1}{\lambda_2} - \frac{1}{\lambda_1}\right)}{\ln\eta - 5\ln\frac{\lambda_2}{\lambda_1}}. \qquad (16.40)$$

Ряд реальних випромінювачів, наприклад оксиди металів, кераміка, вогнетривкі вироби, графіт та інші, фактично є сірими тілами.

Кольорову температуру $T_к$ вимірюють за червоно-синім відношенням інтенсивності для $\lambda_1 = 0{,}665$ мкм і $\lambda_2 = 0{,}470$ мкм. Для несірих тіл кольорова температура здебільшого вища за справжню температуру, тобто $T_к > T$. Кольорова температура більшості твердих тіл і рідин значно менше відрізняється від їх справжньої температури порівняно з радіаційною або яскравісною.

Отже, за допомогою оптичних пірометрів вимірюють не справжні, а умовні температури: радіаційну, яскравісну та кольорову.

16.5. Зовнішній фотоефект і його закони. Рівняння Ейнштейна для зовнішнього фотоефекту

Фотоелектричним ефектом *називають електричні явища, які супроводжуються поглинанням світлового випромінювання в речовині.*

Розрізняють:

1) *зовнішній фотоефект* – виривання електронів з речовини під дією світла у вакуум або в інше середовище. Практичне значення має фотоефект із твердого тіла у вакуум;

2) *внутрішній фотоефект*, за якого відбувається лише перерозподіл електронів за енергетичними станами в конденсованому середовищі внаслідок поглинання електромагнітного випромінювання. У напівпровідниках і діелектриках фотоефект проявляється у зміні електропровідності середовища та діелектричної проникності;

3) *фотогальванічний ефект*, за якого на межі поділу напівпровідника і металу або на межі поділу двох напівпровідників під впливом опромінювання виникає електрорушійна сила (за відсутності зовнішнього електричного поля);

4) *фотоефект в газоподібному середовищі*, який полягає у фотоіонізації окремих молекул або атомів.

Розглянемо закономірності зовнішнього фотоефекту. Фотоефект відкрив у 1887 р. Г. Герц, який виявив, що під час освітлення негативного електрода іскрового розрядника ультрафіолетовими променями розряд відбувається за меншої напруги між електродами, ніж без такого освітлення.

Перші фундаментальні дослідження фотоефекту виконав у 1888–1889 рр. О. Г. Столєтов за допомогою установки, що показана на рис. 16.11.

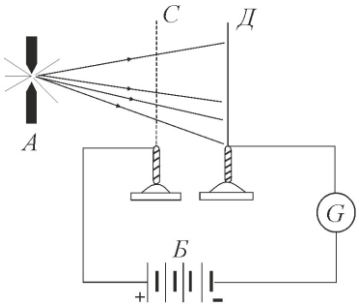


Рис. 16.11

Конденсатор, який утворений дротяною сіткою з міді *С* (анод) і суцільною цинковою пластинкою *Д* (катод), був послідовно ввімкнений з гальванометром *G* у коло акумуляторної батареї *Б*. Під час освітлення негативно зарядженої пластини *Д* світлом від джерела *А* в колі виникає електричний струм, який називають фотострумом. На основі своїх дослідів Столєтов дійшов таких висновків:

1) найбільшу дію чинять ультрафіолетові хвилі;

2) сила струму зростає зі збільшенням освітленості пластини;

3) заряди, які випускаються із пластини *Д* під дією світла, мають від'ємний знак.

У 1898 р. Леонард і Томсон методом відхилення зарядів в електричному і магнітному полях визначили питомий заряд q/m частинок, що вириваються світлом з катода, довівши, що ці частинки є електронами.

Леонард та інші дослідники удосконалили прилад Столєтова, помістивши електроди у вакуумну трубку (рис. 16.12). Катод *К* покривали металом, який досліджували та освітлювали монохроматичним світлом, яке проходило через кварцове вікно. Напругу *U* між катодом *К* та анодом *А* можна регулювати за допомогою потенціометра *П* і вимірювати вольтметром *V*. Дві акумуляторні батареї $Б_1$ і $Б_2$ дають можливість за допомогою потенціометра *П* змінити не лише абсолютну величину, а й знак напруги U. Цей пристрій дав можливість дослідити вольт-амперну характеристику фотоефекту – залежність фотоструму *I* від напруги між електродами (рис. 16.13).

Така залежність відповідає двом різним освітленостям E катода. Зі збільшенням напруги U фотострум I поступово зростає, тобто дедалі більша кількість фотоелектронів досягає анода. Максимальне значення фотоструму $I_н$, яке називають фотострумом насичення, відповідає таким значенням U, за яких усі електрони, що вибиваються з катода світлом, досягають анода: $I_н = en$, де n – кількість електронів, які вилітають з катода за 1 с.

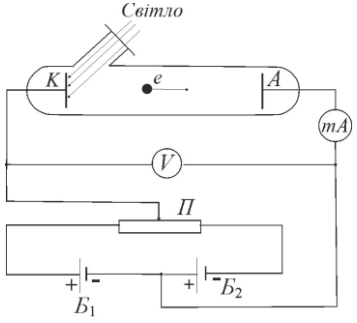


Рис. 16.12

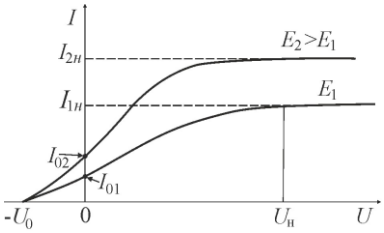


Рис. 16.13

З вольт-амперної характеристики видно, що за $U = 0$ фотострум не зникає. Електрони, вибиті з катода світлом, мають деяку початкову швидкість v, а отже, і відмінну від нуля кінетичну енергію і можуть досягати анода без зовнішнього поля. Щоб фотострум став нульовим, слід прикласти ***затримуючу напругу*** U_0 (на катоді K позитивний потенціал по відношенню до анода A). За $U = U_0$ жодний з електронів не може подолати затримуючого поля і досягнути анода ($I = 0$, рис. 16.13). Отже,

$$\frac{m_e v_{max}^2}{2} = eU_0, \qquad (16.38)$$

тобто вимірявши U_0, можна знайти максимальне значення швидкості v_{max} і кінетичної енергії фотоелектронів, маса яких m_e.

Дослідами встановлено такі основні ***закони зовнішнього фотоефекту***:

1. За фіксованої частоти світла, яке падає на фотокатод, кількість фотоелектронів, що вириваються з нього за одиницю часу, пропорційна інтенсивності світла (сила фотоструму насичення $I_н$ пропорційна енергетичній освітленості E катода $-$ $I_н = \gamma E$, γ – спектральна чутливість фотокатода).

2. Максимальна початкова швидкість фотоелектронів визначається лише частотою ν світла і не залежить від його інтенсивності. Величина v_{max} зростає зі збільшенням частоти ν.

3. Для кожної речовини існує «червона межа» фотоефекту, тобто

максимальна довжина хвилі λ_{max}, за якої ще спостерігається фотоефект. Величина λ_{max} залежить від хімічної природи речовини і стану її поверхні.

Було встановлено фактичну безінерційність зовнішнього фотоефекту. Проміжок часу між початком освітлення і початком фотоструму не перевищує 10^{-9}с.

Фотоефект – результат трьох послідовних процесів:

1) поглинання фотона і поява електрона з великою (порівняно із середньою) енергією;

2) рух цього електрона до поверхні, за якого частина енергії може розсіятись;

3) вихід електрона в інше середовище через поверхню поділу.

Кількісною характеристикою фотоефекту є квантовий вихід β – кількість електронів, які вилітають під дією одного фотона, що падає на поверхню тіла. Величина β залежить від властивостей тіла, стану його поверхні та енергії фотона. Поблизу порога фотоефекту ($\lambda \approx \lambda_{max}$) для більшості матеріалів $\beta \sim 10^{-4}$ електрон/фотон. Така мала величина β обумовлена тим, що світло проникає в метал на глибину $\sim 10^{-7}$ м і там переважно поглинається. Фотоелектрони під час свого руху до поверхні сильно взаємодіють з електронами провідності, яких у металі багато, і швидко розсіюють енергію, отриману від випромінювання. Енергію, яка достатня для виконання роботи виходу, зберігають тільки ті фотоелектрони, які утворились біля поверхні на глибині, яка не перевищує 10^{-9} м. Крім того, поверхні металів сильно відбивають видиме і ближнє ультрафіолетове випромінювання. Різкого збільшення β і зсуву λ_{max} металів у видиму область спектра досягають за допомогою покриття чистої поверхні металу одноатомним шаром електропозитивних атомів або молекул (Cs, Rb, Cs_2O), які утворюють на поверхні дипольний електричний шар.

Для фотоефекту важливе значення має залежність спектральної чутливості приймача випромінювання від довжини світлової хвилі (спектральна чутливість приймача γ оптичного випромінювання – це відношення величини, яка характеризує рівень реакції приймача (наприклад, фотоструму) до потоку енергії монохроматичного випромінювання, яке викликає цю реакцію; вимірюється γ в амперах на ватт А/Вт). Із рис. 16.14 видно, що, починаючи від «червоної межі», зі зменшенням λ відбувається зростання чутливості фотокатода.

Фотоефект не можна пояснити з точки зору хвильової теорії світла.

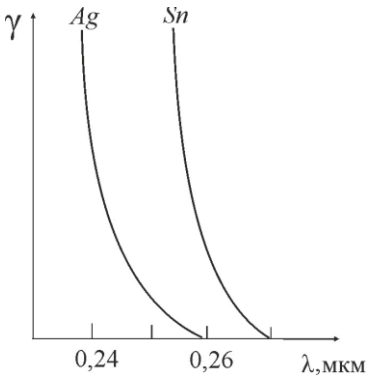


Рис. 16.14

А. Ейнштейн у 1905 р. показав, що явище фотоефекту та його закономірності можуть бути пояснені на основі запропонованої ним квантової теорії фотоефекту.

Згідно з Ейнштейном світло з частотою ν не лише випромінюється, але і поширюється у просторі та поглинається речовиною окремими порціями, енергія яких $\varepsilon = h\nu$. Поширення світла треба розглядати не як безперервний хвильовий процес, а як потік локалізованих у просторі дискретних світлових квантів, що рухаються зі швидкістю поширення світла у вакуумі. Ці кванти електромагнітного випромінювання отримали назву ***фотонів***.

За Ейнштейном, кожний фотон поглинається лише одним електроном, тому кількість вирваних фотоелектронів має бути пропорційна кількості поглинутих фотонів, тобто пропорційна інтенсивності світла (1-й закон фотоефекту). Безінерційність фотоефекту пояснюють тим, що передавання енергії при зіткненні фотона з електроном відбувається майже миттєво.

Енергія падаючого фотона витрачається на виконання електроном роботи виходу Φ з металу і на надання електрону, який вилетів, кінетичної енергії $m_e v_{max}^2/2$.

За законом збереження енергії

$$h\nu = \Phi + \frac{m_e v_{max}^2}{2}. \quad (16.39)$$

Це **рівняння Ейнштейна для зовнішнього фотоефекту**.

Перевірку формули Ейнштейна з достатньою точністю здійснив у 1916 р. Міллікен, який створив прилад, в якому досліджувані поверхні очищались у вакуумі. Після цього вимірювали їх роботу виходу Φ і досліджували залежність максимальної кінетичної енергії фотоелектронів від частоти ν світла. Результати вимірювань показано на рис. 16.15.

Відповідні графіки залежності $eU_0 = m_e \frac{v_{max}^2}{2}$ подано для трьох різних металів. Графіки мають дві цікаві характерні особливості. По-перше, залежність eU_0 від частоти ν – лінійна, і по-друге, графіки для різних металів не збігаються, хоч вони і паралельні один до

одного. Лінійна залежність eU_0 від ν підтверджує рівняння Енштейна (16.39). Тангенс кута нахилу прямої дорівнює сталій Планка h, а відрізок Φ, який відтинається прямою на координатній осі eU_0, – роботі виходу електрона з металу. (**Це можливо, якщо прямі на рис. 16.15 побудовані із врахуванням контактної різниці потенціалів, яка спостерігається у вакуумних фотоелементах**).

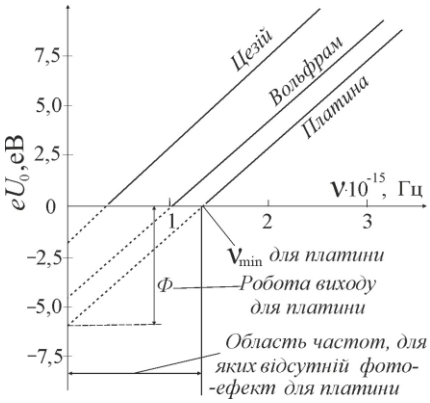


Рис. 16.15

Стає зрозуміло, чому прямі паралельні (кожна має однаковий нахил h) і чому не зливаються в одну лінію (кожна характеризується своїм певним значенням роботи виходу Φ). Коли $eU_0 = 0$, то $\nu = \nu_{min}$. Тут ν_{min} є *граничною частотою*, тобто мінімальною частотою світла, яке падає на метал, за якої ще можливе вибивання електронів з поверхні металу. Цю частоту також називають *червоною межею*. Якщо $\nu = \nu_{min}$, кінетична енергія фотоелектронів буде дорівнювати нулю. Енергії кванта вистачає тільки щоб вирвати електрон з металу. Отже,

$$h\nu_{min} = h(c/\lambda_{max}) = \Phi. \quad (16.40)$$

Величина λ_{max} залежить лише від роботи виходу електрона, тобто від хімічної природи речовини і стану її поверхні.

Якщо створити поле, яке затримуватиме рух електронів, тобто на катод подати «+», а на анод «−», то фотострум буде зменшуватись, і за напруги U_0 струм дорівнюватиме нулю (рис. 16.13). Цю напругу називають затримуючим потенціалом. Тоді формула Ейнштейна (16.39) з урахуванням (16.38) матиме вигляд

$$h\nu = \Phi + eU_0. \quad (16.41)$$

Якщо замість роботи виходу підставити значення енергії фотона, що відповідає червоній межі фотоефекту, то

$$h\nu = h\nu_{min} + eU_0,$$

або

$$eU_0 = h(\nu - \nu_{min}). \quad (16.42)$$

Прямолінійна залежність $eU_0(\nu)$ підтверджується експериментально (рис. 16.15).

Традиційно вважають, що фотоефект дає найбільш прямий експериментальний доказ квантової природи випромінювання.

16.6. Маса та імпульс фотона

Менш ніж через двадцять років після відкриття фотоефекту він виявився одним із явищ, які стали основою квантової теорії світла.

У 1905 р. А. Ейнштейн у роботі «Про одну евристичну точку зору, що стосується виникнення і перетворення світла» довів гіпотезу квантів Планка до логічного кінця і висловив ідею про те, що *світло – це потік «частинок – квантів* – фотонів» (фотонами у 1929 р. назвав кванти світла фізико-хімік Льюїс).

Теорія відносності вимагала, щоб квант випромінювання, який поглинається (або випромінюється), ніс із собою не тільки енергію, але й імпульс, тобто

$$p = \varepsilon/c = \hbar k, \qquad (16.43)$$

де $\varepsilon = \hbar\omega$ – енергія кванта; c– швидкість світла у вакуумі; $k = 2\pi/\lambda$. У загальному випадку $\vec{p} = \hbar\vec{k}$, де $\vec{k}$ – хвильовий вектор.

Звідси уже випливає, що електромагнітне поле не тільки поглинається і випромінюється порціями, але його можна розглядати як сукупність квантів (частинок), які мають енергію та імпульс. У цьому остаточно переконались тільки через два десятиліття – у 1925 р. – фізик Бозе прислав Ейнштейну статтю, в якій довів, що теплові статистичні властивості випромінювання збігаються із властивостями газу із квантів-фотонів.

Може видатися дивним, що фотон має імпульс, який відрізняється від енергії ε тільки множником c. Утім, формули для енергії та імпульсу фотона не потрібно порівнювати з аналогічними формулами для частинки в механіці Ньютона. Справа в тому, що фотон – це частинка, яка летить зі швидкістю світла, а до таких частинок треба застосовувати формули теорії відносності, в якій енергія пов'язана з імпульсом співвідношенням

$$\varepsilon^2 = c^2p^2 + m_0^2c^4, \qquad (16.44)$$

де m_0 – маса спокою частинки, тобто маса частинки за швидкості, що дорівнює нулю. Якщо врахувати формулу (16.43), то із (16.44) випливає, що маса спокою фотона $m_\lambda = m_0 = 0$. Якщо ми схильні розглядати фотони як частинки, то цей результат на перший погляд є дивним. І хоча у квантах Планка реалізувалось передбачення Ньютона, що світло – це пучок частинок-фотонів, але вони все-таки не є корпускулами Ньютона: у *фотонів не виявилось маси спокою*. Крім того вони, на відміну від класичних частинок, могли інтерферувати. Такою властивістю наділила електрони наука XX ст.– квантова механіка. Фотон у будь-якій системі відліку має швидкість c і не існує систем відліку, в якій він перебував би у стані спокою. Рівність

нулю маси спокою означає, *що фотон у стані спокою – поняття, яке не має змісту*.

Боротьба за утвердження нового погляду на природу світла була довгою і складною. Її гострота пояснювалась також і тією обставиною, що зміни, які пропонувалися, зачіпали не тільки оптику (світло), але і всю електродинаміку. Із часів Максвелла і Герца ці області не можна було розглядати вже у відриві одна від одної. Якщо погляди на світло зазнавали кардинальних змін, то на закони електродинаміки вже не можна було дивитись по-старому.

Зусиллями багатьох вчених було побудовано квантову електродинаміку, яка принципово по-іншому, ніж теорія Максвелла, пояснювала взаємодію зарядів. Утім, це не означає, що закони класичної електродинаміки були повністю відкинуті. Взаємовідношення квантових і класичних законів *визначаються* ***принципом відповідності****: класичні закони є граничним випадком квантових і, як наслідок, між ними має існувати зв'язок*. Одним із проявів цього зв'язку є твердження: *якщо закон Кулона не є точним законом «обернених квадратів* ($F \sim 1/r^2$)*», то масса спокою фотона* m_λ *має відрізнятися від нуля.* Варто підкреслити, що питання про рівність нулю m_λ – це питання експериментальне. На цю обставину вперше звернув увагу один із творців квантової механіки Л. де Бройль.

Багато важливих положень сучасної фізики ґрунтується на припущені про те, що $m_\lambda = 0$, тому проблему експериментального підтвердження цього припущення слід розглядати як принципово важливу. Дослідна перевірка закону Кулона – один із шляхів її розв'язку.

У доквантовий час відхилення від закону Кулона шукали, вважаючи, що потенціал заряду треба описувати не просто формулою $\varphi = q/(4\pi\varepsilon_0 r)$, а з експоненціальним множником:

$$\varphi = \frac{q}{4\pi\varepsilon_0 r}\, e^{r/a}, \qquad (16.45)$$

де a – деяка стала.

Якби потенціал електричного заряду мав вигляд (16.45), то всередині зарядженої сфери існувало б електричне поле. Саме таке поле всередині зарядженої сфери і намагалися знайти експериментатори. Ніякого поля знайти не вдалось. Результати дослідів виражали в оцінці параметра a. Так, із правильності закону Кулона випливало, що принаймні $a > 2 \cdot 10^7$ м.

Виникає запитання: а яке відношення має вимірювання довжини a до гіпотетичної маси фотона? Зрозуміло, що у класичній фізиці про жодну масу

фотона говорити не можна, і точність закону Кулона можна характеризувати тільки відстанню, у межах якої був перевірений цей закон. Після того як у фізику ввійшла стала Планка, маса частинки m виявилась пов'язаною з довжиною a формулою

$$a = \frac{\hbar}{mc}\cdot \qquad (16.46)$$

Не важко перевірити, що така комбінація фундаментальних величин має розмірність довжини. За теорією, якби $m = m_\lambda \neq 0$, то на відстані a від заряду потенціал електричного поля був би в $e = 2{,}72\ldots$ рази більшим, ніж дає закон Кулона.

Вимірювання поля всередині зарядженої сфери – не найбільш точний спосіб перевірки закону Кулона. Значно точніше було б перевірити закон Кулона, вимірюючи магнітне поле Землі на різних відстанях або, що виявилось ще більш чутливим, вивчаючи магнітне поле Юпітера. Ці вимірювання дали оцінку $a > 0{,}5 \cdot 10^8$ м ($m_\lambda < 7 \cdot 10^{-51}$ кг).

Утім, перевірка закону Кулона – не єдиний спосіб оцінити верхню межу m_λ. Теоретичний аналіз показує, що якщо $m_\lambda \neq 0$, то має спостерігатись дисперсія електромагнітних хвиль у вакумі, тобто швидкість їх поширення v повинна залежати від довжини хвилі λ (відсутність подібної дисперсії – один із наслідків теорії Максвелла). У цьому випадку для групової швидкості електромагнітних хвиль $v_{\text{гр}}$ можна було б записати так:

$$v_{\text{гр}} = \frac{c}{\sqrt{1+(\lambda/(2\pi a))^2}}, \qquad (16.47)$$

де c – гранична швидкість руху тіл, яка входить у перетворення Лоренца (очевидно, що за $m_\lambda \neq 0$ швидкість $v_{\text{гр}}$ не збігається із швидкістю світла c); λ – довжина електромагнітної хвилі; $2\pi a = h/(m_\lambda c)$ – комптонівська довжина хвилі фотона. Із виразу (16.47) випливає, що за $m_\lambda \neq 0$ швидкість синього світла ($\lambda \sim 0{,}4$ мкм) була би більша, ніж червоного ($\lambda \sim 0{,}7$мкм). Той факт, що насправді дисперсії електромагнітних хвиль у вакуумі за досягнутої точності вимірювань не виявлено, теж свідчить на користь твердження, що $m_\lambda = 0$. Утім, оцінка верхньої межі m_λ, яка отримана з аналізу даних про поширення електромагнітних хвиль, дає значення дещо гірше, ніж досліди з перевірки закону Кулона.

Ще більш точну оцінку маси m_λ отримують із того факту, що в галактиках спостерігаються магнітогідродинамічні хвилі в зарядженій плазмі. В земних умовах такі хвилі спостерігаються в установках, де створюються умови для термоядерних реакцій. Хвильові процеси у плазмі переносять із собою періодично змінне електромагнітне поле. Існування

маси спокою у фотона привело б до «згасання» довгохвильових коливань. У галактиках магнітогідродинамічні процеси захоплюють величезні відстані. Звідси випливає фантастична оцінка $a > > 10^{20}$ м ($m_\lambda < 3{,}5 \cdot 10^{-63}$ кг) – величина, достатня для того, щоб не сумніватись у правильності гіпотези про рівність нулю m_λ.

У 1980 р. американські фізики Дж. Примак і М. Шер висловили припущення, що у разі зниження температури може відбутись фазовий перехід, за якого фотон із безмасової частинки перетвориться на частинку з масою спокою $m_\lambda \approx 2 \cdot 10^{-38}$ кг. Група експериментаторів із Прістонського університету в 1985 р. перевірила гіпотезу про «холодний масивний» фотон за допомогою досліду. Ідея нової перевірки закону Кулона залишилась попередньою, однак виявилось, що зручніше використовувати не систему вкладених куль (метод Кавендиша), а три вкладені коаксіальні циліндри. На два зовнішніх циліндри подавалась змінна напруга. У випадку відхилення закону Кулона від точного закону «обернених квадратів» між двома внутрішніми циліндрами, які пов'язані між собою кінцевою ємністю (приблизно 20 мкФ), мав протікати змінний струм, який експериментатори намагались зареєструвати. Основну частину експериментальної установки розміщували у кріостаті, де підтримували низьку температуру. Досліди проводили за $T =$ $= 4{,}2$ К $(-268{,}8\,℃)$ та 1,36 К $(-271{,}64\,℃)$. Результати аналізу отриманих даних показали, що у разі переходу від 4,2 К до 1,36 К фазового переходу не спостерігалось і обмеження на можливу масу «холодного» фотона визначалось нерівністю:

$m_\lambda \leq (1{,}50 \ldots 1{,}38) \cdot 10^{-45}$ кг (за $T =$ $=1{,}36$ К).

Таким чином, досліди, виконані в земних умовах, а також оцінки, зроблені шляхом аналізу астрофізичних даних, не дають підстав вважати масу спокою фотона відмінною від нуля, і отже, закон взаємодії точкових електричних зарядів зберігає статус точного закону «обернених квадратів».

16.7. Тиск світла

Фотон, подібно до частинки, яка рухається, має енергію $\varepsilon = \hbar\omega$ та імпульс $p = \hbar\omega/c$. Усі ці корпускулярні характеристики пов'язані із хвильовою характеристикою світла – його частотою ω. Одним із підтверджень наявності у фотонів імпульсу є світловий тиск. Тиск світла, який передбачала електромагнітна теорія Максвелла, вперше експериментально виявив і виміряв у 1901 р.

П. М. Лебедєв.

Прилад Лебедєва – це досить чутливі крутильні терези, рухома система яких складається з легкого каркаса із закріпленими на ньому тонкими кружками (крильцями), які розміщені симетрично відносно осі підвісу (рис. 16.16).

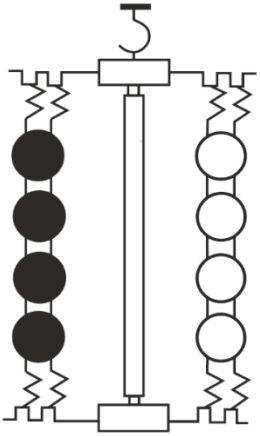


Рис. 16.16

Деякі крильця були дзеркальними, а поверхня інших була зачорнена. Усю цю систему підвішували на тонкій пружній нитці всередині закритого скляного балона, в якому був створений високий вакуум. Крильця освітлювали вольтовою дугою. Світло від дуги направляли на них за допомогою системи лінз і дзеркал. Величину світлового тиску на крильця визначали за кутом закручування нитки підвісу.

Тиск світла настільки малий, що для його надійного вимірювання треба було усунути вплив на крильця всіх інших факторів. Вплив конвекційних струмів повітря Лебедєв усунув, створивши в балоні досить високий вакуум, однак і в такому разі не було усунуто радіометричного ефекту. Причина його в тому, що зачорнене крильце нагрівається внаслідок поглинання світла, яке падає на нього, причому температури освітленої і неосвітленої (задньої) поверхонь крильця неоднакові. Молекули розрідженого повітря в балоні відбиваються від нагрітої поверхні крилець з більшою швидкістю, передаючи їм, відповідно, більший імпульс. Тиск, зумовлений такою дією, значно більший за світловий. Радіометричний ефект може призвести до того, що в досліді тиск на зачорнене крильце буде більший за тиск на дзеркальне крильце тих самих розмірів. Лебедєв усунув вплив радіометричного ефекту, використавши у своїх дослідах дуже тонкі крильця різної товщини від 0,01 до 0,1 мм.

Тиск світла на дзеркальне крильце з коефіцієнтом відбиття $R = 1$ виявився у два рази більший, ніж тиск на зачорнене крильце ($R = 0$).

З точки зору корпускулярних уявлень тиск світла інтерпретується як

результат передачі імпульсів фотонами поверхні, яка поглинає або відбиває їх. За нормального падіння кожний фотон, який поглинається, передає поверхні імпульс $\hbar\omega/c$, а який відбивається $\frac{\hbar\omega}{c} - \left(-\frac{\hbar\omega}{c}\right) = 2\hbar\omega/c$. На одиничну площину поверхні за одиницю часу падає енергія випромінювання $U \cdot c$, де U – об'ємна густина енергії (формула (16.2)). Отже, кількість фотонів, які падають, буде $N = Uc/(\hbar\omega)$. Із цієї кількості NR фотонів відіб'ється, а $N(1-R)$ – погли-неться (R – енергетичний коефіцієнт відбивання поверхні). Традиційно тиск світла дорівнює повному імпульсу, який передається фотонами, які падають на 1 м^2 поверхні за 1 с, тобто

$$p = NR \cdot \frac{2\hbar\omega}{c} + \frac{N(1-R)\hbar\omega}{c} = \frac{(1+R)N\hbar\omega}{c}.$$

Врахувавши, що $U = N\hbar\omega/c$, отримаємо:

$$p = U(1+R), \qquad (16.48)$$

що збігається із виразом, отриманим на основі електромагнітної теорії.

Часто для визначення тиску світла на поверхню з коефіцієнтом відбиття R використовують формулу

$$p = \frac{I}{c}(1+R)\cos^2\varphi, \qquad (16.49)$$

де I – інтенсивність світла, Вт/м^2; φ – кут падіння світла на плоску поверхню (кут між нормаллю до поверхні та променем світла, який падає на цю поверхню).

Ізотропне випромінювання, яке поширюється у всіх напрямках, чинить на поверхню стінки тиск

$$p = U/3\,. \qquad (16.50)$$

Цей вираз можна отримати, розглядаючи випромінювання у замкненій оболонці як ідеальний газ фотонів.

Світловий тиск досить малий. Так, за межами атмосфери Землі інтенсивність сонячного випромінювання дорівнює 1400 Вт/м^2. Відповідний тиск за нормального падіння світла на поверхню, для якої $R = 0{,}5$, дорівнює $7\cdot 10^{-6}$ Па. Ця величина у 10^{10} менша від атмосферного тиску. Попри те, що тиск світла малий, його треба враховувати у деяких випадках. Так, сила гравітаційної взаємодії частинок пропорційна кубу їх радіуса, а сила світлого тиску пропорційна квадрату радіуса частинки. Для частинок досить малих розмірів ці сили можуть бути однакових порядків. Це дає підставу зробити припущення, що хвости комет зумовлені світловим тиском.

Всередині зірок температура досягає мільйонів градусів. Таким температурам відповідає велика густина енергії випромінювання, і тиск світла перешкоджає гравітаційному стисканню зірок. Можливо, цим і пояснюється наявність верхньої межі маси зірок, яка має порядок 10^{30} кг. Крім цього,

випромінювання і поглинання його всередині зірки зумовлює швидке перенесення деякої маси з одних областей в інші. Випромінювання, що виходить із центральних областей, має менший момент імпульсу, ніж ті маси периферійних областей, що поглинають випромінювання. Це зумовлює сповільнене обертання зірки.

Нерівномірність освітлення поверхонь штучних супутників Землі викликає небажане їх обертання навколо деякої осі.

Фокусування лазерного пучка у «пляму», радіус якої дорівнює довжині хвилі, дає змогу отримати порівняно великий тиск. За їх допомогою мікроскопічним частинкам можна надати прискорення, яке в мільйони разів більше за прискорення вільного падіння, і це має практичне застосування.

16.8. Ефект Комптона

Відомі явища, в яких імпульс фотона проявляє себе явно, і співвідношення $\vec{p} = \hbar\vec{k}$ допускає експериментальну перевірку. Особливо чітко проявляються корпускулярні властивості світла в явищі розсіяння рентгенівських променів (про рентгенівські промені детальніше йдеться в підрозд. 17.12). А. Комптон, досліджуючи у 1923 р. розсіяння монохроматичних рентгенівських променів речовинами з легкими атомами (парафін, бор), виявив, що у складі розсіяного випромінювання поряд з випромінюванням початкової довжини хвилі λ спостерігається також випромінювання більшої довжини хвилі λ'.

Схему досліду Комптона зображено на рис. 16.17. Пучок монохроматичних рентгенівських променів джерела *R*, пройшовши систему діафрагм *Д*, падає на речовину *P*, яка розсіює їх, і після розсіяння на кут θ потрапляє в рентгенівський спектрограф C, де вимірюється довжина хвилі розсіяного випромінювання. Досліди Комптона показали, що довжина хвилі λ' розсіяного випромінювання більша за довжину хвилі λ випромінювання, яке падало, причому різниця $\Delta\lambda = \lambda' - \lambda$ не залежить від довжини хвилі випромінювання, що падає, і природи речовини, яка розсіює, а визначається лише величиною кута розсіяння θ.

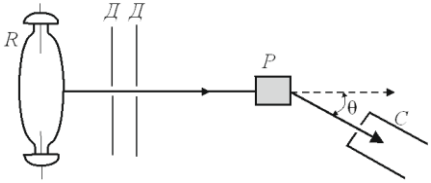


Рис. 16.17

Пояснити ефект Комптона можна на основі квантових уявлень про природу світла. Якщо вважати, що

випромінювання є потоком фотонів, то ефект Комптона – це результат пружного зіткнення рентгенівських фотонів з вільними електронами. В атомах легких елементів енергія зв'язку електрона (енергія іонізації) становить 10 еВ, що приблизно в тисячу разів менше енергії рентгенівського фотона ($\hbar\omega \approx 10$ кеВ), тому електрони в цих дослідах фактично можна вважати вільними. Енергія спокою електрона $m_e c^2 \approx 0{,}5$ МеВ, що набагато більше енергії фотона, тому електрон, який до зіткнення перебував у стані спокою або рухався в атомі з нерелятивістською швидкістю, після зіткнення з таким фотоном залишається нерелятивістським. Нехтуючи початковою енергією електрона, запишемо закон збереження енергії як

$$\hbar\omega = \hbar\omega' + \frac{p^2}{2m_e}, \qquad (16.51)$$

де $\omega = \frac{2\pi c}{\lambda}$; $\omega' = 2\pi c/\lambda'$; p – імпульс електрона після зіткнення з фотоном, а закон збереження імпульсу

$$\vec{p}_\Phi = \vec{p}'_\Phi + \vec{p}, \qquad (16.52)$$

де $\vec{p}_\Phi$ і $\vec{p}'_\Phi$ – імпульс фотона до і після розсіювання (рис. 16.18). Згідно з теоремою косинусів вираз (16.52) переписуємо так:

$$p^2 = p_\Phi^2 + (p'_\Phi)^2 - 2p_\Phi p'_\Phi \cos\theta. \quad (16.53)$$

Підставимо у (16.53) вирази для імпульсу фотона через його частоту, $p_\Phi = \hbar\omega/c$ і $p'_\Phi = \hbar\omega'/c$, і квадрат імпульсу електрона із закону збереження енергії (16.51). Тоді (16.53) набуде вигляду

$$2m_e\hbar(\omega - \omega') = \frac{\hbar^2}{c^2}\{\omega^2 + ((\omega')^2 - \\ -2\omega\omega'\cos\theta\}.$$

Перепишемо це співвідношення у вигляді

$$\omega - \omega' = \left\{\frac{\hbar\omega}{2m_e c^2}\right\} \times \\ \times \{(\omega^2 + (\omega')^2)/\omega - 2\omega'\cos\theta\} \quad (16.54)$$

і враховуючи, що $\hbar\omega/(m_e c^2) \ll 1$, бачимо, що зміна частоти $\Delta\omega = \omega' - \omega$ мала порівняно із самою частотою, тому у правій частині (16.54) можна замінити ω' на ω. Тоді для відносного зсуву частоти $\Delta\omega/\omega$ під час розсіяння отримаємо

$$\frac{\Delta\omega}{\omega} = -\frac{\hbar\omega}{m_e c^2}(1 - \cos\theta) = -\frac{2\hbar\omega}{m_e c^2}\sin^2\frac{\theta}{2}.$$

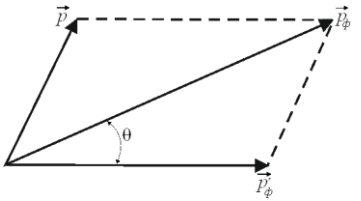


Рис. 16.18

Знак «мінус» показує, що частота випромінювання під час розсіяння зменшується. Це зрозуміло, оскільки частину своєї енергії фотон передає електрону. Переходячи тепер від частот до довжин хвилі $\lambda = \frac{2\pi c}{\omega}$, $\frac{\Delta\lambda}{\lambda} = -\frac{\Delta\omega}{\omega}$, отримаємо

$$\Delta\lambda = \frac{2h}{m_e c}\sin^2\frac{\theta}{2} = 2\Lambda\sin^2\frac{\theta}{2}. \quad (16.55)$$

Співвідношення (16.55) дуже добре описує залежність збільшення довжини хвилі від кута розсіяння, яка спостерігається на досліді. Константу λ, яка входить у (16.55), називають *комптонівською довжиною хвилі* електрона, що виражається у вигляді комбінації трьох універсальних сталих: $\lambda = h/(m_e c)$ (порівняйте формулу для λ і формулу (16.46) для a). Підставляючи числові значення сталих, отримаємо $\lambda = 0{,}24$ нм, що збігається з величиною, яка знайдена з вимірювань зміщення довжини хвилі під час розсіяння.

Під час розсіяння фотона частина його енергії передається електрону. Електрони віддачі відіграють важливу роль у процесі іонізації газів рентгенівським випромінюванням. Використовуючи камеру Вільсона, яка розміщена в магнітному полі, можна знайти імпульс та енергію електронів віддачі, що дозволяє ще більш повно перевірити виконання законів збереження. Вимірювання показали, що в кожному елементарному акті розсіяння електрон набуває якраз таких енергії та імпульсу, які втрачає фотон.

Досліди Комптона яскраво продемонстрували, що енергія та імпульс фотона справді пов'язані формулою (16.43) і що закони збереження енергії та імпульсу виконуються в елементарних актах розсіяння.

Контрольні запитання

1. Чим відрізняється теплове випромінювання від інших видів випромінювання?

2. Які є характеристики теплового випромінювання?

3. Які є характеристики випромінюваної поверхні тіла?

4. Яке тіло вважають абсолютно чорним?

5. Опишіть закон Кірхгофа для теплового випромінювання.

6. Опишіть закони випромінювання абсолютно чорного тіла.

7. Яке практичне застосування формули Релея–Джінса?

8. Як пов'язані формула Планка і квантування енергії?

9. Опишіть безконтактні способи вимірювання температури нагрітих тіл. Які є типи температур?

10. Які закони зовнішнього фотоефекту? Що означає рівняння Ейнштейна для фотоефекту?

11. У чому суть корпускулярної природи світла?

12. Чому є принциповою відсутність маси спокою у фотона?

13. Опишіть тиск світла на поверхню. Яка природа такого тиску?

14. У чому суть ефекту Комптона? Що цей дослід довів?

Розділ 17. Будова атома

17.1. Дослід Резерфорда. Планетарна модель атома

Починаючи з часів Демокріта була поширена думка, що атом – нескінченно мале абсолютно пружне тіло, яке існує вічно у незмінному вигляді. До початку XX ст. мало у кого, навіть у такого геніального фізика, як Дж. Максвелл, таке твердження породжувало сумніви.

Наприкінці XIX ст. особливо бурхливий розвиток фізики почав гальмуватися проблемами, пов'язаними із внутрішньою структурою атомів та їхніми властивостями. На той час було встановлено, що:

– атоми є подільними складними електронейтральними об'єктами, в яких одними із структурних компонентів є електрони;

– найістотніша за величиною взаємодія має електричний характер;

– атом одного елемента може перетворюватися в атом іншого.

Залишалося нез'ясованим питання про внутрішню будову атома, адже за теоремою Ірншоу ніяка конфігурація позитивних і негативних зарядів, пов'язаних між собою кулонівською взаємодією, не є стійкою. Навіть якщо створити таку стійку систему, то нескінченно мале зміщення будь-якого із зарядів незворотно цю стійкість знищує. Запропонована Кельвіном у 1902 р. модель, за якою позитивний заряд рівномірно розподілений по всьому об'єму атома, а електрони рівномірно розподілені на тлі позитивного заряду, забезпечувала стійкість атома. Така модель нагадує кекс з родзинками – позитивна основа кекса із вкрапленими в нього електронами.

Англійський фізик Томсон доповнив модель Кельвіна припущенням, що електрони в такому «кексі» рухаються по замкнутих орбітах. Остаточно модель набула назви *модель атома Томсона*. Упродовж 10 років така модель дала змогу описати складний характер явищ дисперсії та поглинання світла, породила вчення про поляризацію світла, передбачила існування ізотопів тощо.

Англійський фізик Резерфорд піддав сумніву правильність моделі Томсона. Він виконав досліди, в яких аналізували зміни в траєкторіях потоку α-частинок, що проходили крізь надтонкі металічні пластинки. Відхилення в траєкторіях фіксували спалахами на флуоресцентному екрані в точках потрапляння відхиленої α-частинки.

Маса α-частинки у 7300 разів більша від маси електрона, тому зіткнення α-частинки з електроном неістотно змінить її траєкторію. Утім, зіткнення α-частинки з масивним атомом може відчутно змінити і траєкторію, і її енергію.

α-частинки утворюються під час природного радіоактивного розпаду деяких важких елементів і є позитивно зарядженими частинками із зарядом $2e$ та масою, яка приблизно в чотири рази більша за масу атома водню.

Спрощену схему досліду Резерфорда зображено на рис. 17.1. Джерело α-частинок розміщене всередині свинцевої порожнини з вузьким каналом. Усі α-частинки, крім тих, що рухаються всередині вузького каналу, поглинаються свинцем. Потік α-частинок, пройшовши крізь вузьку діафрагму *Д*, потрапляє на тонку золоту фольгу *Ф* завтовшки 10^{-6} ... 10^{-7} м, що складається з декількох атомних шарів. Під час проходження крізь фольгу α-частинки відхиляються на різні кути θ і потрапляють на екран *Е*, який можна поміщати в різних положеннях щодо фольги. Екран покритий флюоресцентною речовиною. За допомогою мікроскопа *М* можна спостерігати місця потрапляння α-частинок за свіченням (спалахами) екрана. Поведінку α-частинок після проходження крізь фольгу вивчали і в камері Вільсона.

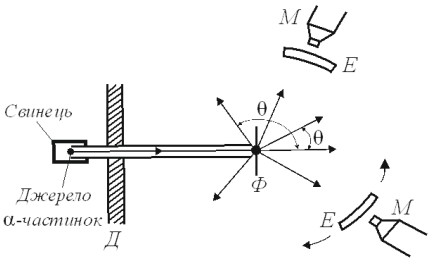


Рис. 17.1. Схема досліду Резерфорда

У результаті численних спостережень таких спалахів було встановлено, що переважна більшість α--частинок відхилялась на кути 1 … 3° відносно їх траєкторії перед пластинкою, проте траплялися відхилені α-частинки на кути до 150°. Відхилені α-частинки на кути, більші ніж 90°, у середньому спостерігалися лише одна на 8000. Попри їхню мізерну кількість вони є важливими із принципового погляду – їх появу ніяк не можна обґрунтувати в межах моделі Томсона.

З метою пояснення несподіваних результатів експерименту Резерфорд висунув нове припущення про будову атома, за яким додатний заряд атома, на противагу моделі Томсона, зосереджений у доволі малому за розмірами об'ємі – ядрі. Електрони обертаються навколо ядра, утримуючись

на орбітах унаслідок зрівноваження відцентрових сил силою кулонівського притягання до ядра. Така картина нагадує планетарну систему, подібну до Сонячної, тому така модель будови атома отримала назву планетарної моделі.

Для обґрунтування планетарної моделі будови атома Резерфорд проаналізував процес розсіяння заряджених частинок на атомах речовини з погляду класичної механіки. Коли частинка пролітає біля ядра, на неї діє кулонівська сила відштовхування:

$$F = \frac{2Ze^2}{4\pi\varepsilon_0 r^2},$$

де $2e$ – заряд α-частинки; Ze – заряд ядра речовини; r – відстань між α-частинкою та ядром. У цьому випадку траєкторією α-частинки є гіпербола. Схему процесу розсіяння зображено на рис. 17.2. Тут p – прицільна відстань, що, фактично, була б відстанню найбільшого наближення α-частинки до атомного ядра якби вона пролітала без розсіяння.

Припускається, що α-частинки, влітаючи в область, зайняту атомом (на рис. 17.2 зображене ядро атома, заряд якого $+Ze$), не розсіюються на електронах атома, а лише на ядрі. Можна навести вагомі аргументи на підтвердження такого припущення. Маса α-частинки, як зазначалося, набагато більша від маси електрона, тому енергетичні втрати α-частинки у разі однократного пружного зіткнення з електроном будуть дуже малими і це не вплине на її траєкторію. Є також інші аргументи, які тут не подаємо, бо це вимагає знання квантових понять, які вивчаються пізніше.

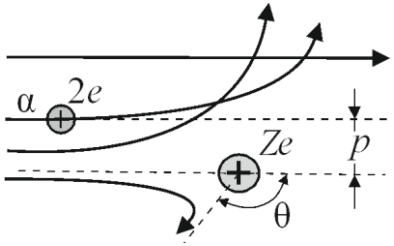


Рис. 17.2. Схема процесу розсіяння α-частинки на атомах речовини

Використовуючи закони збереження енергії та імпульсу в замкнутій системі «α-частинка – ядро» та беручи до уваги, що маса ядер значно більша за масу α-частинки, Резерфорд знайшов таке співвідношення між параметрами задачі розсіяння:

$$\operatorname{ctg}\frac{\theta}{2} = W_{\text{к}}\frac{2\pi\varepsilon_0 p}{Ze^2}, \qquad (17.1)$$

де $W_{\text{к}} = \frac{mv^2}{2}$ – кінетична енергія α-частинки (v, m – її швидкість та маса відповідно), а Z – зарядове число. Експериментально перевірити справедливість формули (17.1)

неможливо через неможливість виміряти прицільний параметр p (рис. 17.2).

Закони динаміки дозволяють розрахувати залежність кута відхилення θ від параметра удару p, а методами теорії ймовірностей можна знайти ймовірність пролітання α-частинки на цій відстані p від частинки і тим самим ймовірність її відхилення на кут θ. Розрахунок показує, що із загальної кількості N падаючих частинок у пучку кількість розсіяних dN_θ α-частинок в тілесному куті $d\Omega$, що міститься між двома конічними поверхнями, твірні яких утворюють кути θ та $\theta + d\theta$ з початковим напрямком руху α-частинок, знаходять за формулою

$$dN_\theta = nN\left(\frac{Ze^2}{8\pi\varepsilon_0 W_к}\right)^2 \cdot \frac{d\Omega}{\sin^4\left(\frac{\theta}{2}\right)}, \quad (17.2)$$

де n – кількість ядер з розрахунку на одиницю площі поверхні мішені. Це і є *формулою Резерфорда* для розсіяння α-частинок. Зауважимо, що формула (17.2) справедлива лише для малих значень прицільних відстаней p, тобто для частинок, розсіяних тільки на великі кути. Якщо прицільна відстань велика, то частинка пролітає поза ділянкою, зайнятою атомом. Атом, загалом, тепер потрібно розглядати як електронейтральний. Характер взаємодії між частинкою, що розсіюється, та атомом змінюється – взаємодія стає відмінною від точкової кулонівської.

Для певної речовини (наприклад, золото), яка розсіює α-частинки за певних енергії та заданої густини потоку, добуток $\frac{dN_\theta}{d\Omega}\sin^4\left(\frac{\theta}{2}\right)$ має бути сталим, що було експериментально підтверджено в дослідах Гейгера і Марсдена. Крім висновку про справедливість планетарної моделі атома, з досліду Резерфорда очевидним був ще один важливий висновок – кількість електронів в атомі збігається з його порядковим атомним числом Z у періодичній системі елементів. Оскільки в нормальному стані атом електронейтральний, то заряд ядра у такому атомі дорівнює $+Z|e|$. До того ж, досліди Резерфорда опосередковано свідчили про те, що закон Кулона для взаємодії заряджених частинок залишається справедливим навіть для дуже малих відстаней між ними.

На підставі результатів дослідів з розсіяння α-частинок тонкими фольгами Резерфорд у 1910 р. запропонував *ядерну модель атома*.

Згідно з цією моделлю в центрі атома міститься ядро, в якому зосереджені позитивний заряд $+Ze$ і фактично вся маса атома. Лінійні розміри ядра $\sim(10^{-14} \ldots 10^{-15})$ м.

Навколо ядра в області з лінійними розмірами $\sim 10^{-10}$ м по замкнених орбітах рухаються Z електронів, утворюючи електронну оболонку (*електронний рій*) атома.

Ядерна модель Резерфорда зовні дуже нагадує Сонячну систему: у центрі – ядро, навколо нього по орбітах рухаються електрони, тому цю модель називають планетарною. Орбіти електронів в атомі стаціонарні, атому властива виняткова стійкість. Стійкість атома не можна погодити із класичним поясненням ядерної моделі.

Розглянемо, наприклад, ядерну модель найпростішого атома – атома водню. Для простоти припустимо, що електрон, маса якого m_e і заряд $-e$, рухається рівномірно по коловій орбіті навколо протона (із зарядом $+e$), який займає центральне положення (рис. 17.3). У першому наближенні рухом протона, маса якого майже у 1836 разів більша за масу електрона, можна знехтувати.

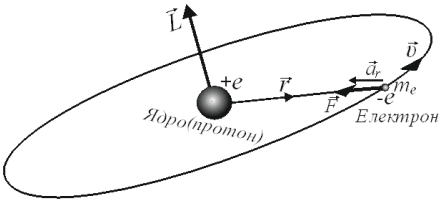


Рис. 17.3. Планетарна ядерна модель атома водню

Сила, яка утримує електрон на орбіті, є силою кулонівського притягання між протоном і електроном. Ця центральна сила $F = \frac{e^2}{4\pi\varepsilon_0 r^2}$, де r – радіус колової орбіти електрона. За другим законом Ньютона запишемо рівняння руху електрона:

$$\frac{m_e v^2}{r} = \frac{e^2}{4\pi\varepsilon_0 r^2}, \qquad (17.3)$$

де $a_r = v^2/r$ – доцентрове прискорення (рис. 17.3). На основі рівняння (17.3) кінетична енергія електрона

$$W_{\text{к}} = \frac{m_e v^2}{2} = \frac{e^2}{8\pi\varepsilon_0 r}. \qquad (17.4)$$

Потенціальна енергія системи «електрон – протон»

$$W_{\text{п}} = -\frac{e^2}{4\pi\varepsilon_0 r}. \qquad (17.5)$$

Знак «мінус» означає, що у цій системі діють сили притягання, а не відштовхування. Повна енергія системи

$$W = W_{\text{к}} + W_{\text{п}} = -\frac{e^2}{8\pi\varepsilon_0 r}. \qquad (17.6)$$

Знак «мінус» показує, що ця система «електрон – протон» є зв'язаною.

Енергію зв'язку електрона визначають як мінімальну енергію, яка потрібна для повного видалення електрона з атома, або, інакше кажучи, для іонізації атома. Експериментальним шляхом було знайдено, що для атома водню енергія зв'язку дорівнює 13,53 еВ.

Підставляючи це значення енергії в (17.6) для W, можна знайти радіус атома r: $r = 0{,}53 \cdot 10^{-10}$ м = 0,53 Å.

Лінійна швидкість v пов'язана з частотою ω обертання електрона на орбіті співвідношенням $v = \omega r$.

Тоді із (17.3) можемо знайти, що

$$\omega = \frac{e}{2r}\sqrt{\frac{1}{\pi\varepsilon_0 m_e r}} = 4{,}125 \cdot 10^{16} \text{ рад/с}. \quad (17.7)$$

Це значення теж збігається із значенням ω, знайденим за допомогою інших методів.

Попри певні успіхи в поясненні будови атома, які досягнуті в межах планетарної моделі та класичної теорії, перед фізиками постав ряд непереборних протиріч. Так, згідно із законами класичної електродинаміки:

а) заряджена частинка, яка рухається із прискоренням, має неперервно випромінювати електромагнітну енергію (див. підрозд. 11.6, формула (11.41));

б) частота цього випромінювання має дорівнювати частоті обертання частинки навколо ядра. Отже, відповідно до цієї моделі повна енергія атома повинна зменшуватись (стаючи все більше від'ємною), тоді як частота обертання (формула (17.7)) має неперервно зростати. Розрахунок за формулою (11.43) показує, що за цих умов за час $\simeq 10^{-8}$ с електрон впаде на ядро й атом перестане існувати (рис. 17.4). Крім цього, якщо така модель правильна, то оптичний спектр випромінювання водню (як і спектри інших елементів) мають бути неперервними, а це означає, що всі атоми за дуже короткий час будуть зруйновані. Обидва висновки перебувають у повному протиріччі з експериментальними даними: атом є дуже стійкою системою, а оптичні спектри газів мають зовсім не неперервний розподіл за частотою, а характеризуються набором дискретних частот, тобто мають лінійчасту структуру. Всі ці факти привели до того, що планетарна модель у класичному розумінні швидко була визнана такою, що не відбулась.

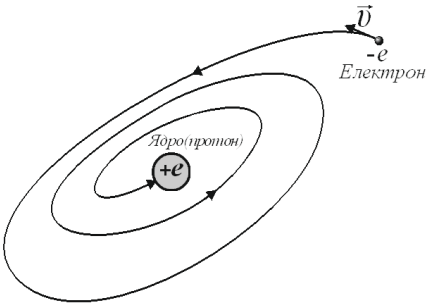


Рис. 17.4. Рух електрона в атомі за класичною фізикою

17.2. Атомні спектри

Електричний розряд у трубці, яка містить одноатомний газ під низьким тиском, є джерелом світла, аналіз якого за допомогою дифракційної ґратки або спектрографа із призмою дозволяють виявити серію дуже чітких ліній випромінювання. Ці лінії, які характеризують газ, який використовують у трубці, називають лінійчастим спектром цього газу. Якщо в розрядній трубці використовують азот, то отримають смугастий спектр, який містить дискретні групи тісно розміщених ліній. Цей смугастий спектр характеризує двохатомні молекули N_2 й має інше походження порівняно з лінійчастими спектрами. Біле світло від джерела випромінювання з ниткою розжарення характеризується суцільним спектром і містить неперервний набір довжин хвиль. Якщо таке світло пропускати крізь одноатомний газ, наприклад водень, то виникає спектр поглинання. При цьому на спектрограмі отримують спектр з темними лініями на світлому фоні. Положення цих ліній відповідає довжинам хвиль яскравих спектральних ліній випромінювання водню, тобто газ поглинає падаюче випромінювання саме з такою довжиною хвиль.

Дослідження атомних спектрів започаткували ще 1860 р., коли було опубліковано працю Г. Кірхгофа і Р. Бунзена «Хімічний аналіз за допомогою спостереження спектру». У наступні роки такі дослідження здійснювали доволі інтенсивно. Було встановлено характерну ознаку атомних спектрів – їхню дискретність. Спектри складаються з набору вузьких смуг (ліній) певної частоти. Так, наприклад, у видиму ділянку спектра найпростішого хімічного елемента атомарного водню потрапляє лише чотири лінії, які позначають $H_\alpha,\ H_\beta,\ H_\gamma,\ H_\delta$.

Упродовж двох десятків років інтенсивних досліджень атомних спектрів було накопичено величезну кількість експериментального матеріалу, складено докладні таблиці. Утім, у великому масиві даних не вдавалося встановити якусь закономірність, хоча всі дослідники розуміли, що в цих таблицях заховано повну інформацію про будову атома.

Перші вагомі успіхи у класифікації атомних спектрів, зокрема водню, були отримані 1885 р. Як це не дивно, але певне значення тут мали ідеї античної філософської піфагорійської школи. Послідовники цієї школи стверджували, що гармонія Всесвіту

завдячує містичній ролі цілих чисел у природі. Під вливом ідей піфагорійців перебував із ранніх років і Й. Бальмер (1825–1898) – учитель фізики жіночої гімназії м. Базеля (Швейцарія). Впродовж багатьох років він шукав закономірності у сукупності атомних спектрів, вважаючи, що цілі числа мають тут мати не останнє значення.

Докладно аналізуючи спектральні лінії $\{H_\alpha, H_\beta, H_\gamma, H_\delta\}$ атома водню у видимій ділянці, Й. Бальмер зауважив, що вони розміщені не випадково, а утворюють серію довжин хвиль, яку можна описати єдиною формулою:

$$\lambda = \lambda_0 \frac{k^2}{k^2-4}, \qquad (17.8)$$

де $\lambda_0 = 3645{,}6$ Å – емпірично визначена стала, а $k = 3, 4, 5, \ldots$ – цілі числа.

Формула (17.8) настільки точно описувала довжину хвиль ліній у видимій ділянці спектра атома водню, що на неї відразу звернули увагу. Стало зрозуміло, що за цією простою емпіричною формулою криється глибокий фізичний зміст. *По суті з формули Бальмера починається справжня наука про атоми та їхню внутрішню електронну будову.*

Незабаром було звернуто увагу на те, що формула (17.8) стає більш прозорішою і змістовнішою, якщо її переписати не для довжини хвиль, а через частоти, використовуючи співвідношення між частотою і довжиною електромагнітної хвилі $\omega = 2\pi\nu = \frac{2\pi c}{\lambda}$, тоді

$$\omega = R_\omega \left(\frac{1}{2^2} - \frac{1}{k^2}\right) \qquad (17.9)$$

або

$$\frac{1}{\lambda} = R_\lambda \left(\frac{1}{2^2} - \frac{1}{k^2}\right), \qquad (17.10)$$

де $R_\omega = \frac{8\pi c}{\lambda_0} = 2{,}07 \cdot 10^{16}$ с$^{-1}$, а $R_\lambda = \frac{4}{\lambda_0} = 1{,}097 \cdot 10^7$ м$^{-1}$. Величини R_ω і R_λ називають *сталими Рідберга.* За формулою (17.10) можна розрахувати довжину хвиль спектральних ліній в серії Бальмера. Для $k = 3$ $\lambda = 6{,}563 \cdot 10^{-7}$ м (червона лінія H_α), для $k = 4$ $\lambda = 4{,}861 \cdot 10^{-7}$ м (блакитна лінія H_β), для $k = 5$ $\lambda = 4{,}339 \cdot 10^{-7}$ м (блакитна лінія H_γ), для $k = 6$ $\lambda = 4{,}101 \cdot 10^{-7}$ м (фіолетова лінія H_δ). За $k = \infty$ ми отримаємо межу серії, наприклад, для серії Бальмера $\lambda = 4 / R_\lambda = 3{,}645 \cdot 10^{-7}$ м. Спектри атомів водню та ртуті зображено на рис. 17.4′.

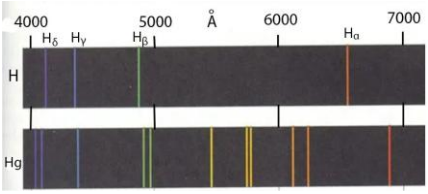


Рис. 17.4′. *Спектри атомів водню та ртуті*

Пізніше подібна аналітична залежність спостерігалась в інших, ультрафіолетовій та інфрачервоній, ділянках спектра атомарного водню. Усіх їх можна об'єднати в *узагальнену формулу Бальмера*:

$$\omega_{nk} = R_\omega \left(\frac{1}{n^2} - \frac{1}{k^2}\right), \qquad (17.11)$$

$$\frac{1}{\lambda} = R_\lambda \left(\frac{1}{n^2} - \frac{1}{k^2}\right), \qquad (17.12)$$

де $n = 1, 2, 3, 4, \ldots$ визначає серію, а $k = n+1, n+2, n+3, \ldots$ визначає окремі лінії серії (табл. 17.1).

Якщо ввести у формулі (17.11) позначення $T_n = \frac{R_\omega}{n^2}$, яке називають термом, то частоту будь-якої лінії спектра водню можна подати як різницю між двома термами, а саме

$$\omega_{nk} = T(n) - T(k). \qquad (17.13)$$

Таблиця 17.1. Спектральні серії водню

Серія	Область спектру	Формула серії	Границя серії ($k=\infty$), λ·10,м
Лаймана	Ультрафіолетова	$\frac{1}{\lambda}=R_\lambda(\frac{1}{1^2}-\frac{1}{k^2})$, $k=2,3,4\ldots$	0,91127
Бальмера	Видима	$\frac{1}{\lambda}=R_\lambda(\frac{1}{2^2}-\frac{1}{k^2})$, $k=3,4,5\ldots$	3,6456
Пашена	Інфрачервона	$\frac{1}{\lambda}=R_\lambda(\frac{1}{3^2}-\frac{1}{k^2})$, $k=4,5,6\ldots$	8,2014
Брекетта	Інфрачервона	$\frac{1}{\lambda}=R_\lambda(\frac{1}{4^2}-\frac{1}{k^2})$, $k=5,6,7\ldots$	14,580
Пфунда	Інфрачервона	$\frac{1}{\lambda}=R_\lambda(\frac{1}{5^2}-\frac{1}{k^2})$, $k=6,7,8\ldots$	22,782
Хемфрі	Інфрачервона	$\frac{1}{\lambda}=R_\lambda(\frac{1}{6^2}-\frac{1}{k^2})$, $k=7,8,9\ldots$	32,806

На перший погляд може здатися, що введення поняття терма є тривіальним, не більше ніж перепозначення. Насправді це не так. Подання частоти як різниці двох термів мало вагомі наслідки для подальшого розвитку теорії атомних спектрів. Важливо, що емпірична формула (17.11) записана для атома водню, проте залишалося нез'ясованим питання про принципову можливість написання подібної простої формули для класифікації спектральних ліній складніших атомів. Враховуючи запис частоти спектральної лінії атома водню у формі (17.13), Й. Рідберг надав спектральним лініям різних елементів форму комбінації термів, застосувавши для терма емпіричну формулу

$$T(n) = \frac{R}{\left(n + \alpha + \frac{\beta}{n^2}\right)^2},$$

де α, β – дійсні числа, які мають значення емпірично підібраних параметрів. На підставі цієї формули вдалося здійснити класифікацію спектральних ліній багатьох елементів таблиці.

Продовжуючи дослідження Рідберга, у 1905 р. Ритц сформулював комбінаційний принцип, за яким будь-яку спектральну лінію будь-якого атома можна подати як різницю термів.

Неважливо, чи відомі аналітичні вирази для термів. Набір спектральних ліній атомів, виглядаючи на перший погляд хаотичною сукупністю, розпався на цілком визначені серії з чіткими закономірностями їх розміщення. Замість великої кількості спектральних ліній достатньо знати набагато меншу кількість термів. Правда, комбінуючи терми, можна передбачувати спектральні лінії, які насправді не спостерігалися. Згодом з'ясувалося, що у спектрі немає окремих спектральних ліній через «правила відбору», які чітко були встановлені та обґрунтовані квантовою теорією.

У результаті виконаної класифікації спектрів стало зрозуміло, що спектральні закономірності мають фундаментальний характер і визначаються внутрішньою будовою атома. Утім, встановити зв'язок структури спектру з будовою атома та фізичними процесами на атомному рівні не вдавалося впродовж багатьох років. Лише 1913 р. у цій проблемі було досягнуто ясності.

17.3. Постулати Бора

Теорія атома водню

Як вже зазначалося, планетарна модель атома не узгоджувалася із законами класичної електродинаміки. Рух електрона по замкнутих орбітах навколо ядра є прискореним. За законами класичної електродинаміки пришвидшений рух заряду завжди супроводжується випромінюванням ЕМХ. Електромагнітні хвилі переносять енергію, тому електрон, втрачаючи енергію на випромінювання, має впасти на ядро. Ніхто не міг спростувати цей важливий наслідок законів електродинаміки, проте існування стабільних атомів також є незаперечним фактом. Данський фізик Нільс Бор перший дійшов висновку, що протиріччя моделі Резерфорда із класичною електродинамікою можна усунути лише тоді, коли відмовитися від класичного (ньютонівського) опису руху електрона по замкнутих орбітах у планетарній моделі атома.

На початку лютого 1913 р. суто випадково Н. Бор зустрів студентського приятеля Ханса Хансена, який розказав йому про спектральні формули Бальмера, Рідберга та Рітца. Н. Бор згадував: «Як тільки я побачив формулу Бальмера, мені все одразу ж стало зрозуміло». Взявши гіпотезу Планка про кванти енергії, формулу (17.11) помножимо на сталу планка $\hbar$. Тоді її перепишемо так:

$$\hbar\omega_{nk} = \frac{\hbar R_\omega}{n^2} - \frac{\hbar R_\omega}{k^2},$$

або

$$\hbar\omega_{nk} = |W_n| - |W_k|, \quad (17.14)$$

де $\hbar\omega_{nk}$ – квант енергії випромінювання дорівнює різниці двох значень енергії атома водню (W_k – енергія до випромінювання, W_n – енергія після випромінювання).

Для усунення протиріччя планетарної моделі атома із законами класичної електродинаміки Нільс Бор припустив, що, рухаючись по орбіті, електрон не випромінює електромагнітних хвиль, а залишається на орбіті досить тривалий час. Упродовж 1913 р. Бор опублікував три праці, в яких він сформулював свої знамениті постулати.

Із цього приводу американський фізик (один із творців теорії надпровідності) Леон Купер сказав: «Звичайно, було дещо самовпевнено висувати припущення, які протирічили електродинаміці Максвелла і механіці Ньютона, але Бор був молодим».

ПОСТУЛАТ 1
(Про стаціонарні стани)

Атоми можуть перебувати тільки у визначених станах, в яких, попри рухи електронів по замкнутих дискретних орбітах, вони не випромінюють енергію. У цих станах атоми мають енергію, яка утворює дискретний ряд: W_1, W_2,..., W_n,... . Ці стани характеризуються стійкістю: будь-яка зміна енергії W_n завдяки випромінюванню чи поглинанню електромагнітного випромінювання або інших форм обміну атома енергією може відбуватися у разі повного переходу з одного стану в інший.

За цим постулатом електрони під час руху навколо ядра майже весь час перебувають на стабільних, незалежних від часу орбітах, які мають цілком визначену енергію. Такі орбіти називають *стаціонарними*.

Далі для пояснення спектральних закономірностей, зумовлених процесами поглинання і випромінювання електромагнітних хвиль, пропонується такий постулат.

ПОСТУЛАТ 2
(Правило частот)

У разі переходу з одного стаціонарного стану в інший атоми випромінюють чи поглинають електромагнітні хвили визначеної частоти. Випромінювання, що виходить або поглинається у разі переходу зі стаціонарного стану W_k у стаціонарний стан W_n, є монохроматичним і його частоту визначають співвідношенням

$$\omega_{nk} = \frac{|W_n| - |W_k|}{\hbar}. \qquad (17.15)$$

Теорія Бора, побудована на сформульованих постулатах, блискуче пояснила структуру спектральних ліній атома водню. Власне, Н. Бор і брав за основу ці експериментальні дані під час формулювання своїх постулатів, тому відразу постало питання про безпосередні, незалежні підтвердження постулатів Бора, про наявність стаціонарних станів і правил переходу між ними.

Дослід Франка і Герца

Першим експериментом, в якому безпосередньо було підтверджено справедливість постулатів Бора, став дослід Дж. Франка і Г. Герца (1914 р.). Установка, яку вони використовували, схематично показано на рис. 17.5, *а*. У трубці з гарячим катодом *K*, живлення якого здійснюється від батареї ε_3, із сіткою *C* та анодом *A*, містяться пари ртуті за низького тиску за температури близько 150 °С.

Між катодом *K* і сіткою *C* прикладена прискорювальна різниця потенціалів U_a, яку можна змінювати від 0 до 60 В. Між анодом і сіткою створюється невелике гальмівне поле з різницею потенціалів U_r (≈ 0,5 В). Дуже чутливий гальванометр Г, включений послідовно в коло анода, вимірює анодний струм (≈ 10^{-9} А).

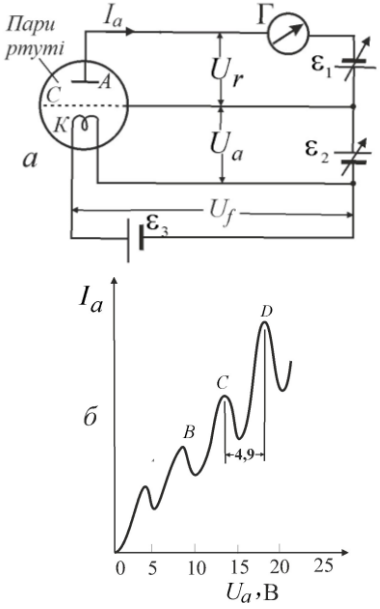


Рис. 17.5. Схема установки і результати дослідів Франка і Герца

Характерну криву залежності анодного струму I_a від прискорювального потенціала U_a зображено на рис. 17.5, *б*. Виявилось, що зі збільшенням U_a, як і в будь-якій електронній лампі, анодний струм зростає, однак цей ріст не є монотонним, а супроводжується різкими спадами струму кожний раз, коли U_a збільшується приблизно на 5 В.

Немонотонность кривої є результатом взаємодії електрона з атомами ртуті. Маса атома ртуті майже в

400 000 разів перевищує масу електрона. Поки енергія бомбардуючого електрона не перевищує 4,9 еВ, його зіткнення з атомами ртуті є пружними (електрон розсіюється на атомах без зміни енергії, змінюючи лише свою траєкторію). Такий електрон не здатний збудити атом ртуті. Ситуація докорінно змінюється після досягнення електроном енергії 4,9 еВ. Тепер електрон може збудити атом ртуті, передавши йому частину своєї кінетичної енергії. Таке зіткнення електрона з атомом ртуті є непружним. У результаті таких зіткнень енергія електронів наскільки зменшується, що електрони не можуть подолати слабке затримуюче поле, яке створюється різницею потенціалів U_r, і не досягають анода. Анодний струм при цьому різко зменшується. Якщо прискорювальний потенціал U_a збільшити ще на 5 В, то електрони можуть зазнати ще одного непружного зіткнення і при цьому повністю втратити свою енергію. Це пояснює другий спад (спостерігається на кривій струму за $U_a = 10$ В), який буде більш різким, ніж перший спад, оскільки відповідає електронам, які зазнали двох непружних зіткнень. Третій спад струму I_a відповідає електронам, які зазнають трьох непружних зіткнень, і т.д. Кожний раз, коли відбувається непружне зіткнення, атом ртуті переходить у збуджений стан і потім, повертаючись в основний стан, випромінює фотон з довжиною хвилі $\lambda = 0{,}2536$ мкм, що відповідає переходам атомів ртуті з першого збудженого стану в основний стан. Енергія такого фотона

$$\varepsilon = hc / \lambda = 4{,}89 \text{ еВ}.$$

Цей результат, а також той факт, що відстань між двома сусідніми спадами струму I_a дорівнює приблизно 4,9 В, переконливо доводить існування дискретних енергетичних рівнів в атомі ртуті. Удосконалена методика експерименту дозволила визначити також енергії збудження для інших енергетичних рівнів. І не дивно, що за ці чудові досліди Дж. Франк і Г. Герц в 1925 р. отримали Нобелівську премію з фізики.

Покажемо, слідом за Бором, до які наслідки мають сформульовані постулати для атома водню. Рівняння (17.3) допускає безмежну кількість замкнутих колових орбіт. Постає питання про існування стаціонарних орбіт серед безмежної кількості можливих. Перший постулат Н. Бора стверджує, що стаціонарні орбіти існують. Тоді відразу виникає таке запитання: «*А як стаціонарні орбіти вирізнити з-поміж безмежної кількості*

всіх можливих, що задовольняють співвідношення (17.3)?». Ось тут Н. Бор зрозумів, а можливо, і вгадав, як відповісти на це запитання. І ця відповідь була простою, тому і геніальною.

Уже від 1905 р. в науці утвердилися і завоювали визнання квантові ідеї, про що йшлося у попередньому розділі. Нагадаємо, що стала Планка h має розмірність дії і від початку названа Планком квантом дії. Але ж поняття дії у класичній механіці відоме ще з часів Гамільтона і Лагранжа. Розмірність дії має орбітальний момент імпульсу $\vec{L} = \vec{r} \times m\vec{v}$ (векторна форма запису).

Величину $\vec{L}$ ще називають моментом кількості руху. Він перпендикулярний як радіусу-вектору $\vec{r}$, так і вектору кількості руху $m\vec{v}$ (див. рис. 17.3). Величиною орбітального моменту імпульсу для колового руху електрона навколо силового центру (ядра) є $L = m_e rv$ (див. рис. 17.3).

Бор відчував, що є глибинний зв'язок між двома «мінімальностями» в природі:

– існуванням найменшої величини для розмірів «електронного рою» в атомі. Меншою ця величина бути не може;

– існуванням найменшої величини для фізичної дії. Меншою вона не буває.

Весною 1912 р. Бор дійшов висновку, що будова «електронного рою» в резерфордівському атомі управляється квантом дії – сталою Планка $\hbar$.

Отже, коли фізична величина має розмірність дії, то найменше числове значення, якого вона може набувати, дорівнює сталій Планка, а більші значення кратні сталій Планка. Після тривалих міркувань і розрахунків Н. Бор дійшов висновку, що стаціонарні можуть бути лише ті колові орбіти, для яких орбітальний момент імпульсу кратний $\hbar$:

$$m_e rv = n\hbar. \qquad (17.16)$$

Ціле число n називають квантовим числом.

Рівняння (17.13) і (17.16) утворюють систему двох рівнянь із двома невідомими r і v. Розв'язок цієї системи дає

$$v_n = \frac{e^2}{4\pi\varepsilon_0 \hbar n}, \qquad (17.17)$$

а

$$r_n = \frac{4\pi\varepsilon_0 \hbar^2}{m_e e^2} n^2, \qquad (17.18)$$

де $n = 1, 2, 3, \ldots$.

У формулі (17.18), яка визначає радіус борівської орбіти, є число n, що

означає залежність радіуса від цього квантового числа.

Радіус r_1 найменшої колової орбіти для атома водню позначають як a_0:

$$a_0 = \frac{4\pi\varepsilon_0\hbar^2}{m_e e^2} = 0{,}53 \cdot 10^{-10}\text{м} = 0{,}53\ \text{Å}, \quad (17.19)$$

його називають *радіусом Бора*. У певному розумінні радіус Бора a_0 визначає розміри атома водню. Задовго до створення теорії Бора розміри атома водню були оцінені експериментально. Теоретичне значення радіуса Бора (17.19) виявилося доволі близьким до експериментального значення радіуса атома водню. Це було ще одним аргументом на користь теорії Бора.

Енергію електрона у кулонівському полі ядра, заряд якого $|e|$, отримаємо, підставивши значення радіусів орбіт (17.18) у формулу (17.6). Тоді

$$W_n = -\frac{e^4 m_e}{32\pi^2\varepsilon_0^2\hbar^2}\cdot\frac{1}{n^2} = -\frac{13{,}6}{n^2}\ \text{еВ}. \quad (17.20)$$

Стан атома з найменшим значенням енергії (n = 1) називають основним. Для атома водню основному стану згідно з (17.20) відповідає енергія $W_1 = -13{,}6$ еВ. Згідно з другим постулатом (17.15) для атома водню частота переходу із стану W_k у стан W_n становитиме

$$\omega_{kn} = \frac{1}{\hbar}(|W_n| - |W_k|) = \frac{e^4 m_e}{32\pi^2\varepsilon_0^2\hbar^3}\left(\frac{1}{n^2} - \frac{1}{k^2}\right) \quad (17.21)$$

Порівняння виразу (17.21) з узагальненою формулою Бальмера (17.11) засвідчує, що вони повністю збігаються, якщо сталу Рідберга R_ω вибрати такою

$$R_\omega = \frac{e^4 m_e}{32\pi^2\varepsilon_0^2\hbar^3} = 2{,}07 \cdot 10^{16}\text{с}^{-1}. \quad (17.22)$$

Отримане значення R_ω збігається з експериментальним значенням. Це був успіх теорії Бора.

Зобразимо енергетичні рівні (рис. 17.6), що описують формулою (17.20), у вигляді ліній, на яких цифри визначають квантові стани *n*. За початок відліку взято стан з $n \to \infty$, який позначимо $W = 0$. Електрон з енергією $W > 0$ здійснює необмежений (інфінітний) рух у полі атомного ядра за гіперболічними траєкторіями. За такого руху електрон може мати довільне додатне значення енергії. Коли $W < 0$, то електрон здійснює обмежений (фінітний) рух по замкнутих орбітах, а його енергії можуть мати тільки дискретні значення (17.20).

Стан n = 1 відповідає основному (нормальному) станові електрона, а всі стани з $n \geq 2$ відповідають збудженим його станам. Вертикальні стрілки, спрямовані догори, відповідають переходам між відповідними станами з

поглинанням атомом енергії, а у зворотному напрямку (рис. 17.6) – з випромінюванням енергії за другим постулатом Бора.

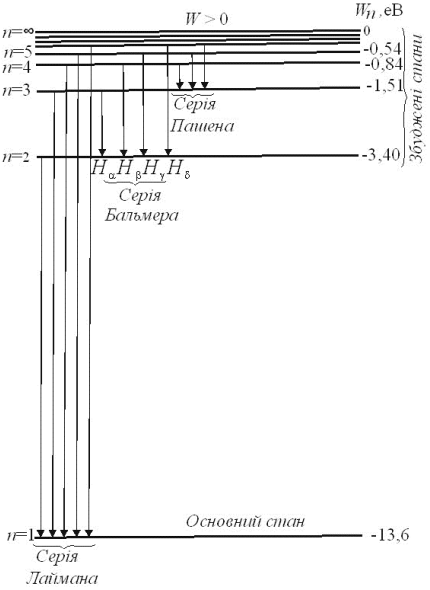


Рис. 17.6. Енергетичні рівні атома водню

Із формули (17.20) і рис. 17.6 видно, що якщо атом перебуває в основному стані ($n = 1$), то для видалення електрона з атома потрібна енергія 13,6 еВ. Отже, енергія зв'язку $W_{зв}$ та енергія іонізації $W_{іон}$ для атома водню, який перебуває в основному стані, $W_{зв} = W_{іон} = 13{,}6$ еВ. За допомогою схеми на рис. 17.6 легко пояснити виникнення серіальних спектральних ліній в табл. 17.1.

Воднеподібні атоми

Теорію Бора можна з успіхом використати і під час розгляду воднеподібних атомів, тобто атомів, які складаються з ядра із зарядом Ze та одного електрона. До таких атомів належать один раз іонізований атом гелію He^+ ($Z = 2$), два рази іонізований атом літію Li^{2+} ($Z = 3$) і т. д.

Для воднеподібних атомів радіуси дозволених орбіт електрона визначають за формулою

$$r_n = \frac{a_0 n^2}{Z}, \qquad (17.23)$$

яка подібна до формули (17.18). Енергію електрона на орбіті визначають за формулою

$$W_n = -\frac{13{,}6Z^2}{n^2}\,\text{еВ}, \qquad (17.24)$$

яку отримують із формули (17.20).

Таким чином, теорія Бора пояснила зміст відомих на той час серіальних формул і передбачила ще серії в інфрачервоній області спектра, які згодом були виявлені експериментально. Стала Рідберга, яку отримали з теорії Бора, дивовижно добре узгоджується з експериментальним значенням цієї величини. Це говорить про те, що теорія Бора правильно описує будову атома водню.

Теорія Бора була важливим кроком у розвитку теорії атома. Вона

показала непридатність класичної фізики для пояснення внутрішньоатомних явищ і головну роль квантових законів у мікросвіті. Разом з тим теорія Бора опиралась на класичну механіку і не була ані послідовно класичною, ані послідовно квантовою і могла бути тільки перехідним станом на шляху створення послідовної теорії атомних явищ. Н. Бор у своїй теорії об'єднав резерфордівську модель атома, емпірично встановлені закономірності в атомних спектрах і квантові уявлення про випромінювання.

Робота Бора – одне з гідних подиву явищ в історії науки. Тільки геніальним просвітленням можна пояснити появу цієї теорії до того, як з'ясувались хвильові властивості частинок. Саме з цього приводу Ейнштейн сказав: «*...вища музикальність у сфері теоретичної думки*».

17.4. Подальший розвиток теорії Бора. Квантування Бора–Вільсона–Зоммерфельда

Постулати Бора, констатуючи наявність в атомі водню стаціонарних орбіт з визначеними значеннями енергії електронів і правил переходу електронів між стаціонарними орбітами, не до кінця пояснюють фізичну природу постулатів, не встановлюють фундаментального зв'язку між стаціонарними станами електронів та їх динамікою по орбітах в кулонівському полі ядра. До того ж, Н. Бор для побудови теорії спектрів атома водню застосував правило квантування (17.16). Відразу виникла потреба у фізичному осмисленні цього правила та його узагальненні. Цю працю проробив, переважно, Вільсон (1915 р.) і незалежно від нього Зоммерфельд (1916 р.). Продемонструємо це на прикладі гармонічного осцилятора, максимально дотримуючись міркувань і висновків праці Зоммерфельда.

Розглянемо матеріальну точку маси m, що здійснює лінійний рух у потенціальному полі $U(x) = kx^2/2$. Це може бути тягарець, маса якого m, що коливається за допомогою пружини жорсткістю k. Повна енергія такого осцилятора

$$W = W_{\text{к}} + W_{\text{п}} = \\ = \frac{mv^2}{2} + \frac{kx^2}{2} = \frac{p^2}{2m} + \frac{m\omega^2 x^2}{2},$$

де $p = mv$ – імпульс точки; x – зміщення від положення рівноваги; $\omega = \sqrt{k/m}$ – власна частота коливання осцилятора, яка не залежить від амплітуди коливання. На рис. 17.7, *а* зображено траєкторію осцилятора у фазовому просторі $\{x;p\}$ для заданого постійного

значення повної енергії W осцилятора. Такою траєкторією є еліпс з півосями $a = \sqrt{2W/(m\omega^2)}$ та $b = \sqrt{2mW}$. Класичний опис руху осцилятора на підставі теорії звичайних диференціальних рівнянь другого порядку вказує, що, якою б не була початкова умова $\{x_0; p_0\}$, через цю точку фазового простору обов'язково проходить одна і лише одна траєкторія.

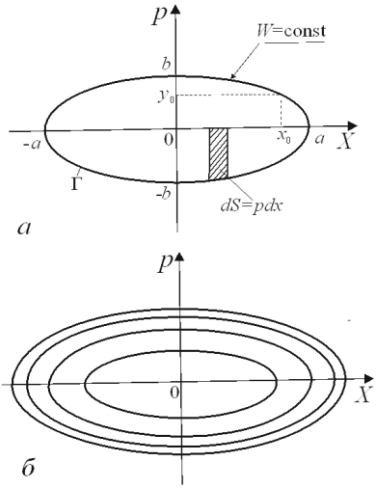


Рис. 17.7. Траєкторія гармонічного осцилятора у фазовому просторі для заданого значення енергії W (а); сім'я траєкторій гармонічного осцилятора для різних значень його енергії (б). Площі, обмежені будь-якими сусідніми траєкторіями, однакові за величиною

Через початкові умови однозначно визначають також енергію осцилятора. Отже, кожному значенню енергії осцилятора відповідає своя траєкторія, яку описують канонічним рівнянням еліпса:

$$\frac{x^2}{a^2} + \frac{p^2}{b^2} = 1. \qquad (17.25)$$

Площа еліпса $S = \pi ab$. Підставивши вирази для довжин півосей a та b, знайдемо

$$S = \frac{2\pi W}{\omega} = \frac{W}{\nu}, \qquad (17.26)$$

де $\nu = \omega/2\pi$ – лінійна частота коливання осцилятора.

Тепер до формулювання правила квантування залишається виконати одну, проте доволі сміливу дію. З формули (17.26) бачимо, що площа еліпса у фазовому просторі має розмірність дії $[S] =$ Дж·с. Але ж розмірність дії має також стала Планка – квант дії. Позначимо через $\triangle W$ різницю між постійними енергіями осцилятора, фазовими траєкторіями яких є два сусідніх еліпси. Тоді відповідно до формули (17.26)

$$\triangle S = \frac{\triangle W}{\nu} = h, \text{ або } \triangle W = h\nu. \qquad (17.27)$$

На рис. 17.7, *б* схематично зображено сімейство траєкторій, що задовольняють цю умову. Перенумеруємо еліпси із вказаного сімейства цілими додатними числами 1, 2, 3, …, n, а відповідні значення енергії осцилятора W_1, W_2, W_3, ..., W_n. Тоді енергія, що відповідає n-му еліпсу,

$W_n = \varepsilon \cdot n$, $n = 1, 2, 3, \ldots$. (17.28)

Квант енергії $\varepsilon = h\nu = \hbar\omega$ – абсолютне значення різниці енергій для двох будь-яких сусідніх фазових траєкторій. Ґрунтуючись на прикладі гармонічного осцилятора та ще декількох простих моделей, було зроблено узагальнення правила квантування (17.27)–(17.28) на довільні механічні системи з одним ступенем вільності, що здійснюють рух по замкненій траєкторії у фазовому просторі. Порівнявши (17.26) і (17.28) знайдемо, що площа еліпса, яка відповідає енергії W_n, буде $S_n = nh$ з однієї сторони, а з другої сторони (рис. 17.7, *а*)

$$S_n = \oint_\Gamma dS = \oint_\Gamma p dx.$$

Порівняння цих двох формул дає правило квантування (редакція Зоммерфельда)

$\oint_\Gamma p dx = nh$, $n = 1, 2, 3, \ldots$, (17.29)

де n – квантове число, за допомогою якого перенумеровують дискретні рівні енергії. Інтегрують у формулі (17.29) по замкненому контуру Γ (рис. 17.7, *а*).

Після створення сучасної квантової теорії правила квантування Бора–Вільсона–Зоммерфельда, або «старої квантової теорії», втратили своє принципове значення для квантової фізики. Утім, інколи ця теорія набуває практичного застосування під час знаходження дискретних рівнів енергії квантових частинок, які здійснюють періодичні рухи у потенціальних силових полях. У загальному випадку правила квантування (17.29) дають задовільні результати тільки тоді, коли рух квантової частинки близький до класичного. Здебільшого це випадки руху частинок у потенціальному полі з великим значенням енергії. Точніше, коли різниця енергій між двома сусідніми дискретними рівнями мала порівняно з енергією цих рівнів, тобто $\frac{|W_n - W_{n+1}|}{W_n} << 1$.

Покажемо, що правило квантування (17.16) очевидне з умови квантування (17.29). Якщо електрон рухається по коловій орбіті, радіус якої r, як це припускається в теорії Бора, то для обчислення інтеграла у (17.29) зручно перейти до полярної системи координат $\{r,\varphi\}$ (рис. 17.8) за формулами $x = r\cos\varphi$, $y = r\sin\varphi$.

Під час руху електрона по колу, радіус якого r (r = const), його просторове положення відносно силового центра (ядра) однозначно визначається тільки однією координатою – полярним кутом φ ($0 \le \varphi \le 2\pi$).

Швидкість $v^2 = \dot{x}^2 + \dot{y}^2 = r^2\dot{\varphi}^2$, $W_к = m_e v^2/2 = m_e r^2\dot{\varphi}^2/2$, а потенціальна енергія $W_п = U(r) = \text{const}$, оскільки $r = \text{const}$. Імпульс електрона $p = m_e v = \frac{dW_к}{dv}$, а імпульс, що відповідає координаті φ, $p_\varphi = \frac{dW_к}{d\dot{\varphi}} = m_e r^2\dot{\varphi}$.

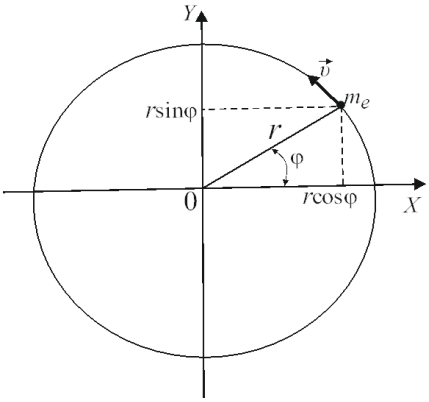


Рис. 17.8. Рух електрона по коловій орбіті

Тоді (17.29) перепишемо так:

$$\oint_\Gamma p_\varphi d\varphi = \oint_\Gamma m_e r^2\dot{\varphi} d\varphi = \oint_\Gamma m_e r v d\varphi = m_e r v \oint_\Gamma d\varphi = 2\pi m_e r v = n_\varphi h.$$

Звідси отримаємо

$$m_e v r = n_\varphi \hbar, \qquad (17.30)$$

що відповідає правилу квантування Бора для колових орбіт ($n_\varphi = 1, 2, 3, \ldots$).

Розвиваючи теорію Н. Бора і ґрунтуючись на дослідженнях В. Вільсона, А. Зоммерфельд припустив, що орбіта електрона в атомі водню є еліптичною. В одному з фокусів еліптичної орбіти перебуває ядро атома. Це припущення ґрунтувалося на поширенні планетарної моделі Сонячної системи на атоми. Це цілком логічно і виправдано, бо планети Сонячної системи та електрони атома рухаються у потенціальних силових полях, які мають схожу математичну структуру. Під час руху електрона по еліпсу поряд з кутом φ відстань r від ядра теж є змінною величиною (рис. 17.9).

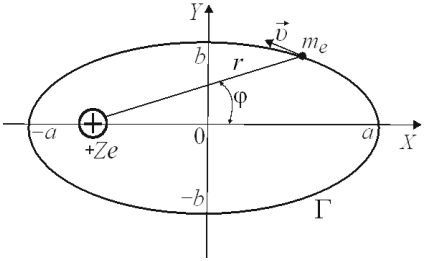


Рис. 17.9. *Рух електрона по еліптичній орбіті*

Правила квантування Зоммерфельда для еліптичних орбіт записують так:

$$\begin{cases} \oint_\Gamma p_\varphi d\varphi = n_\varphi h, \\ \oint_\Gamma p_r dr = n_r h, \end{cases} \qquad (17.31)$$

де n_φ – азимутальне, а n_r – радіальне квантове число.

Перша квантова умова, як і для колових орбіт, також приводить до квантування орбітального моменту імпульсу (17.16). Обчислення інтеграла у другій умові є технічно значно складнішою задачею. Докладніше про це

можна дізнатись з додаткової літератури. Наведемо тут лише основні фізичні висновки. Розв'язання А. Зоммерфельдом задачі (17.31) для еліптичних орбіт вказує на те, що квантові числа n_r та n_φ задовольняють умову

$$n = n_r + n_\varphi. \qquad (17.32)$$

Тут азимутальне n_φ та радіальне n_r квантові числа можуть набувати таких значень: n_φ =1, 2, 3, …, n; $n_r = 0$, 1, …, (n–2), (n–1).

Повна енергія електрона в полі атомного ядра набуває дискретних значень, які визначають формулою (17.24). Велика піввісь еліпса

$$a = a_0(n_r + n_\varphi)^2, \qquad (17.33)$$

а мала –

$$b = a_0(n_r + n_\varphi)n_\varphi, \quad (17.34)$$

де a_0 – радіус першої боровської орбіти.

Для основного стану n = 1 азимутальне квантове число може набувати значення n_φ= 1, а n_r = 0. Орбіта електрона в цьому стані є коловою $a = b = a_0$. Для першого збудженого стану n = 2 допустимими є такі набори квантових чисел: {n_φ = 2, $n_r = 0$}, що відповідає $a = b = 4\,a_0$ і {n_φ= 1, n_r = 1}, що відповідає $a = 4a_0$, $b = 2a_0$ (тобто рух по еліпсу). Для n = 3 можливими є три орбіти, які визначають таким набором квантових чисел: {n_φ = 3, n_r = 0}; {$n_\varphi = 2$, n_r = 1}; {$n_\varphi = 1$, $n_r = 2$}. Ці орбіти зображено на рис. 17.10. У теорії Зоммерфельда з (17.33) і (17.34) випливає, що відношення півосей еліптичних орбіт задають виразом

$$\frac{b}{a} = \frac{n_\varphi}{n_r + n_\varphi}. \qquad (17.35)$$

Із (17.35) випливає, що чим меншим є азимутальне квантове число n_φ, тим більше витягнутий еліпс (рис. 17.10). Слід зазначити, що під час руху різними еліпсами з одним і тим самим квантовим числом n_φ електрон має однакову енергію, тому говорять, що відбувається виродження за азимутальним квантовим числом n_φ. Якщо є, наприклад, зовнішнє електричне поле, еліптичні орбіти з однією енергією, але різної форми, будуть збуджені порізному, і це впливає на енергетичний спектр.

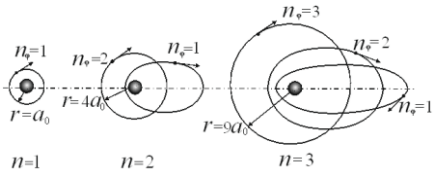


Рис. 17.10. Стаціонарні орбіти основного (n = 1) та двох збуджених рівнів (n = 2; 3) атома водню

17.5. Просторове квантування

Ми розглянули рух електрона у площині еліптичної орбіти, тобто за двох ступеней вільності (два квантові числа:

азимутальне n_φ і радіальне n_r). Положення електрона у просторі характеризується трьома координатами, тому треба враховувати просторову орієнтацію самої орбіти. В атомі всі напрямки рівноймовірні, однак фізично можна створити напрямки, які відрізняються від решти. Такими напрямками можуть бути напрямки зовнішнього магнітного та електричного полів. Припустимо, що атом з орбітальним електроном перебуває в магнітному полі. Атом має магнітний момент і в зовнішньому магнітному полі повинен поводити себе як магніт, тобто має зайняти положення, за якого магнітний момент буде або паралельний або антипаралельний полю (див. підрозд. 5.8). Утім, атом має також і момент кількості руху $\vec{L}$, який перешкоджає орієнтації магнітного момента. Звідси випливає, що в магнітному полі атом має поводити себе так, як звичайне веретено (дзиґа) поводить себе в полі тяжіння, тобто рух електрона по еліпсу супроводжується одночасним обертанням еліпса навколо силової лінії поля (відбувається прецесія еліпса). У результаті еліптична орбіта виходить не замкненою, а подібною на розетку (рис. 17.11).

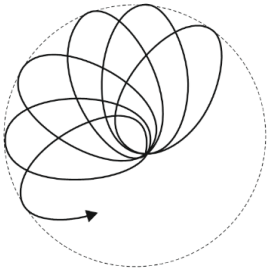


Рис. 17.11. Прецесія еліптчної орбіти електрона (розетка)

Положення площини орбіти електрона щодо виділеного напрямку визначається кутом α між напрямком моменту кількості руху $\vec{L}$ і виділеним напрямком (напрямок індукції магнітного поля $\vec{B}$). Цей тип руху в атомі теж квантований. На рис. 17.12 зображено положення орбіти електрона та орієнтація $\vec{B}$ у просторі координат *X*, *Y*, *Z*. Згідно з рис. 17.12 $L_z = L\cos\alpha$. Враховуючи умову квантування Бора (17.30), отримаємо, що проекція $\vec{L}$ на напрямок $\vec{B}$ (вісь *Z*), буде

$$L_z = \hbar n_\varphi \cos\alpha.$$

Позначимо

$$n_\varphi \cos\alpha = m, \qquad (17.36)$$

й отримаємо

$$L_z = m\hbar, \qquad (17.37)$$

де m отримало назву магнітного квантового числа. Оскільки $-1 \leq \leq \cos\alpha \leq +1$, то можливі значення $m = = n_\varphi \cos\alpha$ перебувають у межах

$$-n_\varphi \le m \le n_\varphi. \qquad (17.38)$$

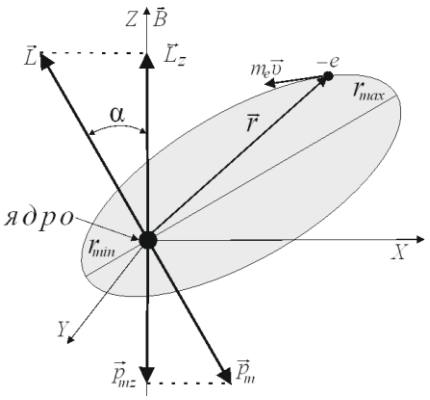


Рис. 17.12. Просторове квантування

Магнітне квантове число m може мати n_φ цілих додатних, від'ємних значень і нульове значення, тобто всього $(2n_\varphi+1)$ значень:

$$m = 0, \pm1, \pm2, \dots, \pm n_\varphi. \qquad (17.39)$$

Таким чином, під час дії на атом зовнішнього магнітного поля можуть зреалізовуватись тільки такі орієнтації орбіт, за яких проєкція на вісь *Z,* що збігається з напрямком магнітного поля $\vec{B}$ моменту кількості руху $\vec{L}$, кратна $\hbar$.

У цьому і полягає суть *просторового квантування*. Квантування проєкції моменту кількості руху L_z на напрямок магнітного поля означає також, що квантується і проєкція орбітального магнітного моменту p_{mz} (рис. 17.12). Зв'язок між *орбітальним магнітним моментом* $\vec{p}_m$ і моментом кількості руху $\vec{L}$, як було встановлено раніше (див. підрозд. 6.2), задають гіромагнітним відношенням

$$p_m = -\frac{e}{2m_e}L, \qquad (17.40)$$

де m_e – маса електрона. Очевидно, що і для проєкції p_{mz} і L_z згідно рис. 17.12 можна записати:

$$p_{mz} = -\frac{e}{2m_e}L_z. \qquad (17.41)$$

Враховуючи умову квантування (17.37), отримаємо:

$$p_{mz} = -\frac{e\hbar}{2m_e}m = -\mu_Б m. \qquad (17.42)$$

Величина $\mu_Б = \frac{e\hbar}{2m_e} = 0{,}93 \cdot 10^{-23}$Дж/Тл є елементарним магнітним моментом, її називають ***магнетоном Бора.***

Таким чином, проєкція магнітного моменту атома на напрямок магнітного поля кратна елементарному магнітному моменту – магнетону Бора.

Підтвердженням теорії просторового квантування є ефект Зеємана й досліди Штерна і Герлаха.

17.6. Ефект Зеємана

Ефект Зеємана був відкритий у 1896 р. Суть його в тому, що в разі дії на атом зовнішнього магнітного поля його енергетичний стан змінюється і стає залежним від магнітного квантового числа m, що проявляться у розщепленні спектральних ліній на декілька

компонентів. Причину такого розщеплення легко пояснити наявністю в електрона, який обертається навколо ядра, магнітного орбітального моменту (17.42).

Магнітний момент у магнітному полі має енергію $W = -\vec{p}_m\vec{B} = -p_{mz}B$, де p_{mz} – проєкція магнітного моменту на напрямок поля (див. підрозд. 5.8). Це означає, що в магнітному полі атом з урахуванням (17.42) отримає додаткову енергію, яка залежить від квантового числа m:

$$\Delta W = \mu_Б B m, \qquad (17.43)$$

де, як відомо, $m = 0, \pm1, \pm2, \ldots, \pm n_\varphi$. Якщо $n_\varphi = 1$, то m набуває значення -1, 0, +1 (тобто знімається виродження по магнітному квантовому числу m, і такий енергетичний рівень розщеплюється на три рівні). Якщо $n_\varphi = 2$, то $m = -2, -1$, 0, +1, +2.

На рис. 17.13 показано розщеплення рівнів і спектральних ліній випромінювання. За відсутності магнітного поля спостерігається одна лінія з частотою ω_0. Якщо включене магнітне поле, крім лінії ω_0 виникають лінії $\omega_0 + \Delta\omega$ та $\omega_0 - \Delta\omega$. Поява цих ліній (або, як говорять, зміщення компонентів) називають нормальним (або Лоренцовим) зміщенням або нормальним ефектом Зеємана, класичне пояснення якого дав Лоренц. Величину нормального зміщення можна оцінити із (17.43), вважаючи, що $\hbar\Delta\omega = \mu_Б B m$. Врахувавши, що $\lambda = \frac{2\pi c}{\omega}$ й $|\Delta\lambda| = \frac{2\pi c}{\omega^2}\Delta\omega$, отримаємо:

$$|\Delta\lambda| = \frac{\lambda^2}{c}\frac{\mu_Б B m}{h}.$$

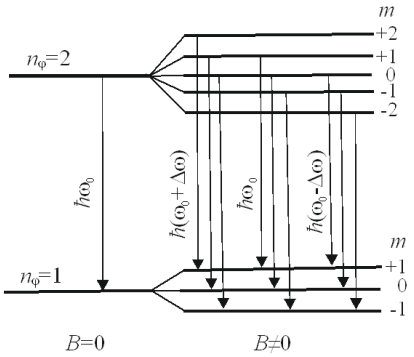


Рис. 17.13.Розщеплення рівнів і спектральних ліній випромінювання в ефекті Зеємана

Експериментально спостерігати ефект Зеємана можна в сильних магнітних полях, коли $B \sim 1$ Тл. Для видимої області спектру ($\lambda \sim 0{,}55$ мкм) і $m = +1$, отримаємо $\Delta\lambda \approx 0{,}14 \cdot 10^{-10}$ м $= 0{,}14$ Å.

Як бачимо, розщеплення дуже мале, і для його спостереження потрібні прилади з великою роздільною здатністю.

Слід зазначити, що таким способом розщеплюється невелика кількість ліній спектру, і в більшості

випадків розщеплення має більш складний характер, особливо у слабких магнітних полях. Складну картину розщеплення спектральних ліній у слабких магнітних полях називають аномальним ефектом Зеємана, який у межах теорії Бора не має пояснення.

17.7. Дослід Штерна і Герлаха

Теорію просторового квантування було підтверджено в дослідах Штерна і Герлаха із розщеплення пучків атомів в неоднорідному магнітному полі у 1922 р. Прилад, який використовували в дослідах Штерна і Герлаха, схематично зображено на рис. 17.14. В балоні з вакуумом розміщено піч для випаровування срібла (атоми срібла воднеподібні). Для створення вузького направленого пучка використовували щілини коліматора. Потім атоми потрапляють у сильно неоднорідне магнітне поле з особливою формою полюсних наконечників. Неоднорідне магнітне поле створюють з таких міркувань. Силу, яка діє на магнітний момент в неоднорідному магнітному полі (див. підрозд. 5.8), виражають формулою $F_x = p_m \frac{dB}{dx} \cos\alpha$, де p_m – магнітний момент атома; α – кут між напрямками векторів $\vec{p}_m$ та $\vec{B}$; $\frac{dB}{dx}$ – градієнт індукції магнітного поля. Щоб сила F_x була якнайбільша, для цього і створювали сильно неоднорідне поле (похідна $\frac{dB}{dx}$ має велике значення). Коли магнітного поля немає, то на екрані (фотопластинці) пучок атомів залишав слід у вигляді смужки, яка відповідала розмірам пучка атомів (рис. 17.15, *а*).

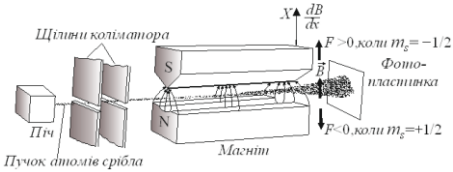
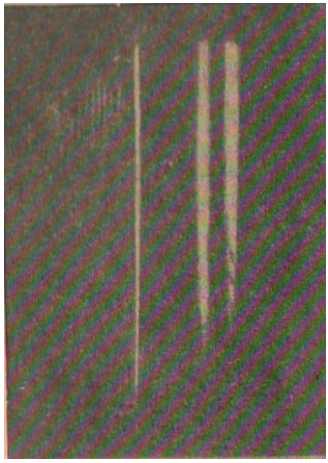


Рис. 17.14. Схема установки у дослідах Штерна і Герлаха



а *б*

Рис. 17.15. Розщеплення пучка парів натрію в магнітному полі: а – поля немає; б – поле наявне

Із класичної фізики відомо, що в магнітному полі площини орбіт відносно напрямку магнітного поля можуть

орієнтуватись як завгодно (кут α може набувати довільних значень), а оскільки на магнітні моменти в неоднорідному полі діє сила, то на екрані можна чекати розширення смужки порівняно з випадком, коли $B = 0$ (поля немає).

Відповідно до теорії просторового квантування площини орбіт атомів срібла можуть орієнтуватись щодо напрямку магнітного поля тільки дискретними способами (17.42). Якщо $n_\varphi = 1$, магнітне квантове число m може набувати значення −1, 0, +1. Атоми, які летять з магнітними моментами, що орієнтовані вздовж напрямку поля, тобто такі, для яких $\cos\alpha = +1$, мають відхилятись в один бік, атоми з магнітними моментами, які орієнтовані так, що для них $\cos\alpha = -1$, мають відхилятись настільки ж у протилежний бік. Атоми, які мають напрямок магнітного моменту, перпендикулярний полю, тобто для яких $\cos\alpha = 0$, не повинні зазнавати відхилення.

Пучок має розпастись на три окремі пучки. Дослід з атомами натрію (рис. 17.15, *б*), срібла (а пізніше і з атомами водню) показав, що атомний пучок справді розпадається, але не на три, а на два пучки (середнього пучка, який не має відхилятись, не було) (рис. 17.16). На цьому ж рис. 17.16 показано розщеплення атомних пучків для різних металів.

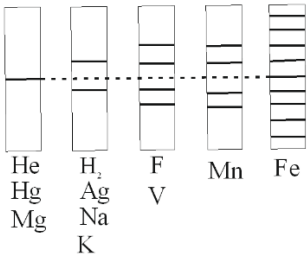


Рис. 17.16. розщеплення атомних пучків для різних металів в дослідах Штерна і Герлаха

Таким чином, ідея просторового квантування підтвердилась, однак механізм розщеплення пучка атомів натрію, срібла та інших елементів був незрозумілий.

Теорія Бора, навіть удосконалена, пояснити цього не могла. Справа в тому, що важливою особливістю атомів першої групи є те, що їхні валентні електрони перебувають в основному стані атома. У цьому стані в електрона немає моменту імпульсу L, а отже і магнітного моменту, тому і виникало питання про пояснення результатів дослідів Штерна і Герлаха. Постають такі запитання: просторове квантування якого магнітного моменту (тобто моменту імпульсу) було виявлено в цих дослідах і проєкція якого магнітного моменту дорівнює одному магнетону Бора?

17.8. Власний момент електрона

У спектрах випромінювання лужних металів, особливо в $^{23}_{11}\mathrm{Na}$, спостерігається одна різка жовта лінія. Утім, детальні дослідження дозволили виявити, що це не проста лінія, а дублет, який складається із двох ліній з довжиною хвиль 5890 і 5896 Å. Таке розщеплення однієї спектральної лінії на дві різні, які близько примикають одна до одної, відоме під назвою тонкої структури атомного спектру. Для її пояснення навіть удосконалена модель атома Бора вимагала доповнень.

Ще у 1921 р. А. Х. Комптон висловив припущення, що електрон має бути частинкою, яка подібна до дзиги. Згодом у 1925 р. С. Гаудсміт і Дж. Уленбек теоретично пояснили розщеплення спектру реальних ліній, припустивши, що електрон здійснює обертальний рух навколо власної осі. Відповідно до цієї гіпотези електрон, що обертається, повинен мати власний магнітний момент імпульсу $\vec{p}_{ms}$ і механічний момент імпульсу $\vec{L}_s$, який називають спіновим моментом (спін) (англ. spin – обертати). У 1928 р. Дірак створив релятивістську квантову теорію електрона, з якої властивості спіна випливали автоматично.

Спостерігач, який рухається разом з електроном по орбіті навколо протона, зазначить, що протон обертається по такій самій орбіті навколо електрона (рис. 17.17). Якщо протон рухається навколо електрона, створюється магнітне поле з індукцією $\vec{B}$, яке показане на рис. 17.17 у місці розташування електрона. На цьому рисунку $\vec{L}$ – орбітальний момент імпульсу електрона, який обертається навколо протона, $\vec{L}_s$ – спіновий момент електрона. Під час орбітального руху електрона створюється магнітний момент $\vec{p}_m$, а під час власного обертання електрона за аналогією зі сферою, яка негативно заряджена та обертається, створюється власний магнітний момент $\vec{p}_{ms}$ (рис. 17.18).

Магнітне поле, яке створюється у результаті обертання протона навколо електрона, взаємодіє з магнітним спіновим моментом $\vec{p}_{ms}$, тому до енергії електростатичної взаємодії електрона і протона треба додати додаткову енергію, яка становить

$$\Delta W_s = -\vec{p}_{ms}\vec{B} = -p_{ms}B\cos\theta' = \\ = -p_{ms}B\cos(\pi-\theta) = p_{ms}B\cos\theta = \\ = p_{msz}B, \qquad (17.43')$$

де p_{msz} – проєкція магнітного спінового моменту на вісь Z.

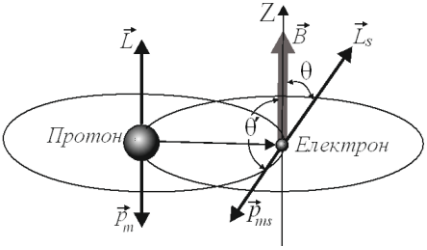


Рис. 17.17. Взаємодія між власними магнітним моментом p_{ms} електрона і магнітним полем

Взаємодія між власними магнітним моментом p_{ms} електрона і магнітним полем, яке створюється орбітальним рухом електрона, називають спін-орбітальною взаємодією. Ця взаємодія є джерелом додаткових енергетичних рівнів, які утворюють тонку структуру спектру.

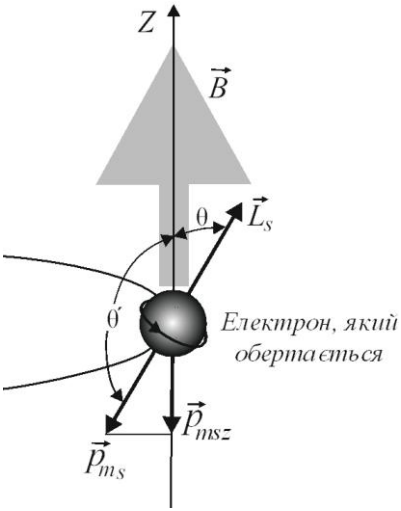


Рис. 17.18. Власний магнітний момент електрона

Примітка. Уявлення про електрон, який обертається навколо своєї осі, використовують тільки для наочності, але воно не має буквального змісту. Якщо припустити, що обертання справді відбувається і взяти для електрона розумні розміри, то для отримання відомого значення власного механічного моменту імпульсу електрона довелось би допустити існування лінійних швидкостей обертального руху, які перевищують швидкість світла.

За допомогою магнітного спінового моменту вдалось пояснити аномальний ефект Зеємана, розщеплення пучка атомів натрію, срібла на два пучки в дослідах Штерна і Герлаха, природу жовтого дублета у спектрах випромінювання натрію та інше.

Отже, треба ввести ще одне квантове число, яке характеризує власний момент імпульсу електрона. Згідно з релятивістською квантовою теорією Дірака власний механічний момент імпульсу квантований відповідно до формули

$$L_s = \sqrt{s(s+1)}\hbar. \qquad (17.44)$$

Спінове квантове число s, яке входить у цю формулу, має тільки єдине значення, що дорівнює $1/2$, тому єдиним можливим значенням L_s буде

$$L_s = \sqrt{\frac{1}{2}\left(\frac{1}{2}+1\right)}\hbar = \sqrt{3}\frac{\hbar}{2}.$$

Число s, як правило, не використовують разом з іншими квантовими числами, оскільки має єдине значення ($\frac{1}{2}$) і не вносить різниці між станами. Власний механічний момент $L_s = \sqrt{3}\frac{\hbar}{2}$ є фундаментальною характеристикою електрона, як його маса або заряд. Величину L_s у напрямку Z (напрямок магнітного поля $\vec{B}$) квантований, її визначають як

$$L_{sz} = L_s\cos\theta = m_s\hbar, \quad (17.45)$$

де $m_s = \pm 1/2$ – магнітне спінове квантове число. Часто $m_s = +1/2$ позначають як «спін-вгору» (↑), а $-m_s = -1/2$ як «спін-вниз» (↓) (рис. 17.19).

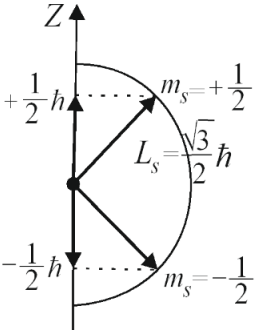


Рис. 17.19. Орієнтація спінового момента відносно осі Z

Гіромагнітне відношення для спінового моменту

$$p_{ms} = -\frac{e}{m_e}L_s,$$

і, відповідно, $p_{msz} = -\frac{e}{m_e}L_{sz}$.

З урахуванням (17.45)

$$p_{msz} = -\frac{e\hbar}{m_e}m_s,$$

або

$$p_{msz} = \pm\frac{e\hbar}{2m_e} = \pm\mu_Б. \quad (17.46)$$

Таким чином, додаткова енергія спін-орбітальної взаємодії (17.43′) становитиме

$$\Delta W_s = \pm\mu_Б B. \quad (17.47)$$

Два значення енергії ΔW_s пояснюють жовтий дублет у спектрі випромінювання натрію та аномальний ефект Зеємана.

17.9. Квантові стани електронів в атомі. Квантові числа

Удосконалена теорія Бора і навіть наявність в електрона спінового магнітного момента не могли пояснити розділення пучка атомів срібла в дослідах Штерна і Герлаха на два пучки. Справа в тому, що відповідно до теорії Бора–Зоммерфельда для найменшого з можливих значень азимутального квантового числа $n_\varphi = 1$ згідно з умовою просторового квантування магнітне квантове число m має набувати трьох значень (−1, 0, 1) (див. підрозд. 17.5) і пучок атомів має розщеплюватись на три пучки. Досліди Штерна і Герлаха, як уже зазначалось, пояснюється

наявністю в електронів спінового магнітного моменту, проєкція якого на напрямок магнітного поля може набувати тільки двох значень: $\pm\mu_Б$. При цьому вважають, що існує тільки спіновий магнітний момент в електрона, але немає орбітального магнітного моменту електрона. Повне пояснення цього явища було отримано за допомогою хвильової (квантової) механіки, основні положення якої ми розглянемо далі. Згідно з квантовою механікою в таких атомах, як водень, а також воднеподібних атомах, зокрема і в атомах срібла, орбітальний момент електрона в нормальному стані дорівнює нулю (у планетарній моделі Бора такого бути не може). Тому одночасно із введенням поняття «спіна» електрона Уленбеку і Гаудсміту довелось ще ввести декілька змін і доповнень в теорію атома. Для пояснення розщеплення пучка атомів срібла (водню) в дослідах Штерна і Герлаха на два пучки (а не на три) азимутальне квантове число n_φ вони замінили на l:

$$l = 0, 1, 2, \dots, n-1. \quad (17.48)$$

Це квантове число l називають орбітальним квантовим числом, яким визначають величину орбітального механічного моменту (орбітального моменту імпульсу L), а також величину орбітального магнітного моменту електрона p_m. У квантовій механіці показують, що орбітальний момент імпульсу, таким чином, виражають через орбітальне квантове число l:

$$L = \sqrt{l(l+1)}\hbar. \quad (17.49)$$

Орбітальний магнітний момент із врахуванням гіромагнітного відношення (17.40) визначають так:

$$p_m = -\frac{e}{2m_e}L = -\frac{e\hbar}{2m_e}\sqrt{l(l+1)} = \\ = -\sqrt{l(l+1)}\mu_Б. \quad (17.50)$$

Магнітне квантове число m, яке характеризує квантування проєкції орбітального моменту імпульсу на напрямок магнітного поля (17.39), може набувати значення цілих чисел від $-l$ до $+l$:

$$m = -l, \ -(l+1), \dots, -1, \dots, \\ 0, +1, \dots +(l-1), +l, \quad (17.51)$$

тобто $2l$ +1 значень.

Таким чином, стан електрона атома характеризується сукупністю чотирьох квантових чисел n, l, m, m_s, які для зручності зведено в табл. 17.2.

Замість наведеного набору квантових чисел n, l, m, m_s, які характеризують стан руху електрона в атомі, можна ввести інший набір квантових чисел n, l, j, m_j, розглядаючи повний момент імпульсу $\vec{L}_j$ електрона в атомі. Повний момент імпульсу

визначається геометричною (векторною) сумою орбітального момента $\vec{L}$ і власного моменту імпульсу електрона $\vec{L}_s$: $\vec{L}_j = \vec{L} + \vec{L}_s$. При цьому моменти $\vec{L}$ і $\vec{L}_s$ можуть складати між собою тільки такі кути, щоб повний момент L_j набував таких значень:

$$L_j = \sqrt{j(j+1)}\hbar, \qquad (17.52)$$

де $j = l + m_s$ повне, або внутрішнє квантове число.

Таблиця 17.2. Квантові числа

Назва	Визначає	Символ	Приймає значення	Повне число значень
Головне квантове число	Розмір орбіти, енергію електрона	n	1,2,3...	∞
Орбітальне квантове число	Витягнутість (форму) орбіти, частково енергію електрона	l	0,1,2..., n-1	n
Магнітне квантове число	Орієнтацію орбіти в просторі, слабо енергію електрона	m	0,±1,±2,..., ±l	2l+1
Спінове квантове число	Орієнтацію спіна, слабо енергію електрона	m_s	-1/2;+1/2	2

Магнітне квантове число m_j визначає проєкцію повного моменту імпульсу на напрямок *Z*. Оскільки $m_s = \pm$½, то за будь-якого $l \neq 0$ квантове число *j* набуває двох значень $j = l \pm$ ½, а за $l = 0$ квантове число j набуває одного значення $|j| =$ ½.

Оскільки можливі значення L_j знаходять за допомогою векторного додавання, то розглядувану модель називають векторною моделлю. На основі векторної моделі можна пояснити аномальний ефект Зеємана і більш складні випадки в дослідах Штерна і Герлаха.

17.10. Електронна оболонка атома. Принцип Паулі

Стан кожного електрона в атомі визначається чотирма квантовими числами: n, l, m, m_s. Енергія електрона в атомі залежить переважно від головного квантового числа n (формула (17.20)). Квантові числа l, m, m_s у формулу (17.20) не входять, однак в реальних умовах в атомах завжди існують магнітні електричні поля, які діють на електрони. Отже, енергія тією чи іншою мірою залежить від квантових чисел l, m, m_s, чого теорія Бора не враховує.

Система електронів в незбуджених атомах намагається зайняти стан з найменшою енергією W_n, що відповідає основному принципу термодинаміки для систем, що перебувають у стані рівноваги, тому електрони мають розміщуватись на найнижчих доступних для них рівнях. Здавалося б, усі електрони мають перебувати на найнижчому (з найменшою енергією) рівні ($n = 1$, $l = 0$), однак практика показує, що це далеко не так.

Влітку 1922 р. Вольфгант Паулі приїхав у Копенгаген, щоб працювати

разом із Бором над поясненням атомного ефекту Зеємана. Спроби класифікувати рівні енергії електрона, який перебуває в сильному магнітному полі, поступово підводили Паулі до формулювання принципу Паулі. Остаточно цей принцип визрів після того, як Уленбек і Гаудсміт висунули ідею про існування спіна електрона і було добавлене квантове число m_s.

Закон фізики, відомий під назвою *принципу Паулі* або *принципу заборони*, встановлює, *що не може бути більше, ніж одного електрона в такому стані, якщо цей стан повністю можна описати чотирма квантовими числами:* n, l, m, m_s.

Принцип Паулі можна сформулювати і в іншій формі: *в одному атомі не може бути двох електронів з одним і тим самим набором квантових чисел* n, l, m, m_s. *Принцип Паулі справедливий для мікрочастинок із спіном* $\hbar/2$ *і є наслідком принципу тотожності частинок у квантовій механіці.*

Дія принципу Паулі й обумовлює послідовність заповнення енергетичних рівнів в атомі.

Цьому значенню квантового числа n відповідають такі значення квантових чисел l, m, m_s:

$$l = 0, 1, 2, \ldots, n-1;$$

$$m = 0, \pm 1, \pm 2, \ , \pm l;$$

$$m_s = +1/2, -1/2.$$

Електрони з енергією, яка відповідає значенню l, утворюють оболонки, які позначають малими латинськими літерами *s*, *p*, *d*, *f*, ... (рис. 17).

l	0	1	2	3
Оболонка	*s*	*p*	*d*	*f*

Рис. 17. Електронні оболонки в атомі

Сукупність оболонок, які відповідають квантовому числу n, утворюють шари, які позначають великими латинськими буквами *K, L, M, N, O,...* (рис. 17′).

n	1	2	3	4	5	6	7
Шар	*K*	*L*	*M*	*N*	*O*	*P*	*Q*

Рис. 17′. *Електронні шари в атомі*

Найбільш близько до ядра розміщений *K*-шар, потім *L*-шар, *M*-шар, *N*-шар, …. .

Стан із заданими n і l записують так: *ns*, *np*, *nd*, *nf* і говорять, наприклад, так: 1*s* – стан, або 1*s* – оболонка (при n = 1), 2*s*-стан, або 2*s*-оболонка (n = 2), 3*p*-стан, або 3*p*-оболонка і т. д.

Кожному шару з конкретним значенням n відповідає певна кількість станів, які визначаються квантовими числами l, m, m_s. Наприклад, за n =

$=1, l = 0, m = 0, m_s = \pm\frac{1}{2}$, тобто в цьому шарі з оболонкою 1*s* можуть перебувати тільки два електрони, які відрізнятимуться квантовими числами $m_s = +1/2, m_s = -1/2$.

За $n = 2$ можуть бути дві оболонки: 2*s* (l = 0) та 2*p* (l = 1). На 2*s*-оболонці може бути тільки два електрони, які відрізнятимуться квантовими числами $m_s = +1/2$ та $m_s = -1/2$. На оболонці 2*p* може бути шість електронів. Справді, за $l = 1, m = -1, 0, +1$ і кожному m відповідають два стани, які відрізняються квантовими числами $m_s = +1/2, m_s = -1/2$. Таким чином, за $n = 2$ кількість станів дорівнює восьми.

Не важко бачити, що для такої комбінації двох чисел n та l можлива $2(2l + 1)$ кількість різних квантових станів, які відрізняються значеннями чисел m та m_s. Для цього значення головного квантового числа *n* кількість різних станів

$$N = \sum_{l=0}^{n-1} 2(2l + 1) = 2n^2.$$

Нижче наведено таблицю квантових станів (табл. 17.3). Для зручності квантові стани $m_s = \pm 1/2$ будемо позначати стрілками (↑↓), які відповідають різним орієнтаціям спіна.

Кількість станів, що зайняті електронами, визначає електронну конфігурацію атома. Наприклад, в атома кисню є 8 електронів, а його електронна конфігурація $1s^2 2s^2 2p^4$ (індекс вгорі показує кількість електронів у цьому стані). В 1*s*-стані перебуває 2 електрони, у 2*s*-стані – 2 електрони, у 2*p*-стані – 4 електрони, таким чином разом – 8 електронів.

Таблиця 17.3. Квантові стани

Шар	*n*	*l*	*m*	m_s	Оболонка	Кількість станів
K	1	0	0	↑↓	1*s*	2
L	2	0	0	↑↓	2*s*	8
		1	-1 0 +1	↑↓ ↑↓ ↑↓	2*p*	
M	3	0	0	↑↓	3*s*	18
		1	-1 0 +1	↑↓ ↑↓ ↑↓	3*p*	
		2	-2 -1 0 +1 +2	↑↓ ↑↓ ↑↓ ↑↓ ↑↓	3*d*	

17.11. Періодична система хімічних елементів

Прослідкуємо побудову таблиці хімічних елементів. Щоб розібратися в розподілі електронів в хімічних елементах, будемо вважати, що кожний черговий елемент утворюється із попереднього шляхом додавання до ядра одного позитивного протона (і відповідного числа нейтронів) й додавання одного електрона, який будемо розміщувати в доступний йому згідно з

принципом Паулі стан з найменшою енергією. Під час заповнення оболонок електронами будемо користуватись правилом зростання суми $n+l$. Оскільки енергія електрона в атомі залежить і від n, і від l, то заповнення оболонок електронами відбувається в порядку зростання суми $n+l$ (табл. 17.4). Із двох однакових сум, наприклад, $n_1+l_1=n_2+l_2$, перевагу слід надавати сумі, в якій менше n, оскільки енергія сильніше залежить від n, ніж від l.

Таблиця 17.4.
Заповнення оболонок електронами відбувається в порядку зростання суми $n+l$

Шар	Оболонка	$n+l$
K	$1s$	1
L	$2s$ $2p$	2 3
M	$3s$ $3p$ $3d$	3 4 5
N	$4s$ $4p$ $4d$ $4f$	4 5 6 7
O	$5s$ $5p$ $5d$ $5f$	5 6 7 8
P	$6s$ $6p$ $6d$ $6f$	6 7 8 9

Наведемо приклад:

– неон $_{10}\mathrm{Ne}-1s^2 2s^2 2p^6$ – інертний газ;

– натрій $_{11}\mathrm{Na}-1s^2 2s^2 2p^6 3s^1$ – метал;

– кремній $_{14}\mathrm{Si}-s^2 2s^2 2p^6 3s^2 3p^2$ – напівпровідник;

– аргон $_{18}\mathrm{Ar}-1s^2 2s^2 2p^6 3s^2 3p^6$ – інертний газ;

– калій $_{19}\mathrm{K}-1s^2 2s^2 2p^6 3s^2 3p^6 3d 4s^1$ – метал;

– нікель $_{28}\mathrm{Ni}-1s^2 2s^2 2p^6 3s^2 3p^6 3d^8\ 4s^2-$ метал;

– мідь $_{29}\mathrm{Cu}-1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^1$ –метал, хороший провідник;

– германій $_{32}\mathrm{Ge}-1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^2 4p^2-$ напівпровідник;

– цезій $_{55}\mathrm{Cs}-1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^2 4p^6 4d^{10} 4f 5s^2 5p^6 6s^1-$ рідкоземельний елемент;

– неодим $_{60}\mathrm{Nd}\ -1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^2 4p^6 4d^{10} 4f^4 5s^2 5p^6 6s^2-$ рідкоземельний елемент.

Для атома міді спостерігається порушення правила зростання суми $n+l$. Згідно з цим правилом електронна конфігурація для міді мала би бути така: $1s^2 2s^2 2p^6 3s^2 3p^6 3d^9 4s^2$, але енергія електронів буде меншою, якщо $3d$-оболонка буде повністю заповнена електронами. Отже, один $4s$-електрон дозаповнює $3d$-оболонку і на $4s$-

оболонці залишається один валентний електрон. Цей $4s^1$-електрон і обумовлює ряд фізичних властивостей міді, як то висока електропровідність і теплопровідність та інше.

Виходячи з цього, можемо стверджувати, що речовини, атоми яких мають валентний s^1-електрон, є металами (провідниками); p^6-електрони є інертними газами; $p^{2\ldots6}$-електрони є напівпровідниками; d^{1-10}-електрони є речовинами групи заліза; f^{1-14}-.електрони є рідкоземельними елементами.

Таким чином, повторюваність будови зовнішньої електронної оболонки, яка випливає із принципу Паулі, і пояснює розміщення хімічних елементів в періодичній таблиці.

17.12. Рентгенівські спектри

У 1895 р. В. Рентген відкрив випромінювання, яке було названо ікс-променями. Рентгенівські хвилі охоплюють широкий інтервал довжини від 0,01 до 10^{-8} мкм. Джерелом рентгенівського випромінювання є рентгенівські трубки. На рис. 17.20 зображено схему рентгенівської трубки. У скляній трубці, в якій створено вакуум ~10^{-6} мм рт.ст., є анод *A* і катод *K* (джерело електронів).

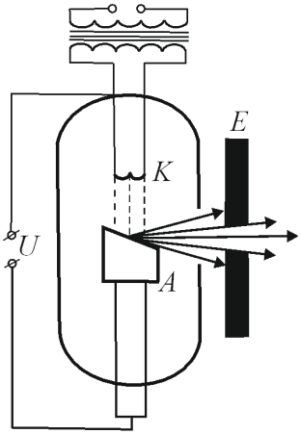


Рис. 17.20 Схема рентгенівської трубки

Між катодом і анодом утворюється сильне електричне поле з різницею потенціалів U ~ 10 … 50 кВ. Електрони, які набули високих енергій у прискорювальному електричному полі, бомбардують анод *A*, площина якого утворює з напрямом руху електронів кут 45°. Унаслідок цього рентгенівське випромінювання, що виникає у процесі гальмуванні електронів анодом, виходить із трубки через отвір захисного свинцевого екрана *E*. Коефіцієнт корисної дії рентгенівських трубок становить близько 1 %, тобто 99 % кінетичної енергії електронів перетворюється в аноді у теплоту, тому в тілі анода роблять канали, в яких циркулює охолоджувальна рідина. За принципом отримання електронних пучків рентгенівські трубки поділяють

на три типи: іонні, електронні, індукційні. Джерелами рентгенівського випромінювання також можуть бути природні або штучні радіоактивні елементи і ряд небесних тіл.

Рентгенівське випромінювання, яке виникає в результаті гальмування електронів, називають гальмівним. Виникнення гальмівного випромінювання можна пояснити так. Електрони, які рухаються, утворюють навколо себе магнітне поле. Різка зміна швидкості електронів при ударі в анод рівносильна ослабленню і зникненню струму, що приводить до зміни магнітного поля, у результаті чого і виникають електромагнітні хвилі.

Гальмівне випромінювання має суцільний спектр. Це пояснюється тим, що одні електрони гальмуються швидше, а інші повільніше. Це приводить до виникнення електромагнітного випромінювання з різними довжинами хвиль. За квантовою теорією енергія фотона рентгенівського випромінювання

$$h\nu = \frac{m_e v^2}{2} - Q,$$

де $m_e v^2/2$ – кінетична енергія електрона перед ударом в анод, а Q – енергія, яка перетворюється в теплоту. Оскільки величина Q має різні значення, то й енергія $h\nu$ може бути різною.

На рис. 17.21 зображено експериментальні криві розподілу інтенсивності рентгенівського випромінювання за довжинами хвиль λ при ударі електронів об вольфрамовий анод за різних значень прискорювальної напруги U, а на рис. 17.22 – при ударі об анод з вольфраму, молібдену і хрому за однакових напруг U. Характерною особливістю суцільних спектрів є наявність чіткої короткохвильової межі λ_{min} та її незалежність від речовин анода. Зі зростанням напруги інтенсивність випромінювання збільшується, а короткохвильова межа зміщується у бік коротких хвиль (рис. 17.21).

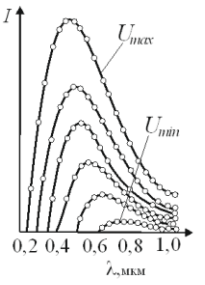


Рис. 17.21

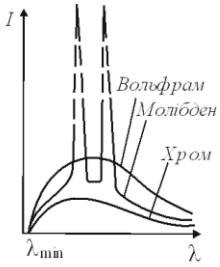


Рис. 17.22

Рис. 17.21. Розподіл інтенсивності рентгенівського випромінювання за довжинами хвиль λ за різних напруг U. Рис. 17.22. Розподіл інтенсивності рентгенівського випромінювання за довжинами хвиль λ при ударі об анод з вольфраму, молібдену і хрому за однакових напруг U

Існування короткохвильової межі λ_{min} в рентгенівському спектрі

пояснюють так. Енергія електрона при ударі об анод може повністю перейти в енергію випромінювання, тобто

$$\frac{m_e v^2}{2} = h\nu_{\max} = \frac{hc}{\lambda_{\min}}.$$

Оскільки $\frac{m_e v^2}{2} = eU$, то

$$eU = \frac{hc}{\lambda_{\min}} \quad \text{та} \quad \lambda_{\min} = \frac{hc}{eU}.$$

Отже, мінімальна довжина хвилі гальмівного випромінювання обернено пропорційна прискорювальній напрузі U.

Зі зростанням напруги на рентгенівській трубці до величини, більшої за деяке критичне значення, що залежить від речовини анода, на фоні суцільного спектра виникають лінійчасті спектри, інтенсивність яких у тисячі разів перевищує інтенсивність суцільного спектра. Оскільки лінійчастий спектр визначається природою речовини, з якої виготовлено анод, то його називають характеристичним.

З рис. 17.22 видно, що для напруги, за якої рентгенівська трубка з анодом із вольфраму і хрому випромінює суцільний спектр, трубка з молібденовим анодом, крім суцільного, випромінює ще й лінійчастий спектр.

Існування *характеристичного випромінювання* підтверджує наявність дискретних рівнів енергії атомів.

Кожний хімічний елемент дає характеристичне рентгенівське випромінювання незалежно від того, чи його атоми вільні, чи входять до складу хімічних сполук.

Рентгенівські спектральні лінії утворюють певні серії, які позначають літерами *K*, *L*, *M*, *N*, *O* і т.д. Кожна серія має невелику кількість ліній, які позначають за порядком зменшення довжини хвилі індексами α, β, γ і т.д. Лінійчасті рентгенівські спектри різних речовин подібні між собою. Це зумовлено подібністю забудови нижніх (найближчих до ядра) шарів електронів в атомах. Зі збільшенням атомного номера речовини анода весь рентгенівський спектр лише зміщується в короткохвильову частину спектра, не змінюючи своєї структури. Це можна пояснити тим, що лінійчасті спектри утворюються внаслідок переходів електронів у внутрішніх шарах атомів. Механізм виникнення лінійчастих спектрів можна пояснити так. Якщо під впливом падаючих електронів високих енергій на атоми анода вибивається один із двох електронів 1*s*-стану *K*-оболонки, то звільнене місце може зайняти електрон з якої-небудь зовнішньої оболонки (*L*, *M*, *N* і т.д.). При цьому виникає *K*-серія. Аналогічно виникають

й інші серії. Схему виникнення спектральних серій показано на рис. 17.23.

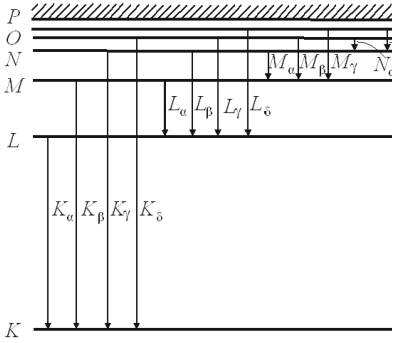


Рис. 17.23. Схема виникнення спектральних серій

Серія K обов'язково супроводжується іншими серіями, оскільки під час випромінювання її ліній звільняються рівні в оболонках L, M, і т.д., які будуть заповнюватися електронами з більш віддалених оболонок. Частота ліній зростає у разі переходу від лінії K_α до K_β, K_γ, …, що пов'язано зі збільшенням енергії, яка вивільняється під час переходу електрона на K-оболонку з дедалі більш віддалених оболонок. Інтенсивність ліній спадає з переходом від лінії K_α до лінії K_β, K_γ, бо ймовірність переходу на K-оболонку з L-оболонки більша, ніж з M-, N- і т.д. Ці закономірності справедливі для L-, M- та інших серій.

Під час дослідження лінійчастих рентгенівських спектрів Г. Мозлі в 1913 р. встановив закон, який виражає зв'язок між частотами спектральних ліній ω та атомним номером Z елемента, який випромінює ці лінії. Так, для K_α-лінії ця залежність має вигляд

$$\omega_{K_\alpha} = R(Z-1)^2\left(\frac{1}{1^2}-\frac{1}{2^2}\right),$$

де R – стала Рідберга. Для лінії K_β

$$\omega_{K_\beta} = R(Z-1)^2\left(\frac{1}{1^2}-\frac{1}{3^2}\right),$$

а для лінії L_α

$$\omega_{L_\alpha} = R(Z-7{,}5)^2\left(\frac{1}{2^2}-\frac{1}{3^2}\right)$$

і т.д. Узагальнений закон Мозлі має вигляд

$$\omega = R(Z-\sigma)^2\left(\frac{1}{k^2}-\frac{1}{n^2}\right), \quad (17.53)$$

де k = 1, n = 2, 3, 4… для ліній K-серії; k = 2, n = 3, 4… для ліній L-серії; k = 3, n = 4, 5, 6… для ліній M-серії; σ – стала екранування, яка має певне значення в межах однієї серії. Мозлі зазначив, що лінії з частотами, які визначають формулою (17.53), збігаються з лініями, які випромінюються під час переходу електрона, що перебуває в полі заряду $(Z-\sigma)e$, з рівня з номером n на рівень з номером m.

Зміст сталої екранування полягає в тому, що у важкому атомі, який має Z електронів, на електрон, що, приміром, здійснює перехід, який відповідає лінії K_α, діє не весь заряд ядра Ze, а заряд

$(Z - \sigma)e$, ослаблений екранувальною дією одного електрона, що залишився на K-оболонці. Для K-серії $\sigma = 1$, для L-серії $\sigma = 7{,}5$.

У загальному вигляді закон Мозлі записують так:

$$\sqrt{\omega} = a(Z - \sigma), \qquad (17.54)$$

де ω – частота; a – стала для певної серії ліній.

Закон Мозлі можна сформулювати так: корінь квадратний з частоти є лінійною функцією атомного номера Z. На рис. 17.24 зображено побудовані по експериментальних точках графіки залежності $\sqrt{\omega}$ від Z для ліній K_α та L_α. За цими графіками можна судити, наскільки точно виконується закон Мозлі. *Закон Мозлі дозволяє за виміряною довжиною хвилі рентгенівських ліній точно встановити атомний номер певного елемента.* Цей закон відіграв значну роль для розміщення елементів в періодичній системі. На основі цього закону вперше було показано, що не атомна маса, а атомний номер, який дорівнює заряду ядра, визначає хімічні властивості атомів. Мозлі дав просте теоретичне пояснення знайденого ним закону. Він зазначив, що лінії з частотами, які визначають формулою (17.53), збігаються з лініями, які випромінюються при переході електрона, що перебуває в полі заряду $(Z - \sigma)e$, з рівня з номером n на рівень з номером k.

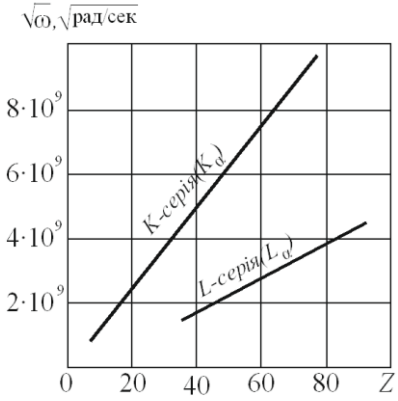


Рис. 17.24 Експериментальна перевірка виконання закону Мозлі (17.54)

Зауважимо, що формула (17.53) ґрунтується на припущенні, що стала екранування σ для обох термів має однакове значення. Реально екранування, наприклад для K-терма буде слабшим, ніж для L-терма, тому що електрон, який перебуває в L-оболонці, екранують обидва електрони K-оболонки і, крім цього, частково беруть участь в екрануванні останні електрони L-оболонки, тоді як для електрона K-оболонки екранування здійснюється тільки одним другим K-електроном. Більш строго формулу (17.53) слід записати у вигляді

$$\omega = R\left\{\frac{(Z-\sigma_1)^2}{k^2} - \frac{(Z-\sigma_2)^2}{n^2}\right\}. \qquad 17.55)$$

Іван Пулюй (1845–1918) у 1896 р. опублікував дві статті: «Про походження рентгенівських променів та їх фотографічну дію» і додаток до праці «Про походження рентгенівських променів та їх фотографічну дію». Він навів близько десяти досконалих знімків, які ілюструють отримані результати. Рентгенограми відрізнялись від знімків Рентгена високою якістю. Пулюєві належить перша вдала спроба з'ясувати механізм виникнення рентгенівських променів, таких він зробив перший знімок цілого людського скелета.

Контрольні запитання

1. У чому суть ядерної моделі атома?

2. У чому полягає неспроможність класичної теорії у поясненні ядерної моделі атома?

3. Сформулюйте постулати Бора. Які досліди підтвердили ці постулати?

4. Які є серії спектральних ліній у спектрі випромінювання атома водню?

5. Що покладено в основу борівської теорії будови атома водню і воднеподібних атомів?

6. Яке співвідношення між механічним і магнітним моментами електрона в атомі?

7. Що означає просторове квантування?

8. Що підтверджують досліди Штерна і Герлаха з атомними пучками?

9. Що підтверджують досліди Зеємана?

10. Які квантові числа описують стан електрона в атомі?

11. Поясніть фізичний зміст кожного квантового числа.

12. Сформулюйте й поясніть принцип Паулі.

13. Яке правило заповнення електронних оболонок атома електронами?

14. Поясніть принцип розміщення хімічних елементів в періодичній таблиці.

15. Що таке рентгенівські спектри? Яка їхня роль в поясненні будови атома?

Розділ 18. Елементи квантової механіки

Квантова механіка (хвильова механіка) – це розділ теоретичної фізики, в якому вивчають закони руху частинок у сфері мікросвіту (в масштабах 10^{-8} ... 10^{-15} м*).* Об'єктами вивчення квантової механіки є кристали, молекули та атоми, атомні ядра, елементарні частинки. Закони квантової

механіки є фундаментом науки, що вивчає будову речовини.

В основу квантової механіки покладено:

1. Уявлення Планка про кванти енергії $W = \hbar\omega = h\nu$.

2. Уявлення Ейнштейна про фотони. Ейнштейн припустив, що світло не тільки випромінюється і поглинається квантами, але й поширюється квантами, тобто дискретність властива самому світлу: світло складається з окремих порцій – світлових квантів, які пізніше назвали фотонами. Енергія фотона $W = h\nu$. На основі цієї гіпотези Ейнштейн пояснив закономірності зовнішнього фотоефекту.

3. Дані про існування дискретних значень деяких фізичних величин, які характеризують стан частинки в мікросвіті (квантування: енергії, момента кількості руху, магнітного момента; спектри випромі-нювання і т.п.).

4. Гіпотеза де Бройля.

18.1. Протиріччя «хвиля – частинка»

З часів філософів античності до Ньютона вчені розглядали світло як потік частинок (корпускул), які швидко рухаються. У 1801 р. Томас Юнг, здійснивши перші досліди з інтерференції світла, підвів тверду експериментальну основу під його хвильову теорію. Здавалось, що фізики можуть остаточно відмовитись від корпускулярної теорії світла, однак наступні дослідження показали, що дедалі більша кількість фізичних властивостей твердих тіл, рідин і газів можна пояснити на основі атомістичних уявлень.

У кінці XIX ст. у результаті експериментів Дж. Дж. Томсона і робіт Х. А. Лоренца було сформульовано концепцію про існування елементарних частинок електрики – електронів. Нарешті «коло замкнулось», коли у 1900 р. Макс Планк встановив закон випромінювання абсолютно чорного тіла, згідно з яким джерело світла випромінює світло не безперервно, а однаковими й кінцевими порціями, які отримали назву квантів. З новою силою протиріччя «хвиля – частинка» проявилось у 1923 р., коли Артур Комптон експериментально довів, що кванти рентгенівського випромінювання поводять себе як частинки з певною енергією і певним імпульсом.

У 1924 р. Луї де Бройль у докторській дисертації «Дослідження з теорії квантів» висловив думку про те, що якщо світло в багатьох випадках

проявляє корпускулярні властивості, то і частинки, зокрема електрони, завдяки симетрії у природі можуть мати хвильові властивості. У 1927 р. ідеї де Бройля були експериментально підтверджені К. Дж. Девіссоном і Л. Х. Джермером й незалежно від них Дж. П. Томсоном, які показали, що для пучка електронів, як і для світлового пучка, спостерігається явище дифракції.

18.2. Хвилі де Бройля

Планк пов'язав енергію частинки світла (фотонів) з частотою світла співвідношенням

$$W = h\nu, \qquad (18.1)$$

де ν – частота фотона, а енергія W збігається із значенням, яке знаходять із спеціальної теорії відносності (із формули $W = W_0 + W_k$, де для фотонів енергія спокою $W_0 = m_\lambda c^2 = 0$, оскільки $m_\lambda = 0$ (див. підрозд. 16.6)) загальна енергія W дорівнює тільки кінетичній енергії W_k. Утім, суто корпускулярна теорія не містить жодного елемента, який дозволив би визначити частоту. Щоб розв'язати цю проблему, де Бройль висловив гіпотезу, згідно з якою з кожним фотоном пов'язаний хвильовий процес. Це дало можливість пояснити таке суто хвильове явище, як інтерференцію. Комптон показав, що світлові хвилі мають імпульс, тобто характеристику, яку найчастіше пов'язують з частинками. Вираз для енергії фотона можна записати у вигляді

$$W = mc^2 = pc, \qquad (18.2)$$

де $p = mc$ – імпульс фотона. Об'єднавши (18.1) і (18.2), отримаємо співвідношення

$$p = \frac{h\nu}{c} = \frac{h}{\lambda}, \qquad (18.3)$$

яке пов'язує довжину хвилі λ і частоту ν, тобто величини, які характеризують фотон як хвилю з імпульсом фотона p. Згодом де Бройль висловив чудову ідею: якщо світлові хвилі мають корпускулярну природу, то частинки, наприклад електрони, мають проявляти хвильові властивості. Таким чином, електрон може мати певний імпульс

$$p = mv = \frac{h}{\lambda}, \qquad (18.4)$$

де m – релятивістська маса; v – швидкість електрона. Із цього співвідношення випливає, що електрону (частинці) приписують певну хвильову характеристику – довжину хвилі λ (частоту ν). Тоді із (18.4) отримуємо, що довжина хвилі

$$\lambda = \frac{h}{mv} = \frac{h}{p}. \qquad (18.5)$$

Співвідношення (18.5) відображає гіпотезу де Бройля.

В теорії відносності, так само як і у класичній механіці, для замкненої

фізичної системи зберігається імпульс $\vec{p}$ та енергія W. Величина $W^2 - (c\vec{p})^2 = = m_0^2c^4$ є інваріантом під час перетворень Лоренца в релятивістській механіці; у цій формулі m_0 – маса спокою частинки; W – повна енергія частинки; c – швидкість світла у вакуумі. Отже,

$$p = (1/c)\sqrt{W^2 - W_0^2},$$

де $W = mc^2$ – повна енергія; $W_0 = m_0c^2$ – енергія спокою. Тоді (18.5) перепишемо так:

$$\lambda = \frac{h}{(1/c)\sqrt{W^2 - W_0^2}} = = \frac{h}{\left(\frac{1}{c}\right)\sqrt{(mc^2)^2 - (m_0c^2)^2}}.$$

Звідси, якщо врахувати, що релятивістська маса $m = m_0/\sqrt{1 - \frac{v^2}{c^2}}$, отримаємо:

$$\lambda = \frac{h}{m_0c\sqrt{1/\left(1 - \frac{v^2}{c^2}\right) - 1}},$$

а після спрощення

$$\lambda = \frac{h\sqrt{1 - v^2/c^2}}{m_0v} = \frac{h}{mv}.$$

Для нерелятивістського випадку ($v \ll c$) релятивістська маса m дорівнює масі спокою частинки m_0.

Хоча у перших роботах де Бройль розглядав проблему на прикладі електрона, усі отримані співвідношення справедливі для будь-яких матеріальних тіл. М'яч, маса якого 0,14 кг, запускають сильним ударом, і він набуває швидкості близько 40 м/с. Довжина хвилі де Бройля, яка пов'язана з таким м'ячем,

$$\lambda = \frac{h}{m_0v} = \frac{6{,}625 \cdot 10^{-34}}{0{,}14 \cdot 40}\text{ м} = = 1{,}2 \cdot 10^{-34}\text{м} = 1{,}2 \cdot 10^{-24}\text{ Å},$$

тобто λ є настільки малою величиною, що її неможливо визначити з досліду. Рух електрона з такою самою швидкістю супроводжується хвильовим процесом з довжиною хвилі

$$\lambda = \frac{h}{m_ev} = \frac{6{,}625 \cdot 10^{-34}}{9{,}1 \cdot 10^{-31} \cdot 40}\text{ м} = = 1{,}8 \cdot 10^{-5}\text{ м} = 1{,}8 \cdot 10^{5}\text{ Å},$$

яка може бути знайдена з експерименту.

Якщо частинка має кінетичну енергію $W_к$, то замінюючи її імпульс p через енергію $W_к$ за формулою $p = = \sqrt{2mW_к}$, можна записати

$$\lambda = \frac{h}{\sqrt{2mW_к}}.$$

Зокрема, для електрона, що прискорюється в електричному полі з різницею потенціалів U, маємо

$$W_к = \frac{m_ev^2}{2} = eU.$$

Тоді для електрона

$$\lambda = \frac{h}{\sqrt{2m_eeU}} = \frac{12{,}25}{\sqrt{U(B)}} \cdot 10^{-10}\text{ м} = = \frac{12{,}25}{\sqrt{U(B)}}\text{ Å}. \qquad (18.6)$$

Для електричних полів з напругою 1... 10^4 В довжини хвиль де Бройля належать інтервалу (0,1 … 10) Å, тобто мають довжини хвиль звичайних рентгенівських променів.

Цікавий результат дало порівняння гіпотези де Бройля і теорії атома Бора. Згідно з постулатом Бора $m_e v r = n\frac{h}{2\pi}$. Із цього рівняння можна визначити довжину стаціонарної орбіти:

$$2\pi r = n\frac{h}{m_e v},$$

де згідно з гіпотезою де Бройля $\frac{h}{m_e v} = \lambda$. Тоді довжина стаціонарної орбіти

$$2\pi r = n\lambda.$$

Отже, головне квантове число n показує, яке число довжин хвиль де Бройля електрона укладається на довжині боровської колової орбіти.

18.3. Експериментальне підтвердження гіпотези де Бройля

Для виявлення хвильових властивостей електронів, що рухаються, необхідно користуватись дифракційними ґратками, період яких за порядком величини дорівнює довжині хвилі де Бройля, тобто $d \sim 1$ Å. Такою ґраткою може бути кристалічна ґратка кристалів, як у випадку вивчення дифракції рентгенівських променів (див. підрозд. 13.8).

У 1925 р. в лабораторії фірми «Белл телефон» (США) К. Дж. Девіссон і К. Х. Кунсман провели експериментальні дослідження явища емісії вторинних електронів, які пізніше продовжили К. Дж. Девіссон і Л. Х. Джермер. Схему досліду показано на рис. 18.1. У цих експериментах спостерігалось явище «випаровування» вторинних електронів під дією пучка електронів, який направлявся на зразок із спресованого порошка нікелю.

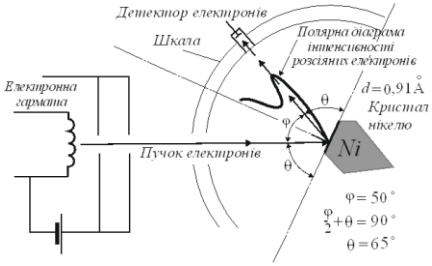


Рис. 18.1. Схема установки Девіссона і Джермера

Одного разу одна із пластинок нікеля окислилась і для видалення окисної плівки пластинку відпалили. Під час повторних дослідів з цією кристалічною мішенню результати різко відрізнялись від тих, які були отримані раніше. Після тривалого прогріву замість дуже дрібних кристаликів утворилось декілька великих монокристалів. Емісія вторинних електронів відбувалась, як і раніше, під будь-якими кутами, але за

певних кутів кількість електронів, які розсіювались, різко зростала.

В одному з дослідів пучок електронів з енергією $W_к$ = 54 еВ (U = 54 В) давав максимум в розподілі інтенсивності розсіяних електронів під кутом φ = 50° (рис. 18.1). За формулою (18.6) знайдемо довжину хвилі де Бройля:

$$\lambda = \frac{12{,}25}{\sqrt{54}}\ \text{Å} = 1{,}67\ \text{Å}.$$

Водночас, використовуючи формулу (13.33) для дифракції хвиль від площин у кристалі за методом Брегга (див. підрозд. 13.8), для максимуму першого порядку ($m = 1$) дифракції електронів від кристала нікеля з періодом ґратки $d = 0{,}91$ Å для $\theta = 65^{\circ}$, за формулою (13.33) отримаємо, що довжина хвилі де Бройля, яка дифрагує,

$$\lambda = 2d\sin\theta = 2 \cdot 0{,}91 \cdot \sin 65^{\circ}\ \text{Å} = 1{,}65\ \text{Å}.$$

Обидва результати добре збігаються, і це переконливо свідчить про те, що електрони проявляють як хвильові властивості, так і властивості частинки. Функцію $\lambda = f(U)$ (формула (18.6)) К. Девіссон передав на графіку як прямолінійну залежність у координатах $\lambda = f(1/\sqrt{U})$ (рис. 18.2), якою підтверджувалась передбачувана де Бройлем обернено пропорційна залежність довжини хвилі де Бройля матеріальної частинки від її швидкості.

Розбіжність між величинами, отриманими в дослідах К. Девіссона та Л. Джермера і підрахованими за теоретичною формулою де Бройля (18.6), не перевищувала прийнятних 2 %. Отже, узгодження теоретичної гіпотези та експерименту було переконливо засвідчено.

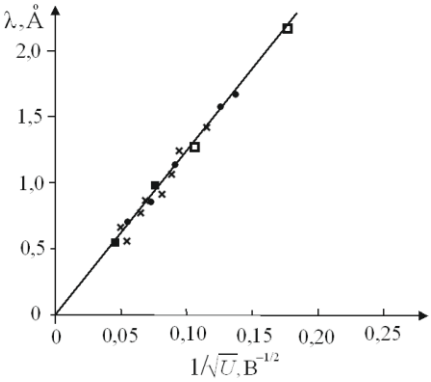


Рис. 18.2. Перевірка формули де Бройля $\lambda = h/mv$. На графіку в координатах λ та $1/\sqrt{U}$ зображено для електронів зв'язок їхньої дебройлевої довжини хвилі λ, обчисленої за даними дифракційної картини, з величиною різниці потенціалів U електричного поля.
(Clinton J. Davisson. The discovery of elektron waves. Nobel Lecture, December 13, 1937, Nobel Lectures, Physics 1922–1941. p. 387–394. 1965. Elsevier Publishing Company, Amsterdam)

У 1927 р. Дж. П. Томсон із своїми студентами здійснив досліди із

швидкими електронами й фактично отримав ті самі результати. Як і в експериментах з дифракції рентгенівських променів, Томсон досліджував розсіювання швидких електронів на дуже тонкій металевій фользі, яку виготовляли з полікристалічного матеріалу з хаотичним розміщенням кристалічних осей. Експериментальну установку схематично показано на рис. 18.3.

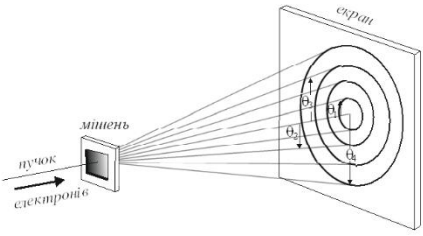


Рис. 18.3. Схема експериментальної установки Томсона

Пропускаючи паралельний пучок електронів (з енергією близько 10^4 еВ) через золоту фольгу завтовшки близько 10^{-5} см = 0,1 мкм, Томсон отримав на екрані ряд дифракційних кілець. Кути розсіювання $\theta_1, \theta_2, \theta_3$ і т.д. відповідали значенням, які отримують з відомого рівняння для дифракції світла, що проходить через ґратку, $d\sin\theta = m\lambda$ ($m = 1,\ 2,\ 3,\ \ldots$), де θ – кут між дифракційним пучком і напрямком падаючого пучка. Експериментальну перевірку хвильової природи частинок дослідники продовжували і в наступні роки. Опишемо деякі найбільш цікаві експерименти.

1. Німецький вчений Рупп у 1929 р. виміряв довжину хвилі де Бройля електронів, здійснивши дифракцію пучка електронів за допомогою простої оптичної дифракційної ґратки за дуже малих кутів ковзання.

2. У 1931 р. Джонсон продемонстрував дифракцію пучка молекул водню під час розсіювання у кристалі.

3. Естерман, Фріш і Штерн у 1938 р. здійснили експеримент з дифракції пучка атомів гелію. Як розсіювальний елемент вони застосували кристали фтористого літію.

4. Протягом тривалого часу залишалось нез'ясованим, чи хвильові властивості є окремо у кожної мікрочастинки, чи їх має лише потік багатьох мікрочастинок. У 1949 р.

В. Фабрикант, М. Сушкін, Л. Біберман провели дослід з дифракції потоку електронів настільки незначної інтенсивності, що інтервал часу між пролітанням наступних електронів був приблизно у 30 000 разів більшим від часу, який потрібний для проходження через увесь прилад, тобто реєструвалось

проходження окремих електронів. За досить тривалої експозиції отримували таку ж дифракційну картину, як і для інтенсивних пучків електронів. Цей дослід переконав, що хвильові властивості притаманні кожному електрону.

Наведені експерименти остаточно довели хвильову природу частинок, але водночас поставили ряд нових питань. Наприклад, оскільки фотони й електрони проявляють властивості і частинок, і хвилі, то може здаватись, що справді немає жодних відмінностей між частинками і хвилями. Утім, використовуючи тільки хвильову теорію, не вдається задовільно інтерпретувати такі експерименти, як комптонівське розсіяння, чи фотоефект, а також пояснити той факт, що частинки, наприклад *електрони*, *не можуть рухатись із швидкістю світла.*

Нільс Бор, намагаючись розв'язати протиріччя «хвиля – частинка», ввів ***принцип доповнюваності***. В основу його покладений той факт, що ані хвилі, ані частинки ніколи не проявляють *одночасно* в одному й тому самому експерименті і хвильові, й корпускулярні властивості. У кожному випадку – чи то яке-небудь випромінювання, чи потік електронів – для повного описання явища треба використовувати і хвильову, і корпускулярну моделі, але кожна з цих моделей має власну сферу застосування.

18.4. Співвідношення невизначеностей як прояв корпускулярно-хвильового дуалізму властивостей матерії

Розглянемо наслідки, що випливають з корпускулярно-хвильової двоїстості електронів, фотонів, протонів та інших об'єктів мікросвіту і які обмеження накладає цей дуалізм на класичний метод описування їх руху й, зокрема, на можливість одночасного визначення їх імпульсів і координат.

Будемо розглядати одновимірний випадок руху частинок. У класичній механіці будь-яка частинка рухається вздовж певної траєкторії так, що фіксовані її координати та імпульси. Сукупність координат x визначає траєкторію руху частинки. У кожній точці траєкторії частинка має конкретні значення координати x та відповідного імпульсу $p_x = m\frac{dx}{dt} = mv_x$. Якщо вважати, що для такої частинки справедливе співвідношення (18.5), то в кожній точці траєкторії буде відповідна довжина хвилі λ, тобто λ буде залежати від координати x. Оскільки довжина

хвилі є характеристикою хвильового процесу, то вона не може залежати від координати.

Таким чином, якщо для частинки відома координата x (тобто невизначеність $\Delta x = 0$) і для неї справедливе співвідношення (18.4), то імпульс для такої частинки невизначений, тобто $\Delta p_x = \infty$.

Мікрочастинки внаслідок наявності в них хвильових властивостей відрізняються від класичних частинок. Одна з основних відмінностей полягає в тому, що мікрочастинка не має чіткої траєкторії, тому неправомірно говорити одночасно про точні значення її координат та імпульсу.

Як відомо, будь-яка хвиля незалежно від природи є процесом, що заповнює більшу чи меншу область простору і через це не може локалізуватися в одній точці. Наприклад, уздовж осі *OX* рухається фотон, для якого точно відомий імпульс $p = p_x$, тобто $\Delta p_x \to 0$. Такому фотону відповідає хвиля, довжина якої однозначно визначається величиною імпульсу $\lambda = h/p$.

Така монохроматична хвиля у просторі нічим не обмежена, область її існування – уся вісь *OX*. Це означає, що в цьому випадку просторовий інтервал Δx, в якому замкнено об'єкт із хвильовими властивостями (фотон), дорівнює нескінченності. Інакше кажучи, якщо $\Delta p_x \to 0$, то $\Delta x \to \infty$. Отже, *якщо точно відомий імпульс фотона, то його локалізація стає цілком невизначеною.*

Розглянемо систему хвиль, суперпозиція яких дає огинальну, яка поширюється зі швидкістю, відмінною від швидкості хвиль, які додаються. Цю огинальну хвиль називають хвильовим пакетом (або групою хвиль). Хвильовий пакет, який побудований відповідною підбіркою великої кількості хвиль, можна інтерпретувати як матеріальну частинку. Як просте наближення розглянемо хвильовий пакет, побудований із двох хвиль, які описують рівняннями

$$\psi_1 = A\cos(\omega_1 t - k_1 x),$$

$$\psi_2 = A\cos(\omega_2 t - k_2 x),$$

де $\omega_1 \approx \omega_2 = \omega$ й $k_1 \approx k_2 = k$.

Результуюча хвиля

$$\psi = \psi_1 + \psi_2 =$$

$$= A_{\text{рез}} \cos\left(\frac{\omega_1 + \omega_2}{2} t - \frac{k_1 + k_2}{2} x\right)$$

з результуючою амплітудою

$$A_{\text{рез}} = 2A \cos\left(\frac{\omega_1 - \omega_2}{2} t - \frac{k_1 - k_2}{2} x\right).$$

На рис. 18.4 зображено дві хвилі, під час складання яких утворюється результуюча хвиля. Зазначимо, що в

результаті суперпозиції цих двох хвиль виникає биття, або хвильові групи. Кожна хвиля буде поширюватись із власною швидкістю (фазовою швидкістю), але групи хвиль, отримувані в результаті суперпозиції, переміщуються з іншою швидкістю (груповою швидкістю, див. підрозд. 10.5 і 15.2). Оскільки $\omega_1 \approx \omega_2$ й $k_1 \approx \approx k_2$, то фазові швидкості двох хвиль будуть приблизно однакові. Тоді

$$v_\text{ф} = \nu\lambda = \omega/k. \qquad (18.7)$$

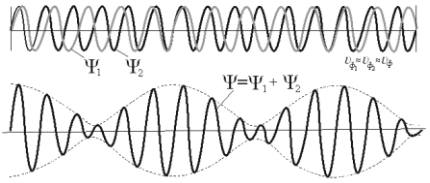


Рис. 18.4. Хвильовий пакет із двох хвиль

У співвідношенні (18.7) величини ν та λ можна асоціювати, відповідно, з частотою і довжиною хвилі де Бройля.

Гіпотеза про хвилі де Бройля набуває ще більш глибокого змісту, якщо врахувати, що швидкість частинки точно дорівнює груповій швидкості хвиль де Бройля. Швидкість частинки знаходять із співвідношення

$$v = dW/dp.$$

Групову швидкість згідно з формулою (10.28) (див. підрозд. 10.5) знаходять з формули $v_\text{гр} = \frac{d\omega}{dk}$. Це співвідношення можна переписати так:

$$v_\text{гр} = \frac{d(2\pi\nu)}{d\left(\frac{2\pi}{\lambda}\right)} = -\lambda^2 \frac{d\nu}{d\lambda}. \qquad (18.8)$$

Із співвідношення (18.5) випливає, що $d\lambda = -hdp/p^2$, а із співвідношення релятивістської механіки $W^2 = W_0^2 + p^2c^2$ знаходимо: $dW = = \frac{pc^2}{W}dp$. Із формули $\nu = W/h$ отримуємо, що $d\nu = = dW/h = \frac{pc^2}{Wh}dp$. Тоді

$$v_\text{гр} = \lambda^2 \frac{d\nu}{dp}\frac{p^2}{h} = \frac{pc^2}{W} = \frac{mvc^2}{mc^2} = v.$$

Групова швидкість, таким чином, дорівнює швидкості частинки, або, інакше кажучи, хвилі де Бройля переміщуються разом із частинкою. Вираз (10.30) (див. підрозд. 10.5) перепишемо так:

$$v_\text{гр} = v_\text{ф} - \lambda \frac{dv_\text{ф}}{d\lambda},$$

в якому групова швидкість пов'язана з фазовою швидкістю. Зауважимо, що якщо хвилі поширюються в середовищі без дисперсії, тобто $v_\text{ф} = \text{const}$, то $v_\text{гр} = = v_\text{ф}$. Щодо групової швидкості фотонів, то для фотона $W = pc$ й $v_\text{гр} = \frac{pc^2}{W} = c$.

Таким чином, ми дістали очікуваний результат – групова швидкість фотонів дорівнює швидкості самого фотона.

На рис. 18.5 зображено випадок, коли складається значна кількість хвиль: $\psi_1, \psi_2, \ldots, \psi_n$. Хвильовий пакет при цьому різко локалізується а його довжина Δx значно скорочується. Оскільки у хвильовому пакеті є набір значень λ, які містяться в деякому інтервалі $\Delta\lambda > 0$, то набір значень імпульсів *p,* як випливає із (18.4), перебуває в інтервалі

$$\Delta p = \Delta\left(\frac{h}{\lambda}\right) = \frac{h}{\lambda^2}|\Delta\lambda|.$$

Чим ширший інтервал можливих значень інтерферуючих хвиль $\Delta\lambda$, а разом з ним інтервал Δp_x можливих значень імпульсів, тим вужча область локалізації Δx результуючого пакета.

Інакше кажучи, чим більша невизначеність імпульсу фотона, тим точніше можна визначити його координати. За $\Delta p \to \infty$ маємо $\Delta x \to 0$, тобто фактичне значення координати фотона можливе тільки за повної невизначеності його імпульсу.

Такі висновки справедливі для фотонів, електронів, протонів та інших мікрочастинок.

Щоб усунути протиріччя «хвиля – частинка», німецький фізик Вернер Гайзенберг у 1927 р. висунув цікаву ідею, ввівши принцип невизначеності (або співвідношення невизначеностей).

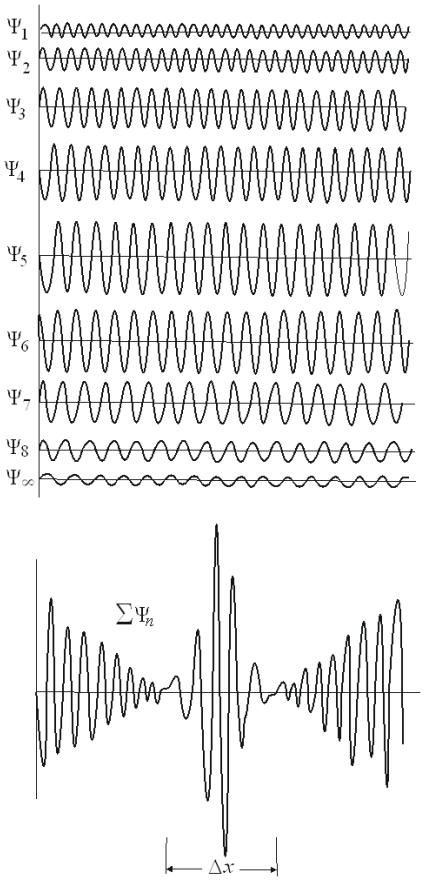


Рис. 18.5. Локалізація хвильового пакета

Цей принцип виражає фундаментальну межу можливості одночасного вимірювання певних пар змінних, наприклад положення частинки та її імпульсу. На рис. 18.6. зображено частинку, яка рухається зі швидкістю v. Частинка локалізована всередині

хвильового пакета, який переміщується із швидкістю $v_{\text{гр}} = v$.

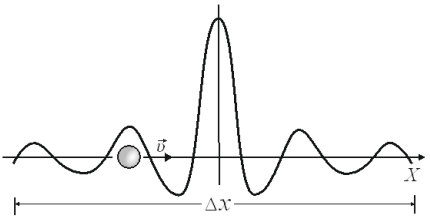


Рис. 18.6

Співвідношення (18.8) можна записати у такій формі:

$$v_{\text{гр}} = \frac{d\omega}{dk} = \frac{d\nu}{d\left(\frac{1}{\lambda}\right)} = \frac{h d\nu}{d\left(\frac{h}{\lambda}\right)} = h\frac{d\nu}{dp}. \quad (18.9)$$

Із рис. 18.6 і співвідношення (18.9) випливає, що

$$v_{\text{гр}} = \frac{\Delta x}{\Delta t} = h\frac{\Delta \nu}{\Delta p}$$

або

$$\Delta p \cdot \Delta x = h \cdot \Delta\nu \cdot \Delta t. \quad (18.10)$$

Якщо потрібно виміряти частоту хвилі, то мінімальним часом, потрібним для вимірювання, буде такий інтервал часу Δt, щоб одна повна довжина хвилі пройшла фіксовану точку. Цей інтервал часу, який повязаний з частотою співвідношенням $\Delta t \geq 1/\Delta\nu$, є час, що відповідає одному повному періоду.

Враховуючи тепер, що $\Delta\nu \cdot \Delta t \geq \geq 1$, із співвідношення (18.10) отримаємо нерівність

$$\Delta p \cdot \Delta x \geq h, \quad (18.11)$$

яка є одним із співвідношень невизначеностей Гайзенберга. Зміст отриманої фундаментальної межі ($\geq h$) не тільки в тому, що є деяка межа для точності вимірювання координати x та імпульсу p, а і в тому, що зі зменшенням області локалізації частинки Δx у просторі збільшується невизначеність у вимірюванні її імпульсу, і навпаки.

Неточність Δp у визначенні імпульсу у співвідношенні (18.11) є неточністю у визначенні проєкції імпульсу на напрямок осі x і більш правильним буде запис

$$\Delta p_x \cdot \Delta x \geq h. \quad (18.12)$$

Такі пари величин (наприклад, p_x та x), для яких невизначеність у значенні однієї величини залежить від точності вимірювання іншої, називають спряженими величинами. Зауважимо, що при русі частинки по осі X, не має жодних обмежень на точність вимірювання координат та проєкцій імпульсу на осі Y та Z (Δp_y та Δp_z).

Співвідношення (18.12) може бути встановлене з розгляду, зокрема, такого прикладу.

Спробуємо визначити значення координати x мікрочастинки, яка вільно летить, поставивши на її шляху щілину, ширина якої Δx і розміщена перпендикулярно до напрямку руху

частинки (рис. 18.7). До проходження частинки через щілину проєкція її імпульсу p_x на вісь X має точне значення, яке дорівнює нулю (щілина за умовою перпендикулярна до імпульсу), так що $\Delta p_x = 0$, зате координата x частинки є зовсім невизначеною. У момент проходження частинкою щілини ситуація змінюється. Замість повної невизначеності координати x виникає невизначеність Δx, але це досягається ціною втрати визначеності значення p_x. Справді, унаслідок дифракції є деяка ймовірність того, що частинка буде рухатись у межах кута 2φ, де φ – кут, що відповідає першому дифракційному мінімуму (максимумами вищих порядків можна знехтувати, оскільки їх інтенсивність мала порівняно з інтенсивністю центрального максимуму). Таким чином, виникає невизначеність $\Delta p_x = p \operatorname{tg} \varphi \approx p\varphi$.

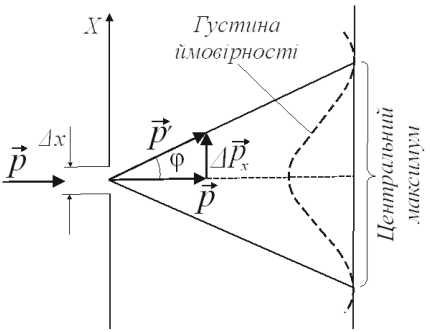


Рис. 18.7. Дифракція хвиль де Бройля на щілині

Краю центрального дифракційного максимуму (першому мінімуму), який отримують від щілини завширшки Δx, відповідає кут φ, для якого згідно з формулою (13.16)

$$\sin \varphi \approx \varphi = \frac{\lambda}{\Delta x}.$$

Отже,

$$\Delta p_x \sim \frac{p\lambda}{\Delta x}.$$

Звідси з урахуванням (18.5) отримуємо співвідношення

$$\Delta x \cdot \Delta p_x \sim p\lambda = h. \qquad (18.13)$$

Отримання співвідношення (18.12) чи (18.13) не є строгими, вони є приблизними. Найбільш строгі співвідношення невизначеностей отримують за допомогою методів квантової механіки і записують для одновимірного випадку так:

$$\langle \Delta x^2 \rangle \cdot \langle \Delta p_x^2 \rangle \geq \frac{\hbar^2}{4}, \qquad (18.14)$$

де $\langle \Delta p_x^2 \rangle$ та $\langle \Delta x^2 \rangle$ – середні квадратичні відхилення (невизначеності) імпульсу p_x і координати x, $(\hbar = h/2\pi)$. Формулу (18.14) часто записують у такому вигляді:

$$\Delta x \cdot \Delta p_x \geq \frac{\hbar}{2}, \qquad (18.15)$$

а для розв'язку задач часто використовують такий запис:

$$\Delta x \cdot \Delta p_x \geq \hbar. \qquad (18.16)$$

Неможливість одночасно точно визначити координату і відповідну

складову імпульсу не пов'язана з недосконалістю методів вимірювання і вимірювальних приладів, а є наслідком подвійної корпускулярно-хвильової природи мікрооб'єктів. Співвідношення невизначеностей отримано з одночасним використанням класичних характеристик руху частинки (координати, імпульсу) і наявності у неї хвильових властивостей.

Щоб глибше зрозуміти суть принципу невизначеності, можна подумки здійснити експеримент, який називають «мікроскопом Гайзенберга» (рис. 18.8). За допомогою уявного мікроскопа з високою роздільною здатністю (рис. 18.8) спробуємо визначити одночасно місце розміщення та імпульс електрона. З геометричної оптики відомо, що за такої роздільної здатності мікроскопа місцеположення об'єкта може бути визначене з точністю

$$\Delta x \approx \frac{\lambda}{2\sin\theta}, \qquad (18.17)$$

де θ – кут, показаний на рис. 18.8.

Коли відстань між двома точками менша за Δx, вони зіллються для спостерігача в одну точку. Отже, величина Δx є мінімальною неточністю у визначенні положення електрона, який рухається, координату й імпульс якого намагаємось визначити за допомогою мікроскопа. Із (18.17) випливає, що чим менша довжина хвилі λ світла, яке падає й освітлює об'єкт, тим менше Δx, й отже, з більшою точністю може бути зафіксоване положення електрона.

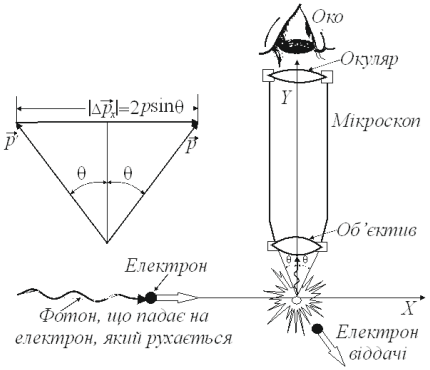


Рис. 18.8. «Мікроскоп Гайзенберга»

Фотон, що падає, буде взаємодіяти з електроном через механізм ефекта Комптона. Щоб побачити електрон, розсіяний фотон має потрапити в мікроскоп у межах кута 2θ. Таким чином, проєкція імпульсу фотона на напрямок x (див. рис. 18.8) може бути визначена з деякою мірою невизначеності:

$$\Delta p_x \approx 2p\sin\theta, \qquad (18.18)$$

де $p = h/\lambda$ – імпульс фотона, який падає.

Відповідно до закону збереження імпульсу співвідношення (18.18) буде також характеризувати мінімальну неточність у визначенні імпульсу електрона віддачі. Якщо перемножити (18.17) й (18.18) і врахувати, що $p = h/\lambda$, то для електрона віддачі будемо мати

$$\Delta x \cdot \Delta p_x \geq (2p \sin\theta)\left(\frac{\lambda}{2\sin\theta}\right) = h,$$

тобто отримуємо співвідношення невизначеностей.

Приклад. Нехай неточність у визначені положення молекули водню з масою близько $2 \cdot 10^{-27}$ кг, становить величину, співвідносну її діаметру, тобто десь 10^{-10}м. Тоді неточність у визначенні імпульсу

$$\Delta p_x \geq \frac{\hbar}{\Delta x} =$$
$$= \frac{1{,}054 \cdot 10^{-34}}{10^{-10}}\,\text{кг} \cdot \text{м/с} \approx 10^{-24}\,\text{кг} \cdot \text{м/с}.$$

Імпульс такої молекули у разі руху її зі швидкістю 2000 м/с (швидкість теплового руху за кімнатної температури)

$$p_x = mv = 2 \cdot 10^{-27} \cdot 2 \cdot 10^{3}\,\text{кг} \cdot \text{м/с} =$$
$$= 4 \cdot 10^{-24}\,\text{кг} \cdot \text{м/с},$$

а відносна неточність імпульсу

$$\frac{\Delta p_x}{p_x} = \frac{10^{-24}}{4 \cdot 10^{-24}} = 0{,}25 = 25\,\%.$$

Таким чином, імпульс цієї молекули не може бути визначений з точністю більшою, ніж 25 % від початкового значення. Якщо, однак, розглянути приклад із кулею, маса якої 50 г, швидкість польоту якої дорівнює 1000 м/с і положення відоме з точністю 1 мм, то неточність визначення імпульсу

$$\Delta p_x \geq \frac{1{,}054 \cdot 10^{-34}}{10^{-3}}\,\text{кг} \cdot \text{м/с} \cong$$
$$\cong 10^{-31}\,\text{кг} \cdot \text{м/с}.$$

Оскільки імпульс

$$p_x = 0{,}05 \cdot 1000\,\text{кг} \cdot \text{м/с} =$$
$$= 50\,\text{кг} \cdot \text{м/с},$$

то відносна неточність імпульсу

$$\frac{\Delta p_x}{p_x} \cong \frac{10^{-31}}{50} = 2 \cdot 10^{-33}.$$

Це число настільки мале, що жодний з відомих лабораторних приладів не може його зафіксувати.

Інше співвідношення невизначеностей встановлює зв'язок між енергією і часом й може бути отримане із співвідношення $\Delta\nu \cdot \Delta t \geq 1$ або $\Delta\left(\frac{W}{h}\right)\Delta t \geq 1$. Звідси

$$\Delta W \cdot \Delta t \geq h. \qquad (18.19)$$

Так само як імпульс і координата не можуть бути визначені однозначно з нескінченно високою точністю, так і енергія W й час t (які теж являють собою пару спряжених змінних) не можуть бути одночасно знайдені з необмеженою точністю. Чим більшої точності досягнуто нами під час вимірюванні часу (тобто чим менше Δt), тим менша точність у визначенні енергії. Співвідношення (18.19) записують у такому вигляді:

$$\Delta W \cdot \Delta t \geq \hbar. \qquad (18.20)$$

Співвідношення невизначеностей є одним з фундаментальних положень квантової механіки. Вони є математичним виразом корпускулярно-

хвильового дуалізму частинок, а останнє розглядають як експериментальний факт.

Істинний зміст співвідношень невизначеностей полягає в тому, що у природі об'єктивно не існує станів частинок з точно визначеними значеннями координат та імпульсів, а отже, втрачає зміст поняття траєкторії.

Співвідношення Гайзенберга обмежують застосування понять класичної фізики до об'єктів з мікросвіту. У класичній механіці, щоб повністю визначити рух тіла протягом всього часу, треба задати його положення (координату x) і швидкість (v_x) в який-небудь початковий момент часу. Згідно із співвідношенням невизначеностей зробити це неможливо. Положення і швидкість (імпульс) не можуть бути визначені одночасно. Доводиться говорити тільки про ймовірність, з якою частинка перебуває в тому чи іншому місці. Як писав Гайзенберг, «*ця невизначеність являє собою справжню основу для появи статистичних зв'язків у квантовій механіці*».

18.5. Статистичне тлумачення хвиль де Бройля

Фізичну суть хвиль де Бройля, які пов'язані з рухом частинки, було розкрито не одразу. Спочатку були спроби розглядати самі частинки як утворення із хвиль, що перебувають у деякій області простору. Інтенсивність хвиль де Бройля було розглянуто у цій концепції як величину, що характеризує густину середовища, з якого утворена частинка. Таке розуміння хвиль де Бройля мало класичний характер. Основою такого розуміння було те, що групова швидкість пакета хвиль збігалася зі швидкістю частинки за законами класичної механіки. Утім, з часом група хвиль розпливається, що пов'язано з дисперсією. Окремі хвилі, з яких утворена група, поширюються з різними швидкостями, що і є причиною розпливання.

Таким чином, побудована із хвиль де Бройля частинка буде нестійкою. Ця нестійкість чітко проявиться, коли частинка рухається в неоднорідному просторі, переходячи з одного середовища в інше. Таким прикладом є класичні досліди з дифракції частинок. Коли пучок електронів проходить крізь тонку фольгу, то він розділяється на систему

конусоподібних дифракційних пучків (рис. 18.3). Якщо розглядати електрон як утворення із хвиль, то кожний дифракційний пучок має представляти деяку частину електрона. Якщо в один реєстраційний прилад направити тільки перший дифракційний пучок, а у другий прилад – тільки другий дифракційний пучок, то це означає, що кожний із приладів прийме тільки частину електрона. Це і є крайнім порушенням атомізму частинок. Справді, частинка завжди діє як ціле й у приладі проявляється вся частинка, а не її частина, тому уявлення про частинки як утворення із хвиль де Бройля протирічить атомізму й має бути відкинутим.

Не можна вважати, що самі хвилі є утворенням із частинок або, точніше кажучи, виникають в середовищі, яке утворене частинками. Дослід показує, що дифракційна картина, що виникає на фотопластині, не залежить від інтенсивності пучка частинок, які падають, а отже, і від густини частинок в одиниці об'єму (див. підрозд. 18.3). Щоб отримати одну й ту саму дифракційну картину, можна зменшити інтенсивність, але збільшити експозицію: важлива тільки загальна кількість частинок, які пройшли. Цей факт однозначно показує, що кожний електрон дифрагує незалежно від інших, тому існування хвильових явищ не можна пов'язувати з наявністю одночасно великої кількості частинок. Варто підкреслити, що хвильові явища в русі мікрочастинок проявляються і в таких випадках, коли говорити про середовище, що утворене сукупністю частинок, зовсім не має змісту. Справді, хвильові властивості електрона проявляються в атомах, де кількість їх кількість зовсім невелика (один електрон у атома водню, два в гелію і т.д.).

Правильне тлумачення хвиль де Бройля знайшов М. Борн, який йшов у науковій роботі зовсім іншим шляхом. Щоб зрозуміти основну думку Борна, уявімо, що ми досліджуємо дифракцію електронів і реєструємо потрапляння «дифракційних» електронів на фотопластині. Кожний з електронів, пройшовши через дифракційний прилад (наприклад, крізь фольгу), проявиться в якому-небудь місці фотопластини і викличе на ній фотохімічну дію. Проходження невеликої кількості електронів дасть на фотопластині картину, подібну на мішень, простріляну поганим стрілком. Тільки за великої кількості електронів, які пройшли крізь фотопластину, виявиться регулярність у

розподілі електронів на ній, і нарешті утвориться розподіл, який відповідає розподілу інтенсивності під час дифракції хвиль (див. підрозд. 13.4).

З такої поведінки частинок Борн вивів статистичне тлумачення хвиль де Бройля, яке дозволило поєднувати атомізм частинок із хвильовими явищами: *інтенсивність хвиль де Бройля в якому-небудь місці простору пропорційна ймовірності виявити частинку в цьому місці.* Так, наприклад, якщо один дифракційний пучок направити на одну фотопластину, а другий на іншу, то за великої кількості електронів, що пройшли через апарат, кількість електронів, які потрапляють на кожну із пластинок, пропорційна інтенсивності хвиль де Бройля, що поширюються в напрямку кожної із фотопластин.

Якщо фотопластину розмістити так, щоб напрямок від дифракційного апарата до пластини збігався з напрямком дифракційного мінімуму (у цьому напрямку хвилі гасять одна одну), то частинки зовсім не будуть потрапляти на таку фотопластину. Якщо мова йде не про велику кількість електронів, а про один, то інтенсивність хвиль де Бройля вказує тільки на ймовірність потрапляння електрона, але зовсім не зобов'язує електрон до тієї чи іншої поведінки.

У такому розумінні хвилі де Бройля не мають нічого спільного із хвилями, які розглядаються у класичній фізиці. У всіх «класичних» хвилях абсолютне значення амплітуди хвилі визначає фізичний стан. Якщо, наприклад, амплітуда коливань повітря в одному випадку скрізь у два рази більша, ніж в іншому, то це означає вчетверо більшу енергію коливань у першому випадку і, разом з тим, інший фізичний стан середовища.

У випадку хвиль де Бройля їх інтенсивність визначає густину ймовірності місцезнаходження частинки, *тому важливе тільки відношення інтенсивностей в різних частинах простору, а не сама їхня абсолютна величина.* Це відношення показує, у скільки разів в одному місці простору ймовірніше виявити частинку, ніж в іншому.

Хвилі де Бройля дають, таким чином, статистичний опис руху мікрочастинки. Відповідно до цього ці хвилі треба розглядати *як хвилі ймовірності: вони визначають густину ймовірності виявити частинку в певному місці простору в цей момент часу*.

Знаючи хвилю де Бройля, яка описує стан частинки, можна знайти ймовірність не тільки місцезнаходження частинки, але й імовірність будь-якого результату вимірювання будь-якої механічної величини, яка належить до частинки, що вивчається.

18.6. Ймовірність місцезнаходження мікрочастинки

Позначимо через x, y, z координати частинки. Згідно з викладеним у підрозд. 18.5 в точний зміст x, y, z вкладене таке: величини x, y, z визначають як координати тієї точки, в якій виявляється частинка. Так, наприклад, це будуть координати плямочки на фотопластині, отриманої в результаті потрапляння на пластину частинки, або, наприклад, координати, які визначають положення щілини, через яку пройшла частинка, і т.п.

Сформулюємо тепер математично статистичну інтерпретацію хвиль де Бройля. Зауважимо, насамперед, що слово «хвиля» ми вживаємо тут досить умовно. Тільки у дуже спеціальних випадках стан частинки буде описуватись простими плоскими хвилями. У загальному випадку те, що ми зараз називаємо хвилями де Бройля, може бути досить складною функцією координат частинки x, y, z і часу t. Тим не менше і для цих складних випадків використовують термін *хвильова функція*, яку позначають літерою ψ:

$$\psi = \psi(x, y, z, t). \qquad (18.21)$$

Як було пояснено у підрозд. 18.5, на основі дослідних *фактів ми вважатимемо, що густина ймовірності місцезнаходження частинки визначається інтенсивністю хвиль,* тобто квадратом амплітуди ψ. Враховуючи, що ψ є *комплексною функцією* (хвилі ймовірності невидимі), а ймовірність має бути завжди справді позитивною величиною, будемо брати за міру інтенсивності не ψ^2, а квадрат модуля ψ, тобто величину

$$|\psi|^2 = \psi\psi^*,$$

де через ψ^* позначено величину, яка комплексно спряжена до ψ.

Ймовірність знайти частинку поблизу точки x, y, z залежить, зазвичай, від розмірів області, яку вибираємо. Розглядаючи нескінченно малу область $x, x + dx;\ y, y + dy;\ z, z + dz$, можемо вважати ψ всередині неї постійною, а тому ймовірність знайти частинку пропорційна об'єму цієї області. Позначимо цей елемент об'єму через $dV = dxdydz$.

Позначаючи ймовірність (нескінченно малу) знайти частинку в

елементі об'єму dV біля точки x, y, z в момент часу t через $dP(x,y,z,t)$, можемо записати статистичне тлумачення хвиль де Бройля у вигляді такої рівності:

$$dP(x,y,z,t) = |\psi(x,y,z,t)|^2 dV. \quad (18.22)$$

Ця рівність дозволяє за відомою хвильовою функцією $\psi(x,y,z,t)$ вирахувати ймовірність місцезнаходження частинки $dP(x,y,z,t)$ в об'ємі dV. Величину

$$\frac{dP}{dV} = |\psi(x,y,z,t)|^2 \quad (18.23)$$

називають густиною ймовірності.

Ймовірність знайти частинку в момент часу t в об'ємі V згідно з теоремою додавання ймовірностей

$$P(V,t) = \int_V dP = \int_V |\psi(x,y,z,t)|^2 dV. \quad (18.24)$$

Якщо виконати інтегрування за всім об'ємом, то отримаємо ймовірність того, що в момент часу t частинка перебуває де-небудь – це ймовірність вірогідної події. В теорії ймовірностей ймовірність вірогідної події беруть за 1. Якщо цього дотримуватись, то інтеграл від $|\psi|^2$ по всьому об'єму слід прирівняти одиниці:

$$\int_\infty |\psi(x,y,z,t)|^2 dV = 1. \quad (18.25)$$

Цю умову називають *нормуванням*, а функція ψ, яка задовольняє цю умову, називають нормованою.

Щоб хвильова функція була об'єктивною характеристикою стану мікрочастинок, вона має задовольняти такі обмежувальні умови:

1) функція ψ, що характеризує ймовірність виявлення мікрочастинки в елементі об'єму, має бути ***скінченною*** (ймовірність не може бути більшою від одиниці), ***однозначною*** (ймовірність не може бути неоднозначною величиною) і ***неперервною*** (ймовірність не може змінюватись стрибком);

2) похідні $\frac{d\psi}{dx}, \frac{d\psi}{dy}, \frac{d\psi}{dz}, \frac{d\psi}{dt}$ мають бути безперервні;

3) функція $|\psi|^2$ має бути інтегрованою;

4) функція ψ не повинна протирічити таким співвідношенням:

$$\lambda = \frac{h}{p}, \nu = \frac{W}{h}, W = \frac{p^2}{2m} + U;$$

5) якщо $x \to \pm\infty$, $y \to \pm\infty$, $z \to \pm\infty$, то $\psi(x,y,z,t)$ має прямувати до нуля.

Хвильова функція задовільняє принцип ***суперпозиції***: якщо система може перебувати у різних станах, що описуються хвильовими функціями $\psi_1, \psi_2, \ldots, \psi_n$, то вона також може перебувати у стані ψ, який описують лінійною комбінацією цих функцій:

$$\Psi = \sum_n C_n \psi_n,$$

де C_n – деякі комплексні числа.

За допомогою хвильової функції у квантовій механіці можна обчислити середні значення фізичних величин, які характеризують певний об'єкт, що перебуває у стані, який описують хвильовою функцією ψ.

18.7. Рівняння Шредінгера. Принцип причинності у квантовій механіці

Зі статистичного трактування хвиль де Бройля і співвідношення невизначеностей Гайзенберга було зроблено висновок, що рівняння руху мікрочастинок в різних силових полях має бути рівнянням, з якого випливали би хвильові властивості частинок, які спостерігаються під час досліду. Основне рівняння має бути диференціальним рівнянням відносно хвильової функції $\psi(x, y, z, t)$, оскільки величина $|\psi|^2$ визначає густину ймовірності перебування частинок у момент часу t в об'ємі dV. Оскільки шукане рівняння повинно враховувати хвильові властивості частинок, то воно має бути хвильовим рівнянням.

Основне рівняння нерелятивістської квантової механіки, яке сформулював у 1926 р. Е. Шредінгер для одновимірного нерелятивістського випадку, має такий вигляд:

$$-\frac{\hbar^2}{2m}\frac{\partial^2\psi(x,t)}{\partial x^2} + U(x,t)\psi(x,t) = i\hbar\frac{\partial\psi(x,t)}{\partial t}, \qquad (18.26)$$

де m – маса мікрочастинки; $U(x,t)$ – її потенціальна енергія в силовому полі, в якому вона рухається; $i = \sqrt{-1}$ – уявна одиниця. Надалі будемо розглядати одновимірний випадок. Це не зменшує значення тих узагальнень, які випливають з рівняння (18.26).

Попри те, що Шредінгер під час отримання свого рівняння користувався класичними уявленнями про «хвилі матерії», це рівняння не можна вивести безпосередньо з фундаментальних законів класичної фізики. Рівняння Шредінгера подібно, наприклад, до другого закону Ньютона, саме по собі є фундаментальним співвідношенням. Коли потенціальна енергія залежить від часу і координат, то рівняння (18.26) відоме як загальне (або часове) рівняння Шредінгера. Розв'язком цього рівняння і є хвильова функція $\psi(x,t)$, яка служить корисним інструментом, засобом для описання фізичних явищ.

З погляду математики рівняння Шредінгера є лінійним диференціальним рівнянням з частинними похідними. З теорії диференційних рівнянь відомо, що кожне лінійне рівняння з частинними похідними має багато розв'язків,

причому таких, що лінійна комбінація будь-якої сукупності розв'язків рівняння також буде розв'язком рівняння.

Для значної кількості фізичних явищ, які відбуваються в мікросвіті, наприклад для опису поведінки електрона в атомі, у деяких випадках важливо вміти знаходити стаціонарні розв'язки рівняння Шредінгера, які не містять часу.

Щоб розв'язати цю задачу, треба знайти стаціонарне рівняння Шредінгера, в якому виключено залежність ψ від часу. Воно має сенс для тих задач, в яких потенціальна енергія U не залежить від часу: $U = U(x)$. Шукатимемо розв'язок рівняння Шредінгера у вигляді добутку:

$$\psi(x,t) = \psi(x)e^{-i\frac{W}{\hbar}t}, \quad (18.27)$$

де $\psi(x)$ є функцією лише координат, а W – повна енергія частинки. Це означає, що стан з певним значенням енергії W (невизначеність $\langle \Delta W^2 \rangle = 0$) гармонічно залежить від часу з частотою $\omega = \frac{W}{\hbar}$.

Підставимо вираз для $\psi(x,t)$ (18.27) у рівняння Шредінгера (18.26):

$$-\frac{\hbar^2}{2m}\ddot{\psi}e^{-i\frac{W}{\hbar}t} + U(x)\psi e^{-i\frac{W}{\hbar}t} = \\ = i^2\hbar\psi\cdot\left(-\frac{W}{\hbar}\right)e^{-i\frac{W}{\hbar}t},$$

де $\ddot{\psi} = \frac{\partial^2\psi}{\partial x^2}$. Скоротивши на множник $e^{-i\frac{W}{\hbar}t}$, отримаємо:

$$\ddot{\psi}(x) + \frac{2m}{\hbar^2}\big(W - U(x)\big)\psi(x) = 0. \quad (18.28)$$

Це і є ***стаціонарне рівняння Шредінгера***. Рівняння (18.28) є найважливішим співвідношенням нерелятивістської квантової механіки. Функції ψ, які задовольняють рівняння Шредінгера за певного W, називають ***власними функціями***. У рівняння Шредінгера як параметр входить повна енергія частинки. В теорії диференціальних рівнянь доводиться, що подібні рівняння мають розв'язок не за довільних значень параметра, а лише за певних значень W. Ці значення ***енергії*** називають ***власними***.

Власні значення W можуть утворювати як безперервний, так і дискретний ряд.

Рівняння Шредінгера дає змогу знайти не тільки конкретний вигляд функції $\psi(x,t)$ у заданому зовнішньому полі $U(x,t)$, а й визначити її зміну з часом $\frac{\partial\psi}{\partial t}$. Знання ж ψ – функції дає змогу знайти усі можливі значення важливих параметрів фізичної системи у будь-який момент часу.

Отже, *рівняння Шредінгера є математичним виразом* ***принципу причинності*** *у квантовій механіці – початковий стан системи (за*

відомих взаємодій у системі) повністю визначає всю наступну її еволюцію.

Утім, на відміну від класичної фізики, квантова механіка не дає чіткої відповіді на запитання, якими є точні значення параметрів фізичної системи у цей момент часу. ψ – функція, вказуючи на цілий спектр можливих значень параметрів системи, дає змогу обчислити лише ймовірність появи кожного значення під час вимірювання.

18.8. Рух вільної частинки. Частинка у прямокутній потенціальній ямі

Розглянемо декілька порівняно простих прикладів застосування рівняння Шредінгера до руху частинки у конкретних умовах.

18.8.1. Рух вільної частинки

Під час руху вільної частинки $(U(x) = 0)$ повна енергія збігається з кінетичною $W = W_к$. Для вільної частинки, що рухається вздовж осі *OX*, $(-\infty < x < +\infty)$, рівняння Шредінгера для стаціонарних станів набуває вигляду

$$\ddot{\Psi} + \frac{2m}{\hbar^2} W\Psi = 0. \qquad (18.29)$$

Прямою підстановкою можна переконатися в тому, що частинним розв'язком рівняння є функція

$$\Psi(x) = Ae^{ikx} + Be^{-ikx},$$

де $k = \frac{2\pi}{\lambda}$ – хвильове число, довжина хвилі де Бройля $\lambda = \frac{h}{p} = \frac{h}{\sqrt{2mW}}$, *A* та *B* – сталі. Розв'язок рівняння (18.29) можна переписати ще й так:

$$\Psi(x) = Ae^{\frac{i\sqrt{2mW}x}{\hbar}} + Be^{-\frac{i\sqrt{2mW}x}{\hbar}}.$$

Функція $\Psi(x)$ є тільки координатною частиною хвильової функції $\Psi(x,t)$ стаціонарного стану. Розв'язок повного рівняння Шредінгера отримаємо у формі (18.27):

$$\Psi(x,t) = Ae^{-\frac{i(Wt-\sqrt{2mW}x)}{\hbar}} + \\ + Be^{-\frac{i(Wt+\sqrt{2mW}x)}{\hbar}}. \qquad (18.30)$$

Цей розв'язок є суперпозицією двох плоских монохроматичних хвиль однакової частоти $\omega = \frac{W}{\hbar}$. Хвиля з амплітудою *A* поширюється в додатному напрямку осі *OX*, хвиля з амплітудою *B* поширюється у протилежному напрямку. Знайдений розв'язок узгоджується із загальним виразом плоскої монохроматичної хвилі $S(x,t) = Ae^{-i(\omega t - kx)}$. Хвильове число *k* для вільної частинки дорівнює $k = \frac{1}{\hbar}\sqrt{2mW}$, а $\omega = \frac{W}{\hbar}$.

Отже, вільній частинці у квантовій механіці відповідає плоска монохроматична хвиля де Бройля. Вона характеризує ймовірність знаходження частинки в певній точці простору.

Справді, взявши лише одну із хвиль, що поширюється в додатному напрямку осі *OX*, отримаємо

$$|\psi|^2 = \psi\psi^* = |A|^2.$$

Власні значення енергії частинки

$$W = \frac{p_x^2}{2m} = \frac{h^2}{2m\lambda^2} = \frac{\hbar^2 k^2}{2m}.$$

Енергія вільної частинки може набувати довільного значення, тобто її енергетичний спектр неперервний.

18.8.2. Частинка в одновимірній прямокутній потенціальній ямі

Нехай частинка може рухатись лише вздовж осі *OX* і перебуває всередині прямокутної потенціальної ями, яка обмежена нескінченно високими бар'єрами (рис. 18.9). У цьому випадку потенціальна енергія частинки $U(x)$ набуватиме таких значень:

$$U(x) = \begin{cases} \infty, & x < 0; \\ 0, & 0 \le x \le l; \\ \infty, & x > l, \end{cases}$$

де l – ширина ями, а енергію відраховують від дна ями. Рівняння Шредінгера (18.28) у випадку одновимірної ями запишемо у вигляді

$$\ddot{\Psi} + \frac{2m}{\hbar^2}(W - U)\Psi = 0. \quad (18.31)$$

За умовою задачі (нескінченно високі «стінки») частинка не проникає за межі ями, тому ймовірність її виявлення за межами ями дорівнює нулю. На межах ями (за $x = 0$ та $x = l$) неперервна хвильова функція має перетворюватися в нуль. Отже, граничні умови мають вигляд

$$\psi(0) = \psi(l) = 0. \quad (18.32)$$

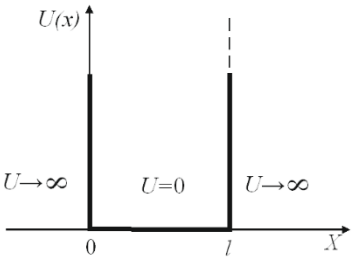


Рис. 18.9. Одновимірна прямокутна потенціальна яма

У межах ями ($0 \le x \le l$) рівняння Шредінгера (18.31) матиме вигляд

$$\ddot{\Psi} + k^2\Psi = 0, \quad (18.33)$$

де $k^2 = \frac{2m}{\hbar^2}W$.

Загальний розв'язок диференціального рівняння (18.33) будемо шукати у вигляді

$$\Psi(x) = Ae^{ikx} + Be^{-ikx}. \quad (18.34)$$

Функція (18.34) повинна задовільняти граничні умови (18.32), тому отримаємо систему рівнянь

$$A + B = 0,$$
$$Ae^{ikl} + Be^{-ikl} = 0. \quad (18.35)$$

Ця однорідна система двох рівнянь буде мати нетривіальний розв'язок відносно A і B тоді, коли дискримінант системи дорівнює нулю, тобто

$$\begin{vmatrix} 1 & 1 \\ e^{ikl} & e^{-ikl} \end{vmatrix} = 0,$$

або $e^{ikl} - e^{-ikl} = 0$. Врахувавши формулу Ейлера ($e^{i\varphi} = \cos\varphi + i\sin\varphi$), отримаємо:

$$\cos kl + i\sin kl - \\ -\cos(-kl) - i\sin(-kl) = 0.$$

Звідси $\sin kl = 0$ виконується лише за $kl = \pi n$, де n – цілі числа (випадок, коли $n = 0$, не розглядаємо, обгрунтування цього буде подано далі). Отже, необхідно, щоб

$$k = \frac{\pi}{l} \cdot n. \qquad (18.36)$$

Тоді $\frac{\pi^2 n^2}{l^2} = \frac{2m}{\hbar^2} W_n$, і власні значення енергії

$$W_n = \frac{\pi^2 \hbar^2}{2ml^2} n^2 = \frac{h^2}{8ml^2} n^2. \qquad (18.37)$$

Рівняння Шредінгера задовольняється лише за значень W_n, що залежать від цілого числа *n*.

Отже, енергія W_n частинки в потенціальній ямі з нескінченно високими стінками не може бути довільною, а набуває лише певних дискретних значень, тобто квантується.

Квантові значення енергії W_n називають **рівнями енергії**, а число *n*, яке визначає енергетичні рівні частинки, називають квантовим числом.

Умова (18.36) має простий фізичний зміст. Оскільки $k_n = \frac{2\pi}{\lambda_n}$, де λ_n – довжина хвилі де Бройля для електрона в ямі, то $\frac{2\pi}{\lambda_n} = \frac{n\pi}{l}$, або $\lambda_n = \frac{2l}{n}$ ($l = (\lambda_n/2)n$), тобто на ширині ями має вкладатись ціле число півхвиль де Бройля. Знайдемо власні хвильові функції. Із (18.35) видно, що $A = -B$, тому (18.34) перепишемо так:

$$\Psi(x) = A\cdot(e^{ikx} - e^{-ikx}) = 2iA\sin kx.$$

\Сталу A знайдемо з умови (18.25):

$$\int_0^l (2iA\sin kx)(-2iA^*\sin kx)\,dx = 1,$$

або $4|A|^2 \int_0^l \sin^2 kx dx = 1$ Врахувавши, що

$$\int_0^l \sin^2 kx dx = \\ = \int_0^l \left(\frac{1}{2} - \frac{1}{2}\cos 2kx\right) dx = \frac{l}{2},$$

отримаємо, що $A = \sqrt{\frac{1}{2l}}$. Тоді власні хвильові функції мають вигляд

$$\psi_n(x) = i\sqrt{\frac{2}{l}}\sin\frac{n\pi}{l}x, (n = 1, 2, 3 \ldots). \qquad (18.38)$$

На рис. 18.10, *а*, зображено графіки функції $\psi_n(x)$ за $n = 1, 2, 3$, а на рис. 18.10, *б*, – густину ймовірності знаходження частинки $|\psi_n(x)|^2$ на різних відстанях від «стінок» ями для $n = 1, 2, 3$.

Наприклад, у квантовому стані з $n = 2$ частинка не може перебувати посередині ями і водночас однаково

часто може перебувати в її лівій або правій частині. Стану з $n = 0$ не існує, оскільки $\psi_n(x) = 0$ за $n = 0$.

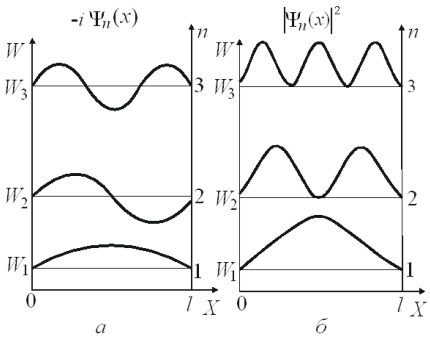


Рис. 18.10. Стани мікрочастинки у потенціальній ямі

Квантовий стан з найменшим значенням енергії (n = 1) називають основним станом. Йому відповідає енергія

$$W_1 = \frac{h^2}{8ml^2}.$$

Оцінимо різницю енергій двох сусідніх рівнів за формулою (18.37):

$$\Delta W_n = \Delta W_{n+1} - W_n = \\ = \frac{h^2}{8ml^2}(n+1)^2 - \frac{h^2}{8ml^2}n^2 = \\ = \frac{h^2}{8ml^2}(2n+1).$$

Для електронів за розмірів ями $l = 10^{-1}$м (вільні електрони в металі) та $m = m_e = 9{,}1 \cdot 10^{-31}$ кг

$$\Delta W_n = (2n+1) \cdot 6 \cdot 10^{-36}\text{Дж} = \\ = (2n+1) \cdot 3{,}8 \cdot 10^{-17}\text{ еВ}.$$

Енергетичні рівні розміщені так тісно, що спектр фактично можна вважати неперервним. Якщо область, в якій рухається електрон, співмірна з атомними розмірами ($l = 10^{-10}$м), то

$$\Delta W_n = (2n+1) \cdot 6 \cdot 10^{-18}\text{Дж} = \\ = (2n+1) \cdot 3{,}8 \cdot 10\text{еВ}.$$

Отже, дискретність енергетичних рівнів у цьому випадку буде досить помітною й отримуємо лінійчастий спектр.

За великих квантових числел ($n \gg 1$)

$$\frac{\Delta W_n}{W_n} = \frac{2n+1}{n^2} \approx \frac{2}{n} \ll 1,$$

тобто сусідні рівні розміщені тим густіше, чим більшим є n. Якщо n дуже велике, то можна говорити про фактичну неперервну послідовність рівнів, і характерна особливість квантових процесів, тобто дискретність, згладжується.

18.9. Розв'язок рівняння Шредінгера для потенціального бар'єра

18.9.1. Постановка проблеми

Якщо є дві області простору, в яких потенціальна енергія частинки менша, ніж на поверхні, яка розділяє ці області, то говорять, що області розділені *потенціальним бар'єром.* Найпростішим прикладом потенціального бар'єра може бути бар'єр в одному

вимірюванні, що зображений на рис. 18.11. По осі ординат відкладено потенціальну енергію $U(x)$ як функцію координати x. У точці x_0 потенціальна енергія має максимум U_0. Весь простір $-\infty < x < +\infty$ ділиться у цій точці на дві області: $x < x_0$ та $x > x_0$, в яких $U < U_0$. Значення терміна «потенціальний бар’єр» з’ясуємо, якщо розглянути рух частинки в полі $U(x)$ на основі класичної механіки. Повна енергія частинки W дорівнює

$$W = \frac{p^2}{2m} + U(x), \qquad (18.39)$$

де $-$ імпульс частинки, а $m -$ її маса. Розв’язуючи (18.39) відносно імпульсу, отримаємо

$$p(x) = \pm\sqrt{2m\big(W - U(x)\big)}. \quad (18.40)$$

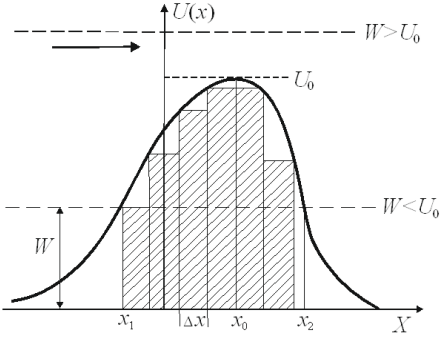


Рис. 18.11. Потенціальний бар’єр

Знаки «+» або «−») треба вибирати залежно від напрямку руху частинки. Якщо енергія частинки W більша за «висоту» бар’єра U_0, то вона безперешкодно пройде бар’єр зліва направо, якщо початковий імпульс $p > 0$, або у протилежному напрямку, якщо початковий імпульс $p < 0$.

Нехай частинка рухається зліва направо й енергія її W менша за U_0. Тоді у деякій точці x_1 потенціальна енергія $U(x_1) = W$, $p(x_1) = 0$, частинка зупиниться. Вся її енергія перетвориться в потенціальну, і рух почнеться в оберненому напрямку: x_1 є *точкою повороту*. За $x > x_1$ імпульс p стає уявною величиною, тому за $W < U_0$ частинка, яка рухається зліва, не пройде через область максимуму потенціалу ($x = x_0$) й не проникне в іншу область, де $x > x_0$. Так само, якщо частинка рухається справа наліво з енергією $W < U_0$, то вона не проникне в область за другою точкою повороту x_2, в якій $U(x_2) = W$ (див. рис. 18.11). Таким чином, потенціальний бар’єр є «непрозорою» перегородкою для всіх частинок, енергія яких менша за U_0 (навпаки, він «прозорий» для частинок з енергією $W > U_0$). Цим і пояснюється назва «потенціальний бар’єр».

Зовсім по-іншому протікають явища біля потенціальних бар’єрів, якщо мова йде про рух мікроскопічних частинок у мікроскопічних полях, тобто про рухи, під час розгляду яких не

можна ігнорувати квантових ефектів (хвильових властивостей частинок). У цьому випадку, як побачимо далі, на противагу висновкам класичної механіки, частинки з енергією W, яка більша від висоти бар’єра U_0, частково відбиваються від бар’єра, а частинки з енергією, яка менша за U_0, частково проникають крізь бар’єр.

18.9.2. Розв’язок рівняння Шредінгера для потенціального бар’єра нескінченої ширини

Розглянемо поведінку частинки на межі двох областей I і II, у кожній з яких потенціальна енергія частинки стала, але ці потенціальні енергії відрізняються на кінцеву величину. Будемо вважати, що на межі областей I та II потенціальна енергія змінюється стрибком, як це показано на рис. 18.12 та 18.13. Виберемо систему координат так, щоб вісь x була паралельна напрямку руху частинки. При цьому потенціальну енергію задають такими умовами:

– область I: за $x < 0$ $U = 0$,

– область II: за $x \geq 0$ $U = U_0 = \text{const}$.

Запишемо рівняння Шредінгера (18.28) для області I:

$$\ddot{\psi}_1 + \frac{2m}{\hbar^2} W\psi_1 = 0, \quad (18.41)$$

і для області II:

$$\ddot{\psi}_2 + \frac{2m}{\hbar^2}(W - U_0)\psi_2 = 0. \quad (18.42)$$

Введемо позначення

$$k_1 = \frac{1}{\hbar}\sqrt{2mW} = \frac{mv_1}{\hbar} = \frac{2\pi}{\lambda_1}, \quad (18.43)$$

$$k_2 = \frac{1}{\hbar}\sqrt{2m(W - U_0)} = \frac{mv_2}{\hbar} = \frac{2\pi}{\lambda_2}, \quad (18.44)$$

де λ_1 та λ_2 – довжини хвиль де Бройля в областях I та II відповідно. Рівняння (18.41) і (18.42) тепер перепишемо у вигляді

$$\ddot{\psi}_1 + k_1^2\psi_1 = 0, \quad (18.41')$$

$$\ddot{\psi}_2 + k_2^2\psi_2 = 0. \quad (18.42')$$

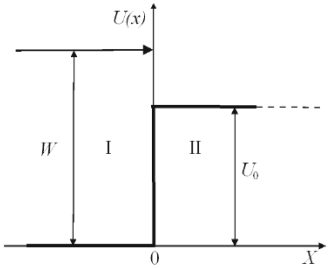


Рис. 18.12. Потенціальний бар’єр нескінченої ширини. Випадок, коли енергія частинки W більша висоти бар’єра U_0

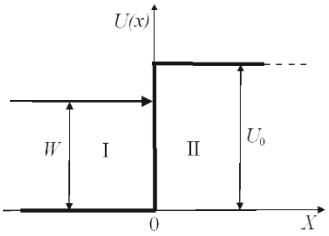


Рис. 18.13. *Потенціальний бар’єр нескінченої ширини. Випадок, коли енергія частинки W менша висоти бар’єра U_0*

Загальні розв’язки цих рівнянь такі:

$$\psi_1 = a_1 e^{ik_1 x} + b_1 e^{-ik_1 x}, \quad (18.45)$$

$$\psi_2 = a_2 e^{ik_2 x} + b_2 e^{-ik_2 x}. \quad (18.46)$$

Величини ψ_1 та ψ_2 подані як суперпозиція плоских хвиль де Бройля, причому $e^{ik_1 x}$ та $e^{ik_2 x}$ характеризують хвилі, які поширюються зліва направо, а $e^{-ik_1 x}$ та $e^{-ik_2 x}$ – хвилі, що поширюються справа наліво.

Розглянемо тепер умови переходу частинки з області I в область II у двох випадках: 1) коли повна енергія частинки W більша за її потенціальну енергію U_0 в області II (рис. 18.12); 2) коли $W < U_0$ (рис. 18.13). Розглянемо ці випадки більш докладно.

1. За $W > U_0$ частинка, яка підпорядковується класичній механіці, обов’язково перейде з області I в область II. Справді, якщо ця частинка несе електричний заряд і рухається в області I зліва направо, то на межі I і II вона потрапляє в поле потенціалу U, яке її затримує. Утім, оскільки її повна енергія $W > U_0$, то частинка подолає це поле затримки і продовжить свій рух в області II з меншою енергією $W - U_0$.

Частинка, рух якої підпорядковується рівнянню Шредінгера, наприклад електрон, має поводити себе зовсім інакше. Це видно навіть з таких якісних міркувань. Рух електрона супроводжується плоскою хвилею де Бройля. На межі двох областей, де відбувається раптова зміна потенціалу, ця хвиля повинна поводити себе так, як веде себе світлова хвиля на межі двох областей з різними показниками заломлення. Це означає, що на межі областей I і II хвиля де Бройля частково відбивається, а частково переходить в область II. Ми можемо також сказати, що переходячи з однієї області в іншу, електрон має ймовірність відбитись і певну ймовірність пройти далі в область II.

Завданням тепер є знайти ці ймовірності. Будемо вважати, що у (18.45) $|a_1|^2$ інтенсивність хвилі, яка падає, а $|b_1|^2$ – інтенсивність відбитої хвилі. В області II поширюється тільки хвиля, що проходить, тому в (18.46) візьмемо $b_2 = 0$. Ймовірність проходження квантової частинки в бар’єр і ймовірність відбиття від бар’єра залежить, таким чином, від коефіцієнтів a_1, b_1, a_2. Ці коефіцієнти можна визначити, врахувавши граничні умови, тобто неперервність хвильової функції і її першої похідної (яка визначає гладкість функції) під час переходу з області I в область II:

$$\begin{matrix}(\Psi_1)|_{\,x=0}=(\Psi_2)|_{\,x=0,}\\ \left(\frac{d\Psi_1}{dx}\right)|_{\,x=0}=\left(\frac{d\Psi_2}{dx}\right)|_{\,x=0.}\end{matrix} \qquad (18.47)$$

З умов (18.47) отримаємо такі співвідношення:

$$a_1 + b_1 = a_2,$$
$$ia_1k_1 - ib_1k_1 = ia_2k_2$$

або

$$a_1 + b_1 = a_2,$$
$$a_1 - b_1 = \frac{k_2}{k_1}a_2. \qquad (18.48)$$

Із системи рівнянь (18.48) отримаємо, що

$$b_1 = \frac{k_1-k_2}{k_1+k_2}a_1, \qquad (18.49)$$

а

$$a_2 = \frac{2k_1}{k_1+k_2}a_1. \qquad (18.50)$$

Тепер ми можемо за аналогією з оптикою визначити коефіцієнт відбиття R і коефіцієнт прозорості D. Коефіцієнт відбиття дорівнює відношенню квадратів амплітуд відбитої і падаючої хвиль:

$$R = \frac{|b_1|^2}{|a_1|^2} = \left|\frac{k_1-k_2}{k_1+k_2}\right|^2. \qquad (18.51)$$

Під час розрахунку коефіцієнта прозорості D слід помножити відношення квадратів амплітуди хвилі, що проходить, до амплітуди хвилі, що падає, на відношення швидкостей частинок в одній та іншій областях з такої причини. Коефіцієнт прозорості дорівнює відношенню потоку, що проходить, до потоку, що падає. Потік частинок дорівнює ρv, де $\rho -$ густина частинок, пропорційна квадрату амплітуди хвилі де Бройля, а швидкість $v \sim p \sim \frac{1}{\lambda}$, де λ – довжина хвилі де Бройля, тому коефіцієнт прозорості

$$D = \frac{\rho_2 v_2}{\rho_1 v_1} = \frac{|a_2|^2\lambda_1}{|a_1|^2\lambda_2} = \frac{|a_2|^2 k_2}{|a_1|^2 k_1}.$$

Врахувавши (18.50), отримаємо

$$D = \frac{4k_1k_2}{(k_1+k_2)^2}. \qquad (18.52)$$

Коефіцієнти R та D можемо тлумачити з корпускулярної точки зору так: R є ймовірність частинці зазнати відбивання на межі областей, а $D -$ ймовірність пройти в область II або, як прийнято говорити, ймовірність подолати потенціальний бар’єр.

Складаючи (18.51) і (18.52), отримаємо $R + D = 1$, що випливає з теореми додавання ймовірностей, оскільки можна однозначно стверджувати, що на межі областей частинка або відіб’ється, або пройде далі.

Вирахуємо тепер коефіцієнти відбиття й пропускання залежно від співвідношення між повною енергією W і потенціальною енергією U_0.

Підставимо у формули (18.51) і (18.52) значення k_1 (формула (18.43)) і k_2 (формула (18.44)). Тоді отримаємо, що

$$R = \left(\frac{1-\sqrt{1-\frac{U_0}{W}}}{1+\sqrt{1+\frac{U_0}{W}}}\right)^2 \qquad (18.53)$$

та

$$D = \frac{4\sqrt{1-\frac{U_0}{W}}}{\left(1+\sqrt{\left(1-\frac{U_0}{W}\right)}\right)^2}. \qquad (18.54)$$

Таблиця 18.1. Коефіцієти прозорості D та відбиття R для різних відношень U_0/W

U_0/W	R	D	W/U_0
0,1	0,007	0,9993	10
0,5	0,0296	0,9704	2
0,8	0,1459	0,8541	1,25
0,9	0,2700	0,2700	1,11
1,0	1,000	0,000	1,0

У табл. 18.1 наведено розрахунки декількох числових значень R і D. Із цієї таблиці видно, що коли енергія частинки вдвічі більша від висоти потенціонального бар’єру U_0, ймовірність відбивання вже має досить значну величину (близько 3 %), а за $U_0 = W$ проникнення в область II взагалі неможливе, тоді як за класичною механікою в цьому випадку, як і в попередніх, частинка достовірно переходить в область II, тільки її кінетична енергія там дорівнює нулю.

2. За $W < U_0$ за класичною механікою перехід з області I в область II неможливий, оскільки за цієї умови (потенціальна енергія більша від повної енергії) кінетична енергія частинки в області II повинна стати від’ємною, а швидкість уявною. Вирахуємо коефіцієнт відбивання R для цього випадку, скориставшись квантовою механікою. Насамперед слід звернути увагу на те, що за $W < U_0$ величина k_2 (формула (18.44)) стає уявною:

$$k_2 = \frac{1}{\hbar}\sqrt{2m(W-U_0)} = \\ = i\frac{1}{\hbar}\sqrt{2m(U_0-W)} = ik,$$

де $k = \frac{1}{\hbar}\sqrt{2m(U_0-W)}$. Тоді згідно з (18.51)

$$R = \left|\frac{k_1-ik}{k_1+ik}\right|^2 = \left(\frac{k_1-ik}{k_1+ik}\right)\left(\frac{k_1+ik}{k_1-ik}\right) = 1,$$

$$D = 1 - R = 0.$$

Отже, за $W < U_0$ коефіцієнт відбивання дорівнює одиниці, тобто відбивання є повним. Такого результату й очікували, однак неочікуваним є те, що, як ми зараз покажемо, хоча відбивання й повне, є певна ймовірність знайти частинку в області II. Інакше кажучи, відбивання не обов’язково відбувається, власне, на межі областей, а деякі частинки заходять в область II для того, щоб потім повернутись в область I. Справді, оскільки за $W < U_0$ коефіцієнт k_2 стає уявним, то розв’язок рівняння Шредінгера (18.46) за умови $b_2 = 0$ набуває вигляду $\psi_2 = a_2 e^{ik_2 x} = \\ = a_2 e^{-kx}$, а ймовірність знайти частинку на одиничній довжині буде:

$$\psi_2\psi_2^* = |\psi|^2 = |a_2|^2 e^{-2kx} = \\ = |a_2|^2 e^{-\frac{2}{\hbar}\sqrt{2m(U_0-W)}\cdot x}. \qquad (18.55)$$

Це означає, що є певна ймовірність знайти частинку в області II. Правда, ця ймовірність експоненціально зменшується зі збільшенням x, але вона відмінна від нуля: мікроскопічні частинки можуть проникати в області, що «заборонені» для частинок макроскопічних величин. Тлумачення отриманих результатів із хвильової точки зору не має труднощів, оскільки випадок $W < U_0$ аналогічний випадку повного внутрішнього відбивання в оптиці. З точки зору геометричної оптики під час переходу світла з більш оптично густого середовища в менш оптично густе, коли кут падіння більший від критичного, світло не може проникати в менш оптично густе середовище. Утім, хвильова оптика показує, а дослід підтверджує, що справді в цьому менш оптично густому середовищі навіть і за кутів падіння, більших від критичного, є хвильове поле з експоненціально спадною амплітудою. Розрахунки показують, що попри проникнення світла з першого у друге середовище середнє значення нормальної складової вектора Пойтинга (потік енергії) до поверхні поділу за достатньо великий проміжок часу дорівнює нулю. Це означає, що немає руху енергії з першого середовища у друге. Детальний аналіз цього випадку, виконаний А. А. Ейхенвальдом, показав, що за повного внутрішнього відбивання лінії вектора Пойтинга виявляються кривими; вони заходять у друге середовище і знову виходять у перше; відбивання продовжує бути повним, незважаючи на встановлення поля у другому середовищі. Відповідно до цього і в нашому випадку коефіцієнт відбивання R дорівнює одиниці, а коефіцієнт пропускання дорівнює нулю: частинки «заходять» в область II, проникають на деяку глибину і знову повертаються в область I.

У квантовій механіці, однак, немає ніякого парадоксу в тому, що частинку можна виявити в області, де $W < U_0$. По-перше, слід завжди пам’ятати, що для мікроскопічних частинок завдяки співвідношенню невизначеностей одночасні точні значення x і p_x неможливі, тому не має сенсу говорити про одночасні, конкретні значення потенціальної (функція x) і кінетичної (функція p) енергії.

Явище, коли електромагнітна (світлова) енергія заходить у друге, оптично менш густе середовище, спостерігав експериментально на досліді Квінке. На рис. 18.14 показано схему

досліду. Дві призми повного внутрішнього відбиття розміщені так, що між ними утворюється зазор дуже малої товщини (співмірний з довжиною падаючої хвилі, рис. 18.14). За великої товщини зазору світло не проникає крізь повітряний прошарок і приймач не реєструє енергію. Утім, якщо товщина зазору менша від глибини проникнення хвилі за повного внутрішнього відбиття, то енергія, яка проникає, проходячи через крізь призму, потрапляє у приймач. Змінюючи товщину зазору, можна регулювати кількість енергії, яка пройшла. На цьому принципі побудовано один зі світлових модуляторів. Товщину зазору змінюють під дією звукових хвиль (мовлення). Приймаючи модульоване світло на фотоелемент, можна отримати змінний струм і, таким чином, відтворити звук (світловий телефон).

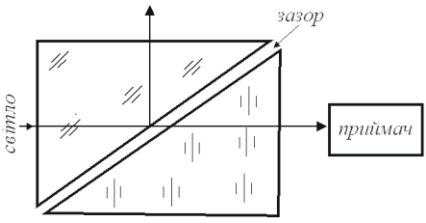


Рис. 18.14. Схема досліду Квінке

Цікаво зазначити, що у цьому досліді товщина зазору залежить від довжини хвилі. Так, наприклад, в ультракороткому діапазоні вона у 10^5 раз більша, ніж в оптичному діапазоні.

18.9.3. Розв'язок рівняння Шредінгера для потенціального бар'єра кінцевої ширини. Тунельний ефект

Нехай частинка, що рухається в додатньому напрямку осі *OX*, зустрічає на своєму шляху прямокутний потенціальний бар'єр, висота якого U_0 й ширина l (рис. 18.15). Отже, величина U вздовж осі *OX* змінюється так:

$$U(x) = \begin{cases} 0, & x < 0; \\ U_0, & 0 \le x \le l; \\ 0, & x > l. \end{cases}$$

За таких умов задачі класична частинка, яка має енергію W, або пройде над бар'єром за $W > U_0$, або відіб'ється від нього за $W < U_0$ й буде рухатись у зворотний бік.

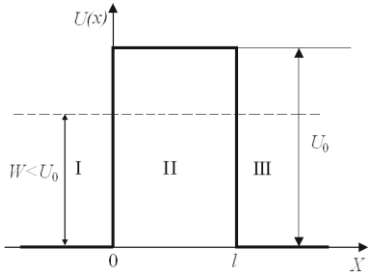


Рис. 18.15. Прямокутний потенціальний бар'єр

Для квантової частинки навіть за $W > U_0$ є відмінною від нуля ймовірністю того, що частинка

відіб'ється від бар'єра і буде рухатись у зворотний бік. За $W < U_0$ є також відмінною від нуля ймовірність того, що частинка проникає через бар'єр і виявиться в області з $x > l$.

Рівняння Шредінгера (18.28) для кожної з виділених областей має такий вигляд:

– для областей I та III

$$\ddot{\psi} + \frac{2m}{\hbar^2} W\psi = 0, \qquad (18.56)$$

– для області II

$$\ddot{\psi} + \frac{2m}{\hbar^2}(W - U_0)\psi = 0. \qquad (18.57)$$

Загальні розв'язки цих диференціальних рівнянь будемо шукати у такому вигляді:

$$\begin{cases} \text{I: } \psi_1 = a_1 e^{ik_1 x} + b_1 e^{-ik_1 x}, \\ \text{II: } \psi_2 = a_2 e^{ik_2 x} + b_2 e^{-ik_2 x}, \\ \text{III: } \psi_3 = a_3 e^{ik_1 x} + b_3 e^{-ik_1 x}, \end{cases} \qquad (18.58)$$

де

$$k_1 = \frac{1}{\hbar}\sqrt{2mW}, \qquad (18.59)$$

$$k_2 = \frac{1}{\hbar}\sqrt{2m(W - U_0)}. \qquad (18.60)$$

В області III є лише хвиля, що пройшла крізь бар'єр і поширюється зліва направо, тому коефіцієнт $b_3 = 0$. За $W < U_0$ в області II k_2 є уявним числом, тобто

$$k_2 = ik = i\frac{1}{\hbar}\sqrt{2m(U_0 - W)}. \qquad (18.61)$$

У випадку широкого бар'єра незалежно від енергії частинки, як ми бачили, є певна ймовірність проникнути під бар'єр. У випадку, який розглядається, потенціальний бар'єр має кінцеву ширину, тому є ймовірність не тільки проникнення, але й проходження частинки крізь бар'єр в область III.

Цей випадок відрізняється від попереднього тим, що відбивання частинок відбувається як на межі областей I та II, так і на межі областей II та III. Для розрахунку коефіцієнтів R і D треба знайти коефіцієнти b_1, b_2, a_2, a_3 (будемо вважати, що $a_1 = 1$). Для цього скористаємось умовами неперервності функції ψ та її першої похідної на межах областей I та II, II й III, тобто за $x = 0$ та $x = l$.

Випишемо ці умови:

$$\begin{matrix} (\psi_1)|_{x=0} = (\psi_2)|_{x=0}, \\ \left(\frac{d\psi_1}{dx}\right)|_{x=0} = \left(\frac{d\psi_2}{dx}\right)|_{x=0}, \end{matrix} \qquad (18.62)$$

$$\begin{matrix} (\psi_2)|_{x=l} = (\psi_3)|_{x=l}, \\ \left(\frac{d\psi_2}{dx}\right)|_{x=l} = \left(\frac{d\psi_3}{dx}\right)|_{x=l}. \end{matrix} \qquad (18.63)$$

Із врахуванням виду хвильових функцій (18.58) ці умови дають таку систему рівнянь (враховано, що $a_1 = 1$, $b_3 = 0$):

$$\begin{cases} 1 + b_1 = a_2 + b_2, \\ k_1 - k_1 b_1 = k_2 a_2 - k_2 b_2, \end{cases}$$

$$\begin{cases} a_2 e^{ik_2 l} + b_2 e^{-ik_2 l} = a_3 e^{ik_1 l}, \\ a_2 e^{ik_2 l} - b_2 e^{-ik_2 l} = a_3 \dfrac{k_1}{k_2} e^{ik_1 l}. \end{cases} \qquad (18.64)$$

Із цієї системи рівнянь можна визначити коефіцієнти b_1, b_2, a_2, a_3. Утім, коефіцієнти b_1, b_2, a_2, а також

коефіцієнт відбиття R визначати не будемо, оскільки це нічого принципово нового порівняно з попереднім розглядом не дає. Для коефіцієнта a_3 отримуємо такий вираз:

$$a_3 = \\ = \frac{4k_1k_2e^{ik_1l}}{(k_1+k_2)^2\,e^{-ik_2l}-(k_1-k_2)^2e^{ik_2l}}. \quad (18.65)$$

Коефіцієнт прозорості бар'єра D знаходять за формулою

$$D = \frac{|a_3|^2}{|a_1|^2} = |a_3|^2 = a_3a_3^*. \quad (18.66)$$

Із врахуванням (18.61) і (18.65), можна отримати такий вираз для D:

$$D = \frac{4k_1^2k^2}{(k_1^2+k^2)^2\mathrm{sh}^2kl+4k_1^2k^2}, \quad (18.67)$$

де гіперболічна функція

$$\mathrm{sh}kl = \frac{e^{kl}-e^{-kl}}{2}.$$

У більшості розглядуваних випадків можна вважати, що $\mathrm{sh}^2kl \cong$ $\cong \frac{1}{4}e^{2kl}$. Тоді

$$D = \frac{4}{\frac{1}{4}\left(\frac{k_1}{k}+\frac{k}{k_1}\right)^2e^{2kl}+4}. \quad (18.68)$$

Доданком 4 у знаменнику можна знехтувати, оскільки k_1 і k одного порядку й $e^{2kl} \gg 4$. З точністю до несуттєвого множника

$$D \sim e^{-2kl} = e^{-\frac{2}{\hbar}\sqrt{2m(U_0-W)}\,l}. \quad (18.69)$$

Формула (18.69) показує, що проникність бар'єра експоненціально залежить від його ширини l.

Таблиця 18.2. Проникність бар'єра

l, Å	1	2	3	5	10
D, відн. один.	$2{,}3\cdot10^{-1}$	$5{,}5\cdot10^{-2}$	$1{,}3\cdot10^{-2}$	$7\cdot10^{-4}$	$5\cdot10^{-7}$

Для електрона за $U_0 - W = 2$ еВ розрахунок дає такі значення D. Як видно з табл. 18.2, для ширини бар'єра в 1 Å (атомні розміри) проникність досить велика (десятки відсотків), але вже за $l =$ $= 10$ Å вона стає нехтовно малою.

Ми розглянули випадок проходження частинки крізь потенціальний бар'єр досить спрощеної прямокутної форми. Це ідеалізований випадок. Насправді потенціальний бар'єр має більш складну форму, ніж зображений на рис. 18.11. Для такого бар'єра коефіцієнт прозорості приблизно виражають такою формулою:

$$D = \\ = D_0\exp\left\{-\frac{2}{\hbar}\int_{x_1}^{x_2}\sqrt{2m(U-W)}\,dx\right\}, \quad (18.70)$$

де $U = U(x)$, а $D_0 = \frac{\left(\frac{4k}{k_1}\right)^2}{\left(1+\left(\frac{k}{k_1}\right)^2\right)^2}$, x_1 і x_2 – поворотні точки, в яких $U(x_1) = U(x_2) = W$; k_1 і k визначають за формулами (18.59) і (18.61); D_0 – константа, яка приблизно дорівнює одиниці, а інтеграл у показнику експоненти формули (18.70) характеризує начебто товщину бар'єра на рівні W.

Формула (18.69) є частинним випадком формули (18.70).

Явище проходження частинки крізь потенціальний бар'єр називають ***тунельним ефектом***. Таку назву дали тому, що частинка проходить крізь бар'єр, начебто через тунель.

На рис. 18.16 показано якісний вигляд функцій $\Psi_1(x)$, $\Psi_2(x)$, $\Psi_3(x)$.

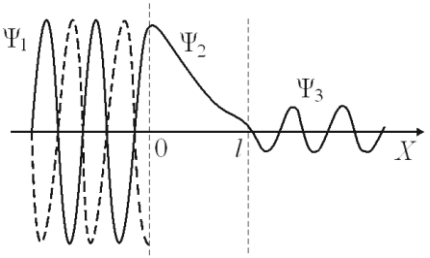


Рис. 18.16. Якісний вигляд функцій $\Psi_1(x)$, $\Psi_2(x)$, $\Psi_3(x)$

Важливо зазначити, що частинки виходять з бар'єра з тією ж енергією, з якою на нього і падають ($W = \hbar\omega$). Це означає, що частота хвилі де Бройля не змінюється (під час дифракції світла теж немає зміни частоти).

Тунельний ефект є специфічним квантовим ефектом. Проходження мікрочастинки крізь область, в яку згідно із законами класичної механіки вона не може проникнути, можна пояснити співвідношенням невизначеностей. Невизначеність імпульсу Δp на відрізку $\Delta x = l$ дорівнює $\Delta p > \frac{\hbar}{l}$. Пов'язана з цим розкидом у значеннях імпульсу невизначеність кінетичної енергії $\Delta W = \frac{p}{m}\Delta p = \sqrt{\frac{2W}{m}}\Delta p = \sqrt{\frac{2W}{m}}\frac{\hbar}{l}$ може виявитися достатньою для того, щоб повна енергія $W + \sqrt{\frac{2W}{m}}\frac{\hbar}{l}$ мікрочастинки виявилася більшою за потенціальну U_0.

Проходження частинок крізь потенціальний бар'єр експериментально підтверджено в явищі холодної емісії електронів з металу. Тунельний ефект відіграє основну роль у явищах радіоактивного α-розпаду, перебігу термоядерних реакцій.

Тунельний ефект лежить в основі відомого ефекту Джозефсона, відкритого в 1963 р. Суть його в тому, що під час проходження тунельного струму крізь тонкий шар діелектрика між двома надпровідниками, відбувається випромінювання електромагнітних хвиль надвисоких частот. Ефект Джозефсона широко застосовують в логічних пристроях, у параметричних перетворювачах та в інших електронних приладах.

На основі тунельного ефекту працюють тунельні діоди, які широко використовують в радіотехніці високих частот, обчислювальній техніці.

18.10. Холодна емісія електронів з металу

Внаслідок тунельного ефекту стає можливою емісія з непідігрітого (холодного) катода під впливом електричного поля ($E \sim 10^8 \ldots 10^9$ В/м). Це *холодна* або *автоелектронна емісія.*

Метал можна розглядати як газ вільних електронів, які можуть переміщуватись «вільно» по кристалу (металу). Границі кристала при цьому є начебто своєрідними стінками, які обмежують об'єм, що заповнений електронним газом. Якщо умовно вибрати за початок відліку енергію вільного електрона, що перебуває у стані спокою у вакуумі, то енергія електрона, який рухається у вакуумі, буде позитивною, а енергія електрона в металі – від'ємною. Таким чином, електрон у металі є начебто в потенціальній ямі (рис. 18.17), і на його шляху до виходу з металу існує потенціальний бар'єр. Як було показано в підрозд. 18.8, енергія електрона в потенціальній ямі дискретна й вільні електрони в металі (їх концентрація в металі $n \sim 10^{28}$ м$^{-3}$) згідно з принципом Паулі заповнюють по два на кожний енергетичний рівень. Навіть за кімнатної температури найвищий із заповнених рівнів утворює різку межу між заповненими й вільними рівнями.

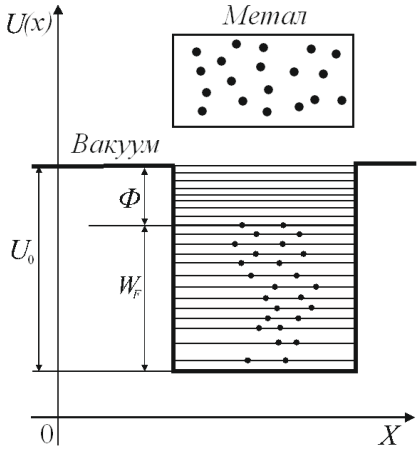


Рис. 18.17. Електрони у металі начебто в потенціальній ямі. Точками позначено електрони

Цей найвищий заповнений рівень називають межею Фермі, а його енергія – енергією Фермі W_F (рис. 18.17). Очевидно, що для звільнення електрона з металу треба надати йому енергію

$$\Phi = U_0 - W_F,$$

де U_0 – глибина потенціальної ями; Φ – робота виходу електрона з металу (входить в рівняння Ейнштейна для фотоефекту). Електрон, однак, може бути звільнений з металу не тільки під дією світла, але і внаслідок нагрівання (термоелектронна емісія). Якщо до металу прикласти електричне поле, як показано на рис. 18.18, то всередині металу і на його поверхні енергія електрона не зміниться (електричне поле не проникає у провідник (див. розд. 3)),

тому енергію електрона, який вийшов з металу і перебуває на відстані x від його поверхні в електричному полі, можна записати так:

$$U(x) = U_0 - eEx, \qquad (18.71)$$

де $(-eEx)$ – потенціальна енергія електрона в однорідному електричному полі, напруженість якого E. Залежність (18.71) на рис. 18.18 зображено прямою лінією *MN*. Прозорість бар'єра *KLMN* визначимо за формулою (18.70), взявши $x_1 = 0$:

$$D = D_0 \exp\left\{-\frac{2}{\hbar}\int_0^{x_2}\sqrt{2m_e(U(x)-W)}dx\right\}.$$

Будемо вважати, що $W = W_F$, тоді

$$\int_0^{x_2}\sqrt{U_0 - eEx - W_F}\,dx = \int_0^{x_2}\sqrt{\Phi - eEx}\,dx =$$

$$= -\frac{2}{3eE}(\Phi - eEx_2)^{\frac{3}{2}} + \frac{2}{3eE}\Phi^{\frac{3}{2}} = \frac{2\Phi^{\frac{3}{2}}}{3eE},$$

за умови, що поворотну точку x_2 знаходять з умови $U_0 - W_F = eEx_2$. Тоді

$$D = D_0 \exp\left\{-4\,\frac{\sqrt{2m_e}\Phi^{\frac{3}{2}}}{3e\hbar}\cdot\frac{1}{E}\right\},$$

або

$$D = D_0 \exp\left\{-\frac{\alpha}{E}\right\}\cdot \qquad (18.72)$$

Струм холодної емісії можна записати так:

$$j = j_0 \exp\left\{-\frac{\alpha}{E}\right\}, \qquad (18.73)$$

де j_0 та $\alpha = 4\,\frac{\sqrt{2m_e}\Phi^{\frac{3}{2}}}{3e\hbar}$ – константи, які характеризують метал; E – напруженість поля. Саме таку залежність «холодного струму» від напруженості поля і було знайдено експериментально.

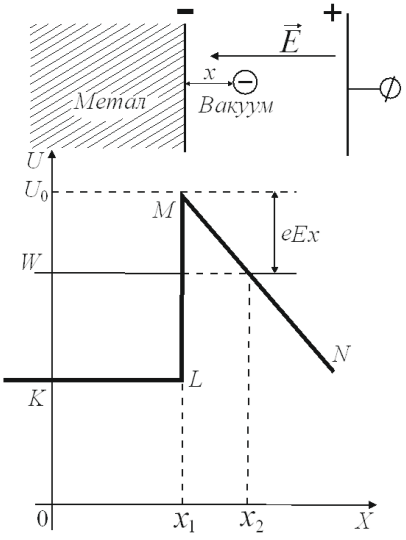


Рис. 18.18. Потенціальний бар'єр для електронів в металі при прикладанні електричного поля

18.11. Ефект Рамзауера

Підтвердженням хвильових властивостей електрона є ефект Рамзауера. Ефект Рамзауера відомий також як ефект Рамзауера–Таундсена – явище аномально слабкого розсіювання повільних електронів атомами нейтральних газів. Це явище вперше

спостерігав у 1921 р. Карл Рамзауер під час вивчення розсіювання електронів в аргоні, а пізніше ефект спостерігали і в інших речовинах.

Рух заряджених частинок (електронів та іонів) в газі є одним з фізичних процесів, що посідає провідне місце у фізиці плазми та газових розрядів і визначає умови роботи всіх газорозрядних та плазмових приладів. При цьому виникають різні зіткнення заряджених частинок з нейтральними частинками газу (атомами, молекулами).

Залежно від енергії електронів розрізняють *пружні* й *непружні* зіткнення. Пружні зіткнення не супроводжуються зміною внутрішньої енергії атомів. Частина кінетичної енергії, яка передається у процесі пружного зіткнення електрона з атомом, приблизно дорівнює відношенню маси електрона m_e та атома M. Це відношення за порядком величини дорівнює: $m_e/M \sim 10^{-4}$, тому пружне зіткнення фактично не змінює кінетичної енергії електронів, а змінює лише напрямок їхнього руху. Пружні зіткнення характерні для електронів з кінетичною енергією, яка дорівнює одиницям електронвольт (еВ).

Зіткнення, за яких внутрішня енергія атома і кінетична енергія електрона змінюються, називають непружними. Непружні зіткнення бувають 1- і 2-го роду. Під час непружних зіткнень 1-го роду електрон віддає частину своєї енергії на збудження або іонізацію атома. Ці явища відбуваються, якщо електрони, які налітають на атом, мають достатню для збудження атома кінетичну енергію – десятки й сотні електронвольт (еВ). Під час непружних зіткнень 2-го роду електрону передається частина енергії збудженого атома, або вся ця енергія збудження, тому такі зіткнення можуть відбуватись між електронами й атомами, які перебувають у збудженому стані.

Основною характеристикою процесу *пружного розсіювання* є його ефективний переріз σ, який характеризує ймовірність пружного зіткнення електрона з окремим атомом. Величину σ можна вважати площею кола з центром в ядрі атома, під час потрапляння в який електрон відхиляється від початкового напрямку. Ефективний переріз пружного розсіювання пов'язаний простим співвідношенням з коефіцієнтом ослаблення електронного пучка за його проходженням через газ:

$$k = \sigma \cdot n, \qquad (18.74)$$

де n – кількість атомів в одиниці об'єму газу. Можна показати, що зменшення інтенсивності пучка I під час його проходження крізь шар газу завтовшки x є експоненціальним:

$$I(x) = I(0) \cdot \exp(-kx). \quad (18.75)$$

Величину ослаблення електронного струму $I(x)/I(0)$ вимірюють експериментально. Тоді можна знайти значення ефективного перерізу σ. Яким чином σ мала б залежати від швидкості (енергії) електронів? З погляду класичної фізики слід очікувати залежність перерізу σ пружного розсіювання електронів від їх швидкості у газі. Під час руху електрона поблизу атома електричне поле електрона зміщує електронну оболонку атома відносно ядра, у результаті атом набуває дипольного моменту. Потенціальне поле цього наведеного диполя викликає силу притягання між електроном та атомом, яке змінює траєкторію руху електрона. Чим повільніше рухаються електрони, тим більше часу вони перебувають у потенціальному полі атома й тим більша зміна напрямку руху відбуватиметься. Інакше кажучи, електрони з більшою кінетичною енергією розсіюватимуться атомами так само, як і електрони з меншою енергією, якщо ці швидкі електрони пролітатимуть на менших відстанях від атома. Таким чином, ефективний переріз розсіювання має монотонно зменшуватися зі збільшенням швидкості електронів, що налітають на атоми.

Це означає, що згідно з формулою (18.75) за $x = \text{const}$ прозорість газу для електронів буде монотонно зростати зі збільшенням їхньої енергії, як схематично показано на рис. 18.19 (пунктирна лінія *1*).

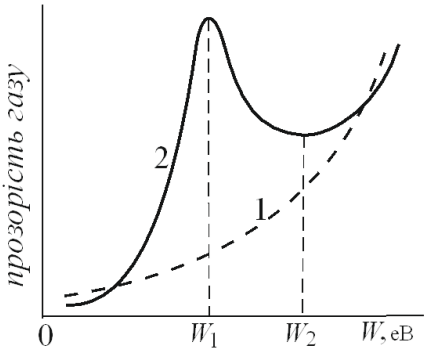


Рис. 18.19. Прозорість нейтрального газу для електронів в залежності від їхньої енергії. Крива 1 – класична теорія; 2 - експеримент

Тим часом досліджуючи у 1921 р. проходження електронного потоку дуже повільних електронів (з енергією 0,75... 1,1 еВ) у газі аргоні, К. Рамзауер виявив, що зі зменшенням енергії електронів розсіювання останніх дуже різко зменшується. Це фізичне явище, коли електрони певної енергії проходять крізь газ фактично безперешкодно

(рис. 18.19, максимум лінії *2*), отримало назву ефекту Рамзауера. Цей ефект, що явно суперечить уявленням класичної фізики про зіткнення електронів з нейтральними атомами, знайшов пояснення при врахуванні хвильової природи мікрочастинок.

Властивість мікрочастинок відбиватися від областей з різким стрибком потенціалу є суто квантово-механічним ефектом, що випливає із хвильових властивостей матерії й не можна пояснити у класичній фізиці. Саме цей ефект відбивання електронного потоку від поверхонь зі стрибком потенціалу дозволяє пояснити ефект Рамзауера.

Електрони під час руху в газі складно взаємодіють з атомами. Головним фактором, що визначає ймовірність зіткнення, є потенціальне поле атома, створюване електронами та ядром, що входять до його складу. Важливу роль відіграє спотворення потенціального поля атома електронами, що рухаються біля нього, внаслідок утворення у частинки-мішені наведеного дипольного моменту (рис. 18.20). Поле цього наведеного диполя $\vec{p}$ викликає додаткове притягання між електроном та атомом і є вирішальним для появи ефекту Рамзауера. Крім цієї поляризаційної сили притягання, існує низка інших сил, що діють під час зіткнення електрона з нейтральною частинкою:

1) додаткова поляризаційна сила відштовхування;

2) кулонівська сила притягання між електроном та атомним ядром;

3) сила відштовхування від електронів, що екранують; ядро;

4) короткодійна обмінна сила відштовхування, обумовлена принципом Паулі для частинок з однаковими спінами. Ця сила перешкоджає просторовому перекриванню електронних оболонок частинок-мішеней і вільного електрона.

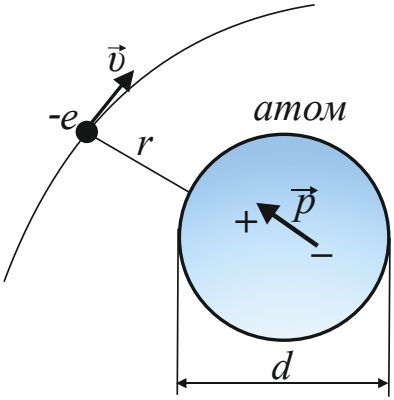


Рис. 18.20. Утворення у частинки-мішені наведеного дипольного моменту

Результуюче потенціальне поле всіх названих взаємодій можна приблизно подати як тривимірну потенціальну яму зі стінками різної

крутизни. Утім, точний квантово-механічний розгляд ефекту Рамзауера з використанням моделі тривимірної сферичної потенціальної ями є досить складним. Зазвичай обмежуються описом цього явища на основі моделі одновимірної прямокутної потенціальної ями, що дозволяє, тим не менше, виявити його основні риси й отримати наближені кількісні співвідношення.

Схематично потенціальну яму разом із зображенням атома мішені зображено на рис. 18.21.

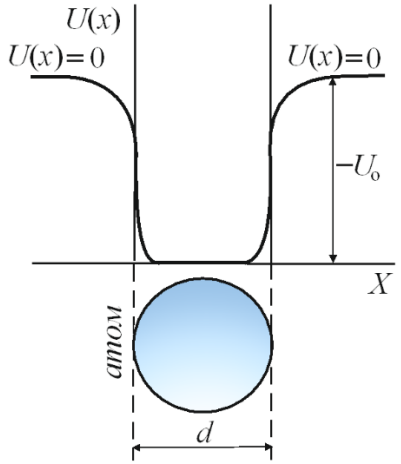


Рис. 18.21. Схема потенціальної ями разом із зображенням атома мішені

Квантовий розгляд пружного розсіювання електронів. Руху електронів у квантовій механіці ставиться в відповідність деякий хвильовий процес – хвиля де Бройля, або Ψ-хвиля. Відповідно, взаємодію електронного пучка з атомами газу розглядають як розсіювання на них первинної «падаючої» хвилі де Бройля. Вторинні Ψ-хвилі, що утворюються під час розсіювання цієї хвилі на різних частинах атома, інтерферують між собою, а кількість розсіяних у тому чи іншому напрямку електронів є пропорційним квадрату модуля амплітуди результуючої розсіяної Ψ-хвилі. Може вийти так, що за певного співвідношення між довжиною хвилі електрона і характерними розмірами атома розсіяні Ψ-хвилі у результаті інтерференції послаблять одна одну й таким чином виникнуть сприятливі умови для практично безперешкодного проходження електронів крізь газ. Так з'являться умови для спостереження ефекту Рамзауера.

Як у класичній, так і у квантовій механіці процеси розсіювання визначають, зрештою, потенціальною енергією взаємодії частинки, що розсіюється, і розсіювача (у нашому випадку – електрона та атома). Правильне якісне уявлення про енергію їх взаємодії можна скласти, враховуючи два основні фактори: для електронів, що пролітають «поза» атомом, – поява в атома наведеного електричного

дипольного моменту, направленого позитивним полюсом до електрона (рис. 18.20), а для електронів, що проходять «крізь» атом, – зменшення екранування ядра електронами оболонки. Обидві ці причини сприяють притяганню електрона до атома, що відповідає появі «потенціальної ями», яка має розміри, співмірні з розміром атома (рис. 18.21). За межами ями потенціальна енергія взаємодії електрона дуже швидко виходить на постійне значення, що зазвичай беруть за нуль. За такого вибору початку відліку енергії потенціальна енергія всередині ями буде негативною. У наближенні, що його розглядаємо, реальний хід потенціалу всередині ями замінюють деяким постійним значенням $U = U_0$, де U_0 – ефективна глибина ями. Зрозуміло, потенціальна яма має не прямовисні, а пологі стінки, *однак крутизна стінок* (рис. 18.21) *збільшується зі зростанням атомного номера елемента й особливо сильно – для атомів благородних газів, які відрізняються найбільш компактною електронною структурою і різкою зовнішньою межею. Саме з цієї причини ефект Рамзауера найкраще проявляється в аргоні, криптоні та* ***ксеноні***. Для них можна використовувати модель сферично-симетричної потенціальної ями з вертикальними стінками і вважати, що вторинні Ψ-хвилі, які інтерферують і визначають ймовірність розсіювання електронів, утворюються в результаті відбиття падаючої хвилі де Бройля від стінок цієї ями. Переходячи до ще більш грубої моделі – до одновимірної прямокутної ями (рис. 18.22), отримуємо, що за такого (одновимірного) розсіювання результат інтерференції безпосередньо визначається співвідношенням між шириною потенціальної ями d і довжиною Ψ-хвилі λ для електронів «всередині» ями.

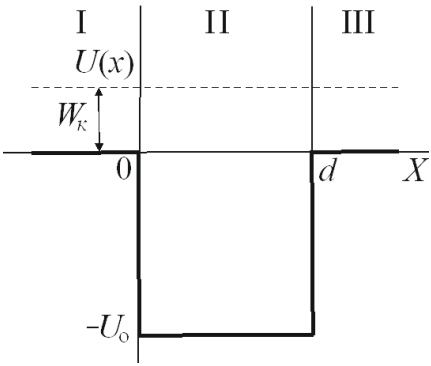


Рис. 18.22. Одновимірна прямокутна потенціальна яма

Згідно з де Бройлем довжина хвилі, що відповідає електрону з імпульсом p,

$$\lambda = \frac{h}{p}. \qquad (18.76)$$

За малих (нерелятивістських) енергій

$$\lambda = \frac{h}{\sqrt{2m_e W_к}}, \qquad (18.77)$$

де h – стала Планка; m_e – маса електрона; $W_к$ – кінетична енергія електрона. Поки електрон перебуває за межами потенціальної ями (рис. 18.22) його кінетична енергія $W_к$ збігається з повною енергією W і довжина хвилі

$$\lambda = \frac{h}{\sqrt{2m_e W}}\cdot \qquad (18.78)$$

Всередині потенціальної ями кінетична енергія електрона буде більшою:

$$W_к = W - (-U_0) = W + U_0, \qquad (18.79)$$

а довжина Ψ-хвилі λ всередині ями – меншою, ніж поза нею:

$$\lambda = \frac{h}{\sqrt{2m_e(W+U_0)}} = \frac{\lambda_0}{\sqrt{1+\frac{U_0}{W}}}\cdot \qquad (18.80)$$

Якщо провести аналогію з оптикою, то знаменник у формулі (18.80) можна інтерпретувати як показник заломлення n для Ψ-хвиль:

$$n = \sqrt{1+\frac{U_0}{W}}\cdot \qquad (18.81)$$

Таким чином, на передній стінці ями Ψ-хвиля переходить в область з більшим показником заломлення, а на задній – з меншим. Як і в оптиці, на межі областей з різними показниками заломлення відбувається відбивання хвилі, причому під час відбивання від передньої стінки відбувається стрибок фази на π («втрата на півхвилі»), а під час відбивання від задньої стінки фаза не змінюється. Розглянемо інтерференцію хвиль де Бройля під час проходження атома (рис. 18.23).

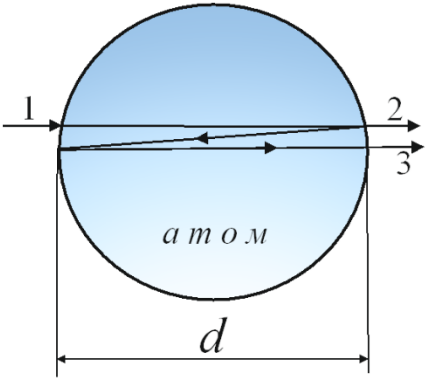


Рис. 18.23. Хвилі: 1 – яка падає на атом; 2 та 3 – які проходять атом

Хвиля *3* зазнає подвійного відбивання всередині атома (потенціальної ями). Хвилі *2* та *3* когерентні й інтерферують між собою. Максимум інтерференції цих хвиль означає, що падаюча хвиля *1* на рис. 18.23 буде проходити атом, і це буде відповідати умові максимуму прозорості газу для електронів (максимуму інтерференції хвиль). Хвилі *2* та *3* на рис. 18.23 набувають різницю ходу $\Delta = 2d$. Максимуму інтерференції буде відповідати умова

$$\Delta = 2d = \lambda. \qquad (18.82)$$

З урахуванням (18.80) формулу (18.82) можна переписати так:

$$\frac{h}{\sqrt{2m_e(W+U_0)}} = 2d. \qquad (18.83)$$

Звідси знайдемо, що максимально прозоро (рис. 18.19) проходити газ будуть електрони з енергією

$$W_1 = \frac{h^2}{8m_e d^2} - U_0. \qquad (18.84)$$

Мінімуму прозорості газу для електронів буде відповідати випадок, коли хвилі *2* та *3* на рис. 18.23 будуть компенсувати одна одну (суперпозиція цих хвиль буде давати мінімум інтерференції). Це відповідає умові

$$\Delta = 2d = \frac{3}{2}\lambda, \qquad (18.85)$$

а з урахуванням (18.80) отримаємо, що

$$\frac{3}{2}\frac{h}{\sqrt{2m_e(W+U_0)}} = 2d. \quad (18.86)$$

Звідси знайдемо, що за енергії електронів

$$W_2 = \frac{9h^2}{32m_e d^2} - U_0 \qquad (18.87)$$

буде мінімум прозорості газу для них або максимум розсіювання електронів на атомах газу (рис. 18.19, лінія *2*).

Квантово-механічний розгляд розсіювання електронів на одновимірній потенціальній ямі (пояснення ефекту Рамзауера)

Для потенціального поля, яке показано на рис. 18.22, аналітичний вираз для потенціальної енергії запишемо так:

$$U(x) = \begin{cases} 0, & x < 0; \\ -U_0, & 0 \le x \le d; \\ 0, & x > d. \end{cases} \quad (18.88)$$

Тоді рівняння (18.28) для кожної з виділених областей мають такий вигляд:

– для областей I і III

$$\ddot{\psi} + \frac{2m}{\hbar^2} W\psi = 0, \quad (18.89)$$

– для області II

$$\ddot{\psi} + \frac{2m}{\hbar^2}(W + U_0)\psi = 0, \quad (18.90)$$

або

$$\ddot{\psi} + k_1^2 W\psi = 0, \qquad (18.91)$$

$$\ddot{\psi} + k_2^2 W\psi = 0, \qquad (18.92)$$

де введено позначення

$$k_1 = \frac{1}{\hbar}\sqrt{2mW} = \frac{2\pi}{\lambda_1}, \quad (18.93)$$

$$k_2 = \frac{1}{\hbar}\sqrt{2m(W + U_0)} = \frac{2\pi}{\lambda_2}, \quad (18.94)$$

k_1 і k_2 – хвильові числа для областей I, III і II відповідно.

Примітка. Для руху мікрочастинки в області потенціального бар'єра, коли $W < U_0$, k_2 є уявним (комплексним) числом (формула (18.61)). У нашому випадку k_2 є дійсним числом, оскільки $W > U_0$.

Загальні розв'язки диференціальних рівнянь (18.91) і (18.92) для областей I, II і III будемо шукати у такому вигляді стоячих хвиль:

$$\begin{cases} \text{I: } \psi_1 = a_1 e^{ik_1 x} + b_1 e^{-ik_1 x}, \\ \text{II: } \psi_2 = a_2 e^{ik_2 x} + b_2 e^{-ik_2 x}, \\ \text{III: } \psi_3 = a_3 e^{ik_1 x} + b_3 e^{-ik_1 x}. \end{cases} \quad (18.95)$$

У цих розв'язках $i = \sqrt{-1}$ – уявна одиниця, $a_1 -$ амплітуда хвилі, яка падає на потенціальну яму, $b_1 -$ амплітуда відбитої від ями хвилі, $a_2 -$ амплітуда хвилі в області II, що поширюється зліва направо; $b_2 -$ амплітуда хвилі в області II, що поширюється справа наліво, $a_3 -$ амплітуда хвилі, що пройшла потенціальну яму (область III). Будемо вважати, що в області III немає хвилі, що поширюється справа наліво, тому $b_3 = 0$.

Нас буде цікавити коефіцієнт прозорості ями:

$$D = \frac{|a_3|^2}{|a_1|^2} = |a_3|^2 = a_3 a_3^*,$$

де $a_3^* -$ комплексно спряжена величина до a_3, вважатимемо, що $a_1 = 1$.

Знайдемо коефіцієнти b_1, b_2, a_2, a_3. Загальна хвильова функція ψ є лінійною комбінацією хвильових функцій ψ_1, ψ_2, ψ_3. Для цього скористаємось умовами неперервності функції ψ та її першої похідної на межах областей I та II, II й III, тобто за $x = 0$ і $x = d$. Запишемо ці умови:

$$\begin{matrix} (\psi_1)|_{x=0} = (\psi_2)|_{x=0}, \\ \left(\frac{d\psi_1}{dx}\right)|_{x=0} = \left(\frac{d\psi_2}{dx}\right)|_{x=0}, \end{matrix} \quad (18.96)$$

$$\begin{matrix} (\psi_2)|_{x=d} = (\psi_3)|_{x=d}, \\ \left(\frac{d\psi_2}{dx}\right)|_{x=d} = \left(\frac{d\psi_3}{dx}\right)|_{x=d}. \end{matrix} \quad (18.97)$$

Із врахуванням виду хвильових функцій (18.95) ці умови дають таку систему рівнянь (враховано, що $a_1 = 1$, $b_3 = 0$):

$$\begin{cases} 1 + b_1 = a_2 + b_2, \\ k_1 - k_1 b_1 = k_2 a_2 - k_2 b_2. \end{cases} \quad (18.98)$$

$$\begin{cases} a_2 e^{ik_2 d} + b_2 e^{-ik_2 d} = a_3 e^{ik_1 d}, \\ a_2 e^{ik_2 d} - b_2 e^{-ik_2 d} = a_3 \frac{k_1}{k_2} e^{ik_1 d}. \end{cases} \quad (18.99)$$

З цієї системи рівнянь можна визначити коефіцієнти b_1, b_2, a_2, a_3. Розв'язок цієї системи рівнянь для коефіцієнта a_3 дає такий вираз:

$$a_3 = \frac{4k_1 k_2 e^{ik_1 d}}{(k_1+k_2)^2 e^{-ik_2 d} - (k_1-k_2)^2 e^{ik_2 d}}. \quad (18.100)$$

(Розв'язок системи рівнянь (18.98) і (18.99) наведено в кінці підручника (див. дод. 1)).

Тоді коефіцієнт прозорості ями буде:

$$D = a_3 a_3^* = \frac{4k_1 k_2 e^{ik_1 d}}{(k_1+k_2)^2 e^{-ik_2 d} - (k_1-k_2)^2 e^{ik_2 d}} \times \frac{4k_1 k_2 e^{-ik_1 d}}{(k_1+k_2)^2 e^{ik_2 d} - (k_1-k_2)^2 e^{-ik_2 d}}. \quad (18.101)$$

Після алгебраїчних перетворень (див. дод. 1) отримаємо:

$$D = \frac{1}{1 + \frac{1}{4}\left(\frac{k_2}{k_1}+1\right)^2 \left(\frac{k_1}{k_2}-1\right)^2 \sin^2 k_2 d}. \quad (18.102)$$

Як, видно, із (18.102) всі величини у знаменнику дробу є додатними величинами, тому перший максимум коефіцієнта проходження D буде за умови

$$\sin^2 k_2 d = 0,$$

або

$$k_2 d = \pi. \qquad (18.103)$$

З урахуванням (18.94) (18.103) перепишемо так:

$$\frac{1}{\hbar}\sqrt{2m(W+U_0)}\cdot d = \pi.$$

Звідси отримаємо, що за енергії електронів

$$W_1 = \frac{h^2}{8m_e d^2} - U_0 \qquad (18.104)$$

будемо мати перший максимум прозорості газу для електронів, як показано на рис. 18.19, лінія *2*. Перший мінімум проходження ями електронами буде за умови

$$\sin^2 k_2 d = 1,$$

або

$$k_2 d = \frac{3}{2}\pi. \qquad (18.105)$$

Звідси із урахуванням (18.94) отримаємо, що за енергії

$$W_2 = \frac{9h^2}{32m_e d^2} - U_0 \qquad (18.106)$$

буде досягнутий перший мінімум проходження електронами газу.

Як бачимо, формули (18.104) і (18.106) повністю збігаються із раніше отриманими формулами (18.84) і (18.87), відповідно.

Співвідношення (18.104) і (18.106) містять дві невідомих величини: розмір атома d і глибину потенціальної ями U_0, Якщо експериментально визначено енергію електронів W_1 і W_2, то розв'язок рівнянь (18.104) і (18.106) для характерного розміру атома дає вираз

$$d = \frac{h\sqrt{5}}{4\sqrt{2m_e(W_2 - W_1)}}, \qquad (18.107)$$

а для глибини потенціальної ями атома

$$U_0 = \frac{4}{5}W_2 - \frac{9}{5}W_1. \qquad (18.108)$$

Таким чином, пояснення ефекту Рамзауера із квантової точки зору не викликає особливих труднощів. Можна вважати, що ефект Рамзауера є одним з наочних підтверджень хвильових властивостей мікрочастинок.

Контрольні запитання

1. У чому суть гіпотези де Бройля і корпускулярно-хвильового дуалізму матерії?
2. Чи у всіх випадках руху матеріальних тіл треба враховувати гіпотезу де Бройля?
3. Що означає координатно-імпульсний метод описання руху частинок?
4. Що визначають співвідношення невизначеностей Гайзенберга?
5. Поясніть поняття хвильової функції та її фізичний зміст. Які властивості хвильової функції?
6. Наведіть розв'язок рівняння Шредінгера для нескінченно глибокої потенціальної ями. Що означає квантування енергії?
7. Наведіть розв'язок рівняння Шредінгера для потенціального бар'єра

кінцевої і нескінченої ширини. Яка природа тунельного ефекту?

8. Поясніть, що таке холодна емісія. Які причини її виникнення?

9. У чому принципова відмінність квантового способу опису розсіювання електронів на атомах від класичного способу?

10. Якою є основна ідея квантового опису розсіювання електронів на атомах?

11. Поясніть, чому взаємодію електрона з атомом моделює потенціальна яма, а не потенціальний горб.

12. Які умови для мінімуму й максимуму розсіювання електронів газом випливають з моделі прямокутної потенціальної ями?

13. За яких значень кінетичної енергії електронів, що налітають на атом, спостерігатиметься гасіння хвиль де Бройля, які пройшли крізь нього?

14. Поясніть ефект Рамзауера на моделі одновимірної потенціальної ями. Яка умова відсутності пружного розсіювання електронів?

15. Яке явище є оптичним аналогом ефекту Рамзауера?

16. Поясніть зміну потенціальної енергії електрона під час прольоту крізь атом.

17. Який вигляд мають розв'язки рівняння Шредінгера за потенціальною ямою, всередині ями і перед нею в задачі про одновимірне розсіювання?

Розділ 19. Основи зонної теорії твердих тіл

19.1. Тверде тіло. Аморфні й кристалічні тверді тіла

Тверде тіло – це агрегатний стан речовини, який характеризується стабільністю форми й характером теплового руху атомів, які здійснюють малі коливання навколо положень рівноваги. Розрізняють кристалічні та аморфні тверді тіла. Кристали характеризуються просторовою періодичністю в розміщенні рівноважних положень атомів. В аморфних тілах атоми коливаються навколо хаотично розміщених точок. Аморфний стан (від грец. amorphous – *безформенний*) – це твердий стан речовини, який характеризується ізотропією властивостей і відсутністю точки плавлення. З підвищенням температури аморфна речовина розм'якає і переходить у рідкий стан поступово. Ці особливості обумовлені відсутністю в речовин в аморфному стані строгої періодичності у розміщенні атомів, іонів, молекул та їх груп на великих відстанях, яка є у кристалах. Кристали мають певну температуру (точку) плавлення.

Стійким станом (з мінімальною внутрішньою енергією) твердого тіла є

кристалічний. З термодинамічної точки зору аморфне тіло перебуває в метастабільному стані та з часом має кристалізуватись. Усі речовини у природі (за винятком рідкого гелію) тверднуть за атмосферного тиску і температури $T > 0$ К.

Дослідження властивостей твердих тіл об'єднались у велику область – фізику твердого тіла, розвиток якої стимулюється потребами техніки. Фізика твердого тіла – джерело нових матеріалів, нові фізичні ідеї, які народжуються у фізиці твердого тіла, проникають в ядерну фізику, астрофізику, біофізику та інші сфери науки. Властивості твердого тіла можна пояснити, виходячи зі знань про його атомно-молекулярну будову і закони руху його атомних частинок (атомів, іонів, молекул), а також субатомних частинок (електрони, атомні ядра).

19.2. Структура кристалів

Кристалами називають тверді тіла, атоми (іони) в яких просторово упорядковані й утворюють тривимірну періодичну структуру, яку називають кристалічною ґраткою. Багато властивостей твердих тіл пояснюється тією періодичністю, з якою розміщені у просторі їх структурні елементи, тому поняття кристалічної ґратки таке важливе. Відстань між атомами у кристалічній ґратці (близько 10^{-10}м = =1 Å) є різною в різних напрямках. У кожній кристалічній ґратці можна виділити основний елемент – *елементарну комірку* (рис. 19.1).

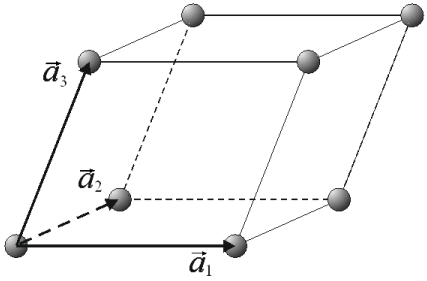


Рис. 19.1. *Елементарна комірка*

Елементарна комірка – це той мінімальний за об'ємом елемент ґратки (групи атомів), послідовним переміщенням (трансляцією) якого можна побудувати всю ґратку кристала. Отже, для кристалічної ґратки характерна періодичність. Виділяючи таку комірку, вдається описати положення атомів чи іонів у речовині, а значить, з її допомогою можна характеризувати структуру кристалів. Положення, які займають атоми чи іони, відповідають точкам (вузлам) ґратки. Розглядаючи структуру кристалів, будемо абстрагуватись від теплового руху атомів і його характера біля вузлів просторової ґратки. Амплітуда цих

коливань мала (десь близько 0,1 міжатомної відстані), тому тепловий рух в середньому не порушує симетрії кристалів.

Таким чином, для описання структури кристала досить задати його елементарну комірку. Елементарну комірку, очевидно, геометрично визначають заданням трьох основних векторів $\vec{a}_1, \vec{a}_2, \vec{a}_3$ (рис. 19.1). У напівпровідниках елементарна комірка складається лише з декількох атомів. Просторово вона дуже мала і може бути розміщена всередині куба зі стороною $5 \cdot 10^{-10}$ м $= 5$ Å.

З-поміж усіх можливих видів ґраток можна виділити кубічну (одну з найпростіших) ґратку, яка має декілька різновидів:

1. *Проста кубічна ґратка*. В кожній вершині такої ґратки розміщений один атом, який належить одночасно восьми сусіднім елементарним коміркам (рис. 19.2, *а*), яких є вісім. У такій формі кристалізується тільки полоній Po.

2. *Кубічна об'ємно центрована ґратка.* Тут, крім атомів у вершинах кубів є ще один атом у центрі (рис. 9.2, *б*). До такого типу належать кристалічні ґратки молібдену і вольфраму.

3. *Кубічна граніцентрована ґратка.* Має шість атомів у центрах граней і, крім того, вісім атомів у вершинах кубів (рис. 19.2, *в*). У такій формі кристалізуються алюміній і низка інших хімічних елементів.

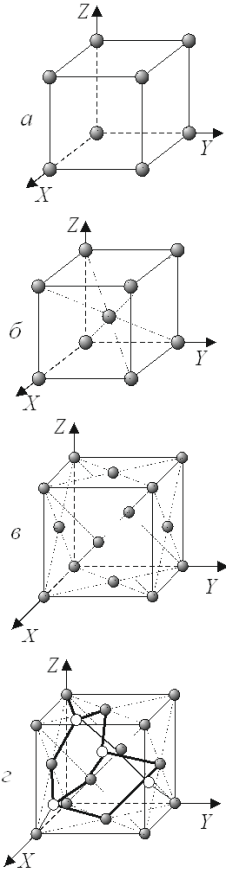


Рис. 19.2. Елементарні комірки кубічних ґраток: *а* – проста кубічна; *б* – кубічна об'ємно центрована; *в* – кубічна граніцентрована; *г* – тип алмазу та арсеніду галію

4. *Ґратка типу алмазу.* Її можна розглядати як дві вкладені одна в одну

кубічні граніцентровані ґратки, які зміщені на відстань чверті діагоналі куба (рис. 19.2, *г*, темними і світлими кружками зображено одні й ті самі атоми). У таку форму кристалізуються вуглець, кремній, германій і сіра модифікація олова.

5. *Ґратка типу арсеніду галію.* Її отримують із ґратки типу алмазу в тому випадку, коли атоми галію збігаються з вузлами однієї граніцентрованої ґратки, а атоми миш'яку – з вузлами іншої (на рис. 19.2, *г*, атоми Ga позначено світлими кружками, а атоми As – темними).

Щоб знайти кількість точок, які належать одній елементарній комірці, треба розглянути граничні точки комірки. Якщо говорити про елементарну комірку простої кубічної ґратки, то тут у вершинах куба є вісім точок, однак кожна із цих восьми точок одночасно належить восьми кубам, які сходяться у вибраній вершині; таким чином, отримуємо $8 \cdot \left(\frac{1}{8}\right) = 1$ точку, яка припадає на елементарну комірку. Для ґратки об'ємно центрованої маємо вісім точок у вершинах, кожна з яких припадає на вісім кубів, і, крім того, всередині кожного куба міститься одна точка. Звідси випливає, що в одній елементарній комірці міститься $8/8 + 1/1 = 2$ точки. Для ґратки граніцентрованої маємо шість точок, які розміщевані на гранях комірок, що торкаються, а також по одній точці у вершинах куба. У результаті знаходимо, що кількість точок, які належать елементарній комірці, становить $8/8 + 6/2 = 4$.

Структуру кристалів досліджують, головним чином, такими методами: дифракції рентгенівських променів (рентгенографія), дифракції електронів (електронографія), дифракції нейтронів (нейтронографія). Дифракція променів або частинок з відповідною довжиною хвилі відбувається на просторовій кристалічній ґратці.

Спрямовані на кристал промені або частинки розсіюються атомами, які розміщені у вузлах кристалічної ґратки. Хвилі, розсіяні окремими атомами, інтерферують між собою. У результаті тільки в деяких напрямках у просторі поширюються підсилені взаємною інтерференцією хвилі (див. рис. 13.28).

Умову появи дифракційного максимуму під час розсіювання рентгенівських променів виражають рівнянням Вульфа–Брегга (13.33). Якщо відомі кути падіння θ променів на кристал (або ж кути, під якими реєструють дифраговані промені) і

довжина хвилі λ випромінювання, можна визначити міжплощинну відстань d у цій системі паралельних атомних площин кристала. Змінюючи кут падіння первинного пучка паралельних променів на кристал можна досягти виконання умови (13.33) для інших сукупностей паралельних атомних площин і визначити міжплощинну відстань для них. У результаті вдається визначити тип ґратки (її симетрію) та її параметри. Рівняння (13.33) справедливе для дифракції на кристалічній ґратці всіх випромінювань з відповідною довжиною хвилі, хоча техніка експерименту у випадку використання рентгенівських променів або прискорених електричним полем електронів зовсім різна.

19.3. Основні типи міжатомної взаємодії у твердих тілах

Сили притягання між частинками, з яких складається тверде тіло, настільки великі, що атоми твердого тіла не можуть рухатись вільно один відносно одного. Утім, ці атоми зовсім не нерухомі. Кожен із них весь час коливається під дією пружних сил, які діють між частинками.

У 1912 р. Фрідріх, Кніппінг і Лауе в дослідах з дифракції рентгенівських променів показали, що кристали справді побудовані з атомів, які симетрично розміщені один відносно одного. Брегг у 1913 р. визначив кристалічні структури KCl, NaCl, KBr і KI в експериментах з дифракції рентгенівських променів. Ці дослідники вперше пояснили фізичний зміст геометричної правильності кристалічних твердих тіл, яка відповідає їхній атомній будові. Структура і жорсткість твердих тіл визначаються відмінністю в електростатичних силах, які зв’язують разом атоми, що утворюють тіло.

Іонний зв’язок та іонні кристали

Типовим представником іонних кристалів є кам’яна сіль і зазвичай іонний зв’язок у кристалах розглядають на цьому прикладі. Електронною структурою ізольованого атома натрію є ${}_{11}\mathrm{Na} - 1s^2 2s^2 2p^6 3s^1$, атома хлору ${}_{17}\mathrm{Cl} - 1s^2 2s^2 2p^6 3s^2 3p^5$. Під час зближення цих атомів, наприклад у процесі конденсації парів, виникає ймовірність переходів електронів зі станів, які належать одному атому, у стан, який належить другому атому (а фактично в цьому випадку вже утворюється система зі спільними станами).

Природно, що перехід зі стану $3s$ електрона натрію у частково порожній стан $3p$ хлору енергетично вигідний,

оскільки при цьому обидва атоми набувають на зовнішньо валентній електронній оболонці структури стійкого октету (вісім електронів). Атом натрію при цьому перетворюється на позитивний іон Na^+, атом хлору – на негативний іон Cl^-. Між ними виникають кулонівські сили притягання, які на близьких відстанях врівноважуються силами відштовхування. Кулонівська взаємодія існує між всіма іонами, які входять у цю конденсовану систему.

У результаті встановлюється деякий рівноважний для таких зовнішніх умов розподіл іонів у просторі, який утворює ґратку іонного кристала (рис. 19.3).

Найбільш характерні властивості іонних кристалів такі:

– висока (до 10^3 K) температура плавлення;

– відсутність електронної і наявність іонної провідності;

– велика механічна міцність, яку, однак, практично не можна використати через велику крихкість іонних кристалів;

– іонні кристали – це діамагнетики (магнітна проникність $\mu < 1$, магнітна сприйнятливість $\chi < 0$), якщо вони не містять іонів феромагнітних елементів.

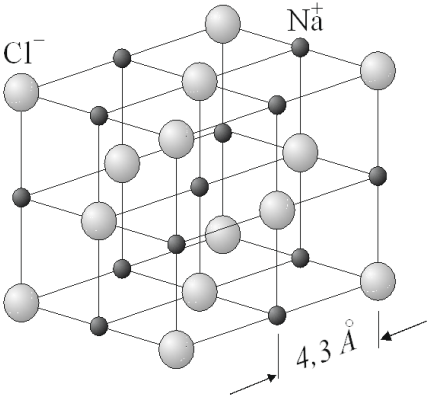


Рис. 19.3. ґратка іонного кристала

Ковалентний зв’язок

Ковалентний (гомеополярний) зв’язок виникає у разі зближення атомів унаслідок перекриття хвильових функцій електронів, що призводить до виникнення виключно квантової обмінної взаємодії між електронами. Природа обмінної взаємодії обумовлена тотожністю електронів, тобто принципово не можна розрізнити електрони один від одного. Класичним прикладом обмінної взаємодії є теорія гомеополярного зв’язку в молекулі водню. Як показує детальний квантово-механічний розгляд, у випадку, коли спіни в системі двох електронів антипаралельні, енергія системи, яка складається зі «звичайної» кулонівської енергії та кулонівської енергії обмінної взаємодії, має мінімум за деякої відстані

між центрами ядер атомів водню, які входять до складу молекули. Це означає, що утворюється стійка система, у цьому випадку – молекула водню H_2. Це приклад найсильнішого хімічного зв’язку у природі, який обумовлює той факт, що атомарного водню у природних умовах не існує.

Увівши поняття ковалентного зв’язку, розглянемо конкретно кристал кремнію. Атом кремнію $_{14}$Si має такий розподіл електронів за станами: $1s^2 2s^2 2p^6 3s^2 3p^2$. У $3p$-стані є лише два електрони й чотири вакантних місця. Кожний атом кремнію у твердому стані намагається притягнути по одному електрону від кожного з чотирьох найближчих сусідніх атомів, щоб дозаповнити $3p$-оболонку. Виникає алмазоподібний тип кристалічної ґратки.

На рис. 19.4 показано двовимірне зображення зв’язків у ґратці кремнію. Тут у вузлах ґратки міститься іон кремнію, якому належать чотири валентних електрони $3s^2 3p^2$. Валентні електрони забезпечують ковалентний зв’язок (на рис. 19.4 їх зображено у вигляді чорних точок).

Кристали з ковалентним зв’язком часто називають валентними кристалами, основні властивості яких такі:

– висока температура плавлення (10^3 К);

– високі твердість і пружність, велика крихкість;

– за електричними властивостями вони належать до напівпровідників і діелектриків (у деяких металах теж трапляється значна частка ковалентного зв’язку).

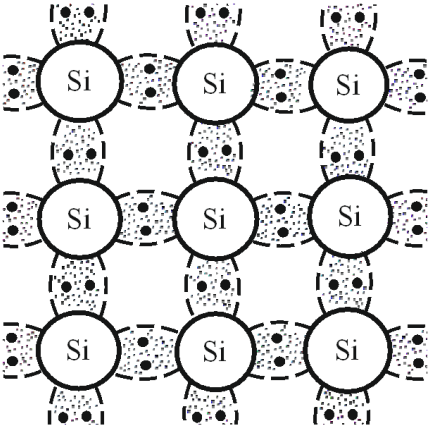


Рис. 19.4. Двовимірне зображення зв’язків у ґратці кремнію

Міжатомна взаємодія в металах. Властивості металів

Для металів характерною є наявність у кристалах майже вільних (слабо зв’язаних з іонами ґратки) електронів. Цим пояснюється високі електро- і теплопровідність металів, типовий металічний блиск, велика пластичність. Металічний зв’язок – це

граничний випадок ковалентного зв'язку. Обмінна взаємодія, яка здійснюється валентними електронами, у цьому випадку відбувається в межах всього кристала, а наявність електронного газу (вільних електронів) визначає ненаправлений характер металічного зв'язку. Цим, зокрема, обумовлена висока пластичність металів: переміщення атомів у ґратці під дією зовнішніх сил менше змінює його енергетичний стан, ніж у випадку, коли за такого переміщення розриваються направлені іонні або ковалентні зв'язки. З точки зору якісних уявлень про утворення стійких октетів на зовнішніх електронних оболонках атомів виникнення металічного зв'язку обумовлене «позбавленням» атомами всіх валентних електронів, у результаті чого кожний атом набуває електронної структури попереднього інертного газу. Наприклад, атом натрію $1s^2 2s^2 2p^6 3s^1$, віддавши у склад електронного газу свій єдиний валентний $3s$-електрон, набуває електронної структури неону $1s^2 2s^2 2p^6$. Це приводить до енергетичного виграшу і до виникнення сил притягання між атомними залишками; кулонівська взаємодія з атомними залишками утримує колективізовані електрони в об'ємі кристала.

Таким чином, структура металу можна уявити як іонний позитивний кістяк, занурений в електронний газ, який компенсує сили відштовхування між іонами і пов'язує їх внаслідок обмінної взаємодії у тверде тіло (рис.19.10). Більшість металів мають щільно упаковану ґратку: кубічну гранецентровану (рис. 19.2, *в*) і гексагональну (рис. 19.5).

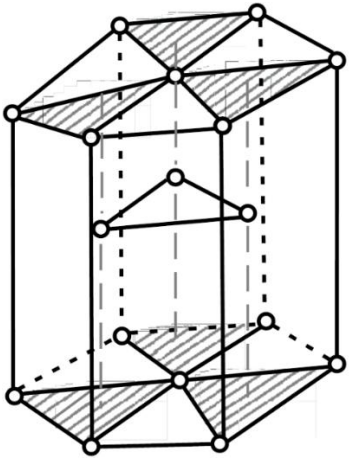
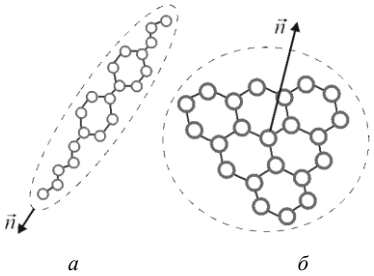


Рис. 19.5. Гексагональна комірка металу (світлі кружечки – іони металу)

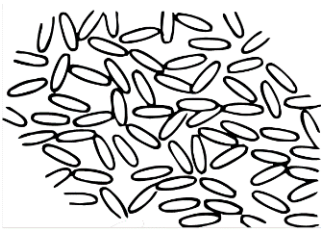
Перехідні тугоплавкі метали, такі як Cr, Mo, W, V, Ta, мають кубічну об'ємно центровану ґратку. Кубічну об'ємно центровану ґратку за температури до 910 ℃ має і залізо. У них, крім звичайного металічного

зв'язку у кристалах, також діє ковалентний зв'язок, обумовлений електронами внутрішніх незаповнених оболонок. Більшість металів – парамагнетики ($\mu \geq 1,\ \chi > 0$). Перехідні метали Fe, Co, Ni та Gd – феромагнетики ($\mu \gg 1$).



а *б*

Рис. 19.6. Молекули рідких кристалів



в

Рис. 19.6. Молекули рідких кристалів

19.4. Рідкі кристали. Сили Ван дер Ваальса

Рідкі кристали – це особливий стан деяких органічних речовин, в якому вони мають реологічні властивості рідини – *плинність*, але зберігають певну впорядкованість у розміщенні молекул і анізотропію ряду фізичних властивостей, яка характерна для твердих кристалів. Рідкі кристали називають ще *мезофазами* (від грец. mezos – середнє).

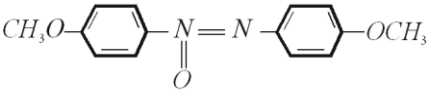
Рідкі кристали відкрив у 1889 р. австрійський ботанік Ф. Рейніцер, який синтезував незвичайні кристали. Під час їх нагрівання отримували рідину, яка залежно від температури була то каламутною, то прозорою, то набувала синюватого відтінку. Німецький фізик Леман почав систематичне вивчення таких речовин і встановив, що відкрито особливий стан, який належить багатьом органічним речовинам. Кількість хімічних сполук, для яких знайдені рідкі кристали, складають декілька тисяч.

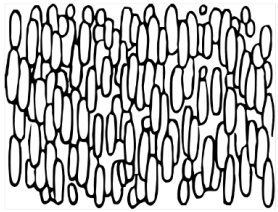
Рідкі кристали утворюються під час нагрівання деяких твердих кристалів (мезогенних): спочатку відбувається фазовий перехід у рідкий кристал, а подальше плавлення рідкого кристала – у звичайну ізотропну рідину. Кожна мезофаза існує в певному температурному інтервалі (термотропні рідкі кристали). Рідкі кристали утворюють речовини, молекули яких мають видовжену палицеподібну форму, часто з чергуванням лінійчастих (рис. 19.6, *а*) і циклічних плоских атомних груп (рис. 19.6, *б*), які паралельні одні одним. На рис. 19.6, *в*, показано

розміщення таких молекул у рідині. Така форма молекул визначає приблизну паралельність їх взаємної укладки, що є основною ознакою рідкого кристала.

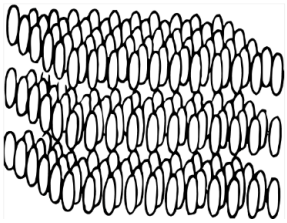
Розрізняють три основні типи рідких кристалів: нематики (рис. 19.7, *а*), смектики (рис. 19.7, *б*) і холерики (рис.19.7, *в*). Найменшу впорядкованість мають нематичні рідкі кристали (назва «нематичний» походить від грец. νημα – нитка). У рідких кристалах у мікроскопі видно тонкі рухомі нитки, які є дефектами структури.

В ідеальному рідкому кристалі таких ниток немає. В нематиках молекули паралельні, але зсунуті вздовж своїх осей одна відносно одної на довільні відстані (рис. 19.7, *а*). У смектичних рідких кристалах молекули паралельні одна одній і розміщені шарами (рис. 19.7, *б*). Структура холестеричних рідких кристалів відрізняється від них додатковим закручуванням молекул у напрямку, який перпендикулярний до їхніх довгих осей (рис. 19.7, *в*). Крок такої спіральної «надструктури» може бути досить великим і досягати декількох мікрометрів. Прикладом нематичного рідкого кристала є параазоксианізол, який існує в мезофазі в інтервалі 116–136 °С. Його молекула має таку структуру:

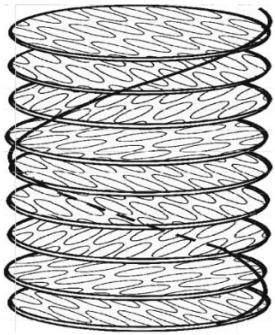




а



б



в

Рис. 19.7. Типи рідких кристалів:
а – нематики; б – смектики; в – холерики

Знайдені смектичні рідкі кристали деяких сполук і їх сумішей

існують в інтервалі від -40 до $+80$ °С, що розширює можливості їх практичного використання. Прикладом холестеричних рідких кристалів є ефіри холестерину. Розглянемо тепер сили, які діють у нематичній рідині. Ці сили мають електричне походження. Сила притягання виникає між двома атомами або молекулами, які самі по собі є електрично нейтральними. Вперше такі сили став враховувати голландський фізик Я. Д. Ван дер Ваальс (1873 р.) для пояснення властивостей реальних газів і рідин, і з того часу вони отримали назву сил Ван дер Ваальса. Він припустив, що на малих відстанях r між молекулами діють сили відштовхування, які зі збільшенням відстані замінюються силами притягання. Спробуємо розібратись, як це відбувається. Уявімо, що через якусь випадкову причину в атомі відбулось зміщення негативно зарядженої електронної хмарини відносно позитивно зарядженого ядра, тобто виникає електричний диполь (рис. 19.8, *а*). Біля атома-диполя виникає електричне поле. Напруженість цього поля швидко зменшується з віддаленням від атома, але біля атома воно достатньо велике. Якщо в поле атома I $\vec{E}_1$ потрапляє нейтральний атом II (рис. 19.8, *б*), то електричне поле атома I має змістити заряди електронів і ядра атома II так, як показано на рис. 19.8, *б*, тобто атом II поляризується. Поле атома II $\vec{E}_2$ підтримує розділення зарядів в атомі I. Із рис. 19.8, *б*, видно, що різнойменно заряджені частинки атомів повинні притягуватись, а отже, й атоми в цілому повинні притягуватись один до одного. Зі зближенням атомів між ними починають діяти сили відштовхування. На відстані, що приблизно дорівнює розміру атома, сили взаємодії між атомами дорівнюють нулю.

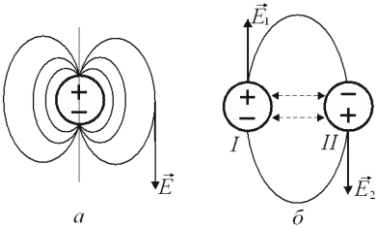


Рис. 19.8. Виникнення сил притягання між молекулами в рідких кристалах

Такі самі міркування можна навести й у відношенні двох молекул, які складаються з декількох десятків атомів. Нейтральні молекули мають притягуватись одна до одної внаслідок виникнення електричних диполів. Сила притягання між двома полярними молекулами максимальна в тому випадку, коли їхні дипольні моменти розміщуються вздовж однієї лінії (рис. 19.9) і залежить від їхньої взаємної орієнтації. Різнойменні

заряди b та c взаємодіють сильніше, ніж однойменні заряди a та c (а також b й d), що перебувають на більшій відстані один від одного. Сила взаємодії таких молекул $F \sim r^{-7}$ зменшується з відстанню значно швидше, ніж кулонівська сила взаємодії заряджених частинок $F_{\text{кул}} \sim r^{-2}$. Через ці причини більша частина атомів однієї молекули прагне наблизитись до атомів іншої молекули, оскільки тільки в цьому випадку сили взаємодії між молекулами перетворюються на нуль. Утім, така ситуація можлива тільки тоді, коли довгі осі молекул або площини молекул паралельні одні до одних. Так виникає певний порядок в орієнтації молекул і з'являється виділений напрямок. Цей напрямок можна характеризувати одиничним вектором $\vec{n}$ (рис. 19.6). Зрозуміло, що таке паралельне розміщення виділених осей (або площин) молекул можливе тільки за досить низької температури, коли теплові поштовхи не наскільки сильні, щоб зруйнувати орієнтаційний порядок у системі молекул. З підвищенням температури обов'язково настає момент, коли хаотичний тепловий рух стає переважаючим і нематичний порядок руйнується.



Рис. 19.9. Сила притягання між двома полярними молекулами максимальна в тому випадку, коли їхні дипольні моменти розміщуються вздовж однієї лінії

Таким чином, система таких особливих молекул може мати два стани: звичайний (ізотропний) рідкий – за високих температур (рис. 19.6, *в*) й анізотропний рідкий – за низьких температур (рис. 19.7).

Теорія рідких кристалів ґрунтується на поєднанні принципів симетрії кристалів (щодо найближчих сусідів) із законами статистичної фізики.

Рідкі кристали мають анізотропію пружності, електропровідності, магнітної сприйнятливості та діелектричної проникності, оптичну анізотропію, сегнетоелектричні властивості й інше. Наприклад, холестеричні рідкі кристали мають велику оптичну активність (у $10^2 - 10^3$ більшу, ніж в органічних речовинах і твердих кристалах).

Найбільші враження створюють оптичні властивості рідких кристалів, які зробили їх такими популярними. Такі оптичні властивості дуже незвичні для рідин і властиві твердому кристалу. Зрозуміло, у чому тут справа: в

нематичній рідині, як і у кристалі, є особливі напрямки. Наявність цих напрямків обумовлює ряд чудових властивостей кристалів і, зокрема, оптичні властивості, тому такі особливі напрямки називають оптичними осями. Напрямок оптичних осей в рідкому кристалі збігається з вектором $\vec{n}$ (рис. 19.6, *а*, *б*). У твердих кристалах оптичні осі жорстко закріплені, а в рідких кристалах напрямки оптичних осей можна змінювати за допомогою найрізноманітніших впливів, зокрема й електричними або магнітними полями.

Ефект зміни напрямку орієнтації молекул в нематичній рідині під дією електричного поля спостерігав В. Фредерікс. Назва цього ефекту тепер носить його ім'я. Користуючись популярними електричними годинниками і калькуляторами з екранами на рідких кристалах, ми спостерігаємо саме явище ефекту Фредерікса. Рідкі кристали мають широке практичне застосування, особливо в системах обробки і відображення інформації, в яких використовують електрооптичні властивості рідких кристалів.

19.5. Класифікація твердих тіл за величиною питомої електропровідності та її залежності від температури

Класифікацію твердих тіл за величною питомої електропровідності наведено у розд. 2 (підрозд. 2.1). Питомий опір у срібла менший, ніж у тефлона, у 10^{24} (!) раз. Жодна фізична характеристика не змінюється так сильно під час переходу від речовини до речовини, як питомий опір. Виникає запитання, чому ж опір цих речовин у твердому стані відрізняється так сильно. Більше того, атоми однієї й тієї самої речовини, наприклад вуглецю С, залежно від того, в яку кристалічну ґратку твердого тіла вони з'єднанні, можуть утворювати і дуже хороший провідник – графіт, і прекрасний ізолятор – алмаз. Олово, здавалось би, є типовим металом. Утім, якщо всім знайоме біле блискуче олово довго тримати за температури −30 ℃, то його кристалічна структура зміниться, а зовнішньо воно посіріє. І це олово (його так і називають – «сіре олово») є напівпровідником, а свинець завжди є металом.

Якщо почати змішувати різні елементи, то картина зовсім ускладнюється. Візьмемо, наприклад, і сплавим

два метали – індій та сурму у пропорції один до одного. Отримаємо напівпровідник InSb, який широко застосовують у техніці.

Приклади з атомами вуглецю й олова наводять на таку думку: чи тверде тіло буде металом або діелектриком залежить не тільки і, можливо, навіть не настільки від властивостей атомів, які утворюють кристал, скільки від того, які відношення складаються у кожного атома з сусідами, тобто від типу зв'язків, котрі об'єднують атоми у кристалічну ґратку твердого тіла. На рис. 19.10 схематично зображено кристалічну ґратку металу.

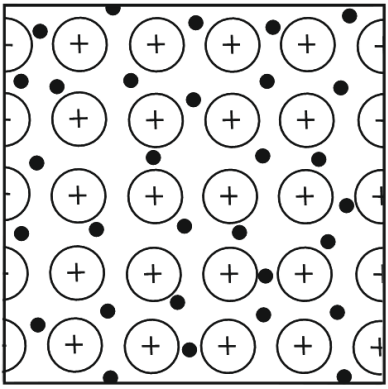


Рис. 19.10. Схематично зображена кристалічна ґратка металу

Правильна ґратка позитивно заряджених іонів занурена в «газ» з вільних електронів (темні кружки), не пов'язаних жорсткими зв'язками з окремими іонами. Електрони, як зазвичай говорять, колективізовані кристалом і можуть вільно переміщуватись під дією зовнішнього електричного поля. У кожному кубічному сантиметрі металу міститься близько 10^{22} вільних носіїв струму – електронів, тому метали є чудовими провідниками.

На рис. 19.4 схематично зображено ґратку (фактично діелектрика) кристалічного кремнію. Кожний атом пов'язаний з найближчими сусідами міцними електронними зв'язками. Щоб розірвати електронні зв'язки між атомами ґратки і створити в такому кристалі вільні електрони, які здатні створювати електричний струм, необхідно затратити енергію. Звідки може взятись ця енергія? Якщо її величина не дуже велика, а кристал досить сильно нагріти, то електронні зв'язки можуть бути розірвані внаслідок теплових коливань ґратки. Світло з відповідною довжиною хвилі теж може розірвати електронні зв'язки і створити у кристалі вільні електрони, але за досить низької температури (і без світла) кристал, який показаний на рис. 19.4, є ідеальним ізолятором – вільних носіїв заряду в ньому немає.

Величину енергії, яка потрібна для того, щоб розірвати електронні зв'язки між атомами і створити в неметалічному кристалі вільні електрони, позначають символом W_g і називають енергією активації. Вона є важливою характеристикою кристала. Значення W_g залежить від структури кристалічної ґратки, від властивостей атомів, які її складають, і у різних матеріалів її значення лежить у межах від нуля до декількох десятків електронвольт. Якщо значення W_g велике, то нагрівання навіть до дуже високої температури не створює у кристалі скільки-небудь помітної кількості вільних електронів. За досить великого значення W_g кристал може навіть розплавитись раніше, ніж у ньому внаслідок теплового руху виникатимуть вільні носії заряду. Такі матеріали з великим значенням W_g є типовими *діелектриками*.

Якщо значення W_g невелике, нагрівання навіть не до дуже високих температур призведе до розриву значної кількості електронних зв'язків і появи відносно великої кількості вільних носіїв заряду (рис. 19.11). Кристал набуває здатності проводити електричний струм, тому матеріали з невеликим значенням W_g будуть займати «начебто середнє місце між провідниками і непровідниками», адже «здатність проводити електрику» (провідність) пропорційна концентрації вільних носіїв заряду. У провідниках (металах) концентрація вільних носіїв заряду дуже велика. Це відповідає ситуації, коли всі електронні зв'язки розірвані і всі валентні електрони вільні. У «непровідниках» (діелектриках) вільних носіїв заряду фактично немає. Зрозуміло, що питомий опір матеріалу з невеликим значенням W_g може дуже сильно змінюватись із температурою. В таких матеріалах концентрація вільних носіїв заряду пропорційна кількості розірваних електронних зв'язків буде дуже різко зростати зі зростанням температури, а питомий опір різко зменшуватись. Ці матеріали називають *напівпровідниками*.

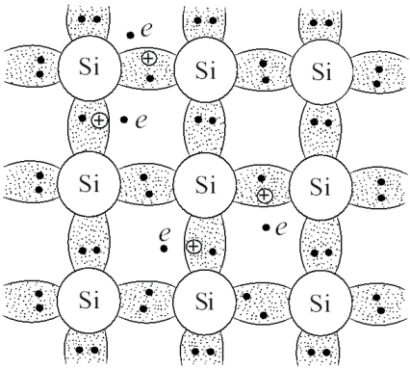


Рис. 19.11. Виникнення носіїв заряду в напівпровіднику: ● – електрони, ⊕ – дірки

Зазначимо, що провідність напівпровідників збільшується не тільки під час нагрівання, але і під час освітлення чи опромінення ядерними частинками; вона змінюється під час прикладання електричних і магнітних полів, при змінні зовнішнього тиску і т.п.

Отже, ***напівпровідники*** – *це речовини, які за кімнатної температури мають питомий опір в інтервалі від* 10^8 *до* 10^{-6} *Ом·м (провідність від* 10^{-8} *до* 10^6 $(\text{Ом}\cdot\text{м})^{-1}$*), який у сильному ступені залежить від зовнішніх умов (наприклад, від температури)*:

$$\rho \sim \exp\left\{\frac{W_g}{2kT}\right\}. \qquad (19.1)$$

Таблиця 19.1. величина енергії W_g для найбільш поширених напівпровідників

Напівпровідник	InSb	Ge	Si
W_g, еВ	0,17	0,72	1,1
Напівпровідник	GaAs	GaAlAs	SiC
W_g, еВ	1,4	1,4 - 2,17	3

Більш коротко, напівпровідники – це неметалічні матеріали з відносно невеликою величиною енергії W_g. Значення W_g для найбільш поширених напівпровідників за кімнатної температури наведено в табл.19.1. Ці напівпровідники широко застосовують на практиці. Так, із InSb виготовляють приймачі інфрачервоного випромінювання ($\lambda \cong 10$ мкм), які лежать в основі роботи приладів нічного бачення. Із германію було виготовлено перші транзистори й діоди. Кремній є основним матеріалом сучасної напівпровідникової електроніки. Арсенід галію GaAs використовують для створення складних сполук (наприклад, GaAlAs), які використовують для створення напівпровідникових світлодіодів і лазерів. З карбіду кремнію SiC створено найбільш надійні і стабільні світлодіоди, здатні працювати за дуже високої температури.

Тут названо тільки декілька найбільш важливих напівпровідникових матеріалів. Всього ж синтезовано, вивчено і використовуються на практиці сотні напівпровідникових сполук.

У металах концентрація електронів не залежить від температури. Навіть за абсолютного нуля «вільні» електрони залишаються вільними і зберігають здатність проводити струм. У хімічно чистих металах питомий опір з підвищенням температури збільшується пропорційно абсолютній температурі:

$$\rho = \rho_0(1 + \alpha t) = \rho_0\, T/T_0, \quad (19.2)$$

де ρ_0 – питомий опір певного металу за 0 °C; α – термічний коефіцієнт опору, який дорівнює $1/273$ K^{-1}, $T_0 = 273$ K.

Таким чином, метали – це такі речовини, які проводять електричний струм за будь-якої температури.

19.6. Енергетичні зони у кристалах

Чому одні тіла є провідниками, а інші напівпровідниками та ізоляторами – на це питання не можна відповісти у межах моделі газу вільних електронів. Необхідно врахувати взаємодію атомів між собою і з електронами.

Якісно задачу розв'язують так. З окремих атомів будують кристалічну ґратку і з'ясовують, як впливають сусідні атоми кристала на поведінку електронів у кожному окремому атомі, тобто досліджують вплив поля сусідніх атомів на енергетичні рівні електронів в атомі. Розмістимо N атомів натрію у вигляді просторової гратки, яка властива кристалу натрію, але на великих відстанях один від одного, щоб взаємодією між ними можна було знехтувати. У цьому випадку енергетичні стани електронів у кожному атомі можна вважати такими, як і в окремо взятому ізольованому атомі. На рис. 19.12 зображено енергетичну схему двох таких атомів. Кожний із них зображений у вигляді потенціальної ями, всередині якої проведені рівні 1s, 2s, 2p, 3s … . Рівні 1s, 2s і 2p у натрію заповнені повністю електронами, рівень 3s заповнений наполовину, а рівні, які розміщені вище 3s, – вільні.

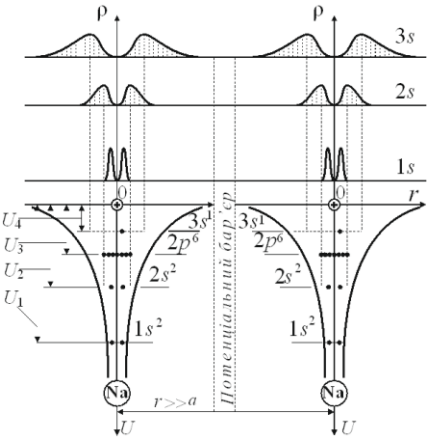


Рис. 19.12. Енергетична схема двох атомів натрію

Ізольовані атоми відокремлені один від одного потенціальними бар'єрами завтовшки $r \gg a$, де a – стала ґратки кристала натрію. Для електронів рівня $1s$ – висота бар'єра U_1, для електронів $2s - U_2$, для електронів $2p - U_3$. Потенціальний бар'єр перешкоджає вільному переходу електронів від одного атома до іншого. У верхній частині рис. 19.12 показано якісну картину розподілу густини ймовірності $\rho = 4\pi r^2 \psi\psi^*$ виявлення електронів на певній відстані від ядра. Максимуми кривих відповідають приблизно положенню боровських орбіт для цих електронів. Тепер будемо повільно

зближувати атоми ґратки, не порушуючи її симетрію. На рис. 19.13 показано картину, яка відповідає такому зближенню. Із цього рисунка видно, що потенціальні криві, які відділяють сусідні атоми (показані пунктиром), частково накладаються одна на одну і дають результуючі криві, які проходять нижче нульового рівня 00 (на рис. 19.12 їх показано суцільними лініями).

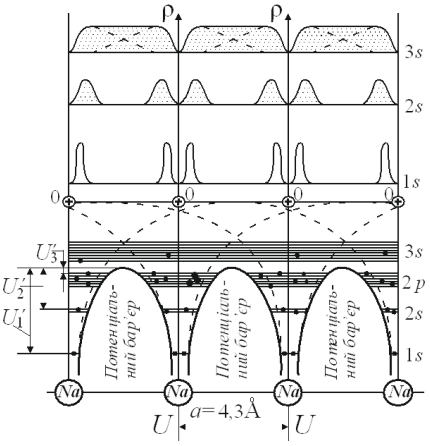


Рис. 19.13. Якісна картина розподілу густини ймовірності $\rho = 4\pi r^2 \psi\psi^*$ *виявлення електронів на певній відстані від ядра*

Отже, зближення атомів зменшує товщину потенціального бар'єра до значення близько a і зменшує його висоту. Для електронів рівня $1s$ висота бар'єра стає U_1', для електронів $2s - U_2'$, для електронів $2p - U_3'$.

Для електронів $3s$-рівня висота бар'єра виявляється нижчою від початкового положення рівня $3s$ в атомі натрію, тому валентні електрони цього рівня отримують можливість фактично без перешкод переходити від одного атома до іншого. Про це свідчить і характер електронних хмар валентних електронів, вони перекриваються настільки сильно, що створюють результуючу хмару рівномірної густини ρ (рис. 19.13). Це відповідає стану повної їх колективізації у ґратці. Такі колективні електрони називають вільними, а їх сукупність – електронним газом.

Унаслідок різкого зменшення висоти і товщини потенціального бар'єра під час зближення атомів можливість переміщення у кристалі отримують не лише валентні електрони, але й електрони, розміщенні на інших рівнях атомів. Переміщення відбувається шляхом тунельного переходу електронів через бар'єри, що відокремлюють сусідні атоми. Якщо колективні електрони зберігають досить сильний зв'язок з атомами, то їх потенціальну енергію можна подати у такому вигляді:

$$U = U_a + \delta U,$$

де U_a – потенціальна енергія електрона в ізольованому атомі. Для кристала вона є

періодичною функцією з періодом, що дорівнює постійній ґратки, оскільки енергія електрона повторюється під час переходу його від одного атома до іншого (рис. 19.13); δU – доданок, який враховує вплив сусідніх атомів на цю енергію.

Якщо знехтувати доданком δU, тобто розглядати нульове наближення, то в ролі хвильової функції та енергії електрона у кристалі слід взяти хвильову функцію ψ_a та енергію $W_a(n,l)$ електрона в ізольованому атомі.

Відмінність між кристалом і атомом у цьому випадку полягає в такому: у той час як в ізольованому атомі цей енергетичний рівень $W_a(n,l)$ є єдиним, то у кристалі, що складається з N атомів, він повторюється N разів. Кожний рівень ізольованого атома у кристалі стає N-кратно виродженим.

Якщо врахувати доданок δU у потенціальній енергії, то в міру зближення ізольованих атомів та утворення з них кристалічної ґратки кожний атом потрапляє у поле своїх сусідів, яке зростає. Така взаємодія приводить до зняття виродження, тому кожний енергетичний рівень, невироджений в ізольованому атомі, розщеплюється на N близько розміщених один від одного підрівнів, що утворюють енергетичну зону. На кожному підрівні згідно з принципом Паулі може перебувати не більше двох електронів з антипаралельними спінами. Якщо енергетичний рівень мав в атомі $(2l + 1)$-кратне виродження, то відповідна йому енергетична зона буде складатися із $N(2l + 1)$ підрівнів. Так, s-рівень дає s-зону, яка складається з N підрівнів і здатна вмістити $2N$ електронів; p-рівень дає p-зону, яка складається із $3N$ підрівнів і здатна вмістити $6N$ електронів і т.д. Найбільший вплив поле ґратки чинить на зовнішні валентні електрони атомів, тому стани цих електронів у кристалі відчувають найбільші зміни, а енергетичні зони, утворені з енергетичних рівнів цих електронів, виявляються найбільш широкими.

Внутрішні ж електрони, які сильно зв’язані з ядром, відчувають лише незначні збурення від інших атомів, унаслідок чого їх енергетичні рівні у кристалі залишаються фактично такими ж вузькими, як і в ізольованих атомах. На рис. 19.14 зображено схематичну картину утворення енергетичних зон у кристалі з дискретних атомних рівнів.

Отже, кожному енергетичному рівню ізольованого атома у кристалі відповідає зона дозволених рівнів: рівню $1s$ – зона $1s$, рівню $2p$ – зона $2p$ і т.д. Зони дозволених енергій розділені областями заборонених енергій – забороненими зонами. Зі збільшенням енергії електрона в атомі ширина дозволених зон збільшується, ширина заборонених – зменшується.

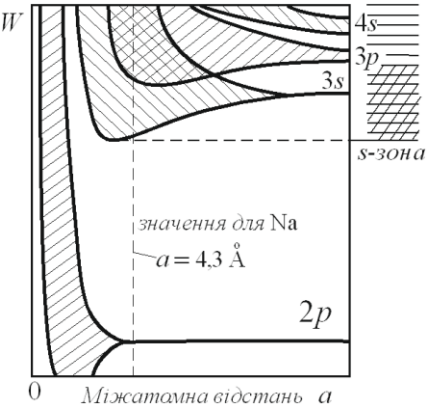


Рис. 19.14. Утворення енергетичних зон у кристалі з дискретних рівнів під час зближення атомів

Дозволені енергетичні зони у твердому тілі можуть бути різним чином заповненими електронами. У граничних випадках вони можуть бути цілком заповнені або цілком вільні. Електрони у твердих тілах можуть переходити з однієї дозволеної зони в іншу. Для переходу електрона з нижньої зони в сусідню верхню зону потрібно затратити енергію, що дорівнює ширині забороненої зони, яка розміщена між ними.

Для внутрішніх переходів електронів у межах зони потрібна дуже мала енергія, тому що відстані між сусідніми енергетичними рівнями в зоні дорівнюють приблизно 10^{-22} еВ. Наприклад, для цього достатньо енергії 10^{-4} ... 10^{-8} еВ, що отримує електрони в металі під дією електричного поля на довжині вільного пробігу за звичайних різниць потенціалів. Під дією спеціального збудження електронам може бути надано енергію, достатню як для внутрішньозонних, так і для міжзонних переходів.

Зонна теорія твердих тіл дозволила з єдиної точки зору пояснити існування металів, діелектриків і напівпровідників, пояснюючи відмінності в їхніх електричних властивостях неоднаковим заповненням електронами дозволених зон і різною шириною заборонених енергетичних зон.

Ступінь заповнення електронами енергетичних рівнів у зоні визначається заповненням відповідного атомного рівня. Якщо, наприклад, якийсь рівень атома повністю заповнений електронами відповідно до принципу Паулі, то

утворена з таких рівнів енергетична зона також повністю заповнена.

Найвищу зону, яка цілком заповнена електронами за $T = 0$ K, називають *валентною зоною*. Зону, яка заповнена електронам частково або вільна за $T = 0$ K, називають *зоною провідності*.

Залежно від ступеня заповнення енергетичних зон електронами і ширини забороненої зони можливі чотири випадки (рис. 19.15). У першому випадку найвища зона, що містить електрони, заповнена лише частково, тобто є вакантні рівні (рис. 19.15, *а*). У такому разі електрон, отримавши малу енергетичну добавку за рахунок теплового руху або електричного поля, зможе перейти на вищий енергетичний рівень зони, тобто стати вільним і брати участь у провідності.

Внутрішній перехід можливий, оскільки навіть за $T = 1$ K енергія теплового руху $kT \approx 10^{-4}$ еВ більша від різниці енергій між сусідніми енергетичними рівнями зони (10^{-22} еВ). Отже, тіло буде провідником електричного струму. Саме це властиве *металам*.

Тверде тіло є провідником електричного струму і в тому випадку, коли валентна зона 3*s* перекривається з вільною від електронів зоною 3*p*, що приводить до неповного заповнення зони (рис. 19.15, *б*). Це справедливе для лужноземельних елементів, що утворюють другу групу періодичної системи елементів. У цьому випадку утворюється «гібридна» зона, яка заповнюється валентними електронами лише частково.

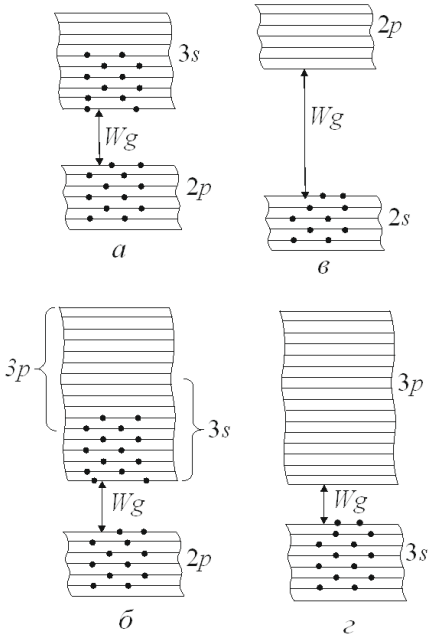


Рис. 19.15. Позначення: темні кружки – електрони

Можливий також перерозподіл електронів між зонами, що виникають з рівнів різних атомів, який може привести до того, що у кристалі виявиться одна цілком заповнена (валентна) зона та одна вільна зона (зона

провідності). Тверді тіла, в яких енергетичний спектр електронних станів складається лише з валентної зони і зони провідності, є діелектриками або напівпровідниками залежно від ширини забороненої зони W_g.

Якщо ширина забороненої зони кристала становить декілька електрон-вольт, то тепловий рух не може перевести електрони з валентної зони у зону провідності, а кристал є *діелектриком*, залишаючись ним за всіх реальних температур (рис. 19.15, *в*). Для типових діелектриків $W_g > 3$ еВ. Так, для алмазу $W_g \approx 5{,}2$ еВ, для нітриду бору $W_g \approx 4{,}6$ еВ, для Al_2O_3 $W_g \approx 7$ еВ і т.д.

Якщо заборонена зона досить вузька, $W_g \leq 1$ еВ, перехід електронів з валентної зони у зону провідності може бути здійснений порівняно легко – або тепловим збудженням, або за рахунок зовнішнього джерела, здатного передати електронам енергію W_g (рис. 19.15, *г*). Такий кристал буде *напівпровідником.* За температур, близьких до 0 К, напівпро-відники ведуть себе як діелектрики, оскільки переходів електронів у зону провідності не відбувається. Для типових напівпровідників $W_g \leq 1$ еВ.

Зонну структуру енергетичного спектру для електрона у кристалі можна отримати в результаті розв'язування рівняння Шредінгера:

$$-\frac{\hbar^2}{2m_e}\Delta\psi(\vec{r}) + U(\vec{r})\psi(\vec{r}) = W\psi(\vec{r}),$$

де $U(\vec{r})$ – потенціальна енергія одного електрона не лише в періодичному полі ядер ґратки, але і в усередненому періодичному полі всіх інших електронів; $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ – оператор Лапласа; W – повна енергія електрона у кристалі; m_e – маса електрона. Через математичні труднощі розв'язування цього рівняння розглядати не будемо.

Для кращого розуміння виникнення енергетичних зон у кристалах розглянемо ще такий підхід. Рух електрона в атомі можна зіставити з будь-якою коливальною системою, наприклад з маятником. Нехай у нас є два зовсім однакових маятники. Поки вони не взаємодіють один з одним, частота коливань обох маятників одна й та сама. Введемо тепер взаємодію між ними – зв'яжемо їх, наприклад, м'якою пружинкою. І зразу ж замість однієї частоти з'явиться дві. Розглянемо рис. 19.16. Зв'язані маятники можуть коливатись синфазно (рис. 19.16, *а*), а можуть на зустріч один одному

(рис. 19.16, *б*). Очевидно, що у другому випадку рух маятників буде більш швидким, тобто частота коливань такої системи вища за власну частоту коливань одного маятника. Таким чином, зв'язок приводить до розщеплення частот. Якщо зв'язати три маятники, то стане вже три власних частоти, у системи з чотирьох зв'язаних маятників чотири власних частоти і так далі до нескінченності.

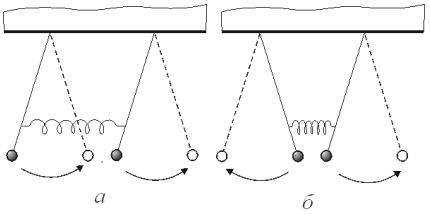


Рис. 19.16. Коливання зв'язаних маятників

Поведінка будь-яких коливальних систем подібна. Якщо ми замінимо маятники, наприклад, на електричні коливальні контури, то при введенні зв'язку між ними їхні власні частоти теж розщеплюються. Електрони в атомі – це теж своєрідна коливальна система. Як і маятник, електрони мають масу, є сила Кулона, яка повертає їх в положення рівноваги, і цим визначається рух електронів в атомі, який згідно із квантовою механікою характеризується власною частотою. Для електронів включення взаємодії під час взаємного зближення атомів приводить до того, що частоти, які були до того однаковими, стають дещо різними. Будемо пам'ятати, що у квантовій механіці є прямий зв'язок між енергією і частотою коливань, який виражають формулою $W = h\nu$.

Розглянемо один із найпростіших елементів, який утворює метал – літій. Порядковий номер Li – три. Це означає, що атом літію має три електрони, і його електронна конфігурація $1s^2 2s^1$. Два електрони утворюють заповнену $1s$-оболонку, яка найближча до ядра і тому сильно зв'язана з ним. Третій електрон перебуває на $2s$-підоболонці, хоч на ній могли б перебувати два електрони. Усі інші дозволені електронні підоболонки вільні, й електрони на них потрапляють тільки під час збудження атома (наприклад, при сильному нагріванні парів літію). Схему рівнів в атомі літію показано на рис. 19.17, 1 *атом*. Під час зближення двох атомів літію кожний із рівнів, які показано на рис. 19.17, 2 *атоми*, розщеплюється на два. Кожному новому рівню енергії буде відповідати своя електронна оболонка тепер уже не окремого атома, а «молекули». Рівні енергії заповнюються електронами за тим же правилом, що й у атома – по два електрони на кожному рівні. Пара

оболонок, яка отримується з найнижнього рівня, буде повністю заповнена електронами. Справді, на них можна розмістити чотири електрони, а їх у двох атомів літію шість. Залишаються два електрони, які тепер розміщуються на нижньому рівні другої пари. Зауважимо, що відбувся якісний стрибок: раніше ці два електрони займали два із чотирьох станів, які мали однакову енергію. Тепер у них з'явилась можливість обирати й вони розмістились так, щоб їх сумарна енергія була найменшою. Неважко уявити, що відбудеться після додавання наступних атомів: для трьох атомів кожний вихідний рівень розщепиться на три (рис. 19.17, 3 *атоми*). Дев'ять електронів розмістяться так: шість на нижній тріаді рівнів, які виникають з рівня найближчої до ядра внутрішньої заповненої оболонки атома, ще два електрони – на нижньому рівні наступної тріади; електрон, що залишився, – на середньому рівні тієї ж тріади. Ще одне місце на цьому рівні залишається вільним, а верхній рівень повністю порожній. Якщо взяти n атомів ($n \gg 1$), то кожний рівень розщепиться на n тісно розміщених рівнів, які утворюють смугу або зону дозволених значень енергії. У нижній зоні всі стани будуть зайнятими, а у другій – тільки половина – тільки ті стани, енергія яких найменша. Наступна зона – повністю порожня.

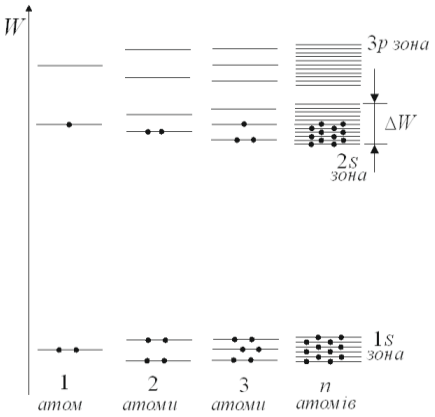


Рис. 19.17. Процес виникнення енергетичних зон в літії

Відстань між сусідніми рівнями в зоні легко оцінити. Можна вважати, що в разі зближення атомів зміна енергії електронів атома приблизно дорівнює теплоті випаровування речовини, розрахованої на один атом. Для металів вона становить декілька електронвольт, а отже й повна ширина зони ΔW (рис. 19.17), яка визначається взаємодією сусідніх атомів, повинна мати той самий порядок, тобто $\Delta W \approx 1$ еВ. Для відстані між рівнями отримуємо $\delta W \approx \Delta W/n$, де n – кількість атомів у зразку. Об'єм, який припадає на один атом, дорівнює $a^3 \sim 10^{-22}$ см^3 (наприклад, для натрію (рис. 19.14) $a = 4{,}3 \cdot 10^{-8}$ см). Для

зразка, об'єм якого $1\ \text{см}^3$, $n \approx 10^{22}$, тому $\delta W \cong 10^{-22}$ еВ.

19.7. Статистика електронів у кристалах

Однією із проблем, яка виникає під час створення приладів електронної техніки, є отримання (генерація) достатньої кількості рухомих носіїв заряду, які можуть забезпечити потрібні електронні процеси.

Функція розподілу електронів

Єдиною можливістю аналізувати поведінку рухомих носіїв заряду є дослідження статистичних закономірностей електронів, тобто дослідити їх функцію розподілу за енергіями, за швидкостями і т.п.

Функцію розподілу електронів за енергіями $N(W)$ задають як відношення кількості електронів в одиниці об'єму dn, що мають енергію в інтервалі від W до $W + dW$, до величини цього інтервалу:

$$N(W) = \frac{dn}{dW}. \qquad (19.3)$$

Ця функція має зміст густини розподілу електронів за енергіями, тобто кількості електронів в одиничному енергетичному інтервалі в одиниці об'єму. Функція $N(W)$ має розмірність кількість електронів$/(\text{м}^3 \cdot \text{Дж})$, але, як правило, у фізиці твердого тіла використовують розмірність кількість електронів$/(\text{см}^3 \cdot \text{еВ})$.

Функцію $N(W)$ розглядають як добуток двох функцій $g(W)$ та $f(W)$, де

$$g(W) = \frac{dn_{рівн}}{dW}, \qquad (19.4)$$

$$f(W) = \frac{dn}{dn_{рівн}}. \qquad (19.5)$$

Функція $g(W)$ характеризує густину станів (рівнів), тобто кількість енергетичних рівнів в одиниці об'єму кристала, які припадають на одиничний енергетичний інтервал близько значення енергії W, тому $dn_{рівн}$ – кількість енергетичних рівнів в одиниці об'єму кристала в інтервалі енергій від W до $W + dW$ (рис. 19.18, *а*). Функція $f(W)$ характеризує ймовірність заповнення енергетичних рівнів електронами. Зручність такого подання функції $N(W)$ у тому, що задача знаходження функції розподілу електронів за енергіями розпадається на дві: квантову задачу визначення характеру розподілу енергетичних рівнів і статистичну задачу знаходження ймовірності заповнення цих рівнів.

Функція густини станів (рівнів) для металів під час вибору початку відліку енергії від нижнього краю зони провідності (рис. 9.18, *а*) має вигляд

$$g(W) = \frac{4\pi}{h^3}(2m_n^*)^{3/2}W^{1/2}, \quad (19.6)$$

де m_n^* – ефективна маса електрона в зоні провідності металу. Для розв'язування задач вважатимемо, що m_n^* у металах дорівнює масі вільного електрона m_e, тобто $m_n^* = 9{,}1 \cdot 10^{-31}$ кг. Функція $g(W)$ є квантовою, оскільки містить сталу Планка h

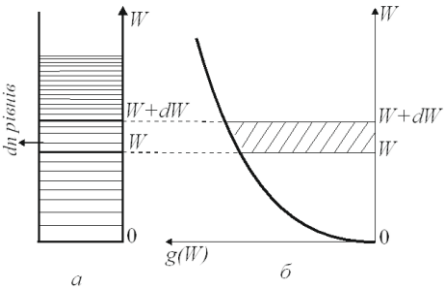


Рис. 19.18. Енергетична діаграма зони провідності (а) і графік функції розподілу густини рівнів (б)

Енергія є важливою характеристикою електрона, який рухається в періодичному полі кристалічної ґратки металу або напівпровідника. Вона є функцією хвильового вектора, модуль якого $|\vec{k}| = 2\pi/\lambda$, тобто $W = W(\vec{k})$. Рівняння $W(\vec{k}) = \text{const}$ визначає у просторі хвильового вектора $\vec{k}(k_x, k_y, k_z)$ поверхню однакового значення енергії або ізоенергетичну поверхню, форма якої переважно визначає електричні властивості твердих тіл.

Ефективну масу електрона визначають за формулою

$$m^* = \hbar^2 \left(\frac{d^2W(k)}{dk^2}\right)^{-1}, \quad (19.7)$$

де $\hbar = \frac{h}{2\pi} = 1{,}054 \cdot 10^{-34}$ Дж · с – стала Планка.

Квантовий розподіл електронів за енергетичними рівнями визначають *статистикою Фермі–Дірака*, він характеризується функцією

$$f(W) = \frac{1}{\exp\left\{\frac{W-W_F}{kT}\right\}+1}, \quad (19.8)$$

де T – абсолютна температура; k – стала Больцмана; W_F – енергія Фермі; W – енергія певного рівня. Одиниця у знаменнику враховує принцип Паулі. Із (19.8) випливає, що $f(W_F) = 1/2$ за будь-якої температури, відмінної від нуля, і це формально означає, що енергія Фермі – це енергія такого енергетичного рівня (стану), ймовірність заповнення якого електронами дорівнює $1/2$. Це твердження можна поширити і на ті випадки, коли значення W_F потрапляють у таку область значень енергії, які заборонені у цій квантовій системі. Графік функції Фермі–Дірака (19.8) зображено на рис. 19.19. За $T = 0$ K він має вигляд розривної ступеневої функції, яка дорівнює 1 у межах значень енергій від 0 до значень W_F й дорівнює 0 за енергій, які більші за W_F. Це означає, що

з імовірністю 1 заповненими електронами є всі енергетичні рівні, енергія яких менша від величини W_F і не має жодного електрона у станах, енергія яких перевищує W_F. З підвищенням температури частина електронів набуває теплової енергії й повна їх енергія може стати більшою від енергії рівня Фермі. На графіку функції розподілу Фермі–Дірака (рис. 19.19) це виглядає таким чином, що ступінчастість функції «розмивається», і тим сильніше, чим вища температура. Аналізуючи (19.8) можна переконатись, що тепловому збудженню піддаються тільки ті електрони, енергія яких відрізняється від W_F на величину приблизно kT в більший чи менший бік, тобто на величину теплової енергії.

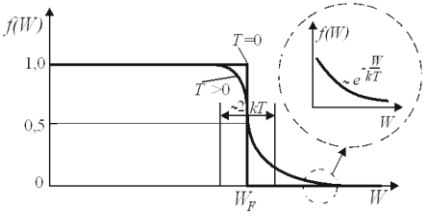


Рис. 19.19. Функція Фермі–Дірака

Вигляд функції розподілу Фермі–Дірака, а також те, що її значення в точках, які симетричні відносно W_F, у сумі дорівнюють 1, дозволяє подати рівень Фермі як деяке середнє значення енергії тих електронів, які беруть участь у тепловому русі. Таке уявлення не дуже строге, але часто зручне для якісного розгляду багатьох фізичних явищ. Знаходження функції розподілу електронів у конкретному колективі зводиться фактично до визначення рівня Фермі та його залежності від температури, концентрації носіїв заряду та інших умов. Кількісне значення енергії Фермі може бути знайдене, наприклад, з умови нормування, суть якої у тому, що інтеграл від певної функції розподілу (19.3) по всьому можливому інтервалу енергій дорівнює концентрації електронів, якщо при цьому використовують функцію густини енергетичних станів (19.6), яка належить до одиниці об'єму, і значення концентрації електронів відоме:

$$n = \int_0^\infty N(W)dW = \int_0^\infty f(W)g(W)dW. \quad (19.9)$$

Розрахунок інтеграла (19.9) в явному вигляді можливий тільки у деяких випадках. Зазвичай використовують наближені методи. Розв'язок подібних задач значно спрощується, якщо розглядати область енергії, яка лежить вище рівня Фермі хоча б на декілька одиниць kT. При цьому через сильну залежність (експоненціальну) функції від аргумента $\exp\left\{\frac{W-W_F}{kT}\right\} \gg 1$ і, нехтуючи одиницею у знаменнику

функції (19.8), отримуємо розподіл вигляду

$$f(W) = \text{const} \cdot \exp\left\{-\frac{W}{kT}\right\}. \quad (19.10)$$

Той факт, що у цьому випадку $f(W) \ll 1$, означає, що біля кожного значення енергії кількість станів набагато більша, ніж кількість електронів, які їх заповнюють. Отже, особливості, які пов'язані із квантовою природою частинки (зокрема, з їх спіном), під час розподілу за станами не проявляються. Ця ситуація характерна для *класичної статистики* Максвелла–Больцмана. Електрони, розподіл яких за енергетичними рівнями підпорядковується закону (19.10), називають ***невиродженим газом***. Відповідно, умову $f(W) \ll 1$ можна розглядати як критерій *невиродженості* електронного газу.

Електрони, розподіл яких за енергетичними рівнями підпорядковується закону (19.8) (тобто враховується принцип Паулі), називають ***виродженим газом***.

Статистика електронів у металах

У металічних кристалах всі колективізовані («вільні») електрони енергетично розміщені в зоні провідності. Їх концентрація визначається концентрацією атомів і кількістю колективізованих валентних електронів від кожного атома. Розподіл цих електронів за енергетичними станами зони провідності визначається, по-перше розподілом станів у зоні і, по-друге, ймовірністю їх заповнення. Розподіл станів характеризується функцією густини станів $g(W)$ (19.6).

На рис. 19.20 зображено зону провідності металу і функції розподілу $g(W)$, $f(W)$, $N(W) = f(W)g(W)$, які розміщені так, щоб вісь їхнього аргумента збігалася із звичайним напрямком енергії на зонних діаграмах. Із рисунка зрозумілий простий фізичний зміст енергії Фермі, яка для металів відповідає найвищому рівню зони провідності, що заповнений електронами за нульової температури.

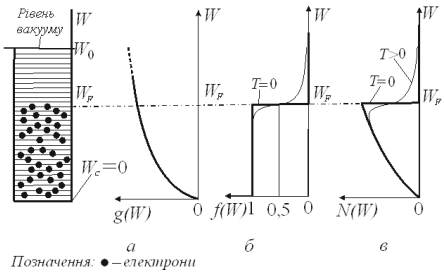


Рис. 19.20. Функції:розподілу густини станів (*а*); ймовірність заповнення станів (*б*); густина електронів (*в*) у зоні провідності металів

Якщо вважати, що вигляд функції (19.6) зберігається не тільки для

нижнього краю зони провідності, але і вище, можна скористатись формулою (19.9) і визначити значення енергії Фермі. Це легко зробити для $T = 0$, коли функція $f(W)$ дорівнює 1 для енергій від 0 до W_F і дорівнює 0 для інших значень енергії (рис. 19.20, *б*). При цьому підінтегральний вираз у (19.9) значно спрощується й інтеграл вираховують елементарно (дорівнює площі криволінійної трапеції для випадку $T = 0$ на рис. 19.20, *в*):

$$n = \int_0^{W_F(0)} g(W)dW = \\ = \frac{4\pi}{h^3}(2m_n^*)^{3/2}\int_0^{W_F(0)} W^{1/2}dW$$

і

$$n = \frac{8\pi}{3h^3}(2m_n^*)^{3/2}\big(W_F(0)\big)^{3/2}. \quad (19.11)$$

Звідси знаходимо значення енергії Фермі за нульової температури

$$W_F(0) = \frac{h^2}{2m_n^*}\left(\frac{3n}{8\pi}\right)^{2/3}, \quad (19.12)$$

яка залежить тільки від концентрації електронів. Це стає більш зрозумілим, якщо уявно збільшити кількість електронів у металі. Додаткові електрони будуть розміщуватись на вільних рівнях, які мають енергію, більшу за W_F, підвищуючи тим самим значення W_F. Для типових металів, коли ефективна маса електронів близька до маси спокою електрона,

$$W_F(0) = 0{,}365 \cdot 10^{-18} n^{2/3}, \quad (19.13)$$

де n визначають у м^{-3}, а $W_F(0)$ – в електронвольтах, і за концентрації електронів у межах 10^{28} ... 10^{29} м^{-3} $W_F(0)$ має величину в декілька електронвольт.

За температур, відмінних від 0, інтегрування (19.9) у загальному вигляді маємо

$$n = \frac{4\pi}{h^3}(2m_n^*)^{3/2} \times \\ \times \int_0^\infty W^{1/2}\left(\exp\left\{\frac{W - W_F}{kT}\right\} + 1\right)^{-1}dW. \quad (19.14)$$

Це означає, що рівень Фермі залежить від температури (рис. 19.21). За низьких температур поведінку $W_F(T)$ описують приблизним виразом

$$W_F(T) \cong W_F(0)\left(1 - \frac{\pi^2}{12}\left(\frac{kT}{W_F(0)}\right)^2\right), \quad (19.15)$$

який отримують під час розкладу підінтегральної функції у (19.14) в ряд.

За високих температур, коли рівень Фермі зміщується в область негативних значень енергій (рис. 19.21), експонента в (19.14) стає набагато більшою від одиниці, що дає можливість перейти до статистики Максвелла – Больцмана і вирахувати інтеграл в явному виді. Вираз для рівня Фермі, який отримують у цьому випадку, характерний для *невиродженого* колективу.

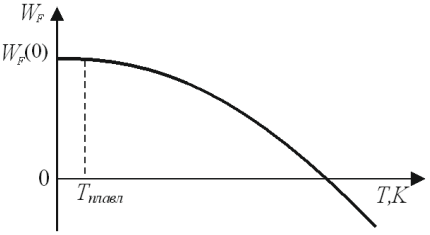


Рис. 19.21. Залежність рівня Фермі від температури

Кількісна оцінка показує, що за температур близько 1000 К рівень Фермі зміщується всього на величину

$$W_F(T) = kT \ln\left(\frac{n}{2}\left(\frac{h^2}{2\pi m_n^* kT}\right)^{3/2}\right), \quad (19.16)$$

$(10^{-3} \ldots 10^{-4}) \cdot W_F(0)$, а (19.16) стає справедливим за температур близько $10^4 \ldots 10^5$ К. Отже, можна вважати, що для металів, які перебувають у твердому стані, електронний газ є завжди *виродженим*, а рівень Фермі майже не залежить від температури й може бути розрахований за формулою (19.13) чи, за потреби, за більш точною формулою (19.15).

З функцією розподілу (19.8) пов'язана ще одна особливість поведінки електронів у металах, яка проявляється в тому, що їх розподіл за енергіями визначається насамперед не тепловим рухом, а принципом Паулі. Це підтверджується виразом для середньої енергії одного електрона:

$$< W \geq W_\Sigma \quad /n = \frac{\int_0^\infty WN(W)dW}{\int_0^\infty N(W)dW,}, \quad (19.17)$$

де W_Σ – сумарна енергія всіх n «вільних» електронів.

За $T = 0$ К вираз (19.17) значно спрощується і

$$< W_0 \geq \frac{\int_0^{W_F(0)} Wg(W)dW}{\int_0^{W_F(0)} g(W)dW} = \frac{3W_F(0)}{5}. \quad (19.18)$$

У більш загальному випадку, якщо розкласти підінтегральні вирази у (19.17) в ряд, обмежившись членами, які містять T^2, то можна отримати

$$< W \geq = \frac{3}{5} W_F(0) \cdot \left(1 + \frac{5\pi^2}{12}\left(\frac{kT}{W_F(0)}\right)^2\right). \quad (19.19)$$

Перший доданок у (19.19) є (19.18) і характеризує середню енергію за нульової температури, відповідно, його називають «нульовою» енергією, що має суто квантову природу. Другий доданок у (19.19) залежить від температури і величина

$$< W(T) > \cong \frac{\pi^2 (kT)^2}{4W_F(0)} \quad (19.20)$$

є середньою тепловою енергією електронів. Порівняння величин (19.18) і (19.20) показує, що навіть за температур, які близькі до температури плавлення металу, теплова енергія електронів становить $10^{-3} \ldots 10^{-4}$ від «нульової енергії», а відповідна теплова швидкість – одиниці відсотків від «нульової швидкості». Така слабка

сприйнятливість електронів у металах до нагрівання випливає з характера функції розподілу електронів. На рис. 19.22 показано функцію Фермі–Дірака для електронів зони провідності срібла ($W_F = 5{,}5$ еВ) для температури 290 К $kT = 0{,}025$ еВ і для температури плавлення 1233 К ($kT = 0{,}106$ еВ).

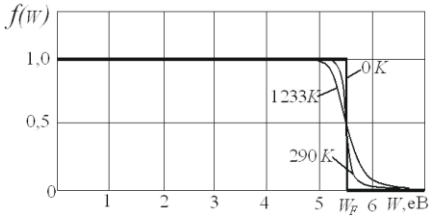


Рис. 19.22 Функція Фермі–Дірака для електронів зони провідності срібла

Із рис. 19.22 видно, що зміни функції під час нагрівання дуже незначні, й тільки ті електрони, енергія яких відрізняється від енергії Фермі на величину близько 0,1 еВ, набувають теплової енергії, беручи участь у тепловому русі. Ці електрони становлять невелику частку близько $\frac{kT}{W_F(0)}$ від усіх електронів, тому й теплова енергія всіх електронів набагато менша від його повної енергії. Ця обставина пояснює, зокрема, малий внесок теплоємності електронного газу в загальну теплоємність металів. Крім того, рис. 19.22 показує, що зглажування ступінчастості функції Фермі–Дірака невелике навіть за високих температур і межа між зайнятими і вільними станами зберігається достатньо різкою. Отже, зміст рівня Фермі, який пов'язаний з цією границею, можна приблизно зберегти не тільки для нульової температури.

Звернімо увагу на те, що розподіл електронів за енергіями в зоні провідності металів подібний до розподілу густини молекул рідини з висотою. Процес переходу електронів на більш високі енергетичні рівні під час нагрівання аналогічний процесу випаровування молекул з рідини. З цієї точки зору більш правильним для електронів в металі була би назва «електронна рідина» (рівень Фермі електронів відповідає висоті рівня поверхні рідини). Така аналогія характерна для сукупності електронів в цілому, а та частина електронів, енергія яких перевищує енергію рівня Фермі на декілька одиниць kT, аналогічна парі над поверхнею рідини і може бути названа «*електронним невиродженим газом*», до якого з достатньою точністю може бути застосована статистика Максвелла–Больцмана. Особливу групу становлять найбільш «гарячі» електрони, енергія яких на декілька електронвольт більша від енергії Фермі. Попри те, що ця група

надто нечисленна (наприклад, ймовірність заповнення електронами рівнів на 1 еВ вище рівня Фермі за $T = 1160$ К не перевищує величину $\exp\{-10\} = 4{,}5 \cdot 10^{-4}$), але саме ці електрони беруть участь у такому, наприклад, явищі, як термоелектронна емісія.

19.8. Електропровідність металів та її залежність від температури. Надпровідність

Класична теорія

Опір провідника протіканню струму зовсім не така проста річ, як може видатися на перший погляд. У розгляд доречно буде ввести поняття кванта, але спочатку спробуємо обійтись без його допомоги.

У класичній теорії не використовують поняття «квант». У багатьох наукових і навчальних книгах написано, що електричний опір металів є результатом зіткнення електронів, які переносять заряд, з іонами кристалічної ґратки. Таке розуміння опору металів виражає класична теорія. Утім, вона не може пояснити деяких основних особливостей електричного опору. У класичній теорії, безумовно, справедливе одне: опір виникає від того, що електрони передають частину своєї енергії та імпульсу (кількості руху) кристалічній ґратці. Але яким чином відбувається ця передача – питання зовсім не просте.

Температурна залежність. Опір чистих металів істотно зростає з температурою. Для багатьох із них зростання в основному пропорційне абсолютній температурі. За низьких температур така проста залежність порушується. Це добре видно із рис. 19.23, де показано, як змінюється з температурою питомий опір міді, провідника, який найбільш широко використовують для передачі струму.

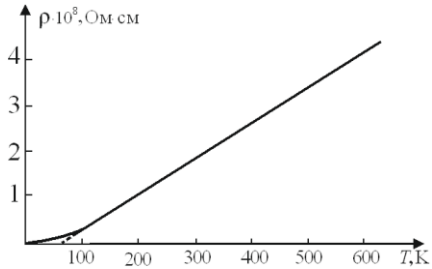


Рис. 19.23. Графік залежності питомого опору міді від абсолютної температури

Якщо пояснювати існування електричного опору зіткненнями електронів з іонами, то температурна залежність опору чистих металів стає зовсім незрозумілою. Справа виглядає так, начебто електрони зіштовхуються тільки з іонами (атомами), які здійснюють тепловий рух, але вільно

пролітають повз нерухомі іони. Виглядає так, що електрону легше потрапити в іон, який рухається. Звичайно, це неправильно. Таким чином, класична теорія не здатна пояснити залежність $\rho(T)$ для металів.

Залишковий опір. Коли вимірюють опір чистих металів за дуже низьких температур, то виявляється цікавий факт. Продовжуючи криву $\rho(T)$ (аналогічно зображеному на рис. 19.23) до абсолютного нуля, отримують опір, який називають залишковим опором. Виявляється, що він сильно змінюється від зразка до зразка. Залишковий опір вкрай чутливий до невеликої кількості домішок, до механічної і термічної обробки металу. Очищуючи метал й обробляючи його так, щоб досягти майже досконалої кристалічної структури, можна зменшити залишковий опір, і цьому зменшенню не має меж. Такий досвід підштовхує до такого висновку: ідеальний кристал за температури абсолютного нуля не повинен мати електричного опору. Інакше кажучи, електричний опір чистих металів обумовлений порушеннями кристалічної ґратки, які викликаються тепловим рухом іонів, домішками, дефектами (неправильностями) кристалічної ґратки.

Атомні коридори. Якби електрони рухались за законами класичної механіки, вони мали б зіштовхуватись з іонами (атомами) незалежно від того, розміщені атоми у строгому порядку (як у кристалі), чи по-іншому. Звичайно, між правильними рядами атомів є начебто коридори (рис. 19.24), але щоб направити електрон по такому коридору, потрібна точна орієнтація кристала щодо електричного поля, яке прикладене до зразка. Утім, зазвичай мають справу не з одним кристалом, а з великою кількістю дрібних кристалів, які орієнтовані випадковим чином (рис. 19.25).

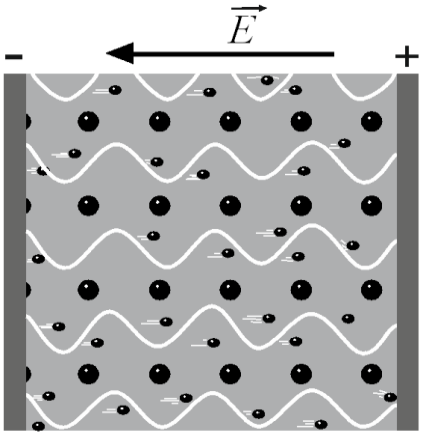


Рис. 19.24. Рух електронів і поширення електронних хвиль в ідеальному кристалі: ● – іони кристалічної ґратки; •– електрони

Одиничний правильний кристал називають монокристалом, а з'єднання багатьох кристалів – полікристалічним агрегатом. Точні вимірювання показують, що опір полікристалічних агрегатів майже не відрізняється від опору монокристалів. Особливо важливо те, що опір полікристалічних агрегатів сильно зменшується зі зниженням температури. Це не можна пояснити рухом електронів атомними коридорами за законами класичної механіки. Під час переходу з одного кристала в інший атомний коридор різко змінює свій напрямок (рис. 19.25) й електрон мав би з розгону вдаритись об «стіну», а він продовжує свій рух атомним лабіринтом.

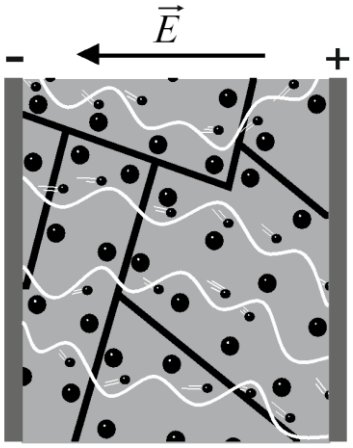


Рис. 19.25. Схема руху електронів атомними лабіринтами у полікристалічному зразку: прямі лінії ▬ – межі мікрокристалів; ● – іони кристалів; ● – електрони

Наведемо *деяку цікаву інформацію.* Давно відомо, що механічна і термічна обробка суттєво впливає на електричний опір металів. За температури 20 °С звичайна технічна мідь має питомий опір 17,2 нОм·м. Після холодного протягування питомий опір мідного дроту зростає до 17,7 нОм·м. Навіть намотування проводу на котушку викликає зростання його опору. Якщо провід відпалити, тобто довгий час нагрівати, а потім знову охолодити до 20 °С, то значення опору повертається до нормальної величини. Очевидно, що опір чутливий до невеликих порушень кристалічної структури. Ще більше опір чутливий до мізерної кількості домішок. Старанне очищення зменшує питомий опір міді за $t = 20$ °С до 16,9 нОм·м. Ця обставина має велике значення для техніки, оскільки вона дозволяє зменшити опір проводів і таким чином зменшити непотрібні витрати електроенергії на нагрівання. Якщо до міді додати 1 % марганцю, то питомий опір її зростає до 48 нОм·м, тобто майже у три рази, і дорівнює опору чистого марганцю (50 нОм·м). Таким самим чином впливають добавки заліза, кобальту, іридію та інші. Якби опір виникав від зіткнення електронів з атомами домішки, то ці домішки мали би

впливати у сто раз слабше. З точки зору класичної теорії такий сильний вплив домішок на опір пояснити неможливо.

Є спеціально розроблені сплави з високим питомим опором – нікелін, манганін, константан, ніхром та інші. Опір цих сплавів у декілька раз більший, ніж у їх складових. Приміром, константан, який містить 60 % міді та 40 % нікелю, має питомий опір 440 нОм·м, тоді як у чистої міді він дорівнює 17 нОм·м, а у нікелю – 72 нОм·м. У ніхрома питомий опір близько 1000 нОм·м. Температурна залежність опору сплавів значно слабша, ніж у чистих металів. Наприклад, у константана в інтервалі температур від 0 до 400 °С питомий опір змінюється тільки від 441 до 448 нОм·м.

З точки зору класичної теорії постійний опір сплавів майже незрозумілий, так само, як і пропорційність між опором та температурою для чистих металів. Опір сплавів мав би за змістом цієї теорії складатись з опору складових частин за простим правилом змішування, а опір чистих металів пропорційний $\sqrt{T}$.

Квантова теорія

Для теоретичного обґрунтування залежності питомого опору металів від температури візьмемо за основу одне з основних тверджень класичної електронної теорії електропровідності, тобто

$$\sigma = enu, \qquad (19.21)$$

де $\sigma = 1/\rho$ – питома електропровідність металу (ρ – питомий опір); n – концентрація електронів провідності («вільних» електронів); u – рухливість електронів.

Вираз (19.21) має дві величини, які принципово можуть залежати від температури – це n та u.

Згідно із зонною теорією металів концентрацію електронів провідності (електронів у зоні провідності) визначають за температури $T = 0$ за формулою (19.11). Для металів енергія Фермі майже не залежить від температури, тому можна вважати, що концентрація «вільних» електронів n не залежить від температури, тобто є сталою. Залишається вияснити залежність від температури рухливості електронів.

Згідно із квантовою механікою електрон має одночасно властивості частинки і хвилі (співвідношення де Бройля $\lambda = h/(m_e v)$). Ширина атомних коридорів близька до довжини хвилі де Бройля електрона, і тому під час руху такими коридорами хвильові властивості електрона проявляються у повній мірі.

У радіотехніці сантиметрових хвиль широко застосовують хвилеводи – металічні трубки круглого або прямокутного перерізу. Радіохвилі поширюються по хвилеводу, йдучи за всіма його згинами. Світло – теж електромагнітні хвилі, і його можна передавати світловодами (трубки із дзеркальними стінками). Дія світловоду із дзеркальними стінками зрозуміла. Світлові хвилі весь час відбиваються від стінок і, таким чином, залишаються всередині світловоду. Радіохвилі теж добре відбиваються від гладких металічних поверхонь. Якщо поверхня шерстка, то хвиля ніби розривається об зазубринки і в результаті розсіюється та поглинається. При цьому енергія хвилі переходить у тепло.

У квантовій теорії протікання електричного струму через метал уявляють як поширення електронних хвиль атомними коридорами, які відіграють роль хвилеводів (рис. 19.24 і 19.25). Якщо атоми розміщені на площині в ідеальному порядку на однакових відстанях один від одного, то така площина повністю відбиває електронні хвилі, як ідеальне дзеркало. Розсіювання і поглинання електронних хвиль відбувається тільки в разі порушення строгого порядку в розміщенні атомів. Цей висновок вперше був отриманий досить складним математичним шляхом з рівнянь квантової механіки. Розглянемо якісно причину існування електричного опору. В ідеально правильному кристалі хвилеводи досконало гладкі. Будь-які порушення атомного порядку діють як шорсткість стінок хвилеводу. Електронні хвилі розсіюються порушеннями ґратки. Частина енергії електронів поглинається і переходить у тепло. Це і є *квантовий механізм електричного опору*.

Правильність атомних хвилеводів порушується тепловими коливаннями іонів кристалічної ґратки, на яких розсіюються електронні хвилі (рис. 19.26). Звідси виникає теплова частина електричного опору, яка залежить від температури. Охолоджуючи метал, її можна зробити скільки завгодно малою. При цьому залишається залишковий, або структурний, опір, пов’язаний з постійними дефектами (порушеннями) кристалічної структури. Дефекти є у будь-якого кристала. Додаткові дефекти виникають під час холодної механічної обробки – протягування проводу і навіть намотування його на котушку. Від цього і збільшується опір. Під час тривалого нагрівання (відпалу) атоми повертаються

на свої місця (дефекти «заліковуються») й опір зменшується.

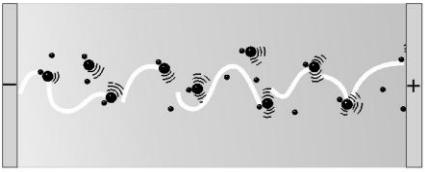


Рис. 19.26. Порушення правильності атомного хвилеводу тепловим рухом атомів: ● – іони кристалічної ґратки; ● – електрони

Проникнення чужорідних атомів викликає істотні порушення кристалічної ґратки, при цьому один атом домішки може вибити з місця сотні атомів кристала-«господаря», тому-то домішки й підвищують опір. Встановлено, що дія домішки тим сильніша, чим більше її атоми відрізняються від атомів «господаря» за розмірами та іншими факторами, чим сильніше вони порушують правильність кристалічної ґратки.

Якщо охолоджувати метал, то теплова (причиною якої є тепловий рух) частина опору зменшується. Коли вона стає значно меншою, ніж структурна частина, яка не залежить від температури, то з подальшим охолодженням опір майже не змінюється. У сплавів з невпорядкованою будовою структурна частина опору настільки велика, що навіть за високих температур більша від теплової. Цим і пояснюється слабка залежність електричного опору сплавів від температури.

Пояснення температурної залежності опору. Тепловий рух у твердому тілі (металі) можна уявити як коливання атомів (іонів) поблизу своїх положень рівноваги, але частинки у твердому тілі не можуть коливатись незалежно. Вони усі міцно пов'язані між собою і здійснюють колективні коливання. Колективні коливання великої кількості частинок речовин – це те саме, що звукові хвилі. Якщо вивести будь-яку частинку із положення рівноваги, то збурення буде поширюватися по тілу із швидкістю звуку. Виявляється, що тепловий рух у твердому тілі можна розглядати як поширення звукових хвиль. Нагріваючи тіло, ми змушуємо його «звучати». На щастя, частота цих коливань у мільярди раз вища від тих, які може чути вухо людини.

Так з'являється зв'язок між електричним опором і звуком. Електрони розсіюються на звукових хвилях, які збуджуються тепловим рухом іонів кристалічної ґратки, але будь-яку хвилю можна вважати *квантами.* У звуковій хвилі це *фонони*, які здатні існувати тільки всередині

речовини (у вакуумі фононів бути не може). Якщо звукову хвилю описувати як потік фононів, то й електрони можна вважати частинками, забувши про їх хвильові властивості. Тепер ми повертаємося до простого пояснення опору: електронам заважає вільно рухатися те, що вони зіштовхуються з іншими частинками – звуковими квантами (фононами).

Будемо вважати для простоти, що всі фонони мають однакову частоту ν. Це припущення неправильне, але помилка через нього невелика. Енергія фонона, як і будь-якого кванта, дорівнює $h\nu$, де h – стала Планка, а ν – частота хвилі. Тоді кількість фононів знаходимо з формули

$$N = \frac{W}{h\nu} = \frac{kT}{h\nu}, \qquad (19.22)$$

де $W = kT$ – енергія теплового руху, яка пропорційна абсолютній температурі T; k – стала Больцмана.

Кількість фононів у твердому тілі пропорційна абсолютній температурі. Отже, рухливість електронів у формулі (19.21) буде обернено пропорційна кількості фононів, тобто

$$u \sim \frac{1}{N} \sim \frac{1}{T}.$$

Це означає, що $\sigma \sim \frac{1}{T}$, а питомий опір металів буде попорційний абсолютній температурі, тому можемо записати

$$\rho = \rho_0(1 + \alpha t) = \alpha\rho_0 T, \qquad (19.23)$$

де t – температура у градусах Цельсія (℃); ρ – питомий опір за температури t; ρ_0 – питомий опір за $t = 0$ ℃; $T = (t + 273)$ К; α – температурний коефіцієнт опору,

$$\alpha = \frac{1}{\rho}\frac{d\rho}{dt}. \qquad (19.24)$$

Коефіцієнт α *характеризує відносну зміну питомого опору при зміні температури на один градус.* Правильність атомних хвилеводів порушується тепловими коливаннями іонів кристалічної ґратки, на яких розсіюються електронні хвилі (рис. 19.26). Звідси виникає теплова частина електричного опору, яка залежить від температури. Охолоджуючи метал, її можна зробити скільки завгодно малою. При цьому залишається залишковий, або структурний, опір, пов'язаний з постійними дефектами (порушеннями) кристалічної структури. Дефекти є у будь-якого кристала. Додаткові дефекти виникають під час холодної механічної обробки – протягування проводу і навіть намотування його на котушку. Від цього і збільшується опір. Під час тривалого нагрівання (відпалу) атоми повертаються

на свої місця (дефекти «заліковуються») й опір зменшується.

Зазначимо, що цей результат справедливий тільки за достатньо високих температур. Виходячи зі змісту квантової теорії кількість фононів не може бути меншим від одиниці. З'ясуємо, за якої температури кількість фононів дорівнює одиниці. Цю температуру називають температурою Дебая і позначають грецько літерою θ. Взявши $N = 1$, з формули (19.22) знаходимо, що

$$\theta = \frac{h\nu}{k}. \qquad (19.25)$$

Температура Дебая (або дебаївська температура) відіграє дуже важливу роль у фізиці твердого тіла. З її допомогою кількість фононів, яка нас цікавить, виражають зовсім просто:

$$N = \frac{T}{\theta}. \qquad (19.26)$$

За температур, нижчих від дебаївської, $N < 1$. Це означає, що можна говорити тільки про середню кількість фононів і розрахунок $\rho(T)$ ускладнюється. Втім величина T/θ, як і раніше, залишається мірою кількості фононів у твердому тілі, а отже, і кількості зіткнень електронів з фононами.

Виходячи з цього було висунуто таку гіпотезу: температурна залежність електричного опору чистих металів виражається універсальною функцією від міри кількості фононів $\frac{T}{\theta}$.

Математично цю гіпотезу записують так:

$$\frac{\rho}{\rho_\theta} = F\left(\frac{T}{\theta}\right), \qquad (19.27)$$

де ρ – опір за температури T; ρ_θ – питомий опір за температури θ; F – функція, вигляд якої може бути знайдено з досліду. Важливо те, що функція $F\left(\frac{T}{\theta}\right)$ є універсальною для різних чистих металів. Як добре ця гіпотеза підтверджується на практиці, видно з рис. 19.27.

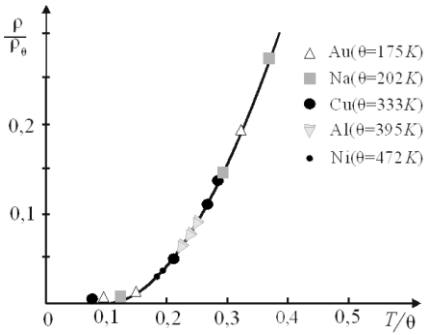


Рис. 19.27. Узагальнена крива температурної залежності питомого опору чистих металів

Надпровідність

У 1911 р. Г. Камерлінг-Оннес відкрив явище стрибкоподібного перетворення в нуль електричного опору деяких металів за досить низьких температур. Вперше це було виявлено в дослідах із ртуттю за температур $T \leqslant 4{,}2$

К. Сучасні методи вимірювання дають підставу зробити висновок, що питомий опір провідника в такому стані менший від 10^{-25} Ом·м. Явище перетворення в нуль опору деяких металів за низьких температур називають *надпровідністю*. Температуру $T_{кр}$, нижче якої спостерігається перехід металів у надпровідний стан, називають *критичною.*

Опір до переходу у надпровідний стан може бути різним. Перехід у надпровідний стан відбувається завжди досить різко. Тепер відомо, що майже половина металевих елементів може переходити у надпровідний стан. Понад тисячу сплавів і сполук металів також здатні переходити у надпровідний стан за певних умов. Речовини, які перебувають у надпровідному стані, мають кілька специфічних властивостей. Основна з них – відсутність електричного опору. У 1914 р. Г. Камерлінг-Оннес виявив, що надпровідний стан руйнується магнітним полем, коли магнітна індукція $\vec{B}$ перевищує деяке критичне значення $\vec{B}_{кр}$. Величина $\vec{B}_{кр}$ залежить від природи надпровідника і його температури. Критичне поле, яке руйнує надпровідність, може бути створене і самим струмом у надпровіднику, тому існує критичне значення сили струму, за якого напровідність зникає. Крім цього, у 1933 р. В. Мейсснер і Р. Оксенфельд встановили, що за температури, нижчої від критичної, відбувається виштовхування магнітного поля з надпровідника. Це явище називають ефектом Мейсснера. Виштовхування магнітного поля з надпровідника означає, що в ньому магнітна індукція $\vec{B}$ дорівнює нулю. Фізично ефект Мейсснера пов'язаний з тим, що в надпровіднику, розміщеному у слабкому магнітному полі, у поверхневому шарі завтовшки $(100 \dots 1000) \cdot 10^{-10}$ м створюються такі колові незгасаючі струми, які повністю компенсують зовнішнє магнітне поле.

Явище надпровідності частіше спостерігається у сплавах, ніж у чистих речовинах. Для чистих речовин проявляється повний ефект Мейсснера. У сплавах відбувається часткове виштовхування магнітного поля з об'єму. Речовини, в яких проявляється повний ефект Мейсснера, називають *надпровід-никами першого роду*, а речовини, в яких не повністю проявляється ефект Мейсснера, *надпровідниками другого роду*. У надпровідниках першого роду надпровідність у певному критичному

полі зникає раптово і металевий провідник переходить у нормальний стан. У надпровідниках другого роду є два критичних магнітних поля – нижнє $H_{\text{н}}$ і верхнє $H_{\text{в}}$. У магнітних полях за $H < H_{\text{в}}$ проявляється повний ефект Мейсснера. Для полів $H_{\text{н}} < H < H_{\text{в}}$ проявляється частковий ефект Мейсснера, а за $H > H_{\text{в}}$ надпровідність зовсім зникає. Явище надпровідності можна пояснити тільки у межах квантової механіки.

Відповідно до теорії Бардіна, Купера і Шріффера виникнення надпровідності зумовлене взаємодією електронів провідності із кристалічною ґраткою. Надпровідність є груповим квантовим ефектом електронів, що взаємодіють між собою внаслідок посередницької ролі ґратки. Окремий електрон під час свого руху поляризує (деформує) ґратку. Інший електрон, потрапляючи в поле поляризованої ґратки, змінює свій стан. Унаслідок цього змінюється його взаємодія з першим електроном. Поляризація ґратки рухомим електроном зумовлює її коливання. Індукована коливаннями ґратки взаємодія між електронами в деяких випадках має характер притягання електронів, які перебувають у станах, близьких до рівня Фермі. Два таких електрони, притягуючись, утворюють своєрідний стан, який називають *куперівською парою*. Між електронами провідності діють сили електростатичного відштовхування, тому утворення куперівських пар можливе у тому випадку, коли притягання електронів унаслідок електрон-фононної взаємодії переважає їх кулонівське відштовхування. Істотним є те, що *куперівська пара* утворюється з електронів з рівними й протилежно спрямованими імпульсами та антипаралельними спінами. Спін такої пари дорівнює нулю. Зазначимо, що за температури $T < T_{\text{кр}}$ існує не одна локалізована пара і не система невзаємодійних пар, а сукупність взаємодійних одна з одною пар електронів. Зв'язок між електронами буде тим сильнішим і стійкішим, чим частіше відбуватимуться процеси обміну фононами між електронами. Утім, за такого процесу обміну електрони, які утворюють пару, змінюють свої імпульси і, щоб утворити пару з нульовим імпульсом, мають «підшуковувати» собі нових «партнерів». Це означає, що пари електронів неперервно обмінюються між собою електронами, тобто весь час відбувається зникнення одних пар і

утворення нових, тому всі електрони в системі пов'язані між собою.

Отже, унаслідок електрон-фононної взаємодії за $T < T_{\text{кр}}$ відбувається перебудова станів і рівнів енергії системи електронів й утворюється надпровідний стан. Це дало змогу стверджувати, що такий тип взаємодії міг би дати основний стан з меншою енергією, ніж повністю невзаємодіючі електрони, і що такий надпровідний основний стан має відмежуватись енергетичною щілиною від стану з нормальною провідністю. Енергетична щілина розміщена близько до рівня Фермі. Наявність скінченної енергетичної щілини у спектрі надпровідного стану пояснює явище надпровідності. Зауважимо, що безпосередні вимірювання заряду носіїв струму у надпровідних вказують на те, що він дорівнює заряду двох електронів.

Якщо електрон-фононна взаємодія істотна, то температура $T_{кр}$ пропорційна температурі Дебая θ для ізотопу певного хімічного елемента. Це означає, що $T_{кр}$ залежатиме від маси M атома ізотопу, а саме $T_{кр} \sim M^{\frac{1}{2}}$, що підтверджено експериментально.

Однією з основних проблем у фізиці надпровідності є проблема підвищення температури переходу в надпровідний стан, тобто створення таких матеріалів, температура переходу яких у надпровідний стан була б значно вищою за температуру рідкого азоту (77K). Майже за 60 років з часу відкриття явища надпровідності вдалося підвищити температуру переходу в надпровідний стан від 4,2 до 23,2 K (Nb_3Ge). У 1986 р. виявлено надпровідність у систем $Ba - La - Cu - O$ при температурах $30 \div 35$ К. У 1987 р. група фізиків США на кераміці $Y - Ba - Cu - O$ досягнула температури переходу у надпровідний стан близько 92 К. У 1988 р. було синтезовано нові вісмутові й талієві сполуки з критичною температурою 100 і 125 К. Нині провідні лабораторії світу ведуть пошуки шляхів створення нових матеріалів, які задовольняли б вимоги високотемпературної надпровідності. Успішне розв'язання цієї проблеми матиме надзвичайно важливе значення для науково-технічного прогресу.

19.9. Статистика електронів і дірок у напівпровідниках

У напівпровідниках колективізовані електрони беруть участь в утворенні ковалентних зв'язків між атомами, і при цьому, пов'язані обмінними силами з

кристалічною ґраткою, не можуть брати участі в електропровідності. Щоб електрони стали рухливими в електричному полі, вони мають подолати сили зв'язку, отримавши певну енергію (рис. 19.11). На мові енергетичних діаграм це означає, що колективізовані електрони містяться у валентній зоні і, щоб стати рухливими, вони мають перейти в зону провідності, в якій є вільні незаповнені енергетичні стани. Для цього електронам слід надати енергію не меншу, ніж величина ширини забороненої зони. Цю енергію вони можуть отримати від різних джерел. Нині ми будемо розглядати генерацію рухливих носіїв заряду (електронів і дірок), які виникають під дією теплового руху.

Розрахувати концентрацію електронів в зоні провідності й дірок у валентній зоні можна, знаючи функції розподілу густини електронів $N_n(W) = g_n(W) \cdot f_n(W)$ і дірок $N_p(W) = g_p(W) \cdot f_p(W)$ для зони провідності та валентної зони відповідно.

Функцію густини дозволених енергетичних рівнів для електронів у зоні провідності напівпровідника визначають за формулою

$$g_n(W) = \frac{4\pi(2m_n^*)^{3/2}}{h^3} \cdot (W - W_c)^{1/2}, \quad (19.28)$$

а для дірок у валентній зоні

$$g_p(W) = \frac{4\pi(2m_p^*)^{3/2}}{h^3} \cdot (W_v - W)^{1/2}. \quad (19.29)$$

У формулах (19.28) і (19.29) m_n^* – ефективна масса електронів у зоні провідності, m_p^* – ефективна маса дірок у валентній зоні, стала Планка $h = 6{,}62 \cdot 10^{-34}$ Дж·с – стала Планка, W_v – енергія «стелі» валентної зони, W_c – енергія «дна» зони провідності (рис. 19.28, *а*, *б*).

Зауваження. За початок відліку енергії у напівпровідниках беруть «стелю» валентної зони, тому будемо вважати, що $W_v = 0$

Вигляд функції розподілу Фермі–Дірака на рис. 19.28, *в*, визначають температурою і рівнем Фермі, положення якого підлягає подальшому уточненню. Функцію Фермі–Дірака для електронів $f_n(W)$ визначають за формулою (19.8), а для дірок

$$f_p(W) = 1 - f_n(W) = \frac{1}{\exp\left\{\frac{W_F - W}{kT}\right\} + 1}, \quad (19.30)$$

де W – енергія електрона (дірки) у зоні; W_F – енергія рівня Фермі; k – стала Больцмана; T – абсолютна температура.

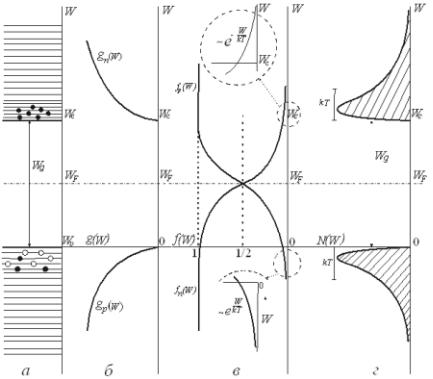


Рис. 19.28. Функція розподілу Фермі–Дірака: а – енергетична діаграма напівпровідника, світлими кружками позначено дірки, чорними – електрони; б – графіки функцій розподілу густини рівнів для електронів $g_n(W)$ і дірок $g_p(W)$; в – графіки функцій розподілу ймовірності заповнення енергетичних рівнів електронами $f_n(W)$ і дірками $f_p(W) = 1 - f_n(W)$; г – графіки функцій N(W) розподілу густини електронів у зоні провідності й дірок у валентній зоні; W_g – ширина забороненої зони; W_F – енергія рівня Фермі; W_c – енергія «дна» зони провідності; W_v – енергія «стелі» валентної зони (нуль відліку енергії)

Концентрація електронів у зоні провідності може бути розрахована як інтеграл (19.9) у межах від нижнього краю зони провідності до верхнього. Утім, для зручності треба розширити верхню межу інтегрування до нескінченності, що майже не впливає на точність розрахунків, оскільки підінтегральна функція під впливом множника $f_n(W)$ дуже швидко зменшується зі збільшенням енергії. Тому

$$n = \int_{W_c}^{\infty} g_n(W)\, f_n(W) dW, \quad (19.31)$$

$$p = \int_{-\infty}^{W_v} g_p\,(W) f_p(W) dW. \quad (19.32)$$

Як, видно, з рис. 19.28, *в*, для розрахунку концентрації електронів і дірок можна з успіхом користуватися класичною функцією розподілу ймовірності Максвелла–Больцмана (19.10). Це пояснюється тим, що вільних рівнів енергії в зоні провідності та валентній зоні набагато більше, ніж електронів і дірок, які там перебувають, тому під час розподілу частинок за енергетичними рівнями принцип Паулі можна не враховувати. Функція $N(W)$ для електронів і дірок на рис. 19.28, *г*, має максимум поблизу дна зони провідності й стелі валентної зони відповідно. Отже, можна вважати, що електрони в зоні провідності переважно перебувають біля його дна, а дірки – біля стелі валентної зони.

Концентрація носіїв заряду n та p у напівпровідниках чисельно дорівнює площам заштрихованих фігур на рис. 19.28, *г*. Для розрахунку n та p зручно перейти до нової змінної $x = (W - W_c)/kT$ для електронів і $x = (W_v - W)/kT$ для дірок. Враховуючи,

що інтеграл $\int_0^\infty \sqrt{x}\exp\{-x\}dx = \frac{\sqrt{\pi}}{2}$, отримаємо:

$$n = N_c \exp\left\{-\frac{W_c - W_F}{kT}\right\}, \quad (19.33)$$

$$p = N_v \exp\left\{-\frac{W_F - W_v}{kT}\right\}, \quad (19.34)$$

де

$$N_c = \frac{2}{h^3}(2\pi m_n^* kT)^{3/2} =$$

$$= 4{,}83 \cdot 10^{21}\left(\frac{m_n^*}{m_e}T\right)^{3/2}, \quad (19.35)$$

$$N_v = \frac{2}{h^3}\left(2\pi m_p^* kT\right)^{3/2} =$$

$$= 4{,}83 \cdot 10^{21}\left(\frac{m_p^*}{m_e}T\right)^{3/2}, \quad (19.36)$$

де $m_e = 9{,}1 \cdot 10^{-31}$кг – маса вільного електрона; T – абсолютна температура; N_c та N_v визначають у метрах кубічних (м$^{-3}$). Простий зв'язок між концентрацією електронів і дірок та положенням рівня Фермі, який виражають формулами (19.33) і (19.34), дуже зручний для аналізу різних ситуацій у напівпровідниках.

Величини N_c і N_v називають *ефективною густиною станів у зоні провідності та валентній зоні* відповідно. Цей термін використовують тому, що для класичного (невирожденого) розподілу концентрація електронів точно дорівнює N_c, помноженій на больцманівську функцію розподілу ймовірності (19.10) за енергії рівня $W = W_c$. Утім, насправді електрони в зоні провідності, а дірки у валентній зоні розподілені в енергетичному інтервалі, що дорівнює декільком величинам kT, як показано на рис. 19.28, *г*. Таким чином зону провідності (й аналогічну валентну зону) можна подати як зону енергій, що складається з N_c (N_v) рівнів, що мають однакову енергію W_c (W_v).

Із співвідношень (19.33) і (19.34) отримаємо

$$\frac{p}{n} = \left(\frac{m_p^*}{m_n^*}\right)^{3/2} \cdot \exp\left\{\frac{2\left(\frac{W_c+W_v}{2} - W_F\right)}{kT}\right\}. \quad (19.37)$$

З формули (19.37) видно, що співвідношення між концентраціями електронів і дірок залежить від їхніх ефективних мас й особливо сильно (*експоненційно*) від положення рівня Фермі відносно середини забороненої зони $\frac{W_c+W_v}{2}$. За рахунок невеликого зміщення рівня Фермі можна отримати зміну співвідношення $\frac{p}{n}$ на декілька порядків. Цей перерозподіл концентрацій рухливих носіїв заряду між зонами може бути здійснений введенням у напівпровідник домішок, оскільки положення рівня Фермі сильно залежить від типу й концентрації домішок.

Закон діючих мас. Із співвідношень (19.33) і (19.34)

отримуємо, що добуток концентрацій рухливих електронів і дірок не залежить від положення рівня Фермі й за певної температури залежить тільки від ширини забороненої зони $W_g = W_c - W_v$ напівпровідника та ефективних мас електронів і дірок:

$$np = N_c N_v \cdot \exp\left\{-\frac{W_g}{kT}\right\}. \quad (19.38)$$

Із співвідношення (19.38) видно, що збільшення концентрації рухливих носіїв заряду одного типу можливе тільки за рахунок зменшення кількості рухливих носіїв іншого типу. Прологарифмувавши (19.37) знайдемо положення рівня Фермі у напівпровіднику:

$$W_F = \frac{1}{2}(W_c + W_v) - \frac{1}{2}kT\ln\frac{p}{n} + \frac{3}{4}kT\ln\frac{m_p^*}{m_n^*}. \quad (19.39)$$

19.10. Електропровідність напівпровідників та її залежність від температури

У загальному випадку електропровідність напівпровідника визначають за такою формулою:

$$\sigma = enu_n + epu_p, \quad (19.40)$$

де n – концентрація електронів провідності; p – концентрація дірок провідності; u_n та u_p – рухливість електронів і дірок відповідно; e – заряд електрона за модулем.

Власний напівпровідник. Власним називають напівпровідник з ідеальною кристалічною ґраткою, яка не містить ані домішок, ані будь-яких дефектів. Енергетичну діаграму такого напівпровідника зображено на рис. 19.29. За температури $T \neq 0$ носії заряду – електрони в зоні провідності (ЗП) і дірки у валентній зоні (ВЗ) – появляються внаслідок «переходу» електронів із ВЗ в ЗП (процеси 1 на рис. 19.29 – процеси генерації).

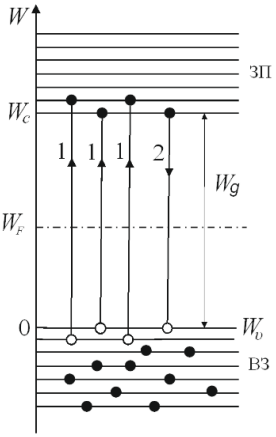


Рис. 19.29. Енергетична діаграма власного напівпровідника: світлі кружки – дірки; темні кружки – електрони; $T \neq 0$; ЗП – зона провідності; ВЗ – валентна зона; W_c – енергія дна зони провідності; W_v – енергія стелі валентної зони, взята за початок відліку енергії; W_F – енергія рівня Фермі; $W_g = W_c - W_v$ – ширина забороненої зони

Одночасно відбуваються і зворотні процеси – процеси рекомбінації

електронів провідності та дірок (електрони заповнюють розірвані зв'язки, процеси 2 на рис. 19.29). За постійної температури між процесами 1 і 2 встановлюється рівновага.

Таким чином, з'являється рівноважна концентрація електронів провідності n_i й дірок провідності p_i, які однакові між собою:

$$n_i = p_i. \qquad (19.41)$$

Цей простий зв'язок (індекс i підкреслює приналежність до власного напівпровідника) є умовою електричної нейтральності речовини й дозволяє визначити положення рівня Фермі. Для цього достатньо у (19.39) відношення $\frac{p}{n}$ прирівняти одиниці, й тоді

$$W_F = W_i + \frac{3}{4}kT\ln\frac{m_p^*}{m_n^*}. \qquad (19.42)$$

Тут через $W_i = \big((W_c + W_v)\big)/2$ позначено середину забороненої зони. Отриманий результат показує, що з ростом температури рівень Фермі лінійно зміщується від середини забороненої зони за $T = 0$ у бік більш легких носіїв. Рівень Фермі відіграє роль «центру мас» між електронами й дірками. Рівень Фермі буде перебувати ближче до тих носіїв заряду, яких більше.

Формулу (19.40) для власного напівпровідника можна записати так:

$$\sigma = en_i\big(u_n + u_p\big), \qquad (19.43)$$

де згідно з (19.38) $n_i^2 = np$, й тоді

$$n_i = \sqrt{N_c N_v}\exp\left\{-\frac{W_g}{2kT}\right\}. \qquad (19.44)$$

У формулі (19.44) W_g – ширина забороненої зони напівпровідника; k – стала Больцмана; N_c, N_v – ефективна густина станів поблизу дна ЗП і біля стелі ВЗ відповідно; їх визначають за формулами (19.35) і (19.36).

Рухливість u_n й u_p різна за величиною внаслідок різниці ефективних мас m_n^* і m_p^* й часу вільного пробігу електрона і дірки (їх буде розглянуто далі), які залежать від механізмів розсіювання електронів і дірок у кристалічній ґратці.

Можна виділити декілька основних механізмів розсіювання носіїв заряду:

1) на теплових коливаннях структурних елементів кристалічної ґратки;

2) на іонізованих домішках (іони домішки);

3) на нейтральних домішках (атом домішки);

4) на дефектах кристалічної ґратки;

5) на носіях заряду.

Четвертим і п'ятим видами розсіювання в теорії напівпровідників

найчастіше нехтують. За високих температур (кімнатній і вище) перший вид розсіювання відіграє основну роль. Оскільки абсолютно чистих з ідеальною кристалічною ґраткою напівпровідників не існує, то, на відміну від металів, рухливіст

$$u_n \sim T^{-3/2} \text{ та } u_p \sim T^{-3/2}. \quad (19.45)$$

Тоді формулу (19.43) з урахуванням (19.45) та (19.44) перепишемо так:

$$\sigma_i = \sigma_{0i} \exp\left\{-\frac{W_g}{2kT}\right\}, \quad (19.46)$$

де $\sigma_{0i} = e\sqrt{N_c N_v}\left(u_n + u_p\right)$ – стала, яка майже не залежить від температури. Якщо формулу (19.46) прологарифмувати, то

$$\ln \sigma_i = \ln \sigma_{0i} - \frac{W_g}{2k} \cdot \frac{1}{T}, \quad (19.47)$$

тобто залежність $\ln \sigma_i$ від оберненої температури $1/T$ є лінійною (за умови, що W_g не залежить від температури), що дає можливість визначити коефіцієнт $\frac{W_g}{2k}$, який містить такий важливий параметр напівпровідника, як ширина забороненої зони W_g.

Домішковий напівпровідник. Найбільш цікаві результати технічного застосування отримали напівпровідники, електричними властивостями яких можна управляти, вводячи в них хімічні домішки. Так, якщо ввести у кристалічну ґратку германію і кремнію незначну кількість домішкових атомів (на 100 000 атомів кристалічної ґратки один атом домішки), електричні властивості напівпровідників різко змінюються.

Атоми домішок називають донорами, якщо, перебуваючи у кристалічній ґратці напівпровідника, вони створюють заповнені електронами рівні енергії, які розміщені у забороненій зоні і з яких електрони можуть перейти у зону провідності. Класичним прикладом донорного напівпровідника є германій ${}_{32}\mathrm{Ge} - 1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^2 4p^2$ і кремній ${}_{14}\mathrm{Si} - 1s^2 2s^2 2p^6 3s^2 3p^2$, в яких частина атомів напівпровідника заміщена п'ятивалентними домішковими атомами (фосфору ${}_{15}\mathrm{P} - 3\mathrm{s}^2 3p^3$, миш'яку ${}_{33}\mathrm{As} - 4\mathrm{s}^2 4p^3$, сурми ${}_{51}\mathrm{Sb} - 5\mathrm{s}^2 5p^3$). Чотири валентних електрони кожного домішкового атома беруть участь в утворенні хімічних зв'язків із сусідніми атомами ґратки, п'ятий валентний електрон залишається зв'язаним із своїм атомом. Утім, енергія цього зв'язку через «відволікаючий» вплив інших атомів всередині кристала набагато менша, ніж для вільного атома, і становить значення приблизно 10^{-2} еВ, яку називають *енергією іонізації (активації)* донора і позначають як ΔW_n (рис. 19.30). У результаті якої-небудь дії

домішковий атом може бути легко іонізований, і п'ятий електрон стає вільним (рис. 19.30, процеси 1). На мові енергетичної моделі напівпровідника процес іонізації означає утворення рухливого електрона в зоні провідності.

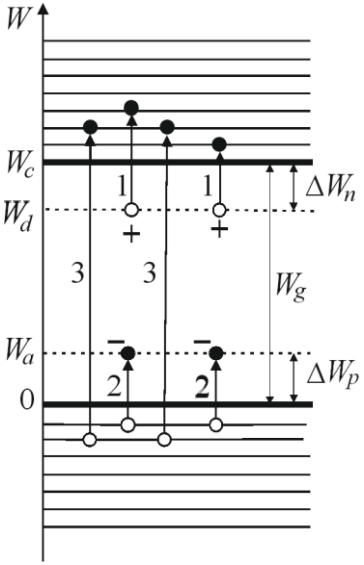


Рис. 19.30. Механізм теплової генерації рухливих носіїв заряду в домішкових напівпровідниках; світлими кружками позначено дірки, чорними – електрони; знаками «+» і «−» позначено іонізовані донори й акцептори; ΔW_n – енергія активації донорних домішок; ΔW_p – енергія активації акцепторних домішок; W_d та W_a – енергія донорних і акцепторних рівнів відповідно

Подібну ситуацію спостерігають у напівпровідниках четвертої групи після заміщення деякої частини їх атомів атомами третьої групи (алюміній ${}_{13}\mathrm{Al} - 2s^2 2p^1$, галій ${}_{31}\mathrm{Ga} - 4s^2 4p^1$, індій ${}_{49}\mathrm{In} - 5s^2 5p^1$). Три валентних електрони домішкового атома беруть участь в утворенні хімічного зв'язку з чотирма сусідніми атомами напівпровідника, при цьому один валентний зв'язок залишається вільним.

Ця вакансія під дією якого-небудь збудження може бути заповнена електроном із другого валентного зв'язку, у результаті чого в системі зв'язків утворюється вакантний стан, який може переміщуватись під дією, наприклад, електричного поля. Цей рух зручно інтерпретувати як рух деякої квазічастинки, яку називають ***діркою***. На мові енергетичної моделі такий процес відповідає переходу електрона з валентної зони на більш високий вакантний стан, пов'язаний з домішковим атомом (рис. 19.30, процеси 2). У результаті домішкий атом стає негативно зарядженим іоном (енергію активації (іонізації) такої домішки позначають як ΔW_p на рис. 19.30), а у валентній зоні утворюється дірка. Вищесказане пояснює, чому такий атом називають *акцептором*, а напівпровідник з такою домішкою *акцепторним*. Значення енергії іонізації (активації) донорних та акцепторних атомів в германії і кремнії наведені у табл. 19.1.

Таблиця 19.1. Енергія іонізації домішок

Акцептор	Енергія іонізації, еВ	
	Германій	Кремній
B	0,0104	0,045
Al	0,0102	0,057
Ga	0,0108	0,065
In	0,0112	0,16
Tl	0,0131	

Донор	Енергія іонізації, еВ	
	Германій	Кремній
P	0,0120	0,044
As	0,0127	0,049
Sb	0,096	0,039
Bi	0,0127	0,039

Таким чином, у домішкових напівпровідниках рухливі носії заряду можуть з'явитись не тільки у результаті порушення валентних зв'язків в основному матеріалі (процеси 3 на рис. 19.30), але і за рахунок іонізації домішкових атомів. Іонізація донорних атомів приводить до утворення вільних електронів у зоні провідності, а іонізація акцепторних атомів – до утворення дірок у валентній зоні. На енергетичній діаграмі (рис. 19.30) ці процеси зображено як переходи електронів з донорних локальних рівнів у зону провідності (процеси 1), переходи електронів із валентної зони на акцепторний рівень (процеси 2), а також міжзонні переходи (процеси 3), за яких генеруються і дірки, й електрони. Інтенсивність різних переходів залежить, насамперед, від температури, енергій іонізації та ширини забороненої зони.

Ймовірність іонізації домішкових атомів, які мають енергію іонізації приблизно 10^{-2} еВ, стає помітною за температур в одиниці й десятки градусів Кельвіна, а інтенсивність міжзонних переходів, які залежать від ширини забороненої зони, суттєві за сотень градусів, тому для домішкових напівпровідників розгляда-ють умовно область низьких температур ($kT < \Delta W_n$, ΔW_p) й область високих температур ($kT > \Delta W_n$, ΔW_p).

Область низьких температур ($kT < \Delta W_n$, ΔW_p). Донорному напівпро-віднику або напівпровіднику n-типу і відповідним величинам будемо присвоювати індекс n. За нульової температури всі стани валентної зони і всі домішкові стани донорних атомів (коли немає акцепторної домішки) зайняті. Підвищення температури приводить до того, що вже за абсолютної температури у декілька градусів з'являється помітна ймовірність іонізації донорних атомів з енергією активації $\Delta W_n \sim 10^{-2}$ еВ. Цей процес викликає появу рухливих носії тільки одного виду

– електронів у зоні провідності. Їх концентрація у цьому випадку буде дорівнювати концентрації нерухомих позитивно іонізованих донорів N_d^+:

$$n_n = N_d^+. \quad (19.48)$$

Генерацією електронно-діркових пар (процеси 3 на рис. 19.30) можна знехтувати. Вважаючи, що концентрація дірок $p_n \ll N_d^+$, (19.48) можна вважати рівнянням електронейтральності. Така якісна картина процесів дозволяє зрозуміти, чому в донорному напівпровіднику електрони називають *основними* носіями, а дірки – *неосновними*.

В акцепторному напівпровіднику *основними* носіями будуть дірки, які утворюються у валентній зоні внаслідок переходів електронів на домішкові акцепторні стани (іонізація домішки) і в зону провідності, а *неосновними* – електрони у зоні провідності.

Концентрацію основних носіїв заряду в цьому випадку визначають за такими формулами:

– для *n*-напівпровідника

$$n_n = \frac{2N_d}{\left(1+\sqrt{1+\frac{8N_d}{N_c}\exp\left\{\frac{\Delta W_n}{kT}\right\}}\right)}; \quad (19.49)$$

– для *p*-напівпровідника

$$p_p = \frac{2N_a}{\left(1+\sqrt{1+\frac{8N_a}{N_v}\exp\left\{\frac{\Delta W_p}{kT}\right\}}\right)}. \quad (19.50)$$

Для області дуже низьких температур замість (19.49) і (19.50) можна записати більш прості вирази:

$$n_n = \sqrt{\frac{N_c N_d}{2}}\exp\left\{-\frac{\Delta W_n}{2kT}\right\}, \quad (19.51)$$

$$p_p = \sqrt{\frac{N_v N_a}{2}}\exp\left\{-\frac{\Delta W_p}{2kT}\right\}, \quad (19.52)$$

де N_d і N_a – концентрації донорних і акцепторних атомів відповідно.

Провідність для області низьких температур для *n*- і *p*-напівпровідників

$$\sigma_n = \sigma_{0n}\exp\left\{-\frac{\Delta W_n}{2kT}\right\}, \quad (19.53)$$

де

$$\sigma_{0n} = e\sqrt{\frac{N_c N_d}{2}}$$

та

$$\sigma_p = \sigma_{0p}\exp\left\{-\frac{\Delta W_p}{2kT}\right\}, \quad (19.54)$$

де

$$\sigma_{0p} = e\sqrt{\frac{N_v N_a}{2}}u_p.$$

У цих формулах σ_{0n} і σ_{0p} – сталі, які майже не залежать від температури.

Якщо формулу (19.53) прологарифмувати, то

$$\ln\sigma_n = \ln\sigma_{0n} - \frac{\Delta W_n}{2k}\frac{1}{T}, \quad (19.55)$$

тобто $\ln\sigma_n$ лінійно залежить від оберненої температури $1/T$. Це дає змогу

визначити коефіцієнт лінійної залежності $\frac{\Delta W_n}{2k}$ і таким чином визначити енергію активації домішки ΔW_n. Описане вище однаковою мірою стосується і напівпровідника *p*-типу.

Область високих температур $\left(kT > \Delta W_n,\ kT > \Delta W_p\right)$**.** У цій області температур усі домішкові атоми іонізовані й рівняння електронейтральності для напівпровідника *n*-типу набуває вигляду

$$n_n = N_d + p_n\,. \qquad (19.56)$$

Із закону діючих мас (19.38) можемо записати:

$$n_n p_n = n_i^2. \qquad (19.57)$$

Розв'язуючи систему рівнянь (19.56) і (19.57) відносно невідомих n_n і p_n, отримуємо

$$n_n = \frac{N_d}{2}\left(1 + \sqrt{1 + \left(\frac{2n_i}{N_d}\right)^2}\right), \qquad (19.58)$$

а

$$p_n = \frac{n_i^2}{n_n}\,. \qquad (19.59)$$

Для напівпровідника *p*-типу за аналогією запишемо відповідні формули:

$$p_p = \frac{N_a}{2}\left(1 + \sqrt{1 + \left(\frac{2n_i}{N_a}\right)^2}\right), \qquad (19.60)$$

$$n_p = \frac{n_i^2}{p_p}. \qquad (19.61)$$

Примітка. Концентрацію власних носіїв заряду n_i для конкретного напівпровідника знаходять за табличними даними.

Практичний інтерес має випадок сильного легування, коли $N_d \gg n_i$ або $N_a \gg n_i$. Тоді концентрація основних носіїв заряду згідно з (19.58) і (19.60) знайдемо за формулами

$$n_n \approx N_d, \qquad (19.62)$$

$$p_p \approx N_a. \qquad (19.63)$$

Згідно із законом діючих мас концентрація неосновних носіїв заряду (дірок) у донорному напівпровіднику становитиме

$$p_n \approx \frac{n_i^2}{N_d}, \qquad (19.64)$$

а, віповідно, концентрація неосновних електронів в акцепторному напівпровіднику

$$n_p \approx \frac{n_i^2}{N_a}. \qquad (19.65)$$

Положення рівня Фермі в домішковому напівпровіднику знайдемо для випадку сильного легування (цей випадок має практичний інтерес). Для напівпровідника *n*-типу, підставляючи (19.62) в (19.33), отримаємо

$$N_d = N_c \exp\left\{-\frac{W_c - W_{F_n}}{kT}\right\}.$$

Звідси знайдемо, що

$$W_{F_n} = W_c - kT\ln\left(\frac{N_c}{N_d}\right). \qquad (19.66)$$

Для напівпровідника *p*-типу, підставляючи (19.63) в (19.34), отримаємо

$$N_a = N_v \cdot \exp\left\{-\frac{W_{F_p} - W_v}{kT}\right\}.$$

Звідси знайдемо:

$$W_{F_p} = W_v + kT\ln\left(\frac{N_v}{N_a}\right). \quad (19.67)$$

Електропровідність домішкового напівпровідника. У загальному випадку електропровідність напівпровідника n-типу буде

$$\sigma = \sigma_{0i}\exp\left\{-\frac{W_g}{2kT}\right\} + \sigma_{0n}\exp\left\{-\frac{\Delta W_n}{2kT}\right\}, \quad (19.68)$$

де перший доданок є власною провідністю в області високих температур, а другий доданок є домішковою провідністю в області низьких температур. На рис. 19.31 показано схематичну залежність $\ln\sigma$ від оберненої температури $1/T$ для домішкового напівпровідника.

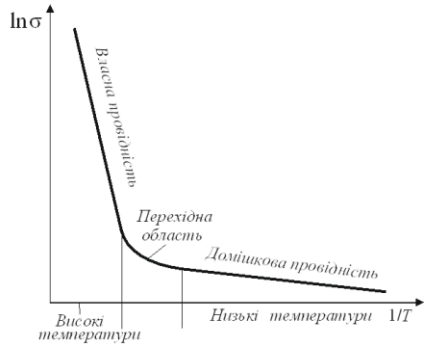


Рис. 19.31. Схематична залежність $\ln\sigma$ *від оберненої температури 1/T для домішкового напівпровідника*

Оскільки $\Delta W_n < W_g$, то домішкова і власна провідність, як правило, реалізуються за різних температур, оскільки електрони використовують енергію теплового руху. Домішкова провідність реалізується за більш низьких температур, за яких енергія руху кристалічної ґратки вже достатня для генерації домішкових носіїв, а тому власною провідністю (перша складова у формулі (19.68)) можна знехтувати.

Навпаки, за більш високої температури, коли домішкові центри вже себе вичерпали (повністю іонізовані), починає діяти механізм генерації власних носіїв. Оскільки власних атомів завжди більше, ніж домішкових ($10^6 : 1$), то за більш високих температур власних носіїв утворюється значно більше, ніж домішкових, і тепер можна знехтувати домішковою провідністю (друга складова у формулі (19.68)). За проміжних температур можуть працювати обидва механізми провідності (рис. 19.31, перехідна область).

19.11. Внутрішній фотоефект

Світло має корпускулярно-хвильову природу: з одного боку йому притаманні хвильові властивості, що обумовлюють явище інтерференції, дифракції, поляризації, з другого –

корпускулярні властивості. Одне з явищ, в якому проявляються корпускулярні властивості – це фотоелектричний ефект. Він виникає у результаті взаємодії квантів електромагнітного випромінювання з електронами речовини.

Залежно від наслідків такої взаємодії розрізняють два види фотоефекту:

1. Зовнішній фотоефект (фотоелектронна емісія), за якого поглинання електромагнітного випромінювання супроводжується випусканням електронів речовиною (див. підрозд. 16.5).

2. Внутрішній фотоефект, за якого через поглинання електромагнітного випромінювання відбувається лише перерозподіл електронів за енергетичними станами без їхнього випускання конденсованим середовищем.

Зовнішній фотоефект (див. підрозд. 16.5) виникає, якщо енергія квантів електромагнітного випромінювання більша або дорівнює роботі виходу електронів з поверхні речовини: $h\nu \geq \Phi$ (формула (16.38)). Електрон, отримавши енергію, може досягнути поверхні тіла, подолати потенціальний бар'єр, створений внутрішнім електричним полем, і вийти у вакуум або в інше середовище.

Внутрішній фотоефект відбувається у тих випадках, коли енергії фотонів недостатньо для вибивання електронів з речовини. Якщо фотон спроможний лише збудити атом, то це може викликати зміну діелектричної проникності середовища – *фотодіелектричний ефект*. Якщо енергії фотонів достатньо для генерації електронів і дірок провідності у напівпровіднику, тоді виникає явище фотопровідності. *Зміну електричного опору напівпровідника під дією випромінювання називають внутрішнім фотоелектричним ефектом, або фоторезистивним ефектом.* Додаткова провідність, яка обумовлена носіями заряду, що створюються випромінюванням, носить назву фотопровідності.

У неоднорідних напівпровідниках в області p-n – переходів у результаті внутрішнього фотоефекту виникає електрорушійна сила (*фото е.р.с*) – *вентильний ефект*.

Розглянемо детальніше механізм виникнення та основні закономірності явища фотопровідності у напівпровідниках. Якщо температура напівпровідника $T > 0$, то завдяки енергії теплового руху навіть в неосвітленому зразку завжди існує деяка кількість вільних електронів і дірок, які перебувають в

термодинамічній рівновазі з кристалічною ґраткою. Їх концентрація n_0 і p_0 визначає електропровідність напівпровідника, коли він неосвітлений (темнова електропровідність) – σ_0. Під час освітлення напівпровідника можуть виникнути додаткові носії заряду, не обумовлені термодинамічною рівновагою із кристалічною ґраткою. Такі носії називають *нерівноважними*. Виникнення цих носіїв проявиться у зміні електропровідності напівпровідника, оскільки він отримає додаткову електропровідність $\Delta\sigma_\phi$.

На рис. 19.32 показано схему утворення електронів і дірок провідності у напівпровідниках різного виду. Під час освітлення напівпровідника деяка частина фотонів поглинається кристалічною ґраткою і вільними носіями, у результаті чого збільшується їхній тепловий рух, але нових носіїв не утворюється. Деяка частина фотонів віддає свою енергію електронам валентної зони. Якщо енергія фотона дорівнює ширині забороненої зони або більша за неї ($h\nu \geq W_g$), тоді електрон, що отримав цю енергію, переходить у зону провідності й може брати участь у створенні фотоструму. Одночасно з переходом електрона в зону провідності у валентній зоні створюється незаповнений рівень, тобто виникає дірка. Таким чином, один поглинутий фотон з енергією ($h\nu \geq W_g$) створює пару «електрон – дірка» (рис. 19.32, *а*). Це процеси генерації фотоносіїв.

Реальними носіями струму в більшості напівпровідників є не електрони й дірки одночасно, а лише ті з них, у яких більший час життя. В одних випадках це можуть бути електрони, в інших – дірки.

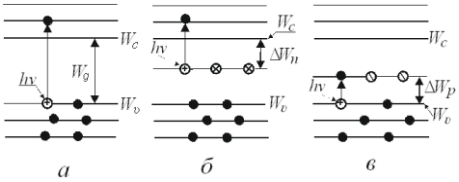


Рис. 19.32. Схема утворення фотоносіїв у напівпровіднику: а – у власному; б – у донорному; в – в акцепторному; W_g – ширина забороненої зони; ΔW_n – енергія активації донорів; ΔW_p – енергія активації акцепторів; ● – електрон; ⊕ – дірка; ⊗ – атом донора; ∅ – атом акцептора

Утворені світлом додаткові носії струму деякий проміжок часу (час життя $\tau = 10^{-3} \ldots 10^{-7}$ с) перебувають у зоні провідності, а потім знову повертаються на одне з місць (дірки), залишене іншими електронами або приєднуються до надлишкових атомів домішки. Це процеси рекомбінації фотоносіїв. Під час неперервного освітлення напівпровідника процеси генерації та рекомбінації

відбуваються одночасно. У результаті встановлюється така динамічна рівновага, за якої постійно існують деякі додаткові концентрації електронів Δn і дірок Δp. Так виникає фотопровідність.

Надлишкові електрони Δn і дірки Δp можуть мати енергію значно більшу, ніж середня енергія рівноважних носіїв заряду. Утім, в результаті взаємодії з фононами (коливаннями) і дефектами кристалічної ґратки енергія нерівноважних носіїв заряду стає однаковою із середньою тепловою енергією рівноважних носіїв заряду. Цей процес відбувається за час 10^{-10} ... 10^{-12} с. Оскільки час життя нерівноважних носіїв заряду значно більший від цієї величини, то протягом всього часу життя кінетична енергія нерівноважних носіїв заряду відповідає середній тепловій енергії рівноважних носіїв заряду. *Отже, можна вважати, що розподіл за енергіями нерівноважних носіїв заряду в зонах є таким самими, як і рівноважних*. Значить і рухливість нерівноважних носіїв заряду не відрізняються від рухливості рівноважних.

Для створення додаткових носіїв у домішкових напівпровідниках фотон повинен мати енергію $h\nu \geq \Delta W_n\,(\Delta W_p)$, де $\Delta W_n\,(\Delta W_p)$ – енергія активації атомів домішки (донорів або акцепторів). У напівпровідниках з донорними домішками фотон викликає перехід електронів з донорних рівнів у зону провідності (рис. 19.32, *б*), унаслідок чого концентрація «вільних» електронів у напівпровіднику підвищується, а електропровідність збільшується. У напівпровідниках з акцепторними домішками фотон викликає перехід електронів з валентної зони на домішкові рівні (рис. 19.32, *в*) з утворенням негативно заряджених іонів домішки. У валентній зоні при цьому утворюються дірки, що забезпечують додаткову фотопровідність.

Таким чином, якщо

$$h\nu \geq W_g \qquad (19.69)$$

для власних напівпровідників і

$$h\nu \geq \Delta W_n\,(\Delta W_p) \qquad (19.69^*)$$

для домішкових напівпровідників, то під час освітлення напівпровідника в ньому з'являються додаткові носії заряду і його провідність збільшується.

Якщо енергія фотона

$$h\nu < W_g \qquad (19.70)$$

для власних і

$$h\nu < \Delta W_n\,(\Delta W_p) \qquad (19.70^*)$$

для домішкових напівпровідників фотопровідність не спостерігається. З умови (19.69) випливає, що має існувати *довгохвильова червона межа* (*границя*) фотопровідності – така довжина хвилі

світла λ_0, що світло з більшою довжиною хвилі фотопровідність не викликає.

Для власних напівпровідників

$$\lambda_0 = \frac{hc}{W_g}. \quad (19.71)$$

Для домішкових напівпровідників

$$\lambda_0 = \frac{hc}{\Delta W_n},\ \lambda_0 = \frac{hc}{\Delta W_p}. \quad (19.72)$$

Іноді довгохвильову межу фотопровідності називають червоною межею. Назва ця умовна, оскільки вперше виміряні значення величини λ_0 для зовнішнього фотоефекту припадали на червону область видимого спектра.

Для власних напівпровідників ширина забороненої зони W_g становить одиниці електронвольт, тому для них довгохвильова межа фотопровідності може належати до видимої області спектра. В домішкових напівпровідниках енергія активації домішок $\Delta W_n W_p$ становить десяті й навіть соті частки електронвольта. Довгохвильова межа фотопровідності для них міститься в інфрачервоній частині спектра.

Розглянемо деякі основні закономірності, які визначають явище фотопровідності й проаналізуємо їх. Зробимо це на прикладі власного напівпровідника.

До освітлення напівпровідника його електропровідність визначають концентрацією і рухливістю електронів і дірок у рівноважному стані:

$$\sigma_0 = en_0u_n + ep_0u_p, \quad (19.73)$$

де u_n та u_p – рухливість; n_0 і p_0 – рівноважна концентрація електронів і дірок відповідно.

Під час освітлення напівпровідника внаслідок поглинання випромінювання відбуваються міжзонні переходи, що збільшує концентрацію електронів на величину Δn, а дірок – на величину Δp. У загальному випадку може змінитись і рухливість носіїв, але вплив зміни рухливості буде помітним лише за дуже низьких температур, тому будемо вважати, що фотопровідність викликана змінами концентрації носіїв.

Електропровідність напівпровідника в умовах освітлення

$$\sigma = e(n_0 + \Delta n)u_n + e(p_0 + \Delta p)u_p, \quad (19.74)$$

а зміну електропровідності або фотопровідність визначають як

$$\Delta\sigma_ф = \sigma - \sigma_0 = e\Delta nu_n + e\Delta pu_p. \quad (19.75)$$

Якщо власний напівпровідник не містить будь-яких домішок і дефектів, а швидкість рекомбінації електронів і дірок однакова, тоді можна вважати, що $\Delta n = \Delta p$, і (19.75) запишемо так:

$$\Delta\sigma_{\phi} = e\Delta p\left(u_n + u_p\right). \quad (19.76)$$

Тут під $\Delta\sigma_{\phi}$ і Δp розуміють стаціонарні значення величин, тобто розглядають $\Delta\sigma_{\phi}$ і Δp після завершення всіх перехідних процесів і встановлення стаціонарної фотопровідності. Як видно з (19.76), величина фотопровідності буде визначатися концентрацією нерівноважних носіїв заряду $\Delta n = \Delta p$, які, у свою чергу, залежать від процесів генерації та рекомбінації.

Нехай на напівпровідникову пластину, товщина якої d, перпендикулярно до поверхні падає монохроматичне випромінювання, частота якого ν, а інтенсивність I (рис. 19.33). Під інтенсивністю розуміють енергію, яка падає за одиницю часу на одиницю площі поверхні зразка. Якщо $I(x)$ – інтенсивність всередині зразка, тоді за одиницю часу в одиниці об'єму буде поглинатися енергія, яка дорівнює $W = \alpha I(x)$, де $\alpha(\nu)$ – коефіцієнт поглинання (розмірність $[\alpha] = \text{м}^{-1}$). Поділимо цю енергію на енергію одного кванта випромінювання. Отримаємо кількість квантів світла (фотонів), що містяться в одиниці об'єму зразка в одиницю часу. Будемо мати на увазі, що не кожен фотон взаємодіє з електронами валентної зони, тому вводять поняття квантового виходу. Тоді швидкість генерації електронно-діркових пар, тобто кількість електронно-діркових пар, що генеруються за одиницю часу в одиниці об'єму,

$$G = \frac{\alpha I(x)}{h\nu}\eta, \quad (19.77)$$

де $\eta(\nu)$ – *квантовий вихід* внутрішнього фотоефекту, який визначає кількість пар носіїв заряду (або кількість носіїв заряду при домішковій фотопровідності), що утворюється одним фотоном, який поглинувся, якщо інтенсивність світла вимірюється кількістю квантів енергії $h\nu$, що падає за секунду на одиницю поверхні. Підкреслимо, що η та α залежать від частоти (довжини хвилі) випромінювання, а I різна залежно від товщини пластини. Квантовий вихід $\eta(\nu)$ враховує, що частина фотонів може передати свою енергію ґратці або вільним носіям зразка, не утворюючи нових електронів і дірок, тому η буде меншим за одиницю. Іноді можливі випадки, коли $\eta > 1$. Наприклад, якщо енергія квантів випромінювання дуже велика ($h\nu >> W_g$), тоді вільний електрон, який з'явився внаслідок поглинання енергії фотона, може мати велику кінетичну енергію і сам буде створювати нові носії заряду завдяки ударній іонізації.

Величина $I(x)$ у загальному випадку змінюється за складним законом

через велику кількість відбивань від різних поверхонь і можливих інтерференційних ефектів.

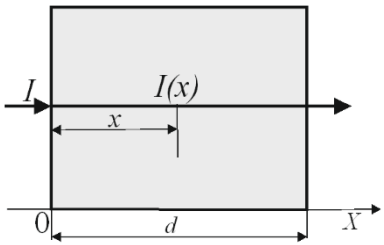


Рис. 19.33. Проходження світла через пластину напівпровідника

У випадку тонкої плівки ($\alpha d << \ll 1$, $d -$ товщина плівки, рис. 19.34) можна вва-жати, що $I(x) = I$, тобто випромінюван-ня буде поглинатись у всьому зразку рівномірно, а швидкість генерації

$$G = \frac{\eta \alpha I}{h\nu}. \qquad (19.78)$$

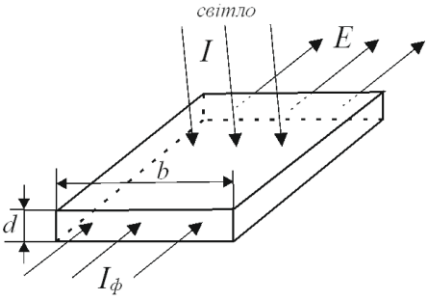


Рис. 19.34

Розглянувши динаміку утворення нерівноважних носіїв (фотоелектронів і фотодірок), можна показати, що у стаціонарних умовах їх концентрацію визначають співвідношенням

$$\Delta n = \Delta p = G\,\tau_{\text{р}} = \frac{\eta \alpha I \tau_{\text{р}}}{h\nu}, \qquad (19.79)$$

де $\tau_{\text{р}} = \tau_{\text{n}}$ – час життя дірок і електронів. Тоді фотопровідність (з урахуванням (19.76) і (19.79)) можна записати так:

$$\Delta\sigma_{\text{ф}} = e\frac{\eta \alpha I \tau_{\text{р}}}{h\nu}\left(u_n + u_p\right), \qquad (19.80)$$

тому для фотоструму (з урахуванням закону Ома $I_ф = jS = \Delta\sigma_{\text{ф}} Ebd$) запишемо

$$I_ф = e\left(u_n + u_p\right)bd\frac{\eta \alpha I \tau_{\text{р}}}{h\nu}IE, \qquad (19.81)$$

де b і d – ширина і товщина зразка відповідно; E – напруженість електричного поля вздовж зразка (рис. 19.34). Формула (19.81) дозволяє проаналізувати основні закономірності явища фотопровідності. Вона має дві величини, які залежать від інтенсивності світла I: по-перше сама інтенсивність світла I, по-друге, час життя нерівноважних пар $\tau_{\text{р}}$. Час життя нерівноважних пар у загальному випадку обернено пропорційний концентрації носіїв: $\tau_{\text{р}} \sim \frac{1}{n_0 + p_0 + \Delta p}$, тому за малих інтенсивностей світла $\Delta p << p_0$, $\tau_{\text{р}}$ буде сталою величиною. Тоді фотострум $I_{\text{ф}}$ має лінійно залежати від інтенсивності світла (рис. 19.35, область *а*). За великих інтенсивностей світла, коли $\Delta p \geq p_0$, час життя

τ_p змінюється зі зміною I, а залежність $I_\phi = f(I)$ стає нелінійною (рис. 19.35, область *б*). З формули (19.81) випливає, що величина фотоструму має залежати від частоти (або довжини хвилі) світла. Ця залежність має назву *спектральної характеристики*. Вигляд її може бути дуже складним, тому що від довжини хвилі залежить і квантовий вихід η, і коефіцієнт поглинання α, й енергія фотонів $h\nu$.

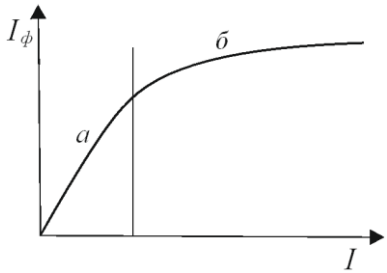


Рис. 19.35. Залежність фотоструму I_ϕ від інтенсивності світла I: а – область малої інтенсивності ($\tau_p = \text{const}$); *б – область великої інтенсивності* ($\tau_p \neq \text{const}$)

Також на вигляд залежності $I_\phi = f(\lambda)$ впливають матеріал напівпровідника, якість поверхні та розміри (товщина зразка). Типовий вигляд спектральної характеристики $I_\phi(\lambda)$ напівпровідникового зразка зображено на рис. 19.36. Спад у довгохвильовій області пов'язаний з існуванням довгохвильової (червоної) межі фотопровідності λ_0. Спад в області малих довжин хвиль визначається збільшенням коефіцієнта поглинання α і інтенсивним поглинанням випромінювання в тонкому поверхневому шарі напівпровідника, що викликає зменшення швидкості об'ємної генерації G і зменшення концентрації нерівноважних носіїв, тобто зниження фотопровідності. Свій внесок у залежність $I_\phi(\lambda)$ робить також складна залежність η(λ).

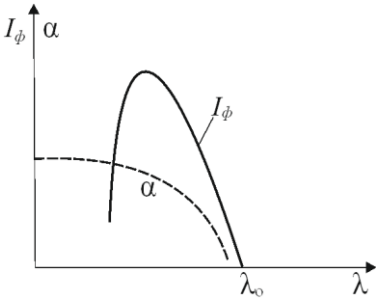


Рис. 19.36. Залежність фотоструму I_ϕ від довжини хвилі випромінювання λ

Релаксація фотопровідності

Розглянемо зміну нерівноважних носіїв заряду концентрацій Δn та Δp, наприклад у результаті опромінення напівпровідникового зразка світлом, інтенсивність якого змінюється так, як зображено на рис. 19.37, *а*, тобто періодично стрибком інтенсивність світла зростає від 0 до I_{max}, потім залишається сталою деякий час Δt, після

чого знову стрибком зменшується до 0. Потім це все багаторазово повторюється.

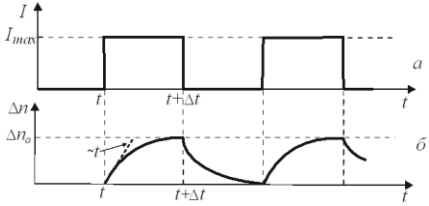


Рис. 19.37. Зміна концентрації нерівноважних носіїв заряду Δn *опроміненням напівпровідникового зразка світлом*

До опромінення зразка концентрація електронів (або дірок) рівноважна, тобто дорівнює n_0 (або p_0). На початку опромінення зразка починається процес генерації нерівноважних носіїв із швидкістю *G*, яка визначається кількістю носіїв, що утворюються в одиниці об'єму за одиницю часу. Якщо інтенсивність світла I_{max} не дуже велика і населеність енергетичних рівнів далека від насичення, то швидкість генерації носіїв визначають формулою (19.77) і не буде змінюватися з часом, оскільки величини, що її визначають, не змінюються з часом. Концентрація нерівноважних носіїв буде зростати за лінійним законом відносно часу (на рис. 19.37, *б* зображено пунктиром):

$$\Delta n = Gt. \qquad (19.82)$$

Утім, одночасно з процесом генерації починається також процес рекомбінації електронів. За малого рівня освітлення, для якого $\Delta n << n_0+p_0$, швидкість рекомбінації

$$R = \frac{\Delta n(t)}{\tau}, \qquad (19.83)$$

де $\tau = \frac{1}{r(n_0+p_0)}$ – час життя фотоносіїв; *r* – коефіцієнт рекомбінації.

Зміна концентрації надлишко-вих носіїв заряду за одиницю часу ($\frac{d(\Delta n)}{dt}$) є різницею між швидкістю генерації і рекомбінації носіїв заряду, тому з урахуванням (19.77) і (19.83)

$$\frac{d(\Delta n)}{dt} = G - R = \frac{\eta\alpha I}{h\nu} - \frac{\Delta n(t)}{\tau}. \qquad (19.84)$$

Розділяючи змінні ($\Delta n(t)$ та t) й інтегруючи з урахуванням початкової умови $\Delta n(0) = 0$ за $t = 0$, отримаємо вираз, який характеризує зміну концентрації надлишкових електронів з часом для процесу наростання кількості фотоносіїв:

$$\Delta n = \eta\alpha\tau\frac{I}{h\nu}\left(1 - \exp\left\{-\frac{t}{\tau}\right\}\right) = \\ = \Delta n_0\left(1 - \exp\left\{-\frac{t}{\tau}\right\}\right). \qquad (19.85)$$

Формула (19.85) пояснює процес наростання фотоносіїв на рис. 19.37, *б*.

Розглянемо, як буде змінюватись Δn після вимкнення опромінення, тобто після моменту часу $t + \Delta t$. Тоді $G = 0$ й (19.84) матиме вигляд

$$\frac{d(\Delta n)}{dt} = -\frac{\Delta n}{\tau}. \qquad (19.86)$$

Інтегрування цього рівняння з урахуванням початкової умови, що за

$t + \Delta t = 0$ $\Delta n = \Delta n_0$, дає для моменту часу, більшого від $t + \Delta t$, розв'язок

$$\Delta n = \Delta n_0 \exp\{-t/\tau\}, \quad (19.87)$$

де

$$\Delta n_0 = \eta\alpha\tau\frac{I}{h\nu}. \quad (19.88)$$

Величина τ має розмірність часу і за своїм значенням в рівнянні (19.87) є характерним часом процесу зменшення нерівноважної концентрації Δn, оскільки за $t = \tau$

$$\Delta n = \frac{\Delta n_0}{e}, \quad (19.89)$$

де e = 2,71828… є основою натурального логарифма, тобто за час τ концентрація фотоносіїв (нерівноважних носіїв) без генерації зменшиться в e разів. Отже, τ може бути *мірою тривалості життя* нерівноважних носіїв заряду.

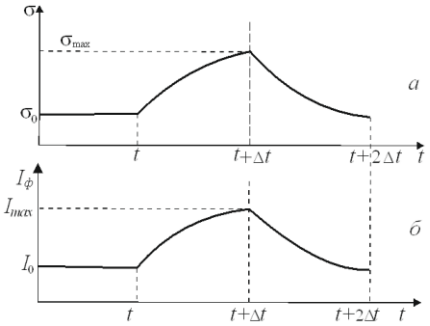


Рис. 19.38. Залежність електропровідності (а) і сили фотоструму (б) через зразок напівпровідника: σ_0 – темнова електропровідність; I_0 – темновий струм; σ_{max} – максимальна провідність; I_{max} – максимальний струм

Оскільки Δn змінюється за експоненційним законом (19.87), то за таким самим законом буде змінюватись і електропровідність σ та фотострум $I_ф$ під впливом світла (рис. 19.38) для моменту часу, який більший за момент часу t +Δt.

Для часу $t' > t + \Delta t$ можемо записати, що струм фотопровідності

$$I_\phi = (I_{max} - I_0)\exp\left\{-\frac{t}{\tau}\right\}. \quad (19.90)$$

Розглянутий вище механізм зміни фотопровідності є досить приблизним й умовним, оскільки не враховує багатьох впливових факторів, зокрема, не враховано можливість різних механізмів рекомбінації, поверхневої рекомбінації і под.

Отже, реєструючи залежність фотоструму від часу, можна, застосувавши відповідний метод обробки даних, визначити (правильніше – оцінити) величину τ – час релаксації фотопровідності, який у випадку міжзонної рекомбінації (19.83) дорівнює часу життя нерівноважних носіїв.

19.12. Дрейф і дифузія носіїв заряду

Дрейф зарядів в електричному полі

Для вільних електронів і дірок життя – це рух, насамперед хаотичний тепловий рух, який існує завжди і відбувається тим енергійніше, чим вища

температура кристала. Вільні електрони й дірки під впливом теплових коливань кристалічної ґратки перебувають у стані безладного хаотичного руху. Середня енергія теплового руху дорівнює $\frac{3kT}{2}$ (k – стала Больцмана, T – абсолютна температура). Прирівнявши величину цієї енергії з кінетичною енергією частинки $\frac{mv^2}{2}$ можна знайти *середню швидкість* v_T хаотичного руху електронів (або дірок):

$$v_T = \sqrt{\frac{3kT}{m}}. \qquad (19.91)$$

Якщо T= 300 К, $m = m_e =$ 9,1× 10^{-31} кг, отримаємо для швидкості вільного електрона значення v_T~10^5 м/с. З такою великою швидкістю носії заряду зазнають 10^{12} ... 10^{13} раз зіткнень із ґраткою кристала, й тому перебувають у тепловій рівновазі з нею. Якщо до напівпровідника (металу) прикладено електричне поле, то на хаотичний тепловий рух вільних носіїв (електронів і дірок) накладається переносний або дрейфовий рух, який є результатом дії електричного поля. Механізм його виникнення такий. Вільні носії заряду при прикладеному електричному полі, як і раніше, з частотою ~ 10^{12} ... 10^{13} разів за секунду зазнають зіткнень із кристалічною ґраткою. Як і раніше, неможливо передбачити, куди полетить носій після чергового зіткнення: вперед, назад, вліво, вправо, вгору, вниз – під будь-яким кутом у будь-якому напрямку. Втім, куди б він не полетів, електричне поле буде, нехай і слабо, тягнути його завжди в одному й тому самому напрямку. Як результат виникає направлений (дрейфовий) рух.

В електричному полі, рухаючись без зіткнень, носій за час t (згідно з другим законом Ньютона $\Delta\vec{p} = \vec{F}\Delta t$), набуває імпульсу $m\vec{v} = e\vec{E}t$ та прискорення $\vec{a} = e\vec{E}/m$, де m – маса носія, $\vec{E}$ – напруженість електричного поля. Щоб правильно вирахувати середню швидкість, якої набуває носій в умовах багатократних зіткнень, слід пам'ятати про дві обставини.

По-перше, після зіткнення носій може з однаковою ймовірністю рухатись в будь-якому напрямку. Це означає, що швидкість направленого руху після зіткнення дорівнює нулю.

По-друге, оскільки процес зіткнення – процес випадковий, носій на проміжку між двома послідовними зіткненнями проводить у «вільному русі» різний час. Обидві ці обставини наочно видно на рис. 19.39, на якому показано залежність швидкості

направленого руху носія від часу. Середня швидкість направленого руху (швидкість дрейфу)

$$v_{др} = a\langle\tau\rangle = (e\langle\tau\rangle/m)E\,, \quad (19.92)$$

де ‹τ› – середній час між зіткненнями, $a = -eE/m$ – прискорення носія. Таким чином, швидкість направленого руху вільних носіїв у кристалі пропорційна напруженості електричного поля. Коефіцієнт пропорційності

$$u = e\langle\tau\rangle/m \quad (19.93)$$

називають *рухливістю носіїв.* Рухливість вільних носіїв є однією з найважливіших характеристик носіїв заряду у кристалі.

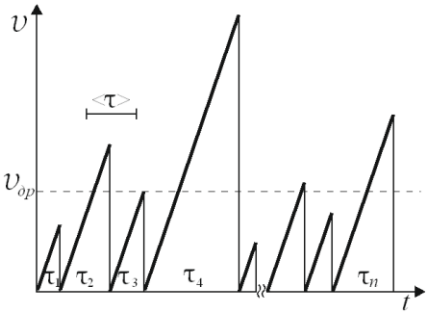


Рис. 19.39. залежність швидкості направленого руху носія заряду від часу

Із формул (19.92) і (19.93) випливає, що

$$v_{др} = uE. \quad (19.94)$$

Чим більша рухливість, тим більша швидкість направленого руху носіїв за тієї самої напруженості електричного поля E. Розмірність рухливості – м^2/(В·с). Із формули (19.94) видно, що рухливість носіїв заряду – це їхня дрейфова швидкість в одиничному електричному полі ($E = 1$ В/м).

Результатом направленого дрейфового руху електронів є дрейфовий струм, густина якого пропорційна направленому потоку частинок і їхньому заряду.

$$j_{др} = env_{др} = enuE, \quad (19.95)$$

де n – концентрація носіїв заряду.

Вираз (19.95) виражає закон Ома в диференціальній формі. Коефіцієнтом пропорційності між $j_{др}$ та E є питома електропровідність

$$\sigma = enu = e^2n\langle\tau\rangle/m. \quad (19.96)$$

У слабких електричних полях, коли швидкість дрейфу мала порівняно з тепловою ($v_{др} << v_T$, наприклад для міді $v_{др} \sim 5 \cdot 10^{-4}$ м/с), наявність чи відсутність поля ніяк не впливає на характер зіткнень носіїв із кристалічною ґраткою. При цьому рухливість u є сталою величиною, яка не залежить від напруженості поля E. Цей випадок відповідає виконанню закону Ома (19.95).

У сильних полях кінетична енергія носіїв, яку вони набувають в електричному полі, може зрівнятись і навіть стати більшою, ніж енергія хаотичного теплового руху. Частота і

характер зіткнень, яких зазнають тепер носії, залежить від їхньої енергії, тому в сильних електричних полях частота й характер зіткнень електронів і дірок (а отже й рухливість) залежать від напруженості поля, тому в сильних полях (коли $v_{др} = v_T$) закон Ома не виконується.

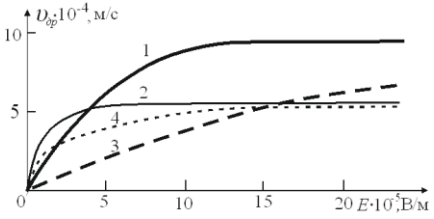


Рис. 19.40. Залежність швидкості направленого руху (дрейфової швидкості) носіїв від напруженості електричного поля в Ge та Si ($T = 300$ K): суцільні лінії – $v_{др}(E)$ для електронів; штрихові лінії – для дірок. 1, 3 – Si; 2, 4 – Ge

У всіх напівпровідниках, які використовують на практиці, за кімнатної температури рухливість у сильних полях зменшується з ростом напруженості поля E. У дуже сильних полях $u \sim \frac{1}{E}$. Це означає, що згідно з (19.94) у цьому випадку $v_{др} = \text{const}$. Залежність дрейфової швидкості від напруженості поля електронів і дірок показано на рис. 19.40. Видно, що у всіх випадках у слабких електричних полях швидкість пропорційна напруженості поля. Цю ділянку залежності $v_{др}$ (E) називають омічною або лінійною. З ростом поля дрейфова швидкість зростає дедалі повільніше, і в досить сильних електричних полях перестає залежати від напруженості поля E.

Дифузія носіїв заряду

Поняття «дифузія» походить від лат. diffusio (розмиття), його застосовують до дуже багатьох явищ у газах, рідинах і твердих тілах.

Якщо на поверхню напівпровідникової пластини, яку старанно очищено від домішок, нанести шар речовини, що містить домішки у великій кількості, то через деякий час цю домішку можна виявити в об'ємі напівпровідникової пластини «далеко» від поверхні напівпровідника. Домішка потрапляє в об'єм через процес дифузії. Швидкість дифузії домішки у напівпровіднику дуже сильно зростає зі зростанням температури, тому за кімнатної температури потрібні десятиліття, щоб домішка проникла в область напівпровідника, а за високої температури процес дифузії може зайняти всього декілька годин або навіть хвилин.

Таким чином, суть процесу дифузії – спонтанне (самопливом), дифузія відбувається без будь-яких зовнішніх впливів, переміщення

речовини з місця, де її багато, туди, де її мало. Процес дифузії є прямим наслідком хаотичного теплового руху атомів (або молекул). На рис. 19.41 показано розподіл концентрації частинок n вздовж осі x. Швидкість зміни концентрації з координатою характеризують похідною dn/dx, яка називається градієнтом концентрації і має напрямок в сторону збільшення концентрації n. Внаслідок хаотичного руху частинки з однаковою ймовірністю будуть проходити через одиничну плоску площину S (рис.19.41) зліва направо і навпаки. Але так як концентрація частинок зліва від площини більша, то буде переважаючий потік частинок через площину зліва направо, ніж справа наліво. Тому можна вважати, що дифузійний потік частинок (носіїв заряду) буде пропорційний dn/dx:

$$dN = -D\frac{dn}{dx} \qquad (19.97)$$

Знак мінус у (19.97) означає, що потік частинок направлений у протилежний бік до напрямку градієнта концентрації dn/dx (рис. 19.41). Коефіцієнт пропорційності D називають коефіцієнтом дифузії. Встановимо розмірність коефіцієнта дифузії. Потік частинок dN за кількістю дорівнює кількості частинок, які за одиницю часу перетинають одиничну площину S. Таким чином, розмірність потоку $[dN] = 1/(\text{м}^2 \cdot \text{с})$. Розмірність концентрації $[n] = \text{м}^{-3}$, а градієнта концентрації $[dn/dx]$, очевидно, м^{-4}, тому коефіцієнт дифузії D має розмірність $[D] = \text{м}^2/\text{с}$.

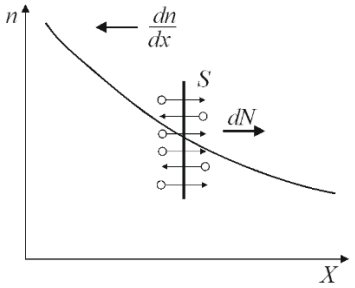


Рис. 19.41. розподіл концентрації частинок n вздовж осі x

Згідно з кінетичною теорією газів

$$D = \frac{1}{3}\langle l\rangle v_T, \qquad (19.98)$$

де $\langle l\rangle = \langle\tau\rangle v_T$ – *довжина вільного пробігу частинки*; ‹τ› – середній час між зіткненнями частинок; v_T – середня швидкість теплового руху.

Унаслідок дифузії виникає потік дірок та електронів з області з більш високою концентрацією в область з більш низькою концентрацією (рис. 19.42), але направлений потік заряджених частинок є електричним струмом. Густину дифузійного струму визначають потоком частинок (19.97) та їхнім зарядом. З урахуванням напрямків

вирази для дифузійних струмів дірок і електронів у напівпровідниках мають такий вигляд (рис. 19.42):

$$j_{D_p} = -eD_p \frac{dp}{dx}, \qquad (19.99)$$

$$j_{D_n} = eD_n \frac{dn}{dx}, \qquad (19.100)$$

де D_p і D_n – коефіцієнти дифузії дірок і електронів відповідно. Мінус у формулі (19.99) означає, що дифузійний струм і dp/dx мають протилежні напрямки.

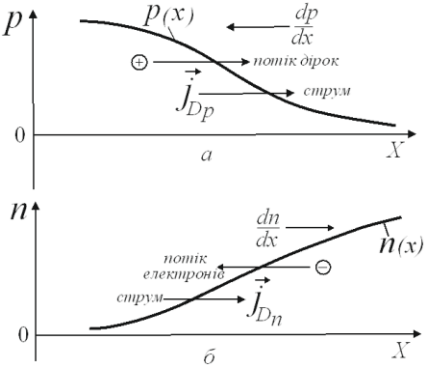


Рис. 19.42. Напрямок дифузійних потоків і струмів для позитивних (дірок) (а) і негативних (електронів) (б) частинок

Під час визначення напрямків дифузійних струмів слід враховувати знак носіїв заряду. Якщо їхній заряд негативний (електрони), то, хоч потік електронів буде направлений у бік зменшення концентрації носіїв, дифузійний струм електронів буде направлений у протилежний бік (рис. 19.42, *б*), у бік зростання концентрації. Для дірок (позитивний заряд) дифузійний струм направлений у той же бік, що й потік дірок (рис. 19.42, *а*) – у бік зменшення концентрації. Дифузійний струм – це справжній струм, який нічим не поступається добре знайомому нам струму, який виникає під дією електричного поля *E*. Під час проходження дифузійного струму виділяється джоулеве тепло і відхиляється магнітна стрілка. Щоб розрахувати дифузійний струм, треба знайти коефіцієнти дифузії електронів D_n і дірок D_p. Між коефіцієнтами дифузії будь-яких частинок та їх рухливістю існує просте співвідношення. Перепишемо (19.98), врахувавши (19.91), у вигляді

$$D = \frac{1}{3} v_T^2 \langle\tau\rangle = \frac{kT}{m} \langle\tau\rangle. \qquad (19.101)$$

Порівнюючи тепер вираз (19.101) з формулою (19.93) для рухливості частинок, отримаємо

$$D = \frac{kT}{e} u. \qquad (19.102)$$

Знаючи величину рухливості за певної температури, за формулою (19.102) легко знайти значення коефіцієнта дифузії. Зв'язок між D та u є наочним відображенням тієї обставини, що і направленому руху частинок під дією сили, і процесу дифузії заважає

один і той самий процес – зіткнення частинок із кристалічною ґраткою, що відображається через середній проміжок часу ‹τ› за середньої теплової швидкості частинок v_T. Співвідношення (19.102) для дірок і електронів може бути записане так:

$$D_p = \frac{kT}{e} u_p, \qquad (19.103)$$

$$D_n = \frac{kT}{e} u_n. \qquad (19.104)$$

Отримані співвідношення (19.103) і (19.104) називають *співвідношеннями Ейнштейна* між коефіцієнтом дифузії та рухливістю.

Особливістю дифузії заряджених частинок є перерозподіл зарядів у результаті дифузійного процесу. Якщо нейтралізації зарядів, які беруть участь в дифузії, не відбувається, виникає об'ємний заряд, що створює електричне поле або потенціальний бар'єр, який стримує процес дифузії. Це явище можна врахувати під час розгляду дрейфового струму, викликаного електричним полем (рис. 19.43). Нехай дифузійний потік електронів направлений справа наліво (рис. 19.43, *б*). У результаті дифузії зліва виникає надлишок електронів, справа нестача і, відповідно, утворюється об'ємний заряд у вигляді подвійного електричного шару, поле якого направлене так, як показано на рис. 19.43, *б*. Створений полем потенціальний бар'єр $U(x)$ перешкоджає дифузійному руху, повертаючи назад ті електрони, енергії яких недостатньо для подолання потенціального бар'єра. Саме ці електрони, які повертаються полем, і створюють дрейфовий струм. У такому випадку повний струм визначається сумою дрейфового і дифузійного струму:

$$j = j_{др} + j_D,$$

де $j_{др}$ знаходять за формулою (19.95).

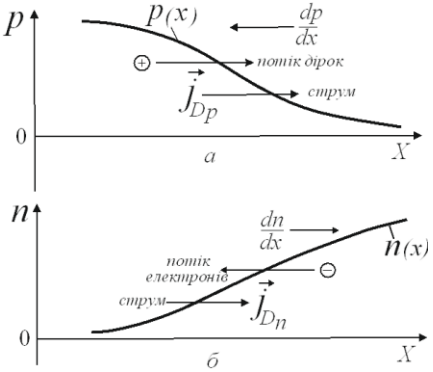


Рис. 19.43. Напрямок дифузійних і внутрішніх дрейфових струмів за дифузії для позитивних (а) і негативних (б) частинок

Вирази для повного струму електронів і дірок мають такий вигляд:

$$j_n = -enu_n E + eD_n \frac{dn}{dx}, \qquad (19.105)$$

$$j_p = -enu_p E + eD_p \frac{dp}{dx}. \qquad (19.106)$$

У виразах (19.105) і (19.106)

струм вважають позитивним, якщо його напрямок збігається з напрямком осі X. Вирази (19.105) і (19.106) справедливі й тоді, коли електричне поле створене не внутрішнім перерозподілом зарядів, а зовнішнім джерелом.

Коли між дифузійними і дрейфовими процесами наступає рівновага, то повний струм j_n, або j_p дорівнюють нулю. Рівність і взаємна компенсація струмів $j_{\partial p}$ та j_D означає існування зв'язку між параметрами, які характеризують процеси дифузії та дрейфу, тобто між коефіцієнтом дифузії D і рухливістю u. Такий підхід дозволяє іншим способом отримати співвідношення (19.103) і (19.104).

Дифузія нерівноважних носіїв

У деяких типах приладів важливу роль відіграє особливий вид дифузії – дифузія надлишкових (нерівноважних) носіїв заряду. Особливість цього виду дифузії в тому, що надлишкові заряди, дифундуючи, віддаляються від того місця, де вони введені у напівпровідник, рекомбінують.

Описати процеси, які при цьому відбуваються, можна за допомогою рівняння неперервності (19.107):

$$\frac{\partial \rho}{\partial t} = -\mathrm{div}\vec{j} = -\frac{\partial j_x}{\partial x} - \frac{\partial j_y}{\partial y} - \frac{\partial j_z}{\partial z}, \quad (19.107)$$

яке виражає закон збереження заряду: зміна густини об'ємного заряду ρ у певній точці можлива тільки внаслідок розходження струму, а поява і зникнення зарядів неможливі.

На перший погляд може видатись дивним, що рівняння (19.107), попри його фундаментальність, не враховує процесів генерації чи рекомбінації носіїв заряду. Утім, ці процеси означають утворення або зникнення рухомих зарядів, тобто їх перехід із зв'язаного стану у вільний або навпаки, так що нейтральність середовища в цілому не порушується і змін об'ємного заряду не відбувається. Явища генерації і рекомбінації мають бути враховані в окремих рівняннях неперервності для носіїв позитивного і негативного зарядів окремо. Якщо для електронів $\rho_n = -en$, а для дірок $\rho_p = ep$, то (19.107) можна переписати так:

$$\frac{\partial n}{\partial t} = \frac{1}{e}\mathrm{div}j_n + G_n - R_n, \quad (19.108)$$

$$\frac{\partial p}{\partial t} = -\frac{1}{e}\mathrm{div}j_p + G_p - R_p, \quad (19.109)$$

де G та R визначають швидкість генерації і рекомбінації відповідно, тобто кількість рухомих носіїв заряду, які з'являються або зникають в одиниці об'єму за одиницю часу.

Розглянемо приклад застосування рівняння неперервності для дослідження дифузії нерівноважних

носіїв заряду. Припустимо, що біля поверхні напівпровідника n-типу якимось чином створюється надлишкова концентрація Δp_0 дірок, які є неосновними носіями заряду для цього напівпровідника (рис. 19.44).

Щоб спростити завдання, розглядаємо одновимірний випадок (без похідних за y та z) і стаціонарний випадок (без похідних за часом). Крім того, будемо вважати, що процесів нерівноважної генерації немає ($G_p = 0$), рекомбінація дірок з рівноважними

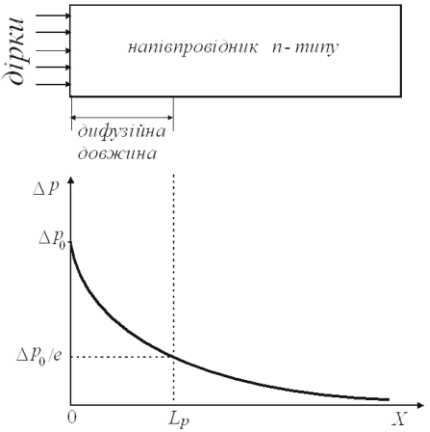


Рис. 19.44. Дифузія нерівноважних дірок у напівпровіднику n-типу

електронами лінійна $R_p = \frac{(p - p_0)}{\tau_p} = -\frac{\Delta p}{\tau_p}$ і напівпровідник має достатньо малий опір, щоб вважати, що всередині напруженість електричного поля $\vec{E} = 0$. Останнє припущення означає, що можна знехтувати дрейфовим рухом дірок порівняно з дифузійним. Ці припущення суттєво спрощують вихідне рівняння (19.109):

$$0 = -\frac{1}{e}\frac{d}{dx}\left(eD_p \frac{d(\Delta p)}{dx}\right) + \frac{\Delta p}{\tau_p}$$

або

$$D_p \frac{d^2(\Delta p)}{dx^2} - \frac{\Delta p}{\tau_p} = 0. \qquad (9.110)$$

Розв'язком цього рівняння є вираз

$$\Delta p(x) = \Delta p_0 \exp\left\{-\frac{x}{L_p}\right\}, \quad (19.111)$$

де

$$L_p = \sqrt{D_p \tau_p}. \qquad (19.112)$$

Результат (19.111) показує, що надлишкові дірки Δp_0 дифундують в об'єм напівпровідника і там рекомбінують з електронами, які є в надлишку в напівпровіднику n-типу (рис. 19.44). Величина L_p, яку називають *дифузійною довжиною*, відповідає відстані, на якій надлишкова концентрація дірок зменшується в $e = 2{,}72\ldots$ рази і характеризує середню глибину, на яку дифундують дірки за час свого життя τ_p. Рух дірок від поверхні створює дифузійний струм, величину якого визначають згідно з (19.99) як

$$j_{D_p} = -eD_p \frac{d(\Delta p)}{dx} \cong \frac{eD_p}{L_p}\Delta p =$$

$$= \frac{eL_p}{\tau_p}\Delta p = j_{D_p}(0)\exp\left\{-\frac{x}{L_p}\right\}, \quad (19.113)$$

де $j_{D_p}(0) = \frac{eL_p}{\tau_p}\Delta p_0$. При цьому враховано, що градієнт $\frac{d(\Delta p)}{dx} \approx \frac{\Delta p}{L_p} < 0$. Із (19.113) видно, що дифузійний струм надлишкових носіїв заряду змінюється з координатою x за тим самим законом, за яким змінюється надлишкова концентрація дірок (19.111).

Величину v_{D_p}, яку знаходять з рівняння

$$v_{D_p} = \frac{L_p}{\tau_p} = \frac{D_p}{L_p}, \quad (19.114)$$

називають *дифузійною швидкістю дірок.* Чисельно вона дорівнює швидкості, з якою нерівноважні носії заряду за час життя проходять шлях, що дорівнює дифузійній довжині.

Вище розглянутий випадок повною мірою належить до напівпровідника p-типу і надлишкових електронів Δn_0.

Аналогічні явища дифузії відбуваються і в тому випадку, якщо нерівноважні носії створюється освітленням поверхні напівпровідника.

19.13. Ефект Ганна

Цікавим ефектом, який спостерігається у напівпровідниках у сильних електричних полях, є *розігрів електронного газу* або утворення «*гарячих*» (нерівноважних) електронів. Це явище привело англійського фізика Джона Ганна на початку шістдесятих років XX ст. до такого відкриття: коли до зразка з напівпровідників арсеніду галію (GaAs) або фосфіду індію (InP) прикласти постійну напругу, яка перевищувала певне значення, струм через напівпровідник починав періодично змінюватись із часом. Інженери відразу зрозуміли, що відкриття Ганна означає появу нового напівпровідникового генератора, який може працювати в області надвисоких частот. Такі генератори були вкрай потрібні у багатьох областях радіоелектроніки, зокрема в радіолокації та системах зв'язку. Десятки спеціалістів розмірковували над відкритим Ганном явищем, однак минув рік, поки його вдалося пояснити.

Ганн досліджував вольт-амперні характеристики зразків із n-напівпровідників. Поки напруга U, яка прикладена до зразка, невелика, струм, як і очікувалось, був пропорційний напрузі – виконувався закон Ома $I = U/R$, де R – опір зразка. Утім, за більших напруг струм починав зростати значно швидше, ніж цього вимагав закон Ома (крива 1 на рис. 19.45). Це

пояснювалось нагріванням зразка джоулевим теплом. Під час нагрівання напівпровідника збільшується концентрація носіїв заряду і зменшується опір зразка (зростає провідність напівпровідника, див. підрозд. 19.10).

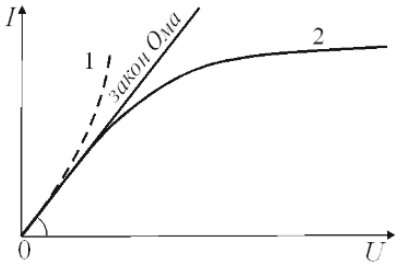


Рис. 19.45. Вольт-амперні характеристмтики

Цікаво оцінити, яка дрейфова швидкість електронів (формула (19.94)) за умови, коли джоулеве тепло починає різко збільшувати струм у зразку з ростом напруги. Діапазон значень рухливості електронів у різних напівпровідниках дуже широкий: від $\sim 10^{-5}$ м^2/(В · с) до $\sim 10^2$ м^2/(В · с). Для Ge $u_n \sim$ 0,4 м^2/(В · с), для Si – 0,13 м^2/(В · с), для GaAS – 1 м^2/(В · с). З підвищенням напруженості електричного поля дрейфова швидкість може виявитись співрозмірною з хаотичною швидкістю, що зменшує час вільного пробігу ‹τ›, а значить і рухливість носіїв (формула (19.93)). Оскільки типове значення рухливості становить 1 м^2/(В · с), а хаотичної швидкості $v_T \sim 10^5$ м/с, то напруженість електричного поля, за якого $v_{др} \sim 0{,}1\, v_T$, не перевищує $\sim 10^4$ В/м.

Важливо підкреслити, що коли внаслідок джоулевого нагріву напівпровідника струм починає різко зростати з ростом напруги, величина рухливості не змінюється. Струм зростає внаслідок того, що розігрів викликає появу «додаткових» електронів. Щоб усунути розігрівання зразка, до зразка прикладали напругу у вигляді коротких імпульсів. Якщо використовувати короткі імпульси (дуже короткочасні) напруги і прикладати їх до зразка так рідко, щоб у проміжку між імпульсами тепло, яке виділяється у зразку, встигало повністю розсіятись, то можна досягти дуже сильних електричних полів у напівпровіднику, не нагріваючи його. Таким чином, вплив розігріву усувався, і вольт-амперна характеристика знову ставала лінійною ($I \sim U$) й залишалась омічною до значно більших значень напруг, ніж за постійного струму. Втім, зі зростанням напруги на зразку знову спостерігається відхилення від закону Ома. Для Ge та Si зі зростанням напруги струм зростає повільніше, ніж за лінійним законом. За дуже великих

напруг струм фактично перестає залежати від напруги – досягає насичення (крива 2 на рис. 19.45). Проводячи експерименти із зразками різної довжини l було встановлено, що відхилення від закону Ома і насичення завжди настає за одних і тих самих значень напруги електричного поля $E = U/l$. Перші спроби пояснити цей факт зменшенням концентрації n електронів у зразку зі збільшенням поля E виявились невдалими. Експерименти з чистими Ge та Si показали, що за тих полів, за яких спостерігається насичення струму, концентрація n від поля не залежить. Оскільки густина дрейфового струму (формула (19.95)) за незмінної n залежить від швидкості дрейфу $v_{др}$, то це означає, що в Ge та Si дрейфова швидкість електронів у сильніших електричних полях від напруженості поля не залежить всупереч формулі (19.94). Пояснення цьому факту приведено у підрозд. 19.12. Можемо тільки додати, що за великих електричних полів ефективність процесів зіткнення електронів провідності із кристалічною ґраткою підвищується, і рухливість електронів у Ge та Si зі зростанням поля зменшується. Насичення дрейфової швидкості електронів у Ge та Si в сильних електричних полях було відкрито задовго до того, як Ганн розпочав свої дослідження. Та обставина, що властивості електронів у сильних полях можуть суттєво відрізнятися від властивостей у слабких електричних полях, теж було добре відомо. Утім, ефект, який Ганн спостерігав на зразках арсеніду галію (n–GaAs) і фосфіду індію (n –InP), був просто вражаючий.

До зразка прикладали імпульси напруги (рис. 19.46, *а*). Поки амплітуди імпульсів були менші за деяку порогову величину U_n (цей випадок показано на рис. 19.46, *а* пунктиром), усе було зрозуміло (приблизно виконувався закон Ома). Струм через зразок зростав, поки наростала напруга, залишався сталим, коли напруга не змінювалась, і зменшувався, коли напруга спадала (пунктирна крива на рис. 19.46, *б*). Коли напруга перевищувала значення U_n (суцільна крива на рис. 19.46, *а*), струм за *постійної напруги* починав періодично змінюватись з часом (цей випадок показано на рис. 19.46, *б* суцільною кривою). Ганн встановив, що порогова напруга пропорційна довжині зразка l: $U_n = E_n l$, де E_n – приблизно однакове значення напруженості поля для всіх зразків. Для n –GaAS величина E_n становила приблизно $3{,}2 \cdot 10^5$ В/м.

Період коливань струму був теж пропорційний довжині зразка і дорівнював $T \cong 2l/v_n$, де $v_n \cong uE_n$.

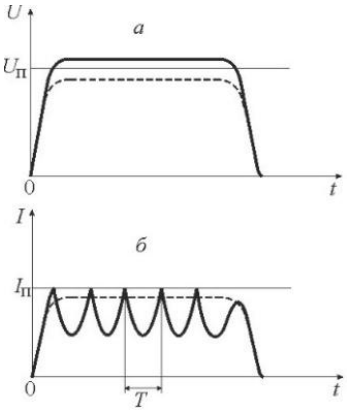


Рис. 19.46. Ефект Ганна

Намагаючись зрозуміти, що відбувається у зразку, коли виникають коливання струму, Ганн виконав дослідження і встановив, що коли напруга на зразку перевищує порогове значення U_n, характер протікання струму і розподіл напруженості електричного поля у зразку різко змінюється. За $U < U_n$ струм у зразку утворює потік електронів від катода до анода, а електричне поле розподілене вздовж зразка однорідно (пунктирна лінія на рис. 19.47). У цих умовах вольт-амперна характеристика зразка відповідає закону Ома, але щойно напруга на зразку досягає порогового значення U_n, характер розподілу поля вздовж зразка різко змінюється. У зразку біля катода виникає вузька область дуже сильного поля – «**домен**» (визначення поняття «домен» наведено у підрозд. 2.3). На рис. 19.47 показано домен сильного електричного поля. Хоча ширина домена невелика, але поле в ньому настільки сильне (E_4), що на ньому падає значна частина прикладеної до зразка напруги (рис. 19.47). Домен рухається від катода до анода із швидкістю, яка приблизно дорівнює $v_n/2$. Досягнувши анодного контакта, домен руйнується. Негайно після цього біля катода зароджується новий домен і все повторюється спочатку.

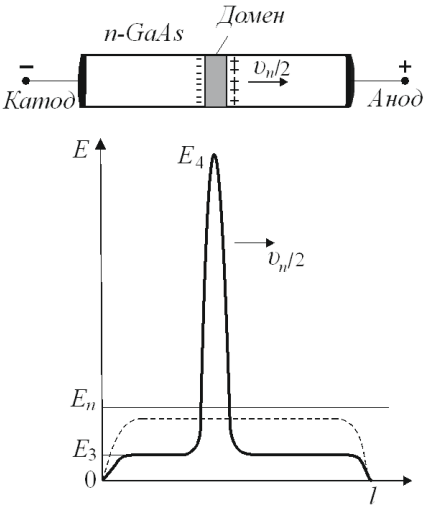


Рис. 19.47. Домен сильного електричного поля

Легко зрозуміти, як пов'язаний періодичний рух домена по зразку з коливаннями струму в ньому. Нехай до

зразка приклали напругу, яка дорівнює U_n. У зразку виник домен, і значна частина напруги впала на домені, але це означає, що на частку решти частини зразка, де домена немає, буде припадати менша напруга; значить, поле зовні домена значно зменшиться (E_3) (порівняти пунктирну і суцільну криві на рис. 19.47). Зменшиться і значення швидкості електронів, яке пропорційне напруженості поля, а отже зменшиться і струм, що протікає через зразок.

Поки домен рухається вздовж зразка, струм не змінюється. Коли домен, досягнувши анода, починає руйнуватись, спад напруги на ньому зменшується. При цьому поле в решті частині зразка зростає, й отже, струм зростає. Щойно напруженість поля у зразку в районі катода стане дорівнювати пороговому значенню $E_n = U_n/l$, у зразку почне формуватись новий домен, і струм знову почне зменшуватись. Поки напруга, яка прикладена до зразка, перевищує порогове значення U_n, цей процес періодично повторюється (рис. 19.46, *б*).

Виникає запитання, чому у зразку утворюється домен, в чому фізична причина утворення домена. На це питання у 1984 р. зумів відповісти американський фізик Герберт Кремер. Він нагадав, що незадовго до відкриття Ганна на основі теоретичних розрахунків було показано, що в арсеніді галію, фосфіді індію і деяких інших напівпровідниках залежність дрейфової швидкості електронів від напруженості поля повинна мати такий вигляд, як показано на рис. 19.48. У певному діапазоні значень напруженості поля E, які починаються із значення E_n, дрейфова швидкість електронів зменшується зі збільшенням напруженості поля. Виходячи із формули (19.94) ми повинні зробити висновок, що в цьому діапазоні значень E рухливість u електронів зменшується зі збільшенням E так різко, що дрейфова швидкість зменшується, хоч напруженість електричного поля зростає.

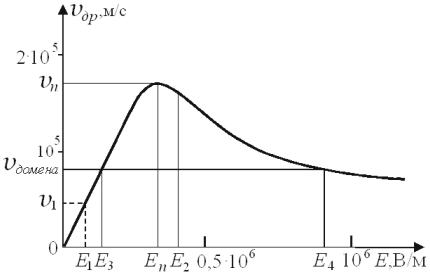


Рис. 19.48. Залежність дрейфової швидкості електронів від напруженості поля

Як же пов'язані ефекти, що спостерігав Ганн, з таким видом залежності швидкості електронів від напруженості поля? Розглянемо простий приклад. Нехай до зразка прикладено напругу $U_1 < U_n$. При цьому поле вздовж зразка розподілене однорідно:

$E_1 = \frac{U_1}{l}$. Дрейфова швидкість електронів $v_1 = uE_1$ (рис. 19.48), й тоді потік електронів від катода до анода створює струм. Припустимо тепер, що десь у зразку випадково утворилась невелика ділянка, в якій напруженість поля буде трохи більшою, ніж значення E_1 (цьому сприяють неоднорідності в легованому напівпровіднику). За рахунок хаотичного характеру руху електронів у кристалі маленькі відхилення поля від середнього значення E_1 (флуктуації поля) виникають неперервно й тепер прослідкуємо за тим, що відбудеться з такою флуктуацією, коли вона виникає. Якщо на якій-небудь ділянці зразка напруженість поля стала дещо більшою, ніж у решті зразка, то це означає, що на межах цієї ділянки накопичились електричні заряди: негативні з боку катода й позитивні з боку анода (рис. 19.47). Інакше кажучи, зліва від виділеної ділянки утворився невеликий «надлишок» електронів, а справа – невеликий «дефіцит» електронів. Поле цих «об'ємних» надлишкових зарядів складається із полем зовнішнього джерела і таким чином підсилюється поле всередині виділеної ділянки.

Що ж відбувається з такого флуктуацією далі?

Оскільки ми розглядаємо випадок, коли $E_1 < E_n$, то швидкість електронів пропорційна напруженості поля. Отже, електрони всередині флуктуації рухаються швидше, ніж електрони в решті зразка. Це означає, що флуктуація буде «тікати» від надлишкових електронів зліва і «наздоганяти» ту ділянку справа, де електронів не вистачає. У результаті флуктуація поля розсмоктується.

А тепер розглянемо, що відбудеться із флуктуацією, коли напруженість E_2 поля у зразку дещо більша, ніж порогове значення E_n (рис. 19.48). Знову будемо вважати, що спочатку поле розподілене вздовж зразка однорідно; унаслідок флуктуації зарядів виникають об'ємні заряди; у результаті в малій ділянці зразка напруженість поля виявиться більшою, ніж у решті зразка, але тепер більшому полю відповідає менша дрейфова швидкість (рис. 19.48). Це означає, що флуктуація рухається повільніше, ніж електрони зліва і справа від неї. У результаті зліва до надлишкових електронів, які там уже є, добавляться нові електрони, що доганяють флуктуацію. Справа, де був «дефіцит» електронів, їх стане ще менше – електрони «втечуть» від флуктуації. Таким чином, об'ємний заряд стане більшим і напруженість поля всередині флуктуації збільшиться. Утім, збільшення напруженості поля за $E > E_n$ відповідає зменшенню дрейфової швидкості. У результаті флуктуація стане рухатись ще повільніше, ще

більше електронів скупчиться зліва від флуктуації і ще менше їх стане справа, а напруженість поля всередині флуктуації стане більшою і т. д. Отже за умови, коли дрейфова швидкість електронів зменшується з ростом напруженості поля, однорідний розподіл поля по зразку нестійкий щодо малої флуктуації поля; раз виникнувши, така флуктуація починає нестримно наростати. Ріст флуктуації припиниться, коли швидкість електронів всередині флуктуації зрівняється зі швидкістю електронів зовні неї. При цьому напруженість поля зовні домена буде E_3, а в домені E_4 (рис. 19.47 і 19.48).

Поки надлишкове поле всередині флуктуації мале, на ній падає мала напруга, і її наявність у зразку фактично не впливає на швидкість електронів поза флуктуацією. Утім, зі зростанням поля всередені флуктуації на ній починає падати все більша й більша напруга. При цьому напруга на решті зразка стає меншою (адже повна напруга, яка прикладена до зразка, постійна). Поле поза флуктуацією зменшується. Така флуктуація, яка сильно виросла, і є доменом поля. Поза доменом швидкість електронів зменшується зі зменшенням напруженості поля, оскільки поза доменом напруженість завжди менша E_n. Таким чином, зі зростанням домена швидкість електронів поза доменом зменшується. Зрештою швидкості електронів всередині та поза доменом зрівняються і ріст домена припиниться. Зразком від катода до анода буде «бігти» незмінна за формою (стабільна) ділянка сильного поля – домен зі швидкістю $v_{\text{др}} \approx v_{\text{п}}/2$ (рис. 19.48).

Ці домени і спостерігав Ганн у своїх експериментах. Максимальне поле в домені E_4 може досягати значень $2\cdot 10^6 \ldots 30\cdot 10^6$ В/м.

Таким чином, з'ясування природи існування ефекту Ганна звелось до пояснення залежності швидкості дрейфу електронів $v_{\text{др}}$ від напруженості електричного поля у напівпровіднику, яку показано на рис. 19.48. Згідно з формулами (19.93) і (19.94) можемо записати:

$$v_{\text{др}} = u_n E = \frac{e\langle\tau_{\text{п}}\rangle}{m_n^*}E, \quad (19.115)$$

де m_n^* – ефективна маса електронів; $\langle\tau_n\rangle$ – середній час їх вільного пробігу.

Для арсеніду галію зона провідності складається із двох зон, енергетична відстань між якими становить 0,36 еВ. Оскільки ширина забороненої зони GaAs за $T = 300$ К дорівнює 1,4 еВ, то енергія дна першої нижньої зони провідності W_{c_1} =1,4 еВ, а другої (верхньої) – $W_{c_2} = 1{,}76$ еВ. Ефективна маса електрона у першій зоні $m_{n1}^* = 0{,}07\, m_e$, а в другій зоні $m_{n2}^* = 1{,}2\, m_e$, де m_e – маса вільного

електрона. Це означає, що рухливість електронів в нижній зоні u_{n1} більша від рухливості електронів у верхній зоні u_{n2}, оскільки рухливість $u_n \sim 1/m_n^*$. Крім того, густина енергетичних станів $g(W) = 4\pi(2m_n^*)^{\frac{3}{2}}W^{\frac{1}{2}}/h^3$ значно більша у верхній зоні, оскільки $m_{n2}^* >> m_{n1}^*$.

За звичайних температур, коли енергія теплового руху $kT \ll 0{,}36$ еВ, у слабких електричних полях електрони заповнюють фактично тільки стани нижньої зони, і маючи велику рухливість, забезпечують значну електропровідність і лінійну вольт-амперну характеристику. У сильних електричних полях ($E > 3{,}5 \cdot 10^5$ В/м) спостерігається «розігрів» електронів провідності. Це означає підвищення енергії дрейфу і перехід електронів у верхню зону провідності. Більша густина станів у верхій зоні забезпечує можливість переходу туди значної частини електронів зони провідності, а мала їхня рухливість u_{n2} призводить до зменшення швидкості дрейфу $v_{\text{др}}$. Так пояснюється залежність $v_{\text{др}}$ від E, яку показано на рис. 19.48.

Ефект Ганна спостерігається більш ніж у двадцяти напівпровідникових сполуках: антимоніді індію InSb, антимоніді миш'яку AsSb, у германії під стискуванням, у кремнії за низьких температур, у потрійних напівпровідникових сполуках GaAlAs, InGaSb, InAsP, у сполуці InGaAsP і в багатьох інших напівпровідниках. Основне застосування ганнівських діодів (так зазвичай називають зразки, в яких спостерігається ефект Ганна) – генератори надвисокої частоти.

Контрольні запитання

1. Дайте визначення твердого тіла.

2. Які тверді тіла є кристалами?

3. Наведіть класифікацію твердих тіл за величиною питомої електропровідності та її залежності від температури.

4. Які існують типи зв'язків атомів у твердих тілах?

5. Чому енергетичний спектр електронів у кристалах має бути зонним?

6. Дайте визначення зон – валентної, провідності, забороненої.

7. Поясніть електричні властивості металів, напівпровідників і діелектриків з точки зору зонної теорії.

8. Дайте визначення функції густини енергетичних рівнів у зоні. Як ця густина залежить від енергії в зоні?

9. У чому відмінність квантового і класичного розподілу електронів за енергетичними рівнями у зоні?

10. Дайте визначення енергії Фермі. Від чого залежить величина енергії Фермі?

11. Дайте визначення функції розподілу густини електронів за енергетичними станами (рівнями).

12. Який фізичний зміст ширини забороненої зони?

13. Як ви думаєте, у чому зміст введення поняття *ефективної* густини станів для електронів і дірок у напівпровідниках?

14. Поясніть, від яких факторів залежить відношення концентрацій електронів і дірок у напівпровідниках.

15. Поясніть фізичний зміст закону діючих мас.

16. Поясніть, чому концентрація електронів і дірок у власних напівпровідниках сильно залежить від температури.

17. Як ви думаєте, чи можливий такий випадок у напівпровідниках, щоб ефективна маса електронів дорівнювала ефективній масі дірок? Де б у цьому випадку розміщувався рівень Фермі?

18. Чому для електричних властивостей домішкових напівпровідників важливим є те, чи це область високих, чи низьких температур?

19. Як ви розумієте вислів «Сильно легований напівпровідник»?

20. Від яких факторів залежить величина часу життя у напівпровідниках?

21. Поясніть роль основних і неосновних носіїв заряду в напівпровідниках з точки зору електричних і фотоелектричних властивостей.

22. Які є методи визначення основних фізичних параметрів напівпровідників?

23. Які види струму існують у напівпровідниках? У чому відмінність між ними?

24. У чому суть ефекту Ганна?

25. Як пояснюється ефект Ганна за допомогою домена електричного поля?

26. Яке практичне застосування ефекта Ганна?

Розділ 20. Термоелектронна емісія

20.1. Історія відкриття термоелектронної емісії

Термоелектронну емісію вперше спостерігав у 1883 р. Томас Алва Едісон (1847–1931). Термоелектронну емісію було відкрито випадково. У кінці 1879 р. Едісон в основному закінчив роботу зі створення електричної лампи (рис. 20.1) розжарювання і наполегливо працював над проблемою збільшення строку її роботи.

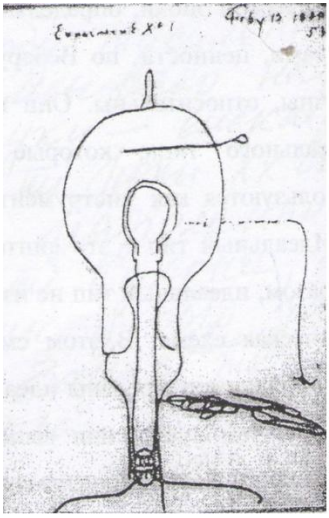


Рис. 20.1. Фотокопія сторінки з лабораторного журналу Едісона

Лампа ця по суті мало чим відрізнялась від сучасних ламп розжарювання – тільки замість вольфрамової спіралі використовували нитку з обвугленого бамбукового волокна. Перші такі лампи (рис. 20.1) могли неперервно горіти близько 40 год, через що для масового застосування вони були непридатні. У пошуках причини швидкого перегорання вуглецевої нитки Едісон звернув увагу на те, що внутрішня поверхня лампового балона покривається темним нальотом з вуглецевих частинок, які, як він думав, вилітали з нитки розжарювання. Сподіваючись якось усунути це небажане явище, Едісон розмістив у лампі біля нитки додатковий електрод, електричний потенціал якого можна було змінювати. При цьому він зі здивуванням виявив, що, коли потенціал цього електрода вищий від потенціалу нитки (позитивний щодо потенціалу нитки), між ним і ниткою – через вакуумний проміжок! – протікав електричний струм. Попри всю технічну винахідливість жодної практичної користі зі свого відкриття Едісон отримати не зміг, і воно було надовго забуто. Ефект Едісона вдалось пояснити тільки після відкриття електрона у 1897 р. Пояснення дав Дж. А. Флемінг, який у 1882–1885 рр. працював у Едісона. Флемінг зрозумів, що електричний струм у балоні лампи створюють електрони, які виходять з розжареної нитки і притягуються позитивно зарядженим додатковим електродом.

Продовжуючи роздумувати над явищем термоелектронної емісії, у 1904 р. Флемінг запатентував першу в історії техніки радіолампу. Це був вакуумний діод, який повністю ґрунтувався на односторонній провідності, пов'язаній з ефектом Едісона, що дозволяла випрямляти змінний струм. Флемінг задумав свою лампу як детектор для радіохвиль.

У 1906 р. американський інженер Л. де Форест (1873–1961) винайшов трьохелектродну лампу – тріод, історично перший електронний підсилювальний прилад. Відкриття ефективного способу підсилення слабких електричних сигналів ознаменувало народження нової області техніки й науки – електроніки. До появи транзисторів (50-ті роки XX ст.) в ній повністю панували радіолампи, робота яких основана на ефекті Едісона, якому знаменитий винахідник свого часу не надав особливого значення.

Нині явище термоелектронної емісії використовують у сотнях мільйонів джерел світла, насамперед в люмінесцентних лампах.

20.2. Робота виходу електрона з металу

Суть цього явища полягає в тому, що з підвищенням температури збільшується енергія теплових коливань атомів металу і, відповідно, теплова енергія електронів. Попри те, що середня теплова енергія електронів навіть за температур у тисячу градусів є незначною (становить десяті частки електрон-вольта) порівняно із загальною середньою енергією електронів в металі (див. підрозд. 19.7), все ж мала частина електронів набуває енергії, яка є достатньою для подолання сил, що перешкоджають виходу електронів з металу.

Хоча таких електронів мало, втім саме вони обумовлюють процеси емісії. За фізичним змістом термоелектронна емісія багато в чому аналогічна процесу випаровування молекул з рідини. Вище сказане проілюструємо енергетичною діаграмою електронів провідності в металі на рис. 20.2, де показано, що за $T = 0$ функція розподілу густини електронів $N(W) = 0$ для $W > W_F$. Це означає, що електрони заповнюють енергетичні рівні, які лежать нижче рівня Фермі (квантовий розподіл). Електрони з енергіями $W > W_0$ (електрони поза металом) з'являються під час підвищення температури, і їх кількість, яка пропорційна заштрихованій області під кривою $N(W)$ для $W > W_0$ (рис. 20.2, *б*), зростає з температурою. Це обумовлено тим, що зі зростанням температури відбувається «розмиття» функції

Фермі–Дірака біля енергії Фермі при $W > W_F$. Оскільки $W > W_F \gg kT$, то функція Фермі–Дірака з великою точністю може бути замінена класичною функцією Больцмана (19.20):

$$f(W) \cong \exp\left\{-\frac{W-W_F}{kT}\right\}. \qquad (20.1)$$

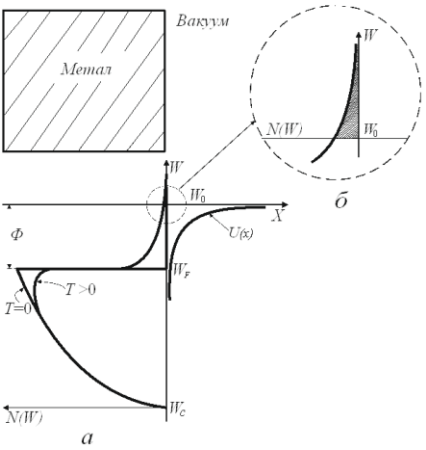


Рис. 20.2. Енергетична діаграма електронів біля поверхні металу: а) $N(W) = g(W)f(W)$ – функція розподілу густини електронів, де $g(W)$ – функція розподілу густини енергетичних рівнів, $f(W)$ – функція Фермі–Дірака (функція розподілу ймовірності заповнення енергетичних рівнів електронами), W_c – енергія дна зони провідності електронів у металах (беруть як початок відліку енергії), W_0 – енергія вакууму (енергія вільного електрона), W_F – енергія рівня Фермі, $\Phi = W_0 - W_F$ – робота виходу електронів з металу

Таким чином, електрони з енергією $W > W_0$ (або «гарячі» електрони) підпорядковуються класичному розподілу, й вони відповідальні за термоелектронну емісію. Виникає запитання, а чому електрони, які вийшли з металу (тобто мають енергію $W > W_0$), не залишають метал зовсім, а утримуються біля поверхні металу.

Справа в тому, що під час виходу електрона за межі металу на нього діє сила зі сторони нескомпенсованого позитивного заряду (рис. 20.3), яка утримує його. Розрахунок сили взаємодії може бути проведений методом дзеркальних відображень, суть якого в тому, що розподілений по поверхні металу позитивний заряд замінюється точковим позитивним зарядом $+e$, який міститься по інший бік поверхні металу на такій самій відстані від поверхні, що і розглядуваний електрон (рис. 20.3). Згідно з законом Кулона ця сила становитиме

$$F(x) = -\frac{e^2}{16\pi\varepsilon_0 x^2}, \qquad (20.2)$$

а потенціальна енергія електрона в полі такої сили буде

$$U(x) = -\frac{e^2}{16\pi\varepsilon_0 x}. \qquad (20.3)$$

Графік функції (20.3) схематично показано на рис. 20.4. Ця функція описує форму потенціального бар'єра, який перешкоджає виходу електрона з металу. Робота, яку виконує електрон для подолання сил цього потенціального бар'єра під час емісії називають *роботою виходу*. Для металів роботу виходу

$$\Phi = W_0 - W_F \qquad (20.4)$$

називають термодинамічною (рис. 20.2, *а*).

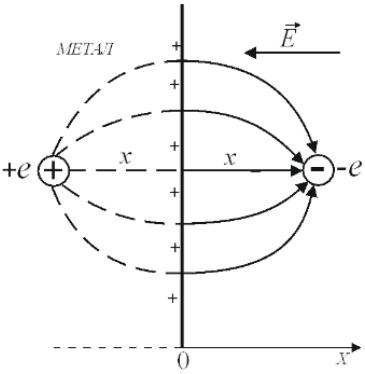


Рис. 20.3. Сили дзеркального відображення

Сили дзеркального відображення (20.2) можуть бути послаблені зовнішнім електричним полем, напруженість якого $\vec{E}$ (рис. 20.3). Сила дзеркального відображення (20.2) і сила електричного поля $e\vec{E}$ направлені назустріч одна одній, а в точці

$$x_0 = \sqrt{e/(16\pi\varepsilon_0 E)} \quad (20.5)$$

(рис. 20.4) взаємно компенсують одна одну. Потенціальна енергія електрона буде дорівнювати сумі енергії (20.3), а його енергії в зовнішньому полі $-\ eEx$ (рис. 20.3), тобто

$$U'(x) = -\frac{e^2}{16\pi\varepsilon_0 x} - eEx. \quad (20.6)$$

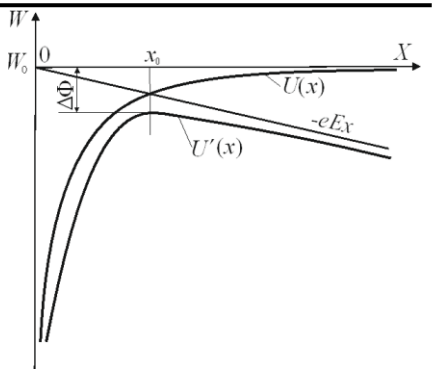


Рис. 20.4. Графіки цих потенціальних енергій електрона

Графіки цих потенціальних енергій схематично показано на рис. 20.4. Точка x_0 відповідає максимуму потенціальної енергії $U'(x)$ і визначає найбільшу висоту потенціального бар'єра. Оскільки $U'(x)$ відраховують від енергії вакууму W_0, то це означає, що робота виходу електрона з металу зменшується на величину

$$\Delta\Phi = U'(x_0) = -e\sqrt{\frac{eE}{4\pi\varepsilon_0}}. \quad (20.7)$$

Таким чином, кількість електронів, які можуть брати участь в термоелектронній емісії, визначають площею заштрихованої фігури на рис. 20.2, *б*, величину якої визначають температурою і напруженістю зовнішнього електричного поля $\vec{E}$ (рис. 20.3), яка зменшує роботу виходу Φ.

20.3. Закони термоелектронної емісії

Термоелектронну емісію широко використовують в електровакуумних приладах, тому для її вивчення може бути використано двох - або трьохелектродну лампу (рис. 20.5). Вивчення законів термоелектронної емісії за допомогою таких ламп зводиться до експериментального дослідження залежності анодного струму I_A від анодної напруги U_A, тобто до дослідження вольтамперних характеристик (ВАХ) (рис. 20.6).

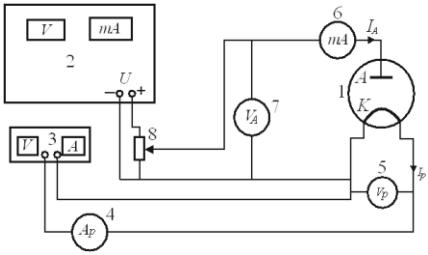


Рис. 20.5. Принципова схема установки для дослідження ВАХ діода: 1 – електронна лампа (діод); 2 – універсальне джерело живлення широкого спектру призначення (воно живить анодне коло лампи); 3 – джерело живлення, за допомогою якого регулюють величину струму розжарення, а відповідно, і температуру катода K; 4 – амперметр, який вимірює струм розжарення катода; 5 – вольтметр, який вимірює напругу на катоді; 6 – міліамперметр, який вимірює струм в анодному колі; 7 – вольтметр, який вимірює напругу між анодом і катодом (анодна напруга); 8 – потенціометр, який дає можливість змінювати анодну напругу в широких межах

На графіках можна побачити три характерні ділянки: початкову ділянку (I), ділянку зростання (II), ділянку насичення струму (III). Звернімо увагу на те, що ВАХ електронної лампи є *нелінійною*, а отже, електричний струм у лампі не підпорядковується закону Ома.

Спочатку (ділянка I ВАХ) зі збільшенням анодної напруги сила струму зростає повільно, потім значно швидше (ділянка II) (рис. 20.6).

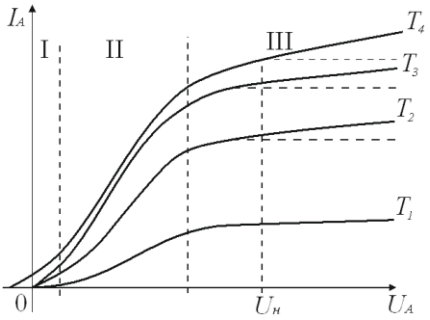


Рис. 20.6. Вольт-амперні характеристики для різних температур катода ($T_4 > T_3 > T_2 > T_1$)

Починаючи з деякого значення напруги $U_н$ подальше зростання струму фактично припиняється (ділянка III) (рис. 20.6). Залежність струму діода від напруги легко пояснити. Якщо наявна термоелектронна емісія, у просторі між катодом та анодом у будь-який момент часу є електрони, що рухаються від катода до анода, які утворюють хмару негативного заряду (*просторовий заряд*).

Цей просторовий заряд змінює розподіл потенціалу в діоді. Якщо катод і анод є плоскими пластинками, паралельними одна одній (рис. 20.7), то коли немає просторового заряду (катод холодний) розподіл потенціалу між катодом і анодом, які утворюють плоский конденсатор, зображують прямою лінією *1*. Якщо наявний термоелектронний струм (розжарений катод), між катодом й анодом виникає просторовий заряд і розподіл потенціалу змінюється – він подається тепер кривою *2*. При цьому значення потенціалу в будь-якій площині x менше, ніж коли немає просторового заряду, а отже, і швидкість руху електронів за наявності просторового заряду зменшується. Зі збільшенням анодної напруги концентрація електронів у хмарі просторового заряду зменшується, тому гальмівна дія просторового заряду зменшується й анодний струм зростає.

Зауважимо, що розподіл потенціалу, зображений кривою *2*, дістаємо тоді, коли початкова швидкість вилітання електронів з катода досить мала, що, звичайно, і спостерігається. А коли початковою швидкістю знехтувати не можна, розподіл потенціалу має складніший вигляд.

Коли потенціал анода стає $U_A > U_{\text{н}}$ (рис. 20.6), то всі електрони, які випускає катод за кожну одиницю часу, потрапляють на анод, і струм I_A має досягати максимального значення (струм насичення) за певної температури катода. Струм насичення характеризує емісійну здатність катода, що залежить від природи катода та його температури, проте в області III насичення анодного струму (рис. 20.6) немає. Анодний струм I_A залежить від потенціалу анода U_A, і ця залежність тим сильніша, чим вища температура катода. Пояснення цієї залежності наведено далі.

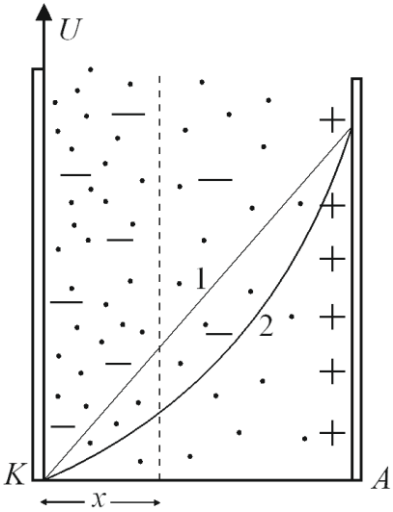


Рис. 20.7. Просторовий заряд у діоді та спричинений ним перерозподіл потенціалу: 1– розподіл потенціалу без просторового заряду; 2 – розподіл потенціалу за наявності просторового заряду

Якщо напруга менша за $U_{\text{н}}$ яанодної напруги U_A описують законом Богуславського–Ленгмюра:

$$I_A = \alpha U_A^{3/2}, \qquad (20.8)$$

де α – стала величина, яка залежить від конструкції електродів лампи.

Закон (20.8) ще часто називають «законом трьох других». Його отримано теоретично для кількох спрощень, а саме:

1) катод і анод − площини, що утворюють плоский конденсатор;

2) на поверхні катода ($x = 0$) потенціал дорівнює нулю ($U = 0$), а на аноді $U = U_A$;

3) напруженість електричного поля ($E = -\frac{dU}{dx}$) на поверхні катода та безпосередньо біля нього дорівнює нулю;

4) початкова швидкість електронів, що вилетіли з поверхні катода, дорівнює нулю.

Зауважимо, що залежності $I_A = f(U_A)$, які спостерігаються на практиці (рис. 20.6), дещо відрізняються від теоретичних. Відомі такі відмінності:

1. Реальні ВАХ (рис. 20.6) не виходять з початку координат. Вони починаються при невеликих від'ємних напругах на аноді, тому за $U_A = 0$ існує деякий невеликий початковий струм. Це можна пояснити тим, що електрони, які вилітають з поверхні катода, мають початкову швидкість, відмінну від нуля. Деякі з них, маючи значну швидкість, можуть навіть за нульової анодної напруги ($U_A = 0$) подолати гальмівне поле просторового заряду й долетіти до анода, створюючи в анодному колі незначний струм.

2. Реальні ВАХ на ділянці II, як правило, є більш пологими, ніж теоретичні. Крім того, для різних температур катода ділянки II реальних кривих розходяться, тобто не збігаються. За законом Богуславського−Ленгмюра на ділянці II має бути єдина крива за будь-яких температур Таке розходження реальних кривих ВАХ пояснюється тим, що зі зміною температури змінюється і характер її ж розподілу вздовж катода. За незначного розжарювання катод має довші охолоджені кінці, і його робоча довжина, а відповідно, і діюча поверхня значно менші за геометричні розміри цього електрода; тому у формулі (20.8) коефіцієнт α і величина струму значно менші від теоретичного значення за заданої U_A. З підвищенням температури розжарювання збільшується робоча довжина катода, що призводить до зростання коефіцієнта α, і згідно з (20.8) струм I_A стає більшим за того самого значення U_A.

3. На крутизну ділянки зростання ВАХ впливає спад потенціалу вздовж катода. Струм розжарювання, що проходить через катод, створює падіння потенціалу по його довжині. Унаслідок цього потенціали різних точок катода неоднакові. Потенціал катода поступово

збільшується від його кінця з меншим потенціалом до його кінця з більшим потенціалом. Завдяки цьому потенціал анода відносно різних ділянок катода різний, а отже, не однакова й густина струму на різних ділянках катода.

Ділянку III ВАХ описують законом Річардсона–Дешмана:

$$j_s = AT^2 \exp\left\{-\frac{\Phi}{kT}\right\}, \qquad (20.9)$$

де j_s – густина струму насичення; $A = \frac{4\pi m_e e k^2}{h^3}$ – стала, що не залежить від виду металу; Φ – робота виходу електронів з металу; k – стала Больцмана; m_e – маса електрона; h – стала Планка; e – заряд електрона.

Формула (20.9) є основним законом термоелектронної емісії. Вона характеризує залежність струму емісії від температури і термодинамічної роботи виходу. Ця залежність експоненціальна, тобто струм насичення дуже різко залежить від температури.

Точне виведення формули (20.9) виходить за межі курсу, але можемо сказати, що струм насичення є пропорційним швидкості й кількості електронів, що виходять за межі металу, тобто площі заштрихованої фігури на рис. 20.2, *б*.

Із формули (20.9) випливає, що *струм насичення* не має залежати від напруги між катодом (нагрітий метал) й анодом.

Утім експериментально (в області III на рис. 20.6) спостерігається залежність струму насичення від анодної напруги всупереч передбаченням формули (20.9). Цей ефект пояснив Шотткі. Зростання струму насичення зі збільшенням анодної напруги пов'язане зі зменшенням роботи виходу електронів з катода у разі збільшення електричного поля між катодом і анодом (формула 20.7).

Отже, замість формули (20.9) можемо записати

$$j = AT^2 \exp\left\{-\frac{\Phi - |\Delta\Phi|}{kT}\right\} = j_s \exp\left\{\frac{|\Delta\Phi|}{kT}\right\}$$

або

$$j = j_s \exp\left\{\frac{e}{kT}\sqrt{\frac{eE}{4\pi\varepsilon_0}}\right\}. \qquad (20.10)$$

Формула (20.10) пояснює збільшення струму за $U > U_н$ унаслідок ефекту Шотткі, що проявляється у відсутності насичення струму на ВАХ, особливо за значних температур катода. Таким чином, можна теоретично пояснити вигляд основних ділянок ВАХ електронної лампи (діода).

Явище термоелектронної емісії використовують для створення катодів у різних електронних приладах. Для виготовлення катодів найбільше поширення отримав вольфрам, рідше використовують молібден, тантал, цирконій і деякі інші метали. Завдяки великій тугоплавкості робоча

температура вольфрамового катода може бути доведена до 2500 K, за якої попри велику роботу виходу (4,5 еВ) густина струму досягає $(0{,}3 \ldots 0{,}5) \cdot 10^4$ А/м^2. Металічні катоди мають малу ефективність, оскільки вимагають великих затрат на нагрівання до високої температури.

Водночас для практики дуже важливо по можливості зменшити робочу температуру катода електронної лампи, бо при цьому зменшується потужність, що витрачається на розжарювання катода, і збільшити строк служби лампи, тому тепер поряд з катодами з чистих тугоплавких металів (вольфрам, молібден) широко застосовують катоди складнішої будови.

Велике застосування у техніці має *оксидний катод*. Він має металеву підкладку (керн) (рис. 20.8), на яку нанесено шар окисів лужноземельних металів (BaO, SrO, CaO або їх суміш). Для розжарювання катода через керн пропускають струм («катоди прямого нагрівання») або нагрівають катод, користуючись допоміжною металевою спіраллю («підігрівні катоди»). Щоб надати катоду високої емісійної здатності, його додатково обробляють (активують); при цьому через електронну лампу за температури катода близько 1300 K протягом деякого часу пропускають термоелектронний струм. Згідно з сучасними поглядами під час активування катода на його поверхні виникає одноатомний шар позитивних іонів лужноземельного металу (рис. 20.8), який значно зменшує роботу виходу і цим збільшує емісійну здатність катода.

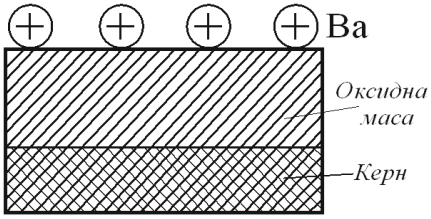


Рис. 20.8. Структура оксидного катода (схематично)

Під час виготовлення оксидних катодів на керн спочатку наносять вуглекислі сполуки лужноземельних металів ($BaCO_3$, $SrCO_3$, $CaCO_3$) і потім прожарюють катод перед активуванням у вакуумі. При цьому вуглекислі сполуки розкладаються за реакцією

$$BaCO_3 \rightleftarrows BaO + CO_2$$

і керн покривається окисами.

Сучасні оксидні катоди мають високі характеристики. Їх робоча температура дорівнює 1070 … 1170 K, а іноді й менше. Їх нормальна емісійна здатність досягає 10^4 А/м^2. Для порівняння зазначимо, що робоча температура вольфрамових катодів лежить близько значення 2500 °C, а

термоелектронні струми, які знімаються з них, на практиці не перевищують 10^3 А/м^2. За дуже короткочасних струмів (імпульси струму тривалістю 10^{-6} ... 10^{-5} с) оксидні катоди здатні давати емісію до 10^6 А/м^2 і вище.

20.4. Контакт двох металів

Утворення контакту між двома твердими тілами неминуче викликає взаємний обмін рухомими носіями заряду. За однакової температури тіл у стаціонарному стані встановлюється термодинамічна рівновага, яка характеризується постійним значенням рівня Фермі, єдиним для всього колективу рухомих носіїв. Встановлення рівноваги супроводжується переходом електронів, який здійснюється у результаті термоелектронної емісії або тунельного проходження, або дифузійних процесів залежно від якості контакту і властивостей тіл, які контактують.

Розглянемо два метали, енергетичні діаграми яких показано на рис. 20.9, *а*. У разі достатньо малій відстані (близько декількох ангстрем (Å)), між ними можливий обмін електронами внаслідок термоелектронної емісії. Якщо робота виходу $\Phi_1 < \Phi_2$, то струм емісії (20.9) j_1 буде перевищувати j_2, і другий метал набуде надлишкового електронного заряду. Зменшення його потенціалу викликає появу в зазорі електричного поля, яке перешкоджає потоку електронів з першого металу у другий. На енергетичній діаграмі (рис. 20.9, *а*) це відобразиться взаємним зміщенням енергетичних рівнів і появою додаткового потенціального бар'єра $eU_к$, який зменшує струм j_1 і фактично не впливає на струм j_2:

$$j_1 = A_1 T^2 \exp\{-(\Phi_1 + eU_к)/(kT)\},$$
$$j_2 = A_2 T^2 \exp\{-\Phi_2/(kT)\}. \quad (20.11)$$

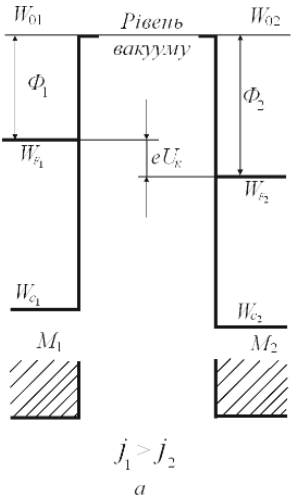


Рис. 20.9, а. Енергетичні діаграми двох металів

Рівновага встановлюється, коли вирівнюються зустрічні потоки електронів. Значення рівноважного потенціального бар'єра знаходимо з умови $j_1 = j_2$ (для простоти вважатимемо, що $A_1 = A_2$):

$$eU_к = Φ_2 - Φ_1. \quad (20.12)$$

Різницю потенціалів $U_к$, яка виникає між металами у рівноважному стані, називають контактною. Зауважимо, що перехід електронів з одного металу в інший змінює концентрацію електронів у них, але настільки незначно, що значення енергії Фермі W_F, яку відраховують від дна зони провідності W_c і яка залежить від концентрації електронів (див. підрозд. 19.7, формула (19.22)), майже не змінюється. Контактну різницю потенціалів визначають відстанню між початковими положеннями рівнів Фермі (рис. 20.9, *а*), або, що те ж саме, різницею початкових значень термодинамічних робіт виходу (рис. 20.9, *б*).

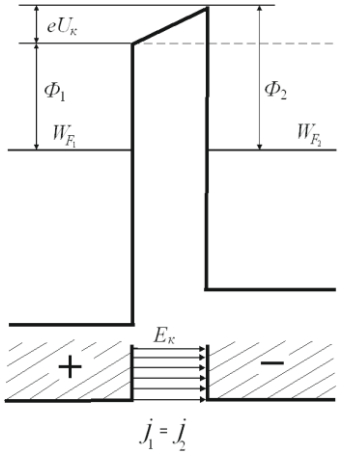


Рис. 20.9, б. Контактну різницю потенціалів визначають різницею початкових значень термодинамічних робіт виходу

Рівновагу металів, які контактують, можна порушити, якщо до них прикласти напругу U від зовнішнього джерела. Залежно від знака цієї напруги (позитивною будемо вважати напругу, яка створює електричне поле, що має напрямок, протилежний до напрямку контактного поля, як показано на рис. 20.9, *в*) буде змінюватись висота бар'єра, а значення струму j_1 знайдемо за формулою

$$j_1 = A_1T^2\exp\{-(Φ_1 + eU_к - eU)/(kT)\}.$$

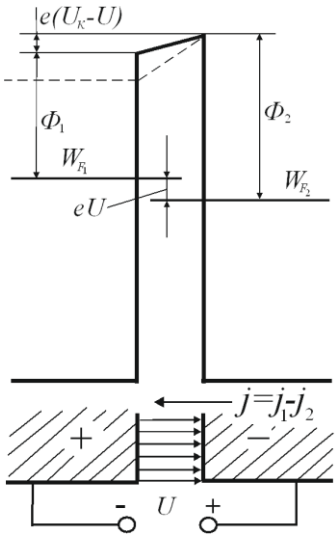


Рис. 20.9, *в. До металів, які контактують, прикладена зовнішня напруга* U

Тоді результуючий струм j, який протікає через контакт, дорівнює різниці $j_1 - j_2$. Враховуючи (20.12), знайдемо

$$j = j_s(\exp\{eU/(kT)\} - 1), \quad (20.13)$$

де $j_s = A_2T^2\exp\{-Φ_2/(kT)\}$ – струм насичення.

Вираз (20.13) описує ВАХ контакту двох металів. Її вигляд, зображений на рис. 20.10, свідчить про появу нелінійних властивостей контакту, тобто про різний опір у прямому і зворотному напрямках включення зовнішнього джерела. Слід мати на увазі, що мова йде не про ідеальний контакт, а контакт з досить малим вакуумним зазором між металами.

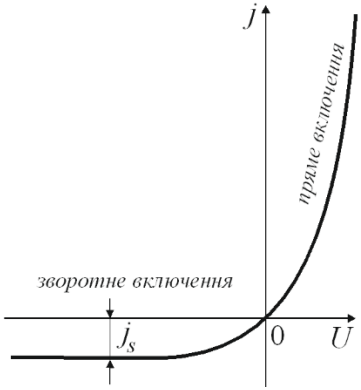


Рис. 20.10. Ідеалізована вольт-амперна характеристика контакту двох металів, які розділені вакуумним проміжком

Розглянуті явища відбуваються у приладах, які використовують структури типу МДМ («метал – діелектрик – метал»), які можна уявляти як контакт двох металів з вакуумним зазором, який замінює тонка діелектрична плівка.

Вольт-амперні характеристики МДМ–структур відрізняються від ВАХ, які визначають за формулою (20.13), оскільки тут не враховано можливість тунельного проходження електронів через бар'єр, а також сили дзеркального відображення, які суттєво змінюють форму бар'єра, що й обумовлює ефект Шотткі (формула (20.7)).

Контрольні запитання

1. Що таке термоелектронна емісія? Який її механізм?
2. Сформулюйте закон термоелектронної емісії Богуславського–Ленгмюра.
3. Сформулюйте закон термоелектронної емісії Річардсона–Дешмана.
4. Що таке робота виходу електрона з металу?
5. Які фізичні фактори визначають величину роботи виходу?
6. Зобразіть ВАХ (теоретичні), дайте їм пояснення.
7. Що таке ефект Шотткі? Як він вливає на вигляд ВАХ?

Розділ 21. Контактні явища

Властивості напівпровідникових пристроїв визначаються параметрами енергетичних бар'єрів, які виникають між двома частинами напівпровідникового кристала. Найбільш важливими параметрами бар'єрів є їхня енергетична висота, ширина і розподіл напруженості електричного поля всередині бар'єра.

21.1. Проникнення зовнішнього електричного поля в метал, діелектрик, напівпровідник

Метал. Розмістимо всередині конденсатора металічну пластину, (як показано на рис. 21.1, *а*). Як відомо з підрозд. 3.1, напруженість електричного поля всередині металу дорівнює нулю. Поки електричного поля біля металу немає у кожному, навіть дуже маленькому, об'ємі всередині металу і на його поверхні кількість позитивних іонів точно дорівнює кількості негативно заряджених електронів, тому будь-яка частина металу електронейтральна.

У металі, який розміщений в електричному полі, електронейтральність біля поверхні порушується. До поверхні, яка повернута до позитивного електрода зовнішнього джерела, наближаються електрони, а від поверхні, повернутої до негативного електрода, вони віддаляються, залишаючи біля поверхні шар позитивно заряджених іонів (рис. 21.1, *а*). Таким чином заряд на поверхні металу створює електричне поле, яке направлене проти зовнішнього поля, яке його викликало. Зовнішнє поле повністю екранується і в результаті поля всередині металу немає.

Це можна зрозуміти ще і з таких міркувань. Із підрозд. 1.5 відомо, що густина силових ліній пропорційна напруженості електричного поля, починаються лінії на позитивних, а закінчуються на негативних зарядах, тому, який би не був густий потік силових ліній поля, що створене зовнішнім джерелом біля поверхні металу, внаслідок великої концентрації електронів у металі ($\sim 10^{28}$ м$^{-3}$), силові лінії поля завжди замикаються на поверхневих зарядах і не проникають у глибину провідника (рис. 21.1, *а*). Утім, якщо поблизу поверхні провідника електронейтральність порушується, то що ж означає поняття «поблизу»? На яку глибину поле проникає в метал?

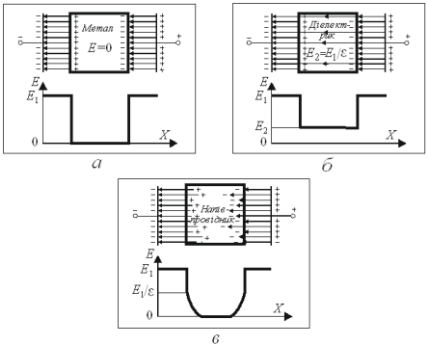


Рис. 21.1. Розподіл поля в металі (а), діелектрику (б) і напівпровіднику n-типу (в), які розміщені у зовнішньому електричному полі

Припустимо, наприклад, що під дією зовнішнього поля для того, щоб екранувати його, усі електрони в металічній пластині вилетіли з шару, товщина якого Δx дорівнює всього тільки постійній ґратці, яка у твердих тілах

дорівнює приблизно $5 \cdot 10^{-10}$ м. Яке поле створюється при цьому біля поверхні металу? Або, що те ж саме, яке зовнішнє поле буде при цьому екрановане? Якщо з якої-небудь ділянки металу «пішли» всі «вільні» електрони, то всередині цієї ділянки виникає позитивний об'ємний заряд з густиною $\rho = |e|N$, де N– концентрація позитивно заряджених іонів, яка точно дорівнює концентрації електронів n в металі. Об'ємному заряду ρ біля поверхні в шарі, товщина якого Δx, відповідає густина поверхневого заряду $\sigma = \rho\Delta x = eN\Delta x$. Тоді згідно з (3.1) напруженість електричного поля E поблизу поверхні металу буде:

$$E = \sigma/\varepsilon_0 = eN\Delta x/\varepsilon_0 =$$

$$= \frac{1{,}6 \cdot 10^{-19} \cdot 10^{28} \cdot 5 \cdot 10^{-10}}{8{,}85 \cdot 10^{-12}} \text{ В/м} =$$

$$= 9 \cdot 10^{10} \text{ В/м}.$$

Отримане значення E дуже велике. Якщо $E \sim 3 \cdot 10^6$ В/м, у повітрі виникає електричний пробій. Утім, щоб екранувати дію навіть такого сильного поля, достатньо, щоб із приповерхневого шару металу, товщина якого становить один атомний шар, «пішла» $3 \cdot 10^6/9 \cdot 10^{10} \approx 3 \cdot 10^{-5}$ – стотисячна частка електронів.

Таким чином, навіть дуже сильні електричні поля не проникають у метал на глибину більшу, ніж стала ґратки, та навіть у межах цієї товщини відносне збіднення електронами зі сторони «мінуса» зовнішнього джерела (і, відповідно, відносне збагачення зі сторони «плюса») не перевищує декількох часток відсотка.

Діелектрик. Проникнення електричного поля в діелектрик було розглянуто у розд. 2. Тут тільки зазначимо, що на мові силових ліній діелектрик в електричному полі показаний на рис. 21.1, *б*. Частина силових ліній зовнішнього поля залишається на зв'язаних поверхневих зарядах діелектрика, але ті силові лінії, які не замкнулись на поверхневих зв'язаних зарядах і проникли всередину діелектрика, пронизують його наскрізь. Всередині діелектрика немає вільних зарядів, які могли би перегрупуватись і, змінивши своє положення, екранувати дію електричного поля, яке проникло в діелектрик.

Напівпровідник. У пластині напівпровідника, яка розміщена в електричному полі (рис. 21.1, *в*), з одного боку, як у діелектрику, на поверхні виникнуть унаслідок поляризації зв'язані заряди. Таким чином, у напівпровіднику біля межі з вакуумом (або повітрям) напруженість поля E_2 буде в ε раз меншою, ніж у вакуумі. З другого боку, в напівпровіднику, як і в металі, є вільні носії заряду, які здатні, перегрупу-

вавшись, екранувати дію електричного поля, яке проникло у напівпровідник.

Розглянемо, на яку глибину поле проникає у напівпровідник, та у чому відмінність між напівпровідниками і металами. Ця відмінність полягає в тому, що концентрація вільних носіїв і, відповідно, заряджених іонів у напівпровідниках зазвичай значно менша, ніж у металах, тому, щоб створити достатньо великий позитивний заряд, який здатний екранувати дію зовнішнього поля, електрони у напівпровіднику n-типу повинні «відтікти» не з тонкого шару, товщина якого дорівнює сталій ґратки, а із шару досить великої товщини (рис. 21.1, *в*). Зрозуміло, що товщина цього шару буде тим більшою, чим менша рівноважна концентрація електронів n у напівпровіднику.

Концентрація електронів n у багатьох випадках визначається концентрацією донорів N_d. Таким чином, глибина, на яку поле проникає всередину напівпровідника, визначається рівнем його легування.

Рівень легування напівпровідників, які використовують для виготовлення різних напівпровідникових приладів, лежить у межах від $\sim 10^{16}$ м$^{-3}$ до $\sim 10^{26}$ м$^{-3}$, тобто може відрізнятись в десятки мільярдів разів. Відповідно, і глибина проникнення поля в напівпровідник, або, інакше кажучи, ширина енергетичних бар’єрів у напівпровідниках може змінюватись у дуже широких межах. У дірковому напівпровіднику ситуація із проникненням електричного поля в напівпровідник аналогічна тій, що і в електронному напівпровіднику.

21.2. Розподіл поля в бар’єрі. Ширина бар’єра

Поверхневі стани. Поверхня напівпровідникового кристала є порушенням в ідеальній періодичній решітці. Ідеальний порядок атомів, які розміщені в об’ємі напівпровідника на строго однаковій один від одного відстані, на поверхні раптово порушується. Зрозуміло, що умови зв’язку між валентними електронами, які належать приповерхновим атомам, будуть дещо іншими, ніж в об’ємі. Обрив кристалічної ґратки є причиною виникнення біля поверхні додаткових енергетичних рівнів. Ці поверхневі стани називають рівнями Тамма. Крім того, на поверхні адсорбуються чужерідні атоми й завжди існують структурні дефекти: спотворення ґратки, вакансії, тобто порожні місця у вузлах ґратки, які виникають, коли атоми напівпровідника зміщуються із своїх законних місць у вузлах ґратки в міжвузловини. Усі ці домішкові та структурні порушення теж є джерелом поверхневих станів.

Густина поверхневих станів на реальних поверхнях перебуває в межах від $\sim 10^{13}$ м$^{-2}$ до $\sim 10^{19}$ м$^{-2}$. Поверхневі рівні можуть бути як донорними, так і акцепторними.

Залежно від того, який тип поверхневих станів переважає (донорний чи акцепторний), поверхня буде або віддавати електрони в об'єм напівпровідника і заряджатись позитивно, або навпаки, захоплювати електрони з об'єму і заряджатись негативно.

Нехай на поверхні напівпровідника n-типу переважають акцепторні стани. Тоді поверхневі рівні будуть захоплювати електрони з частини об'єму напівпровідника, який прилягає до поверхні. У результаті поверхня заряджається негативно, а у приповерхневій області напівпровідника виникне шар, збіднений електронами, тобто заряджений позитивно. Таким чином, поблизу поверхні утворюється подвійний заряджений шар, поле якого буде відштовхувати електрони у глибину кристала. Це поле максимальне біля поверхні і спадає у міру віддалення від неї у глибину кристала. Це поле відштовхує електрони від поверхні кристала, а позитивно заряджені дірки притягуються до поверхні (рис. 21.2).

Біля поверхні всередині кристала утворюється, через подвійний заряджений шар, енергетичний бар'єр, висота якого W_s. Величину $U_s = W_s / e$ називають поверхневим потенціалом. Біля самої поверхні концентрація електронів провідності буде:

$$n_s = n_n \exp\left\{-\frac{W_s}{kT}\right\}, \qquad (21.1)$$

де $n_n -$ концентрація електронів у глибині кристала далеко від поверхні.

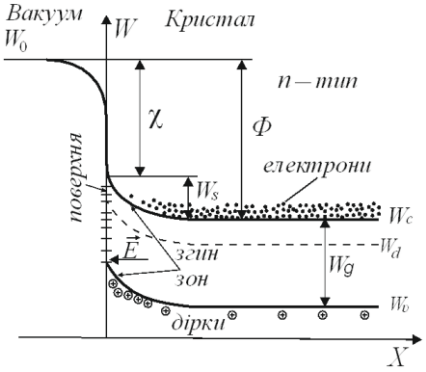


Рис. 21.2. Поверхня у напівпровіднику n -типу

Енергія дірок біля поверхні менша, ніж енергія дірок в об'ємі кристала, на величину W_s, тому концентрація дірок p_s біля поверхні становитиме

$$p_s = p_n \exp\left\{\frac{W_s}{kT}\right\}, \qquad (21.2)$$

де $p_n -$ концентрація дірок у глибині кристала.

З'ясуємо, до чого приводить наявність на поверхні кристала поверхневого потенціалу і згину зон, який супроводжує його. Нехай у кремнії n-типу концентрація електронів

$n_n = 10^{21}$ м$^{-3}$. Тоді згідно із законом діючих мас (19.48) рівноважна концентрація дірок у такому кристалі за $T = 300$ K становитиме $p_n = n_i^2/n_n = \frac{(1{,}1\cdot 10^{16})^2}{10^{21}} = 1{,}21\cdot 10^{11}$ м$^{-3}$. Електронів в об'ємі кристала в 10^{10} (10 міліардів) разів більше, ніж дірок.

Нехай величина поверхневого бар'єра $W_s = 0{,}1$ еВ. Тоді за кімнатної температури ($T = 300$ K) $kT = 0{,}026$ еВ, і концентрація електронів біля поверхні згідно з (21.1) становитиме $n_s = 10^{21}\cdot \exp\left\{\frac{-0{,}1}{0{,}026}\right\} \approx 2\cdot 10^{19}$ м$^{-3}$, а концентрація дірок згідно з (21.2) $p_s = 5{,}66\cdot 10^{12}$ м$^{-3}$. Електронів поблизу поверхні приблизно в 50 разів менше, ніж в об'ємі, а дірок у стільки ж раз більше. Утім, дірок біля поверхні все ще в $3{,}5\cdot 10^6$ раз менше, ніж електронів, і їхній внесок у провідність незначний, тому наявність поверхневого потенціалу в цьому випадку означає виникнення біля поверхні збідненого електронами шару з питомим опором, який значно більший, ніж об'ємний питомий опір. *Таке виникнення біля поверхні кристала збіднених шарів з високим питомим опором лежить в основі роботи діодів і транзисторів*. За $W_s = 0{,}2$еВ $n_s = 4{,}6\cdot 10^{17}$ м$^{-3}$, $p_s = 2{,}7\cdot 10^{14}$ м$^{-3}$. Якісно ситуація залишилась тією самою $n_s/p_s = 1{,}7\cdot 10^3$. За $W_s = 0{,}4$ еВ виникає якісно нова ситуація: тепер $n_s = 2\cdot 10^{14}$м$^{-3}$, $p_s = 6\cdot 10^{17}$м$^{-3}$. Дірок поблизу поверхні майже у три тисячі раз більше, ніж електронів! На поверхні напівпровідника *n*-типу утворився шар з дірковою провідністю – шар з інверсним типом провідності, або *інверсний шар.*

На рис. 21.3 зображено енергетичну діаграму для діркового напівпровідника за позитивного заряду поверхневих станів. На рис. 21.2 і 21.3 наведено такі позначення: χ —електронна спорідненість напівпровідника і, як видно з рисунків, $\chi \approx W_0 - W_c$ (для цього напівпровідника є величиною сталою); Φ — робота виходу електронів із напівпровідника.

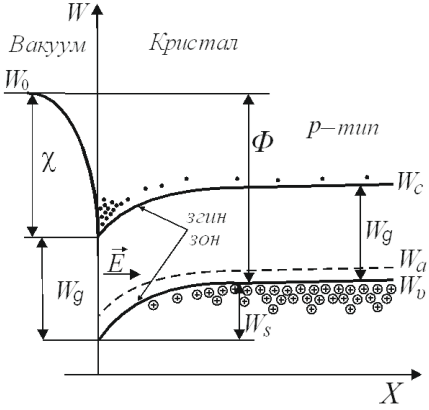


Рис. 21.3. Енергетична діаграма для діркового напівпровідника за позитивного заряду поверхневих станів

Для n- і p-кремнію величина поверхневого бар'єра W_s приблизно дорівнює $W_g/2$ (цей випадок трапляється дуже часто). Тоді для n-кремнію (рис. 21.2) робота виходу $\Phi_n \approx \chi + W_s = \chi + W_g/2$, а для p-кремнію (рис. 21.3) $\Phi_p \approx \chi + W_g - W_s \approx \chi + W_g/2$. Таким чином, $\Phi_n \approx \Phi_p$, тобто робота виходу електрона із напівпровідника n- і p-типу (кремнію) приблизно дорівнює значенню роботи виходу у власного кремнію.

Визначимо розподіл електричного поля і ширину енергетичного бар'єра в області поверхневого потенціалу W_s (рис. 21.4). Для цього використаємо рівняння Пуассона (1.38) і деякі припущення про характер бар'єра та напівпровідника.

По-перше, будемо вважати, що напівпровідник легований однорідно, тобто концентрація домішки скрізь однакова і не залежить від координати. Якщо це напівпровідник n-типу, то концентрація іонізованих донорів N_d скрізь однакова. Подруге, будемо вважати, що висота потенціального бар'єра W_s, з одного боку, достатньо велика, $W_s \gg kT$, але з другого боку, недостатня для того, щоб викликати інверсію провідності біля поверхні кристала. За цих двох умов розрахунок поля в бар'єрі стає дуже простим.

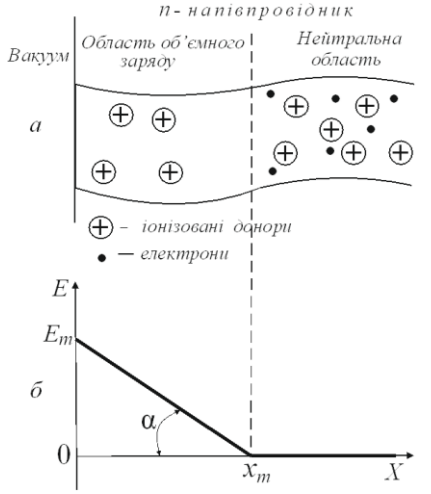


Рис. 21.4. Формування поверхневого потенціального бар'єра (напівпровідник n-типу): а – негативний заряд поверхневих станів спричинив відтік електронів із приповерхневого шару, де залишився нескомпенсований позитивний заряд іонізованих донорів; б – поле приповерхневої області напівпровідника лінійно зменшується від величини E_m (на межі «вакуум – кристал», $x = 0$) до нуля (на межі «область об'ємного заряду – нейтральна область», $x = x_m$), $\tan\alpha = eN_d/(\varepsilon\varepsilon_0)$

Справді, потенціальний бар'єр на поверхні обумовлений тим, що рухомі носії (у цьому разі електрони) відтекли від поверхні, залишивши нескомпенсованим позитивний об'ємний заряд іонізованих донорів (рис. 21.4, *а*). Якщо $W_s \gg kT$, із приповерхневого шару «йдуть» фактично всі вільні електрони. Тоді густина об'ємного заряду, природно, дорівнюватиме нескомпенсованому

заряду іонізованих донорів, які залишились: $\rho = eN_d$, де $N_d -$ концентрація донорів, $e -$ величина заряду електрона за модулем. Тоді згідно з рівнянням Пуассона (1.38) поле всередині поверхневого шару буде змінюватись за законом

$$\frac{dE}{dx} = \frac{\rho}{\varepsilon\varepsilon_0} = \frac{eN_d}{\varepsilon\varepsilon_0}. \qquad (21.3)$$

Оскільки $N_d -$ величина постійна, то відповідно до формули (21.3) швидкість зміни поля з координатою теж є величиною сталою, тобто напруженість поля E лінійно спадає з координатою x (рис. 21.4, *б*). Із формули (21.3) і (рис. 21.4, *б*) легко отримати, що

$$E(x) = \frac{eN_d}{\varepsilon\varepsilon_0}(x_m - x), \qquad (21.4)$$

де x_m – межа об'ємного заряду.

Якби нам була відома величина $E_m -$ поле на поверхні кристала, то із (21.4) ми легко визначили б ширину бар'єра x_m. Із (21.4) видно, що за $x = 0$, $E = E_m$:

$$E_m = \frac{eN_d}{\varepsilon\varepsilon_0}x_m. \qquad (21.5)$$

Утім, значення E_m невідоме. Відомо, що між поверхнею кристала і точкою $x = x_m$, де немає поля, існує потенціальний бар'єр висотою W_s, що означає, що між цими точками існує різниця потенціалів $U_s = W_s / e$. Цього достатньо, щоб визначити і величину E_m, і ширину бар'єру x_m. Справді, оскільки

$$E = -dU_s / dx,$$

то $U_s = -\int_0^{x_m} E(x)dx$.

Із рис. 21.4, *б* видно, що U_s формально дорівнює площі трикутника $OE_m x_m$, тобто

$$U_s = \frac{1}{2}E_m x_m. \qquad (21.6)$$

Тоді із співвідношень (21.5) і (21.6) отримаємо, що

$$x_m = \left(\frac{2\varepsilon\varepsilon_0 U_s}{eN_d}\right)^{\frac{1}{2}};$$
$$E_m = \left(\frac{2eN_d U_s}{\varepsilon\varepsilon_0}\right)^{\frac{1}{2}}. \qquad (21.7)$$

Проведемо деякі числові оцінки. Приймемо, що $U_s = 0{,}2$ В. Значення діелектричної проникності для найбільш широко розповсюджених напівпровідників Ge, Si і GaAs є: 16, 12,5 і 12 відповідно. Тоді, якщо $N_d = 10^{16}$ м$^{-3}$, то за формулами (21.7) отримаємо: $E_m = 2{,}4 \cdot 10^3$ В/м, $x_m = 170$ мкм. Якщо ж при тій самій приповерхневій різниці потенціалів $U_s = 0{,}2$ В рівень легування $N_d = 10^{26}$ м$^{-3}$, то $E_m = 2{,}4 \cdot 10^8$ В/м, $x_m = 0{,}0017$ мкм.

В першому випадку поле приповерхневого бар'єра проникає в глибину напівпровідника на товщину порядку 300 000 атомних шарів. У другому випадку – 3–4 шарів.

Зрозуміло, якщо ми маємо справу з дірковим напівпровідником, то всі приведені судження і всі отримані формули є справедливими. Необхідно тільки замінити в них величину N_d на

$(-N_a)-$ концентрацію акцепторної домішки.

21.3. *p-n*–перехід та його властивості

Якщо кристал напівпровідника легований так, що одна частина, має провідність *p*-типу (діркову), а інша – провідність *n*-типу (електронну), то на границі між цими частинами виникає шар з особливими властивостями, який називається *p–n*-переходом або, по-іншому, електронно-дірковим переходом.

21.3.1. Способи отримання $p-n$-переходів

Якщо взяти два куски напівпровідника (рис. 21.5, *а*): один *p*-, інший *n*-типу, акуратно їх відшліфувати і міцно притиснути один до одного, то ми отримаємо об'єкт, властивості якого не мають нічого спільного із властивостями *p–n*-переходу. Із підрозд. 21.2 відомо, що на поверхні, як би її акуратно не обробляли, завжди будуть порушення кристалічної грвтки. Утворюється спектр поверхневих станів, тобто виникають приповерхневі потенціальні бар'єри. Крім цього, на поверхні будуть окисли, адсорбовані сторонні атоми. Як щільно не притиснуті один до одного куски *p*- і *n*-типу напівпровідників, «рани», які нанесені кристалічній ґратці на поверхні, не «залікуєш».

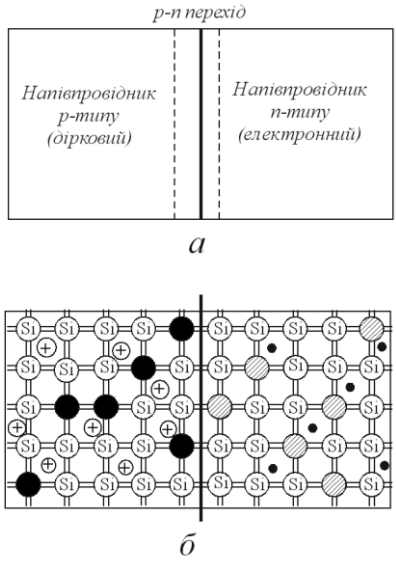


Рис. 21.5. На межі p- і n-типів напівпровідника виникає p–n-перехід (а). Показаний кристал кремнію, в лівій частині якого акцепторна домішка (темні кружечки), в правій – донорна домішка (заштриховані кружечки). Позначення: ⊕ – дірки, ● – електрони (б)

Яким же повинен бути контакт *p*- і *n*-напівпровідників? Звернувшись до рис. 21.5, *б*, зазначимо важливу обставину. Кристалічна ґратка напівпровідника нічого про *p–n*-перехід «не знає», вона просто його не помічає. На самій межі, яка розділяє *p*- і *n*-області, чи зліва, чи справа від неї, ми бачимо все ту ж кристалічну решітку кремнію, з тими ж зв'язками, які ніяк не порушуються тим, що в різні частини кристалу введені різні

домішки. Тільки такий перехід має всі ті чудові властивості, про які будемо говорити в подальшому і саме такий перехід будемо вивчати.

Проблема створення якісних $p-n$-переходів вирішувалась і багато в чому вирішена протягом більше півстоліття зусиллями тисяч спеціалістів. Розроблено досить багато способів отримання переходів і всі ці способи доведені до високої технічної доскона-лості. *Для всіх способів характерний спільний прийом*. Береться кристал напівпровідника *p*- або *n*-типу, а потім якимось способом на певну глибину кристалу вводяться домішки, які створюють в даному напівпровіднику провідність протилежного характеру. Якщо вихідний кристал був *n*-типу, то вводиться акцепторна домішка, якщо *p*-типу, то донорна. Домішка вводиться в кристал в більшій кількості, ніж вихідна. Тому в цій частині кристалу, куди попала домішка, тип провідності змінюється. Відбувається, як говорять, перекомпенсація вихідної домішки. На межі між областю, де тип провідності кристалу змінився і областю, де він залишився вихідним, виникає $p-n$-перехід.

Сплавлення. Сплавлення один із самих старих способів створення *p–n*-переходів. Перші *p–n*-переходи були отримані цим методом в 1950 р. В основі методу лежить утворення сплаву напівпровідникового матеріалу і металу, атоми якого є донорами або акцепторами в напівпровіднику.

На рис. 21.6 показаний приклад утворення сплавного $p-n$-переходу в $n-$Ge при вплавленні в нього індія (In). Таблетка індію, яка має форму тонкого диску, клалась на очищену поверхню кристалу германію (рис. 21.6, *а*). Потім кристал разом з таблеткою повільно нагрівають. Як тільки температура перевищить температуру плавлення індію (156 ℃) таблетка починає плавитись, перетворюючись в краплю рідини. При підвищенні температури форма краплі, як видно із рис. 21.6, *а*, змінюється. Сили поверхневого натягу рідини стараються надати краплі форму кулі, а сили змочування – розмазати краплю по поверхні кристалу. Індій при $t \simeq 300$ ℃ має форму кульки, при більш високій температурі (≈ 500 ℃) розтікається по поверхні германію.

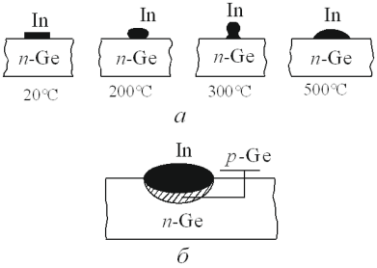


Рис. 21.6. Отримання p–n-переходу сплавленням

Кристалічний германій при температурі ~ 500 ℃ дуже добре розчиняється в рідкому індії. Розчинення продовжується до тих пір, поки не утвориться насичений розчин германію і індію. Після цього пластинку германія починають повільно охолоджувати. Із зменшенням температури кількість германію, здатного розчинятись в індії, теж зменшується; надлишковий германій, який випадає із розчину, знову кристалізується на нерозчиненій частині пластинки (рекристалізується). Цей рекристалізований шар (рис. 21.6, *б*) є збагачений атомами індію. Атоми індію в германію є акцепторами. Концентрація індію в рекристалізованому шарі складає $\sim 10^{25}$ м$^{-3}$ і значно перевищує концентрацію донорної домішки у вихідному кристалі. Тому на межі рекристалізованої області і нерозчиненої частини кристалу утворюється $p-n$-перехід.

В 50-ті роки 20 ст. метод сплавлення відігравав важливу роль в розвитку напівпровідникової електроніки. Саме на його основі були створені перші промислові германієві діоди і транзистори.

Дифузія. Дифузійний метод виготовлення $p-n$-переходів з'явився незабаром після методу сплавлення як деяка його модифікація. Метод дифузії виявився більш ефективним і на довгі роки був основним методом створення $p-n$-переходів, а потім і інтегральних схем.

Розглянемо процес дифузії детальніше. Нехай напівпровідниковий кристал розміщений в атмосфері газу, який містить велику концентрацію атомів речовини, яка є в даному напівпровіднику донорною (або акцепторною) домішкою. На поверхні будь-якого кристалу є безліч порушень кристалічної ґратки. Тому атоми домішки із газу – дифузанта легко захоплюється поверхнею напівпровідника. Поверхня стає збагаченою домішками. Оскільки ідеальних кристалів не існує, то дифузія в глибину кристалу буде відбуватись, головним чином, по вакансіям, міжвузловинах та інших порушеннях в ґратці кристалу при його нагріванні. Дифузійний потік домішкових атомів направлений від поверхні, де домішкових атомів багато, в глибину напівпровідника, де їх немає.

На рис. 21.7 показана схема проведення процесу дифузії. Кварцева труба 1 розміщена в пічці, яка створює всередині труби температуру $1000-1300$ ℃. В трубі знаходиться кварцева касета 3 із закріпленими на ній пластинками кремнію 4. Речовина домішки (так званий дифузант) поступає

в трубу у вигляді газу, концентрація якого строго контролюється. Протягом всього часу, поки поверхня напівпровідника обдувається дифузантом, атоми домішки дифундують в глибину пластинки. Чим більший час дифузії, тим глибше проникають домішки в кристал. $p-n$-перехід при дифузії виникає внаслідок того, що кристал, в якому проводиться дифузія, має вихідну об'ємну домішку, яка протилежна по типу від домішки, яка вводиться при дифузії. Якщо вихідна домішка, наприклад, донорного типу, то для створення p–n-переходу використовують дифузію акцепторів.

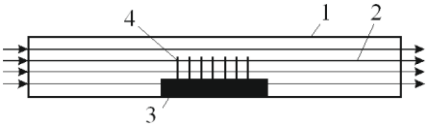


Рис. 21.7. Отримання p–n-переходу дифузією

Глибина залягання $p-n$-переходів, які виготовляються методом дифузії, залежить від об'ємної концентрації вихідної домішки в кристалі та лежить в межах від долей мікрометра до сотень мікрометрів.

Іонна імплантація. Це один із сучасних методів виготовлення $p-n$-переходів, який дозволяє отримувати p–n-переходи дуже маленьких розмірів з глибиною залягання, яка точно контролюється і є досить малою (до сотих долей мікрометра). Суть методу іонної імплантації, або, як його деколи називають методом іонного легування, в тому, що в напівпровідник проникають атоми домішки, які рухаються строго в одному напрямку з певною і досить великою швидкістю, на відміну від методу дифузії, в якому атоми домішки рухаються хаотично.

Установка для проведення іонної імплантації схематично показана на рис. 21.8. У камері 1, де відбувається процес, створюється і підтримується глибокий вакуум. Іони домішки виникають в джерелі іонів 2, яке складається із матеріалу домішки, який нагрівається до температури, щоб почалось випаровування атомів з поверхні. Біля поверхні матеріалу домішки, горить електрична дуга, яка іонізує практично всі частинки, які в неї попадають.

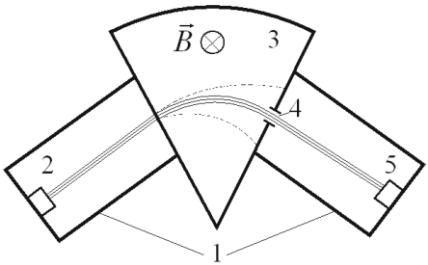


Рис. 21.8. Отримання p–n-переходу іонною імплантацією

Позитивні іони домішки, які утворились, витягуються через вихідний отвір джерела сильним електричним полем. Особлива система електродів (яку

називають електростатичною лінзою) надає пучку іонів необхідну голкоподібну форму. Прискорені полем до великих швидкостей і сфокусовані іони попадають в магнітний аналізатор 3. Це найважливіший вузол, який забезпечує одну із найголовніших переваг методу іонного легування: практично повне очищення пучка іонів від сторонніх домішок. В аналізаторі пучок іонів попадає в магнітне поле (силові лінії магнітного поля $\vec{B}$ на рис. 21.8 перпендикулярні площині рисунка). На іони діє сила Лоренца. Радіус траєкторії іонів в магнітному полі залежить від їх швидкості і від маси іонів. Тому, якщо в пучку є сторонні частинки крім іонів легуючої домішки, вони відхиляються в аналізаторі не на той кут, на який відхиляються іони домішки.

В вихідній стінці аналізатора є щілина 4, через яку проходять тільки ті іони, які відхилились строго на заданий кут. «Вибирати» тип іонів в пучку дуже легко: достатньо змінити величину індукції магнітного поля аналізатора.

«Відфільтровані» іони, які вилітають із аналізатора, підлітають до напівпровідникової пластинки 5. При зіткненнях з решіткою напівпровідника іони поступово витрачають свою енергію, віддаючи її атомам, які попадаються на їхньому шляху, і в кінці кінців зупиняються. Шлях, який іон проходить до повної зупинки, або глибина проникнення, залежить від маси і енергії іона. Енергію іонів можна точно регулювати, змінюючи прискорюючий іони електричний потенціал. Таким чином, проникнення іонів домішок в решітку відбувається на строго задану глибину. Наприклад, іони бора B^+ при енергії 20 кеВ проникають в кремнієву пластинку на глибину 0,071 мкм, при енергії 100 кеВ – на 0,327 мкм, а при енергії 300 кеВ – на 0,559 мкм.

Концентрація введеної домішки визначається сумарною дозою опромінення. Ця величина з високою точністю контролюється часом іонного бомбування. Таким чином, іонна імплантація – один із найточніших методів виготовлення *p–n*-переходів.

На шляху іонного пучка розміщують електростатичні пластини, такі, як в електронно-променевій трубці осцилографів. Подаючи на ці пластини напругу, яка змінюється за заданим законом, який виробляє і управляє ним ЕОМ, можна відхиляти іонну «голку» в потрібну точку пластини. Таким чином, іонний промінь «малює» необхідну конфігурацію $p-n$-переходів. Закінчивши весь «малюнок» для однією домішки і змінивши магнітне поле аналізатора і енергію пучка, можна зразу ж виконати

інший рисунок, який створить іншу систему $p-n$-переходів на іншій глибині, і т.д. Такі установки використовуються для виготовлення інтегральних схем, в яких на одній напівпровідниковій пластині розміщується сотні тисяч і більше $p-n$-переходів.

21.3.2. Потенціальний бар'єр в області контакта p- і n-напівпровідників

З'ясуємо, чому між двома областями одного і того самого напівпровідника, які відрізняються один від одного тільки характером введеної домішки, виникає потенціальний бар'єр. На рис.21.9, *а* зліва показаний напівпровідник p-типу, справа n-типу. Дірковий напівпровідник містить при звичайних умовах позитивно заряджені дірки – рухомі частинки, які здатні переносити заряд (створювати струм), і негативно заряджені акцептори, закріплені в кристалічній гратці. Електронний напівпровідник, навпаки, містить негативно заряджені електрони провідності і позитивно заряджені іонізовані донори, які закріплені в гратці.

На рис. 21.9, *б* наведена енергетична діаграма напівпровідників p- і n-типу. Це, як правило, сильно леговані напівпровідники, в яких концентрація основних носіїв заряду визначається

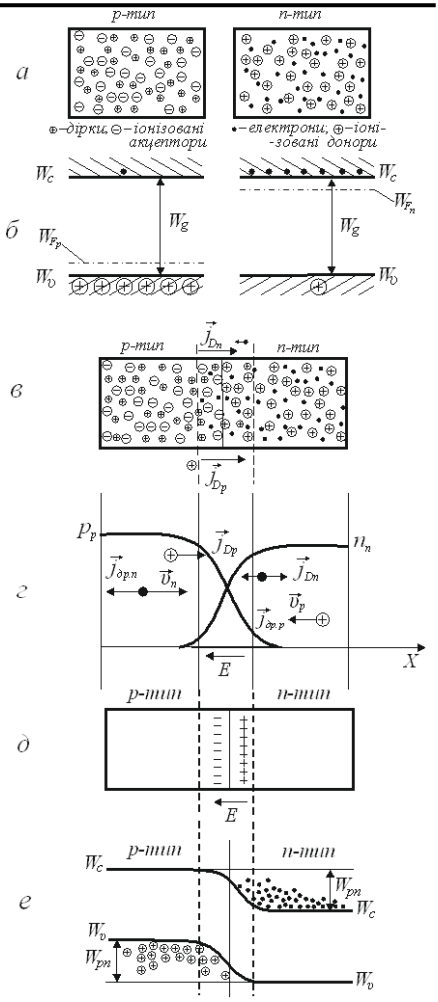


Рис.21.9 Утворення потенціального бар'єру на межі $p-n$-переходу: а– напівпровідники p- і n-типу, в цілому електронейтральні; б– енергетична діаграма напівпровідників, W_{F_p} і W_{F_n} – енергії Фермі, W_g – ширина забороненої зони; в – дифузія електронів в напівпровідник $p-$типу і дірок в напівпровідник n-типу; г– напрямки дифузійних струмів основних і дрейфових

струмів неосновних носіїв заряду; д – біля межі p–n-переходу виникає подвійний заряджений шар; е – наявність подвійного зарядженого шару на межі p–n-переходу еквівалентно існуванню на границі енергетичного бар'єра. Цей бар'єр перешкоджає проникненню електронів із напівпровідника n-типу в напівпровідник p-типу і проникненню дірок із напівпровідника p-типу в напівпровідник n-типу

рівнем легування (див. підрозд. 19.10; формули (19.62); (19.63)). Концентрація неосновних носіїв заряду в напівпровідниках визначається формулами (19.64) і (19.65) із підрозд. 19.10. Положення рівня Фермі в напівпровіднику *p*-типу визначається формулою (19.67), а в напівпроводнику *n*-типу формулою (19.66). Рівні Фермі в таких напівпровідниках знаходяться близько біля «стелі» валентної зони W_v і «дна» зони провідності W_c. Таке розташування рівнів Фермі відображає той факт, що напівпровідники сильно леговані і концентрація основних носіїв заряду значно переважає концентрацію неосновних носіїв заряду. Напівпровідники однакові (наприклад, кремній), тому ширина забороненої зони W_g в них однакова.

На рис. 21.9, *в* показано, що відбудеться, коли такі напівпровідники утворюють контакт. Повинен виникнути дифузійний потік електронів із напівпровідника *n*-типу в напівпровідник *p*-типу. І навпаки – дифузійний потік дірок із *p*- в *n*-напівпровідник. Густини дифузійних струмів (рис. 21.9, *г*) визначаються формулами (19.99) і (19.100) із підрозд. 19.12. При цьому частина напівпровідника *p*-типу, біля $p-n$-переходу буде заряджатись негативно, а частина напівпровідника *n*-типу біля $p-n$-переходу буде набувати позитивний заряд, оскільки в ній залишається нескомпенсований позитивний заряд нерухомих донорів. Дифузія дірок із *p*- в *n*-область сприяє заряджанню біля p–n-переходу напівпровідника *p*-типу негативно, а *n*-типу – позитивно. Таким чином, біля p–n-переходу утворюється подвійний заряджений шар, негативний зі сторони *p*-області і позитивний зі сторони *n*-області (рис. 21.9, *д*), який створює контактне електричне поле $\vec{E}$. Це поле буде протидіяти процесу дифузії. Воно виштовхує дірки із цього шару в *p*-напівпровідник, а електрони в *n*-напівпровідник.

Контактне поле $\vec{E}$ сприяє дрейфу електронів із *p*-напівпровідника в *n*-область і дрейфу дірок із *n*-напівпровідника в *p*-область (рис. 21.9, *г*). Густини дрейфових струмів визначається формулою (19.95) і в нашому випадку можуть бути записані так:

$$j_{\partial p.n} = en_p v_n = e\frac{n_i^2}{N_a}v_n, \qquad (21.8)$$

$$j_{\partial p.p} = ep_n v_p = e\frac{n_i^2}{N_d}v_p. \qquad (21.9)$$

При написанні цих виразів враховані формули (19.64) і (19.65). Розглянемо процеси детальніше. Неосновні носії в p-області з концентрацією n_p у своєму хаотичному русі попадають в область контактного поля, яке є для них прискорюючим, і переходять в n-область напівпровідника. Якщо в області переходу їхня кількість не змінюється в результаті, наприклад, процесів рекомбінації або генерації, то величина струму, який створюється електронами, визначається тільки кількістю електронів, що попали на межу прискорюючого поля і майже не залежить від напруженості електричного поля. Аналогічна ситуація і із дрейфом дірок із n-області в p-область. Тому дрейфовий струм через $p-n$-перехід буде дорівнювати:

$$j_s = j_{\partial p.n} + j_{\partial p.p} = e\left(\frac{v_n}{Na} + \frac{v_p}{N_a}\right)n_i^2. \qquad (12.10)$$

Враховуючи формулу (19.44) можемо зробити висновок, що дрейфовий струм неосновних носіїв

$$j_s \sim \exp\left\{-\frac{W_g}{kT}\right\}, \qquad (21.11)$$

тобто він різко залежить від температури і величини ширини забороненої зони W_g.

Таким чином, в області $p-n$-переходу встановиться рівновага між потоками електронів і дірок. Але, якщо контактне електричне поле $\vec{E}$ (рис. 21.9, *д*) сприяє переходу неосновних носіїв заряду через перехід, то для основних носіїв заряду воно є гальмівним, тобто бар'єром (рис. 21.9, *е*). Проте, внаслідок хаотичного теплового руху, серед основних носіїв заряду, яких багато, знайдуться дуже енергійні, енергія яких буде більша висоти бар'єру W_{pn}. Концентрація їх визначається розподілом Максвелла–Больцмана, тобто

$$n \approx n_n \exp\left\{-\frac{W_{pn}}{kT}\right\}, \qquad (21.12)$$

$$p \approx p_p \exp\left\{-\frac{W_{pn}}{kT}\right\}. \qquad (21.13)$$

Концентрація цих «гарячих» електронів і дірок дуже мала, але і концентрація неосновних носіїв заряду, які утворюють струм (21.10), теж дуже мала. Тому <u>в області $p-n$-переходу встановлюється *динамічна* рівновага між потоками основних і неосновних носіїв заряду, результатом чого є виникнення контактної різниці потенціалів</u>

$$U_{pn} = W_{pn} / e.$$

Рис. 21.9, *е* пояснює, що означає існування в місці контакту p- і n-напівпровідників енергетичного бар'єра. Із рисунка видно, що електрону для того, щоб попасти із n-області в p-область, необхідно подолати енергетичний бар'єр,

висота якого W_{pn}. Такий же бар'єр необхідно подолати і дірці для того, щоб потрапити з p- в n-область $p-n$-переходу.

Висота бар'єра. Як уже обговорювалось у підрозд. 19.9 рівень Фермі в напівпровіднику відіграє роль «центру мас» між електронами і дірками і його розташування в забороненій зоні напівпровідника завжди ближче до тих носіїв заряду, яких більше. Тому можна вважати, що енергії мілких акцепторних і донорних рівнів W_a і W_d майже збігаються з енергіями Фермі W_{F_p} і W_{F_n} (рис. 21.10).

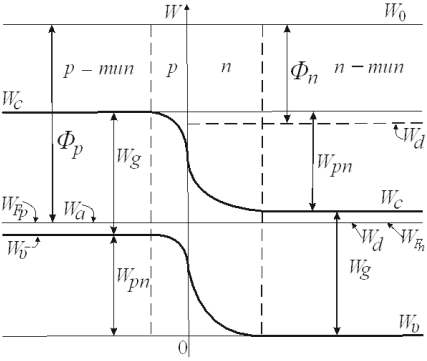


Рис. 21.10. Енергетичний бар'єр на межі p–n-переходу. $\Phi_p = W_0 - W_a$, $\Phi_n = W_0 - W_d$. W_0 – енергія вакууму (енергія вільного електрона). W_a і W_d – енергії мілких акцепторних і донорних рівнів, які майже збігаються з енергіями Фермі (W_{F_p} і W_{F_n}) в p- і n-напівпровідниках відповідно

Робота виходу електронів із напівпровідника p-типу $\Phi_p = W_0 - W_a = W_0 - W_{F_p}$, а для напівпровідника $n-$ типу $\Phi_n = W_0 - W_d = W_0 - W_{F_n}$, де W_0 – енергія вакууму (енергія вільного електрону). Електрону, щоб потрапити з n-області в p-область необхідно подолати енергетичний бар'єр, висота якого

$$W_{pn} = \Phi_p - \Phi_n = W_{F_n} - W_{F_p}.$$

З урахуванням формул (19.66) і (10.67), отримаємо

$$W_{pn} = W_g - kT \ln \frac{N_c N_v}{N_a N_d}, \quad (21.14)$$

де враховано, що $W_g = W_c - W_v$. Цій формулі можна надати і такий вигляд:

$$W_{pn} = kT \ln \frac{n_n p_p}{n_i^2} \cdot \quad (21.15)$$

При цьому враховано, що $n_n \approx \approx N_d$, $p_p \approx N_a$, $n_i^2 = N_c N_v \exp\left\{-\frac{W_g}{kT}\right\}$.

Із (21.14) видно, що величина потенціального бар'єра *тим більша, чим сильніше леговані* p- і n-області напівпровідників. Якщо $N_d \approx N_c$, а $N_a \approx \approx N_v$, то контактна різниця потенціалів $U_{pn} = W_{pn} / e$ матиме максимальне значення W_g/e. У підрозд. 19.5 приводяться значення W_g при кімнатній температурі для різних напівпровідників. Із приведеної таблиці видно, що U_{pnmax} може набувати значення від 0,17 В у InSb до 3 В у SiC. Зрозуміло, що такі значення U_{pnmax} є чисто теоретичні, на практиці U_{pn} є меншими.

Збіднений шар. Між p- і n-областями в p–n-переході існує енергетичний бар'єр, максимальна висота якого

складає ~ 0,7 еВ для германієвого, ~ 1 еВ для кремнієвого, ~ 1,4 еВ для GaAs переходу. Більшість електронів в зоні провідності (і дірок в валентній зоні) мають енергію порядку kT (тобто всього 0,026 еВ при кімнатній температурі). Зрозуміло, що для більшості носіїв бар'єр на межі *p*– і *n*– областей є непроникний. Кількість електронів і дірок, які здатні подолати бар'єр на межі *p*- і *n*-областей дуже не значна і за допомогою формул (21.12) або (21.13) можемо оцінити їх долю для Ge при $T = 300$ К:

$$\frac{n}{n_n} \sim \exp\left\{-\frac{0{,}70}{0{,}026}\right\} \sim 10^{-12}.$$

21.3.3. Розподіл електричного поля в бар'єрі. Ширина бар'єра

Оскільки область енергетичного бар'єра практично повністю збіднена вільними носіями, можна досить просто розрахувати всі основні характеристика бар'єра. Справа від межі $p-n$-переходу, в *n*-області, в кожному кубічному метрі напівпровідника міститься N_d позитивно заряджених донорів. Оскільки концентрація електронів і дірок в області бар'єра дуже незначна, то густина об'ємного заряду ρ в бар'єрі буде визначатись концентрацією донорів $\rho = eN_d$. Точно так само в *p*-області (зліва від переходу) густина об'ємного заряду визначається концентрацією негативно заряджених акцепторів: $\rho = -eN_a$ (рис. 21.11, *а*, *б*). У випадку, якщо густина об'ємного заряду ρ постійна, визначення розподілу поля в бар'єрі $E(x)$ і ширина бар'єру x є проста задача, враховуючи розв'язану задачу в підрозд. 21.2. Залежність $E(x)$ в *p–n*-переході показана на рис. 21.11, *в*. Напруженість електричного поля максимальна в площині *p–n*-переходу і лінійно спадає при віддаленні від переходу.

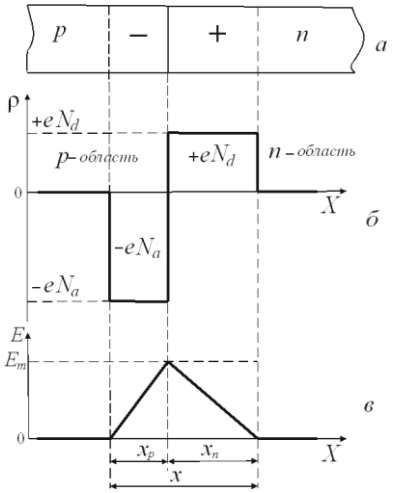


Рис. 21.11. Розподіл об'ємного заряду і електричного поля в p–n-переході: а – якісний розподіл заряду; б – залежність густини об'ємного заряду $\rho(x)$; в – розподіл напруженості електричного поля $E(x)$; N_d – концентрація донорів в n-області; N_a – концентрація акцепторів в p-області; x_p – і x_n – глибина проникнення електричного поля в p-область і n-область відповідно; $x = x_p + x_n$ – ширина p–n-переходу

Скориставшись формулою (21.4) отримаємо, що глибина проникнення поля в p-область буде

$$x_p = \frac{\varepsilon\varepsilon_0 E_m}{eN_a}, \qquad (21.16)$$

а в n-область –

$$x_n = \frac{\varepsilon\varepsilon_0 E_m}{eN_d}. \qquad (21.17)$$

Враховуючи це, із трикутника на рис. 21.11, *в* легко знайти, що

$$U_{pn} = \frac{1}{2} E_m \cdot x, \qquad (21.18)$$

де x – ширина $p-n$-переходу (основа трикутника). Тому легко отримати, що

$$x = x_p + x_n = \sqrt{\frac{2\varepsilon\varepsilon_0 U_{pn}}{e}\left(\frac{1}{N_d} + \frac{1}{N_a}\right)}. \qquad (21.19)$$

На практиці використовуються різко несиметричні $p-n$-структури. Так називаються $p-n$-переходи, в яких одна область легована сильніше, ніж інша. На рис. 21.12 показаний розподіл поля в $p-n$-переході, в якому p-область легована значно сильніше, ніж n-область, тобто $N_a \gg N_d$. Із рисунка видно, що в цьому випадку ширина області об'ємного заряду x_p в p-області значно менша, ніж x_n, і розподіл поля $E(x)$ практично має вид прямокутного трикутника. Практично вся напруга на бар'єрі $U_{pn} = \frac{1}{2} E_m x_n$ падає на n-область. В цьому випадку розрахунок основних параметрів бар'єра можна здійснити за допомогою формул (21.17) і (21.18):

$$x = x_n \cong \left(\frac{2\varepsilon\varepsilon_0 U_{pn}}{eN_d}\right)^{1/2}, \qquad (21.20)$$

$$E_m \cong \left(\frac{2eN_d U_{pn}}{\varepsilon\varepsilon_0}\right)^{1/2}. \qquad (21.21)$$

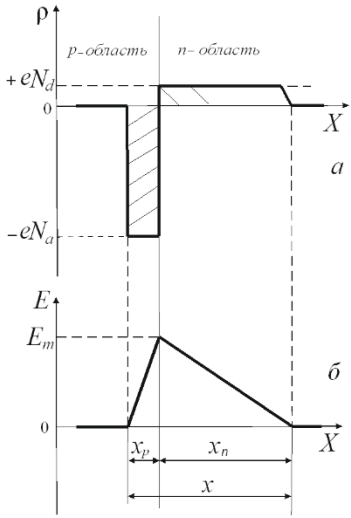


Рис. 21.12. Розподіл об'ємного заряду $\rho(x)$ *(а) і поля* $E(x)$ *(б) в різко несиметричному p–n-переході* ($N_a \gg N_d$)

У кремнієвому потужному високовольтному випрямляючому діоді концентрація донорів в n-області $N_d = 10^{19}$ м$^{-3}$. (Концентрація акцепторів в p-області $N_a \approx 10^{24}$ м$^{-3}$, тобто в 100 000 разів більша і перехід дійсно різко несиметричний). Підставляючи в формули (21.20) і (21.21) значення $U_{pn} = 1{,}1$ В, $\varepsilon = 12{,}5$, $N_d = 10^{19}$ м$^{-3}$, отримаємо $x \approx x_n \approx 12$ мкм, а $E_m \approx 1{,}8 \cdot 10^5$ В/м.

У кремнієвих діодах, які призначенні для випрямлення низьких

напруг, концентрація донорів в *n*-області може бути порядку 10^{23} м$^{-3}$. В такому діоді ширина області об'ємного заряду складає всього $\sim 0{,}12$ мкм, а максимальне поле в *p–n*-переході $\sim 1{,}8 \times \times 10^7$ В/м.

21.3.4. Зворотне зміщення

Порушимо рівновагу між дрейфовим і дифузійним струмами в області *p–n*-переходу. Для цього досить прикласти до діода, який містить *p–n*-перехід, напругу, або як часто говорять, зміщення U_0 від зовнішнього джерела (рис. 21.13). Якщо «плюс» джерела під'єднаний до *n*-області діода, а «мінус» – до *p*-області, то до *p–n*-переходу прикладена зворотня напруга (або зворотне зміщення). Саме такий випадок показаний на рис. 21.13.

Діод має три області: *p*-область, *n*-область і область об'ємного заряду (сам *p–n*-перехід), який знаходиться між ними. По відношенню до зовнішнього джерела U_0 всі ці три області включені послідовно. Область об'ємного заряду збіднена вільними носіями заряду, електронами і дірками. Тому вона має найбільший опір. Значить, практично вся прикладена напруга U_0 падає на область об'ємного заряду.

Звертаємо увагу на те, що поле $\vec{E}$ від зовнішнього джерела U_0 при зворотньому зміщенні на діоді співпадає за напрямком з полем всередині області об'ємного заряду (рис. 21.13). Це означає, що спад напруги від зовнішнього джерела U_0 додається до напруги на переході U_{pn}, яка існувала і до прикладення зовнішнього зміщення. У результаті висота потенціального бар'єра між *p*- і *n*-областями стає $U_{pn} + U_0$, а висота енергетичного бар'єра $e(U_{pn} + + U_0)$ (рис. 21.13, *а*).

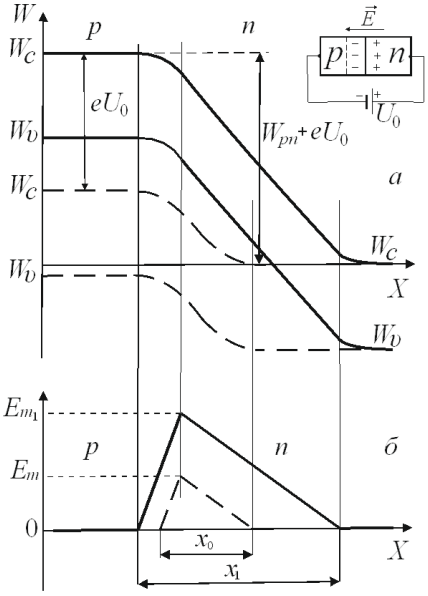


Рис. 21.13. Енергетична діаграма (а) і розподіл електричного поля (б) в p–n-переході. Штрихові лінії – випадок відсутності зовнішньої напруги. Суцільні лінії – до діода прикладена зворотна напруга U_0

Для різко несиметричного переходу ширину бар'єра і максимальну

напруженість електричного поля знайдемо із формул (21.20) та (21.21):

$$x_1 \cong x_{n_1} \approx \left(\frac{2\varepsilon\varepsilon_0(U_{pn}+U_0)}{eN_d}\right)^{1/2}, \quad (21.22)$$

$$E_{m_1} = \left(\frac{2eN_d(U_{pn}+U_0)}{\varepsilon\varepsilon_0}\right)^{1/2}. \quad (21.23)$$

Із формули (21.22) видно, що у разі прикладання до діода зворотної напруги вільні носії заряду витісняються з областей, які прилягають до *p–n*-переходу: область об'ємного заряду стає ширшою (рис. 21.13, *б*). Для високовольтного кремнієвого випрямляючого діода за $U_0 = 1000$ В за формулами (21.22) і (21.23) знайдемо $x = x_{n_1} = 370$ мкм, $E_{m_1} = 5400$ кВ/м.

Зі збільшенням U_0 максимальна напруженість E_m зростає, проте для кожного напівпровідникового матеріалу існує граничне значення напруженості електричного поля E_i, перевищувати яке не слід. У полі, напруженість якого E_i, для Si та GaAs вона досягає значень $\sim 3 \times 10^7$ В/м, електрони й дірки набувають такої великої енергії, що під час зіткнення з атомами кристалічної ґратки напівпровідника здатні створювати нові електрони й дірки. Напруженість E_i називають *напруженістю пробою.* Напругу пробою *p–n*-переходу можна визначити за допомогою формули (21.23), а саме:

$$U_i \approx \frac{\varepsilon\varepsilon_0 E_i^2}{2eN_d}. \quad (21.24)$$

Якщо $N_d \approx 10^{23}$ м$^{-3}$, $E_i = 3 \cdot 10^7$ В/м, $\varepsilon = 12$, то $U_i = 3$ В, а якщо $N_d \approx 10^{20}$ м$^{-3}$, то $U_i = 3000$ В.

Зворотний струм. Коли до *p–n*-переходу зовнішню напругу не прикладено, то струм неосновних носіїв заряду врівноважується зустрічним потоком «гарячих» основних носіїв заряду, але коли до *p–n*-переходу прикладене зворотне зміщення і висота бар'єра W_{pn} збільшиться, то «гарячі» носії вже не можуть подолати бар'єр. На потік неосновних носіїв прикладена напруга ніяк не впливає, тому зворотний струм визначається потоком неосновних носіїв заряду (21.10) і дуже різко залежить від температури. Вище сказане належить до германієвого діода. Для кремнієвого та арсенід-галієвого переходів зворотний струм не насичується, а монотонно зростає пропорційно $U_0^{1/2}$. У цих діодах домінувальну роль під час зворотного зміщення відіграє струм, який генерується в області об'ємного заряду. Деколи його називають просто *генераційним струмом*. Носії, які створюють генераційний струм, виникають, власне, у *p–n*-переході, в області об'ємного заряду, де існує сильне електричне поле. Будь-який носій, електрон або дірка, які виникають у цій області, будуть негайно підхоплені електричним полем і викинуті: електрони

в n-область, дірки – у p-область. Генераційний струм пропорційний ширині бар'єра (області об'ємного заряду) x. Оскільки згідно з формулою (21.22) $x \sim U_0^{1/2}$, то зрозуміло, чому генераційний струм пропорційний $U_0^{1/2}$.

Відношення густини генераційного струму j_g до густини струму насичення j_s (формула (21.10))

$$\frac{j_g}{j_s} = \frac{N \cdot x}{n_i L}, \qquad (21.25)$$

де $N_d = N_a = N$, $D_n = D_p = D$, $L_n = L_p = L$ і час життя $\tau_p = \tau_n = \tau$. Вираз (21.25) дає можливість порівняти відносні величини j_g та j_s для будь-якого p–n-переходу за будь-якої температури та будь-якої прикладеної напруги. Із (21.25) видно, що зі зростанням ширини забороненої зони напівпровідника W_g роль j_g різко, експоненціально зростає, оскільки згідно з (19.44), концентрація n_i експоненціально зменшується. Це пояснює, чому внесок генераційного струму у зворотний струм у кремнієвих, а особливо в арсенід-галієвих переходах значно більший, ніж у германієвих.

21.3.5. Пряме зміщення

На рис. 21.14 зображено p–n-перехід, до якого прикладене пряме зміщення. «Плюс» джерела під'єднаний тепер до p-області діода, «мінус» – до n-області. Поле від зовнішнього джерела U_0 направлене назустріч полю, яке існує всередині p–n-області. Значить, спад напруги від зовнішнього джерела U_0 вираховують з бар'єрної різниці потенціалів U_{pn}, яка існувала до прикладання зовнішнього зміщення.

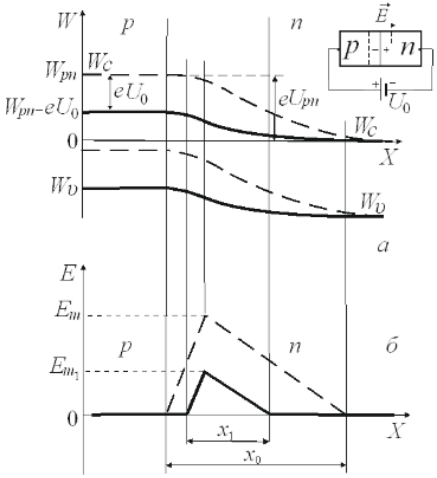


Рис. 21.14. Енергетична діаграма (а) й розподіл напруженості електричного поля (б) в p–n-переході. Штрихові лінії – випадок відсутності зовнішньої напруги. Суцільні лінії – до діода прикладене пряме зміщення

У результаті висота потенціального бар'єра між p- і n-областям стане дорівнювати $U_{pn} - U_0$, а висота енергетичного бар'єра, відповідно, $e(U_{pn} - U_0) = W_{pn} - eU_0$. Якщо зміщення немає, «прориваються» тільки ті електрони з n-області в p-область і дірки із p-області в n-область, енергія яких більша від висоти бар'єра eU_{pn} (формули (21.12) і (21.13)). Струм, який

вони створюють, компенсується дрейфовим струмом неосновних носіїв (формула (21.10)). Якщо висота бар'єра зменшується на величину eU_0, то тепер вже з *n*- у *p*-область здатні «прориватись» усі електрони, енергія яких більша від висоти бар'єра $W_{pn} - eU_0$. Концентрація таких електронів згідно з (21.12) становить

$$n_1 \cong n_n \exp\left\{-\frac{W_{pn} - eU_0}{kT}\right\}.$$

Таким чином, кількість електронів, здатних подолати бар'єр, що знизився, зростає у $\frac{n_1}{n_n} = \exp\left\{\frac{eU_0}{kT}\right\}$ разів, наприклад, за $U_0 = 0{,}1$ В, $T = 300$ К, в $\frac{n_1}{n_n} \cong 6 \cdot 10^{16}$ разів. Струм, який утворюють ці електрони, у стільки ж разівзростає і становить тепер величину

$$j_{n_1} = j_{ns} \exp\left\{\frac{eU_0}{kT}\right\}, \qquad (21.26)$$

де враховано, що без зміщення дрейфовий струм неосновних електронів j_{ns} (формула (21.8)) компенсується струмом «гарячих» електронів (формула (21.12)). Таким чином, електронна складова струму через *p–n*-перехід під час його прямого включення буде:

$$j_n(U_0) = j_{n_1} - j_{ns} = \\ = j_{ns}\left(\exp\left\{\frac{eU_0}{kT}\right\} - 1\right). \qquad (21.27)$$

Аналогічно, діркова складова прямого струму

$$j_p(U_0) = j_{ps}\left(\exp\left\{\frac{eU_0}{kT}\right\} - 1\right). \qquad (21.28)$$

Повний струм через *p–n*-перехід дорівнює сумі електронної й діркової складових і становить

$$j(U_0) = j_n(U_0) + j_p(U_0) = \\ = j_s\left(\exp\left\{\frac{eU_0}{kT}\right\} - 1\right), \qquad (21.29)$$

де $j_s = j_{ns} + j_{ps}$.

Формула (21.29) описує залежність струму через *p–n*-перехід від прикладеної напруги за будь-яких зміщень – і прямих, і зворотних. Тільки за прямої прикладеної напруги величину U_0 треба вважати додатною, а за зворотної – від'ємною.

Оцінимо відношення струму під час прямого зміщення до струму під час зворотного зміщення за допомогою формули (21.29) за кімнатної температури $T = 300$ К ($kT = 0{,}026$ еВ) за $U_0 = 1$ В:

$$\frac{j_{\text{пр.}}}{|j_{\text{зв.}}|} = \frac{\exp\left\{\frac{eU_0}{kT}\right\} - 1}{\left|\exp\left\{-\frac{eU_0}{kT}\right\} - 1\right|} = \\ = \frac{\exp\{38\} - 1}{|\exp\{-38\} - 1|} \approx \exp\{38\} \simeq 10^{16}.$$

Таким чином, струм при прямому зміщенні значно перевищує струм при зворотному зміщенні, тобто *p–n*-перехід має односторонню провідність. Цю властивість *p–n*-переходу в основному і використовують на практиці. На рис. 21.15 показано ВАХ германієвого діода з *p–n*-переходом. Зверніть увагу на різницю в масштабах по осях струму і напруги, які відповідають прямому і

зворотному зміщенню. За $U_0 \simeq -200$ В, як видно з рис. 21.15, струм змінюється через *p–n*-перехід, а напруга на діоді стала. У такому режимі пробою працює стабілітрон, який стабілізує напругу на *p–n*-переході.

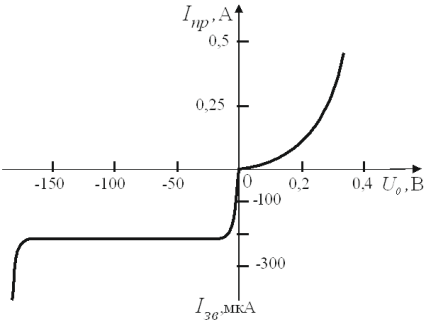


Рис. 21.15. Вольт-амперна характеристика германієвого p−n-переходу

Примітка. 1. Під час отримання формули (21.29) вважалось, що вся прикладена до діода напруга U_0 падає на *p–n*-переході, тобто що опір *p–n*-переходу набагато більший, ніж усі інші послідовно ввімкнені з ним опори: опір *n*-області, *p*-області, опір контактів до діода і т.п. За зворотної напруги таке припущення справедливе. Опір бар'єра (*p–n*-переходу) є великий вже за нульового зміщення. Під час прямого зміщення справа дещо інша. Зі збільшенням прямої напруги опір бар'єра різко зменшується! Отже, не вся прикладена напруга U_0 падає на *p–n*-переході. Значна частина напруги падає на послідовно ввімкнених з переходом контактах, а також *n*- і *p*-областях, тому висота бар'єра *p–n*-переходу знизиться не на eU_o, а на дещо меншу величину. В результаті ситуація, коли

$$eU_0 > W_{pn}$$

$$\left(n_1 \cong n_n \exp\left\{-\frac{W_{pn} - eU_0}{kT}\right\} > n_n\right)$$

стає неможливою. Найбільше, чого можна досягти, збільшуючи пряму напругу на діоді, – це дуже сильно зменшити висоту бар'єра (як деколи говорять – «спрямити» бар'єр). Але досягнутий ефект вимагає, щоб густина струму через діод складала десятки й навіть сотні тисяч амперів на сантиметр квадратний (А/см^2).

2. Увипадку різко несиметричного переходу (сильно легована *p*-область) за будь-яких зміщень основний внесок у загальний струм в області бар'єра належить діркам. Цю область діода називають емітером (від лат. emitto – випускаю). Вона здатна випускати (інжектувати) у слаболеговану *n*-область таку кількість дірок (неосновних носіїв для *n*-області), що їх концентрація може значно перевищувати вихідну концентрацію основних носіїв (електронів в *n*-області).

Слаболеговану область діода називають базою (від грец. basis – основа). Саме у слаболегованій області переходу, базі, переважно розміщується

область об'ємного заряду при зворотному зміщенні (рис. 21.13). Параметри бази – її ширина і рівень легування – визначають напругу пробою *p–n*-структури (формули (21.22), (21.23), (21.24)). Відповідно до чинної термінології на рис. 21.15 аналізують роботу діода з p-емітером та n-базою.

21.4. Ємність електронно-діркового переходу

Якщо до *p–n*-переходу прикладено зворотну напругу U_0, то щодо змінної напруги зворотний зміщений *p–n*-перехід поводить себе як конденсатор. Ємнісний опір *p–n*-переходу визначають за формулою (9.77) $X_C = 1 / (\omega C)$, де ω – циклічна частота напруги; C – ємність *p–n*-переходу. Ємність $p-n$-переходу можна визначити за формулою ємності плоского конденсатора

$$C = \varepsilon\varepsilon_0 S/x, \qquad (21.30)$$

де x – ширина області об'ємного заряду; S – площа переходу. Оскільки x визначають за формулою (21.22), то можна визначати ємність переходу за будь-якої прикладеної напруги і різних рівнів легування. Із розд. 8 (підрозд. 8.2) відомо, що конденсатор проводить змінний електричний струм тому, що лінії струму провідності всередині конденсатора переходять в лінії струму зміщення (змінне електричне поле) і, таким чином, електричне коло стає замкненим. Пояснення проходження змінного струму через конденсатор може бути й таким. Коли до конденсатора прикладено змінну напругу, то він протягом певного періоду накопичує електричний заряд, а потім протягом іншої частини періоду віддає його в зовнішнє коло. Періодичне заряджання і розряджання конденсатора підтримує в колі змінний струм.

Тепер розглянемо зворотний зміщений *p–n*-перехід, до якого одночасно з постійною зворотною напругою U_0 прикладена ще й невелика змінна напруга, амплітуда якої ΔU_0 (рис. 21.16). Суцільна крива на рис. 21.16 відповідає розподілу об'ємного заряду ρ в *p–n*-переході за зворотної напруги U_0. Штриховою лінією показано розподіл об'ємного заряду, ρ, коли до діода прикладено напругу $U_0 + \Delta U_0$.

Напруга на діоді стала більшою – область об'ємного заряду розширилась (формула (21.22)). Вільні носії – електрони в *n*-області й дірки в *p*-області виштовхуються в зовнішнє коло з області об'ємного заряду, яка звільняється від носіїв. В *n*-області зі збільшенням напруги від U_0 до $U_0 + \Delta U_0$ в зовнішнє коло виштовхуються всі електрони, які за напруги U_0 містяться на ділянці між суцільною і штриховою лініями. В *p*-

області так само з ділянки між суцільною і штриховою кривими витісняються дірки.

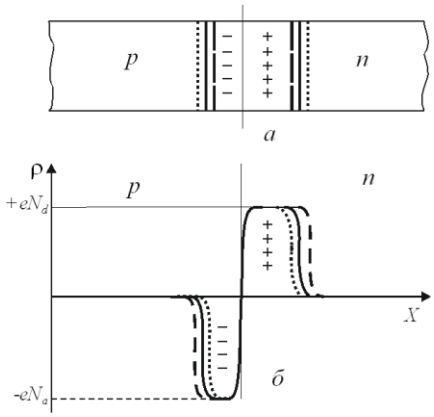


Рис. 21.16. Розподіл об'ємного заряду в p–n-переході: суцільна лінія – до діода прикладено зворотну напругу U_0;штрихова лінія – зворотна напруга дорівнює $U_0 + \Delta U_0$; лінія точками – $U_0 - \Delta U_0$

Точковою кривою показано розподіл об'ємного заряду, коли до діода прикладена напруга $U_0 - \Delta U_0$. Напруга на діоді стала меншою – ширина області об'ємного заряду зменшилась, а це означає, що із зовнішнього кола забираються вільні носії, які заповнюють тепер ділянки діода, де їх не було за напруги U_0. Якщо напруга зменшилась від U_0 до $U_0 - \Delta U_0$, то вільні носії, які відбираються із зовнішнього кола, заповнюють області між суцільною і точковими кривими.

Таким чином, зрозуміло, що *p–n*-перехід має здатність у певну частину періоду накопичувати вільні носії заряду, а в іншу – віддавати їх у зовнішнє коло, тобто має ємність. Цей тип ємності називають бар'єрною.

Бар'єрну ємність визначимо як $C = \Delta q \,/\, \Delta U_0$, де Δq є величиною заряду, яка накопичується або віддається *p–n*-переходом під час зміни напруги на величину ΔU_0. Для різко несиметричного переходу вся прикладена напруга падає на слаболеговану область і повну ширину об'ємного заряду практично визначають за формулою (21.22). Зі збільшенням напруги на величину ΔU_0 ширина області об'ємного заряду збільшується на величину Δx_n. Значення Δx_n можна визначити з формули (21.22). При цьому у зовнішнє коло буде витіснений заряд електронів на одиницю площі поперечного перерізу $\Delta q = eN_d \cdot \Delta x_n$ (такий самий заряд дірок буде витіснений з *p*-області). Відношення $\Delta q / \Delta U_0$ визначить величину ємності *p–n*-переходу на одиницю площі C_0. Тоді

$$C_0 = \frac{dq_0}{dU_0} = eN_d \frac{dx_n}{dU_0} = eN_d \cdot \frac{1}{2}\left(\frac{2\varepsilon\varepsilon_0}{eN_d}\right)^{1/2} \times$$

$$\times \frac{1}{\left(U_{pn} + U_0\right)^{1/2}} = \left(\frac{e\varepsilon\varepsilon_0 N_d}{2\left(U_{pn} + U_0\right)}\right)^{1/2} = \frac{\varepsilon\varepsilon_0}{x}, \quad (21.31)$$

де U_0 – напруга на *p–n*-переході; x – ширина області об'ємного заряду, яку визначають формулою (21.22). Із

формули (21.31) випливають два важливих висновки. *По-перше*, бар'єрна ємність *p–n*-переходу залежить від напруги. Чим більшим є прикладене до діода зворотне зміщення U_0, тим меншою буде ємність. Конденсатори, ємність яких залежить від напруги, називають нелінійними, тому часто говорять, що *p–n*-перехід є нелінійною ємністю. *По-друге*, за однакової зворотної напруги на переході бар'єрна ємність тим більша, чим вищим є рівень легування: $C \sim N_d^{1/2}$. Оцінимо величину бар'єрної ємності для високовольтного випрямляючого діода з кремнію, для якого $S \sim 10 - 30$ см^2. За $S = 10$ см^2, $N_d = 10^{19}$ м$^{-3}$ і без зовнішньої напруги на діоді ємність C визначимо як $C = S\left(\frac{e\varepsilon\varepsilon_0 N_d}{2U_{pn}}\right)^{1/2} \approx 1 \times$ $\times 10^{-8}$ Ф $= 10^4$пФ. Навіть за $U_0 =$ $= 1000$ В ємність діода буде значною величиною: $C \approx 300$ пФ.

Порівняємо ємнісний опір діода з його активним опором, який визначають зворотним струмом. Ці опори можна вважати з'єднаними паралельно. За кімнатної температури й напруги на діоді $U_0 = 100$ В типове значення зворотного струму $I_{зв}$ для кремнієвого діода такого типу становить ~0,5 мА. Таким чином, активний опір $R = U_0/I \simeq 2 \cdot 10^5$ Ом. Ємність за $U_0 = 100$ В за формулою (21.31) становитиме:

$$C = SC_0 = S\left(\frac{e\varepsilon\varepsilon_0 N_d}{2U_0}\right)^{1/2} \sim 10^{-9}\text{Ф}.$$

Якщо частота $\nu \geq 800$ Гц, то $X_C = 1/(C \cdot 2\pi\nu) = 2 \cdot 10^5$ Ом $\sim R$, тобто за частоти напруги $\nu > 800$ Гц опір діода буде визначатись не тільки зворотним струмом, а і ємністю структури. За промислової частоти $\nu = 50$ Гц $X_C \sim 3 \times$ $\times \cdot 10^6$ Ом опір R визначає в основному опір діода. Утім, є діоди, призначені для роботи в колах високих частот. Для таких діодів наявність бар'єрної ємності відіграє дуже важливу роль. Деколи ємнісний струм, обумовлений бар'єрною ємністю, заважає діоду виконувати ту роботу, для якої він призначений. У цьому випадку величину C намагаються зробити по можливості меншою, щоб збільшити опір. Цього досягають за допомогою зменшення площі діодів, які призначені для роботи у колах НВЧ. Площа S для таких діодів становить 10^{-4} ... 10^{-6} см^2 (10^{-2} ... 10^{-4} мм^2). Величина бар'єрної ємності таких *p–n*-структур дорівнює десятим, а деколи і сотим часткам пікофарада. Утім, є й такі діоди, дія яких прямо ґрунтується на властивостях бар'єрної ємності переходу. Їх використовують як конденсатори, ємністю яких можна керувати прикладеною напругою в колах високої і надвисокої частоти.

21.4.1. Варикапи

Назва такого приладу, як варикап (від англ. variable capacitor – змінний конденсатор), дуже точно характеризує його призначення. Варикапи використовують як конденсатори змінної ємності, величина якої змінюється залежно від прикладеної напруги. Чим більша напруга, тим менша ємність. Кожний, хто налаштовує радіоприймач на потрібну станцію, з точки зору радіотехніки зайнятий тим, що, змінюючи ємність змінного конденсатора, налаштовує вхідний контур радіоприймача на потрібну частоту. Одним із перших застосувань варикапів була заміна механічного змінного конденсатора в радіоприймачі.

Напівпровідникові ємності, керовані напругою, лежать в основі надчутливих перетворювачів постійної напруги у змінну. З їх допомогою здійснюється приймання й підсилення надслабих сигналів, зокрема і сигналів з космосу. Їх використовують для генерації сигналів з частотою в сотні мільярдів герців. У параметричних підсилювачах в *LC*-контурі, що налаштований на частоту сигналу, який підсилюється, як ємність C включають варикап. Такий параметр варикапа, як ємність, два рази за період коливань сигналу, який підсилюється, у потрібній фазі змінюється зовнішнім джерелом коливань. У результаті досягають підсилення, яке називають параметричним. Основна перевага параметричних підсилювачів – їх рекордно висока чутливість. Головним параметром варикапа є коефіцієнт зміни ємності K_C, який визначають як відношення максимальної ємності варикапа C_{max} до мінімальної C_{min}: $K_C = C_{max} / C_{min}$. Оскільки ємність варикапа залежить від прикладеної до нього напруги, то зрозуміло, що величина K_C залежить від діапазону, в якому можна змінювати зворотнє зміщення на варикапі. Мінімальну зворотну напругу U_0 вважатимемо нульовою. При цьому напруга на p–n-переході мінімальна й дорівнює U_{pn} і згідно з формулою (21.21)

$$C_{max} = S\left(\frac{e\varepsilon\varepsilon_0 N_d}{2U_{pn}}\right)^{1/2}. \quad (21.32)$$

Максимальна напруга на p–n-переході обмежується процесом ударної іонізації, її визначають за формулою (21.24). Враховуючи, що $U_i \gg U_{pn}$, з (21.31) отримаємо значення мінімальної ємності варикапа:

$$C_{min} = S\left(\frac{e\varepsilon\varepsilon_0 N_d}{2U_i}\right)^{1/2} = S\frac{eN_d}{E_i}. \quad (21.33)$$

Тоді коефіцієнт зміни ємності K_C будемо визначати як

$$K_C = \frac{C_{max}}{C_{min}} = \left(\frac{\varepsilon\varepsilon_0 E_i^2}{2eU_{pn}N_d}\right)^{1/2}. \quad (21.34)$$

Із (21.34) видно, що чим менша концентрація домішки у слаболегованій області, тим більша величина K_C. На практиці значення N_d беруть з діапазону 10^{21} ... 10^{22} м$^{-3}$. Величина K_C в реальних варикапах лежить у межах від ~2 до 15.

21.5. Фотодіоди і світлодіоди

В основі роботи будь-якого фотоелектричного пристрою лежить елемент, електричні параметри якого – опір, струм або напруга – змінюються під дією світла. Одним з найбільш поширених елементів такого типу є фотодіод – *p*–*n*-перехід, зміщений у зворотному напрямку.

Якщо на напівпровідник падає світло, енергія фотона якого $W_Ф$ більша, ніж ширина забороненої зони W_g, то кожний поглинутий фотон породжує у напівпровіднику електронно-діркову пару (див. рис. 19.32, *а*). Якщо електрон і дірка виникли під дією світла в області об'ємного заряду зворотно зміщеного *p*–*n*-переходу, то будь-який носій, електрон чи дірка, які народжені в області об'ємного заряду, будуть негайно підхоплені електричним полем і викинуті: електрон – в *n*-область, дірка – в *p*-область. У пітьмі, без освітлення, генераційний струм *p*–*n*-переходу (темновий струм) визначається кількістю електронів і дірок, які є в області об'ємного заряду внаслідок теплової генерації. Під час освітлення зворотно зміщеного переходу світлом з енергією квантів $W_Ф > W_g$ струм через перехід збільшується. Фотострум, який виникає під дією світла, буде у стільки раз більшим від темнового струму, у скільки разів кількість носіїв, які створюються в області об'ємного заряду світлом, буде більшим, ніж кількість електронів і дірок, які виникають внаслідок теплової генерації. Фотострум, підсилений і перетворений відповідною електронною схемою, приводить у дію захисні, охоронні, сигнальні, блокувальні й подібні пристрої.

На рис. 21.17, *а*, показано типову конструкцію фотодіода. Із рисунка видно, що *p*–*n*-перехід зміщений у зворотному напрямку. Верхній контакт до діода виконаний у формі кільця (рис. 21.17, *б*). Область *p*–*n*-переходу всередині кільця, тобто вікно, доступна для світла. У фотодіоді площина *p*–*n*-переходу розміщується близько до поверхні кристала: це потрібно для того, щоб світло, яке падає на кристал, не встигло поглинутись на шляху до області об'ємного заряду. Зазвичай площина *p*–*n*-переходу міститься у фотодіодах на відстані в декілька десятих часток мікрометра від поверхні. Такий *p*–*n*-

перехід можна виготовити за допомогою дифузії або іонної імплантації.

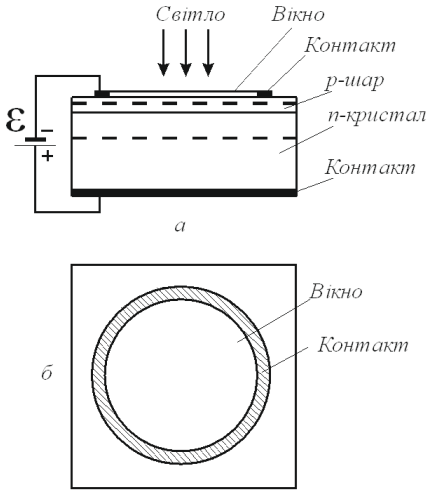


Рис. 21.17. Конструкція типового фотодіода

Величина зворотної напруги U_0, яка прикладається до p–n-переходу, становить зазвичай 10... 30 В. Зі зростанням зворотної напруги, як відомо із формули (21.22), область об'ємного заряду розширюється. Величину U_0 і концентрацію домішки у слаболегованій n-області кристала N_d вибирають таким чином, щоб майже весь світловий потік, який падає на поверхню світлодіода, поглинався в області об'ємного заряду і робив внесок у фотострум. Найчастіше потрібна для цього товщина збідненої області x_n становить декілька мікрометрів.

Найбільше поширення мають фотодіоди на основі Ge та Si, однак розроблені й використовуються також фотодіоди із GaAs, InAs, InSb, GaP, трикомпонентних напівпровідникових сполук GaInAs, GaAlAs, чотирикомпонентної сполуки GaInAsP.

Одна з найважливіших причин для розроблення фотодіодів на основі різних напівпровідникових сполук полягає в тому, що фотодіод, який виготовлений із певного матеріалу, здатний ефективно реагувати на світло з довжиною хвилі, що лежить тільки у певному діапазоні. Цю властивість фотодіодів називають *селективністю* (від лат. selection – вибір, відбір), що показано на рис. 21.18. На рис. 21.18 для фотодіода, який виготовлений з напівпровідника з шириною забороненої зони W_g, показано якісну залежність фотоструму від енергії падаючих фотонів $W_\Phi = h\nu$. Якщо $W_\Phi < W_g$, енергії фотонів недостатньо для створення в матеріалі електронно-діркових пар і фотоструму тому не виникає. За $W_\Phi > W_g$ фотострум спочатку швидко збільшується із зростанням W_Φ, а потім починає зменшуватись. Справа в тому, що зі збільшенням W_Φ різко зростає коефіцієнт поглинання світла в матеріалі, тобто падаючий світловий потік зі зростанням енергії фотонів поглинається все ближче і ближче до поверхні фотоприймального вікна (рис. 21.17, *б*). Як відомо з підрозд.

21.2, на поверхні завжди існує велика кількість дефектів і домішок, які створюють ефективні центри рекомбінації, тому електрони й дірки, які виникають під дією світла біля поверхні, швидко рекомбінують і не дають внеску у фотострум.

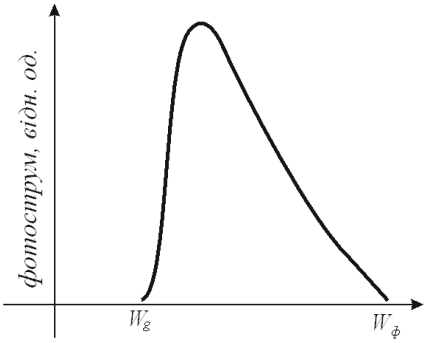


Рис. 21.18. Якісна залежність фотоструму від енергії W_Φ фотонів світла, яке падає на фотодіод

Таким чином, область максимальної фоточутливості діода визначають величиною W_g і відповідною умовою $W_\Phi \geq W_g$. Напівпровідники з невеликою шириною забороненої зони, або *вузькозонні напівпровідники*, використовують для створення фотодіодів, які чутливі до світла з малою енергією квантів, тобто з великою довжиною хвилі. Напівпровідники з великим значенням W_g (*широкозонні напівпровідники*) використовують для створення приймачів ультрафіолетового світла.

Важливою властивістю фотодіодів є швидкодія. Швидкодію фотоелектричних приладів оцінюють за тим, як швидко спадає фотострум через прилад після вимикання світла. Якщо різко перекрити світловий потік, який падає на приймальне вікно фотодіода, то фотострум зникне щойно електрони й дірки, збуджені світлом, будуть винесені полем з області об'ємного заряду.

Як відомо, типове значення напруженості електричного поля у зворотно зміщеному *p–n*-переході велике, воно становить $\sim 10^6 \ldots 10^7$ В/м. У таких сильних полях дрейфова швидкість електронів і дірок $v_{\text{др}}$ не залежить від напруженості поля (див. підрозд. 19.12) і дорівнює $\sim 10^5$ м/с (див. рис. 19.40). Шлях, який мають пройти електрон і дірка, перш ніж вони будуть винесені полем із *p–n*-переходу, дорівнює, очевидно, ширині області об'ємного заряду x. Для типових фотодіодів значення x складає декілька мікрометрів ($(1 \ldots 10) \cdot 10^{-6}$ м).

Таким чином, час, який визначає швидкодію фотодіода τ, приблизно становить $\tau \sim x/v_{\text{др}} \sim 10^{-10} \ldots 10^{-11}$ с. Фотодіод здатний реагувати на зміни інтенсивності світла за стомільярдні частки секунди. Висока швидкодія дозволяє використовувати фотодіоди у швидкісних системах передачі та

обробки інформації волоконно-оптичного зв'язку. Приймачами швидкомодульованих світлових сигналів у таких системах служать саме фотодіоди. Фотоелектричні пристрої дозволяють вводити в ЕВМ тисячі чисел і команд за секунду.

Світлодіод (світловипромінювальний діод) – напівпровідниковий прилад, який перетворює електричну енергію в енергію оптичного випромінювання на основі явища інжекційної електролюмінесценції, що відбувається в напівпровідниковому кристалі з *p–n*-переходом або гетеропереходом чи контактом «метал – напівпровідник».

Будемо розглядати напівпровідниковий *p–n*-перехід за умови прямому зміщенні. Можна сказати, що будь-який прямо зміщений *p–n*-перехід є світлодіодом. Основні носії (електрони й дірки), які потрапляють з емітера (інжектуються) у базу, рекомбінують. Частина цих носіїв рекомбінують з випромінюванням фотонів. Якась частина фотонів, не зазнавши поглинання в діоді, вилітають назовні. Для виготовлення світлодіодів придатні матеріали з достатньо великою ймовірністю випромінювальної рекомбінації. Германій і кремній цим вимогами не відповідають – в них переважна частина електронів і дірок рекомбінує без випромінювання фотона. В арсеніді галію GaAs і трикомпонентних напівпровідникових сполуках на його основі (GaAsP і GaAlAs) ймовірність випромінювальної рекомбінації дуже висока (близька до одиниці), тому більшу частину світлодіодів випускають на основі цих матеріалів.

Колір, тобто довжина хвилі світла, яку випромінює світлодіод, визначається енергією фотона W_Φ, який випромінюється під час рекомбінації. У більшості випадків ця енергія близька до ширини забороненої зони напівпровідника W_g. Так, довжина хвилі випромінювання арсенід-галієвого $(W_g \simeq 1{,}4\ \text{еВ})$ світлодіода $\lambda = hc/W_g = 1{,}24/1{,}4$ мкм = $= 0{,}89$ мкм. Такого випромінювання людське око не бачить – воно відповідає інфрачервоній області спектра. Приймачами такого інфрачервоного випромінювання є кремнієві фотодіоди, для яких енергія фотона $W_\Phi \approx 1{,}4$ еВ приблизно відповідає максимуму їх спектральної чутливості (рис. 21.18).

Додавання у ґратку GaAs атомів фосфору P або алюмінію Al приводить до збільшення ширини забороненої зони W_g у сполуці $GaAs_{1-x}P_x$. Зі збільшенням у напівпровідниковій сполуці частки відносного вмісту фосфору x, ширина забороненої зони напівпровідника

збільшується від 1,4 еВ (GaAs, $x = 0$) до 2,26 еВ (GaP, $x = 1$). Виготовляючи твердий розчин GaAsP з різним вмістом P, отримують світлодіоди, які випромінюють червоне світло

$$W_Ф = (1{,}6 \ldots 1{,}9)\ \text{еВ},\ \lambda_ч = $$
$$= (0{,}653 \ldots 0{,}775)\ \text{мкм}.$$

Отримати таким способом жовті або зелені світлодіоди не вдається, хоч енергія квантів, що відповідає цим кольорам, менша, ніж ширина заборненої зони чистого GaP ($W_g = 2{,}26$ еВ, $\lambda = 0{,}549$ мкм відповідає зеленій області спектра). Справа в тому, що за $x \geq 0{,}45$ ймовірність випромінювальної рекомбінації в GaAsP різко падає і світлодіоди стають малоефективними. Для збільшення ефективності свічення у твердий розчин GaAsP або в чистий GaP додають спеціальні домішки (азот N, цинк Zn, кисень O), які збільшують ймовірність випромінювальної рекомбінації. Сьогодні для створення «різнокольорових» світлодіодів використовують широкий вибір напівпровідникових матеріалів. З використанням GaAlAs отримують інфрачервоні й червоні джерела світла. На основі сполук GaAsP та GaP виготовляють світлодіоди з кольором свічення від червоного до зеленого. Чотирикомпонентна напівпровідникова сполука GaInAsP дозволяє створювати інфрачервоні світлодіоди з довжиною хвилі $\lambda \sim 1 \ldots 1{,}3$ мкм, які є найкращими джерелами для волоконно-оптичних ліній зв'язку, а на основі карбіду кремнію SiC отримують світлодіоди всіх основних кольорів – червоного, зеленого і синього.

На сьогодні головна сфера використання світлодіодів – це різні індикатори. У наш час інформаційного вибуху поява зручних, універсальних, дешевих та ефективних приладів для індикації інформації, якими є світлодіоди, дозволяє вирішувати багато важливих проблем.

Світлодіоди випромінюють некогерентне випромінювання з вузьким спектром. Довжина хвилі випромінювання залежить від напівпровідникового матеріалу і його легування. Яскравість випромінювання більшості світлодіодів перебуває на рівні $10^{-3} \ldots 10^{-5}$ кд / м^2 **(кандела на квадратний метр – одиниця СІ яскравості плоскої поверхні, площа якої 1 м^2, що світиться у перпендикулярному до неї напрямку, за сили світла 1 кд. Кандела (від лат. c a n d e l a – свічка) – одиниця сили світла в СІ).**

Коефіцієнт корисної дії світлодіода видимого випромінювання становить від 0,01 % до декількох відсотків. У світлодіодів інфрачервоного випромінювання з метою зниження витрат на повне внутрішнє відбивання і поглинання в тілі кристала для

останнього вибирають напівсферичну форму, а для покращення характеристик направленості випромінювання світлодіод розміщують у параболічному або конусному відбивачі. Коефіцієнт корисної дії світлодіода з напівсферичною формою кристала досягає 40 %.

Сонячні батареї. За кожну секунду Земля отримує від Сонця у вигляді випромінювання $\sim 3 \cdot 10^{17}$ Дж енергії, що у 30 000 разів більше енергії, яка виробляється за цей час всіма електростанціями світу. Способи перетворення сонячної енергії в електричну досліджують дуже давно. Перша пропозиція з використання для цієї мети напівпровідників була у 1884 р., коли спробували виготовити сонячний елемент на основі сполуки селену із золотом. Вже більше 100 років триває наполеглива робота зі створення і вдосконалення напівпровідникових перетворювачів сонячної енергії.

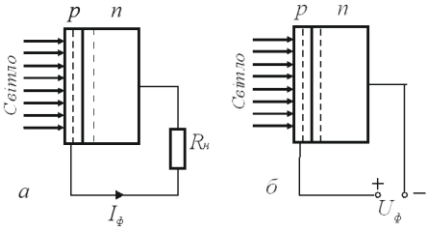


Рис. 21.19. Виникнення фотоструму (а) і фото-е.р.с. (б) у р–n-переході під дією світла

Сонячний елемент – це той самий *р–n*-перехід, і робота його дуже подібна до роботи вже відомого фотодіода. Утім, на відміну від фотодіода, зовнішнього зміщення до сонячного елемента не прикладається.

Повернімось до рис. 21.17 й уявімо, що зовнішнього зовнішнього зміщення до діода не прикладено, ніякого джерела у зовнішньому колі немає, а нижній (суцільний) і верхній (кільцевий) контакти просто з'єднані один з одним через опір навантаження $R_н$ (рис. 21.19, *а*). Як відомо, і без зміщення між *р*- і *n*-областями існує потенціальний бар'єр. Біля *р–n*-переходу є область об'ємного заряду (рис. 21.12, *а*), в якому існує електричне поле (рис. 21.12, *б*). Це означає, що і без зовнішнього зміщення електронно-діркові пари, які виникатимуть під дією світла в області об'ємного заряду, будуть розділятися полем: дірки будуть викидатись в *р*-область, електрони – в *n*-область. Такий рух дірок і електронів викликає у зовнішньому колі фотострум, який буде направлений від *р*-області до *n*-області. Фотострум $I_ф$, який протікає через опір навантаження $R_н$ (цим опором може бути нагрівальний елемент, двигун, акумулятор, який заряджається, і под.), здійснює корисну роботу. Енергія світла перетворюється в електричну енергію. Якщо розімкнути провід, який з'єднує *р*- і *n*-області діода, то струм протікати не буде. Під час неперервного

падіння світла на вікно сонячного елемента в матеріалі неперервно створюються електронно-діркові пари, які розділяються полем *p–n*-переходу. Дірки викидаються в *p*-область і створюють там надлишковий позитивний заряд. Електрони переміщуються в *n*-область і створюють негативний заряд. Між *p*- і *n*-областями елемента виникає різниця потенціалів – фото-е.р.с. U_ϕ (рис. 21.19, *б*). Може видатись, що якщо *p–n*-перехід освітлювати достатньо довго, то величина фото-е.р.с. може стати досить великою. Справді, чим довше світло падає на перехід, тим більше дірок накопичується в *p*-, а електронів – в *n*-області, і, як наслідок, тим більша має бути різниця потенціалів між *p*- і *n*-парами переходу. Насправді такого не відбувається. Полярність фото-е.р.с – «плюс» з боку *p*- і «мінус» з боку *n*-області – відповідають прямому зміщенню на *p–n*-переході (рис. 21.14). Пряме зміщення на переході знижує висоту бар'єра й полегшує відтік дірок з *p*-області в *n*-область, а електронів – з *n*-області в *p*-область (рис. 21.14, *а*, *б*). У результаті дії двох протилежних механізмів – накопичення носіїв під дією світла і їхній відтік через зниження висоти бар'єра – при кожному значенні інтенсивності світла, яке падає на елемент, у стаціонарному стані встановлюється певна величина фото-е.р.с. U_ϕ. Зі збільшенням інтенсивності світла U_ϕ росте дуже слабо, пропорційно логарифму інтенсивності. За дуже великої інтенсивності світла, коли під дією фото-е.р.с. бар'єр стає фактично «спрямленим», величина U_ϕ дорівнює висоті бар'єра неосвітленого переходу:

$$U_{\phi max} \simeq U_{pn} \simeq W_g \,/\, e.$$

Коефіцієнт корисної дії сучасних сонячних елементів виріс до 10 … 20 %. Головна складність на шляху до збільшення ефективності сонячних елементів у тому, що сонячне світло складається з фотонів різних енергій. До того ж, для кожного сонячного елемента, з якого б напівпровідника він не був виготовлений, існує крива спектральної чутливості (рис. 21.18). Із кривої легко зрозуміти, що ефективно перетворюються у фотострум тільки кванти світла з енергією $W_\Phi \approx W_g$. Якщо вибрати для виготовлення сонячного елемента напівпровідник з великою шириною забороненої зони W_g, то фактично весь спектр сонячного світла пройде крізь такий елемент, не створивши електронно-діркових пар ($W_\Phi < W_g$). Якщо, навпаки, виготовити сонячний елемент з матеріалу з малим значенням W_g, то майже всі фотони, що містяться в сонячному світлі, будуть мати енергію $W_\Phi \gg W_g$. Із

рис. 21.18 видно, що такі фотони роблять дуже малий внесок у фотострум. Вони поглинаються біля самої поверхні елемента, далеко від *p–n*-переходу, і генеровані ними електронно-діркові пари швидко «гинуть» через поверхневу рекомбінацію. Розрахунки показують, що максимальну ефективність перетворення енергії сонячного світла в електричну енергію повинні мати елементи, які виготовлені з напівпровідника з $W_g \approx$ $\approx 1{,}5$ еВ ($\lambda = \frac{1{,}24}{1{,}5}$ мкм $\approx 0{,}83$ мкм). Біля цієї оптимальної величини W_g перебуває значення забороненої зони добре вивчених та освоєних у промисловості матеріалів Si та GaAs. Із них, головним чином, і виготовляють сонячні елементи.

Типові значення фото-е.р.с. сонячного елемента становлять ~0,5 В, максимальна сила фотоструму $I_ф \sim 50$ мА за робочої площі елемента $S \sim 2$ см^2. Для отримання більших напруг (10 ... 20 В) елементи з’єднують послідовно, а для отримання робочих струмів – паралельно. Так отримують сонячні батареї.

21.6. Інші види контактів

Важливим елементом більшості напівпровідникових приладів є контакт металу і напівпровідника. Історично цей контакт був основою одного з перших напівпровідникових випрямляючих діодів, який на сьогодні не має широкого застосування. Утім, завдяки досягненням сучасних технологій інтерес до такого типу контактів все більше зростає. Це стосується насамперед випрямляючого контакту «метал – напівпровідник», який називають **бар’єром Шотткі**.

Бар’єр Шотткі утворюється біля межі поділу металу і напівпровідника (рис. 21.20). Якщо робота виходу електронів з металу $Φ_M$ більша, ніж із напівпровідника $Φ_n$, то рівноважний стан встановлюється за контактної різниці потенціалів $U_к = (Φ_M - Φ_n.) / e$, яка виникає у результаті переходу електронів із напівпровідника в метал. Контактне поле $E_к$, направлене так, як показано на рис. 21.20, *а*, і фактично не буде проникати у метал. Воно локалізується в деякому приповерхневому шарі напівпровідника, товщина якого може бути визначена, як і в електронно-дірковому переході, за формулою (21.20). Цей шар збіднюється основними носіями заряду, і просторовий заряд в ньому створюється насамперед позитивно іонізованими донорними атомами. Збіднений шар, який називають деколи запірним, забезпечує контакту нелінійні властивості.

Під час підключення зовнішньої напруги U_0 у прямому напрямку (рис. 21.20, *б*) поле джерела послаблює контактне поле і зменшує висоту

потенціального бар’єра на величину eU_0. Електрони, основні носії *n*-напівпровідника, переходячи в метал, створюють прямий струм тим більший, чим нижчий потенціальний бар’єр. Підвищення потенціального бар’єра за зворотної напруги (рис. 21.20, *в*) зменшує прямий струм, і через контакт буде протікати зворотний струм, який обумовлений рухом термічно збуджених електронів металу у прискорювальному полі контактної області.

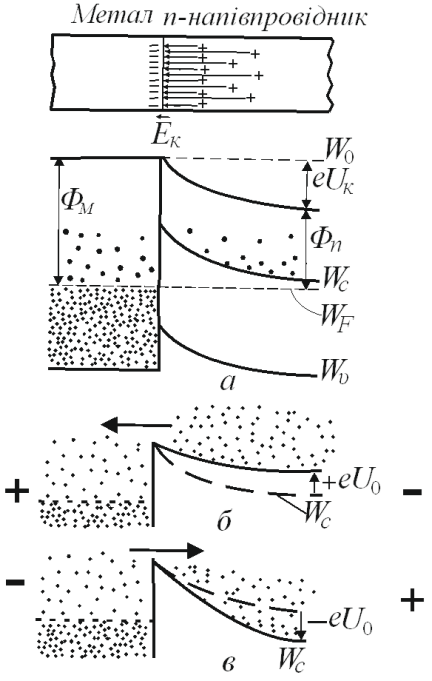


Рис. 21.20. Енергетичні діаграми біля контакту металу з напівпровідником n-типу $(\Phi_M > \Phi_n)$ *за рівноваги (а). за прямого (б) і зворотного (в) зміщення;* W_0 *– енергія вакууму*

Таким чином, у створенні струму беруть участь тільки електрони і перехід їх із напівпровідника в метал не супроводжується явищем дифузії і рекомбінації. Надлишковий заряд електронів фактично миттєво розподіляється в об’ємі, створюючи дрейфовий струм. Це і забезпечує відсутність у контакті «метал – напівпровідник» бар’єрної ємності і, відповідно, його високу швидкодію. Діоди, створені на основі такого контакту (діоди Шотткі), можуть працювати на частотах до 100 ГГц. Вольт-амперна характеристика контакту «метал – напівпровідник» майже така сама, як і для електронно-діркового переходу. Утім, на її вигляд суттєво впливає форма потенціального бар’єра біля поверхні металу. Сили дзеркального відображення згладжують бар’єр, і зовнішня напруга викликає ефект, аналогічний ефекту Шотткі при емісії електронів у вакуум (див. розд. 20). Зовсім інші властивості має контакт металу й електронного напівпровідника, якщо $\Phi_M < \Phi_n$ (рис. 21.21). У цьому випадку контактна різниця потенціалів має протилежний напрямок і електричне поле викликає збагачення контактного шару напівпровідника електронами, тому такий контакт має малий опір, а його ВАХ майже лінійна. Контакти з лінійною

характеристикою називають омічними, їх широко використовують для створення виводів напівпровідникових приладів.

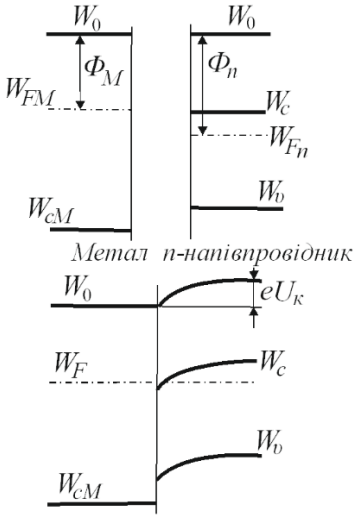


Рис. 21.21. Енергетична діаграма контакту «метал – напівпровідник» за $\Phi_M < \Phi_n$. Омічний контакт

Аналогічні властивості проявляє і контакт типу n^+n (знак «+» означає високий ступінь легування напівпровідника домішкою). Оскільки зі збільшенням концентрації донорної домішки рівень Фермі розміщується ближче до зони провідності і зменшується робота виходу, а за концентрацією електронів n^+ напівпровідник наближається до металів, то енергетична діаграма (рис .21.22) і електричні властивості n^+n контакту майже не відрізняються від омічного контакту «метал – напівпровідник».

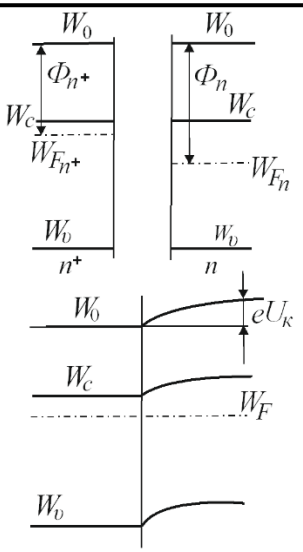


Рис. 21.22. Енергетична діаграма контакту n^+n

Досить різноманітні властивості мають гетеропереходи – контакт двох різних напівпровідників, які мають, відповідно, і різну ширину заборонених зон. Вид енергетичної діаграми і властивості гетеропереходу залежать від значень ширини забороненої зони W_{g1} і W_{g2}, енергії електронної спорідненості $W_0 - W_c$ та $W_0 - W_v$, де $W_0 -$ енергія вакууму; $W_c -$ енергія «дна» зони провідності напівпровідника n-типу; $W_v -$ енергія «стелі» валентної зони напівпровідника p-типу), від виду і концентрації домішки в областях, які контактують. Розглянемо контакт двох напівпровідників з різним характером

провідності (неізотипний) або з однаковим типом провідності (ізотипний). Утворення контактної різниці потенціалів викликає, як і у звичайному електронно-дірковому переході, викривлення енергетичних зон (рис. 21.23, *а*).

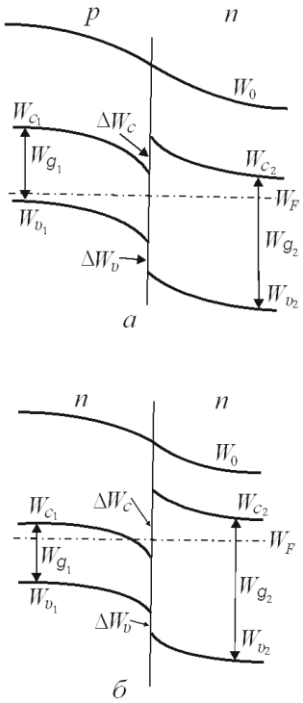


Рис. 21.23. Енергетична діаграма гетеропереходу: а – неізотипний, б – ізотипний

Утім.ю через різницю в ширині заборонених зон у місці контакту виникають розриви дозволених енергетичних рівнів ΔW_c та ΔW_v. Наявність таких розривів забезпечує різну висоту потенціального бар'єра для різнойменних носіїв: $eU_к + \Delta W_v$ – для дірок, які переходять з *p*-області в *n*-область, та $eU_к - \Delta W_c$ – для електронів, які рухаються у зворотному напрямку. Така різниця дозволяє забезпечити, по суті, одностороннюю інжекцію, у цьому випадку електронів з *n*- у *p*-напівпровідник. Слід мати на увазі, що потенціальний пік, який утворюється на межі в зоні провідності, може бути досить вузьким і тому прозорим для електронів через тунельний ефект.

Властивості односторонньої провідності можуть виявлятись і в ізотипних *nn*- або *pp*-переходах. На рис. 21.23, *б*, зображено енергетичну діаграму електронно-електронного *nn*-гетеропереходу. Як і у звичайних електронно-діркових переходах, умови переходу носіїв у протилежних напрямках тут будуть різними, що викликано різницею у ширині забороненої зони. Головною особливістю такого переходу є те, що прямий струм створюється фактично тільки електронами і процесу дифузії неосновних носіїв немає. Відсутність перехідних процесів, які пов'язані з кінцевим часом життя неосновних носіїв, дозволяє використовувати такі переходи, як і переходи з бар'єром Шотткі, для дуже високих частот.

Наведені приклади звичайно не вичерпують усіх можливих типів гетеропереходів. Їх розмаїття пов'язане з відмінністю в енергії електронної спорідненості, в ширині забороненої зони, концентрації і типі домішки напівпровідників, які контактують. Розглядаючи властивості гетеропереходів слід мати на увазі одну особливість: попри те, що гетеропереходи виготовляють із напівпровідників, які мають однотипну будову кристалічної ґратки, відмінність в їх параметрах неминуче викликає існування великої кількості поверхневих станів у місці контакту, вплив яких на властивості гетеропереходу може бути дуже значним. Наприклад, на поверхневих станах може накопичуватись заряд, поле якого буде суттєво визначати форму потенціальних бар'єрів у переході. Поверхневі стани можуть виявитись ефективними центрами рекомбінації електронів і дірок у місці контакту. Цілеспрямоване використання поверхне-вих станів дозволяє ще більше урізноманітнити властивості гетеропереходів, спектр використання яких сьогодні дуже широкий (наприклад, у напівпровідникових лазерах).

Контрольні запитання

1. Чим визначається глибина проникнення електричного поля в напівпровідник?

2 Як визначають контактну різницю потенціалів *p*- та *n*-області напівпровідника?

3. Які *p*–*n*-переходи називають різко несиметричними?

4. Які ви знаєте основні параметри *p*–*n*-переходу? Чим вони обумовлені?

5. Від чого суттєво залежить густина струму неосновних носіїв заряду через *p*–*n*-перехід?

6. Що означає зворотне зміщення для *p*–*n*-переходу?

7. Які значення ширини бар'єра і струму через *p*–*n*-перехід бувають при зворотному зміщенні?

8. Що означає пряме зміщення для *p*–*n*-переходу?

9. Які значення висоти бар'єра і струму через *p*–*n*-перехід при прямому зміщенні?

10. Чим визначають максимальне значення висоти потенціального бар'єра *p*–*n*-переходу?

11. Чим обумовлена бар'єрна ємність *p*–*n*-переходу?

12. Які основні принципи роботи фотодіодів і світлодіодів?

13. Що таке контакт «метал – напівпровідник»? В якому випадку такий контакт є омічним?

14. У чому полягає різниця між *p*−*n*-переходом і гетеропереходом?

Розділ 22. Елементи квантової електроніки

Квантова електроніка – галузь фізики, яка вивчає методи підсилення і генерації електромагнітних коливань та хвиль, які ґрунтуються на використанні вимушеного випромінювання, а також властивості квантових підсилювачів і генераторів та їх застосування. Лазери – це генератори і підсилювачі когерентного випромінювання в оптичному діапазоні довжин хвиль, робота яких ґрунтується на індукованому (вимушеному) випромінюванні квантових систем – атомів, іонів, молекул, що перебувають у станах, які суттєво відрізняються від стану термодинамічної рівноваги. Слова «лазер» і «мазер» (квантовий генератор або підсилювач діапазону НВЧ) походять з початкових букв англійської фрази Light (Microwave) Amplification by Stimulated Emission of Radiation, що означає підсилення світла (НВЧ-хвиль) індукованим (вимушеним) випромінюванням. Лазери, як і мазери, генератори й підсилювачі НВЧ-діапазону, називають ще *квантовими* генераторами (підсилювачами), оскільки поведінка частинок, які беруть участь в їх роботі, описується законами квантової механіки. Принципова відмінність лазерів від усіх інших джерел світла (теплових, газорозрядних та інших), які по суті справи є джерелами оптичного шуму, у високому ступені когерентності лазерного випромінювання. Зі створенням лазерів в оптичному діапазоні з'явились джерела випромінювання когерентних хвиль в радіодіапазоні, які можна успішно використовувати для цілей зв'язку та передавання інформації. Вони за багатьма властивостями – направленістю випромінювання, смугою частот, які передаються, низьким рівнем шумів, концентрацією енергії у просторі та в часі і под. – переважають класичні пристрої радіодіапазону.

22.1. Фізичні основи квантової електроніки

Процеси випромінювання і поглинання світла реальними атомами, молекулами та іншими квантовими системами описує квантова електродинаміка. Утім, ці процеси можна дослідити на основі простої феноменологічної теорії, яку розвинув Ейнштейн у 1916 р. У цій теорії вперше використано уявлення про *вимушене випромінювання.*

Будемо вважати, що вільний атом може перебуває тільки у стаціонарних станах з визначеними енергіями W_1, W_2, ... Перехід атома з одного стаціонарного стану в інший може відбуватись стрибком у результаті *поглинання* або

випромінювання електромагнітного випромінювання, причому для такого елементарного процесу виконується закон збереження енергії: $W_m - W_n = \hbar\omega$ – енергія фотона, який поглинається або випромінюється, дорівнює різниці енергії відповідних стаціонарних станів атома. Ці квантові уявлення про будову атома і характер його взаємодії з випромінюванням, які узагальнили гіпотезу Планка про гармонічний осцилятор, увів Бор у 1913 р., вони повністю підтверджуються сучасною квантовою теорією.

Розглянемо систему N_0 частинок (атоми, молекули) за постійної температури T (рис. 22.1). Енергія електронів в атомах дискретна. Кожному значенні енергії відповідає енергетичний рівень. Є рівень з найменшою енергією – такий рівень енергії називають основним. Усі інші рівні енергії називаються збудженими.

Якщо електрон в атомі міститься на основному рівні (або стані), то будемо вважати, що атом перебуває в основному стані. Якщо електрон міститься на збудженому рівні, то будемо вважати, що атом перебуває у збудженому стані. У рівноважному стані ($T = \text{const}$) розподіл атомів за енергетичними рівнями (станами) підпорядковується класичному розподілу Больцмана:

$$N = N_0 \exp\left\{-\frac{W}{kT}\right\}, \qquad (22.1)$$

де N_0 – загальна кількість атомів; N – кількість атомів на енергетичному рівні, енергія якого W; k – стала Больцмана.

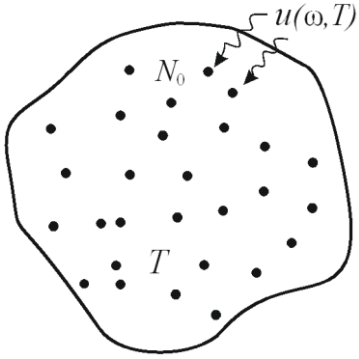


Рис. 22.1. Точками позначено частинки (атоми, молекули і под.)

Обмежимося розглядом двох енергетичних рівнів: основного і першого збудженого (рис. 22.2). Енергія основного рівня W_1, а збудженого W_2. За формулою (22.1) кількість атомів, що перебувають в основному стані (на рівні W_1),

$$N_1 = N_0 \exp\left\{-\frac{W_1}{kT}\right\}, \qquad (22.2)$$

а в збудженому (W_2)

$$N_2 = N_0 \exp\left\{-\frac{W_2}{kT}\right\}. \qquad (22.3)$$

У рівноважному стані завжди $N_2 < N_1$, оскільки $W_2 > W_1$. Кількість атомів, що містяться на цьому енергетичному рівні (у цьому енергетичному стані), називають його *заселеністю*, тобто згідно з (22.1) чим

більша енергія рівня, тим менша його заселеність. Нехай вільний атом, який не піддається зовнішнім впливам, перебуває у збудженому стані з енергією W_2. Відповідно до квантових уявлень самочинний перехід атома за відсутності зовнішніх впливів зі збудженого стану W_2 в основний W_1 з випромінюванням фотона (*спонтанне випромінювання*) відбувається миттєво, стрибком (рис. 22.2, перехід 1). В який саме момент відбувається цей перехід, передбачити неможливо. Момент випромінювання фотона є випадковою величиною, судження про яку можуть мати тільки статистичний характер. Позначимо через A_{21} ймовірність спонтанного переходу атома за одиницю часу зі збудженого стану в основний. Розглянемо сукупність дуже великої кількості однакових атомів, взаємодією між якими можна знехтувати. Нехай у момент часу t у першому збудженому стані W_2 перебувають N_2 атомів. На проміжку часу від t до $t + dt$ частина з них спонтанно перейде в основний стан. Неможливо вказати, які саме атоми здійснюють перехід, але, знаючи ймовірність A_{21}, можна вказати середню кількість dN_{21} таких переходів:

$$dN_{21} = A_{21}N_2dt. \qquad (22.4)$$

Якщо при цьому жодних процесів збудження атомів не відбувається, то зміна кількості N_2 збуджених атомів за проміжок часу від t до $t + dt$ згідно з (22.4)

$$dN_2 = -A_{21}N_2dt. \qquad (22.5)$$

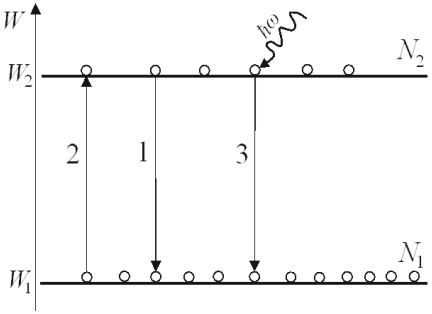


Рис. 22.2. Кружками позначено атоми на енергетичних рівнях

Коефіцієнт Ейнштейна A_{21} характеризує розглядуваний атом як такий, що не залежить від часу. Розв'язок рівняння (22.5) має вигляд

$$N_2\,(t) = N_{20}\exp\{-t/\tau\}, \qquad (22.6)$$

де N_{20} – кількість збуджених атомів за $t = 0$; $\tau = 1/A_{21}$.

Таким чином, коли немає зовнішніх впливів, у результаті спонтанного випромінювання кількість збуджених атомів зменшується за експоненціальним законом. Проміжок часу $\tau = 1/A_{21}$, протягом якого N_2 зменшується в $e = 2{,}72\ldots$ раз, дорівнює середньому часу життя атомів у збудженому стані. За таким самим законом (22.6) має зменшуватись із часом світіння газу збуджених атомів. Згідно з квантовими уявленнями спонтанне випромінювання – це сукупність

незалежних переходів: один зі збуджених атомів може повернутись в основний стан через короткий проміжок часу, інший може перебувати у збудженому стані значно довше, але середній для великої сукупності атомів час життя τ має цілком певну величину. На квантовій мові це означає, що спонтанному випромінюванню атома під час переходу із збудженого стану в основний відповідає вузький, але кінцевий інтервал частот $\Delta\nu \sim \sim 1/\tau$.

Оскільки частота випромінювання визначається умовою Бора $h\nu = = W_2 - W_1$, то розклад частот свідчить про кінцеву ширину ΔW енергетичного рівня збудженого стану W_2. Цей висновок повністю відповідає співвідношенню невизначеності Гайзенберга між енергією і часом $\Delta W \cdot \Delta t \geq \hbar$ або $\tau \geq \hbar/\Delta W$, яке пов'язує невизначеність у зміні енергії системи з невизначеністю моменту часу, коли ця зміна відбудеться.

Експериментальна перевірка закону (22.6) дала для червоної лінії водню ($\lambda = 0{,}6562$ мкм) $\tau = 1{,}5 \cdot 10^{-8}$ с, а для лінії ртуті ($\lambda = 0{,}2537$ мкм) $\tau = 9{,}8 \cdot 10^{-8}$ с.

Статистичний, випадковий характер процесів спонтанного випромінювання приводить до того, що фази, напрямки поширення і стан поляризації світлових хвиль, випромінювані окремими атомами, не узгоджені одні з одними. Це означає, що спонтанне випромінювання некогерентне. В електромагнітному полі, крім спонтанного випромінювання, будуть відбуватись і процеси збудження атомів, тобто переходи з основного стану у збуджений з поглинанням фотонів з енергією $\hbar\omega = = W_2 - W_1$. Ймовірність такого переходу за одиницю часу пропорційна густині енергії електромагнітного поля $u(\omega, T)$ на частоті переходу ω і деякому коефіцієнту B_{12}, який характеризує ймовірність збудження атома (рис. 22.1, перехід 2). Середня кількість dN_{12} переходів з основного стану в збуджений за проміжок часу від t до $t + dt$ пропорційна також кількості N_1 атомів в основному стані:

$$dN_{12} = N_1 B_{12} u(\omega, T) dt. \qquad (22.7)$$

Коефіцієнт Ейнштейна B_{12} залежить від вибраних станів W_1 та W_2 атома й може бути розрахований методами квантової механіки. Тут його розглядатимемо як феноменологічну сталу. Коефіцієнт B_{12} не залежить від спектральної густини енергії $u(\omega, T)$ поля. Пропорційність кількості dN_{12} актів поглинання (22.7) величині $u(\omega, T)$ спостерігається тоді, коли спектральна густина $u(\omega, T)$ плавно залежить від частоти ω біля частоти переходу. Припустимо тепер, що атоми

перебувають у стані термодинамічної рівноваги з полем випромінювання. Тоді на основі принципу детальної рівноваги кількість переходів з випромінювання фотонів і з поглинанням має бути однаковою. Якщо прирівняти праві частини (22.4) і (22.7), можна знайти вигляд функції $u(\omega, T)$, тобто спектральну густину рівноважного випромінювання. Водночас при цьому для $u(\omega, T)$ отримано не формулу Планка (16.21), а її окремий випадок за $\frac{\hbar\omega}{kT} \gg 1$, тобто формулу Віна (16.30). Щоб таким способом отримати формулу Планка (16.21), яку можна підтвердити на практиці, слід припустити, як вперше показав Ейнштейн, що електромагнітне поле викликає не тільки переходи 2 (рис. 22.2) з основного стану у збуджений, але й обернені переходи 3 (рис. 22.2) зі збудженого стану в основний, який супроводжується випромінюванням фотонів. Такі переходи під дією зовнішнього поля, на відміну від спонтанного, отримали назву *індукованого*, або *вимушеного* (*стимульованого*) випромінювання.

Кількість вимушених переходів $dN_{21}^{в}$ за проміжок часу від t до $t + dt$ пропорційна спектральній густині енергії $u(\omega, T)$ поля на частоті переходу, кількості N_2 атомів у збудженому стані W_2 і деякому коефіцієнту B_{21}, який характеризує ймовірність такого переходу в атомі. Із врахуванням спонтанного випромінювання повна кількість переходів за проміжок часу від t до $t + dt$ із збудженого стану в основний

$$dN_{21} = N_2(A_{21} + B_{21}\, u(\omega, T))dt. \quad (22.8)$$

У *стані термодинамічної рівноваги* слід прирівняти праві частини виразів (22.7) і (22.8). Тоді

$$u(\omega, T) = \frac{A_{21}/B_{12}}{\frac{N_1}{N_2} - \frac{B_{21}}{B_{12}}}. \quad (22.9)$$

Легко переконатись у тому, що для цієї пари рівнів енергії W_1 та W_2 коефіцієнти Ейнштейна B_{12} і B_{21} дорівнюють один одному. Справді, за дуже високої температури густина енергії $u(\omega, T)$ є настільки великою, що у формулі (22.8) можна знехтувати першим доданком порівняно із другим. Це означає, що за високої температури вимушене випромінювання переважає спонтанне. Прирівнявши для цих умов праві частини (22.7) і (22.8), отримуємо $N_1 B_{12} = N_2 B_{21}$, але при рівновазі і за $kT/(\hbar\omega) \to \infty$ відношення населеності рівнів, як видно з (22.2) і (22.3, буде

$$\frac{N_2}{N_1} = \exp\left\{-\frac{W_2 - W_1}{kT}\right\} = \\ = \exp\{-\hbar\omega/(kT)\} \Rightarrow 1. \quad (22.10)$$

Це означає, що за цих умов $N_2 = N_1$, тому $B_{12} = B_{21}$. Коефіцієнти B_{12} і B_{21} залежать тільки від властивостей атома і не залежать від зовнішніх умов, за яких відбуваються переходи, тому

рівність $B_{12} = B_{21}$, яка отримана для окремого випадку $T \to \infty$, справедлива завжди, зокрема і за відсутності теплової рівноваги.

Тоді для довільної температури з урахування (22.10) і $B_{12} = B_{21}$ формулу (22.9) перепишемо так:

$$u(\omega, T) = \frac{A_{21}/B_{21}}{\exp\{\hbar\omega/(kT)\}-1}. \quad (22.11)$$

Цей вираз збігається з формулою Планка (16.21) за $\frac{A_{21}}{B_{21}} = \frac{\hbar\omega^3}{\pi^2 c^3}$. Таким чином, усі три коефіцієнти Ейнштейна пов'язані між собою.

Найбільш важливий результат, який випливає з виведення формули Планка за Ейнштейном, – це висновок про існування вимушеного випромінювання. Тільки у цьому випадку теорія Ейнштейна не суперечить законам теплового випромінювання. За $\hbar\omega >> kT$ (справедлива формула Віна (16.30)) для підтримки термодинамічної рівноваги фактично достатньо спонтанного випромінювання. Вимушене випромінювання у таких умовах значно менше за поглинання, оскільки за $kT << \hbar\omega$ термічно збуджених атомів мало. Навпаки, в умовах низьких частот і високих температур, коли $\frac{\hbar\omega}{kT} << 1$, за теплової рівноваги переходи з вимушеним випромінюванням відбуваються майже так само часто, як і спонтанні переходи з випромінюванням.

Вимушене випромінювання має чудові властивості, на що вперше звернув увагу П. Дірак у 1927 р., застосувавши квантову механіку до поля випромінювання. У кожному акті вимушеного випромінювання відбувається збільшення на одиницю кількості фотонів у тій моді (частоті) випромінювання, під дією якої відбувався перехід. Усі фотони однієї моди тотожні. Це означає, що новий фотон не відрізняється від фотонів, які викликали це випромінювання. Частота, фаза, напрямок і поляризація хвиль, які випромінюються за вимушених переходів, точно такі самі, як і у випромінювання, яке викликає переходи.

Розглянемо зв'язок між коефіцієнтами Ейнштейна і коефіцієнтом поглинання α у формулі (15.6). Нехай у розрідженому середовищі поширюється паралельний пучок монохроматичного випромінювання, частота якого $\omega = (W_2 - W_1)/\hbar$. Зміна кількості dN фотонів у потоці під час проходження шару, товщина якого $dx = cdt$ (c – швидкість випромінювання в середовищі) унаслідок процесів вимушеного поглинання і вимушеного випромінювання відповідно до (22.7) і (22.8)

$$dN = (N_2 B_{21} - N_1 B_{12}) u(\omega, T) dx/c. \quad (22.12)$$

Внеском спонтанного випромінювання у зміну кількості фотонів направленного пучка можна знехтувати, оскільки спонтанне випромінювання поширюється у всіх напрямах і у фіксований напрямок пучка, який розглядається, потрапляє незначна його частина. Формула (22.12) дозволяє з'ясувати, які умови потрібні для безпосереднього експериментального виявлення вимушеного випромінювання. Оскільки B_{21}=B_{12}, то

$$dN = (N_2 - N_1)B_{12}u(\omega,T)dx/c.$$

Звідси видно, що у поширенні направленого пучка внесок поглинання переважає внесок вимушеного випромінювання й інтенсивність пучка зменшується у всіх випадках, коли $N_2 < N_1$, тобто кількість збуджених атомів менша від кількості атомів в основному стані (закон розподілу (22.1)). Так буває, якщо пучок поширюється у середовищі, яке перебуває у стані термодинамічної рівноваги або близькому до нього стані. Щоб спостерігати наростання інтенсивності направленого пучка (негативне поглинання) і тим самим експериментально виявити вимушене випромінювання, треба створити в середовищі *нерівноважний стан*, за якого кількість атомів на більш високому енергетичному рівні була би більшою, ніж на низькому ($N_2 > N_1$). Першу спробу виявити вимушене випромінювання у видимій області спектра у дослідах з парами ртуті, атоми якої переводились в нерівноважний стан електричним розрядом, зробив В. А. Фабрикант (1939 р.). Він же вперше висловив ідею використання вимушеного випромінювання для підсилення світла.

Знайдемо закон зміни інтенсивності пучка у міру його поширення в середовищі. Середній потік енергії випромінювання (інтенсивність $\langle \Pi \rangle = I$) за одиницю часу дорівнює добутку $u(\omega,T)$ на швидкість світла c: $I = u(\omega,T)\cdot c$. Оскільки зміна *потоку* фотонів на шляху dx дорівнює cdN, то зміна потоку енергії dI на тому ж відрізку dx дорівнює $\hbar\omega c\cdot dN$. Помножимо обидві частини рівності (22.12) на $\hbar\omega c$ і знайдемо, що

$$dI = (\hbar\omega/c)(N_2 - N_1)B_{12}u(\omega,T)cdx$$

або

$$dI = -\left(\frac{\hbar\omega}{c}\right)(N_1 - N_2)B_{12}\,Idx.$$

Звідси

$$dI/dx = -\alpha I(x), \qquad (22.13)$$

де

$$\alpha = \left(\frac{\hbar\omega}{c}\right)(N_1 - N_2)B_{12}. \qquad (22.14)$$

Якщо населеність рівнів N_1 і N_2 не залежить від x (однорідне середовище), то розв'язок (22.13) має вигляд

$$I(x) = I_0\exp\{-\alpha x\}, \qquad (22.15)$$

де I_0 – інтенсивність направленого пучка за $x = 0$. Якщо $\alpha > 0$, що буває за $N_2 < N_1$, то інтенсивність випромінювання експоненціально зменшується у міру поширення пучка і ми маємо закон Бугера (15.82), а формула (22.14) виражає коефіцієнт поглинання α через коефіцієнт Ейнштейна B_{12} і населеність рівнів N_1 та N_2.

Для тих випадків, коли у середовищі виконується умова $N_2 > N_1$, відома як *інверсна населеність*, у (22.14) $\alpha < 0$ (негативний коефіцієнт поглинання) інтенсивність хвилі відповідно до (22.15) наростає у міру її поширення. Підсилення падаючого пучка світла здійснюється унаслідок того, що за $N_2 > N_1$ переходи з вимушеним випромінюванням фотонів відбуваються частіше, ніж переходи з поглинанням. Оскільки фотони, що виникають під час вимушеного випромінювання, тотожні з фотонами, які викликають випромінювання, то когерентні властивості початкового пучка повністю зберігаються. Таким є принцип дії *квантового підсилювача* випромінювання. Різні способи створення потрібного для роботи підсилювача середовища з інверсною населеністю (*активного середовища*) розглянуто далі. Важливо зауважити, що для створення активного середовища завжди потрібно ззовні підводити додаткову енергію, яка потім під час вимушеного випромінювання частково перетворюється в енергію електромагнітного випромінювання, яке підсилюється. Квантовий підсилювач можна перетворити у генератор випромінювання, якщо ввести в ньому *позитивний зворотний зв'язок*. Для цього частина випромінювання, яке виходить з підсилювача, має повертатись в активне середовище. Тоді відпадає необхідність у вхідному сигналі, оскільки відбувається самозбудження системи. Обернений зв'язок можна здійснити, розмістивши активне середовище в резонатор, який утворений двома паралельними дзеркалами.

22.2. Лазери

Принцип роботи лазера або оптичного квантового генератора (ОКГ) оснований на трьох фундаментальних ідеях, які народились у різний час в різних сферах фізики.

Перша ідея пов'язана з використанням *вимушеного випромінювання* світла атомними системами, яке відкрив Ейнштейн у 1917 р., теоретично вивчаючи некогерентне теплове випромінювання. Як показав Дірак, фотони вимушеного випромінювання не відрізняються від фотонів, які викликали це випромінювання.

Друга ідея полягає у застосуванні *термодинамічно нерівноважних систем*, в яких можливе підсилення, а не поглинання світла. Її висловив В. А. Фабрикант у 1940 р.

Третя ідея, яка бере початок у радіофізиці, полягає у використанні *позитивного оберненого зв'язку* для перетворення підсилювальної системи в генератор когерентного випромінювання.

Те, що перші квантові прилади з'явились у радіодіапазоні НВЧ, пояснюється тим, що класична радіофізика не могла вирішити традиційними для неї методами ряд дуже важливих із практичної точки зору проблем, таких як освоєння більш коротких хвиль, створення у короткохвильовому діапазоні високостабільних, малошумних і достатньо потужних приладів. Це наполегливо змушувало шукати інший підхід розв'язку цих проблем і призвело до створення в НВЧ-діапазоні принципово нових приладів – квантових підсилювачів і генераторів.

Перший квантовий генератор (лазер), який працював на квантових переходах з довжиною хвилі 1,25 см, був створений у 1954 р. Активним середовищем у них був пучок молекул аміаку – інверсна населеність на робочому переході досягалась просторовим розподілом молекул аміаку, які перебували в різних квантових станах, під час проходження пучка через неоднорідне електричне поле (див. формулу (2.8)). Перший оптичний квантовий генератор, який працював в імпульсному режимі, створив у 1960 р. Мейман. Активним середовищем у ньому служив стрижень із кристала рубіна, який збуджувався світлом від лампи спалаху. Роком пізніше Джаван, Беннет і Ерріот створили перший лазер неперервної дії з газоподібним активним середовищем (суміш неону і гелію). Сьогодні є багато типів лазерів, які відрізняються способами збудження активного середовища, робочою спектральною областю, потужністю, часовими і спектральними характеристиками випромінювання і под.

Існують різні способи отримання потрібного для лазера активного середовища, яке підсилює випромінювання. Згідно з рис. 22.2 для створення стаціонарної інверсії ($N_2 > N_1$) потрібно, щоб швидкість накачки (процеси 2) і час життя атомів у збудженому стані W_2 були більшими, ніж для нижнього рівня W_1. Із цього випливає, що віддати перевагу необхідно селективній накачці і що інверсія може бути досягнута не тільки внаслідок переважного заселення верхнього рівня W_2 (рис. 22.2), але й унаслідок високої швидкості очистки нижнього рівня W_1.

Є такі способи створення активного середовища (інверсна заселеність енергетичних рівнів W_1 і W_2 (рис. 22.2)):

1. Унаслідок поглинання світла (оптична накачка). Підбираючи джерело світла з відповідним спектром випромінювання, можна забезпечити високу селективність (вибірковість) накачки. Найбільш успішно цей вид накачки використовують у твердотільних (на кристалах і стеклах) лазерах і в лазерах на барвниках.

2. Унаслідок непружних зіткнень атомів і молекул з вільними електронами, за яких частина енергії електрона йде на збудження атома або молекули. Вільні електрони можуть створюватись або під час газового розряду, або вводитись у газ у вигляді пучка, який формується у прискорювачі.

3. Унаслідок непружних зіткнень атомів робочої речовини зі збудженими атомами або іонами допоміжного газу з передаванням енергії збудження від них робочій речовині. У деяких типах зіткнень передавання енергії має резонансний характер і досягається високий ступінь селективності заселеності рівнів.

4. У процесі спеціально підібраних хімічних реакцій (хімічна накачка); при цьому збуджуються коливальні рівні молекул, причому збудження може бути селективним.

5. Унаслідок нагріву (теплова накачка). Цей метод використовують для накачки коливальних рівнів у молекулах, інверсія на переходах між якими здійснюється за рахунок різного часу життя (часу релаксації) для верхнього і нижнього лазерних рівнів під час швидкого адіабатичного (без теплообміну) розширення газу. На цьому принципі ґрунтується робота газодинамічних лазерів.

6. У напівпровідникових лазерах неперервної дії нерівноважного (інверсного) стану досягають під час пропускання електричного струму через $p-n$-перехід.

Щоб перетворити підсилювач (активне середовище) у генератор, треба організувати зворотний позитивний зв'язок. Для його здійснення частина випромінювання, яке генерується, має залишатись всередині активного середовища і викликати вимушене випромінювання все новими і новими збудженими атомами. З цією метою активне середовище уміщують в оптичний резонатор (рис. 22.3), який утворений двома паралельними плоскими або сферичними дзеркалами, одне з яких напівпрозоре. Розглянемо, як

відбувається генерація («запалювання») ОКГ.

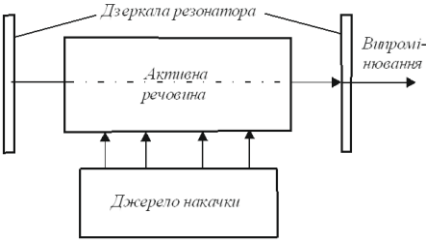


Рис. 22.3. Схема будови лазера

Нехай є активне середовище (рис. 22.4), тобто середовище з інверсною населеністю енергетичних рівнів ($N_2 > > N_1$). Якщо якийсь атом спонтанно випромінив хвилю у напрямку під кутом до оптичної осі (рис. 22.4, хвиля 1), то ця хвиля в цьому напрямку індукує вимушене випромінення, яке рано чи пізно покине активне середовище, тобто будь-яке спонтанне випромінення у напрямку під кутом до оптичної осі резонатора не має шансів бути підсиленим і вийти з ОКГ. Тільки хвиля, спонтанно випромінена атомом у напрямку оптичної осі резонатора (рис. 22.4, хвиля 2), буде підсилюватись, зустрічаючи на своєму шляху поширення збуджені атоми (атоми на енергетичному рівні W_2). Підсилення хвилі є тим більшим, чим більший шлях в активному середовищі вона пройде. Після відбивання від дзеркала її підсилення продовжуватиметься. Цей процес багаторазово повторюється, а енергія хвилі в резонаторі стрімко наростає. У цьому і проявляється позитивний зворотний зв'язок, який здійснюється за допомогою дзеркал. Підсилення хвилі відбувається згідно з формулою (22.15) стільки разів, скільки разів хвиля проходить крізь активне середовище. Якщо активне середовище заповнює простір між дзеркалами, то, відбившись від дзеркала, хвиля знову на довжині резонаторами L підсилюється в $\exp\{|\alpha|L\}$ разів, де $|\alpha|$ – коефіцієнт підсилення (формула (22.14)). Слід зазначити, що система із двох паралельних дзеркал має резонансні властивості – резонує тільки на певних частотах і виконує в лазері ту роль, яку у звичайних низькочастотних генераторах відіграє коливальний контур.

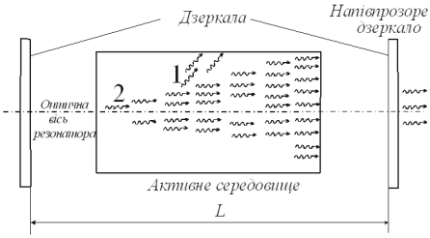


Рис. 22.4. Схема виникнення лазерного випромінювання

Нехай на подвійній довжині резонатора вкладається ціле число довжин хвиль: $2L = \lambda q$, де q – ціле число. Цим забезпечується позитивний обернений звязок: під час другого і всіх наступних проходжень через резонатор

хвиля узгоджена за фазою з вихідною, тобто фактично не відрізняється від неї. Усі елементарні хвилі, які виникають у різні моменти часу під час вимушених переходів атомів, що розміщені в різних точках активного середовища, створюють одну когерентну хвилю. Таким чином, оптичний резонатор збільшує ефективну відстань, яку хвиля, що поширюється вздовж його осі, проходить в активному середовищі і сприяє формуванню когерентного монохроматичного випромінювання, яке виходить крізь напівпрозоре дзеркало.

Якщо оптичний резонатор розглядати як коливальну систему, в якій власні нормальні коливання (*моди*) мають вигляд стоячих електромагнітних хвиль з вузлами на дзеркалах, то частоти ν таких мод визначають з умови $2L = q\lambda_q$, де q – ціле число, тобто утворюють еквідистантний спектр:

$$\nu_q = v/\lambda_q = c/(n\lambda_q) = q(c/(2nL)), \quad (22.16)$$

де n – показник заломлення активного середовища; c – швидкість світла у вакуумі. Отже, лазер – це автоколивальна система, в якій можливі незгасаючі коливання на одній із власних частот резонатора. Частотна відстань між сусідніми модами (q відрізняється на одиницю)

$$\Delta\nu = c/(2nL). \quad (22.17)$$

Водночас існує низка причин, які будуть зменшувати амплітуду хвилі всередині резонатора. Це втрати під час відбивання від дзеркал, а також втрати на розсіювання в середовищі та деякі інші. Якщо ввести для описання сумарних втрат коефіцієнт втрат β, то формулу (22.15) перепишемо так:

$$I = I_0 \exp\{-(\alpha + \beta)x\}. \quad (22.18)$$

Для виникнення стійкої генерації потрібно, щоб $(\alpha + \beta < 0)$, тобто втрати випромінювання мають бути меншими, ніж підсилення.

Слід також підкреслити, що корисна дія дзеркал виявляється і в тому, що неперервно збільшується кількість атомів, які випромінюють вимушено, оскільки за кожний цикл відбивання зростає густина енергії електромагнітних хвиль $u(\omega, T)$. Зазначимо, що цей процес не буде продовжуватися нескінченно, оскільки починаючи з деякого значення $u(\omega, T)$ настає явище насичення в ОКГ. Інверсія в заселеності енергетичних рівнів та кількість вимушених переходів з випромінюванням зменшується, а отже, і підсилення уповільнюється. Ефект насичення – це одне з основних явищ нелінійної оптики, який відіграє важливу роль у розумінні механізму генерації випромінювання ОКГ і визначає умови стандартності процесу підсилення у лазерах.

На рис. 22.5 показано частотну залежність коефіцієнта підсилення α в робочому середовищі (вона визначається шириною і формою спектральної лінії робочого середовища, наприклад лінії під час переходу атома з рівня W_2 на рівень W_1 (рис. 22.2), і набором власних частот відкритого резонатора). Для відкритих резонаторів з високою добротністю пропускання резонатора $\Delta\nu_р$, яка визначає ширину резонансних кривих окремих мод (криві 2−4 на рис. 22.5), і навіть відстань між сусідніми модами $\Delta\nu$ є меншою від ширини лінії підсилення $\Delta\nu_п$ (рис. 22.5), тому у контур підсилення потрапляє декілька типів мод коливань резонатора.

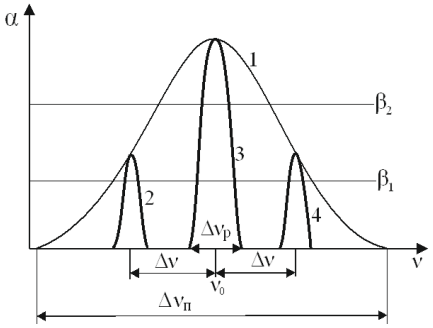


Рис. 22.5. Частотна залежність коефіцієнта підсилення α в робочому середовищі (крива 1) і типи коливань резонатора (2−4), які потрапляють у контур підсилення ($\Delta\nu_п$ − ширина лінії підсилення). За рівня втрат β_1 можуть генеруватися три моди (2,3,4), а за β_2 – одна центральна (3) ($\beta_2 > \beta_1$)

Таким чином, лазер не обов'язково генерує на одній частоті, частіше навпаки, генерація відбувається одночасно на декількох типах коливань, для яких підсилення α є більшим за втрати в резонаторі β (три моди на рис. 22.5 за $\beta = \beta_1$). Щоб лазер працював на одній частоті (в одночастотному режимі), зазвичай потрібно вживати спеціальні заходи (наприклад, збільшити втрати, як показано на рис. 22.5 за $\beta = \beta_2$) або змінити відстань між дзеркалами так, щоб у контур підсилення потрапила тільки одна мода. Це забезпечить високу монохроматичність ($\Delta\lambda/\lambda$) лазерного випромінювання. Тобто, якщо коефіцієнт підсилення в робочому середовищі перекриває втрати в резонаторі для певних типів коливань, то на них виникає генерація. Приводом для її виникнення є, як і в будь-якому генераторі, шум. У лазерах цим шумом є спонтанне випромінювання.

Необхідно зазначити ще одну властивість лазерного випромінювання – його вузьку спрямованість (малу розбіжність). Вона є наслідком того, що хвиля, яка виходить з ОГК, є результатом складання величезної кількості висококогерентних хвиль від багатьох атомів.

Тут доречно згадати явище дифракції на ґратці, де розподіл

інтенсивності світла на екрані є результатом інтерференції когерентних хвиль, які йдуть від великої кількості щілин, причому чим більше щілин ґратки, тим більша інтенсивність максимуму ($\sim N^2$) і менша його ширина ($\sim \frac{1}{N}$), і тим швидше спадає інтенсивність інших максимумів. Випромінювання лазера з певним допущенням можна вважати граничним випадком дифракції. Випромінювання атомів активного середовища когерентне, кількість їх величезна, а тому випромінювання активного середовища зосереджене у вузькому дифракційному куті, а інтенсивність на осі променя в N^2 разів більша від інтенсивності хвилі окремого атома, однак цей кут через дифракцію на вихідній щілині отвору ОКГ не може бути зведений до нуля у принципі. Таким чином, значна когерентність лазера і визначає незначну розбіжність його випромінювання.

22.3. Властивості лазерного випромінювання

Опишемо властивості лазерного випромінювання.

1. *Потужність* лазерного випромінювання в неперервному режимі може бути 10^5 ... Вт, в імпульсному – до 10^{12} ... 10^{16} Вт, при цьому вдається досягнути інтенсивності 10^{12} ... 10^{16} Вт/см^2. Суттєво, що ці потужності можуть бути сконцентровані у надзвичайно вузьких спектральних ($\Delta\lambda$ – мале) і часових інтервалах.

2. *Тривалість* імпульсу випромінювання в лазерах може становити 10^{-12} ... 10^{-13} с і спеціальними методами її можна довести до 10^{-15} с (за цей час світло проходить всього $3 \cdot 10^{-5}$ см), тобто лазер має дивовижно високий ступінь концентрації енергії за часом.

3. *Монохроматичність* лазерного випромінювання визначають як $\Delta\nu_p/\nu_0$ (або $\Delta\lambda_p/\lambda_0$), де $\Delta\nu_p$ – ширина лінії генерації (рис. 22.5), яка у принципі обумовлена шумами, ν_0 – її центральна частота, під час роботи лазера на одній частоті та в неперервному режимі. Використовуючи спеціальні методи стабілізації, вдається отримати відносну стабільність частоти $\Delta\nu_p/\nu_0 =$ $= 10^{-12}$ ... 10^{-13} ($\Delta\nu_p = 50$ ... 500 Гц).

Високий ступінь монохроматичності лазерного випромінювання визначає високу спектральну густину енергії – високий ступінь концентрації світлової енергії у дуже малому спектральному інтервалі. Висока монохроматичність полегшує фокусування лазерного випромінювання, оскільки при цьому хроматична аберація лінзи (залежність показника заломлення матеріалу лінзи n від довжини хвилі λ) стає несуттєвою.

4. *Когерентність*. Лазери мають надзвичайно високий, порівняно з іншими джерелами світла, ступінь когерентності – часовий і просторовий (див. підрозд. 12.5 і 12.6). Під час роботи лазера в одномодовому режимі досягають повної просторової когерентності, що визначає високу направленість лазерного випромінювання і робить можливим його фокусування у пляму надзвичайно малих розмірів (з довжиною хвилі λ). Часову когерентність (формули (12.37) і (12.38)) можна записати так:

$$t_{когер} = \lambda^2/(c \cdot \Delta\lambda) = 1/\Delta\nu_р, \quad (22.19)$$

тобто характеризується монохроматичністю і виявляється дуже високою. Так, для лазера на $He - Ne$, який працює неперервно в одномодовому режимі, $\Delta\nu_р \cong 1$ кГц і $t_{когер} = 10^{-3}$ с. Тоді довжина когерентності $l_{когер} = t_{когер} \cdot c$ (с – швидкість світла) становить $3 \cdot 10^5$ м (300 км), тоді як для нелазерних джерел світла (наприклад, натрієва лампа $t_{когер} = 10^{-10}$ с) $l_{когер} = 0{,}03$ м. Таким чином, з використанням лазерів можна спостерігати інтерференційну картину навіть за різниці ходу променів у декілька кілометрів.

5. *Направленість* лазерного випромінювання визначається переважно тим, що у відкритому резонаторі можуть збуджуватись тільки такі хвилі, які напрямлені по осі резонатора (рис. 22.4) або під дуже малими кутами до неї. У разі високого ступеня просторової когерентності кут розбіжності лазерного променя θ може бути доведений до граничного, який визначається дифракцією на вихідному отворі лазера. Типові значення θ становлять:

– для газових лазерів $(0{,}5 \ldots 5) \cdot 10^{-3}$ радіан,

– у твердотільних $(2 \ldots 20) \cdot 10^{-3}$ радіан,

– у напівпровідникових $(5 \ldots 50) \cdot 10^{-3}$ радіан.

6. *Яскравість*. Через високу направленість випромінювання лазерні джерела світла мають дуже високу яскравість, завдяки чому на мішені можна отримати дуже велику інтенсивність світла. Так, гелій-неоновий лазер з потужністю всього 10 мВт і розбіжністю випромінювання 10^{-4} радіан при поперечній площі пучка 0,1 $см^2$ має яскравість 10^6 $Вт/(см^2 \cdot стерадіан)$, що набагато перевищує яскравість Сонця (130 $Вт/(см^2 \cdot стерадіан)$).

Примітка. Яскравість L, поверхнево-просторова густина світлового потоку, який йде від поверхні, дорівнює відношенню світлового потоку $d\Phi$ до геометричного фактора $d\Omega dS\cos\theta$:

$$L = d\Phi/(d\Omega dS\cos\theta),$$

де $d\Omega$ – заповнений випромінюванням тілесний кут; dS – площина, яка

випромінює або приймає випромінювання; θ – кут між перпендикуляром до цієї площини і напрямком випромінювання. У нашому випадку $d\Omega = \pi \mathrm{tg}^2\,\theta \cong$ $\cong 3 \cdot 10^{-7}$ стерадіан, $\cos\theta \approx 1$, тому

$$L = \frac{10 \cdot 10^{-3}}{3 \cdot 10^{-7} \cdot 10^{-1}} \frac{\text{Вт}}{\text{см}^2 \cdot \text{стерадіан}} \sim$$
$$\sim 0{,}3 \cdot 10^{6} \frac{\text{Вт}}{\text{см}^2 \cdot \text{стерадіан}}.$$

22.4. Особливості конструкції і роботи деяких типів лазерів

Розглянемо деякі характеристики окремих лазерів.

1. *Лазер на кристалі рубіна*. Рубін – це кристал окису алюмінію Al_2O_3 (корунд), в якому невелика частина іонів алюмінію (0,05 %) під час вирощування кристала заміщена іонами хрому Cr^{3+}. Сам корунд у видимій області прозорий, а основну роль в роботі лазера відіграють іони хрому. Спрощену схему енергетичних рівнів хрому показано на рис. 22.6, *а*. Інверсія населеності створюється між основним станом (рівнем) W_1 і збудженим станом (рівнем) W_2. Цьому переходу відповідає довжина хвилі $\lambda = 0{,}6943$ мкм у червоній області спектра. Вище рівня W_2 лежать широкі смуги енергетичних рівнів W_3 і W_3'. Переходам на них з основного стану відповідають дві широкі смуги поглинання в зеленій і синій областях спектра. З цим поглинанням пов'язане рожеве забарвлення рубіна.

Циліндричний стрижень рубіна, довжина якого декілька сантиметрів і діаметр близько 1 см, освітлюється потужним імпульсом білого світла від лампи спалаху, яка обвиває його у вигляді спіралі (рис. 22.6, *б*). Тривалість імпульсу становить близько 1 мс. За достатньої енергії спалаху більша частина іонів хрому, поглинаючи світло, переходить у стани W_3 та W_3'. Потім іони хрому за час близько 10^{-8} с без випромінювання переходять на рівень W_2, передаючи надлишок енергії коливанням кристалічної ґратки. Час життя (час релаксації) іонів хрому у збудженому стані W_2 становить декілька мілісекунд, що на декілька порядків перевищує типову величину часу життя збуджених станів $(10^{-7} \ldots 10^{-8})$ с. Збуджені рівні з таким великим часом життя називають *метастабільними*. За недостатньої енергії спалаху на рівні W_2 виявиться менше половини всіх іонів хрому. Тоді зворотний їхній перехід в основний стан відбувається спонтанно (за час ≈ 1 мс) і супроводжується люмінесценцією (свіченням) на довжині хвилі 0,6943 мкм.

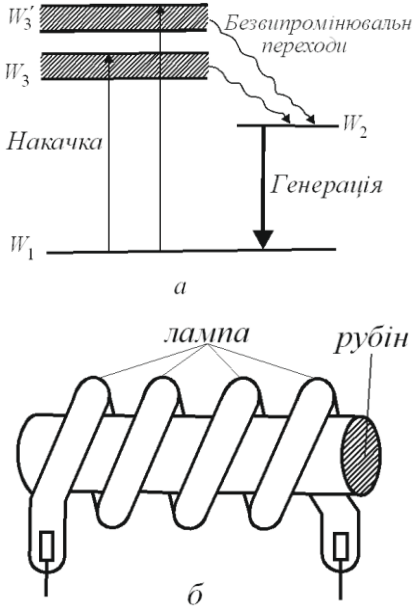


Рис. 22.6. Схема рубінового лазера

Якщо ж у результаті поглинання світла спалаху і наступних безвипромінювальних переходів на рівні W_2 накопичиться більше половини всіх іонів хрому, то між рівнями W_1 і W_2 виникає інверсія населеностей ($N_2 > N_1$). Такий механізм утворення інверсії населеності називають *оптичною накачкою.* Якщо рубіновий стрижень розміщений у зовнішньому оптичному резонаторі або має посріблені плоскопаралельні торці, в ньому виникає короткий імпульс лазерної генерації на довжині хвилі 0,6943 мкм. Лазерний імпульс має складну часову структуру і складається з нерегулярної послідовності окремих імпульсів тривалістю близько 1 мкс. Через малу тривалість імпульсу (близько 1 мс) потужність рубінового лазера в імпульсі досягає декількох кіловат за порівняно невеликої енергії (декілька джоулів).

2. *Лазер на неодимовому склі*. Крім рубінового лазера, який працює за *трирівневою схемою*, значного поширення отримали *чотирирівневі лазери* на іонах рідкоземельних елементів (неодим, самарій), які впроваджені у кристалічну (флюорит кальцію) або скляну матрицю. Ці іони мають структуру енергетичних рівнів (рис. 22.7), яка надзвичайно добре підходить для лазерів. За кімнатної температури різниця енергій $W_2 - W_1$ рівнів 2 й 1 в декілька разів перевищує kT, тому рівень W_2 фактично порожній (слабо заселений). За допомогою сильного світла лампи накачки іони неодиму збуджуються в широку смугу рівнів 4 (W_4). Час життя іона у стані 4 дуже малий ($\tau \geq \hbar/\Delta W$, а ΔW порівняно значна величина), тому з великою ймовірністю відбуваються безвипромінювальні переходи у метастабільний стан 3 (W_3). Інверсія населеності між рівнями 2 і 3 лазерного переходу досягається дуже легко, оскільки нижній рівень 2 майже незаселений. Лазер на неодимовому склі генерує випромінювання на довжині хвилі 1,06

мкм у вигляді імпульсів з дуже великими енергіями (близько 1 кДж). За чотирирівневою схемою працює лазер неперервної дії на кристалах ітрій-алюмінієвого граната (YAG) з домішкою неодима і з вихідною потужністю до 1 кВт.

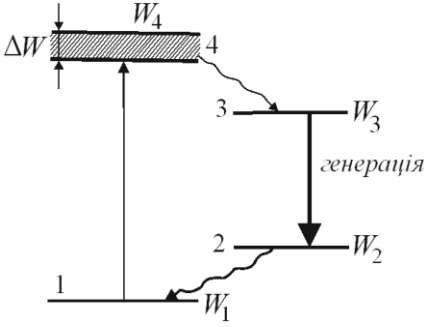


Рис. 22.7. Структура енергетичних рівнів лазерів на іонах рідкоземельних елементів

3. *Газові лазери* за багатьма характеристиками переважають лазери інших типів. Вони перекривають широкий спектральний діапазон (від субміліметрового до ультрафіолетового). Серед газових лазерів завжди можна знайти лазер, який має принаймні одну з таких властивостей: високий ступінь монохроматичності випромінювання, малу розбіжність випромінювання (аж до дифракційної межі), граничне значення потужності в неперервному режимі, високий ККД тощо.

Будову найбільш поширеного гелій-неонового лазера схематично показано на рис. 22.8. Газорозрядна трубка із внутрішнім діаметром 1... 10 мм і довжиною від декількох десятків сантиметрів до 1,5... 3 м має торцеві плоскопаралельні скляні або кварцові вікна, які встановлені під кутом Брюстера до її осі. Для лінійно поляризованого світла з електричним вектором у площині падіння коефіцієнт відбивання дорівнює нулю (див. підрозд. 14.2, рис. 14.8 і 14.10), тому брюстеровські вікна забезпечують лінійну поляризацію випромінювання лазера і виключають втрати енергії під час поширення світла з активного середовища до дзеркал і назад. Трубку розміщують у резонаторі, який утворений двома дзеркалами. Пропускання дзеркала, крізь яке виводиться випромінювання, вибирають близько 1−2 %, іншого − менше 1 %.

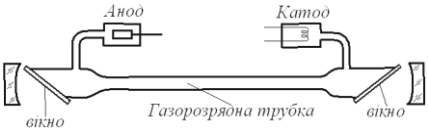


Рис. 22.8. Розрядна трубка газового лазера з вікнами, які встановлені під кутом Брюстера

Розрядну трубку заповнюють сумішшю гелію і неону в молярному відношенні 5:1 за тиску 10^2 ... 10^3 Па. До електродів вмикається джерело з напругою у декілька кіловольтів. Типова сила струму під час розряду – десятки міліамперів. Тліючий розряд створює

умови для виникнення інверсії населеності рівнів у неоні. Гелій використовують для резонансного збудження неону.

На рис. 22.9 показано спрощену схему енергетичних рівнів неону Ne. Випромінювання з довжиною хвиль 0,6328 і 1,150 мкм відповідають переходам $W_4 \to W_2$ та $W_3 \to W_2$. Атоми неону збуджуються (накачуються) у результаті їхнього зіткнення з електронами в газовому розряді, при цьому вони переходять зі стану W_1 у стани W_3 і W_4.

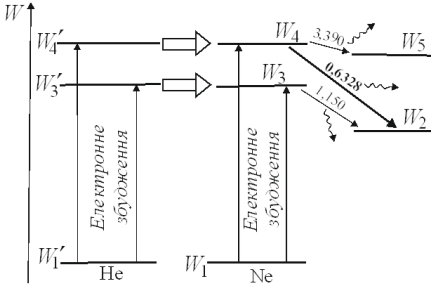


Рис. 22.9. Схема енергетичних рівнів Не *та* Ne *– сисла означають довжину хвиль випромінювання в мікрометрах*

Енергії електронів газового розряду вистачає для створення інверсного заселення рівнів W_3 і W_2. Випромінювання, яке виникає при цьому, припадає на інфрачервону область спектру ($\lambda = 1{,}15$ мкм). Рівні $W_4 \to W_2$ та $W_4 \to W_5$ залишаються заселеними неінверсно.

Інтенсивне заселення цих рівнів можна забезпечити, якщо додати до неону атоми гелію Не. Гелій має два метастабільних (наче стабільних) стани W_3' і W_4', які «заселяються» під час зіткнення атомів з електронами і за величиною енергії близькі до рівнів W_3 і W_4 неону. Час життя атомів у цих станах набагато більший від часу життя у звичайних станах (10^{-8} с). Пов'язано це з тим, що для гелію переходи $W_3' \to W_1'$, $W_4' \to W_1'$ є забороненими, тобто не можуть бути здійсненими шляхом випромінювання, а будь-яким іншим, наприклад у процесі зіткнення з атомами неону (горизонтальні стрілки на діаграмі). У результаті таких зіткнень атоми неону набувають енергії, і їх кількість на рівнях W_3 та W_4 значно зростає. Виникає інверсне заселення рівнів W_4 і W_2, а отже, можливе підсилення випромінювання ($\lambda = 0{,}6328$ мкм). Висока оптична однорідність газового активного середовища дозволяє отримувати випромінювання з дуже високим ступенем часової і просторової когерентності.

4. *Напівпровідниковий лазер*, лазер на основі напівпровідникового кристала. На відміну від лазерів інших типів, у напівпровідниковому лазері використовують випромінювальні квантові переходи

між дозволеними енергетичними зонами, а не дискретними рівнями енергії. У напівпровідниковому *активному середовищі* може досягатись дуже великий показник оптичного підсилення α до 10^6 м$^{-1}$ (формула (22.14), завдяки чому розміри активного елемента напівпровідникового лазера виключно малі (довжина резонатора $L \sim 50$ мкм ... 1 мм). Крім компактності, особливостями напівпровідникового лазера є мала інерційність ($\sim 10^{-9}$ с), високий коефіцієнт корисної дії (до 50 %), можливість спектрального перетворювання і великий вибір речовин для генерації в широкому спектральному діапазоні від $\lambda = 0{,}3$ мкм до $\lambda = 30$ мкм. Активними частинками у напівпровідниковому лазері є надлишкові (нерівноважні) електрони провідності й дірки, тобто вільні носії заряду, які можуть інжектуватись, дифундувати і дрейфувати в активному середовищі. Найважливішим способом накачки у напівпровідниковому лазері є інжекція носіїв через $p-n$-перехід або гетероперехід, яка дозволяє здійснювати безпосереднє перетворення електричної енергії в когерентне випромінювання (інжекційний лазер). Іншим способом накачки служить електричний пробій (наприклад у стримерних лазерах), бомбування електронами (напівпровідниковий лазер з електричною накачкою) та освітлення (напівпровідниковий лазер з оптичною накачкою). Напівпровідникові лазери вперше створені на $p-n$-переході у кристалі арсеніду галію GaAs з електричною накачкою.

Оптичне підсилення у напівпровідниках виникає під дією інтенсивної накачки при виконанні умов інверсії населеності рівнів біля дна W_c у зоні провідності та стелі W_v у валентній зоні (рис. 22.10). При цьому ймовірність заповнення електронами верхніх робочих рівнів у дозволеній зоні (зоні провідності) більша, ніж нижніх рівнів (валентної зони). У цьому випадку вимушені випромінювальні переходи переважають над поглинальними переходами (рис. 22.11). Величина оптичного підсилення залежить не тільки від інтенсивності накачки, але й від інших факторів: ймовірності випромінювальної рекомбінації, внутрішнього квантового виходу випромінювання, температури.

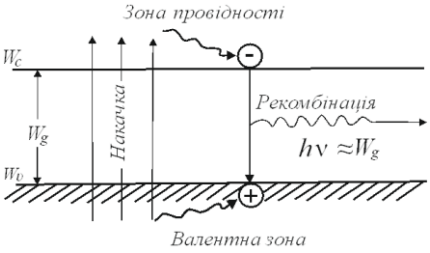


Рис. 22.10. Схема напівпровідникового лазера з оптичною накачкою

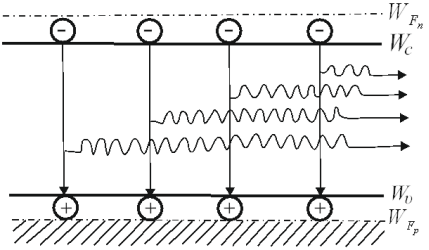


Рис. 22.11. Генерація випромінювання в напівпровідниковому лазері з оптичною накачкою

Як лазерні матеріали використовують прямозонні напівпровідники (наприклад, GaAs, CdS, PbS), в яких квантовий вихід випромінювання може досягти 100 %. На непрямозонних напівпровідниках (Ge, Si) створити лазер не вдається.

Різноманітність напівпровідникових лазерних матеріалів дозволяє перекрити широкий спектральний діапазон з допомогою напівпровідникових лазерів. Інжекційний напівпровідниковий лазер – це напівпровідниковий діод, дві плоскопаралельні грані якого перпендикулярні площині $p-n$-переходу (гетеропереходу), служать дзеркалами оптичного резонатора (коефіцієнт відбивання ~30 %, рис. 22.12.). Деколи застосовують зовнішні резонатори. Інверсія заселеності досягається при великому прямому струмі через діод унаслідок інжекції надлишкових носіїв у шар, який прилягає до переходу. Генерація когерентного випромінювання виникає у смузі краєвої люмінесценції (випромінювання), якщо оптичне підсилення здатне перевищити втрати енергії, які пов'язані з виходом випромінювання назовні, поглинанням і розсіянням всередині резонатора. Струм, який відповідає початку генерації, називають пороговим. Густина порогового струму в інжекційних напівпровідникових лазерах, як правило, становить ~1 кА/см².

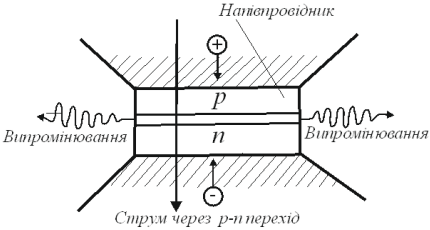


Рис. 22.12. *Схема напівпровідникового лазера з електричною накачкою*

Найбільше поширення отримали напівпровідникові лазери на основі гетероструктур (гетеролазери), які мають найбільш низьку порогову густину струму за температури 300 К. Гетеролазер містить два гетеропереходи, один типу $p-n$, що інжектує електрони (емітер), і другий типу $p-p$, що обмежує дифузійне розтікання носіїв заряду з активного шару – активна область міститься між ними. У смугових лазерах активна область у формі вузької смужки, ширина якої 1... 20 мкм, тягнеться вздовж

осі резонатора від одного дзеркала до другого.

Через малі розміри активної області пороговий струм смугових гетеролазерів досить малий (5 … 100 мА) для отримання неперервної генерації за $T = 300$ К. Потужність випромінювання таких напівпровідникових лазерів (~100 мВт) обмежена перегріванням активної області. В коротких імпульсах напівпровідникові лазери випромінюють більшу потужність (до 100 Вт), яка обмежена оптичним руйнуванням торцевих граней. Багатоелементні інжекційні напівпровідникові лазери дають в імпульсі потужність до 10 кВт. Напівпровідники, з яких можуть бути виготовлені гетеролазери, попри різний хімічний склад повинні мати однаковий період кристалічної ґратки. Використовують багатокомпонентні тверді розчини, серед яких можна знайти неперервні ряди речовин з постійним періодом ґратки (ізоперіодичні системи). Наприклад, у гетеролазері на основі твердих розчинів $Al_xGa_{1-x}As$ гетероструктура складається із шарів (рис. 22.13) $p(Al_xGa_{1-x}As)$; $p(GaAs)$ і $(Al_xGa_{1-x}As)$. У структурі із двома гетеропереходами носії зосереджені всередині активної області ($p-GaAs$ на рис. 22.13, *г*), яка обмежена з обох сторін потенціальними бар'єрами; випромінювання теж обмежене цією областю внаслідок стрибкоподібного зменшення показника заломлення за її межами. Ці обмеження сприяють підсиленню стимульованого випромінювання і, відповідно, зменшенню порогової густини струму.

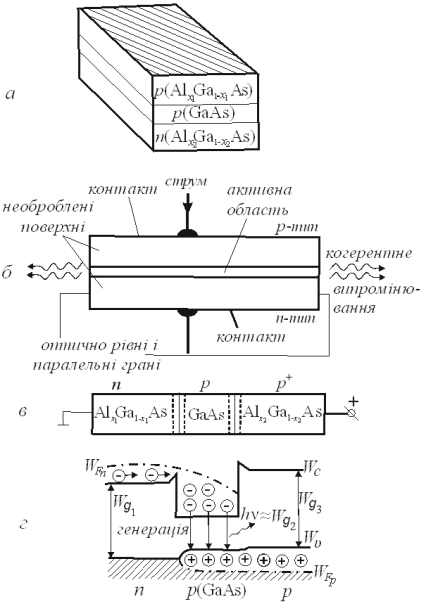


Рис. 22.13. Схема гетеролазера із двосторонньою гетероструктурою на основі AlGaAs (а, б, в) і його енергетичною діаграмою (г): W_c і W_v – краї зони провідності і валентної зони відповідно; W_{F_n} і W_{F_p} – енергія Фермі для електронів і дірок

22.5. Фізичні основи волоконно-оптичних ліній зв'язку

Телефонні розмови, телевізійні програми, комп'ютерна інформація – все це закодовані у певній послідовності ультракороткі лазерні спалахи, які

передаються на великі відстані по дуже тонкому кварцевому волокну. Хоч суто зовнішньо волоконно-оптична лінія зв'язку (ВОЛС) нічим не відрізняється від традиційної провідної: кабель, приймальний і передавальний модулі по її кінцях, ретранслятори (регенератори), які вбудовуються в кабель через певні відстані для відновлення сигналу, який послабшав, комутатори там, де лінія розгалужується, стики, які з'єднують все в єдине ціле (рис. 22.14). Звичайно, «начинка» усіх цих блоків – волокна, лазери, фотодіоди, оптико-електронні перемикачі – надають цим лініям зв'язку нових якостей.

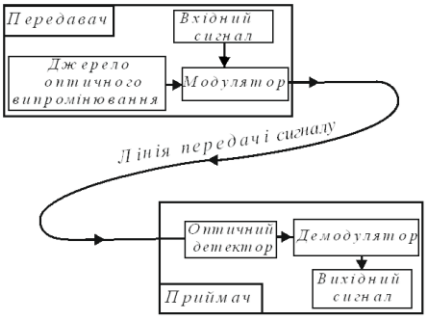


Рис. 22.14. Структурна схема оптичної системи зв'язку

Так з'явився ідеальний засіб для підсвітки важкодоступних місць для ендоскопії, для передавання зображень, для перетворення світлових полів. Волоконна оптика кардинально перетворила багато напрямів приладобудування, медицини, вимірювальної техніки, нею зацікавились спеціалісти із засобів відображення інформації, логіко-математичного аналізу.

Революційну роль волоконної оптики порівнювали з напівпровідниками, які в той період уже зробили крок від транзисторів до мікросхем. Але про оптичний зв'язок чути було мало. Збільшення довжини волокна від одного метра до десяти зменшувала світлопередачу до десятих і сотих часток відсотка. У 1970 р. в одному з випусків Праць американського інституту радіоінженерів навіть було висловлено резюме: оптичний зв'язок не відбувся. Утім, саме в цьому році з'явилось повідомлення американської фірми «Corning Glass» про виготовлення кварцового волокна із такою значною прозорістю, що лазерний сигнал міг пробігти по ньому більше кілометра з невеликим ослабленням. Світловод для оптичного зв'язку був створений – оптичне волокно.

Досвід свідчить, що кожний новий технічний прорив, як правило, ґрунтується на «трьох китах»: «фізика – матеріал – технологія». Фізика світлопередачі оптичним волокном ґрунтується на законах відбивання і заломлення світла.

Розглянемо проходження світлового випромінювання крізь межу поділу двох середовищ. В геометричній

оптиці світлові хвилі зображують променями, які зазнають змін на межі поділу середовищ з різними оптичними властивостями, що характеризуються показником заломлення $n = c/v$, який у загальному випадку показує, у скільки разів швидкість c поширення світла у вакуумі більша від швидкості v поширення світла у розглядуваному середовищі (див. підрозд. 11.3).

Як відомо, середовище, у якого показник заломлення більший, називають оптично більш густим, в іншому випадку – менш густим, тому під час падіння променя світла (світлової хвилі) на межу поділу таких середовищ у загальному випадку виникає відбита і заломлена хвилі (рис. 22.15, *а*). Відповідно до закону Снелліуса кути падіння $\alpha_{пад}$, відбивання $\alpha_{відб}$ і заломлення β пов'язані такими співвідношеннями: $\alpha_{пад} = \alpha_{відб}$;

$$n_1 \sin \alpha_{пад} = n_2 \sin \beta \,. \quad (22.20)$$

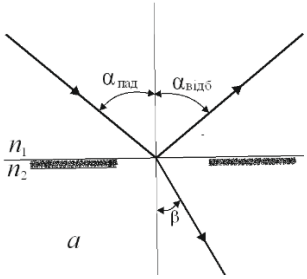


Рис. 22.15, а. Відбивання і заломлення світлового променя на межі двох середовищ: $n_1 < n_2$

Якщо промінь переходить з оптично більш густого середовища у менш густе $n_1 > n_2$ (рис. 22.15, *б*), то $\beta > \alpha_{пад}$. Шляхом збільшення кута падіння можна досягнути випадку, за якого заломлений промінь буде ковзати по межі поділу середовищ, не переходячи у друге середовище (рис. 22.15, *б*). Кут падіння, за якого виникає цей ефект, називають граничним кутом $\alpha_{гр}$ повного внутрішнього відбивання. Очевидно, що для всіх кутів падіння, більших від критичного $(\alpha_{пад} > \alpha_{гр})$, буде спостерігатись тільки відбивання, а заломлення не буде. Це явище називають *повним внутрішнім відбиттям*, й оскільки вся потужність світлового променя майже повністю повертається в область більш густого середовища, на цьому ефекті ґрунтується принцип передачі оптичного випромінювання по оптичному волокну.

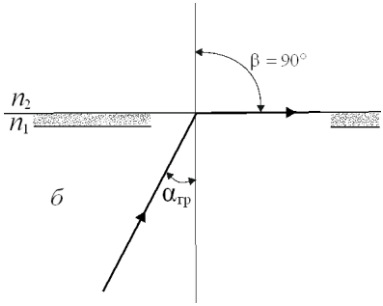


Рис. 22.15, б. Повне внутрішнє відбивання світлового променя на межі двох середовищ $n_1 > n_2$

Поширення світлових променів в оптичних волокнах

У загальному випадку оптичне волокно (рис. 22.16) складається із серцевини, по якій відбувається поширення світлових хвиль, та оболонки, яка призначена, з одного боку, для створення кращих умов відбивання на межі поділу «серцевина – оболонка», а з другого – для зниження випромінювання енергії в навколишній простір. З метою підвищення міцності й тим самим надійності волокна поверх оболонки зазвичай накладають первинне захисне зміцнювальне покриття. Така конструкція оптичного волокна є типовою, її використовують у багатьох оптичних волокнах як базову. Розглянемо тепер структуру світловоду, по якому передається інформація в сучасних оптичних лініях зв'язку. Оптичне волокно складається із серцевини (1 на рис. 22.16) й оболонки (2 на рис. 22.16), які є вісесиметричним діелектричним хвилеводом. Хвилеводна структура утворюється внаслідок зменшення показника заломлення від осі хвилеводу до його оболонки для виконання умови повного внутрішнього відбиття (рис. 22.15, *б*). Якщо зміна показника заломлення відбувається стрибком (рис. 22.17, *а*, $n_2 < n_1$.), то волокна називають ступеневими. Волокна з плавною зміною показника заломлення називають градієнтними. Для таких волокон показник заломлення n залежить від відстані r від осі волокна до його периферії, причому $dn/dr < 0$, тобто зі збільшенням r величина n зменшується (рис. 22.17, *б*).

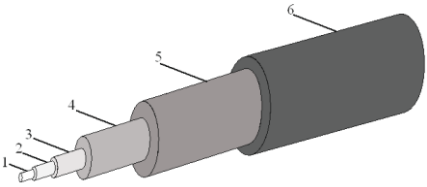


Рис. 22.16. Загальний вигляд типового оптичного волокна: 1 – серцевина; 2 – оболонка; 3 –силіконове покриття; 4 – буферна облонка; 5 – зміцнююча оболонка; 6 – зовнішня оболонка

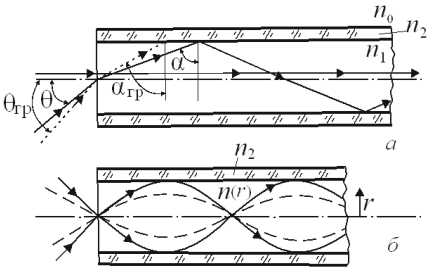


Рис. 22.17 Поширення світла у ступеневому (а) і градієнтному (б) волоконному світловоді

У ступеневому волокні промені зазнають багатократного відбивання від межі між серцевиною та оболонкою. У градієнтному хвилеводі відбивання немає, а відбувається викривлення траєкторії променя внаслідок оптичної неоднорідності середовища (зміна показника заломлення у напрямку, який перпенди-

кулярний осі). У результаті промінь, який перетинає вісь, знову повертається до неї, причому повертання це є плавним. Траєкторія променя у такому хвилеводі задовольняє принцип мінімуму оптичної довжини (принцип Ферма). Із рис. 22.17, *а*, видно, що промені, які поширюються по серцевині ступеневого світловоду під кутом $\alpha > \alpha_{гр}$ до його осі, зазнають повного внутрішнього відбиття на межі поділу «серцевина – оболонка», і тому поширюються вздовж світловоду. Інші промені проникають в оболонку. Граничний кут можна знайти за формулою, яку визначають із співвідношення (22.20):

$$\sin\alpha_{гр} = n_2/n_1.$$

Кожному значенню кута α відповідає значення $\theta -$ кута падіння на торець світловоду (рис. 22.17, *а*). Кут $\theta_{гр}$, за якого промінь буде поширюватись по світловоду під кутом $\alpha_{гр}$, називають вхідним кутом світловоду. Зазвичай його характеризують апертурою $A = n_0 \sin\theta_{гр}$, де $n_0 -$ показник заломлення зовнішнього середовища. Для ступеневого світловоду апертура $A = \sqrt{n_1^2 - n_2^2}$, а для градієнтного справедливі співвідношення $\sin\theta_{кр} = \sqrt{n^2 - n_2^2}$ та $A = \sqrt{n_{max}^2 - n_2^2}$, де $n_{max} -$ найбільше значення показника заломлення серцевини світловоду. У ступеневих світловодах, які використовують у системах зв'язку, апертура дорівнює 0,18 … 0,23, у градієнтних – 0,13 … 0,18.

Променеве наближення справедливе тільки для описання поширення випромінювання у волокнах з поперечним перерізом, набагато більшим від довжини хвилі світла, оскільки у таких світловодах світло поширюється відповідно до законів геометричної оптики. Якщо ж розміри поперечного перерізу волокна можна порівняти з довжиною хвилі випромінювання, то для аналізу поширення випромінювання через такі світловоди треба використати електромагнітну теорію світла і рівняння Максвелла. Строгий розв'язок цих рівнянь показує, що по світловоду малого перерізу одночасно може поширюватись певний дискретний набір електромагнітних хвиль, які називають модами або власними хвилями. Хвилі у серцевині, які відповідають променям, що утворюють малі й великі кути з віссю світловоду, називають *модами нижчих і вищих порядків* відповідно. На рис. 22.18 показано, що частина променів, які увійшли у світловод, поширюється тільки в його серцевині. Це направлені моди *а* та *b*. Частина променів поширюється в оболонці – це моди оболонки *с*. Нарешті

частина променів виходить назовні – це моди *d*, які витікають.

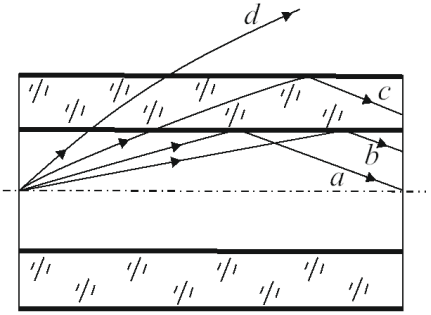


Рис. 22.18. Моди у ступеневому волоконному світловоді

Таким чином, кожна мода є коливанням, яке характеризується певною просторовою структурою електричного і магнітного полів і відповідною фазовою швидкістю. Будь-яке оптичне випромінювання, яке поширюється по світловоду, можна подати як суперпозицію мод, тобто як результат додавання коливань, що поширюються в ньому. Кількість цих мод можна розрахувати, якщо відомі довжина хвилі, радіус серцевини світловоду і різниця показників заломлення серцевини й оболонки.

У волоконних світловодах, товщина яких становить від 1 до 5 мкм, поширюється одна мода – одне коливання. У більш товстих (від 10 до 100 мкм) світловодів може одночасно поширюватись декілька коливань (мод). На практиці використовують як одномодові, так і багатомодові світловоди.

Технологія виготовлення світловодів

Революційність технологічного прориву у створенні надпрозорих волокон пов'язана з вибором кварцу як основного матеріалу. Хімічно кварц – це діоксид кремнію SiO_2, у надчистому вигляді його можна синтезувати методом парафазного осаджування, який розроблений у 1970 р. (рис. 22.19).

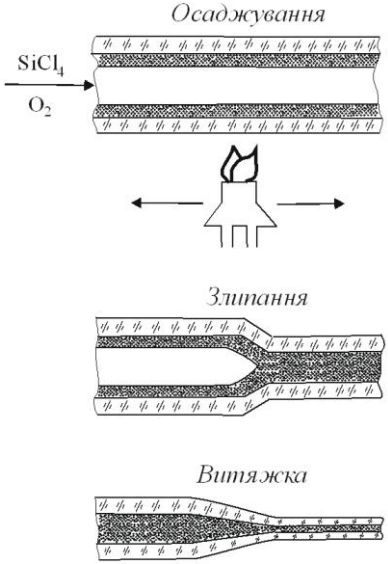


Рис. 22.19. Основні етапи виготовлення кварцового двошарового світловоду

Через розпечену кварцову вихідну трубу пропускають суміш чотирихлористого кремнію $SiCl_4$ і кисню. Під дією високої температури (~1400 °C) у цій суміші відбувається реакція з

утворенням кварцу, який осідає на внутрішній поверхні труби. У процесі осаджування у потік додають газоподібні реактиви, які містять бор, фосфор, германій – ці присадки трохи збільшують або зменшують показник заломлення кварцу. Після отримання шарів заданого складу і товщини трубу нагрівають ще більше до розм'якання і злипання. Отриману таким чином вихідну заготовку потім перетягують через серію *фільєр* (франц. filiere, від fil – волокно, нитка, деталь машин у вигляді ковпачка або пластини з отворами, діаметр яких 0,8 … 0,06 мм, для формування волокон), які послідовно звужуються. Установку для витяжки волокна за розмірами можна порівняти із триповерховим будинком. Отримати волокна з високою прозорістю – для цілей оптичного зв'язку умова необхідна, але недостатня, адже інформація кодується послідовністю імпульсів, отже, треба мати можливість передавати по волокну саме світлові імпульси, і чим коротшими вони будуть, тим вищою буде пропускна здатність каналу зв'язку. Важливим параметром оптичного волокна є дисперсія, яка визначає його пропускну здатність для передавання інформації.

Дисперсія оптичних волокон

Дисперсія – це розсіювання з часом спектральних або модових складових оптичного сигналу, яке призводить до збільшення тривалості імпульса оптичного випромінювання під час поширення його по оптичному волокну (рис. 22.20). Дисперсію визначають різницею квадратів тривалості імпульсів на виході та вході оптичного волокна:

$$\tau = \sqrt{\tau_{вих}^2 - \tau_{вх}^2}, \qquad (22.21)$$

де значення $\tau_{вих}$ і $\tau_{вх}$ визначають на рівні половини амплітуди імпульсів.

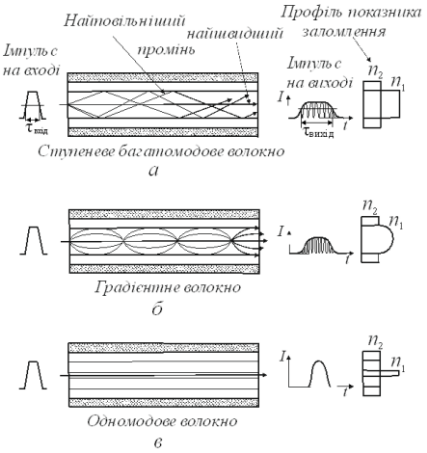


Рис. 22.20. Поширення випромінювання по ступеневому (а), градієнтному багатомодовому (б) та одномодовому оптичному волокну (в)

Як видно з рис. 22.18 та 22.20, *а*, *б*, різні апертурні промені на шляху поширення проходять до вихідного торця різні відстані: досить порівняти промінь, який падає вздовж осі, з променем, який падає під кутом $\theta_{гр}$ (рис. 22.17, *а*). Це призводить до розмиття імпульсу – із прямокутного на вході він стає

дзвоноподібним, розмазаним на виході (рис. 22.20, *а*), Сусідні імпульси набігають один на одний, виникають помилки, збої, а то й повна втрата інформації.

До цієї *хвилевідної* (*міжмодової*) *дисперсії* додається ще і звичайна «*хроматична*» *дисперсія* – розмиття імпульсу світла, що обумовлене залежністю фазової швидкості $v_\phi = c/(n(\lambda))$ і групової швидкості $v_{гр} = v_\phi - \lambda\frac{dv_\phi}{d\lambda} = v_\phi\left(1+\frac{\lambda}{n}\frac{dn}{d\lambda}\right)$ (див. підрозд. 15.2) світла у кварці від довжини хвилі (рис. 22.20, *б*).

Причиною міжмодової дисперсії є некогерентність джерел випромінювання, які реально працюють у спектрі довжини хвиль від λ до $\lambda + \Delta\lambda$. Міжмодова дисперсія переважає у багатомодових оптичних волокнах. Усе це в сукупності призвело до того, що в перших волокнах 70-х років XX ст. при використанні тодішніх не зовсім досконалих лазерів розмиття імпульсу характеризувалось дисперсією $\tau \sim 40$ нс/км. Це означає, що десятикілометрова лінія з такого волокна придатна для одночасного передавання тільки декількох телефонних розмов. Міжмодова дисперсія у градієнтних оптичних волокнах зазвичай на порядок нижча, ніж у ступеневих волокнах. Це обумовлено тим, що внаслідок зменшення показника заломлення від осі оптичного волокна (рис. 22.20, *б*) до оболонки швидкість $(v = c/n)$ поширення променів вздовж їхньої траєкторії змінюється. Так, на траєкторіях, які близькі до осі, швидкість менша (більший n), а на віддалених траєкторіях – більша (менший n). Отже промені, які поширюються найкоротшими траєкторіями, що близькі до осі, мають меншу швидкість, а промені, які поширюються по більш довгих траєкторіях, мають більшу швидкість. У результаті час поширення променів вирівнюється і збільшення тривалості імпульсу стає меншим. У 1973 р. було встановлено, що випромінювання з довжиною хвиль $\lambda = 1{,}3 \ldots 1{,}5$ мкм майже не зазнають дисперсії у кварці. Світлопропускання у цій області спектру набагато вище, ніж за $\lambda = 0{,}85$ мкм, а саме на такій довжині хвилі працювали тогочасні лазери. Тому і постало питання про перехід у нову спектральну область, що вимагало створення нових джерел і приймачів світла. Для боротьби із хвилеводною (модовою) дисперсією довелось звернутись до одномодових волокон (рис. 22.20, *в*) з діаметром серцевини близько 5 мкм. Саме вони і становлять основу сучасного високошвидкісного волоконно-оптичного зв'язку. Для створення джерел світла для ВОЛЗ

було синтезовано новий клас напівпровідникових сполук (підрозд. 22.2), які включають чотири компоненти: індій In, галій Ga, миш'як As і фосфор P. Варіюючи склад, отримують потрібну довжину хвилі випромінювання лазера.

Приймачі оптичного випромінювання

Одним із базових компонентів ВОЛЗ, за допомогою яких здійснюють реєстрацію оптичного випромінювання, є приймачі оптичного випромінювання. Останні виготовлені на основі фоточутливих напівпровідникових елементів, що використовують явище фотопровідності (внутрішнього фотоефекту, підрозд. 19.11). Принцип роботи фотодіодів розглянуто у підрозд. 21.4.

На сьогоднішній день найбільш поширеним типом детектора оптичного випромінювання є $p{-}i{-}n$-структура фотодіода – **PIN** фотодіод (рис. 22.21). Основною відмінністю цього типу діодів від звичайних діодів з *p*–*n*-переходами є наявність між сильно легованими p^+ і n^+ шарами напівпровідника *i*-шару, який є слабо легованим напівпровідником *n*-типу, товщина якого становить декілька десятків мікрометрів (рис. 22.21, *а*).

В *i*-шарі вільних носіїв заряду немає, тому його деколи називають збідненим шаром. Оскільки сильне легування збільшує провідність шарів p^+ і n^+, то зовнішня різниця потенціалів *U*, яка прикладена до них, створює в *i*-шарі градієнт електричного поля (рис. 22.21, *б*). Під час подачі на *p*–*i*–*n*-структуру зворотної напруги, то через те, що в *i*-шарі немає вільних носіїв, цей шар поляризується, і через опір навантаження протікає малий постійний струм I_m, який називають темновим струмом. Значення цього струму визначається параметрами матеріалу напівпровідника, топологією *p*–*i*–*n*-структури і температурою навколишнього середовища.

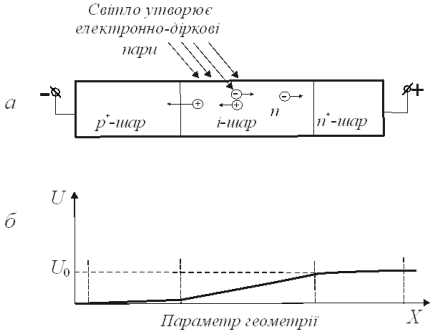


Рис. 22.21. Структура включення і розподіл потенціалу p–i–n-фотодіода

Якщо впливати на *i*-шар зовнішнім випромінюванням, у ньому утворюються вільні електрон-діркові пари, які швидко розділяються і, прискорюючись прикладеним електричним полем, рухаються до електродів у протилежних напрямках, створюючи у зовнішньому колі діода електричний струм $I_ф$ (фотострум). PIN-діоди на

велику довжину хвилі (1,3... 1,5мкм) виготовляють із Ge, InGaAs, InGaAsP.

Більш високу чутливість, ніж PIN-діоди, мають лавинні фотодіодні детектори **APD** (avalanche photon detector). Основна відмінність APD-фотодіодів від PIN-діодів, які мають $p^+ - i - n^+$-структуру, полягає в тому, що вони містять додатковий *p*-шар (рис. 22.22, *а*). При цьому створюється $p^+ - i - p - n^+$-структура з таким профілем розподілу легувальних домішок, за якого найбільша напруженість електричного поля спостерігаєжться в *p*-шарі (рис. 22.22, *б*), що забезпечує найменшу провідність цього шару. З цієї причини зворотне зміщення в APD-діодах настільке велике, що збіднений шар збільшується до розмірів *i*-шару не знижуючи, а навіть збільшуючи напруженість електричного поля. У зв'язку із цим електронно-діркові пари, які створюються світлом, розділяються і прискорюються цим полем у збідненому шарі, отримуючи енергію, достатню для іонізації атомів кристала. Зіштовхуючись з нейтральними атомами, носії викликають зростаючу в геометричній прогресії концентрацію електронів і дірок – лавинний ефект (рис. 22.22, *а*). Цей процес називають помноженням первинного фотоструму.

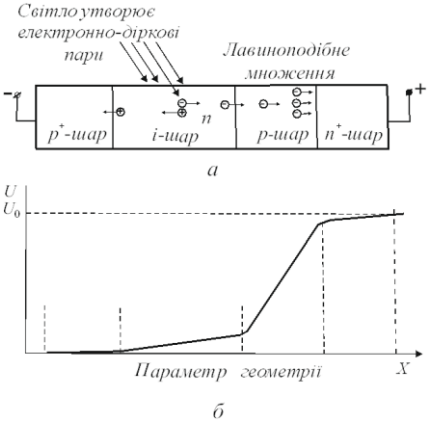


Рис. 22.22. Структура включення і розподіл потенціалу APD-фотодіода

Нині використовують коротко-хвильові Si APD-діоди та довгохвильові InGaAs, Ge, APD, InGaAsP /InP та інші.

Введення і виведення світла у світловодах

Для введення випромінювання у світловод і виведення світла на фотодіод використовують лінзи. Найбільш ефективні лінзи з малою фокусною відстанню, однак у таких лінз великі спотворення. У зв'язку із цим замість звичайних лінз застосовують градієнтні лінзи, в яких фокусна відстань зменшується не внаслідок геометрії, а внаслідок використання матеріалів, в яких показник заломлення плавно змінюється. У градієнтних стрижневих лінзах, які називають граданами або *селфоками,* промінь, що ввійшов у лінзу,

поширюється криволінійною траєкторією.

На рис. 22.23 показано поширення променів у градієнтній стрижневій лінзі, в якій показник заломлення змінюється за параболічним законом. Залежно від довжини градани можуть бути багато-, одно-, напів- і чверть кроковими.

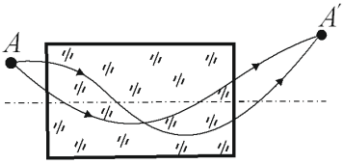


Рис. 22.23. Поширення світла у градані

Напів- і чвертькрокові градани зображено на рис. 22.24. Градієнтні стрижневі лінзи використовують для з'єднання окремих волокон, джерела або приймача випромінювання з волокном, у різних розгалужувачах світлових потоків, оптичних перемикачах і под.

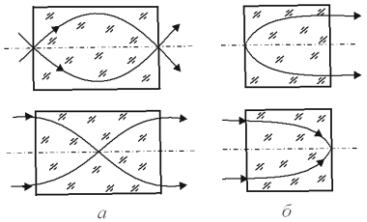


Рис. 22.24. Напівкрові (а) і чвертькрокові (б) градани

Найбільш часто застосовують чвертькрокові градієнтні стрижневі лінзи, які, як видно з рис. 22.24, дозволяють перетворити розбіжний пучок у паралельний, або навпаки, – сфокусувати паралельний пучок.

На рис. 22.25 показано з'єднання двох світловодів 1 за допомогою чвертькрокової градієнтної стрижневої лінзи 2. Якщо їх розміщення співвісне, світло з одного волокна потрапляє у друге. Розмістивши між градієнтними лінзами світлофільтр або послаблювач потужності 3, можна отримати елемент, який називають атенюатором.

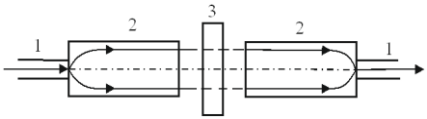


Рис. 22.25. З'єднання двох світловодів за допомогою граданів

На рис. 22.26 показано схему спрямованого відгалужувача. У ньому між градієнтними лінзами 2 розміщене напівпрозоре дзеркало 3, яке частину випромінювання пропускає, а частину направляє у лінзу 4, що пов'язана ще з одним світловодом. Втрати світла у таких розгалужувачах мінімальні.

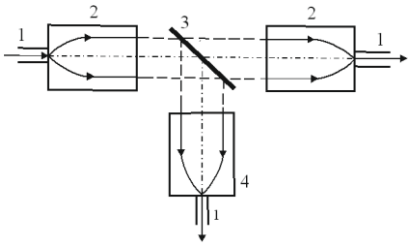


Рис. 22.26. Спрямований відгалужувач на граданах

Переваги ВОЛЗ

Розглянемо переваги ВОЛЗ.

Перше – дуже мале загасання сигналу у волокні:

$$B = \frac{10}{L}\lg(P_{вх}/P_{вих}),$$

де B – величина загасання в децибелах на кілометр; $P_{вх}$ і $P_{вих}$ – вхідна і вихідна потужність сигналу; L – довжина світловоду. На сьогоднішній день створені світловоди із втратами 0,5... 1 дБ/км. Це означає, що потужність сигналу зменшується у два рази на відстані $L \sim 6$ км, тобто відстань між ретрансляторами може досягати 100 ... 200 км, що набагато перевищує той самий показник у провідних лініях. Але і це не межа – теоретичні оцінки показують, що у спектральній області $\lambda \simeq 2 \ldots 2{,}5$ мкм можна досягнути передавання сигнала без ретранслятора на 1000 і більше кілометрів. Звичайно, для цього доведеться замінити і кварц, і лазери, і фотодіоди.

Друге – це надвисока пропускна здатність: швидкість передачі десятки (Мбіт/с) вже давно перестала бути рекордною, а типові значення складають сотні і тисячі (Мбіт/с). Виявилось, що якщо одночасно пропускати по волокну потоки світлових імпульсів з дещо відмінними довжинами хвиль, то вони не змішуються один з одним і на кінці приймання можуть бути розділені. Завдяки цьому можна збільшити в десятки, а то і в сотні разів пропускну здатність вже прокладених комунікацій. Але й це ще не все. Якщо вводити у волокно лазерні імпульси підвищеної потужності, то утворюються солітони (поодинокі хвилі), які під час поширення не тільки не «розлазяться», але можуть навіть звужуватись. Швидкість передавання інформації може досягати 1 Тбіт/с. Сьогодні лабораторні досліди дають швидкість передавання інформації по оптоволокну 1,125 Тбіт/с – це у 50 тис. разів перевищує середню швидкість підключення приватних користувачів до інтернету. За таких умов передавання інформації, щоб отримати всі дані, зібрані космічним телескопом Хаббл за 20 років спостереження, потрібно трохи більше, ніж півхвилини.

Третє – переносником інформації у ВОЛЗ є фотон, який не має електричного заряду, тому світловий сигнал, який поширюється волокном, зовсім не чутливий до зовнішніх електромагнітних впливів: розряди блискавок, численні радіопередавачі, електродвигуни, силові ядерні установки і т. п. Волоконним лініям непотрібне важке і дороге екранування, вони захищені від завад самим принципом своєї роботи. Традиційна радіоапаратура не тільки піддається впливу

електромагнітних хвиль, вона і сама їх випромінює (носієм інформації є заряджена частинка – електрон, нерівномірний рух якого супроводжується випромінюванням електромагнітних хвиль). На цій основі й виник радіошпіонаж. І лише ВОЛЗ, в яких втаємниченість передавання повідомлень закладена у самій фізиці, дозволяє гарантовано убезпечитись від несанкціонованого доступу до інформації.

Четверте – волоконні кабелі легкі й компактні, у десятки і сотні разів легші й дешевші, ніж металічні. Основу металічних кабелів становлять мідь і свинець – не тільки дуже дорогі, але й досить рідкісні в земній корі. Волокно виготовляють із кварцу, а він утворений зі звичайного піску, запаси якого величезні. Один з апологетів волоконного зв'язку за важливістю зіставляє його з винайденням парової машини, електричної лампочки чи транзистора.

Контрольні запитання

1. Назвіть фізичні принципи, які лежать в основі роботи будь-якого квантового генератора.

2. Як можна пояснити необхідність існування квантового переходу (вимушеного випромінювання), поняття про який вперше ввів
А. Ейнштейн?

3. У чому полягає суть ідеї В. Фабриканта щодо можливості підсилення випромінювання, яке проходить крізь певне середовище?

4. Чи можна здійснити підсилення випромінювання в середовищі, яке перебуває у рівноважному стані? Відповідь пояснити.

5. Як конкретно реалізуються принципи, які лежать в основі роботи будь-якого квантового генератора, в лазері (на прикладі гелій-неонового лазера)?

6. Поясніть трирівневу енергетичну схему створення інверсної населеності, яка реалізована в гелій-неоновому, рубіновому лазері та інших. Чим вона ефективніша від дворівневої схеми?

7. Як підсилюється випромінювання в активному середовищі?

8. Яке призначення мають дзеркала в оптичному генераторі?

9. Які властивості притаманні лазерному випромінюванню? Чому саме тільки лазерному?

10. Опишіть, як застосовують ОКГ?

11. Які фізичні основи волоконного зв'язку?

12. Які є типи оптичних волокон?

13. Які є типи дисперсії в оптичних волокнах? Як дисперсія впливає на передавання інформації?

14. Які оптичні волокна є одномодовими, а які багатомодовими?

15. Які є джерела світла і фотоприймачі в оптоволоконних лініях зв'язку?

16. Який тип лінз використовують в оптоволоконних лініях зв'язку?

Розділ 23. Атомні ядра

23.1. Атомні ядра та їх характеристики

Атомне ядро, або центральна масивна частина атома, складається із протонів і нейтронів, які називають нуклонами (від лат. nucleus – ядро). Маса ядра атома приблизно у $4 \cdot 10^3$ разів більша від маси всіх електронів, що входить до складу атома. Розміри атомного ядра становлять $\sim 10^{-14}$ ... 10^{-15} м. Електричний заряд ядра позитивний і за абсолютною величиною дорівнює сумі зарядів атомних електронів нейтрального атома. Атомне ядро відкрив Резерфорд (див. підрозд. 17.1) у дослідах з розсіювання α-частинок під час проходження їх крізь речовину. У кінці 1913 р. учень Резерфорда Г. Мозлі показав експериментально, що електричний заряд ядра атома (в одиницях абсолютної величини заряду електрона e) дорівнює порядковому номеру Z елемента у періодичній системі елементів. Після Мослі факт існування атомного ядра утвердився у фізиці остаточно.

В ядерній фізиці масу виражають в атомних одиницях маси (а.о.м.) Одна а.о.м. дорівнює $1{,}66 \cdot 10^{-27}$ кг. Частіше массу виражають в одиницях енергії, для цього одну а.о.м. помножують на c^2 і виражають у мільйонах електронвольт (Мев), де c – швидкість світла у вакуумі. Одній а.о.м відповідає 931,2 МеВ. Маса спокою електрона (в Мев)

$$m_{\text{МеВ}} = \frac{m_e c^2}{e} = \frac{9{,}1 \cdot 10^{-31} \cdot (3 \cdot 10^8)^2}{1{,}6 \cdot 10^{-19}}\ \text{МеВ} = 0{,}51\ \text{МеВ}. \quad (23.1)$$

Протон. Протон p має позитивний заряд, що дорівнює величині заряду електрона, і масу спокою

$$m_p = 1{,}672 \cdot 10^{-27}\text{кг} = 1836 m_e = 938{,}26\ \text{МеВ}.$$

Протон має спін, що дорівнює половині ($S = 1/2$) і власний магнітний момент $\mu_p = +2{,}79\ \mu_я$, де

$$\mu_я = \frac{e\hbar}{2m_p} = 5{,}05 \cdot 10^{-27}\ \text{Дж/Тл} \quad (23.2)$$

є одиницею магнітного момента, яку називають *ядерним магнетоном*, що у 1836 разів менший від магнетона Бора $\mu_Б$ (див. підрозд. 17.5). Отже, власний магнітний момент протона приблизно у 660 разів менший, ніж магнітний момент електрона.

Нейтрон. Нейтрон n відкрив у 1932 р. англійський фізик Д. Чедвик. Електричний заряд цієї частинки дорівнює нулю, а маса

$$m_n = 1{,}675 \cdot 10^{-27}\ \text{кг} = 1839\, m_e = 939{,}55\ \text{МеВ} \quad (23.3)$$

дуже близька до маси протона. Різниця мас нейтрона і протона $m_n - m_p$ становить 1,3 МеВ, тобто 2,5 m_e. Нейтрон має спін, який дорівнює половині ($S = 1/2$), і попри відсутність електричного заряду власний магнітний момент

$$\mu_n = -1{,}91\mu_я,$$

де знак мінус вказує на те, що напрямки власних механічного і магнітного моментів нейтрона і протона протилежні.

У вільному стані нейтрон нестабільний (радіоактивний) – він самовільно розпадається, перетворюючись у протон з випусканням електрона (e^-) і ще однієї частинки, яку називають антинейтрино ($\tilde{\nu}$).

Характеристики атомного ядра

Опишемо характеристики атомного ядра.

1. Зарядове число Z дорівнює кількості протонів, які входять у склад ядра, і визначає його заряд, який дорівнює $+Ze$. Число Z визначає порядковий номер хімічного елемента в періодичній таблиці, тому його також називають *атомним номером ядра.*

2. Масове число A – число нуклонів (тобто сумарна кількість протонів і нейтронів) в ядрі. Число нейтронів в ядрі $N = A - Z$.

3. Для позначення ядер застосовують символ

$$_ZX^A \text{ або } {}^{A}_{Z}X,$$

де під X розуміють хімічний символ певного елемента.

Ядра утворюють такі групи:

1. *Ізотопи.* Ядра з однаковими Z, але різними A. Більшість хімічних елементів має по декілька стабільних ізотопів. Наприклад, у кисню є три стабільних ізотопи ${}^{16}_{8}O$, ${}^{17}_{8}O$, ${}^{18}_{8}O$, в олова – десять і т.д. Водень має три ізотопи: ${}^{1}_{1}H$ – звичайний водень, або протій ($Z = 1$, $N = 0$), ${}^{2}_{1}H$ – важкий водень, або дейтерій ($Z = 1$, $N = 1$), ${}^{3}_{1}H$ –тритій ($Z = 1$, $N = 2$). Протій і дейтерій стабільні, тритій радіоактивний.

2. *Ізобари.* Ядра з однаковим масовим числом A, наприклад ${}^{40}_{18}Ar$ і ${}^{40}_{20}Ca$.

3. *Ізотони.* Ядра з однаковою кількістю нейтронів $N = A - Z$, наприклад ${}^{13}_{6}C$, ${}^{14}_{7}N$.

4. *Ізомери.* Радіоактивні ядра з однаковими Z та A, але з різним періодом напіврозпаду. Приміром, існують два ізомери ядра ${}^{80}_{35}Ba$, в одного з них період напіврозпаду дорівнює 18 хв, в іншого – 4,4 год.

Відомо близько 1500 ядер, які відрізняються або Z, або A, або й тим, і іншим. Приблизно 1/5 частина цих ядер стійкі, інші радіоактивні. Багато ядер

було отримано штучним шляхом за допомогою ядерних реакцій.

У природі трапляються елементи з атомними номерами Z від 1 до 92, крім технецію (Tc, $Z = 43$) і прометію (Pm, $Z = 61$).

Розміри ядра. З експериментів з використанням заряджених частинок, які розсіювались на ядрах, отримали висновок, що ядро можна вважати кулею, радіус якої може бути виражений формулою

$$R = r_0 A^{1/3}, \qquad (23.4)$$

де A – масове число певного ядра, $r_0 = (1{,}3 \ldots 1{,}7) \cdot 10^{-15}$ м = (1,3... 1,7) Ф. ($\Ф = 10^{-15}$м – Фермі – назва в ядерній фізиці одиниці довжини). У задачах і прикладах беруть компромісне значення $r_0 = 1{,}3$ Ф.

Густину маси ядра можна розрахувати за формулою

$$\rho_{\text{ядра}} = \frac{\text{маса ядра}}{\text{об'єм ядра}},$$

де масса ядра приблизно дорівнює A (а.о.м), а об'єм

$$\frac{4\pi R^3}{3} = \frac{4\pi\left(r_0 A^{1/3}\right)^3}{3} = \frac{4}{3}\pi r_0^3 A.$$

Тоді густина ядерної речовини

$$\rho_{\text{ядра}} = \frac{3 \cdot \text{а.о.м.}}{(4\pi r_0^3)} \approx 2 \cdot 10^{17} \text{ кг/м}^3.$$

Це неймовірно велика густина ядерної речовини, яка приблизно однакова для всіх ядер, оскільки не залежить від масового числа A.

23.2. Ядерні сили

Згідно із законом Кулона позитивно заряджені протони, які перебувають на дуже близьких відстанях в атомному ядрі, сильно відштовхуються один від одного так, що вони мали б миттєво розлетітись, тому не можна пояснити стабільне існування атомного ядра, якщо не зробити припущення, що нуклони в ньому піддаються впливу потужних ядерних сил притягання. Ці ядерні сили є проявом сильної взаємодії.

Розглянемо *основні властивості ядерних сил.*

1. Ядерні сили є силами притягання.

2. Ядерні сили досить значні, тому таку взаємодію називають сильною. Енергія зв'язку, що припадає на один нуклон у ядрі, досягає 7... 8,5 МеВ. Сили взаємодії ядра з електронами атома забезпечують енергію зв'язку від десятків до тисяч електрон-вольтів, а сили зв'язку між атомами в молекулах забезпечують енергію зв'язку в кілька електрон-вольтів.

3. Ядерні сили є короткодіючими, на відміну від електромагнітних і гравітаційних сил. Радіус r дії ядерних сил приблизно дорівнює розміру нуклона (10^{-15} м). За $r > 10^{-14}$ м ядерні сили фактично дорівнюють нулю.

4. Ядерні сили мають властивість зарядової незалежності, тобто величина

ядерних сил не залежить від електричного заряду взаємодіючих нуклонів. Ядерні взаємодії між двома протонами або між двома нейтронами, або між протоном і нейтроном, однакові.

5. Ядерні сили мають властивість насичення, тобто кожний нуклон у ядрі взаємодіє лише з обмеженою кількістю найближчих до нього нуклонів. Насичення полягає у тому, що питома енергія зв'язку нуклонів у ядрі зі збільшенням кількості нуклонів не зростає, а залишається приблизно сталою (якщо не враховувати легких ядер).

Із властивостей насичення ядерних сил можна зробити такий важливий висновок: взаємодія між нуклонами не зводиться тільки до сил притягання. Якщо відстань між нуклонами $r < 0{,}5 \cdot 10^{-15}$ м, притягання між ними переходить у відштовхування. Якщо у ядрі немає сил відштовхування, між нуклонами ядра мали б колапсувати, тобто стискатись у точку.

6. Ядерні сили залежать від *орієнтації спінів нуклонів*, які взаємодіють. Система із протона і нейтрона утворює «ядро – дейтрон» лише у випадку, коли спіни протона і нейтрона паралельні. Якщо ж спіни протилежно напрямлені, то нейтрон і протон ядра не утворюють.

7. Величина ядерних сил також залежить від взаємної орієнтації спіна та орбітального моменту кожного нуклона.

8. Ядерні сили *не є центральними силами*, тобто силами, які можна уявити такими, що діють вздовж лінії, яка з'єднує центри взаємодійних нуклонів.

Щоб пояснити властивості ядерних сил – насичення і короткодію – В. Гайзенберг висунув гіпотезу про те, що ядерні сили є «обмінними силами», тобто ядерні сили між двома нуклонами забезпечуються третьою частинкою.

Довжина хвилі $\lambda_{\text{о.ч.}}$ обмінної частинки має відповідати радіусу $r_{\text{я}}$ дії ядерних сил. Виходячи з формули де Бройля (18.5)

$$\lambda_{\text{о.ч.}} \approx \frac{h}{m_{\text{о.ч.}} c} = r_{\text{я}} \leq 10^{-14} \text{ м}$$

можна оцінити масу обмінної частинки:

$$m_{\text{о.ч.}} \approx \frac{h}{r_{\text{я}} c} = 243\, m_e.$$

Ця маса має у 300 разів перевищувати масу електрона.

У 1935 р. японський фізик Х. Юкава висунув гіпотезу про те, що ядерні сили зумовлені невідомою на той час частинкою, маса якої становить 200... 300 мас електрона. Оскільки ці гіпотетичні частинки за величиною маси посідали проміжне місце між масою електрона і протона, то їх назвали мезонами (від грец. mezos – середній, проміжний). Такі частинки справді було

виявлено у 1947 р. у космічному випромінюванні. Виявилось, що існує три типи π-мезонів: π^+, π^-, π^0, Заряди π^+ і π^- за абсолютною величиною дорівнюють заряду електрона. Маса зарядженого мезона $m_{\pi^\pm} = 273m_e$, маса π^0-мезона $m_{\pi^0} = 264m_e$. Спін π-мезонів дорівнює нулю.

Ядерну взаємодію двох нуклонів, що містяться на відстані радіуса дії ядерних сил, можна подати у такий спосіб. Один нуклон випускає π-мезон, а другий нуклон його поглинає упродовж 10^{-23} с (час прольоту частинки зі швидкістю світла відстані між нуклонами – ядерний час). Частинки, які існують лише в області дії ядерних сил протягом ядерного часу, називають *віртуальними*.

Розглянемо обмінну взаємодію між нуклонами. У результаті віртуальних процесів

$$p \Leftrightarrow n + \pi^+, \quad n \Leftrightarrow p + \pi^-,$$

$$p \Leftrightarrow p + \pi^0, \quad n \Leftrightarrow n + \pi^0.$$

Нуклон наче оточений хмарою віртуальних π-мезонів, які утворюють поле ядерних сил. Поглинання цих мезонів іншим нуклоном приводить до сильної взаємодії між нуклонами, яка здійснюється за однією з таких схем:

Схема 1:

$$p + n \Leftrightarrow n + \pi^+ + n \Leftrightarrow n + p.$$

Протон випускає віртуальний π^+-мезон і перетворюється в нейтрон. Мезон поглинається нейтроном, який унаслідок цього перетворюється у протон. Потім такий процес відбувається у зворотному напрямку. Кожний із взаємодійних нуклонів частину часу проводить у зарядженому стані, а частину – у нейтральному.

Схема 2:

$$n + p \Leftrightarrow p + \pi^- + p \Leftrightarrow p + n.$$

Нейтрон і протон обмінюються π^--мезонами.

Схема 3:

$$p + p \Leftrightarrow p + \pi^0 + p \Leftrightarrow p + p.$$

Схема 4:

$$n + n \Leftrightarrow n + \pi^0 + n \Leftrightarrow n + n.$$

Два протони або два нейтрони обмінюються нейтральними π^0-мезонами. Перша із цих чотирьох схем обміну π-мезонами нуклонів експериментально була підтверджена під час вивчення розсіяння нейтронів на протонах. Безперечно, процес взаємодії між нуклонами значно складніший, і наведені схеми не що інше, як намагання унаочнити дії обмінних сил. Обмінні сили – це суто квантово-механічне поняття.

23.3. Енергія зв'язку ядра

Енергія зв'язку ядра $W_{зв}$ – це енергія, яку треба затратити, щоб розщепити ядро на окремі нуклони. Маса ядра $m_я$ завжди менша від суми мас частинок, які входять в його склад. Це

обумовлено тим, що під час об’єднання нуклонів в ядро виділяється енергія зв’язку нуклонів один з одним.

Енергія спокою частинки пов’язана з її масою співвідношенням $W_0 = mc^2$. Значить, енергія ядра, яке перебуває у стані спокою, менша від сумарної енергії нуклонів, які перебувають у стані спокою і не взаємодіють між собою на величину

$$W_{зв} = \left(Zm_p + (A - Z)m_n - m_я\right)c^2. \quad (23.5)$$

Ця величина і є енергією зв’язку нуклонів в ядрі.

Співвідношення (23.5) фактично не порушується, якщо вважати, що $m_p = m_н - m_e$, а $m_я = m_a - Zm_e$, де $m_н$ – маса атома водню; m_a – маса атома, який відповідає певному ядру; m_e – маса електрона. Отже, формулу (23.5) можна переписати так:

$$W_{зв} = (Zm_н + (A - Z)m_n - m_a)c^2. \quad (23.6)$$

Формула (23.6) зручна тим, що в таблицях зазвичай наводять не маси ядер, а маси атомів.

Величину

$$\Delta m = Zm_н + (A - Z)m_n - m_a \quad (23.7)$$

називають *дефектом маси* ядра. Якщо дефект маси виразити в а.о.м., то енергію зв’язку знаходять за формулою

$$W_{зв} = \Delta m(\text{а.о.м.}) \cdot 931{,}2 \text{ МеВ}. \quad (23.8)$$

Енергію зв’язку, яка припадає на один нуклон, тобто

$$W_{пит} = \frac{W_{зв}}{A}, \quad (23.9)$$

називають *питомою енергією* зв’язку нуклонів у ядрі.

Якщо $W_{зв} > 0$, то ядро стабільне (щоб розщепити його на складові частини, йому треба надати енергію ззовні). Якщо $W_{зв} < 0$, то ядро нестабільне (буде розпадатись само по собі, або, як говорять, спонтанно).

Вирахуємо питому енергію зв’язку ядра $^{16}_{8}O$. Скористаємось довідниковими даними.

Розв’язок:

$$Zm_н = 8 \cdot 1{,}007\,825 \text{ а.о.м.} = 8{,}062\,600 \text{ а.о.м.},$$

$$Nm_n = 8 \cdot 1{,}008\,665 \text{ а.о.м.} = 8{,}069\,320 \text{ а.о.м.}$$

Маса атома $^{16}_{8}O$ дорівнює 16,000 000 а.о.м., і тому дефект маси $\Delta =$ $= +0{,}131\,920$ а.о.м. Тоді енергія зв’язку $W_{зв} = 0{,}131\,920 \times 931{,}2$ МеВ $=$ $= 122{,}84$ МеВ, а питома енергія зв’язку $W_{пит} = 122{,}84/18 = 7{,}68$ МеВ/нуклон.

Для порівняння зауважимо, що енергія зв’язку валентних електронів в атомах має величину близько 10 еВ, що в 10^6 разів менше від питомої енергії зв’язку ядер. На рис. 23.1 зображено графік, який показує залежність питомої енергії зв’язку $W_{зв}/A$ від масового числа A. Сильніше за все пов’язані нуклони в ядрах з масовими числами 50... 60 (тобто для елементів від Cr до Zn). Питома

енергія зв’язку для цих ядер досягає 8,7 МеВ/нуклон. За винятком декількох нерегулярностей, які спостерігаються для ядер $^{4}_{2}\mathrm{He}$, $^{12}_{6}\mathrm{C}$, $^{16}_{8}\mathrm{O}$, крива на рис. 23.1 порівняно плавна. Вона різко зростає за малих значень A і вирівнюється за $A \geq 30$. Зі зростанням A величина $W_{\text{пит}}$ поступово зменшується; для найважчого природного елемента урану вона становить 7,5 МеВ/нуклон. Така залежність $W_{\text{пит}}$ від A пояснюється таким чином. За малих A нуклони на поверхні ядра не з усіх сторін оточені іншими нуклонами, що зменшує ефективність взаємодії із сусідніми нуклонами. Виникає поверхневий ефект, подібний до поверхневого натягу в рідинах.

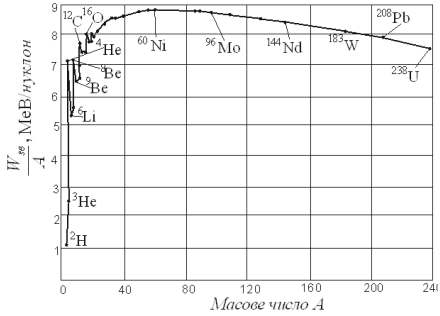


Рис. 23.1. Залежність питомої енергії зв’язку від масового числа

Ще один ефект, який зменшує енергію зв’язку ядер, – це кулонівське електростатичне відштовхування між протонами. Цей ефект особливо сильно проявляється для ядер з великим масовим числом A. Зв’язок між нуклонами стає слабшим, а ядра менш міцними.

Така залежність питомої енергії зв’язку ядра від масового числа робить енергетично можливими два процеси:

1) поділ важких ядер на декілька більш легких ядер;

2) злиття (синтез) легких ядер в одне ядро. Обидва ці процеси повинні супроводжуватись виділенням великої кількості енергії. Так, наприклад, поділ одного ядра з масовим числом A = 240 (питома енергія зв’язку дорівнює 7,5 МеВ/нуклон) на два ядра з масовими числами A = 120 (питома енергія зв’язку дорівнює 8,5 МеВ/нуклон) призвів би до виділення енергії в 240 МеВ. Злиття двох ядер важкого водню $^{2}_{1}\mathrm{H}$ в ядро гелію $^{4}_{2}\mathrm{He}$ призвело би до виділення енергії 24 МеВ. Для порівняння скажемо, що у разі з’єднанні одного атома вуглецю із двома атомами кисню (згорання вугілля до CO_2) виділяється енергія близько 5 еВ.

Ядра із значеннями масового числа A від 50 до 60 є енергетично найбільш вигідними. У зв’язку з цим виникає питання: чому ядра з іншими значеннями A є стабільними? Відповідь полягає в такому. Щоб розділитись на декілька частин, важке ядро має пройти через ряд проміжних станів, енергія яких перевищує енергію основного стану ядра. Отже, для процесу ділення ядру потрібна додаткова енергія (*енергія активації*), яка

потім повертається назад, додаючись до енергії, яка виділяється під час поділу ядра внаслідок зміни енергії зв'язку. У звичайних умовах ядру не має звідки взяти енергію активації, унаслідок чого важкі ядра не зазнають спонтанного поділу. Енергія активації може бути надана важкому ядру додатковим нейтроном, який захоплюється ним. Процес поділу ядер урану або плутонію під дією нейтронів, які захоплюються ядрами, лежить в основі роботи ядерних реакторів і звичайної атомної бомби.

Щодо легких ядер, то злиття їх в одне ядро можливе, якщо вони підійдуть одне до одного на досить близьку відстань ($\sim 10^{-15}$ м $= 1$ Ф). Такому зближенню ядер заважає кулонівське відштовхування між ними. Щоб подолати це відштовхування, ядра мають рухатись з великими швидкостями, які відповідають температурам у декілька сотень мільйонів кельвінів. З цієї причини процес синтезу легких ядер називають *термоядерною реакцією*. Термоядерні реакції протікають у надрах Сонця і зірок. У земних умовах поки що були здійснені некеровані термоядерні реакції під час вибуху водневих бомб.

23.4. Моделі атомного ядра

Спроби створити теорію ядра наштовхуються на серйозні труднощі: 1) недостатні знання про сили, які діють між нуклонами, 2) надзвичайна громіздкість квантової задачі багатьох тіл (ядро з масовим числом A – це система з A тіл). Ці труднощі змушують іти шляхом створення ядерних моделей, які дозволяють за допомогою порівняно простих математичних засобів описувати певну сукупність властивостей ядра. Жодна з подібних моделей не може дати вичерпного описання ядра, тому зазвичай використовують моделі, кожна з яких описує певні властивості ядра і своє коло явищ. У кожній моделі містяться довільні параметри, значення яких підбирають так, щоб отримати узгодження з експериментом. З багатьох моделей найбільш використовувані дві – краплинна та оболонкова.

Краплинна модель. Цю модель запропонував Я. І. Френкель у 1939 р. і потім розвинув Н. Бор та інші вчені. Вона ґрунтується на аналогії між поведінкою нуклонів у ядрі та молекул у краплині рідини. В обох випадках сили, що діють між складовими частинами, короткодійні та мають властивість насичення. Рідина має сталу густину, ядерна речовина також характеризується сталою густиною незалежно від числа нуклонів у ядрі. Об'єм краплини, як і об'єм ядра, пропорційний кількості частинок, що в них містяться.

Під час отримання певної енергії краплина ядерної рідини переходить у збуджений стан. Отримана енергія у результаті зіткнень нуклонів швидко перерозподіляється між ними. Утім, можуть виникнути такі умови, коли отримана енергія концентрується на поверхневому нуклоні або на групі нуклонів. Якщо ця енергія більша від енергії зв'язку частинки в ядрі, частинка може подолати поверхневі сили ядерного притягання і вийти з ядра. Такий процес аналогічний випаровуванню молекули з поверхні рідини.

Істотна відмінність ядра від краплини рідини у цій моделі полягає в тому, що вона трактує ядро як краплину електрично зарядженої нестискуваної рідини (з такою самою густиною, як ядерна), яка підпадає під закони квантової механіки. Краплинна модель ядра дозволила отримати напівемпіричну формулу для енергії зв'язку нуклонів у ядрі, пояснила механізм реакцій поділу ядер, однак не змогла пояснити, наприклад, підвищену стійкість ядер, які містять магічні числа протонів і нейтронів.

Оболонкова модель ядра (1949–1950 рр., німецький фізик Х. Йенсен, американський фізик М. Гепперт-Майєр) ґрунтується на припущенні, що нуклони в ядрі розміщуються на дискретних рівнях (оболонках), заповнюючи їх згідно з принципом Паулі. Стійкість ядер залежить від ступеня заповнення таких оболонок. Вважають, що ядра з повністю заповненими оболонками найбільш стійкі. Відповідно до дослідів особливо стійкими є ядра, в яких число протонів або число нейтронів (або обидва ці числа) дорівнюють 2, 8, 20, 28, 50, 82, 126. Ці числа отримали назву ***магічних***. Ядра, в яких число протонів Z або число нейтронів N є магічним (тобто особливо стійкі ядра), теж називають магічними. Ядра, в яких магічними є і Z, і N, називають *два рази магічними*.

Два рази магічних ядер відомо всього п'ять:

$$^{4}_{2}\mathrm{He}\ (Z=2, N=2),\ ^{16}_{8}\mathrm{O}\ (Z=8, N=8)$$
$$^{40}_{20}\mathrm{Ca}\ (Z=20, N=20),$$
$$^{48}_{20}\mathrm{Ca}\ (Z=20, N=28),$$
$$^{208}_{82}\mathrm{Pb}\ (Z=82, N=126).$$

Ці ядра особливо стійкі. Зокрема, особлива стійкість ядра гелію $^{4}_{2}\mathrm{He}$ проявляється в тому, що це єдина складна частинка, яка випромінюється важкими ядрами під час радіоактивного розпаду (її називають α-частинкою).

Нові експериментальні дані про властивості атомних ядер, які з'явились у результаті подальших досліджень, виходили за межі обох теорій. Так виникла узагальнена модель (синтез краплинної та оболонкової) – *оптична*

модель ядра (пояснює взаємодію ядер з елементарними частинками та ін.).

23.5. Радіоактивність. Основний закон радіоактивного перетворення атомних ядер

Радіоактивністю (від лат. radio – випромінюю, radius – промінь і activus – діючий) *називають самовільне і спонтанне перетворення ядер нестійких ізотопів одних елементів у ядра ізотопів інших елементів, що зумовлено внутріш-німи причинами та супроводжується* α-, β-, γ-*випромінюванням, а також інших частинок (нейтронів, протонів).*

Опишемо процеси, які характеризують різні типи реакцій ядерних розпадів природно радіоактивних ядер:

α-розпад: ${}^{A}_{Z}\mathrm{X} \to {}^{A-4}_{Z-2}\mathrm{X} + {}^{4}_{2}\mathrm{He}$;

β-розпад: ${}^{A}_{Z}\mathrm{X} \to {}_{Z+1}{}^{A}\mathrm{X} + {}_{-1}{}^{0}\beta + \bar{\nu}$;

β^{+}-розпад: ${}^{A}_{Z}\mathrm{X} \to {}_{Z-1}{}^{A}\mathrm{X} + {}^{0}_{1}\beta + \nu$;

β^{-}-захоплення: ${}^{A}_{Z}\mathrm{X} + {}_{-1}{}^{0}\beta \to {}_{Z-1}{}^{A}X + \nu$;

γ-розпад: $({}^{A}_{Z}\mathrm{X})^{*} \to {}^{A}_{Z}\mathrm{X} + \gamma$.

Тут ν та $\bar{\nu}$ позначають нейтрино і антинейтрино. Символом $({}^{A}_{Z}\mathrm{X})^{*}$ позначено ядро, яке перебуває у збудженому стані. Коли воно повертається у свій основний стан ${}^{A}_{Z}\mathrm{X}$, то випромінює γ-квант.

Радіоактивність, яка спостерігається в ядрах, що існують у природних умовах, називають *природною*. Радіоактивність ядер, які отримані за допомогою ядерних реакцій, називають *штучною*. Між природною і штучною радіоактивністю немає принципової відмінності.

Природні радіоактивні перетворення ядер, які відбуваються самочинно, називають *радіоактивним розпадом*. Ядро, що виникло внаслідок розпаду, називають *дочірним* ядром, а ядро, яке розпалось – *материнським*.

Теорія, яка пояснює це явище, ґрунтується на припущенні, що радіоактивний розпад є спонтанним процесом.

Численні досліди показали, що на швидкість радіоактивного розпаду не впливають жодні зовнішні зміни температури, тиск, наявність електричних і магнітних полів, вид хімічної сполуки, її агрегатний стан. Із цього випливає, що радіоактивний процес відбувається у глибинних частинах атома, тобто в його ядрі. Радіоактивний розпад – це властивість самого атомного ядра, вона залежить тільки від його внутрішнього стану.

Унаслідок динамічності цього процесу природно припустити, що кількість ядер $|dN|$, які розпадаються за інтервал часу від t до $t + dt$,

пропорційна проміжку часу dt і кількості ядер, які ще не розпались на момент часу t:

$$-dN = \lambda N dt, \qquad (23.10)$$

де λ – стала величина, яку називають *сталою розпаду*, або *радіоактивною сталою*. Знак «−» вказує на те, що загальна кількість радіоактивних ядер під час розпаду зменшується ($dN < 0$). Стала розпаду

$$\lambda = \frac{\frac{|dN|}{N}}{dt} \qquad (23.11)$$

дорівнює відносному зменшенню кількості ядер, які зазнають розпаду за одиницю часу. Стала λ визначає швидкість радіоактивного розпаду.

Розділивши в рівнянні (23.10) змінні та проінтегрувавши, отримаємо:

$$\int_{N_0}^{N} \frac{dN}{N} = -\lambda \int_0^t dt\,; \ln\frac{N}{N_0} = -\lambda t.$$

Звідси

$$N = N_0 e^{-\lambda t}. \qquad (23.12)$$

Цей вираз виражає *закон радіоактивного розпаду*, згідно з яким кількість ядер, які не розпались, зменшується експоненціально. У цій формулі N_0 – початкова кількість ядер, які не розпались на момент часу $t = 0$, N – кількість ядер, які не розпались на момент часу t.

Кількість ядер, які розпались за час t, визначають виразом

$$N_0 - N = N_0\left(1 - e^{-\lambda t}\right). \qquad (23.13)$$

На рис. 23.2 зображено залежність $\ln\frac{N}{N_0}$ від часу t. Вона дає змогу експериментально за нахилом прямої визначити сталу розпаду λ, тому що $\Delta(\ln\frac{N}{N_0})/\Delta t = \lambda$.

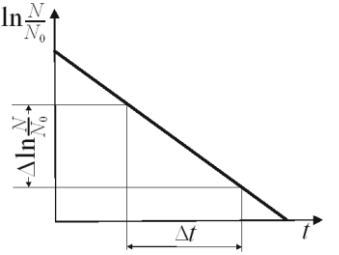


Рис. 23.2

Середній час життя. Середнім часом життя радіоактивного ядра називають середній час життя ядер в якому-небудь зразку, який містить ці радіоактивні ядра.

Кількість ядер $|dN(t)|$, що розпалась за проміжок часу від t до $t+dt$, визначають модулем виразу (23.10)

$$|dN(t)| = \lambda N(t) dt.$$

Час життя кожного з цих ядер дорівнює t. Отже, сумарний час життя всіх N_0 ядер, що були спочатку, отримують інтегруванням виразу $tdN(t)$. Поділивши цю суму на число ядер N_0, отримаємо середній час життя τ радіоактивного ядра:

$$\tau = \frac{1}{N_0}\int_0^\infty t|dN(t)| = \frac{1}{N_0}\int_0^\infty t\lambda N(t)dt =$$
$$= \frac{1}{N_0}\int_0^\infty t\lambda N_0 e^{-\lambda t}dt = \int_0^\infty t\lambda e^{-\lambda t}dt = \frac{1}{\lambda},$$

$$\tau = 1/\lambda. \qquad (23.14)$$

Стала розпаду λ є оберненою величиною до середнього часу життя τ певного радіоактивного елемента, тому (23.12) перепишемо так:

$$N = N_0 e^{-\frac{t}{\tau}}, \qquad (23.15)$$

де τ – це час, за який кількість ядер під час радіоактивного розпаду зменшиться в e = 2,718 28... раз.

Для характеристики стійкості ядер відносно розпаду і для оцінювання тривалості життя певного радіоактивного ізотопу вводять поняття періоду піврозпаду $T_{1/2}$.

Періодом піврозпаду називають час, протягом якого початкова кількість ядер певної речовини розпадається наполовину.

З визначення $T_{1/2}$ випливає, що

$$\frac{N_0}{2} = N_0 e^{-\lambda T_{1/2}}\ ,\ e^{-\lambda T_{1/2}} = 2.$$

Звідси отримуємо:

$$T_{1/2} = \frac{\ln 2}{\lambda} = \frac{0{,}693}{\lambda} = 0{,}693\tau. \quad (23.16)$$

Періоди піврозпаду різних природних ізотопів досить різняться один від одного: $T_{1/2} = 4{,}51 \cdot 10^9$ років для $^{238}_{92}U$, $T_{1/2} = 3 \cdot 10^{-7}$ с для $^{212}_{84}$Po.

На рис. 23.3 показаннно графіки кривих розпаду за різних значень λ. *Кількість атомів, що розпадається за одну секунду, називають активністю елемента.*

Активність A визначають такими формулами:

$$A = \left|\frac{dN}{dt}\right| = \lambda N = \lambda N_0 e^{-\lambda t} = \frac{N\ln 2}{T}. \quad (23.17)$$

Отже, активність обернено пропорційна періоду піврозпаду і зменшується з часом за експоненціальним законом.

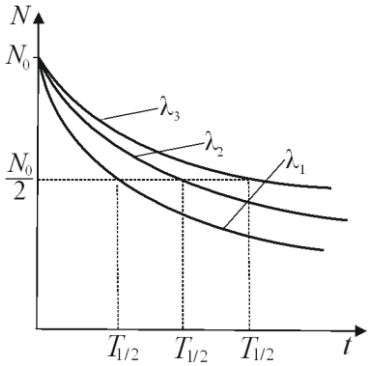


Рис. 23.3. Графіки кривих розпаду за різних значень λ

За одиницю активності препарату беруть один розпад за секунду – беккерель (Бк). Часто користуються позасистемною одиницею, яку називають Кюрі (Кі):

$$1\ \text{Кі} = 3{,}7 \cdot 10^{10}\ \text{Бк}.$$

($3{,}7{\cdot}10^{10}$ розпадів за 1 секунду відбувається в 1 грамі чистого радію ^{226}Ra).

Якщо відбувається ланцюжок радіоактивних перетворень, то закон розпаду дає змогу робити висновок лише про зменшення з часом материнської

речовини. Якщо ж дочірня речовина також радіоактивна, то одночасно відбувається і нагромадження цієї речовини, і її зменшення внаслідок розпаду. За час dt розпадеться $\lambda_м N_м dt$ ядер атомів материнської речовини із загальної кількості $N_м$ цих ядер, що були на момент часу t. Якщо з кожного ядра материнської речовини утворюється одне ядро дочірньої, то за проміжок часу dt кількість ядер дочірньої речовини збільшиться на $\lambda_м N_м dt$. Водночас з кількості ядер $N_д$ дочірньої речовини, що були на момент часу t, за проміжок часу dt розпадеться $\lambda_д N_д dt$.

Загальна зміна $dN_д$ кількості ядер дочірньої речовини за одиницю часу

$$\frac{dN_д}{dt} = \lambda_м N_м - \lambda_д N_д.$$

У випадку рухомої рівноваги між материнською і дочірньою речовинами $\frac{dN_д}{dt} = 0$ й $\lambda_м N_м = \lambda_д N_д$ – умова радіоактивної рівноваги. Звідси

$$\frac{N_м}{N_д} = \frac{\lambda_д}{\lambda_м} = \frac{T_{1/2м}}{T_{1/2д}}.$$

Числа атомів обох речовин прямо пропорційні до періодів їх піврозпадів.

Радіоактивний розпад записують у вигляді рівняння:

$$^{A_1}_{Z_1}\mathrm{X} \rightarrow {}^{A_2}_{Z_2}\mathrm{Y} + b,$$

де $^{A_1}_{Z_1}\mathrm{X}$ – материнське радіоактивне ядро, його маса m_x; $^{A_2}_{Z_2}\mathrm{Y}$ – дочірнє ядро (продукт розпаду), його маса m_y; b – частинка, яка випускається, її маса m_b.

Під час радіоактивного розпаду справджується *закон збереження енергії*, згідно з яким повна енергія материнського ядра дорівнює повній енергії продуктів розпаду:

$$m_м c^2 = (m_д + m_b)c^2 + W_p,$$

де W_p – повна енергія розпаду, яка дорівнює зміні енергії спокою материнського ядра $m_м$, що перетворюється в кінетичну енергію дочірнього ядра $m_д$, яке вилітає, і енергію ($m_b\ c^2$) γ-квантів.

Під час радіоактивного розпаду ядер виконується *закон збереження зарядових чисел*:

$$Z_я = \sum_i Z_i, \qquad (23.19)$$

де $Z_я$ – зарядове число ядра, яке зазнало розпаду; Z_i – зарядові числа ядер і частинок, що виникли внаслідок розпаду.

Крім того, виконується *правило збереження масових чисел*:

$$A_я = \sum_i A_i. \qquad (23.20)$$

Ці співвідношення під час радіоактивного розпаду формують у вигляді правил зміщення, які дають змогу встановити, яке ядро виникає внаслідок розпаду певного материнського ядра. При цьому розрізняють правила зміщення для випадків можливого α- або β-розпаду.

Ядра, які виникають унаслідок радіоактивного розпаду, також можуть бути радіоактивними. Послідовність такого процесу приводить до виникнення

ланцюжка радіоактивних перетворень, який закінчується стабільним елементом. Сукупність елементів, яка утворює такий ланцюжок, називають радіоактивним сімейством.

Сімейства називають за найбільш довгоживучим «родоначальником»: сімейство торію $^{232}_{90}$Th, нептунію $^{237}_{93}$Np, урану $^{238}_{92}$U та актинідію $^{235}_{89}$Ac. Кінцевими продуктами такого розпаду, відповідно, є $^{208}_{82}$Pb, $^{209}_{83}$Bi, $^{206}_{82}$Pb, $^{207}_{82}$Pb, тобто єдине сімейство нептунію (штучно радіоактивного ядра) закінчується нуклідом Bi. Усі інші (природно-радіоактивні ядра) – нуклідами Pb.

Дія γ-випромінювання (а також інших видів іонізувального випромінювання) на речовину характеризується дозою іонізувального випромінювання. Розрізняють такі одиниці випромінювання:

Поглинута доза випромінювання (*D*) – фізична величина, що дорівнює відношенню енергії поглинутого випромінювання до маси опромінюваної речовини. Одиниця поглинутої дози випромінювання – грей (Гр): 1 Гр = = 1 Дж/кг – доза випромінювання, за якої опромінюваній речовині масою 1 кг передається енергія довільного іонізувального випромінювання в 1 Дж.

Експозиційна доза випромінювання (*X*) – фізична величина, що дорівнює відношенню суми електричних зарядів усіх іонів одного знака, створених електронами, звільненими в опромінюваному повітрі (у разі повного використання іонізувальної здатності електронів), до маси цього повітря.

Одиниця експозиційної дози випромінювання – кулон, поділений на кілограм (Кл/кг), часто користуються позасистемною одиницею – **рентген** (*P*):

$$1\ \text{Р} = 2{,}58 \cdot 10^{-4}\ \text{Кл/кг}.$$

При експозиційній дозі, яка дорівнює одному рентгену, в 1 м^3 сухого повітря за нормального атмосферного тиску виникає сумарний заряд іонів одного знака, величина яких 0,33· мКл.

Еквівалентна доза іонізувального випромінювання $H-$ це добуток поглинутої дози на середній коефіцієнт якості k іонізувального випромінювання у певному елементі об'єму біологічної тканини стандартного складу: (O – 76,2 %, C – 11,1 %, H – 10,1 %, N – 2,6 %). Оскільки коефіцієнт k є безрозмірною величиною, то розмірність H збігається з розмірністю D [H] 1 Дж/кг = 1 Зв (**зіверт**).

Зіверт дорівнює еквівалентній дозі, за якої добуток поглинутої дози в біологічній тканині стандартного складу на середній коефіцієнт якості дорівнює 1 Дж/кг.

Еквівалентна доза іонізувального

випромінювання є основною величиною, яка визначає рівень радіаційної небезпеки при хронічному опроміненні людини в малих дозах. Більш широко використовують такі одиниці вимірювання H, як мілізіверт (мЗв) та мікрозіверт (мкЗв). У разі опромінення всього тіла людини протягом року допускається еквівалентна доза іонізувального випромінювання не більше 250 Зв.

Потужність еквівалентної дози іонізувального випромінювання (потужність еквівалентної дози) $dH/dt = \dot{H}$. $[\dot{H}] =$ 1 Зв/с. Зіверт за секунду дорівнює потужності еквівалентної дози, за якої за 1 с створюється еквівалентна доза в 1 Зв; 1 Зв = 1 Дж/кг = 10^2 бер.

Час перебування у полі випромінювання за низьких рівнів іонізувального випромінювання вимірюють, як правило, годинами (шестигодинний робочий день, 36-годинний робочий тиждень), тому надають перевагу такій одиниці потужності еквівалентної дози, як мікрозіверт за годину (мкЗв/год) незалежно від розміру величини.

Допустима середньорічна потужність еквівалентної дози під час опромінення всього тіла людей, які працюють, дорівнює 28 мкЗв/год за 36-годинного робочого тижня.

Природний фон створює потужність еквівалентної дози в межах 0,05... 0,2 мкЗв/год.

Біологічна доза – величина, що вказує, який вплив має випромінювання на організм. Одиниця біологічної дози – біологічний еквівалент рентгена (**бер**): 1 бер – доза довільного виду іонізувального випромінювання, яка здійснює таку саму біологічну дію, яку здійснює доза рентгенівського або γ-випромінювання в 1 Р: (1 бер = =10^{-2} Дж/кг = 10 мЗв).

Потужність дози випромінювання – величина, яка дорівнює відношенню дози випромінювання до часу опромінювання.

23.6. Ядерні реакції

Ядерною реакцією називають процес сильної взаємодії атомного ядра з елементарною частинкою або з іншим ядром, що призводить до перетворення ядра (ядер). Частинки починають взаємодіяти під час наближення їх до відстані 10^{-15} м, за якої починають діяти ядерні сили.

Характер ядерної реакції визначається видом частинки, її енергією, властивостями речовини та умовами взаємодії. Зокрема, зіткнення частинок з ядром викликає різні ядерні реакції, у результаті яких відбувається зміна заряду чи маси частинки або випускаються γ-

кванти.

Типовим прикладом ядерної реакції є процес взаємодії швидких α-частинок з ядрами азоту, за якої спостерігається виліт протонів:

$$^{4}_{2}\mathrm{He} + {}^{14}_{7}\mathrm{N} \rightarrow {}^{18}_{9}\mathrm{F} \rightarrow {}^{1}_{1}\mathrm{H} + {}^{17}_{8}\mathrm{O}.$$

Вперше ядерну реакцію такого типу здійснив Резерфорд у 1919 р. Кількість відомих нині ядерних реакцій сягає уже тисячі.

За механізмом взаємодії ядерні реакції можна поділити на два види:

1) прямі ядерні реакції;

2) реакції з утворенням проміжного ядра.

Прямі ядерні реакції відбуваються за дуже високих енергій частинок і за дуже короткий час, потрібний для того, щоб частинка пролетіла крізь ядро (10^{-23} ... 10^{-21} с). Наприклад, швидкий протон може вибити з поверхні ядра один із нуклонів і полетіти разом з ним.

Більшість ядерних реакцій з кінетичною енергією частинок, меншою за 10 МеВ, проходять з утворенням проміжного ядра. Така ядерна взаємодія відбувається у два етапи. На першому етапі відбувається захоплення частинки ядром і виникнення проміжного ядра, яке перебуває у збудженому стані. Енергія збудження W^* складається з кінетичної енергії частинки $W_к$ та енергії зв'язку нуклона $W_{зв}$, який приєднався:

$$W^* = W_к + W_{зв}.$$

Енергія зв'язку нуклона в середньому дорівнює 8 МеВ, тому складове ядро отримує досить велику енергію збудження.

В ядерній взаємодії беруть участь лише ближні до частинки нуклони, тому до них переходить основна частина енергії збудження. Енергія, отримана такими нуклонами, передається потім за участю ядерних сил сусіднім і поступово статистично розподіляється між всіма нуклонами ядра. При цьому відбувається значний розігрів ядра. Енергія збудження, розподілена рівномірно між всіма нуклонами, недостатня для подолання енергії зв'язку нуклона. Утім, з часом флуктуації збудженого ядра приводять до того, що в одному нуклоні зосереджується енергія, достатня для відривання його від ядра ($W_{нук} > W_{зв}$). Тоді настає другий етап ядерної реакції – викидання нуклона з ядра.

Під час захоплення нейтрона з утворенням складового ядра швидкість нейтрона, який вилетів, звичайно менша, ніж швидкість захопленого первинного нейтрона. Оскільки первинні та вторинні нейтрони не можна розрізнити, таке явище сприймають як дещо сповільнене розсіяння нейтрона із втратою частини його енергії. Такий процес має назву

непружного (резонансного) розсіяння частинок.

Проміжок часу, який потрібний нуклону з енергією 1 МеВ ($v \sim 10^7$ м/с), щоб пройти відстань, яка дорівнює діаметру ядра ($\sim 10^{-14}$ м), називають ядерним часом (або *ядерним часом прольоту* $\tau_я$):

$$\tau_я \cong \frac{10^{-14}\text{м}}{10^7\text{м/с}} = 10^{-21}\text{ с}.$$

Середній час життя складового ядра (10^{-1} ... 10^{-2} с) на багато порядків перевищує ядерний час $\tau_я$. Отже, розпад проміжного ядра є процесом, який не залежить від першого етапу реакції, що полягає в захопленні частинки (проміжне ядро ніби «забуває» спосіб свого утворення). Одне й те саме ядро може розпадатися різними шляхами, причому характер цих шляхів і їх відносна ймовірність не залежать від способу утворення проміжного ядра.

Під час ядерної реакції зберігається кількість нуклонів і сумарний заряд. Відбувається лише перерозподіл нуклонів і заряду між ядрами та частинками. Збереження заряду і кількості нуклонів дає змогу якісно визначити можливі напрямки перебігу ядерної реакції.

Ядерні реакції супроводжуються зміною кінетичної енергії взаємодійних частинок. *Для ядерних реакцій виконуються закони збереження енергії та імпульсу.*

23.7. Поділ ядер

В експериментах, які провів Енріко Фермі, Еміліо Сегре та їх співробітниками у 1934 р., ядра урану бомбардували нейтронами. Опромінений нейтронами уран проявляв при цьому штучну β-радіоактивність з декількома невеликими періодами піврозпаду. Оскільки на той час було відомо, що ядра урану розпадаються з випромінюванням α-частинок з дуже великим періодом піврозпаду, то Фермі та його співробітники запропонували, що в них в урані утворились трансуранові елементи (із $Z > 92$). Це припущення, однак, виявилось неправильним.

Отто Ган і Фріц Штрассман незабаром показали за допомогою дуже ретельного хімічного аналізу, що одним з елементів, які утворюються після бомбірдування урану нейтронами, був ізотоп барію $^{141}_{56}$Ba. І тоді у 1939 р. Отто Фріш і Ліза Мейтнер висунули важливу гіпотезу, що ядра урану після бомбардування повільними (тепловими) нейтронами зазнають реакції поділу, у результаті якої одне ядро урану розщеплюється на два ядра приблизно однакової маси, але менших розмірів.

Ці експерименти стали можливими внаслідок відкриття Фермі та співробітників, суть якого у тому, що

повільні (теплові) нейтрони отримуються шляхом гальмування нейтронних пучків за допомогою пропускання їх крізь матеріали, багаті воднем, наприклад крізь воду або парафін.

Реакція поділу ядер урану схематично може бути подана рівнянням

$$^{235}_{92}\mathrm{U} + {}^{1}_{0}n \rightarrow {}^{236}_{92}\mathrm{U}^{*} \rightarrow X + Y + \text{нейтрони}, \quad (23.21)$$

де ${}^{1}_{0}n$ – повільний (тепловий) нейтрон; ${}^{236}_{92}\mathrm{U}^{*}$ – надзвичайно нестабільний ізотоп урану; X та Y – ядра-фрагменти, які отримуються під час поділу ядра урану. Ядра-фрагменти в реакції поділу можуть бути, звичайно, різними, як це насправді й проявляється в експерименті. Типовими реакціями ядерного ділення є такі реакції:

$$^{235}_{92}\mathrm{U} + {}^{1}_{0}n \rightarrow {}^{236}_{92}\mathrm{U}^{*} \rightarrow \rightarrow {}^{144}_{56}\mathrm{Ba} + {}^{89}_{36}\mathrm{Kr} + 3{}^{1}_{0}n + Q, \quad (23.22)$$

$$^{235}_{92}\mathrm{U} + {}^{1}_{0}n \rightarrow {}^{236}_{92}\mathrm{U}^{*} \rightarrow \rightarrow {}^{140}_{54}\mathrm{Xe} + {}^{94}_{38}\mathrm{Sr} + 2{}^{1}_{0}n + Q, \quad (23.23)$$

де Q – енергія, яка виділяється під час реакції.

β-активність, яку спостерігали Фермі та його співробітники, відбувалась від радіоактивних розпадів ядер-фрагментів, наприклад від ядерних реакцій:

$$^{140}_{54}\mathrm{Xe} \xrightarrow{\beta^-} {}^{140}_{55}\mathrm{Cs} \xrightarrow{\beta^-} {}^{140}_{56}\mathrm{Ba} \xrightarrow{\beta^-} \rightarrow {}^{140}_{57}\mathrm{La} \xrightarrow{\beta^-} {}^{140}_{58}\mathrm{Ce},$$

$$^{94}_{38}\mathrm{Sr} \xrightarrow{\beta^-} {}^{94}_{39}\mathrm{Y} \xrightarrow{\beta^-} {}^{94}_{40}\mathrm{Zr} \quad (23.24)$$

для реакції (23.23).

Коли ядро ${}^{235}_{92}\mathrm{U}$ захоплює тепловий нейтрон, енергія захопленого нейтрона розподіляється між нуклонами, які входять у склад ядра. Потім, подібно до рідкої краплини, ядро починає сильно деформуватись, поки кулонівське відштовхування його протонів не розірве це ядро на два ядра-фрагменти, як це показано на рис. 23.4. Розподіл ядер-фрагментів, які отримуються під час поділу ядра урану, що захопило нейтрон, за масовими числами показано на рис. 23.5.

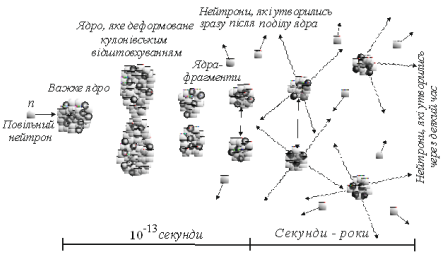


Рис. 23.4. Поділ ядра ${}^{235}_{92}\mathrm{U}$ після захоплення повільного нейтрона

Усі ядра-фрагменти, які утворюються під час поділу ядра урану, мають надлишкове число нейтронів. Через цей надлишок нейтронів вони є сильно нестабільними. Збуджені ядра-фрагменти майже одразу випромінюють нейтрони протягом дуже короткого проміжку часу близько 10^{-13} с. Крім цих одразу випущених нейтронів є, однак, інша група *затриманих за часом* нейтронів, які випромінюються ядрами-

фрагментами з періодом півроспаду у декілька секунд (рис. 23.4). Ці затримані за часом нейтрони дуже важливі для здійснення контролю за ядерними реакціями, які відбуваються в атомному реакторі.

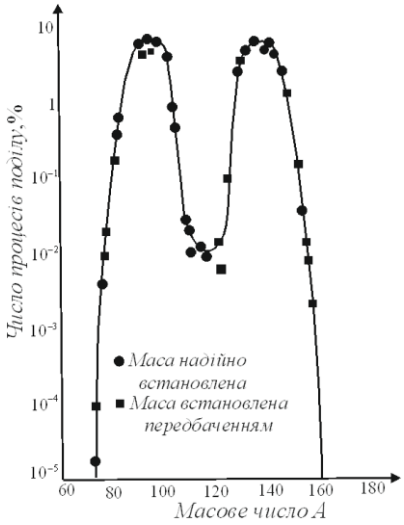


Рис. 23.5. Крива виходу продуктів на один поділ. Під час поділу на теплових нейтронах утворюються переважно ядра-фрагменти із співвідношенням мас 2:3. Найімовірнішими продуктами поділу з виходом ≈6,5 % є ядра-фрагменти з масовими числами 95 і 139. Поділ на однакові ядра-фрагменти є малоймовірними (≈0,01 %)

Під час ядерної реакції поділу виділяється досить велика кількість енергії. Ще трохи енергії виділяється під час радіоактивного розпаду ядер-фрагментів.

Енергія ядерного поділу вивільняється у формі кінетичної енергії ядер-фрагментів та у вигляді кінетичної енергії електронів, фотонів, нейтрино і нейтронів, які випускаються ядрами-фрагментами.

Оцінимо значення величини енергії, яка виділяється під час реакції ядерного поділу (23.22) і (23.23). У цих реакціях нестабільне ядро $^{235}_{92}\mathrm{U}$ розщеплюється, зрештою, на ядра-фрагменти $^{140}_{58}\mathrm{Ce}$, $^{94}_{40}\mathrm{Zr}$, шість β-частинок і два нейтрони. Маси до і після поділу наведено у табл. 23.1.

Таблиця 23.1

Маси частинок до і після реакції ядерного поділу (23.22) і (23.23)

До поділу	Після поділу
235,0439а.о.м.($^{235}_{92}$U)	139,9054а.о.м.($^{140}_{58}$Ce)
1,0087а.о.м.($^{1}_{0}n$)	93,9036а.о.м.($^{94}_{40}$Zr)
	2,0173а.о.м.($2\,^{1}_{0}n$)
	0,0033а.о.м.($6\beta^-$)
$\sum$ 236,0526а.о.м.	$\sum$ 235,8296а.о.м.

Дефект маси реакції $\Delta m = (236{,}0526 - -235{,}8296)$ а. о. м. $= 0{,}223$ а.о.м. Енергетичний ефект реакції знайдемо за формулою (23.8):

$$Q = 0{,}223 \cdot 931{,}2 \text{ МеВ} = 208 \text{ Мев}.$$

На один нуклон, який бере участь у реакції, припадає $\frac{208}{236} = 0{,}88 \frac{\text{Мев}}{\text{нуклон}}$.

Для порівняння зазначимо, що енергія, яка виділяється під час α -розпаду одного ядра, становить 5 МеВ, а енергія, яка виділяється в одному акті в хімічній реакції згорання, дорівнює 4 еВ.

Нейтрони, які отримуються в реакції поділу, можна знову використовувати для здійснення нових актів реакції поділу ядер, і при цьому буде звільнятись все більше і більше енергії. В «атомному котлі», який побудував Фермі в Чикагському університеті, вперше було проведено самопідтримуючу ланцюгову ядерну реакцію 2 грудня 1942 р.

Тепловими нейтронами діляться ще ядра плутонію $^{239}_{94}\mathrm{Pu}$, урану $^{233}_{92}\mathrm{U}$, технецію $^{230}_{90}\mathrm{Th}$. Двох останніх ізотопів у природі немає, їх отримують штучно. Одна зі схем реакції поділу ядер плутонію така:

$$^{239}_{94}\mathrm{Pu} + {}^{1}_{0}n \rightarrow$$
$$\rightarrow {}^{140}_{54}\mathrm{Xe} + {}^{96}_{40}\mathrm{Zr} + 4{}^{1}_{0}n + 188\text{МеВ}. \quad (23.25)$$

Зазначимо, що в цій реакції в одному акті поділу виділяється чотири нейтрони, що приблизно у два рази більше, ніж в реакції поділу (23.23).

Реакцію поділу атомних ядер можна викликати, бомбардуючи ядра не тільки нейтронами, але й іншими частинками, наприклад протонами, α-частинками та γ-квантами. В реакції поділу ядер можуть брати участь, крім урану, плутонію, також і більш легкі елементи. Типовий приклад такої реакції поділу – реакція поділу ядер міді, яка викликається протонами з енергією 60 МеВ, прискорених на циклотроні:

$$^{63}_{29}\mathrm{Cu} + {}^{1}_{1}p \rightarrow {}^{38}_{17}\mathrm{Cl} + {}^{25}_{13}\mathrm{Al} + {}^{1}_{0}n. \quad (23.26)$$

23.8 Ланцюгова реакція поділу. Ядерний реактор

Для практичного застосування поділу важких ядер найважливіше значення має виділення великої кількості енергії під час кожного акту поділу і появі при цьому 2−3 нейтронів. Кожен із цих нейтронів взаємодіє із сусідніми ядрами речовини, що, у свою чергу, спричиняє в них реакцію поділу, тобто відбувається лавиноподібне зростання кількості актів поділу. *Таку реакцію поділу називають ланцюговою.*

Схематично ланцюгову реакцію поділу ядер зображено на рис. 23.6. Ланцюгова реакція поділу характеризується *коефіцієнтом розмноження k нейтронів, який дорівнює відношенню кількості нейтронів у певному поколінні до їх кількості у попередньому поколінні.*

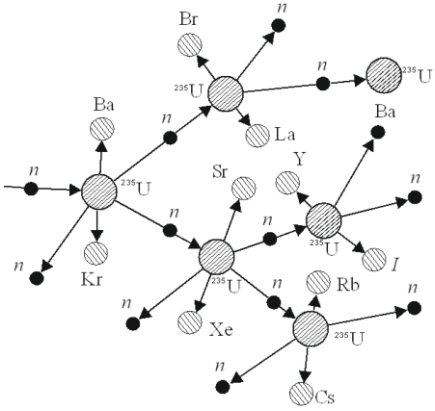


Рис. 23.6. Схема ланцюгової реакцію поділу ядер урану

Виявляється, що не всі вторинні нейтрони, які утворюються, викликають наступний поділ ядер, що призводить до зменшення *k*. По-перше, унаслідок скінченних розмірів активної зони (простір, де відбувається ланцюгова реакція) і великої проникної здатності нейтронів частина з них покине активну зону раніше, ніж буде захоплена яким-небудь ядром. По-друге, частина нейтронів захоплюється ядрами домішок, що не діляться, і завжди присутні в активній зоні.

Коефіцієнт розмноження залежить від природи речовини, що ділиться, а для певного ізотопа – від його кількості, а також від розмірів і форми активної зони.

Мінімальні розміри активної зони, за яких можливе здійснення ланцюгової реакції, називають *критичними розмірами*. Мінімальну масу речовини, що ділиться, яка потрібна для здійснення ланцюгової реакції, називають *критичною масою*. Для зменшення втрат нейтронів і критичних параметрів речовини її оточують відбивачем – шаром неподільної речовини, яка має малий ефективний поперечний переріз для захоплення нейтронів і великий переріз для їх розсіяння.

Обчислимо швидкість розвитку ланцюгової реакції. Нехай τ – середній час життя одного покоління нейтронів, а N – кількість нейтронів у попередньому поколінні. У цьому поколінні їх кількість дорівнює kN, тобто приріст кількості нейтронів за одне покоління $dN = kN - -N = N(k-1)$. Приріст кількості нейтронів за одиницю часу, тобто швидкість наростання ланцюгової реакції,

$$\frac{dN}{dt} = \frac{N(k-1)}{\tau},\ \frac{dN}{N} = \frac{k-1}{\tau}dt.$$

Інтегруючи цей вираз, отримуємо

$$N = N_0 e^{(k-1)\frac{t}{\tau}},$$

де N_0 – кількість нейтронів у початковий момент часу, а N – їх кількість у момент часу t. За $k > 1$ йде наростаюча реакція, кількість поділів неперервно росте й реакція може стати вибуховою.

За $k = 1$ йде самопідтримуюча реакція, у процесі якої кількість нейтронів з часом не змінюється. За $k < 1$ йде загасаюча реакція.

Ланцюгові реакції поділяють на *керовані* й *некеровані*. Вибух атомної бомби є некерованою реакцією. Якщо маса близька до критичної, нейтрони швидко розмножуються і реакція набуває вибухового характеру. На цьому ґрунтується дія атомної бомби. Ядерним зарядом такої бомби є два або більше шматків майже чистого ^{235}U або ^{239}Pu. Маса кожного шматка менша від критичної, унаслідок чого ланцюгової

реакції не виникає.

В земній атмосфері завжди є деяка кількість нейтронів, народжених космічними променями, тому щоб викликати вибух, досить з'єднати шматки ядерного заряду в один шматок з масою, яка більша від критичної. Це має виконуватись дуже швидко, і з'єднання шматків має бути дуже щільним. Інакше ядерний заряд розлетиться на частини, перш ніж встигне прореагувати значна частина речовини, яка ділиться. Для з'єднання шматків використовують звичайну вибухову речовину. Ланцюгова реакція в атомній бомбі відбувається на швидких нейтронах. Під час вибуху встигає прореагувати тільки частина ядерного заряду.

Керовані ланцюгові реакції здійснюються в ядерних реакторах. Як сировинні й подільні речовини в реакторах використовують $^{235}_{92}\mathrm{U}$, $^{239}_{94}\mathrm{Pu}$, $^{238}_{92}\mathrm{U}$ а також $^{232}_{90}\mathrm{Th}$. У природній суміші ізотопів урану ізотопа $^{238}_{92}\mathrm{U}$ у 140 разів більше, ніж ізотопа $^{235}_{92}\mathrm{U}$.

Велике значення в ядерній енергетиці має не лише здійснення ланцюгової реакції поділу, але й керування нею. *Пристрій, в якому здійснюється і підтримується керована ланцюгова реакція поділу, називають ядерним реактором*.

Перший ядерний реактор, який схематично зображено на рис. 23.7, містив 52 тонни урану у вигляді блоків, які були перекладені приблизно 472 тоннами графіту, що відігравав роль сповільнювача нейтронів. У реакторі було п'ять контролюючих стрижнів (покритих кадмієм – дуже в'язким матеріалом). Для забезпечення безпеки було ще три стрижні з постійно прикладеним до них механічним навантаженням, яке заштовхувало їх у реактор(у випадку, якщо під час аварії відкаже електромережа).

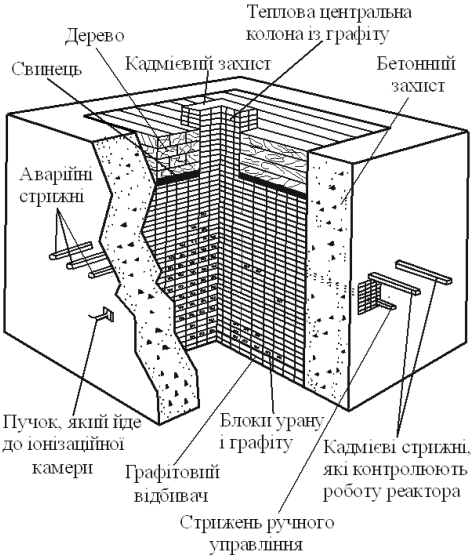


Рис. 23.7. Схема першого ядерного реактора

Речовиною, яка ділиться в реакторі, є природний (або дещо збагачений ізотопом $^{235}\mathrm{U}$) уран. Щоб перешкодити радіаційному захопленню

нейтронів ядрами ^{238}U (яке стає особливо інтенсивним для енергії нейтронів ~7 МеВ), порівняно невеликі блоки урану розміщують на деякій відстані один від одного, а проміжки між блоками заповнюють сповільнювачем, тобто речовиною, в якій нейтрони сповільнюються до теплових швидкостей. Такі теплові нейтрони ефективно взаємодіють з ядрами ^{235}U.

Сповільнення нейтронів здійснюється внаслідок пружного розсіювання. У цьому випадку енергія, яка втрачається частинкою, що сповільнюється, залежить від співвідношення мас частинок, які зіштовхуються. Максимальна кількість енергії втрачається у випадку, якщо обидві частинки мають однакову масу. З цієї точки зору ідеальним сповільнювачем має бути речовина, яка містить звичайний водень, наприклад вода (маси протона і нейтрона приблизно однакові). Ядра сповільнювача мають сильно пружно розсіювати нейтрони. Цю умову задовольняють дейтрон (ядро важкого водню – дейтерію D), а також ядра графіту C й берилію Be. Щоб зменшити енергію нейтрона від 2 МеВ до теплових енергій (~0,03 Мев), у важкій воді D_2O достатньо близько 25 зіткнень, у C і Be – приблизно 100 зіткнень.

Тепловиділяючі елементи (твели) – це блоки з радіоактивного матеріалу, що містяться в герметичній оболонці, яка слабо поглинає нейтрони. За рахунок енергії, що виділяється у процесі поділу ядер, твели розігріваються, і їх поміщають у потік теплоносія. Проходячи через активну зону, теплоносій у вигляді газу, води або розплавленого металу нагрівається і передає теплоту через спеціальний пристрій робочому тілу, наприклад воді в паротурбогенераторі, а потім знову надходить в активну зону реактора. Активну зону оточують тепловідбивачем, що зменшує витік нейтронів.

Керування ланцюговою реакцією здійснюють спеціальними керуючими стрижнями з матеріалів, що сильно поглинають нейтрони (наприклад, кадмій). Якщо стрижні вставлені повністю, реакції не буде. Під час поступового виймання стрижнів k росте і за певного положення доходить до одиниці. У цей момент реактор починає працювати. У міру його роботи кількість матеріалу, який ділиться в активній зоні, зменшується, і відбувається її забруднення фрагментами поділу. Щоб реакція не припинилась, з активної зони за допомогою автоматичного пристрою поступово виймають керуючі стрижні.

В реакторі є аварійні стрижні, введення яких зі зростанням

інтенсивності реакції одразу її припиняє. Кожний реактор має біологічний захист – систему екранів із захисних матеріалів.

23.9. Реакції синтезу легких ядер

Сонце втрачає величезну кількість енергії, але з часом помітно не охолоджується. Цей факт був загадкою для фізиків до того часу, поки вони не відкрили існування ядерної реакції.

У 1938 р. Ханс Бете висунув гіпотезу, що за величезну кількість енергії, яка вивільняється на Сонці і в інших зорях, відповідальні ядерні реакції, в яких два легких ядра об'єднуються або зливаються й утворюють одне більш важке ядро. Такі реакції називають ядерними реакціями синтезу.

Як видно з рис. 23.1, об'єднання двох легких ядер, які належать до швидко зростаючої ділянки цієї кривої, в одне більш важке ядро дає ядро з більшою енергією зв'язку в розрахунку на один нуклон, ніж у ядер, які об'єднуються. Оскільки отримується більша питома енергія зв'язку, сумарна маса вихідних ядер, які зливаються, має зменшуватись, і тому енергетичний ефект реакції має бути позитивним і в реакціях синтезу енергія справді буде вивільнятися.

Оскільки для синтезу ядер потрібні дуже високі температури (див. підрозд. 23.3), цей *процес називають термоядерною реакцією*.

Для прикладу наведемо деякі реакції синтезу легких ядер і значення теплових ефектів Q:

$$^2_1\mathrm{H}+^2_1\mathrm{H} = ^3_1\mathrm{H} + ^1_1\mathrm{H}, \qquad (Q = 4{,}0\ \text{Мев}),$$

$$^2_1\mathrm{H}+^2_1\mathrm{H} = ^3_2\mathrm{H} + ^1_0n, \qquad (Q = 3{,}3\ \text{Мев}),$$

$$^2_1\mathrm{H}+^3_1\mathrm{H} = ^4_2\mathrm{He} + ^1_0n, \quad (Q = 17{,}6\ \text{Мев}),$$

$$^6_3\mathrm{Li} + ^2_1\mathrm{H} = ^4_2\mathrm{He} + ^4_2\mathrm{He}, \quad (Q = 22{,}4\ \text{Мев}).$$

Виділення великої кількості енергії пояснюється тим, що питома енергія зв'язку ядер різко збільшується під час переходу від ядер водню ($^2_1\mathrm{H}$, $^3_1\mathrm{H}$) до літію $^6_3\mathrm{Li}$ й особливо до гелію $^4_2\mathrm{He}$.

Реакції синтезу атомних ядер мають ту особливість, що в них енергія, яка виділяється на один нуклон, значно більша, ніж у реакціях поділу важких ядер. Справді, якщо під час поділу ядра $^{235}_{92}\mathrm{U}$ виділяється енергія приблизно 0,84 МеВ на один кулон, то в реакції синтезу приблизно 3,5 МеВ на один нуклон, тобто в чотири рази більше.

Оцінимо температуру перебігу термоядерної реакції на прикладі синтезу ядра дейтерію $^2_1\mathrm{H}$. Для об'єднання ядер дейтерію їх треба наблизити на відстань, що дорівнює радіусу дії ядерних сил $r = 2\cdot 10^{-15}$ м, долаючи при цьому кулонівський бар'єр $\frac{e^2}{4\pi\varepsilon_0 r} \approx 0{,}7$ Мев. На кожне ядро, що стикається, припадає 0,35 МеВ. Середній енергії теплового руху $3kT/2$, яка дорівнює 0,35 МеВ, відповідає температура $2\cdot 10^9$ К. Отже,

реакція синтезу ядер дейтерію може відбуватися лише за температури, що перевищує на два порядки температуру центральних ділянок Сонця ($1{,}3\cdot 10^7$ К).

Утім, для перебігу реакцій синтезу атомних ядер достатньо температур близько 10^7 К. Це пов'язано із двома факторами: 1) за температур, характерних для реакцій синтезу атомних ядер, речовина перебуває у стані плазми, розподіл частинок за швидкостями якої описується законом Максвелла, тому завжди є деяка кількість ядер, енергія яких набагато перевищує середнє значення; 2) синтез ядер може відбуватися внаслідок тунельного ефекту.

За високих температур і тисків, які існують всередині Сонця й зірок, молекули речовини дисоційовані на атоми, а високоіонізовані атоми утворюють гарячу й активну плазму.

У цих умовах кулонівське відштовхування ядер може бути легко подолане і ядра можуть зливатись одне з одним. У результаті вивільняється величезна енергія.

У зв'язку з тим, що водень – найбільш поширений елемент у Всесвіті, Бете припустив, що в зірках здійснюється «вуглецевий цикл». Вважають, що це найбільш важливий ланцюжок ядерних реакцій, за якого енергія виділяється внаслідок реакції злиття ядер:

$$\text{а) } {}^1_1\mathrm{H}+{}^{12}_6\mathrm{C}\rightarrow {}^{13}_7\mathrm{N}+\gamma;$$
$$\text{б) } {}^{13}_7\mathrm{N}\rightarrow {}^{13}_6\mathrm{C}+{}^0_{+1}e+\nu;$$
$$\text{в) } {}^1_1\mathrm{H}+{}^{13}_6\mathrm{C}\rightarrow {}^{14}_7\mathrm{N}+\gamma; \quad (23.27)$$
$$\text{г) } {}^1_1\mathrm{H}+{}^{14}_7\mathrm{N}\rightarrow {}^{15}_8\mathrm{O}+\gamma;$$
$$\text{д) } {}^{15}_8\mathrm{O}\rightarrow {}^{15}_7\mathrm{N}+{}^0_{+1}e+\nu;$$
$$\text{е) } {}^1_1\mathrm{H}+{}^{15}_7\mathrm{N}\rightarrow {}^{12}_6\mathrm{C}+{}^4_2\mathrm{He}.$$

Розглядаючи уважно цей ланцюжок реакцій, переконуємось, що ядро вуглецю виступає як каталізатор; з нього ланцюжок реакцій починається, ним же він і закінчується. У наведеному ланцюжку реакцій витрачається чотири ядра водню (протони – ${}^1_1\mathrm{H}$) й утворюється три γ-кванти, два нейтрино (ν), два позитрона (${}^0_{+1}e$) та одне ядро гелію (${}^4_2\mathrm{He}$), так що по суті у вуглецевому циклі йде реакція

$$4\,{}^1_1\mathrm{H}\rightarrow {}^4_2\mathrm{He}+2\,{}^0_{+1}e+2\nu+3\gamma+Q. \quad (23.27')$$

Вирахуємо різницю мас Δm для цієї реакції (табл. 23.2).

Таблиця 23.2

Маси частинок до і після реакції (23.27′)

До реакції	Після реакції
4,031300а.о.м.($4\,{}^1_1\mathrm{H}$)	4,002603а.о.м.(${}^4_2\mathrm{He}$) 0,001098а.о.м.($2\,{}^0_{+1}e$)
	$\sum$ 4,003701а.о.м.

Тоді $\Delta m = (4{,}031\,300 - 4{,}003\,701) = 0{,}0276$ а. о. м., що дає тепловий ефект реакції $Q = 0{,}0276 \times 931{,}2 = 25{,}7$ МеВ. Цю енергію виносять із собою γ-кванти, два нейтрино, ядро гелію (α-частинка).

Ще в одному циклі ядерних реакцій, яка, ймовірно, існує, беруть участь переважно протони, і тому його називають «протонно-протонним циклом». У цьому відбувається такий ланцюжок ядерних реакцій:

а) ${}_1^1\mathrm{H} + {}_1^1\mathrm{H} \rightarrow {}_1^2\mathrm{H} + {}_{+1}^0e + \nu$;

б) ${}_1^2\mathrm{H} + {}_1^1\mathrm{H} \rightarrow {}_2^3\mathrm{He} + \gamma$; (23.28)

в) ${}_2^3\mathrm{He} + {}_2^3\mathrm{He} \rightarrow {}_2^4\mathrm{He} + 2{}_1^1\mathrm{H}$.

У цьому ланцюжку реакції а) та б) повинні відбуватись по два рази, щоб було отримано два ядра ${}_2^3\mathrm{He}$, які потрібні для реакції *в*). У результаті всіх цих реакцій із чотирьох ядер водню (протонів) утворюється одна α-частинка, два позитрони і два γ-кванти. Енергія, яка виділяється у цьому циклі, дорівнює 26,2 МеВ.

Вперше штучну термоядерну реакцію було здійснено у 1953 р., а потім у вигляді вибуху водневої (термоядерної бомби), що є некерованою реакцією. Вибуховою речовиною, в якій проходила реакція ${}_1^3\mathrm{H} + {}_1^2\mathrm{H} \rightarrow {}_2^4\mathrm{He} + {}_0^1n$, є суміш дейтерію і тритію, а запалом – звичайна атомна бомба, під час вибуху якої виникає потрібна для перебігу термоядерної реакції температура.

Теоретичною основою для перебігу штучних керованих термоядерних реакцій є реакції типу ${}_1^2\mathrm{H} + {}_1^3\mathrm{H} \rightarrow {}_2^4\mathrm{He} + {}_0^1n$, що відбуваються у високотемпературній плазмі. Утім, завдання полягає не тільки у створенні умов, потрібних для інтенсивного виділення енергії в термоядерних процесах, а здебільшого у підтриманні цих умов. Для здійснення термоядерної реакції, яка сама себе підтримує, потрібно, щоб швидкість виділення енергії в системі, де відбувається реакція, була не менша за швидкість відведення енергії від системи.

Розрахунки показують, що для забезпечення самопідтримної керованої термоядерної реакції температуру дейтерієвої плазми треба довести до кількох сотень мільйонів градусів. За температур ~10^8 К реакція має повну інтенсивність і супроводжується виділенням великої енергії. Так, за температури $1{,}17\cdot10^8$ К потужність, яка виділяється в одиниці об'єму плазми під час злиття дейтерієвих ядер, становить 3 кВт/м^3, тоді як за температури $1{,}17\cdot10^6$ К вона дорівнює лише 10^{-17} кВт/м^3.

Основною причиною втрат енергії високотемпературною плазмою є її велика теплопровідність, яка швидко зростає ($\sim T^{5/2}$) за високих температур. Енергія з плазми може відводитися завдяки дифузії гарячих частинок з області, де відбувається реакція, на стінки апарата, в якому міститься плазма. Якщо плазму не теплоізолювати від контакту з будь-якими навколишніми речовинами, то її можна нагрівати лише до кількох сотень тисяч градусів, тому що енергія,

яка виділяється внаслідок реакції синтезу, виходитиме на стінки. Інакше кажучи, треба втримати плазму в заданому об'ємі, не допускаючи її розширення. Ідею ефективної магнітної термоізоляції плазми стосовно проблеми керованого термоядерного синтезу запропонували А. Д. Сахаров та І. Є. Тамм у 1950 р. Якщо пропустити через плазму у формі стовпа вздовж його осі великий електричний струм, то магнітне поле цього струму створює електродинамічні сили, які прагнутимуть стиснути плазмовий стовп. Отже, стовп плазми буде відірваний від стінок і стягнутий у плазмовий шнур. Утім, утримувати плазмовий шнур у такому стані не вдається: відбуваються швидкі радіальні коливання – він то розширюється, то знову стискається. Унаслідок нестабільності, нестійкості плазми у плазмовому шнурі виникають деформації, які змінюють геометричну форму шнура. Результатом цього є порушення термоізоляції, інтенсивна взаємодія плазми зі стінками, що приводить до забруднення дейтерію речовиною стінок і швидкого охолодження плазми.

Основним питанням, розв'язання якого дозволить здійснити керовані термоядерні реакції, є з'ясування умов, за яких високотемпературна плазма в магнітному полі належної конфігурації може зберегти стійкість. Розв'язання цього питання поряд з пошуками шляхів підвищення температури плазми є головним напрямом, в якому розвиваються дослідження керованих термоядерних реакцій. Можливість реалізації термоядерної реакції зводиться до необхідності виконання двох вимог: наявності деякої мінімальної температури і певного обмеження для добутку концентрації n частинок у плазмі на час τ їх утримання у плазмі. Ці умови називають *критерієм Лоусона*:

$$\begin{cases} n\tau > 10^{14}\,\dfrac{\text{с}}{\text{см}^3}, \qquad T > 10^8\ \text{К}, \\ {}^2_1\text{H} + {}^3_1\text{H} \to {}^4_2\text{He} + {}^1_0n; \end{cases}$$

$$\begin{cases} n\tau > 10^{15}\,\dfrac{\text{с}}{\text{см}^3}, \quad T > 10^9\ \text{К}, \\ {}^2_1\text{H} + {}^2_1\text{H} \to {}^3_1\text{H} + {}^1_1\text{H}. \end{cases}$$

Контрольні запитання

1. Які є сучасні уявлення про будову атомного ядра?
2. Що таке дефект маси ядра?
3. Що таке енергія зв'язку ядра? Як її визначають практично?
4. Накресліть приблизний графік залежності питомої енергії зв'язку від масового числа ядра. Який висновок можна зробити з такої залежності питомої енергії зв'язку від масового числа?
5. Які є види радіоактивності?
6. Сформулюйте закон радіоактивного розпаду. Які є характеристики радіоактивного розпаду?

7. Які існують дози радіоактивного випромінювання?

8. Дайте визначення активності радіоактивної речовини. В яких фізичних одиницях вимірюють активність радіоактивної речовини?

9. Як отримується енергія під час поділу важких ядер. Який принцип дії атомної бомби?

10. Як отримують енергію під час синтезу легких ядер? Який принцип дії водневої (термоядерної) бомби?

11. Як отримують енергію в ядерних реакторах на атомних електростанціях?

Список використаної літератури

1. Лопатинський І. Є. Фізика : підручник / І. Є. Лопатинський, І. Р. Зачек, Г. А. Ільчук, Б. М. Романишин. – Львів : Афіша, 2009. – 386 с.

2. Кучерук І. М. Загальна фізика. Електрика і магнетизм: навч. посіб. / І. М. Кучерук, І. Т. Горбачук; за заг. ред. В. Й. Сугакова. – Київ : Вища шк., 1990. – 367 с.

3. Кучерук І. М. Загальна фізика. Оптика. Квантова фізика : навч. посіб. / І. М. Кучерук, В. П. Дущенко. – Київ : Вища шк., 1991. – 463 с.

4. Feynman R. The Feynman Lectures on Physics. Electrodynamics / R. Feynman, R. Leighton, M. Sands. – Vol. 2. – Addison-Wesley Publishing Company, INC., Reading, Massachusetts, Palo Alto, London, 1964.

5. Feynman R. The Feynman Lectures on Physics. Electricity and Magnetism / R. Feynman, R. Leighton, M. Sands. – Vol. 2. – Addison-Wesley Publishing Company, INC., Reading, Massachusetts, Palo Alto, London, 1964.

6. Feynman R. The Feynman Lectures on Physics. Continuum Physics / R. Feynman, R. Leighton, M. Sands. – Vol. 2. – Addison-Wesley Publishing Company, INC., Reading, Massachusetts, Palo Alto, London, 1964.

7. Воловик П. М. Фізика для університетів / П. М. Воловик. – Київ; Ірпінь : Перун, 2005. – 864 с.

8. Purcell E. M. Electricity and Magnetism / E. M. Purcell. – 2nd ed. – McGraw-Hill, New York, 1985. – 416 p.

9. Осадчук В. С. Волоконно-оптичні системи передачі : навч. посіб. / В. С. Осадчук, О. В. Осадчук. – Вінниця : ВНТУ, 2005. – 225 с.

10. Герасимов С. М. Физические основы электронной техники: учеб. пособие для вузов / С. М. Герасимов, М. В. Белоус, В. А. Москалюк. – Киев : Вища шк., 1981. – 386 с.

11. Калашніков С. Г. Електрика / С. Г. Калашніков. – Пер. з 2-го рос. вид. – Київ : Рад. шк., 1964. – 630 с.

12. Лукіянець Б. А. Планетарна модель атома. Квантування Бора–Вільсона–Зоммерфельда / Б. А. Лукіянець, Г. В. Понеділок, Ю. К. Рудавський // Світ фізики. – 2009. – № 4. – с. 3–14.

Додаток 1

Отримання формули (18.102) для коефіцієнта прозорості потенціальної ями

Розглянемо розв'язок системи рівнянь (18.98), (18.99):

$$1 + b_1 = a_2 + b_2, \tag{Д.1.1}$$

$$k_1 - k_1 b_1 = k_2 a_2 - k_2 b_2, \tag{Д.1.2}$$

$$a_2 e^{ik_2 d} + b_2 e^{-ik_2 d} = a_3 e^{ik_1 d}, \tag{Д.1.3}$$

$$a_2 e^{ik_2 d} - b_2 e^{-ik_2 d} = a_3 \frac{k_1}{k_2} e^{ik_1 d}. \tag{Д.1.4}$$

З рівняння (Д.1.1) виразимо b_1 через a_2 та b_2 й отримують

$$b_1 = a_2 + b_2 - 1.$$

Підставимо цей вираз для b_1 у рівняння (Д.1.2) й виразимо b_2 через a_2. Отримаємо

$$b_2 = \frac{k_1 + k_2}{k_2 - k_1} a_2 - \frac{2k_1}{k_2 - k_1}. \tag{Д.1.5}$$

Додамо ліві й праві частини рівнянь (Д.1.3) і (Д.1.4) й виразимо a_2 через a_3:

$$2a_2 e^{ik_2 d} = a_3 \frac{k_1 + k_2}{k_2} e^{ik_1 d},$$

звідки отримують, що

$$a_2 = a_3 \frac{k_1 + k_2}{2k_2} e^{ik_1 d} \cdot e^{-ik_2 d}. \tag{Д.1.6}$$

Підставимо (Д.1.5) і (Д.1.6) в рівняння (Д.1.3). Тоді отримують вираз:

$$a_3\frac{k_1+k_2}{2k_2}e^{ik_1d}+a_3\frac{k_1+k_2}{k_2-k_1}e^{-ik_2d}\cdot\frac{k_1+k_2}{2k_2}e^{ik_1d}\cdot e^{-ik_2d}-$$
$$-\frac{2k_1}{k_2-k_1}e^{-ik_2d}=a_3e^{ik_1d}.$$

З цього виразу отримують:

$$a_3\left(\frac{k_1+k_2}{2k_2}e^{ik_1d}-e^{ik_1d}+\frac{(k_1+k_2)^2}{2k_2(k_2-k_1)}e^{ik_1d}\cdot e^{-2ik_2d}\right)=\frac{2k_1}{k_2-k_1}e^{-ik_2d}.$$

Помножимо ліві й праві сторони цього виразу на $e^{-ik_1d}\cdot e^{ik_2d}$. Тоді отримують:

$$a_3\left(\frac{k_1+k_2}{2k_2}e^{ik_2d}-e^{ik_2d}+\frac{(k_1+k_2)^2}{2k_2(k_2-k_1)}\cdot e^{-ik_2d}\right)=\frac{2k_1}{k_2-k_1}e^{-ik_1d},$$

звідки після нескладних перетворень отримують:

$$a_3\left(\frac{k_1-k_2}{2k_2}e^{ik_2d}+\frac{(k_1+k_2)^2}{2k_2(k_2-k_1)}\cdot e^{-ik_2d}\right)=\frac{2k_1}{k_2-k_1}e^{-ik_1d},$$

або

$$a_3\big((k_1+k_2)^2\cdot e^{-ik_2d}-(k_2-k_1)^2e^{ik_2d}\big)=4k_1k_2e^{-ik_1d}. \quad (Д.1.7)$$

Із формули (Д.1.7) отримують вираз (18.100) для a_3:

$$a_3=\frac{4k_1k_2e^{ik_1d}}{(k_1+k_2)^2\,e^{-ik_2d}-(k_1-k_2)^2e^{ik_2d}}. \quad (Д.1.8)$$

Тоді коефіцієнт прозорості ями знайдемо з виразу

$$D=a_3a_3^*=\frac{4k_1k_2e^{ik_1d}}{(k_1+k_2)^2\,e^{-ik_2d}-(k_1-k_2)^2e^{ik_2d}}\times$$
$$\times\frac{4k_1k_2e^{-ik_1d}}{(k_1+k_2)^2\,e^{ik_2d}-(k_1-k_2)^2e^{-ik_2d}}=$$
$$=\frac{16{k_1}^2{k_2}^2}{(k_1+k_2)^4-(k_1+k_2)^2(k_1-k_2)^2e^{-i2k_2d}-(k_1+k_2)^2(k_1-k_2)^2e^{i2k_2d}+(k_1-k_2)^4}. \quad (Д.1.9)$$

Враховують такі формули

$$e^{-i2k_2d}=\cos 2k_2d-i\sin 2k_2d,$$
$$e^{i2k_2d}=\cos 2k_2d+i\sin 2k_2d.$$

Тоді вираз (Д.1.9) записують так:

$$D=\frac{16{k_1}^2{k_2}^2}{(k_1+k_2)^4+(k_1-k_2)^4-2(k_1+k_2)^2(k_1-k_2)^2\cdot\cos 2k_2d},$$

або

$$D = \frac{16{k_1}^2{k_2}^2}{(k_1+k_2)^4+(k_1-k_2)^4-2(k_1+k_2)^2(k_1-k_2)^2\cdot(\cos^2 k_2 d-\sin^2 k_2 d)} =$$

$$= \frac{16{k_1}^2{k_2}^2}{(k_1+k_2)^4+(k_1-k_2)^4-2(k_1+k_2)^2(k_1-k_2)^2\cdot(1-2\sin^2 k_2 d)} =$$

$$= \frac{16{k_1}^2{k_2}^2}{(k_1+k_2)^4+(k_1-k_2)^4-2(k_1+k_2)^2(k_1-k_2)^2+4(k_1+k_2)^2(k_1-k_2)^2\sin^2 k_2 d)} =$$

$$= \frac{16{k_1}^2{k_2}^2}{((k_1+k_2)^2-(k_1-k_2)^2)^2+4(k_1+k_2)^2(k_1-k_2)^2\sin^2 k_2 d)} =$$

$$= \frac{16{k_1}^2{k_2}^2}{16{k_1}^2{k_2}^2+4(k_1+k_2)^2(k_1-k_2)^2\sin^2 k_2 d)}. \qquad \text{(Д.1.10)}$$

Поділимо чисельник і знаменник виразу (Д.1.10) на $16{k_1}^2{k_2}^2$, й після нескладних перетворень отримують формулу (18.102):

$$D = \frac{1}{1+\frac{1}{4}\left(\frac{k_2}{k_1}+1\right)^2\left(\frac{k_1}{k_2}-1\right)^2\sin^2 k_2 d}.$$

Предметний покажчик

А

Абсолютне чорне тіло 391
Автоколивання 232
Активність елемента 682
Акцептори 556
Аналізатори 352
Анізотропія штучна 339, 354
Антиферомагнетик 162
Аперіодичний процес 220
Апертура
- інтерференції 297
- світловода 662
Атом водню 311
Атомна одиниця маси 671

Б

Бар'єр
- висота 487
- потенціальний 485
- Шотткі 632
Батарея сонячна 630
Бер 685
Биття 213
Бідзеркала Френепя 298
Білінза Бі'є 288
Біпризма Френеля 288
Боровський радіус 435

В

Варикап 624
Вектор
- електричного зміщення (індукції) 47
- намагнічування 136
- Пойтинга 268
- поляризації 45
Вестона елемент 87 ─ 90
Взаємна енергія струмів 186
Видність 286
Випромінювальна здатність 390
Випромінювання
- вимушене 637, 641
- рентгенівське 455
- елементарного диполя 273 - 278
- рівноважне 388
- спонтанне 639
- теплове 386
- характеристичне 457
Висота звуку 190
Вітер електричний 64
Власна
- енергія струму 182
- частота 156
Власний напівпровідник 552
Внутрішній фотоефект 560
Вузол електричного кола 90

Г

Гальванічний елемент 88
Генератор квантовий 637
- оптичний 461,463
Герца досліди 270

Гетероперехід 635
Гіпотеза Планка 398
Гіромагнітне відношення 138
Гістерезис 54
- діелектричний 56
- магнітний 152
Головна площина 348
Головний максимум 328
Градани 668
Група хвиль 245
Групова швидкість 184,273
Густина
- енергії випромінювання 283
- об'ємна 242
- спектральна 389
- потоку енергії 244

Д

Дефект маси 676
Децибел 255
Диполь
- електричний 39
- магнітний 98, 109
- елементаргий 273
Дисперсії класична теорія 365
Дисперсійна область 330
Дисперсія..362
- аномальна 364
- відносна 365
- кутова 329
- лінійна 330
- міжмодова 664
- нормальна 368
- оптичних волокон 664
- речовини 364
- світла 362
- середня 364
- хроматична 665
Дифракція 297
- рентгенівського випромінювання 333
- Фраунгофера 322
- Френеля 305, 307, 309
Дифракційна ґратка 327
Дифузія 572
- носіїв заряду 572
Дифузійна
- довжина 577
- швидкість 577
Дихроїзм 354
Діаграма випромінювання диполя 277
Діамагнетизм 144
Діелектрики 37
- іонні 42
- неполярні 42
- полярні 43
Дзеркала Френеля 288
Дірка (носій заряду) 556
Ділянка
- кола неоднорідна 85
- однорідна 86
Добротність контура 219
Довгохвильова (червона) межа фотоефекту 411, 563

Довжина
- вільного пробігу 573
- когерентності 294
Додавання
- взаємно перпендикулярних коливань 213
- однаково напрямлених коливань 210, 213
Доза
- іонізувального випромінювання 684
- біологічна 685
- еквівалентна 684
- поглинута 684
- експозиційна 684
Домен 42
Домени
- електричні 51
- магнітні 160
Домішковий напівпровідник 556
Донори 556
Досліди
- Герца 261
- Девіссона і Джермера 464
- Квінке 492
- Резерфорда 422
- Франка і Герца 432
- Штерна і Герлаха 445
Дрейф носіїв заряду 569
Дуалізм корпускулярно-хвильовий 461–463

Е

Еквіпотенціальні поверхні 31
Екстраструм 180
Електрична стала 16
Електрети 58
Електричний
- вітер 63
- диполь 39
- заряд 13
Електризація через вплив 59
Електроємність відокремленого провідника 66
Електрон 14
Електронна оболонка 452
Електронний газ
- вироджений 536
- невироджений 537
Електрорушійна сила 85
Елемент струму 96
Елементарна комірка 508
Електропровідність
- металів 536 … 546
- питома 82
- темнова 562, 563
Емісія
- автоелектронна 496
- термоелектронна 586
- холодна 496
Емітер 620
Енергетична світність 390
Енергетичні зони 526
Енергія
- активації 555

- взаємна 186
- власна струму 183
- електричного поля 74
- зарядженого провідника 74
- зв'язку ядра 675
- конденсатора 75
- магнітного поля 184
- «нульова» 533
- пружної хвилі 242
- струму 182
- Фермі 552

Ефект
- Ганна 577
- Доплера 257, 384
- Зеємана 443
- Керра 356
- Комптона 418
- Коттона–Мутона 357
- п'єзоелектричний 56
- Поккельса 358
- Рамзауера 498
- тунельний 495
- Шотткі 592
- Фарадея 361
- Фредерікса 519
- Холла 124

Ефективна
- густина станів 551
- маса 532

Є

Ємність: 67
- конденсатора:
- плоского 70
- сферичного 72
- циліндричного 71
- кулі 68
- *pn*-переходу 622

З

З'єднання конденсаторів
- паралельне 73
- послідовне 73

Закон
- Ампера 98
- Біо–Савара–Лапласа 100
- Богуславського–Ленгмюра 591
- Брюстера 342
- Бугера–Ламберта–Бера 380
- Джоуля–Ленца 84
- діючих мас 552
- збереження електричних зарядів 14
- зміщення Віна 396
- Кірхгофа 392
- Кулона 15
- Кюрі 153
- Ленца 170
- Малюса 339
- Мозлі 452
- Ома в диференціальній формі для неоднорідної ділянки кола 87
- Ома для однорідної ділянки кола 83

- Ома для неоднорідної ділянки кола 87
- Ома для повного струму 139
- термоелектроннної емісії 592, 593
- радіоактивного розпаду 681
- Релея 382
- Річардсона–Дешмана 593
- Стефана–Больцмана 394
- Фарадея 170
- фотоефекту 406

Заселеність
- енергетичних рівнів 638
- інверсна 643

Захист електростатичний 64

Звук 253:
- гучність 255
- інтенсивність 254
- тиск 254
- тональний 254
- швидкість 257

Зв'язок
- гомеополярний (ковалентний) 512
- іонний 511
- металічний 513

Здатність
- випромінювальна 382
- поглинальна 383

Зіверт 684

Зміщення
- електричне 46
- зворотне 616
- пряме 618

Зона
- валентна 527, 528
- заборонена 527, 528
- провідності 527, 528

Зони
- енергетичні 524 … 530
- Френеля 308

Зонна пластинка 313

І

Ізобари 672

Ізомери 672

Ізотони 672

Ізотопи 672

Імпеданс 230

Інверсна населеність 644

Інверсний шар 596, 597

Індуктивність 178
- взаємна 181

Індукція
- електромагнітна 166
- електростатична 46
- магнітного поля 99

Інтенсивність
- звуку 254
- світла 268, 280
- хвилі 243, 268

Інтерференційний максимум, мінімум 249, 284

Інтерференція

- Жамена 301
- Майкельсона 302
- світла 293
- хвиль 247
Іонний зв'язок 509

К

Кандела 629
Квазімонохроматичне світло 285
Квазістаціонарні процеси 197
Квантування Бора–Вільсона–Зоммерфельда 438
Квант 400
Квантові числа 451
Квантовий
- генератор 644
- підсилювач 644
Кільця Ньютона 292
Класична теорія електропровідності металів 534
Ковалентний зв'язок 510
Когерентність 248
- часова 294
- просторова 296
Коефіцієнт
- дисперсії 365
- дифузії 574
- екстинкції 382
- загасання 216
- питомого обертання 359
- поглинання 380, 639
- потужності 232
Коливання 204
- вимушені 221
- гармонічні (власні) 206, 208-228
- електромагнітні 208
- загасаючі 216
Кольори тонких плівок 291
Конденсатори 70:
- з'єднання паралельне 72
- послідовне 72
Контакт двох металів 595
Контактна різниця потенціалів 596
Контур
- ідеалізований 208
- коливальний 207
Крива намагнічування 149, 164
Кристали 508
Критична маса, розміри 691
Куперовська пара 548
Кут
- Брюстера 342
- повороту площини поляризації 359, 360
Кюрі температура (точка) 153

Л

Лазер 644
Лазери
- газові 655
- на неодимовому склі 653
- напівпровідникові 656–658
- рубіновий 653

Ларморова прецесія 144, 375
Лінії
- індукції магнітного поля 93
- напруженості електричного поля 18
Лінійчастий спектр поглинання 380
Логарифмічний декремент загасання 218

М

Магічні числа 676
Магнетон Бора 137
Магнетрон 126
Магнітострикція 160
Магнітна напруженість поля 132
Магічні числа 679
Мазер 637
Маса фотона 411
Масове число 672
Мезони 674
Метали 511
Метод зон Френеля 308
Мікроінтерферометр Майкельсона 303
Моди у світловоді 663
Модель
- ядра краплинна 678
- оболонкова 679
Момент
- диполя електричного 38
- магнітного 101
- механічний власний 136, 443
- орбітальний магнітний 137, 443
Монохроматична хвиля 249, 293

Н

Надпровідність 546
Намагніченість 149
Намагнічування спонтанне 163
Напруженість
- електричного поля 17
- магнітного поля 132
- поблизу зарядженого провідника 62
Напруженість
- магнітного поля 139
- пробою 617
Напівпровідники 527, 549:
- власний 553
- домішковий 554
Напруга діюча 232
Напрямок електричного струму 70
Населеність інверсна 644
Нейтрон 671
Носії
- заряду неосновні 557
- нерівноважні 562
- основні 557
Нуклон 671

О

Об'ємна густина енергії
- електричного поля 74
- магнітного поля 184
- теплового випромінювання 389

Опір
- від'ємний 235
- ємнісний 231
- індуктивний 231
- критичний 221

Оптична:
-активність 644
- вісь 348
-густина 271
-довжина шляху 284
- накачка 653
- пірометрія 402
- різниця ходу 284

П

Парамагнетизм 148

Перехід електронно-дірковий 605 … 616

Період
- дифракційної ґратки 327
- півроспаду 682

Періодична система хімічних елементів 449

Петля гістерезису 151

П'єзоелектрики 57

Питома
- енергія зв'язку 677
-теплова потужність 84

Питомий заряд електрона 14, 131

Площина поляризації хвилі 338

Повне внутрішнє відбиття 660

Поверхні еквіпотенціальні 29

Поглинальна здатність 391

Поглинання світла 379

Подвійне променезаломлення 346

Поле електричне:
- в діелектрику 47
- у провіднику 60
- вихрове 175
- зарядженого циліндра 26
- зарядженої кулі 27
- двох площин заряджених різнойменно 26
- лінійного заряду 26
- однорідне 20
- площини 25
- потенціальне 31

Поле магнітне 93

Поляризатори 352, 353

Поляризаційні призми 351

Поляризація
- діелектриків 41
- електронна 42
- залишкова 55
- іонна 44
- орієнтаційна 43
- світла 337
- спонтанна 54

Поляроїди 354

Постулати
- Бора 431
- Ейнштейна 411, 637
Потік
- енергії хвилі 243
- вектора електричного зміщення 47
- магнітної індукції 111
- напруженості електричного поля 20
Потокозчеплення 179
Потужність дози випромінювання 685
Пояснення ефекту Фарадея 375
Правила Кірхгофа 91
Правило
- «лівої руки» 101
- «правої руки» 101
Принцип
- відповідності 413
- Гюйгенса–Френеля 306
- доповнюваності 467
- невизначеності 472
- Паулі 452
- причинності 481
- суперпозиції:
- електричних полів 16
- магнітних полів 100
- хвиль 245
Провідник відокремлений 66
Провідники в електричному полі 60
Провідність
- власна 548
- діркова 551
- домішкова 552, 553
- електронна 553
- напівпровідників 548
- питома 76
Промені рентгенівські 332, 450
Промінь
- звичайний 347
- незвичайний 347
Проникність
- відносна діелектрична 47
- магнітна 140
Просвітлення оптики 301
Просторова когерентність 295
Просторове квантування 443
Протон 671
Пуассона рівняння 31

Р

Радіоактивність 680
Радіус Бора 435
Реактанс 231
Реакція
- поділу ланцюгова 690, 692
- термоядерна 695, 696
- ядерна 688
- ядерного поділу 688
- синтезу 694
Резонанс 224
- напруги, струму 225
- параметричний 228
Резонатор оптичний 647

Рентген (Р) 684
Рівень:
- гучності 255
- Фермі в напівпровідниках 551, 559
Рівні
- акцепторні 556
- донорні 556
Рівняння
- Ейнштейна для фотоефекту 410
- Максвелла 193, 198, 199
- плоскої хвилі 229
- Пуасонна 34
- хвильове 241
- Шредінгера 480
Рідкі кристали 515

Різниця
- потенціалів 30
- ходу хвиль геометрична 249, 284
Робота
- виходу електронів 587
- сил електростатичного поля 29
- сторонніх сил над зарядами 79
Роздільна здатність 331
Розміри критичні 691
Розпад
- альфа 680
- бета 680
- радіоактивний 680
Розподіл Больцмана 534
Розподіл поля в бар'єрі 600, 614
Рухливіть носіїв заряду 82

С

Самоіндукція 178
Світло квазімонохроматичне 293
Світловолокно
- градієнтне 661
- ступеневе 661
Світлодіоди 628

Світло поляризоване:
- еліптично 320 … 330
- лінійно 320 … 330
- по колу 320 … 330
- плоско (лінійно) 330
- частково 329
Світло природне 336
Сегнетоелектрики 52
Селективність 626
Середній час життя 681
Середовище
- активне 644
- пружне 236
Селфоки 667
Сила
- Ван дер Ваальса 517
- затримуюча 152
- коерцитивна 55
- консервативна 29
- Лоренца 173

Сили
- сторонні 84
- ядерні 673
Силова лінія 19
Скін-ефект 189
Смуги
- однакового нахилу 290
- однакової товщини 292
Соленоїд 105, 112
Сонячні батареї 630
Співвідношення
- Ейнштейна 574
- невизначеностей Гайзенберга 472
Спін електрона 137, 448
Спіраль Корню 320
Сприйнятливість
- діелектрична 44
- магнітна 141
Стала
- Верде 378
- Керра 356
- радіоактивного розпаду 681
- Рідберга 423
- Холла 124
Стан
- системи нерівноважний 643
- рівноважний 641
Статистика
- класична 534
- Фермі–Дірака 532
Струм
- вихровий 187
- генераційний 617
- діючий 232
- дифузійний 77, 572
- ефективний 232
- зворотний 617
- змінний 230
- зміщення 194
- індукційний 180
- конвекційний 77
- насичення 593
- повний 198
- постійний 78
- провідності 78
- термоелектронний 592, 593
Ступінь поляризації 338

Т

Тверде тіло 505–508
Температура:
- Дебая 545
- кольорова 405
- критична 546
- Кюрі 1154
- радіаційна 402
- яскравісна 403
Температурний коефіцієнт опору 81
Теорема
- єдиності в електростатиці 65
- Лармора 138

- Остроградського-Гаусса
для електростатичного поля 21–28,
для магнітного поля 112
Тиск світла 415
Тіло сіре 391
Тороїда магнітне поле 114
Тунельний ефект 494

У

Умови максимумів та інімумів дифракційної ґратки 328

Ф

Фактор дисперсії 378
Фарада 62
Ферити 157, 162
Феромагнетизм 150
Фон 256
Фонони 543
Формула
- Бальмера 429
- Больцмана 530
- Віна 396
- Вульфа–Брега 334
- де Бройля 462
- Лоренца 122
- Планка 401
- Релея–Джінса 397
-Річардсона–Дешмана 593
- Томсона 209
Фотодіоди 626
- *APD* 667
- *p–i–n* 666
Фотоефект
- вентильний 561
- внутрішній 560
- зовнішній 408
Фотон 411
Фронт хвилі 239
Функції власні 481
Функція:
- густини станів 549, 531
- Фермі–Дірака 532
- хвильова 478

Х

Характеристика спектральна 562, 621
Хвилі
- де Бройля 462, 464
- електромагнітні 261
- довжина 238
- звукові 253
- когерентні 248
- некогерентні 248
Хвильова поверхня 239
Хвильовий
- опір 210
- пакет 470
Холодна емісія 496

Ц

Циркулярція
- вектора індукції магнітного поля 102

- напруженості електростатичного поля 32
- напруженості магнітного поля 139
Цуг хвильовий 271

Ч

Час:
- життя нерівноважних неосновних носіїв заряду 568
- середній 681
- когерентності 294
- релаксації 218
- ядерного прольоту 687
Частинки віртуальні 675
Частота
- власна 208
- резонансна 225
- хвилі 238
Числа квантові 451
Число
- зарядове 672
- масове 672
- хвильове 238
Червона (довгохвильова) межа 563

Ш

Шар
- збіднений 613
- інверсний 602
Швидкість
- групова 244, 373
- дрейфова 78, 570
- звуку 244
- середня 569
- світла 265
- фазова 240
- хвилі 240
Ширина
- інтерференційної смуги 285
- *p*–*n*-переходу 614
Шкала електромагнітних хвиль 261

Я

Явище електромагнітної індукції 157
Ядерна модель атома 424
Ядерний реактор 692
Ядро атома 672, 673
Яма потенціальна 483, 502
Яскравість 651

Н а в ч а л ь н е в и д а н н я

Скіцько Іван Федорович
Скіцько Олексій Іванович

Фізика для інженерів

Підручник

2-ге видання, перероблене і доповнене

Коректура *Наталії Мурашової*

Надруковано з оригінал-макета замовника

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Свідоцтво про державну реєстрацію: серія ДК № 5354 від 25.05.2017 р.
просп. Берестейський, 37, м. Київ, Україна, 03056

Формат $60\times84^{1}/_{16}$. Гарнітура Times.
1 файл (21,3 Мб). Обл.-вид. арк. 35,75. Поз. 25-1-1-010.

Видавництво «Політехніка» КПІ ім. Ігоря Сікорського
вул. Політехнічна, 14, корп. 15,
м. Київ, Україна, 03056
тел. (044) 204-81-7

715